

(12) STANDARD PATENT
(19) AUSTRALIAN PATENT OFFICE

(11) Application No. **AU 2010295739 B2**

- (54) Title
A system for transformation of the chloroplast genome of Scenedesmus sp. and Dunaliella sp.
- (51) International Patent Classification(s)
A01H 13/00 (2006.01) **C12N 15/05** (2006.01)
C12N 1/13 (2006.01)
- (21) Application No: **2010295739** (22) Date of Filing: **2010.09.14**
- (87) WIPO No: **WO11/034863**
- (30) Priority Data
- (31) Number (32) Date (33) Country
61/242,735 **2009.09.15** **US**
- (43) Publication Date: **2011.03.24**
(44) Accepted Journal Date: **2016.02.25**
- (71) Applicant(s)
Sapphire Energy, Inc.
- (72) Inventor(s)
Botsch, Kyle;Szyjka, Shawn;Levine, Wendy;Curran, Amy;O'Neill, Bryan;Mendez, Michael
- (74) Agent / Attorney
Freehills Patent Attorneys, Level 43 101 Collins Street, Melbourne, VIC, 3000
- (56) Related Art
WO 2010/104763 A1 (SAPPHIRE ENERGY, INC.) 16 September 2010
WO 2009/045550 A2 (SAPPHIRE ENERGY) 09 April 2009
WO 2009/036067 A2 (SAPPHIRE ENERGY & THE SCRIPPS RESEARCH INSTITUTE) 19 March 2009
WO 2010/111707 A1 (SAPPHIRE ENERGY, INC.) 30 September 2010
WO 2010/051489 A1 (SAPPHIRE ENERGY, INC.) 06 May 2010

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

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
24 March 2011 (24.03.2011)

PCT

(10) International Publication Number
WO 2011/034863 A1

(51) International Patent Classification:

A01H 13/00 (2006.01) C12N 15/05 (2006.01)
C12N 1/13 (2006.01)

(21) International Application Number:

PCT/US2010/048828

(22) International Filing Date:

14 September 2010 (14.09.2010)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

61/242,735 15 September 2009 (15.09.2009) US

(71) Applicant (for all designated States except US): **SAPPHIRE ENERGY, INC.** [US/US]; Legal Department, 3115 Merryfield Row, San Diego, California 92121 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **BOTSCH, Kyle** [US/US]; Sapphire Energy, INC., 3115 Merryfield Row, San Diego, California 92121 (US). **SZYJKA, Shawn** [US/US]; Sapphire Energy, INC., 3115 Merryfield Row, San Diego, California 92121 (US). **LEVINE, Wendy** [US/US]; Sapphire Energy, INC., 3115 Merryfield Row, San Diego, California 92121 (US). **CURRAN, Amy** [US/

US]; Sapphire Energy, INC., 3115 Merryfield Row, San Diego, California 92121 (US). **O'NEILL, Bryan** [US/US]; Sapphire Energy, INC., 3115 Merryfield Row, San Diego, California 92121 (US). **MENDEZ, Michael** [US/US]; Sapphire Energy, INC., 3115 Merryfield Row, San Diego, California 92121 (US).

(74) Common Representative: **SAPPHIRE ENERGY, INC.**; Legal Department, 3115 Merryfield Row, San Diego, California 92121 (US).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ,

[Continued on next page]

(54) Title: A SYSTEM FOR TRANSFORMATION OF THE CHLOROPLAST GENOME OF *SCENEDESMUS SP.* AND *DUNALIELLA SP.*

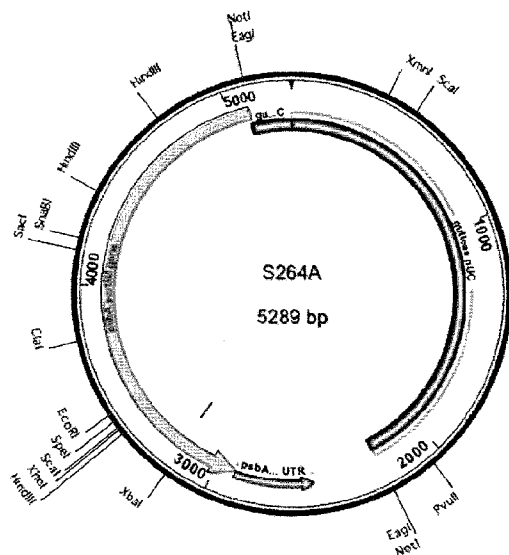


FIG. 1

(57) Abstract: The present disclosure relates to methods of transforming various species of algae, for example, algae from the genus *Scenedesmus* and the genus *Dunaliella*, vectors and nucleic acid constructs useful in conducting such transformations, and recombinant algae, for example, *Scenedesmus* and *Dunaliella* produced using the vectors and methods disclosed herein.



TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

— *as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii))*

Published:

— *with international search report (Art. 21(3))*

— *with sequence listing part of description (Rule 5.2(a))*

Declarations under Rule 4.17:

— *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*

**A SYSTEM FOR TRANSFORMATION OF THE CHLOROPLAST
GENOME OF *SCENEDESMUS SP.* AND *DUNALIELLA SP.***

CROSS REFERENCE TO RELATED APPLICATION

[0001] This application claims the benefit of United States Provisional Application Number 61/242,735, filed September 15, 2009, the entire contents of which are incorporated by reference for all purposes.

INCORPORATION BY REFERENCE

[0002] All publications, patents, patent applications, public databases, public database entries, and other references cited in this application are herein incorporated by reference in their entirety as if each individual publication, patent, patent application, public database, public database entry, or other reference was specifically and individually indicated to be incorporated by reference.

BACKGROUND

[0003] Algae are unicellular organisms, producing oxygen by photosynthesis. One group, the microalgae, are useful for biotechnology applications for many reasons, including their high growth rate and tolerance to varying environmental conditions. The use of microalgae in a variety of industrial processes for commercially important products is known and/or has been suggested. For example, microalgae have uses in the production of nutritional supplements, pharmaceuticals, natural dyes, a food source for fish and crustaceans, biological control of agricultural pests, production of oxygen and removal of nitrogen, phosphorus and toxic substances in sewage treatment, and pollution controls, such as biodegradation of plastics or uptake of carbon dioxide.

[0004] Microalgae, like other organisms, contain lipids and fatty acids as membrane components, storage products, metabolites and sources of energy. Some algal strains, diatoms, and cyanobacteria have been found to contain proportionally high levels of lipids (over 30%). Microalgal strains with high oil or lipid content are of great interest in the search for a sustainable feedstock for the production of biofuels.

[0005] Some wild-type algae are suitable for use in various industrial applications. However, it is recognized that by modification of algae to improve particular characteristics useful for the aforementioned applications, the relevant processes are more likely to be commercially viable. To this end, algal strains can be developed which have improved characteristics over wild-type strains. Such developments have been made by traditional techniques of screening and mutation and selection. Further, recombinant DNA technologies have been widely suggested for algae. Such approaches may increase the economic validity of production of commercially valuable products.

[0006] One area in which algae have received increasing attention is the production of fuel products. Fuel products, such as oil, petrochemicals, and other substances useful for the production of petrochemicals are increasingly in demand. Much of today's fuel products are generated from fossil fuels, which are not considered renewable energy sources, as they are the result of organic material being covered by successive layers of sediment over the course of millions of years. There is also a growing desire to lessen dependence on imported crude oil. Public awareness regarding pollution and environmental hazards has also increased. As a result, there has been a growing interest and need for alternative methods to produce fuel products. Thus, there exists a pressing need for alternative methods to develop fuel products that are renewable, sustainable, and less harmful to the environment. One potential source of

alternative production of fuel and fuel precursors is genetically modified organisms, such as bacteria and plants, including algae. To date, algae have yet to be successfully developed as a commercially viable platform for biofuel production, due mainly to the high cost of harvesting and processing of algae for recovery of the biofuel. Thus, a need exists to develop host organisms such as algae (for example, *Scenedesmus* sp., *Chlamydomonas* sp., and *Dunaliella* sp.) and bacteria for which such costs are reduced. One way of genetically modifying an organism is to transform the organism with a nucleic acid that encodes for a protein, wherein expression of the protein results, for example, in the increased production of a product, or in the production of a product that the organism does not usually make.

SUMMARY

[0007] 1. An isolated *Scenedesmus* sp. comprising a chloroplast genome that has been transformed with an exogenous nucleotide sequence, wherein the exogenous nucleotide sequence comprises a nucleic acid sequence encoding at least one protein. 2. The isolated *Scenedesmus* sp. of claim 1, wherein the protein is involved in the isoprenoid biosynthesis pathway. 3. The isolated *Scenedesmus* sp. of claim 2, wherein the protein is a synthase. 4. The isolated *Scenedesmus* sp. of claim 3, wherein the synthase is a farnesyl-diphosphate (FPP) synthase. 5. The isolated *Scenedesmus* sp. of claim 4, wherein the FPP synthase is from *G. gallus*. 6. The isolated *Scenedesmus* sp. of claim 3, wherein the synthase is a fusicoccadiene synthase. 7. The isolated *Scenedesmus* sp. of claim 6, wherein the fusicoccadiene synthase is from *P. amygdali*. 8. The isolated *Scenedesmus* sp. of claim 3, wherein the synthase is a bisabolene synthase. 9. The isolated *Scenedesmus* sp. of claim 8, wherein the bisabolene synthase is from *A. grandis*. 10. The isolated *Scenedesmus* sp. of claim 1, wherein the exogenous nucleotide sequence is at least 0.5 kb, at least 1.0 kb, at least 2 kb, at least 3 kb, at least 5 kb, at least 8 kb, at least 11 kb, or at least 19 kb. 11. The isolated *Scenedesmus* sp. of claim 1, wherein the nucleic acid sequence encodes for two proteins, three proteins, or four proteins. 12. The isolated *Scenedesmus* sp. of claim 1, wherein the exogenous nucleotide further comprises a second nucleic acid sequence encoding a selectable marker. 13. The isolated *Scenedesmus* sp. of claim 12, wherein the marker is chloramphenicol acetyltransferase (CAT), erythromycin esterase, or cytosine deaminase. 14. The isolated *Scenedesmus* sp. of claim 1, wherein the *Scenedesmus* sp. is *S. dimorphus*. 15. The isolated *Scenedesmus* sp. of claim 1, wherein the *Scenedesmus* sp. is *S. obliquus*. 16. The isolated *Scenedesmus* sp. of claim 1, wherein the nucleic acid sequence encodes for a biomass-degrading enzyme. 17. The isolated *Scenedesmus* sp. of claim 16, wherein the biomass-degrading enzyme is a galactanase, a xylanase, a protease, a carbohydrase, a lipase, a reductase, an oxidase, a transglutaminase, or a phytase. 18. The isolated *Scenedesmus* sp. of claim 16, wherein the biomass degrading enzyme is an endoxylanase, an exo- β -glucanase, an endo- β -glucanase, a β -glucosidase, an endoxylanase, or a lignase. 19. The isolated *Scenedesmus* sp. of claim 1, wherein the nucleic acid sequence encodes for an esterase. 20. The isolated *Scenedesmus* sp. of claim 19, wherein the esterase is an erythromycin esterase. 21. The isolated *Scenedesmus* sp. of claim 1, wherein the nucleic acid sequence encodes for a deaminase. 22. The isolated *Scenedesmus* sp. of claim 1, wherein the nucleic acid sequence encodes for a betaine aldehyde dehydrogenase. 23. The isolated *Scenedesmus* sp. of any of claims 1 to 22, wherein the nucleic acid sequence is codon optimized for expression in the chloroplast genome of the *Scenedesmus* sp. 24. An isolated *Scenedesmus* sp. comprising a chloroplast genome transformed with an exogenous nucleotide sequence wherein the transformed *Scenedesmus* sp. has an isoprenoid content that is different than an untransformed *Scenedesmus* sp. that is the same species as the isolated *Scenedesmus* sp., and wherein the exogenous nucleotide sequence comprises a nucleic acid encoding for an enzyme involved in isoprenoid biosynthesis. 25. The isolated *Scenedesmus* sp. of claim 24, wherein the

nucleic acid does not encode for an ent-kaurene synthase. 26. The isolated *Scenedesmus sp.* of claim 24, wherein the nucleic acid is codon optimized for expression in the chloroplast genome of the *Scenedesmus sp.*

[0008] 27. An isolated *Scenedesmus sp.* comprising a chloroplast genome transformed with an exogenous nucleotide sequence wherein the transformed *Scenedesmus sp.* has an increased accumulation of fatty acid based lipids and/or a change in the types of lipids, as compared to an untransformed *Scenedesmus sp.* that is the same species as the isolated *Scenedesmus sp.*, and wherein the exogenous nucleotide comprises a nucleic acid sequence encoding for an enzyme involved in fatty acid synthesis. 28. The isolated *Scenedesmus sp.* of claim 27, wherein the nucleic acid is codon optimized for expression in the chloroplast genome of the *Scenedesmus sp.*

[0009] 29. A method of transforming a chloroplast genome of a *Scenedesmus sp.* with a vector, wherein the vector comprises: i) a first nucleotide sequence of a *Scenedesmus sp.* chloroplast genome; ii) a second nucleotide sequence of a *Scenedesmus sp.* chloroplast genome; iii) a third nucleotide sequence comprising an exogenous nucleotide sequence, wherein the exogenous nucleotide sequence comprises a nucleic acid encoding a protein of interest, wherein the third nucleotide sequence is located between the first and second nucleotide sequences, and wherein the vector is used to transform the chloroplast genome of the *Scenedesmus sp.*; and iv) a promoter configured for expression of the protein of interest. 30. The method of claim 29, wherein the third nucleotide sequence further comprises a second nucleic acid sequence encoding a second protein of interest. 31. The method of claim 29, wherein the promoter is a psbD or a tufA promoter. 32. The method of claim 29, wherein the *Scenedesmus sp.* is *S. dimorphus*. 33. The method of claim 29, wherein the *Scenedesmus sp.* is *S. obliquus*. 34. The method of claim 29, wherein the first nucleotide sequence is at least 500 bp, at least 1000 bp, or at least 1,500 bp in length, and the first nucleotide sequence is homologous to a first portion of the genome of the *Scenedesmus sp.*, and the second nucleotide sequence is at least 500 bp, at least 1000 bp, or at least 1,500 bp in length, and the second nucleotide sequence is homologous to a second portion of the genome of the *Scenedesmus sp.*. 35. The method of claim 29, wherein the third nucleotide sequence is at least 0.5 kb, at least 1.0 kb, at least 2 kb, at least 3kb, at least 5 kb, at least 8 kb, at least 11 kb, or at least 19 kb in size. 36. The method of claim 29, wherein the nucleic acid is codon optimized for expression in the chloroplast genome of the *Scenedesmus sp.* 37. A transformed chloroplast genome of a *Scenedesmus sp.*, transformed by the method of claim 29.

[0010] 38. A method of transforming a chloroplast genome of a *Scenedesmus sp.* with at least one exogenous nucleotide sequence, comprising: i) obtaining the exogenous nucleotide sequence, wherein the exogenous nucleotide sequence comprises a nucleic acid sequence encoding a protein; ii) binding the exogenous nucleotide sequence onto a particle; and iii) shooting the exogenous nucleotide sequence into the *Scenedesmus sp.* by particle bombardment, wherein the chloroplast genome is transformed with the exogenous nucleotide sequence. 39. The method of claim 38, wherein the exogenous nucleotide sequence is at least 0.5 kb, at least 1.0 kb, at least 2 kb, at least 3kb, at least 5 kb, at least 8 kb, at least 11 kb, or at least 19 kb in size. 40. The method of claim 38, wherein the nucleic acid is codon optimized for expression in the chloroplast genome of the *Scenedesmus sp.* 41. The method of claim 38, wherein the particle is a gold particle or a tungsten particle. 42. The method of claim 41, wherein the gold particle is about 550 nm to about 1000 nm in diameter. 43. The method of claim 38, wherein the particle bombardment is carried out by a biolistic device. 44. The method of claim 43, wherein the biolistic device has a helium pressure of about 300 psi to about 500 psi. 45. The method of claim 43, wherein the biolistic device has a helium pressure of at least 300psi, at least 350psi, at least 400psi, at least 425psi, at least 450psi, at least 500psi, or at least 500psi. 46. The method of claim 38, wherein the exogenous nucleotide sequence bound to the particle is shot at a distance of about 2 to about 4 cm from the

Scenedesmus sp. 47. The method of claim 43, wherein the biolistic device is a Helicos Gene Gun or an Accell Gene Gun. 48. The method of claim 38, wherein the nucleic acid encodes for a protein involved in isoprenoid biosynthesis. 49. The method of claim 38, wherein the nucleic acid encodes for a protein involved in fatty acid biosynthesis. 50. The method of claim 38, wherein the *Scenedesmus sp.* is *S. dimorphus*. 51. The method of claim 38, wherein the *Scenedesmus sp.* is *S. obliquus*. 52. A transformed chloroplast genome of a *Scenedesmus sp.*, transformed by the method of claim 38.

[0011] 53. A method for obtaining a region of a chloroplast genome of a green algae, wherein the region is useful in the transformation of the green algae, comprising: 1) obtaining genomic DNA of the green algae; 2) obtaining a degenerate forward primer, wherein the forward primer is directed towards a psbB gene of the green algae; 3) obtaining a degenerate reverse primer, wherein the reverse primer is directed towards a psbH gene of the green algae; and 4) using the primers of step 2) and step 3) to amplify the region of the chloroplast genome of the green algae, wherein the nucleotide sequence of the amplified region is obtained. 54. The method of claim 53, wherein the amplified region is amplified by PCR. 55. The method of claim 53, wherein the sequenced region is cloned into a vector. 56. The method of claim 53, wherein the degenerate forward is primer 4099 (SEQ ID NO: 129) or forward primer 4100 (SEQ ID NO: 130), and wherein the degenerate reverse primer is primer 4101 (SEQ ID NO: 131) or reverse primer 4102 (SEQ ID NO: 132). 57. The method of claim 53, wherein the forward primer is primer 4099 (SEQ ID NO: 129) and the reverse primer is primer 4102 (SEQ ID NO: 132). 58. The method of claim 53, wherein at least a portion of the sequence of the amplified region is known. 59. The method of claim 53, wherein the amplified region of the chloroplast genome is from *C. reinhardtii*, *C. vulgaris*, *S. obliquus*, or *P. purpurea*. 60. The method of claim 53, wherein the sequence of the amplified region is unknown. 61. The method of claim 53, wherein the amplified region of the chloroplast genome is: from *D. tertiolecta* and comprises the nucleic acid sequence of SEQ ID NO: 133; from a *Dunaliella* of unknown species comprising the nucleic acid sequence of SEQ ID NO: 134; from *N. abudans* and comprising the nucleic acid sequence of SEQ ID NO: 135; from *C. vulgaris* and comprising the nucleic acid sequence of SEQ ID NO: 136; or from *T. suecica* and comprising the nucleic acid sequence of SEQ ID NO: 137. 62. The method of claim 53, wherein the amplified region of the chloroplast genome comprises a nucleotide sequence encoding a gene cluster psbB-psbT-psbN-psbH. 63. The method of claim 53, wherein the amplified region of chloroplast genome comprises a nucleotide sequence encoding a gene cluster psbB-psbT. 64. The method of claim 63, wherein a nucleic acid encoding a gene is inserted between the nucleotide sequence encoding psbB and psbT. 65. The method of claim 53, wherein the amplified region of chloroplast genome comprises a nucleotide sequence encoding a gene cluster psbT-psbN. 66. The method of claim 65, wherein a nucleic acid encoding a gene is inserted between the nucleotide sequence encoding psbT and psbN. 67. The method of claim 53, wherein the amplified region of chloroplast genome comprises a nucleotide sequence encoding a gene cluster psbN-psbH. 68. The method of claim 67, wherein a nucleic acid encoding a gene is inserted between the nucleotide sequence encoding psbN and psbH. 69. The method of claim 53, wherein the amplified region of chloroplast genome comprises a nucleotide sequence encoding a gene cluster psbH-psbK. 70. The method of claim 69, wherein a nucleic acid encoding a gene is inserted between the nucleotide sequence encoding psbH and psbK. 71. The method of claim 53, wherein the amplified region of the chloroplast genome comprises a nucleotide sequence encoding a region 3' of psbK. 72. The method of claim 53, wherein the sequence is a nucleic acid sequence. 73. The method of claim 53, wherein the sequence is an amino acid sequence.

[0012] 74. A region of a chloroplast genome of a green algae, obtained by the method of: 1) obtaining genomic DNA of the green algae; 2) obtaining a degenerate forward primer, wherein the forward primer is directed towards a psbB gene of the green algae; 3) obtaining a degenerate reverse primer, wherein the reverse primer is directed towards a psbH gene of the green algae; and 4) using the primers of step 2) and step 3) to amplify the region of the chloroplast genome of the green algae, wherein the amplified region is sequenced and comprises a nucleotide sequence, and wherein the nucleotide sequence is modified to comprise a nucleic acid sequence encoding for at least one protein.

[0013] 75. A vector useful in the transformation of the chloroplast genome of *Scenedesmus obliquus*, comprising a 5.2 kb region from the *Scenedesmus obliquus* chloroplast genome (*Scenedesmus* chloroplast sequence NCBI reference sequence: NC_008101, 057,611-062850bp), wherein the region comprises the nucleic acid sequence of SEQ ID NO: 125, or comprising a nucleotide sequence that is at least 80%, at least 85%, at least 90%, at least 95%, or at least 99% homologous to at least a 500 bp sequence of the nucleic acid sequence of SEQ ID NO: 125.

[0014] 76. An isolated nucleotide sequence comprising the nucleic acid of SEQ ID NO: 125, or comprising a nucleic acid sequence that is at least 80%, at least 85%, at least 90%, at least 95%, or at least 99% homologous to at least a 500 bp sequence of the nucleic acid sequence of SEQ ID NO: 125, wherein the isolated nucleotide sequence can be used to transform a chloroplast genome a *Scenedesmus* sp. 77. The isolated nucleotide sequence of claim 76, wherein the nucleic acid sequence of SEQ ID NO: 125 is modified to comprise a second nucleic acid encoding a protein.

[0015] 78. A host cell comprising a nucleic acid sequence of SEQ ID NO: 125, or comprising a nucleic acid sequence that is at least 80%, at least 85%, at least 90%, at least 95%, or at least 99% homologous to at least a 500 bp sequence of the nucleic acid sequence of SEQ ID NO: 125. 79. The host cell of claim 78, wherein the host cell is a host cell from a *Scenedesmus* sp. 80. The host cell of claim 78, wherein the nucleic acid sequence of SEQ ID NO: 125 is modified to comprise a second nucleic acid sequence encoding a protein.

[0015a] As used herein, except where the context requires otherwise, the term "comprise" and variations of the term, such as "comprising", "comprises" and "comprised", are not intended to exclude other additives, components, integers or steps.

[0015b] Reference to any prior art in the specification is not, and should not be taken as, an acknowledgment, or any form of suggestion, that this prior art forms part of the common general knowledge in Australia or any other jurisdiction or that this prior art could reasonably be expected to be ascertained, understood and regarded as relevant by a person skilled in the art.

BRIEF DESCRIPTION OF THE DRAWINGS

[0016] These and other features, aspects, and advantages of the present disclosure will become better understood with regard to the following description, appended claims and accompanying figures where:

[0017] Figure 1 shows a graphical representation of the mutated psbA fragment and 3'UTR vector used to engineer DCMU^r into *S. dimorphus*.

- [0018] Figure 2 shows amplification and digestion of DNA from psbA 264A DCMU^r transformants (3 and 4) and *S. dimorphus* wildtype (WT). U = uncut DNA and C = cut DNA (digested with Xba1).
- [0019] Figure 3 shows a psbA S264 transformant that is DCMU^r and atrazine^r.
- [0020] Figure 4 shows a graphical representation of p04-38.
- [0021] Figure 5 shows a graphical representation of p04-21.
- [0022] Figure 6 shows PCR amplification of DNA from *S. dimorphus* CAM^r transformants.
- [0023] Figure 7 shows a graphical representation of vector p04-31 used to transform *S. dimorphus*.
- [0024] Figure 8 shows a multiscreen of BD11 clones.
- [0025] Figure 9 shows a Western of SE0070 expressing BD11 (subclones of parent 2 and 4).
- [0026] Figure 10 shows endoxylanase activity in clarified lysates from *S. dimorphus* transformants containing the endoxylanase gene (parent 2 and 4).
- [0027] Figure 11 shows a graphical representation of p04-28.

[0028] Figure 12 shows verification of homoplasmy in several lines of *S. dimorphus* engineered with FPP-synthase.

[0029] Figure 13 shows an anti-flag Western with farnesyl diphosphate (FPP) synthase protein expression in 7 transformants.

[0030] Figure 14A shows an overlay of the Total Ion Chromatogram (TIC) for wild type negative control (untransformed *Scenedesmus dimorphus*), the engineered strain (*S. dimorphus* transformed with FPP synthase (avian), and a positive control of FPP. The y axis is abundance. The x axis is time.

[0031] Figure 14B shows the TIC of the FPP positive control. The retention time of FPP is 11.441 minutes. The y axis is abundance. The x axis is time.

[0032] Figure 14C shows the mass spectrum of the FPP positive control at 11.441 minutes. The y axis is abundance. The x axis is m/z (mass to charge ratio).

[0033] Figure 14D shows the TIC of the engineered strain (*S. dimorphus* transformed with FPP synthase (avian) incubated with IPP and DMAPP. The retention time of the product (FPP) is 11.441 minutes. The y axis is abundance. The x axis is time.

[0034] Figure 14E shows the mass spectrum of the engineered strain (*S. dimorphus* transformed with FPP synthase (avian) at 11.441 minutes. The y axis is abundance. The x axis is m/z (mass to charge ratio).

[0035] Figure 14F shows the TIC of the untransformed wild type *S. dimorphus* strain incubated with IPP and DMAPP. The y axis is abundance. The x axis is time.

[0036] Figure 14G shows the mass spectrum of the untransformed wild type *S. dimorphus* strain at 11.441 minutes. The y axis is abundance. The x axis is m/z (mass to charge ratio).

[0037] Figure 15A shows an overlay of the Total Ion Chromatogram (TIC) for wild type negative control (untransformed *Scenedesmus dimorphus*) incubated with IPP and DMAPP, the engineered strain (*S. dimorphus* transformed with IS09 (FPP synthase)) incubated with IPP and DMAPP, and a positive control of FPP. All three enzymatic reactions were incubated with amorphaadiene synthase to form amorphadiene in a coupled enzyme assay. The y axis is abundance. The x axis is time.

[0038] Figure 15B shows the TIC of the FPP positive control incubated with amorphaadiene synthase. The retention time of amorphadiene is 9.917 minutes. The y axis is abundance. The x axis is time.

[0039] Figure 15C shows the mass spectrum of the amorphadiene positive control at 9.917 minutes. The y axis is abundance. The x axis is m/z (mass to charge ratio).

[0040] Figure 15D shows the TIC of the engineered strain (*S. dimorphus* transformed with IS09 (FPP synthase), incubated with IPP and DMAPP and amorphaadiene synthase. The retention time of the product of the reaction (amorphadiene) is 9.917 minutes. The y axis is abundance. The x axis is time.

[0041] Figure 15E shows the mass spectrum of the product produced by the engineered strain (*S. dimorphus* transformed with IS09 (FPP synthase)) when incubated with IPP, DMAPP, and amorphaadiene synthase. The retention time of the product (amorphadiene) is 9.917 minutes. The y axis is abundance. The x axis is m/z (mass to charge ratio).

[0042] Figure 15F shows the TIC of the untransformed wild type *S. dimorphus* strain incubated with IPP and DMAPP and amorphaadiene synthase. The y axis is abundance. The x axis is time.

[0043] Figure 15G shows the mass spectrum (at 9.917 minutes) of the enzymatic reaction with untransformed wild type *S. dimorphus* strain incubated with IPP, DMAPP, and amorphaadiene synthase. The y axis is abundance. The x axis is m/z (mass to charge ratio).

[0044] Figure 16 shows a graphical representation of p04-196.

[0045] Figures 17A and 17B show a comparison of SIM monitored GC/MS chromatogram of *S. dimorphus* transformed with IS-88 (17A) and wild-type *S. dimorphus* (17B). Chromatograms were monitored with ions m/z=229, 135, and 122 (diagnostic for fusicoccadiene). The retention time of the peak in (17A) (7.617 min) matches that of purified fusicoccadiene.

[0046] Figures 18A and 18B show a mass spectra at the retention time of the fusicoccadiene peak (t=7.617 min) for *S. dimorphus*-IS88 (A) and wild type *S. dimorphus* (B). The mass spectrum of *S. dimorphus*-IS88 matches well with the known spectrum of fusicoccadiene. The mass spectrum of wild-type shows only background ions. Figure 18C shows the mass spectrum of purified fusicoccadiene.

[0047] Figure 19 shows a graphical representation of p04-118.

[0048] Figure 20 shows an anti-flag Western blot of *S. dimorphus* engineered with a gene encoding phytase (FD6).

[0049] Figure 21 shows a graphical representation of p04-162.

[0050] Figure 22 shows that the *EreB* gene (SEQ ID NO: 25) is amplified from DNA derived from several potential transformants but not from DNA derived from wild type *S. dimorphus*. Controls: W = no DNA; + = plasmid DNA; and D and O are *S. dimorphus* DNA.

[0051] Figure 23 shows a graphical representation of p04-161.

[0052] Figure 24 shows *codA* plates.

[0053] Figure 25 shows a graphical representation of p04-267.

[0054] Figure 26 shows an anti-flag Western blot of *S. dimorphus* engineered with FPP synthase (Is09) and bisabolene synthase (Is11) genes showing expression of both proteins.

[0055] Figure 27 shows a graphical representation of p04-116.

[0056] Figure 28 shows that endoxylanase is produced as a single peptide (not a fusion with CAT) in engineered *S. dimorphus* cells.

[0057] Figure 29 shows endoxylanase activity in engineered *S. dimorphus* (operon 1__1, 2__1, 2__2, 2__3). + is *S. dimorphus* engineered with *psbD* driving xylanase and "wt" is wild type.

[0058] Figure 30A and Figure 30B show that endoxylanase and CAT are transcribed as a single transcript. Figure 30A shows the primer design and Figure 30 B is an agarose gel showing amplification of cDNA from 4 of the 5 transformants corresponding to the endoxylanase-CAT transcript.

[0059] Figure 31 shows a graphical representation of transforming DNA with different RBS sequences. In both cases, the *psbD* promoter and the *psbA3'*UTR from *S. dimorphus* are used to regulate CAT-RBS-BD11 expression. BD11 encodes the endoxylanase gene from *T. reesei*. These cassettes were subcloned into vector p04-166 between region Homology A and homology region B.

[0060] Figure 32A shows xylanase activity of p04-231 from TAP plates. Figure 32B shows xylanase activity of p04-232 from TAP plates. Endoxylanase activity was detected in cells engineered with RBS1 linking CAT and endoxylanase (p04-231) but not with RBS2 (p04-232).

[0061] Figure 33 shows a graphical representation of p04-142.

- [0062] Figure 34 shows verification of homoplasmy in clone 52, an engineered *S. dimorphus* line containing a CAT cassette in the region between psbT and psbN.
- [0063] Figure 35 shows a graphical representation of the transforming DNA (A) and loopout product (B) that results from recombination at the identical D2 (psbD) promoter segments.
- [0064] Figure 36 shows failure to amplify a CAT fragment in a multiplex PCR of *S. dimorphus*.
- [0065] Figure 37 shows a graphical representation of p04-291 and p04-294. BAD1 and BAD4 are the betaine aldehyde dehydrogenase genes from spinach and sugar beet, respectively.
- [0066] Figure 38 shows an anti-HA western blot showing expression of betaine aldehyde dehydrogenase from spinach (291 clones 1, 2, 3, BAD1) or from sugar beet (294 4-1, 5-1, 6-1, 7-1, BAD4) in *S. dimorphus*.
- [0067] Figure 39 shows a graphical representation of p45-5 and p45-6.
- [0068] Figure 40 shows an agarose gel of EreB amplification from *D. tertiolecta* transformants in lanes 4, 5, 6.
- [0069] Figure 41 shows a graphical representation of p45-12.
- [0070] Figure 42 shows an agarose gel of EreB amplification from *D. tertiolecta* transformant 12-3.
- [0071] Figure 43 shows that Xylanase protein (BD11) is detected in *D. tertiolecta* transformant 12-3 via an anti-flag Western blot.
- [0072] Figure 44 shows that Xylanase activity is detected in *D. tertiolecta* transformant 12-3. Positive control is *S. dimorphus* engineered with endoxylanase.
- [0073] Figure 45 shows vector gutless pUC (2,436 bp).
- [0074] Figure 46 shows vector p04-35 (4,304 bp).
- [0075] Figure 47 shows vector pSS-007 (6,132 bp).
- [0076] Figure 48 shows vector pSS-013 (7,970 bp).
- [0077] Figure 49 shows vector pSS-023 (10,322 kb).
- [0078] Figure 50 shows Gene Vector 1 (5,774 bp).
- [0079] Figure 51 shows Gene Vector 2 (10.198 kb).
- [0080] Figure 52 shows Gene Vector 3 (7,111 bp).
- [0081] Figure 53 shows vector pRS414 (4,784 bp).
- [0082] Figure 54 shows vector pBeloBAC11 (7,507 bp).
- [0083] Figure 55 shows vector pLW001 (10.049 kb).
- [0084] Figure 56 shows vector pLW092 (13.737 kb).
- [0085] Figure 57 shows vector pBeloBAC-TRP (10.524 kb).
- [0086] Figure 58 shows vector pLW100 (18.847 kb).
- [0087] Figure 59 shows vector p04-198.
- [0088] Figure 60 shows vector pSS-035 (6,491 bp).
- [0089] Figure 61 shows vector pSS-023 CC93 CC94 (15.083 kb).
- [0090] Figure 62 shows vector pSS-023 CC93 CC97 (15.077 kb).
- [0091] Figure 63 shows vector pLW100 CC90 CC91 CC92 (26.319 kb).
- [0092] Figure 64 shows vector pLW100 four gene assembly (34.509 kb).
- [0093] Figure 65 shows pSS-023 restriction digest mapping with *NdeI*, *PacI*, *PstI*, *ScaI*, *SnaBI*, and *SpeI*.
- [0094] Figure 66 shows pLW001 restriction digest mapping with *EcoRV*, *NotI*, *PmlI*, *PvuI*, and *SnaBI*.

- [0095] Figures 67A-E show pLW092 restriction digest mapping with *PacI* (c), *PstI* (e), *ScaI* (b), and *XhoI* (d), and uncut (a).
- [0096] Figure 68 shows pLW100 restriction digest mapping with *EcoRV*, *NdeI*, *NotI*, *PacI*, *PstI*, *ScaI* and *XhoI*.
- [0097] Figure 69 shows pSS-035 restriction digest mapping with *EcoRI*, *EcoRV*, *KpnI*, *NotI*, *PvuII*, and *ScaI*.
- [0098] Figure 70 shows plasmid DNA comprising four two-gene contigs digested with *NdeI*.
- [0099] Figure 71 shows plasmid DNA comprising two three-gene contigs digested with *NdeI*.
- [0100] Figure 72 shows plasmid DNA comprising four four-gene contigs digested with *NdeI*.
- [0101] Figure 73 shows PCR amplification of the conserved *psbB-psbT-psbH-psbN* gene cluster from *S. dimorphus*.
- [0102] Figure 74 shows PCR amplification of the conserved *psbB-psbT-psbH-psbN* gene cluster from a strain of genus *Dunaliella*; an unknown species.
- [0103] Figure 75 shows PCR amplification of the conserved *psbB-psbT-psbH-psbN* gene cluster from *N. abundans*.
- [0104] Figure 76 shows vector p04-128.
- [0105] Figure 77 shows vector p04-129.
- [0106] Figure 78 shows vector p04-130.
- [0107] Figure 79 shows vector p04-131.
- [0108] Figure 80 shows vector p04-142.
- [0109] Figure 81 shows vector p04-143.
- [0110] Figure 82 shows vector p04-144.
- [0111] Figure 83 shows vector p04-145.
- [0112] Figure 84 shows a homoplasmy PCR screen for clones from *S. dimorphus* that have a resistance cassette between either *psbT* and *psbN* (p04-142) or between *psbN* and *psbH* (p04-143).
- [0113] Figure 85 shows a homoplasmy PCR screen for clones from *S. dimorphus* that have a resistance cassette between either *psbH* and *psbK* (p04-144) or 3' of *psbK* (p04-145).
- [0114] Figure 86 is a nucleotide alignment of the *psbB* gene from four different algae species.
- [0115] Figure 87 is a nucleotide alignment of the *psbH* region from four different algae species.
- [0116] Figure 88 is an alignment of the genome region from the *psbB* gene to the *psbH* gene of four different algae species.
- [0117] Figure 89 is vector p04-151.
- [0118] Figure 90A-D shows restriction enzyme mapping results.
- [0119] Figure 91 shows vector pLW106.
- [0120] Figure 92A and B depict 4 clones that screen PCR positive for both BD11 and IS99.
- [0121] Figure 93A-C depict 4 clones that screen PCR positive for CC90, CC91, and CC92.
- [0122] Figure 94A and B depict 2 clones that screen PCR positive for IS61, IS62, IS57 and IS116.

DETAILED DESCRIPTION

[0123] The following detailed description is provided to aid those skilled in the art in practicing the present disclosure. Even so, this detailed description should not be construed to unduly limit the present disclosure as modifications and variations in the embodiments discussed herein can be made by those of ordinary skill in the art without departing from the spirit or scope of the present disclosure.

[00124] As used in this specification and the appended claims, the singular forms “a”, “an” and “the” include plural reference unless the context clearly dictates otherwise.

[00125] Endogenous

[00126] An endogenous nucleic acid, nucleotide, polypeptide, or protein as described herein is defined in relationship to the host organism. An endogenous nucleic acid, nucleotide, polypeptide, or protein is one that naturally occurs in the host organism.

[00127] Exogenous

[00128] An exogenous nucleic acid, nucleotide, polypeptide, or protein as described herein is defined in relationship to the host organism. An exogenous nucleic acid, nucleotide, polypeptide, or protein is one that does not naturally occur in the host organism or is at a different location in the host organism.

[00129] Examples of genes, nucleic acids, proteins, and polypeptides that can be used in the embodiments disclosed herein include, but are not limited to:

[00130] SEQ ID NO: 1 is a PCR primer.

[00131] SEQ ID NO: 2 is a PCR primer.

[00132] SEQ ID NO: 3 is a PCR primer.

[00133] SEQ ID NO: 4 is a PCR primer.

[00134] SEQ ID NO: 5 is a PCR primer.

[00135] SEQ ID NO: 6 is a PCR primer.

[00136] SEQ ID NO: 7 is a PCR primer.

[00137] SEQ ID NO: 8 is a PCR primer.

[00138] SEQ ID NO: 9 is a PCR primer.

[00139] SEQ ID NO: 10 is a PCR primer.

[00140] SEQ ID NO: 11 is a PCR primer.

[00141] SEQ ID NO: 12 is a PCR primer.

[00142] SEQ ID NO: 13 is a PCR primer.

[00143] SEQ ID NO: 14 is a PCR primer.

[00144] SEQ ID NO: 15 is a PCR primer (#4682).

[00145] SEQ ID NO: 16 is a PCR primer (#4982).

[00146] SEQ ID NO: 17 is a PCR primer.

[00147] SEQ ID NO: 18 is a PCR primer.

[00148] SEQ ID NO: 19 is a PCR primer.

[00149] SEQ ID NO: 20 is a nucleotide sequence of an artificial FLAG epitope tag linked to a MAT epitope tag by a TEV protease site.

[00150] SEQ ID NO: 21 is a gene encoding an endoxylanase from *T. reesei* codon optimized for chloroplast expression in *C. reinhardtii*.

[00151] SEQ ID NO: 22 is a nucleotide sequence of an artificial TEV protease site linked to a FLAG epitope tag.

[00152] SEQ ID NO: 23 is a gene encoding an FPP synthase from *G. gallus* codon optimized for chloroplast expression in *C. reinhardtii*.

[00153] SEQ ID NO: 24 is a nucleotide sequence of an artificial streptavidin epitope tag.

[00154] SEQ ID NO: 25 is a gene encoding a fusicoccadiene synthase from *P. amygdali* codon optimized according to the most frequent codons in the *C. reinhardtii* chloroplast.

[00155] SEQ ID NO: 26 is a gene encoding a phytase from *E. coli* codon optimized for chloroplast expression in *C. reinhardtii*.

[00156] SEQ ID NO: 27 is a nucleotide sequence of an artificial FLAG epitope tag linked to a MAT epitope tag by a TEV protease site.

[00157] SEQ ID NO: 28 is a modified chloramphenicol acetyltransferase gene from *E. coli* with the nucleotide at position 64 changed from an A to a G, the nucleotides at positions 436, 437, and 438 were changed from TCA to AGC, and the nucleotide at position 516 was changed from a C to a T.

[00158] SEQ ID NO: 29 is a modified erythromycin esterase gene from *E. coli* with the nucleotide at position 153 changed from a C to a T, the nucleotide at position 195 changed from a T to a C, the nucleotide at position 198 changed from a A to a C, the nucleotide at position 603 changed from a T to a A, the nucleotide at position 1194 changed from a C to a T, and the nucleotide at position 1203 changed from a T to an A.

[00159] SEQ ID NO: 30 is a fragment of genomic DNA from *S. dimorphus* that encodes a region containing a portion of the 3' end of the *psbA* gene and some untranslated region, with nucleotide 1913 of the fragment mutated from a T to a G for the S264A mutation, and nucleotides 1928 to 1930 mutated from CGT to AGA to generate a silent *XbaI* restriction site.

[00160] SEQ ID NO: 31 is a gene encoding a cytosine decarboxylase from *E. coli* codon optimized for expression in the chloroplast of *C. reinhardtii*.

[00161] SEQ ID NO: 32 is a gene encoding a betaine aldehyde dehydrogenase from *S. oleracea* codon optimized according to the tRNA usage of the chloroplast of *C. reinhardtii*.

[00162] SEQ ID NO: 33 is a nucleotide sequence of an artificial 3xHA tag linked to a 6xHIS tag by a TEV protease site.

[00163] SEQ ID NO: 34 is a gene encoding a betaine aldehyde dehydrogenase from *B. vulgaris* codon optimized for expression in the chloroplast of *C. reinhardtii*.

[00164] SEQ ID NO: 35 is a gene encoding an E-alpha-bisabolene synthase from *A. grandis* codon optimized for expression in the chloroplast of *C. reinhardtii*.

[00165] SEQ ID NO: 36 is a modified nucleotide sequence that is the reverse complement of SEQ ID NO: 37 with extra nucleotides on the 5' and 3' ends; nucleotides 1-43 are extra on the 3' end and nucleotides 532-541 are extra on the 5' end.

[00166] SEQ ID NO: 37 is a nucleotide sequence of the endogenous promoter from the *psbA* gene of *S. dimorphus* that was cloned into an integration vector.

[00167] SEQ ID NO: 38 is a modified nucleotide sequence that is the reverse complement of SEQ ID NO: 39 with extra nucleotides on the 5' end and a nucleotide insertion; nucleotides 535-716 are extra 5' sequence and nucleotides 176-188 are the insertion.

[00168] SEQ ID NO: 39 is a nucleotide sequence of the endogenous promoter for the *psbB* gene of *S. dimorphus* that was cloned into integration vectors.

[00169] SEQ ID NO: 40 is a sequence of the endogenous promoter for the *psbD* gene of *S. dimorphus* that was cloned into integration vectors.

[00170] SEQ ID NO: 41 is a modified nucleotide sequence that is the reverse complement of SEQ ID NO: 42 with extra sequence on the 5'; nucleotides 537-464 are extra sequences on the 5' end, nucleotide 308 is changed from a C to a T, nucleotide 310 is changed from a C to a T, and nucleotide 259 is changed from an A to a G.

[00171] SEQ ID NO: 42 is a nucleotide sequence of the endogenous promoter for the *tufA* gene of *S. dimorphus* that was cloned into integration vectors.

[00172] SEQ ID NO: 43 is a modified nucleotide sequence that is the reverse of SEQ ID NO: 44 with extra sequences on the 5' end; nucleotides 550-557 are extra sequences on the 5' end.

[00173] SEQ ID NO: 44 is a nucleotide sequence for the endogenous promoter of the *rpoA* of *S. dimorphus*.

[00174] SEQ ID NO: 45 is a nucleotide sequence for the endogenous promoter of the *cemA* gene in *S. dimorphus* that was cloned into integration vectors.

[00175] SEQ ID NO: 46 is a modified nucleotide sequence that is the reverse complement of SEQ ID NO: 47 with an insertion at nucleotides 233-266.

[00176] SEQ ID NO: 47 is a nucleotide sequence for the endogenous promoter of the *fisH* gene in *S. dimorphus* that was cloned into integration vectors.

[00177] SEQ ID NO: 48 is a modified nucleotide sequence of SEQ ID NO: 49 that has extra sequences on the 5' end; nucleotides 1-19 are extra sequences, nucleotide 404 has been changed from an A to a T.

[00178] SEQ ID NO: 49 is a nucleotide sequence for the endogenous promoter of the *rbcL* gene in *S. dimorphus* that was cloned into integration vectors.

[00179] SEQ ID NO: 50 is a modified nucleotide sequence of SEQ ID NO: 51 that has 24 nucleotides truncated on the 5' end; the nucleotide at position 2 is changed from a G to a C, position 5 is changed from an A to a G, at positions 199 and 200 two T's are inserted, and at position 472 it is changed from an A to a G.

[00180] SEQ ID NO: 51 is the nucleotide sequence of the endogenous promoter for the *chlB* gene from *S. dimorphus* that was cloned into integration vectors.

[00181] SEQ ID NO: 52 is a modified nucleotide sequence of SEQ ID NO: 53 where nucleotides 1-3 are extra sequence, the nucleotide at position 442 is a G insertion, and the R at position 482 is a result of poor sequencing.

[00182] SEQ ID NO: 53 is a nucleotide sequence for the endogenous promoter of the *petA* gene in *S. dimorphus* that was cloned into integration vectors.

[00183] SEQ ID NO: 54 is a modified nucleotide sequence that is the reverse complement of SEQ ID NO: 55 where nucleotides 3, 8, 18, 21, 49, 57, and 82 are insertions, nucleotides 484-503 are extra on the 5' end, nucleotide 26 is changed from a C to a T, and nucleotide 30 is changed from an A to a C.

[00184] SEQ ID NO: 55 is the nucleotide sequence of the endogenous promoter for the *petB* gene from *S. dimorphus* that was cloned into integration vectors.

[00185] SEQ ID NO: 56 is the modified nucleotide sequence of SEQ ID NO: 57 that has 3 nucleotides truncated on the 5' end.

[00186] SEQ ID NO: 57 is the nucleotide sequence of the endogenous terminator region for the *rbcL* gene from *S. dimorphus* that was cloned into integration vectors.

[00187] SEQ ID NO: 58 is the nucleotide sequence of the endogenous terminator region for the *psbA* gene from *S. dimorphus* that was cloned into integration vectors.

[00188] SEQ ID NO: 59 is the nucleotide sequence of the endogenous terminator region for the *psaB* gene from *S. dimorphus* that was cloned into integration vectors.

[00189] SEQ ID NO: 60 is a nucleic acid linker sequence (RBS1).

[00190] SEQ ID NO: 61 is a nucleic acid linker sequence (RBS2).

[00191] SEQ ID NO: 62 is the nucleotide sequence of the endogenous promoter region for the *psbD* gene from *D. tertiolecta* that was cloned into integration vectors.

[00192] SEQ ID NO: 63 is the nucleotide sequence of the endogenous promoter region for the *tufA* gene from *D. tertiolecta* that was cloned into integration vectors.

[00193] SEQ ID NO: 64 is the nucleotide sequence of the endogenous terminator region for the *rbcL* gene from *D. tertiolecta* that was cloned into integration vectors.

[00194] SEQ ID NO: 65 is the nucleotide sequence of the endogenous terminator region for the *psbA* gene from a *Dunaliella* isolate of unknown species that was cloned into integration vectors.

[00195] SEQ ID NO: 66 is PCR primer 1.

[00196] SEQ ID NO: 67 is PCR primer 2.

[00197] SEQ ID NO: 68 is PCR primer 3.

[00198] SEQ ID NO: 69 is PCR primer 4.

[00199] SEQ ID NO: 70 is PCR primer 5.

[00200] SEQ ID NO: 71 is PCR primer 6.

[00201] SEQ ID NO: 72 is PCR primer 7.

[00202] SEQ ID NO: 73 is PCR primer 8.

[00203] SEQ ID NO: 74 is PCR primer 9.

[00204] SEQ ID NO: 75 is PCR primer 10.

[00205] SEQ ID NO: 76 is PCR primer 11.

[00206] SEQ ID NO: 77 is PCR primer 12.

[00207] SEQ ID NO: 78 is PCR primer 13.

[00208] SEQ ID NO: 79 is PCR primer 14.

[00209] SEQ ID NO: 80 is PCR primer 15.

[00210] SEQ ID NO: 81 is PCR primer 16.

[00211] SEQ ID NO: 82 is PCR primer 17.

[00212] SEQ ID NO: 83 is PCR primer 18.

[00213] SEQ ID NO: 84 is PCR primer 19.

[00214] SEQ ID NO: 85 is PCR primer 20.

[00215] SEQ ID NO: 86 is PCR primer 21.

[00216] SEQ ID NO: 87 is PCR primer 22.

[00217] SEQ ID NO: 88 is PCR primer 23.

[00218] SEQ ID NO: 89 is PCR primer 24.

[00219] SEQ ID NO: 90 is PCR primer 25.

[00220] SEQ ID NO: 91 is PCR primer 26.

[00221] SEQ ID NO: 92 is PCR primer 27.

- [00222] SEQ ID NO: 93 is PCR primer 28.
- [00223] SEQ ID NO: 94 is PCR primer 29.
- [00224] SEQ ID NO: 95 is PCR primer 30.
- [00225] SEQ ID NO: 96 is PCR primer 31.
- [00226] SEQ ID NO: 97 is PCR primer 32.
- [00227] SEQ ID NO: 98 is PCR primer 33.
- [00228] SEQ ID NO: 99 is PCR primer 34.
- [00229] SEQ ID NO: 100 is PCR primer 35.
- [00230] SEQ ID NO: 101 is PCR primer 36.
- [00231] SEQ ID NO: 102 is PCR primer 37.
- [00232] SEQ ID NO: 103 comprises a nucleic acid sequence encoding for URA3.
- [00233] SEQ ID NO: 104 comprises a nucleic acid sequence encoding for ADE2.
- [00234] SEQ ID NO: 105 comprises a nucleic acid sequence encoding for URA3-ADE2.
- [00235] SEQ ID NO: 106 is a nucleic acid linker sequence with engineered restriction sites.
- [00236] SEQ ID NO: 107 comprises a nucleic acid sequence encoding for TRP1-ARS1-CEN4 (from pYAC4).
- [00237] SEQ ID NO: 108 comprises a nucleic acid sequence encoding for LEU2.
- [00238] SEQ ID NO: 109 comprises a nucleic acid sequence encoding for CC-93.
- [00239] SEQ ID NO: 110 comprises a nucleic acid sequence encoding for CC-94.
- [00240] SEQ ID NO: 111 comprises the contig sequence (CC93-CC94) that was inserted into pSS-023.
- [00241] SEQ ID NO: 112 comprises the contig sequence (CC93-CC97) that was inserted into pSS-023.
- [00242] SEQ ID NO: 113 comprises a nucleic acid sequence encoding for CC-97.
- [00243] SEQ ID NO: 114 comprises the contig sequence (CC90-CC91-CC92) that was inserted into pLW100.
- [00244] SEQ ID NO: 115 comprises a nucleic acid sequence encoding for CC-90.
- [00245] SEQ ID NO: 116 comprises a nucleic acid sequence encoding for CC-91.
- [00246] SEQ ID NO: 117 comprises a nucleic acid sequence encoding for CC-92.
- [00247] SEQ ID NO: 118 comprises a nucleic acid sequence encoding for HIS3.
- [00248] SEQ ID NO: 119 comprises a nucleic acid sequence encoding for LYS2.
- [00249] SEQ ID NO: 120 comprises the contig sequence (IS57-IS116-IS62-IS61) that was inserted into pLW100.
- [00250] SEQ ID NO: 121 comprises a nucleic acid sequence encoding for IS57.
- [00251] SEQ ID NO: 122 comprises a nucleic acid sequence encoding for IS116.
- [00252] SEQ ID NO: 123 comprises a nucleic acid sequence encoding for IS62.
- [00253] SEQ ID NO: 124 comprises a nucleic acid sequence encoding for IS61.
- [00254] SEQ ID NO: 125 is a 5,240 base pair sequence from *Scenedesmus obliquus*.
- [00255] SEQ ID NO: 126 is the A3 homology region.
- [00256] SEQ ID NO: 127 is the B3 homology region.
- [00257] SEQ ID NO: 128 comprises a sequence encoding for rblcL-CAT-psbE.
- [00258] SEQ ID NO: 129 is a degenerate PCR primer.
- [00259] SEQ ID NO: 130 is a degenerate PCR primer.
- [00260] SEQ ID NO: 131 is a degenerate PCR primer.

[00261] SEQ ID NO: 132 is a degenerate PCR primer.

[00262] SEQ ID NO: 133 is genomic sequence of the region encoding the *psbB*, *psbT*, *psbN*, and *psbH* genes from *D. tertiolecta*.

[00263] SEQ ID NO: 134 is genomic sequence of the region encoding the *psbB*, *psbT*, *psbN*, and *psbH* genes from a *Dunaliella* of unknown species.

[00264] SEQ ID NO: 135 is a partial genomic sequence of the region encoding the *psbB*, *psbT*, *psbN*, and *psbH* genes from *N. abundans*; the stretch of N's represents a gap in the sequence.

[00265] SEQ ID NO: 136 is genomic sequence of the region encoding the *psbB*, *psbT*, *psbN*, and *psbH* genes from an isolate of *C. vulgaris*.

[00266] SEQ ID NO: 137 is genomic sequence of the region encoding the *psbB*, *psbT*, *psbN*, and *psbH* genes from *T. suecica*.

[00267] SEQ ID NO: 138 is PCR primer (#4682).

[00268] SEQ ID NO: 139 is PCR primer (#4982).

[00269] SEQ ID NO: 140 is PCR primer 4684.

[00270] SEQ ID NO: 141 is PCR primer 4685.

[00271] SEQ ID NO: 142 is PCR primer 4686

[00272] SEQ ID NO: 143 is PCR primer 4687.

[00273] SEQ ID NO: 144 is PCR primer 4688.

[00274] SEQ ID NO: 145 is PCR primer 4689.

[00275] SEQ ID NO: 146 comprise a nucleotide sequence encoding BD11.

[00276] SEQ ID NO: 147 comprise a nucleotide sequence encoding IS99.

[00277] SEQ ID NO: 148 comprise a nucleotide sequence encoding CAT.

[00278] SEQ ID NO: 149 to SEQ ID NO: 170 are PCR primers.

[00279] The present disclosure relates to methods of transforming various species of algae, for example, algae from the genus *Scenedesmus* and from the genus *Dunaliella*, vectors and nucleic acid constructs useful in conducting such transformations, and recombinant *Scenedesmus* and *Dunaliella* organisms produced using the vectors and methods disclosed herein. In one embodiment, the *Scenedesmus* sp. utilized is *Scenedesmus dimorphus*. *Scenedesmus* sp. are members of the Chlorophyceans a diverse assemblage of green algae. *Scenedesmus* is a genus consisting of unicells or flat coenobial colonies of 2, 4, 8 or 16 linearly arranged cells. Cells contain a single plastid with pyrenoid and uninucleate. *Scenedesmus* sp. are common inhabitants of the plankton of freshwaters and brackish waters and occasionally form dense populations. In one embodiment, the organism utilized is from the genus *Dunaliella*. In another embodiment, the *Dunaliella* sp. is *D. tertiolecta*.

[00280] One embodiment, the disclosure provides vectors useful in the transformation of *Scenedesmus* sp., for example, *Scenedesmus dimorphus* or *Scenedesmus obliquus*. In another, embodiment, the disclosure provides vectors useful in the transformation of *Dunaliella* sp., for example, *Dunaliella tertiolecta*. An expression cassette can be constructed in an appropriate vector. In some instances, the cassette is designed to express one or more protein-coding sequences in a host cell. Such vectors can be constructed using standard techniques known in the art. In a typical expression cassette, the promoter or regulatory element is positioned on the 5' or upstream side of a coding sequence whose expression is desired. In other cassettes, a coding sequence may be flanked by sequences which allow for

expression upon insertion into a target genome (e.g., nuclear or plastid). For example, a nucleic acid encoding an enzyme involved in the synthesis of a compound of interest, for example an isoprenoid, such that expression of the enzyme is controlled by a naturally occurring regulatory element. Any regulatory element which provides expression under appropriate conditions such that the mRNA or protein product is expressed to a level sufficient to produce useful amount of the desired compound can be used.

[00281] One or more additional protein coding sequences can be operatively fused downstream or 3' of a promoter. Coding sequences for single proteins can be used, as well as coding sequences for fusions of two or more proteins. Coding sequences may also contain additional elements that would allow the expressed proteins to be targeted to the cell surface and either be anchored on the cell surface or be secreted to the environment. A selectable marker is also employed in the design of the vector for efficient selection of algae transformed by the vector. Both a selectable marker and another sequence which one desires to introduce may be introduced fused to and downstream of a single promoter. Alternatively, two protein coding sequences can be introduced, each under the control of a promoter.

[00282] One approach to construction of a genetically manipulated strain of *Scenedesmus* or *Dunaliella* involves transformation with a nucleic acid which encodes a gene of interest, typically an enzyme capable of converting a precursor into a fuel product or precursor of a fuel product (e.g., an isoprenoid or fatty acid), a biomass degrading enzyme, or an enzyme for the improvement of a characteristic of a feedstuff. In some embodiments, a transformation may introduce nucleic acids into any plastid of the host alga cell (e.g., chloroplast). In other embodiments, a transformation may introduce nucleic acids into the nuclear genome of the host cell. In still other embodiments, a transformation introduces nucleic acids into both the nuclear genome and a plastid. In some instances, the nucleic acids encoding proteins of interest (e.g., transporters or enzymes) are codon-biased for the intended site of insertion (e.g., nuclear codon-biased for insertion in the nucleus, chloroplast codon-biased for insertion in the chloroplast).

[00283] To construct the vector, the upstream DNA sequences of a gene expressed under control of a suitable promoter may be restriction mapped and areas important for the expression of the protein characterized. The exact location of the start codon of the gene is determined and, making use of this information and the restriction map, a vector may be designed for expression of an endogenous or exogenous protein by removing the region responsible for encoding the gene's protein but leaving the upstream region found to contain the genetic material responsible for control of the gene's expression. A synthetic oligonucleotide is typically inserted in the location where the protein sequence once was, such that any additional gene could be cloned in using restriction endonuclease sites in the synthetic oligonucleotide (i.e., a multicloning site). An unrelated gene (or coding sequence) inserted at this site would then be under the control of an extant start codon and upstream regulatory region that will drive expression of the foreign (i.e., not normally there) protein encoded by this gene. Once the gene for the foreign protein is put into a cloning vector, it can be introduced into the host organism using any of several methods, some of which might be particular to the host organism. Variations on these methods are amply described in the general literature.

[00284] The term "exogenous" is used herein in a comparative sense to indicate that a nucleotide sequence (or polypeptide) being referred to is from a source other than a reference source and is different from the sequence of the reference, or is linked to a second nucleotide sequence (or polypeptide) with which it is not normally associated, or is modified such that it is in a form that is not normally associated with a reference material. For example, a polynucleotide encoding an enzyme is exogenous with respect to a nucleotide sequence of a chloroplast, where the polynucleotide is not normally found in the chloroplast (e.g., a mutated polynucleotide encoding a chloroplast sequence

or a nuclear sequence). As another example, a polynucleotide encoding an enzyme is exogenous with respect to a host organism where the polynucleotide comprises operatively linked sequences (e.g., promoters, homologous recombination sites, selectable markers, and/or termination sequences), that are not normally found in the reference organism.

[00285] Polynucleotides encoding enzymes and other proteins useful in the present disclosure may be isolated and/or synthesized by any means known in the art, including, but not limited to cloning, sub-cloning, and PCR. A vector herein may encode polypeptide(s) having a role in the mevalonate pathway, such as, for example, thiolase, HMG-CoA synthase, HMG-CoA reductase, mevalonate kinase, phosphomevalonate kinase, and mevalonate-5-pyrophosphate decarboxylase. In other embodiments, the polypeptides are enzymes in the non-mevalonate pathway, such as DOXP synthase, DOXP reductase, 4-diphosphocytidyl-2-C-methyl-D-erythritol synthase, 4-diphosphocytidyl-2-C-methyl-D-erythritol kinase, 2-C-methyl-D-erythritol 2,4,-cyclodiphosphate synthase, HMB-PP synthase, HMB-PP reductase, or DOXP reductoisomerase.

[00286] One embodiment is directed to a vector comprising a nucleic acid encoding an enzyme capable of modulating a fusicoccadiene biosynthetic pathway. Such a vector may further comprise a promoter for expression of the nucleic acid in algae. Nucleic acid(s) included in such vectors may contain a codon biased form of a gene, optimized for expression in a host organism of choice. In one embodiment, the fusicoccadiene produced is fusicocca-2,10(14)-diene. Another aspect of the present disclosure is directed to a vector comprising a nucleic acid encoding an enzyme that produces a fusicoccadiene when the vector is integrated into a genome of an organism, such as photosynthetic algae, wherein the organism does not produce fusicoccadiene without the vector and wherein the fusicoccadiene is metabolically inactive in the organism.

[00287] Further provided herein is a method of producing a fuel product, comprising: a) transforming a *Scenedesmus* sp or *Dunaliella* sp., wherein the transformation results in the production or increased production of a fusicoccadiene; b) collecting the fusicoccadiene from the organism; and c) using the fusicoccadiene to produce a fuel product.

[00288] The present disclosure also contemplates host cells making polypeptides that contribute to the secretion of fatty acids, lipids or oils, by transforming host cells (e.g., algal cells) and/or organisms comprising host cells with nucleic acids encoding one or more different transporters. In some embodiments, the host cells or organisms are also transformed with one or more 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, ketoacylsynthase, enoylreductase and a desaturase. In some embodiments, the enzymes are catabolic or biodegrading enzymes. In some embodiments, a single enzyme is produced.

[00289] 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 synthesis pathway. One example of a pathway might include genes encoding an acetyl CoA carboxylase, a malonyltransferase, a ketoacylsynthase, and a thioesterase. Cells transformed with entire pathways and/or enzymes extracted from those cells, can synthesize complete fatty acids or intermediates of the fatty acid synthesis pathway. Constructs may contain multiple copies of the same gene, multiple genes encoding the same enzyme from different organisms, and/or multiple genes with one or more mutations in the coding sequence(s).

[00290] In some instances, the host cell will naturally produce the fatty acid, lipid, triglyceride or oil of interest. Thus, transformation of the host cell with a polynucleotide encoding a transport protein will allow for secretion or increased

secretion of the molecule of interest from the cell. In other instances, the host cell is transformed with a polynucleotide encoding one or more enzymes necessary for the production of the molecule of interest. The enzymes produced by the modified cells result in the production of fatty acids, lipids, triglycerides or oils that may be collected from the cells and/or the surrounding environment (e.g., bioreactor, growth medium). In some embodiments, the collection of the fatty acids, lipids, triglycerides or oils is performed after the product is secreted from the cell via a cell membrane transporter.

[00291] Synthesis of fatty acids, lipids or oils can also be accomplished by engineering a cell to express an accessory molecule or modulation molecule. In certain embodiments, the accessory molecule is an enzyme that produces a substrate utilized by a fatty acid synthesizing enzyme. In some embodiments the accessory or modulation molecule contributes to the growth or nourishment of the biomass.

[00292] An additional aspect of the present disclosure provides a vector comprising a nucleic acid encoding a biomass degrading enzyme and a promoter configured for expression of the nucleic acids in a non-vascular photosynthetic organism, for example a *Scenedesmus* sp. and more particularly *S. dimorphus* or *Dunaliella* sp.. Vectors of the present disclosure may contain nucleic acids encoding more than one biomass degrading enzyme and, in other instances, may contain nucleic acids encoding polypeptides which covalently link biomass degrading enzymes. Biomass degrading enzymes may include cellulolytic enzymes, hemicellulolytic enzymes and ligninolytic enzymes. More specifically, the biomass degrading enzymes may be exo- β -glucanase, endo- β -glucanase, β -glucosidase, endoxylanase, or lignase. Nucleic acids encoding the biomass degrading enzymes may be derived from fungal or bacterial sources, for example, those encoding exo- β -glucanase in *Trichoderma viride*, exo- β -glucanase in *Trichoderma reesei*, exo- β -glucanase in *Aspergillus aculeatus*, endo- β -glucanase in *Trichoderma reesei*, endo- β -glucanase in *Aspergillus niger*, β -glucosidase in *Trichoderma reesei*, β -glucosidase in *Aspergillus niger* endoxylanase in *Trichoderma reesei*, and endoxylanase in *Aspergillus niger*. Other nucleic acids encoding biomass degrading enzymes may be endogenous to the organisms.

[00293] Also provided is a composition containing a plurality of vectors each of which encodes a different biomass degrading enzyme and a promoter for expression of said biomass degrading enzymes in a chloroplast. Such compositions may contain multiple copies of a particular vector encoding a particular enzyme. In some instances, the vectors will contain nucleic acids encoding cellulolytic, hemicellulolytic and/or ligninolytic enzymes. More specifically, the plurality of vectors may contain vectors capable of expressing exo- β -glucanase, endo- β -glucanase, β -glucosidase, endoxylanase and/or lignase. Some of the vectors of this embodiment are capable of insertion into a chloroplast genome and such insertion can lead to disruption of the photosynthetic capability of the transformed chloroplast. Insertion of other vectors into a chloroplast genome does not disrupt photosynthetic capability of the transformed chloroplast. Some vectors provide for expression of biomass degrading enzymes which are sequestered in a transformed chloroplast.

[00294] Another vector encodes a plurality of distinct biomass degrading enzymes and a promoter for expression of the biomass degrading enzymes in a non-vascular photosynthetic organism. The biomass degrading enzymes may be one or more of cellulolytic, hemicellulolytic or ligninolytic enzymes. In some vectors, the plurality of distinct biomass degrading enzymes is two or more of exo- β -glucanase, endo- β -glucanase, β -glucosidase, lignase and endoxylanase. In some embodiments, the plurality of enzymes is operatively linked. In other embodiments, the plurality of enzymes is expressed as a functional protein complex. Insertion of some vectors into a host cell genome does not disrupt photosynthetic capability of the organism. Vectors encoding a plurality of distinct enzymes, may lead to production of

enzymes which are sequestered in a chloroplast of a transformed organism. The present disclosure also provides an algal cell and in particular a *Scenedemus sp.* or *Dunaliella sp.* transformed with a vector encoding a plurality of distinct enzymes. For some embodiments, the organism may be grown in the absence of light and/or in the presence of an organic carbon source.

[00295] Yet another aspect provides a genetically modified chloroplast of a *Scenedemus sp.* or *Dunaliella sp.* producing one or more biomass degrading enzymes. Such enzymes may be cellulolytic, hemicellulolytic or ligninolytic enzymes, and more specifically, may be an exo- β -glucanase, an endo- β -glucanase, a β -glucosidase, an endoxylanase, a lignase and/or combinations thereof. The one or more enzymes are sequestered in the chloroplast in some embodiments. The present disclosure also provides photosynthetic organisms containing the genetically modified chloroplasts of the present disclosure.

[00296] Yet another aspect provides a method for preparing a biomass-degrading enzyme. This method comprises the steps of (1) transforming a photosynthetic, non-vascular organism and in particular a *Scenedemus sp.* or *Dunaliella sp.* to produce or increase production of said biomass-degrading enzyme and (2) collecting the biomass-degrading enzyme from said transformed organism. Transformation may be conducted with a composition containing a plurality of different vectors encoding different biomass degrading enzymes. Transformation may also be conducted with a vector encoding a plurality of distinct biomass degrading enzymes. Any or all of the enzymes may be operatively linked to each other. In some instances, a chloroplast is transformed. This method may have one or more additional steps, including: (a) harvesting transformed organisms; (b) drying transformed organisms; (c) harvesting enzymes from a cell medium; (d) mechanically disrupting transformed organisms; or (e) chemically disrupting transformed organisms. The method may also comprise further purification of an enzyme through performance liquid chromatography.

[00297] Still another method of the present disclosure allows for preparing a biofuel. One step of this method includes treating a biomass with one or more biomass degrading enzymes derived from a photosynthetic, nonvascular organism for a sufficient amount of time to degrade at least a portion of said biomass. The biofuel produced may be ethanol. The enzymes of this method may contain at least traces of said photosynthetic nonvascular organism from which they are derived. Additionally, the enzymes useful for some embodiments of this method include cellulolytic, hemicellulolytic and ligninolytic enzymes. Specific enzymes useful for some aspects of this method include exo- β -glucanase, endo- β -glucanase, β -glucosidase, endoxylanase, and/or lignase. Multiple types of biomass including agricultural waste, paper mill waste, corn stover, wheat stover, soy stover, switchgrass, duckweed, poplar trees, woodchips, sawdust, wet distiller grain, dry distiller grain, human waste, newspaper, recycled paper products, or human garbage may be treated with this method of the disclosure. Biomass may also be derived from a high-cellulose content organism, such as switchgrass or duckweed. The enzyme(s) used in this method may be liberated from the organism and this liberation may involve chemical or mechanical disruption of the cells of the organism. In an alternate embodiment, the enzyme(s) are secreted from the organism and then collected from a culture medium. The treatment of the biomass may involve a fermentation process, which may utilize a microorganism other than the organism which produced the enzyme(s). In some instances the non-vascular photosynthetic organism may be added to a saccharification tank. This embodiment may also comprise the step of collecting the biofuel. Collection may be performed by distillation. In some instances, the biofuel is mixed with another fuel.

[00298] An additional method provides for making at least one biomass degrading enzyme by transforming a chloroplast to make a biomass degrading enzyme. The biomass degrading enzyme may be a cellulolytic enzyme, a

hemicellulolytic enzyme, or a ligninolytic enzyme, and specifically may be α -glucanase, α -glucanase, β -glucosidase, endoxylanase, or lignase. In some instances, the biomass degrading enzyme is sequestered in the transformed chloroplast. The method may further involve disrupting, via chemical or mechanical means, the transformed chloroplast to release the biomass degrading enzyme(s). In some instances, multiple enzymes will be produced by a transformed chloroplast. The biomass degrading enzymes may be of fungal or bacterial origin, for example, α -glucanase, α -glucanase, β -glucosidase, endoxylanase, lignase, or a combination thereof.

[00299] 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 an entire biodegradation pathway. One example of a pathway might include genes encoding an α -glucanase (acts on the cellulose end chain), an α -glucanase (acts on the interior portion of a cellulose chain), β -glucosidase (avoids reaction inhibitors by / degrades cellobiose), and endoxylanase (acts on hemicellulose cross linking). Such cells transformed with entire pathways and/or enzymes extracted from them, can degrade certain components of biomass. 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.

[00300] Alternately, biomass degradation pathways can be created by transforming host cells with the individual enzymes of the pathway and then combining the cells producing the individual enzymes. This approach allows for the combination of enzymes to more particularly match the biomass of interest by altering the relative ratios of the multiple transformed strains. For example, two times as many cells expressing the first enzyme of a pathway may be added to a mix where the first step of the reaction pathway is the limiting step.

[00301] Following transformation with enzyme-encoding constructs, the host cells and/or organisms are grown. The biomass degrading enzymes may be collected from the organisms/cells. Collection may be by any means known in the art, including, but not limited to concentrating cells, mechanical or chemical disruption of cells, and purification of enzymes from cell cultures and/or cell lysates. Cells and/or organisms can be grown and then the enzyme(s) collected by any means. One method of extracting the enzyme is by harvesting the host cell or a group of host cells and then drying the host cell(s). The enzyme(s) from the dried host cell(s) are then harvested by crushing the cells to expose the enzyme. The whole product of crushed cells is then used to degrade biomass. Many methods of extracting proteins from intact cells are well known in the art, and are also contemplated herein (e.g., introducing an exogenous nucleic acid construct in which an enzyme-encoding sequence is operably linked to a sequence encoding a secretion signal - excreted enzyme is isolated from the growth medium).

[00302] Extracting and utilizing the biomass-degrading enzyme can also be accomplished by expressing a vector containing nucleic acids that encode a biomass production-modulation molecule in the host cell. In this embodiment, the host cell produces the biomass, and also produces a biomass-degrading enzyme. The biomass-degrading enzyme can then degrade the biomass produced by the host cell. In some instances, vector used for the production of a biomass-degrading enzyme may not be continuously active. Such vectors can comprise one or more inducible promoters and one or more biomass-degrading enzymes. Such promoters activate the production of biomass-degrading enzymes, for example, after the biomass has grown to sufficient density or reached certain maturity.

[00303] The present methods can also be performed by introducing a recombinant nucleic acid molecule into a chloroplast, wherein the recombinant nucleic acid molecule includes a first polynucleotide, which encodes at least one polypeptide (i.e., 1, 2, 3, 4, or more). In some embodiments, a polypeptide is operatively linked to a second, third,

fourth, fifth, sixth, seventh, eighth, ninth, tenth and/or subsequent polypeptide. For example, several enzymes in a biodegradation pathway may be linked, either directly or indirectly, such that products produced by one enzyme in the pathway, once produced, are in close proximity to the next enzyme in the pathway.

[00304] Another aspect provides host organisms or cells disclosed herein (e.g. *Scenedesmus sp.* or *Dunaliella sp.*) that have been genetically modified or modified (e.g. by methods disclosed herein) for use as a feedstock. The compositions of genetically modified algae disclosed here can be used directly as a feedstock or can be added to a feedstock to generate a modified or improved feedstock. For example a composition can comprise a feedstock and a genetically modified algae. Genetic modification of an algae can comprise engineering an algae to express one or more enzymes. In some aspects the enzyme can be a biomass degrading enzyme and in some aspects the enzyme can be a biosynthetic enzyme. Genetically modified algae can also express both types of enzymes (e.g. a biomass degrading enzyme and a biosynthetic enzyme). The enzyme expressed can be one that is naturally expressed in the algae or not naturally expressed in the algae. In some aspects the enzyme produced is not naturally expressed in the algae. For example an enzyme (e.g. a biomass degrading enzyme) can be an exogenous enzyme. In another example a composition can comprise a feedstock and a genetically modified algae wherein the algae is modified to increase the expression of a naturally occurring enzyme (e.g. a biomass-degrading enzyme). In some aspects an enzyme can be secreted from a genetically modified algae or added to the feedstock as an independent ingredient.

[00305] Biomass degrading enzymes can improve the nutrient value of an existing feedstock by breaking down complex components of the feedstock (e.g. indigestible components) into components that can be absorbed and used by the animal. A biomass-degrading enzyme can be expressed and retained in the algae or secreted or expelled (i.e. produced ex vivo) from the algae. Genetically modified algae that provide the biomass degrading enzymes can also be utilized by the animal for the inherent nutrient value of the algae. For example a composition can comprise a feedstock, a genetically modified algae, and a biomass-degrading enzyme that is ex vivo to the genetically modified algae. In another example a genetically modified algae is modified to increase expression of a naturally occurring biomass-degrading enzyme.

[00306] The expression of certain exogenous biosynthetic enzymes in an algae can allow the biosynthesis of nutrient rich lipids, fatty acids and carbohydrates. Genetically modified algae that express such nutrient rich components can be added to an existing feedstock to supplement the nutritional value of the feedstock. In some aspects such genetically modified algae can comprise as much as 100% of the feedstock. Algae can be genetically modified to produce or increase production of one or more fatty acids, lipids or hydrocarbons. In one example a genetically modified algae comprise an exogenous nucleic acid encoding an enzyme in an isoprenoid biosynthesis pathway. In some aspects a genetically modified algae can have a higher content, or an altered content, or a different content of, for example, fatty acids, lipids or hydrocarbons (e.g. isoprenoids) than an unmodified algae of the same species. For example, the modified algae can produce more of a desired isoprenoid, and/or produce an isoprenoid that the algae does not normally produce, and/or produce isoprenoids that are normally produced but at different amounts than are produced in an unmodified algae.

[00307] Therefore in one aspect a composition can comprise a feedstock and a genetically modified algae wherein the algae has a higher lipid, fatty acid, or isoprenoid content relative to an unmodified algae of the same species. The biosynthetic enzymes can also be one found in a mevalonate pathway. For example the enzyme can be farnesyl

pyrophosphate synthase, geranyl geranyl phosphate synthase, squalene synthase, thioesterase, or fatty acyl-CoA desaturase.

[00308] An improved feedstock can be comprised entirely or partially of a genetically modified algae. In some aspects a genetically modified algae can be added to a composition to generate an improved feedstock. The composition may not be considered a feedstock suitable for consumption by animals until after the addition of a genetically modified algae. In some aspects a genetically modified algae can be added to an existing feedstock to generate an improved feedstock. In some aspects a genetically modified algae can be added to an existing feedstock at a ratio of at least 1:20 (weight of algae/wt of feedstock). In some aspects an improved feedstock can comprises up to 10, 20, 30, 40, 50, 60, 70, 80, 90, 95, or 100 percent of a genetically modified algae. In some aspects a viable genetically modified algae can be added to a feedstock (e.g. as a seed culture) at a concentration of less than 5% (w/w) of the feedstock wherein the genetically modified algae multiplies to become up to 10, 20, 30, 40, 50, 60, 70, 80, 90 or 95% percent of the feedstock (w/w). A feedstock or improved feedstock can also comprise additional nutrients, ingredients or supplements (e.g. vitamins). An improved feedstock comprising a genetically modified algae can also comprise any normal ingredient of an animal feed including but not limited any vegetable, fruit, seed, root, flower, leaf, stem, stalk or plant product of any plant. An improved feedstock comprising a genetically modified algae can also comprise any animal parts or products (e.g. meat, bone, milk, excrement, skin). An improved feedstock comprising a genetically modified NVPO can also comprise any product or bi-product of a manufacturing process (e.g. sawdust or brewers waste). Additional non limiting examples of ingredients of a feedstock or an improved feedstock as disclosed herein include alfalfa, barley, blood meal, grass, legumes, silage, beet, bone meal, brewer grain, brewer's yeast, broom grass, carrot, cattle manure, clover, coffee, corn, corn gluten meal, distiller grains, poultry fat, grape, hominy feed, hop leaves, spent hops, molasses, oats, algae, peanuts, potato, poultry litter, poultry manure, rape meal, rye, safflower, sorghum, soybean, soy, sunflower meal, timothy hay, or triticale. Therefore in one aspect a composition can comprise a feedstock and a genetically modified NVPO wherein the feedstock comprises one or more of alfalfa, barley, blood meal, beet, bone meal, brewer grain, brewer's yeast, broom grass, carrot, cattle manure, clover, coffee, corn, corn gluten meal, distiller grains, poultry fat, grape, hominy feed, hop leaves, spent hops, molasses, oats, algae, peanuts, potato, poultry litter, poultry manure, rape meal, rye, safflower, sorghum, soybean, soy, sunflower meal, timothy hay, or triticale.

[00309] In some aspect a genetically modified algae can be used for a purpose (e.g. in producing a recombinant product or biofuel) and the remaining portion thereof can be used for an improved feedstock. Therefore an improved feedstock can comprise a portion of a genetically modified algae. For example a composition of an animal feed ingredient can comprise of whole and/or defatted algae (e.g. after removal of fatty acids, lipids or hydrocarbons, e.g. after hexane extraction) or a mixture of whole and defatted algae, which provides both the feed enzyme and the inherent nutritive value of the algae. In another example a genetically modified algae can be washed, dehydrated, centrifuged, filtered, defatted, lysed, dried, processed (e.g. extracted), or milled. The remaining portion thereof can be used as a feedstock, as an improved feedstock or as a supplement to improve a feedstock. For example a composition can comprise a feedstock and a portion of a genetically modified algae wherein the genetically modified algae is at least partially depleted of a lipid, fatty acid, isoprenoid, carotenoid, carbohydrate, or selected protein. The genetically modified algae can also be genetically modified to produce a biomass-degrading enzyme as disclosed herein.

[00310] Methods of generating, modifying, supplementing or improving a feedstock composition are also disclosed herein. The methods can comprise combining a genetically modified algae or a portion thereof with a feedstock to

generate the improved feedstock. In one example the method comprises removing a lipid, fatty acid, isoprenoid, or carbohydrate from a genetically modified algae. The remaining genetically modified algae, or a portion thereof, can be combined with a feedstock to generate the improved feedstock composition. In one example of the method the modified algae does not express an exogenous phytase. The genetically modified algae or a portion thereof can comprise a nucleic acid (e.g. an exogenous nucleic acid). The nucleic acid can be a vector. In one example of the method, the nucleic acid encodes a biomass degrading enzyme. The biomass-degrading enzyme can be a galactanase, xylanase, protease, carbohydrase, lipase, reductase, oxidase, transglutaminase, or phytase. The biomass-degrading enzyme can be a carbohydrase wherein the carbohydrase is an α -amylase, β -amylase, endo- β -glucanase, endoxylanase, β -mannanase, α -galactosidase, or pullulanase. The biomass-degrading enzyme can be a protease wherein the protease is a subtilisin, bromelain, or fungal acid-stable protease. The biomass-degrading enzyme can be a phytase. In another example of the method the genetically modified NVPO further comprises an exogenous nucleic acid encoding an enzyme in an isoprenoid biosynthesis pathway. The enzyme in the isoprenoid biosynthesis pathway can be farnesyl pyrophosphate synthase, geranyl geranyl phosphate synthase, squalene synthase, thioesterase, or fatty acyl-CoA desaturase. The enzyme in the isoprenoid biosynthesis pathway can be in a mevalonate pathway. In yet another example of the method, the method can further comprise removing a lipid, fatty acid, or isoprenoid, from the genetically modified NVPO prior to combining with a feedstock to generate the improved feedstock.

[00311] Candidate genes for directing the expression of proteins (e.g. enzymes) in genetically modified algae for use in animal feeds can be obtained from a variety of organisms including eukaryotes, prokaryotes, or viruses. In some instances, an expressed enzyme is one member of a metabolic pathway (e.g. an isoprenoid biosynthesis pathway). Several enzymes may be introduced into the algae to produce increased levels of desired metabolites, or several enzymes may be introduced to produce a algae containing multiple useful feed enzyme activities (e.g. simultaneous production of xylanase, endo- β -glucanase, and phytase activities).

[00312] Feed enzymes can be expressed in host organisms (e.g. *Scenedesmus sp.*) and purified to a useful level. The purified enzymes can be added to animal feed in a manner similar to current practice. Feed enzymes can also be expressed in host organisms (e.g. algae), and the resulting host organisms can be added as a feed ingredient, adding both nutritive value and desired enzyme activity to the animal feed product. In this application, the genetically modified host organisms can be added to a feedstock alive, whole and non-viable or as a lysate wherein the host organisms are lysed by any suitable means (e.g. physical, chemical or thermal).

[00313] Many animal feeds can contain plant seeds, including soybeans, maize, wheat, and barley among others. Plant seeds can contain high levels of myo-inositol polyphosphate (phytic acid). This phytic acid is indigestible to non-ruminant animals, and so feeds with high levels of phytic acid may have low levels of bioavailable phosphorous. The phytic acid can also chelate many important nutritive minerals, such as calcium and magnesium. Incorporation of a phytase into the feed, which can act in the animals upper gut, can release both the chelated mineral nutrients and significant levels of bioavailable phosphorous. The net result is that less free phosphorous needs to be added to the animal feed product. In addition, phosphorous levels in the excreta can be reduced, which can reduce downstream phosphorous pollution.

[00314] Genetically modified algae that express phytases or similar enzymes can be added to a feedstock to improve the nutrient or digestible properties of the feed. Phytases contemplated for use herein can be from any organism (e.g. bacterial or fungal derived). Non limiting examples of types of phytases contemplated for use herein include 3-phytase

(alternative name 1- phytase; a myo-inositol hexaphosphate 3-phosphohydrolase, EC 3.1.3.8), 4- phytase (alternative name 6-phytase, name based on 1 L-numbering system and not I D- numbering, EC 3.1.3.26), and 5-phytase (EC 3.1.3.72). Additional non limiting examples of phytases include microbial phytases, such as fungal, yeast or bacterial phytases such as disclosed in EP 684313, US 6139902, EP 420358, WO 97/35017, WO 98/28408, WO 98/28409, JP 1 1000164, WO98/13480, AU 724094, WO 97/33976, US 6110719, WO 2006/038062, WO 2006/038128, WO 2004/085638, WO 2006/037328, WO 2006/037327, WO 2006/043178, US 5830732 and under UniProt designations P34753, P34752, P34755, 000093, 031097, P42094, 066037 and P34754 (UniProt, (2008) <http://www.uniprot.org/>). Polypeptides having an amino acid sequence of at least 75% identity to an amino acid sequence (comprising the active site) of any one of the phytases disclosed above are also contemplated for use herein. In one example a composition can comprise a feedstock and a genetically modified algae. The genetically modified algae can be genetically modified to produce a biomass-degrading enzyme such as a phytase. In one aspect the phytase is a phytase of bacterial or fungal origin. In one aspect the biomass-degrading enzyme is an enzyme other than a phytase.

[00315] Many plant parts (e.g. seeds, fruits, stems, roots, leaves and flowers) from plants such as, for example, soybeans, wheat, and barley contain polysaccharides that are indigestible by some animals (e.g. non-ruminant animals). Non limiting examples of such carbohydrates include xylans, raffinose, stachyose, and glucans. The presence of indigestible carbohydrates in animal feed can reduce nutrient availability. Indigestible carbohydrates in poultry feed can result in sticky feces, which can increase disease levels. The presence of one or more carbohydrate degrading enzymes (e.g. α -amylase) in the animal feed can help break down polysaccharides, increase nutrient availability, increase the bio-available energy content of the animal feed, and reduce health risks. Non limiting examples of carbohydrate degrading enzymes contemplated for use herein include amylases (e.g. α -amylase and β -amylase), β -mannanase, maltase, lactase, β -glucanase, endo- β -glucanase, glucose isomerase, endoxylanase, α -galactosidase, glucose oxidase, pullulanase, invertase and any carbohydrate digesting enzyme of bacterial, fungal, plant or animal origin. In one example a composition can comprise a feedstock and a genetically modified algae. The genetically modified algae can be genetically modified to produce a biomass-degrading enzyme such as a carbohydrase. In one aspect the carbohydrase can be an α -amylase, β -amylase, endo- β -glucanase, endoxylanase, β -mannanase, α -galactosidase, or pullulanase.

[00316] Many feedstocks contain plant parts (e.g. seeds) with anti-nutritive proteins (e.g. protease inhibitors, amylase inhibitors and others) that reduce the availability of nutrients in an animal feed. Addition of a broad spectrum protease (e.g. bromelain, subtilisin, or a fungal acid-stable protease) can break down these anti-nutritive proteins and increase the availability of nutrients in the animal's feed. Non limiting examples of proteases contemplated for use herein include endopeptidases and exopeptidases. Non limiting examples of proteases contemplated for use herein include serine proteases (e.g. subtilisin, chymotrypsins, glutamyl peptidases, dipeptidyl-peptidases, carboxypeptidases, dipeptidases, and aminopeptidases), cysteine proteases (e.g. papain, calpain-2, and papain-like peptidases and bromelain), aspartic peptidases (e.g. pepsins and pepsin A), glutamic proteases, threonine proteases, fungal acid proteases and acid stable proteases such as those disclosed in (US 6855548). In one example a composition can comprise a feedstock and a genetically modified algae. The genetically modified algae can be genetically modified to produce a biomass-degrading enzyme such as a protease. In one aspect the protease can be a subtilisin, bromelain or fungal acid-stable protease.

[00317] Non limiting examples of lipases contemplated for use herein include pancreatic lipase, lysosomal lipase, lysosomal acid lipase, acid cholesteryl ester hydrolase, hepatic lipase, lipoprotein lipase, gastric lipase, endothelial lipase, pancreatic lipase related protein 2, pancreatic lipase related protein 1, lingual lipase and phospholipases (e.g.

phospholipase A1 (EC 3.1.1.32), phospholipase A2, phospholipase B (lysophospholipase), phospholipase C and phospholipase D).

[00318] An improved feedstock can be generated by combining a feedstock with a algae that is genetically altered to produce an enzyme (e.g. a carbohydrase, protease or lipase). In some aspects the enzyme is produced ex vivo to the organism. In some aspects the enzyme is secreted. Enzymes produced ex vivo to the organisms can break down components of a feedstock prior to ingestion by an animal. Therefore an improved feedstock can be generated by combining a feedstock with an algae that is genetically altered to produce an enzyme (e.g. a carbohydrase, protease or lipase) and subjecting the mixture to a holding period. A holding period can allow the genetically altered algae to multiply and to secrete more enzyme into the feedstock. A holding period can be from several hours up to several days. In some aspects a holding period is for up to 1, 2, 3, 4, 5, 6, 7, 8, 9 or 10 days. In some aspects a holding period is for up to several days to several weeks. In some aspects a holding period is indefinite. An indefinite holding period allows intermittent removal and use of the improved feedstock and intermittent addition of the base feedstock.

[00319] Host Cells or Host Organisms

[00320] Biomass useful in the methods and systems described herein can be obtained from host cells or host organisms.

[00321] A host cell can contain a polynucleotide encoding a polypeptide of the present disclosure. In some embodiments, a host cell is part of a multicellular organism. In other embodiments, a host cell is cultured as a unicellular organism.

[00322] Host organisms can include any suitable host, for example, a microorganism. Microorganisms which are useful for the methods described herein include, for example, photosynthetic bacteria (e.g., cyanobacteria), non-photosynthetic bacteria (e.g., *E. coli*), yeast (e.g., *Saccharomyces cerevisiae*), and algae (e.g., microalgae such as *Chlamydomonas reinhardtii*).

[00323] Examples of host organisms that can be transformed with a polynucleotide of interest (for example, a polynucleotide that encodes a protein involved in the isoprenoid biosynthesis pathway) include vascular and non-vascular organisms. The organism can be prokaryotic or eukaryotic. The organism can be unicellular or multicellular. A host organism is an organism comprising a host cell. In other embodiments, the host organism is photosynthetic. A photosynthetic organism is one that naturally photosynthesizes (e.g., an alga) or that is genetically engineered or otherwise modified to be photosynthetic. In some instances, a photosynthetic organism may be transformed with a construct or vector of the disclosure which renders all or part of the photosynthetic apparatus inoperable.

[00324] By way of example, a non-vascular photosynthetic microalga species (for example, *C. reinhardtii*, *Nannochloropsis oceanica*, *N. salina*, *D. salina*, *H. pluvialis*, *S. dimorphus*, *D. viridis*, *Chlorella* sp., and *D. tertiolecta*) can be genetically engineered to produce a polypeptide of interest, for example a fusicoccadiene synthase or an FPP synthase. Production of a fusicoccadiene synthase or an FPP synthase in these microalgae can be achieved by engineering the microalgae to express the fusicoccadiene synthase or FPP synthase in the algal chloroplast or nucleus.

[00325] In other embodiments the host organism is a vascular plant. Non-limiting examples of such plants include various monocots and dicots, including high oil seed plants such as high oil seed *Brassica* (e.g., *Brassica nigra*, *Brassica napus*, *Brassica hirta*, *Brassica rapa*, *Brassica campestris*, *Brassica carinata*, and *Brassica juncea*), soybean (*Glycine max*), castor bean (*Ricinus communis*), cotton, safflower (*Carthamus tinctorius*), sunflower (*Helianthus annuus*), flax (*Linum usitatissimum*), corn (*Zea mays*), coconut (*Cocos nucifera*), palm (*Elaeis guineensis*), oil nut trees

such as olive (*Olea europaea*), sesame, and peanut (*Arachis hypogaea*), as well as *Arabidopsis*, tobacco, wheat, barley, oats, amaranth, potato, rice, tomato, and legumes (e.g., peas, beans, lentils, alfalfa, etc.).

[00326] The host cell can be prokaryotic. Examples of some prokaryotic organisms of the present disclosure include, but are not limited to, cyanobacteria (e.g., *Synechococcus*, *Synechocystis*, *Athrospira*, *Gleocapsa*, *Oscillatoria*, and, *Pseudoanabaena*). Suitable prokaryotic cells include, but are not limited to, any of a variety of laboratory strains of *Escherichia coli*, *Lactobacillus* sp., *Salmonella* sp., and *Shigella* sp. (for example, as described in Carrier et al. (1992) J. Immunol. 148:1176-1181; U.S. Pat. No. 6,447,784; and Sizemore et al. (1995) Science 270:299-302). Examples of *Salmonella* strains which can be employed in the present disclosure include, but are not limited to, *Salmonella typhi* and *S. typhimurium*. Suitable *Shigella* strains include, but are not limited to, *Shigella flexneri*, *Shigella sonnei*, and *Shigella dysenteriae*. Typically, the laboratory strain is one that is non-pathogenic. Non-limiting examples of other suitable bacteria include, but are not limited to, *Pseudomonas putida*, *Pseudomonas aeruginosa*, *Pseudomonas mevalonii*, *Rhodobacter sphaeroides*, *Rhodobacter capsulatus*, *Rhodospirillum rubrum*, and *Rhodococcus* sp.

[00327] In some embodiments, the host organism is eukaryotic (e.g. green algae, red algae, brown algae). In some embodiments, the algae is a green algae, for example, a Chlorophyceae. The algae can be unicellular or multicellular. Suitable eukaryotic host cells include, but are not limited to, yeast cells, insect cells, plant cells, fungal cells, and algal cells. Suitable eukaryotic host cells include, but are not limited to, *Pichia pastoris*, *Pichia finlandica*, *Pichia trehalophila*, *Pichia kockii*, *Pichia membranaefaciens*, *Pichia opuntiae*, *Pichia thermotolerans*, *Pichia salictaria*, *Pichia guercuum*, *Pichia pijperi*, *Pichia stipitis*, *Pichia methanolica*, *Pichia* sp., *Saccharomyces cerevisiae*, *Saccharomyces* sp., *Hansenula polymorpha*, *Kluyveromyces* sp., *Kluyveromyces fragilis*, *Candida albicans*, *Aspergillus nidulans*, *Aspergillus niger*, *Aspergillus oryzae*, *Trichoderma reesei*, *Chrysosporium lucknowense*, *Fusarium* sp., *Fusarium gramineum*, *Fusarium venenatum*, *Neurospora crassa*, and *Chlamydomonas reinhardtii*. In other embodiments, the host cell is a microalga (e.g., *Chlamydomonas reinhardtii*, *Dunaliella salina*, *Haematococcus pluvialis*, *Nannochloropsis oceanica*, *N. salina*, *Scenedesmus dimorphus*, *Chlorella* spp., *D. viridis*, or *D. tertiolecta*).

[00328] In some instances the organism is a rhodophyte, chlorophyte, heterokontophyte, tribophyte, glaucophyte, chlorarachniophyte, euglenoid, haptophyte, cryptomonad, dinoflagellum, or phytoplankton.

[00329] In some instances a host organism is vascular and photosynthetic. Examples of vascular plants include, but are not limited to, angiosperms, gymnosperms, rhyniophytes, or other tracheophytes.

[00330] In some instances a host organism is non-vascular and photosynthetic. As used herein, the term "non-vascular photosynthetic organism," refers to any macroscopic or microscopic organism, including, but not limited to, algae, cyanobacteria and photosynthetic bacteria, which does not have a vascular system such as that found in vascular plants. 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. For example, the microalgae *Chlamydomonas reinhardtii* may be transformed with a vector, or a linearized portion thereof, encoding one or more proteins of interest (e.g., a protein involved in the isoprenoid biosynthesis pathway).

[00331] Methods for algal transformation are described in U.S. Provisional Patent Application No. 60/142,091. The methods of the present disclosure can be carried out using algae, for example, the microalga, *C. reinhardtii*. The use of microalgae to express a polypeptide or protein complex according to a method of the disclosure 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.

[00332] The vectors of the present disclosure may be capable of stable or transient transformation of multiple photosynthetic organisms, including, but not limited to, photosynthetic bacteria (including cyanobacteria), cyanophyta, prochlorophyta, rhodophyta, chlorophyta, heterokontophyta, tribophyta, glaucophyta, chlorarachniophytes, euglenophyta, euglenoids, haptophyta, chrysophyta, cryptophyta, cryptomonads, dinophyta, dinoflagellata, pyrmnesiophyta, bacillariophyta, xanthophyta, eustigmatophyta, raphidophyta, phaeophyta, and phytoplankton. Other vectors of the present disclosure are capable of stable or transient transformation of, for example, *C. reinhardtii*, *N. oceanica*, *N. salina*, *D. salina*, *H. pluvalis*, *S. dimorphus*, *D. viridis*, or *D. tertiolecta*.

[00333] Examples of appropriate hosts, include but are not limited to: bacterial cells, such as *E. coli*, *Streptomyces*, *Salmonella typhimurium*; fungal cells, such as yeast; insect cells, such as *Drosophila* S2 and *Spodoptera* Sf9; animal cells, such as CHO, COS or Bowes melanoma; adenoviruses; and plant cells. The selection of an appropriate host is deemed to be within the scope of those skilled in the art.

[00334] Polynucleotides selected and isolated as described herein are introduced into a suitable host cell. A suitable host cell is any cell which is capable of promoting recombination and/or reductive reassortment. The selected polynucleotides can be, for example, in a vector which includes appropriate control sequences. The host cell can be, for example, a higher eukaryotic cell, such as a mammalian cell, or a lower eukaryotic cell, such as a yeast cell, or the host cell can be a prokaryotic cell, such as a bacterial cell. Introduction of a construct (vector) into the host cell can be effected by, for example, calcium phosphate transfection, DEAE-Dextran mediated transfection, or electroporation.

[00335] Recombinant polypeptides, including protein complexes, can be expressed in plants, allowing for the production of crops of such plants and, therefore, the ability to conveniently produce large amounts of a desired product. Accordingly, the methods of the disclosure can be practiced using any plant, including, for example, microalga and macroalgae, (such as marine algae and seaweeds), as well as plants that grow in soil.

[00336] In one embodiment, the host cell is a plant. The term "plant" is used broadly herein to refer to a eukaryotic organism containing plastids, such as 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, and roots. A part of a plant useful for propagation includes, for example, seeds, fruits, cuttings, seedlings, tubers, and rootstocks.

[00337] A method of the disclosure can generate a plant containing genomic DNA (for example, a nuclear and/or plastid genomic DNA) that is genetically modified to contain a stably integrated polynucleotide (for example, as described in Hager and Bock, *Appl. Microbiol. Biotechnol.* 54:302-310, 2000). Accordingly, the present disclosure

further provides a transgenic plant, e.g. *C. reinhardtii*, which comprises one or more chloroplasts containing a polynucleotide encoding one or more exogenous or endogenous polypeptides, including polypeptides that can allow for secretion of fuel products and/or fuel product precursors (e.g., isoprenoids, fatty acids, lipids, triglycerides). A photosynthetic organism of the present disclosure comprises at least one host cell that is modified to generate, for example, a fuel product or a fuel product precursor.

[00338] Some of the host organisms useful in the disclosed embodiments are, for example, are extremophiles, such as hyperthermophiles, psychrophiles, psychrotrophs, halophiles, barophiles and acidophiles. Some of the host organisms which may be used to practice the present disclosure are halophilic (e.g., *Dunaliella salina*, *D. viridis*, or *D. tertiolecta*). For example, *D. salina* can grow in ocean water and salt lakes (for example, salinity from 30-300 parts per thousand) and high salinity media (e.g., artificial seawater medium, seawater nutrient agar, brackish water medium, and seawater medium). In some embodiments of the disclosure, a host cell expressing a protein of the present disclosure can be grown in a liquid environment which is, for example, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0, 1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9, 2.0, 2.1, 2.2, 2.3, 2.4, 2.5, 2.6, 2.7, 2.8, 2.9, 3.0, 3.1, 3.2, 3.3, 3.4, 3.5, 3.6, 3.7, 3.8, 3.9, 4.0, 4.1, 4.2, 4.3 molar or higher concentrations of sodium chloride. One of skill in the art will recognize that other salts (sodium salts, calcium salts, potassium salts, or other salts) may also be present in the liquid environments.

[00339] Where a halophilic organism is utilized for the present disclosure, it may be transformed with any of the vectors described herein. For example, *D. salina* may be transformed with a vector which is capable of insertion into the chloroplast or nuclear genome and which contains nucleic acids which encode a protein (e.g., an FPP synthase or a fusicoccadiene synthase). Transformed halophilic organisms may then be grown in high-saline environments (e.g., salt lakes, salt ponds, and high-saline media) to produce the products (e.g., lipids) of interest. Isolation of the products may involve removing a transformed organism from a high-saline environment prior to extracting the product from the organism. In instances where the product is secreted into the surrounding environment, it may be necessary to desalinate the liquid environment prior to any further processing of the product.

[00340] The present disclosure further provides compositions comprising a genetically modified host cell. A composition comprises a genetically modified host cell; and will in some embodiments comprise one or more further components, which components are selected based in part on the intended use of the genetically modified host cell. Suitable components include, but are not limited to, salts; buffers; stabilizers; protease-inhibiting agents; cell membrane- and/or cell wall-preserving compounds, e.g., glycerol and dimethylsulfoxide; and nutritional media appropriate to the cell.

[00341] For the production of a protein, for example, an isoprenoid or isoprenoid precursor compound, a host cell can be, for example, one that produces, or has been genetically modified to produce, one or more enzymes in a prenyl transferase pathway and/or a mevalonate pathway and/or an isoprenoid biosynthetic pathway. In some embodiments, the host cell is one that produces a substrate of a prenyl transferase, isoprenoid synthase or mevalonate pathway enzyme.

[00342] In some embodiments, a genetically modified host cell is a host cell that comprises an endogenous mevalonate pathway and/or isoprenoid biosynthetic pathway and/or prenyl transferase pathway. In other embodiments, a genetically modified host cell is a host cell that does not normally produce mevalonate or IPP via a mevalonate pathway, or FPP, GPP or GGPP via a prenyl transferase pathway, but has been genetically modified with one or more polynucleotides comprising nucleotide sequences encoding one or more mevalonate pathway, isoprenoid synthase pathway or prenyl

transferase pathway enzymes (for example, as described in U.S. Patent Publication No. 2004/005678; U.S. Patent Publication No. 2003/0148479; and Martin et al. (2003) Nat. Biotech. 21(7):796-802).

[00343] Culturing of Cells or Organisms

[00344] An organism may be grown under conditions which permit photosynthesis, however, this is not a requirement (e.g., a host organism may be grown in the absence of light). In some instances, the host organism may be genetically modified in such a way that its photosynthetic capability is diminished or destroyed. In growth conditions where a host organism is not capable of photosynthesis (e.g., because of the absence of light and/or genetic modification), typically, the organism will be provided with the necessary nutrients to support growth in the absence of photosynthesis. For example, a culture medium in (or on) which an organism is grown, may be supplemented with any required nutrient, including an organic carbon source, nitrogen source, phosphorous source, vitamins, metals, lipids, nucleic acids, micronutrients, and/or an organism-specific requirement. Organic carbon sources include any source of carbon which the host organism is able to metabolize including, but not limited to, acetate, simple carbohydrates (e.g., glucose, sucrose, and lactose), complex carbohydrates (e.g., starch and glycogen), proteins, and lipids. One of skill in the art will recognize that not all organisms will be able to sufficiently metabolize a particular nutrient and that nutrient mixtures may need to be modified from one organism to another in order to provide the appropriate nutrient mix.

[00345] Optimal growth of organisms occurs usually at a temperature of about 20°C to about 25°C, although some organisms can still grow at a temperature of up to about 35°C. Active growth is typically performed in liquid culture. If the organisms are grown in a liquid medium and are shaken or mixed, the density of the cells can be anywhere from about 1 to 5×10^8 cells/ml at the stationary phase. For example, the density of the cells at the stationary phase for *Chlamydomonas* sp. can be about 1 to 5×10^7 cells/ml; the density of the cells at the stationary phase for *Nannochloropsis* sp. can be about 1 to 5×10^8 cells/ml; the density of the cells at the stationary phase for *Scenedesmus* sp. can be about 1 to 5×10^7 cells/ml; and the density of the cells at the stationary phase for *Chlorella* sp. can be about 1 to 5×10^8 cells/ml. Exemplary cell densities at the stationary phase are as follows: *Chlamydomonas* sp. can be about 1×10^7 cells/ml; *Nannochloropsis* sp. can be about 1×10^8 cells/ml; *Scenedesmus* sp. can be about 1×10^7 cells/ml; and *Chlorella* sp. can be about 1×10^8 cells/ml. An exemplary growth rate may yield, for example, a two to four fold increase in cells per day, depending on the growth conditions. In addition, doubling times for organisms can be, for example, 5 hours to 30 hours. The organism can also be grown on solid media, for example, media containing about 1.5% agar, in plates or in slants.

[00346] One source of energy is fluorescent light that can be placed, for example, at a distance of about 1 inch to about two feet from the organism. Examples of types of fluorescent lights includes, for example, cool white and daylight. Bubbling with air or CO₂ improves the growth rate of the organism. Bubbling with CO₂ can be, for example, at 1% to 5% CO₂. If the lights are turned on and off at regular intervals (for example, 12:12 or 14:10 hours of light:dark) the cells of some organisms will become synchronized.

[00347] Long term storage of organisms can be achieved by streaking them onto plates, sealing the plates with, for example, Parafilm™, and placing them in dim light at about 10°C to about 18°C. Alternatively, organisms may be grown as streaks or stabs into agar tubes, capped, and stored at about 10°C to about 18°C. Both methods allow for the storage of the organisms for several months.

[00348] For longer storage, the organisms can be grown in liquid culture to mid to late log phase and then supplemented with a penetrating cryoprotective agent like DMSO or MeOH, and stored at less than -130°C. An

exemplary range of DMSO concentrations that can be used is 5 to 8%. An exemplary range of MeOH concentrations that can be used is 3 to 9%.

[00349] Organisms can be grown on a defined minimal medium (for example, high salt medium (HSM), modified artificial sea water medium (MASM), or F/2 medium) with light as the sole energy source. In other instances, the organism can be grown in a medium (for example, tris acetate phosphate (TAP) medium), and supplemented with an organic carbon source.

[00350] Organisms, such as algae, can grow naturally in fresh water or marine water. Culture media for freshwater algae can be, for example, synthetic media, enriched media, soil water media, and solidified media, such as agar. Various culture media have been developed and used for the isolation and cultivation of fresh water algae and are described in Watanabe, M.W. (2005). *Freshwater Culture Media*. In R.A. Andersen (Ed.), *Algal Culturing Techniques* (pp. 13-20). Elsevier Academic Press. Culture media for marine algae can be, for example, artificial seawater media or natural seawater media. Guidelines for the preparation of media are described in Harrison, P.J. and Berges, J.A. (2005). *Marine Culture Media*. In R.A. Andersen (Ed.), *Algal Culturing Techniques* (pp. 21-33). Elsevier Academic Press.

[00351] Organisms may be grown in outdoor open water, such as ponds, the ocean, seas, rivers, waterbeds, marshes, shallow pools, lakes, aqueducts, and reservoirs. When grown in water, the organism can be contained in a halo-like object comprised of lego-like particles. The halo-like object encircles the organism and allows it to retain nutrients from the water beneath while keeping it in open sunlight.

[00352] In some instances, organisms can be grown in containers wherein each container comprises one or two organisms, or a plurality of organisms. The containers can be configured to float on water. For example, a container can be filled by a combination of air and water to make the container and the organism(s) in it buoyant. An organism that is adapted to grow in fresh water can thus be grown in salt water (i.e., the ocean) and vice versa. This mechanism allows for automatic death of the organism if there is any damage to the container.

[00353] Culturing techniques for algae are well known to one of skill in the art and are described, for example, in *Freshwater Culture Media*. In R.A. Andersen (Ed.), *Algal Culturing Techniques*. Elsevier Academic Press.

[00354] Because photosynthetic organisms, for example, algae, require sunlight, CO₂ and water for growth, they can be cultivated in, for example, open ponds and lakes. However, these open systems are more vulnerable to contamination than a closed system. One challenge with using an open system is that the organism of interest may not grow as quickly as a potential invader. This becomes a problem when another organism invades the liquid environment in which the organism of interest is growing, and the invading organism has a faster growth rate and takes over the system.

[00355] In addition, in open systems there is less control over water temperature, CO₂ concentration, and lighting conditions. The growing season of the organism is largely dependent on location and, aside from tropical areas, is limited to the warmer months of the year. In addition, in an open system, the number of different organisms that can be grown is limited to those that are able to survive in the chosen location. An open system, however, is cheaper to set up and/or maintain than a closed system.

[00356] Another approach to growing an organism is to use a semi-closed system, such as covering the pond or pool with a structure, for example, a "greenhouse-type" structure. While this can result in a smaller system, it addresses many of the problems associated with an open system. The advantages of a semi-closed system are that it can allow for a greater number of different organisms to be grown, it can allow for an organism to be dominant over an invading organism by allowing the organism of interest to out compete the invading organism for nutrients required for its

growth, and it can extend the growing season for the organism. For example, if the system is heated, the organism can grow year round.

[00357] A variation of the pond system is an artificial pond, for example, a raceway pond. In these ponds, the organism, water, and nutrients circulate around a "racetrack." Paddlewheels provide constant motion to the liquid in the racetrack, allowing for the organism to be circulated back to the surface of the liquid at a chosen frequency. Paddlewheels also provide a source of agitation and oxygenate the system. These raceway ponds can be enclosed, for example, in a building or a greenhouse, or can be located outdoors.

[00358] Raceway ponds are usually kept shallow because the organism needs to be exposed to sunlight, and sunlight can only penetrate the pond water to a limited depth. The depth of a raceway pond can be, for example, about 4 to about 12 inches. In addition, the volume of liquid that can be contained in a raceway pond can be, for example, about 200 liters to about 600,000 liters.

[00359] The raceway ponds can be operated in a continuous manner, with, for example, CO₂ and nutrients being constantly fed to the ponds, while water containing the organism is removed at the other end.

[00360] If the raceway pond is placed outdoors, there are several different ways to address the invasion of an unwanted organism. For example, the pH or salinity of the liquid in which the desired organism is in can be such that the invading organism either slows down its growth or dies.

[00361] Also, chemicals can be added to the liquid, such as bleach, or a pesticide can be added to the liquid, such as glyphosate. In addition, the organism of interest can be genetically modified such that it is better suited to survive in the liquid environment. Any one or more of the above strategies can be used to address the invasion of an unwanted organism.

[00362] Alternatively, organisms, such as algae, can be grown in closed structures such as photobioreactors, where the environment is under stricter control than in open systems or semi-closed systems. A photobioreactor is a bioreactor which incorporates some type of light source to provide photonic energy input into the reactor. The term photobioreactor can refer to a system closed to the environment and having no direct exchange of gases and contaminants with the environment. A photobioreactor can be described as an enclosed, illuminated culture vessel designed for controlled biomass production of phototrophic liquid cell suspension cultures. Examples of photobioreactors include, for example, glass containers, plastic tubes, tanks, plastic sleeves, and bags. Examples of light sources that can be used to provide the energy required to sustain photosynthesis include, for example, fluorescent bulbs, LEDs, and natural sunlight. Because these systems are closed everything that the organism needs to grow (for example, carbon dioxide, nutrients, water, and light) must be introduced into the bioreactor.

[00363] Photobioreactors, despite the costs to set up and maintain them, have several advantages over open systems, they can, for example, prevent or minimize contamination, permit axenic organism cultivation of monocultures (a culture consisting of only one species of organism), offer better control over the culture conditions (for example, pH, light, carbon dioxide, and temperature), prevent water evaporation, lower carbon dioxide losses due to out gassing, and permit higher cell concentrations.

[00364] On the other hand, certain requirements of photobioreactors, such as cooling, mixing, control of oxygen accumulation and biofouling, make these systems more expensive to build and operate than open systems or semi-closed systems.

[00365] Photobioreactors can be set up to be continually harvested (as is with the majority of the larger volume cultivation systems), or harvested one batch at a time (for example, as with polyethylene bag cultivation). A batch photobioreactor is set up with, for example, nutrients, an organism (for example, algae), and water, and the organism is allowed to grow until the batch is harvested. A continuous photobioreactor can be harvested, for example, either continually, daily, or at fixed time intervals.

[00366] High density photobioreactors are described in, for example, Lee, et al., *Biotech. Bioengineering* 44:1161-1167, 1994. Other types of bioreactors, such as those for sewage and waste water treatments, are described in, Sawayama, et al., *Appl. Micro. Biotech.*, 41:729-731, 1994. Additional examples of photobioreactors are described in, U.S. Appl. Publ. No. 2005/0260553, U.S. Pat. No. 5,958,761, and U.S. Pat. No. 6,083,740. Also, organisms, such as algae may be mass-cultured for the removal of heavy metals (for example, as described in Wilkinson, *Biotech. Letters*, 11:861-864, 1989), hydrogen (for example, as described in U.S. Patent Application Publication No. 2003/0162273), and pharmaceutical compounds from a water, soil, or other source or sample. Organisms can also be cultured in conventional fermentation bioreactors, which include, but are not limited to, batch, fed-batch, cell recycle, and continuous fermentors. Additional methods of culturing organisms and variations of the methods described herein are known to one of skill in the art.

[00367] Organisms can also be grown near ethanol production plants or other facilities or regions (e.g., cities and highways) generating CO₂. As such, the methods herein contemplate business methods for selling carbon credits to ethanol plants or other facilities or regions generating CO₂ while making fuels or fuel products by growing one or more of the organisms described herein near the ethanol production plant, facility, or region.

[00368] The organism of interest, grown in any of the systems described herein, can be, for example, continually harvested, or harvested one batch at a time.

[00369] CO₂ can be delivered to any of the systems described herein, for example, by bubbling in CO₂ from under the surface of the liquid containing the organism. Also, spargers can be used to inject CO₂ into the liquid. Spargers are, for example, porous disc or tube assemblies that are also referred to as Bubblers, Carbonators, Aerators, Porous Stones and Diffusers.

[00370] Nutrients that can be used in the systems described herein include, for example, nitrogen (in the form of NO₃⁻ or NH₄⁺), phosphorus, and trace metals (Fe, Mg, K, Ca, Co, Cu, Mn, Mo, Zn, V, and B). The nutrients can come, for example, in a solid form or in a liquid form. If the nutrients are in a solid form they can be mixed with, for example, fresh or salt water prior to being delivered to the liquid containing the organism, or prior to being delivered to a photobioreactor.

[00371] Organisms can be grown in cultures, for example large scale cultures, where large scale cultures refers to growth of cultures in volumes of greater than about 6 liters, or greater than about 10 liters, or greater than about 20 liters. Large scale growth can also be growth of cultures in volumes of 50 liters or more, 100 liters or more, or 200 liters or more. Large scale growth can be growth of cultures in, for example, ponds, containers, vessels, or other areas, where the pond, container, vessel, or area that contains the culture is for example, at least 5 square meters, at least 10 square meters, at least 200 square meters, at least 500 square meters, at least 1,500 square meters, at least 2,500 square meters, in area, or greater.

[00372] *Chlamydomonas sp.*, *Nannochloropsis sp.*, *Scenedesmus sp.*, and *Chlorella sp.* are exemplary algae that can be cultured as described herein and can grow under a wide array of conditions.

[00373] One organism that can be cultured as described herein is a commonly used laboratory species *C. reinhardtii*. Cells of this species are haploid, and can grow on a simple medium of inorganic salts, using photosynthesis to provide energy. This organism can also grow in total darkness if acetate is provided as a carbon source. *C. reinhardtii* can be readily grown at room temperature under standard fluorescent lights. In addition, the cells can be synchronized by placing them on a light-dark cycle. Other methods of culturing *C. reinhardtii* cells are known to one of skill in the art.

[00374] Polynucleotides and Polypeptides

[00375] Also provided are isolated polynucleotides encoding a protein, for example, an FPP synthase, described herein. As used herein "isolated polynucleotide" means a polynucleotide that is free of one or both of the nucleotide sequences which flank the polynucleotide in the naturally-occurring genome of the organism from which the polynucleotide is derived. The term includes, for example, a polynucleotide or fragment thereof that is incorporated into a vector or expression cassette; into an autonomously replicating plasmid or virus; into the genomic DNA of a prokaryote or eukaryote; or that exists as a separate molecule independent of other polynucleotides. It also includes a recombinant polynucleotide that is part of a hybrid polynucleotide, for example, one encoding a polypeptide sequence.

[00376] The novel proteins of the present disclosure can be made by any method known in the art. The protein may be synthesized using either solid-phase peptide synthesis or by classical solution peptide synthesis also known as liquid-phase peptide synthesis. Using Val-Pro-Pro, Enalapril and Lisinopril as starting templates, several series of peptide analogs such as X-Pro-Pro, X-Ala-Pro, and X-Lys-Pro, wherein X represents any amino acid residue, may be synthesized using solid-phase or liquid-phase peptide synthesis. Methods for carrying out liquid phase synthesis of libraries of peptides and oligonucleotides coupled to a soluble oligomeric support have also been described. Bayer, Ernst and Mutter, Manfred, *Nature* 237:512-513 (1972) ; Bayer, Ernst, et al., *J. Am. Chem. Soc.* 96:7333-7336 (1974); Bonora, Gian Maria, et al., *Nucleic Acids Res.* 18:3155-3159 (1990). Liquid phase synthetic methods have the advantage over solid phase synthetic methods in that liquid phase synthesis methods do not require a structure present on a first reactant which is suitable for attaching the reactant to the solid phase. Also, liquid phase synthesis methods do not require avoiding chemical conditions which may cleave the bond between the solid phase and the first reactant (or intermediate product). In addition, reactions in a homogeneous solution may give better yields and more complete reactions than those obtained in heterogeneous solid phase/liquid phase systems such as those present in solid phase synthesis.

[00377] In oligomer-supported liquid phase synthesis the growing product is attached to a large soluble polymeric group. The product from each step of the synthesis can then be separated from unreacted reactants based on the large difference in size between the relatively large polymer-attached product and the unreacted reactants. This permits reactions to take place in homogeneous solutions, and eliminates tedious purification steps associated with traditional liquid phase synthesis. Oligomer-supported liquid phase synthesis has also been adapted to automatic liquid phase synthesis of peptides. Bayer, Ernst, et al., *Peptides: Chemistry, Structure, Biology*, 426-432.

[00378] For solid-phase peptide synthesis, the procedure entails the sequential assembly of the appropriate amino acids into a peptide of a desired sequence while the end of the growing peptide is linked to an insoluble support. Usually, the carboxyl terminus of the peptide is linked to a polymer from which it can be liberated upon treatment with a cleavage reagent. In a common method, an amino acid is bound to a resin particle, and the peptide generated in a stepwise manner by successive additions of protected amino acids to produce a chain of amino acids. Modifications of the technique described by Merrifield are commonly used. See, e.g., Merrifield, *J. Am. Chem. Soc.* 96: 2989-93 (1964). In

an automated solid-phase method, peptides are synthesized by loading the carboxy-terminal amino acid onto an organic linker (e.g., PAM, 4-oxymethylphenylacetamidomethyl), which is covalently attached to an insoluble polystyrene resin cross-linked with divinyl benzene. The terminal amine may be protected by blocking with t-butyloxycarbonyl. Hydroxyl- and carboxyl- groups are commonly protected by blocking with O-benzyl groups. Synthesis is accomplished in an automated peptide synthesizer, such as that available from Applied Biosystems (Foster City, California). Following synthesis, the product may be removed from the resin. The blocking groups are removed by using hydrofluoric acid or trifluoromethyl sulfonic acid according to established methods. A routine synthesis may produce 0.5 mmole of peptide resin. Following cleavage and purification, a yield of approximately 60 to 70% is typically produced. Purification of the product peptides is accomplished by, for example, crystallizing the peptide from an organic solvent such as methyl-butyl ether, then dissolving in distilled water, and using dialysis (if the molecular weight of the subject peptide is greater than about 500 daltons) or reverse high pressure liquid chromatography (e.g., using a C¹⁸ column with 0.1% trifluoroacetic acid and acetonitrile as solvents) if the molecular weight of the peptide is less than 500 daltons. Purified peptide may be lyophilized and stored in a dry state until use. Analysis of the resulting peptides may be accomplished using the common methods of analytical high pressure liquid chromatography (HPLC) and electrospray mass spectrometry (ES-MS).

[00379] In other cases, a protein, for example, a protein involved in the isoprenoid biosynthesis pathway or in fatty acid synthesis, is produced by recombinant methods. For production of any of the proteins described herein, host cells transformed with an expression vector containing the polynucleotide encoding such a protein can be used. The host cell can be a higher eukaryotic cell, such as a mammalian cell, or a lower eukaryotic cell such as a yeast or algal cell, or the host can be a prokaryotic cell such as a bacterial cell. Introduction of the expression vector into the host cell can be accomplished by a variety of methods including calcium phosphate transfection, DEAE-dextran mediated transfection, polybrene, protoplast fusion, liposomes, direct microinjection into the nuclei, scrape loading, biolistic transformation and electroporation. Large scale production of proteins from recombinant organisms is a well established process practiced on a commercial scale and well within the capabilities of one skilled in the art.

[00380] It should be recognized that the present disclosure is not limited to transgenic cells, organisms, and plastids containing a protein or proteins as disclosed herein, but also encompasses such cells, organisms, and plastids transformed with additional nucleotide sequences encoding enzymes involved in fatty acid synthesis. Thus, some embodiments involve the introduction of one or more sequences encoding proteins involved in fatty acid synthesis in addition to a protein disclosed herein. For example, several enzymes in a fatty acid production pathway may be linked, either directly or indirectly, such that products produced by one enzyme in the pathway, once produced, are in close proximity to the next enzyme in the pathway. These additional sequences may be contained in a single vector either operatively linked to a single promoter or linked to multiple promoters, e.g. one promoter for each sequence. Alternatively, the additional coding sequences may be contained in a plurality of additional vectors. When a plurality of vectors are used, they can be introduced into the host cell or organism simultaneously or sequentially.

[00381] Additional embodiments provide a plastid, and in particular a chloroplast, transformed with a polynucleotide encoding a protein of the present disclosure. The protein may be introduced into the genome of the plastid using any of the methods described herein or otherwise known in the art. The plastid may be contained in the organism in which it naturally occurs. Alternatively, the plastid may be an isolated plastid, that is, a plastid that has been removed from the cell in which it normally occurs. Methods for the isolation of plastids are known in the art and can be found, for

example, in Maliga et al., *Methods in Plant Molecular Biology*, Cold Spring Harbor Laboratory Press, 1995; Gupta and Singh, *J. Biosci.*, 21:819 (1996); and Camara et al., *Plant Physiol.*, 73:94 (1983). The isolated plastid transformed with a protein of the present disclosure can be introduced into a host cell. The host cell can be one that naturally contains the plastid or one in which the plastid is not naturally found.

[00382] Also within the scope of the present disclosure are artificial plastid genomes, for example chloroplast genomes, that contain nucleotide sequences encoding any one or more of the proteins of the present disclosure. Methods for the assembly of artificial plastid genomes can be found in co-pending U.S. Patent Application serial number 12/287,230 filed October 6, 2008, published as U.S. Publication No. 2009/0123977 on May 14, 2009, and U.S. Patent Application serial number 12/384,893 filed April 8, 2009, published as U.S. Publication No. 2009/0269816 on October 29, 2009, each of which is incorporated by reference in its entirety.

[00383] Introduction of Polynucleotide into a Host Organism or Cell

[00384] To generate a genetically modified host cell, a polynucleotide, or a polynucleotide cloned into a vector, is introduced stably or transiently into a host cell, using established techniques, including, but not limited to, electroporation, calcium phosphate precipitation, DEAE-dextran mediated transfection, and liposome-mediated transfection. For transformation, a polynucleotide of the present disclosure will generally further include a selectable marker, e.g., any of several well-known selectable markers such as neomycin resistance, ampicillin resistance, tetracycline resistance, chloramphenicol resistance, and kanamycin resistance.

[00385] A polynucleotide or recombinant nucleic acid molecule described herein, can be introduced into a cell (e.g., alga cell) using any method known in the art. A polynucleotide can be introduced into a cell by a variety of methods, which are well known in the art and selected, in part, based on the particular host cell. For example, the polynucleotide can be introduced into a cell using a direct gene transfer method such as electroporation or microprojectile mediated (biolistic) transformation using a particle gun, or the "glass bead method," or by pollen-mediated transformation, liposome-mediated transformation, transformation using wounded or enzyme-degraded immature embryos, or wounded or enzyme-degraded embryogenic callus (for example, as described in Potrykus, *Ann. Rev. Plant. Physiol. Plant Mol. Biol.* 42:205-225, 1991).

[00386] As discussed above, microprojectile mediated transformation can be used to introduce a polynucleotide into a cell (for example, as described in 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 cell using a device such as the BIOLISTIC PD-1000 particle gun (BioRad; Hercules Calif.); a Helios Gene Gun (Cat. #165-2431 and #165-2432; BioRad, U.S.A.); or an Accell Gene Gun (Auragen, U.S.A.). Methods for the transformation using biolistic methods are well known in the art (for example, as described in 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 microprojectile mediated delivery (for example, as described in Duan et al., *Nature Biotech.* 14:494-498, 1996; and 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, and the glass bead agitation method.

[00387] The basic techniques used for transformation and expression in photosynthetic microorganisms are similar to those commonly used for *E. coli*, *Saccharomyces cerevisiae* and other species. Transformation methods customized for a photosynthetic microorganisms, e.g., the chloroplast of a strain of algae, are known in the art. These methods have been described in a number of texts for standard molecular biological manipulation (see Packer & Glaser, 1988, "Cyanobacteria", Meth. Enzymol., Vol. 167; Weissbach & Weissbach, 1988, "Methods for plant molecular biology," Academic Press, New York, Sambrook, Fritsch & Maniatis, 1989, "Molecular Cloning: A laboratory manual," 2nd edition Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.; and Clark M S, 1997, Plant Molecular Biology, Springer, N.Y.). These methods include, for example, biolistic devices (see, for example, Sanford, Trends In Biotech (1988) 6: 299-302, U.S. Pat. No. 4,945,650, electroporation (Fromm et al., Proc. Nat'l. Acad. Sci. (USA) (1985) 82: 5824-5828); use of a laser beam, electroporation, microinjection or any other method capable of introducing DNA into a host cell.

[00388] Plastid transformation is a routine and well known method 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.

[00389] A further refinement in chloroplast transformation/expression technology that facilitates control over the timing and tissue pattern of expression of introduced DNA coding sequences in plant plastid genomes has been described in PCT International Publication WO 95/16783 and U.S. Patent 5,576,198. This method involves the introduction into plant cells of constructs for nuclear transformation that provide for the expression of a viral single subunit RNA polymerase and targeting of this polymerase into the plastids via fusion to a plastid transit peptide. Transformation of plastids with DNA constructs comprising a viral single subunit RNA polymerase-specific promoter specific to the RNA polymerase expressed from the nuclear expression constructs operably linked to DNA coding sequences of interest permits control of the plastid expression constructs in a tissue and/or developmental specific manner in plants comprising both the nuclear polymerase construct and the plastid expression constructs. Expression of the nuclear RNA polymerase coding sequence can be placed under the control of either a constitutive promoter, or a tissue-or developmental stage-specific promoter, thereby extending this control to the plastid expression construct responsive to the plastid-targeted, nuclear-encoded viral RNA polymerase.

[00390] When nuclear transformation is utilized, the protein can be modified for plastid targeting by employing plant cell nuclear transformation constructs wherein DNA coding sequences of interest are fused to any of the available transit peptide sequences capable of facilitating transport of the encoded enzymes into plant plastids, and driving expression by employing an appropriate promoter. Targeting of the protein can be achieved by fusing DNA encoding plastid, e.g., chloroplast, leucoplast, amyloplast, etc., transit peptide sequences to the 5' end of DNAs encoding the enzymes. The sequences that encode a transit peptide region can be obtained, for example, from plant nuclear-encoded plastid proteins, such as the small subunit (SSU) of ribulose biphosphate carboxylase, EPSP synthase, plant fatty acid

biosynthesis related genes including fatty acyl-ACP thioesterases, acyl carrier protein (ACP), stearoyl-ACP desaturase, β -ketoacyl-ACP synthase and acyl-ACP thioesterase, or LHCPH genes, etc. Plastid transit peptide sequences can also be obtained from nucleic acid sequences encoding carotenoid biosynthetic enzymes, such as GGPP synthase, phytoene synthase, and phytoene desaturase. Other transit peptide sequences are disclosed in Von Heijne et al. (1991) *Plant Mol. Biol. Rep.* 9: 104; Clark et al. (1989) *J. Biol. Chem.* 264: 17544; della-Cioppa et al. (1987) *Plant Physiol.* 84: 965; Romer et al. (1993) *Biochem. Biophys. Res. Commun.* 196: 1414; and Shah et al. (1986) *Science* 233: 478. Another transit peptide sequence is that of the intact ACCase from *Chlamydomonas* (genbank EDO96563, amino acids 1-33). The encoding sequence for a transit peptide effective in transport to plastids can include all or a portion of the encoding sequence for a particular transit peptide, and may also contain portions of the mature protein encoding sequence associated with a particular transit peptide. Numerous examples of transit peptides that can be used to deliver target proteins into plastids exist, and the particular transit peptide encoding sequences useful in the present disclosure are not critical as long as delivery into a plastid is obtained. Proteolytic processing within the plastid then produces the mature enzyme. This technique has proven successful with enzymes involved in polyhydroxyalkanoate biosynthesis (Nawrath et al. (1994) *Proc. Natl. Acad. Sci. USA* 91: 12760), and neomycin phosphotransferase II (NPT-II) and CP4 EPSPS (Padgett et al. (1995) *Crop Sci.* 35: 1451), for example.

[00391] Of interest are transit peptide sequences derived from enzymes known to be imported into the leucoplasts of seeds. Examples of enzymes containing useful transit peptides include those related to lipid biosynthesis (e.g., subunits of the plastid-targeted dicot acetyl-CoA carboxylase, biotin carboxylase, biotin carboxyl carrier protein, α -carboxy-transferase, and plastid-targeted monocot multifunctional acetyl-CoA carboxylase (Mw, 220,000); plastidic subunits of the fatty acid synthase complex (e.g., acyl carrier protein (ACP), malonyl-ACP synthase, KASI, KASII, and KASIII); steroyl-ACP desaturase; thioesterases (specific for short, medium, and long chain acyl ACP); plastid-targeted acyl transferases (e.g., glycerol-3-phosphate and acyl transferase); enzymes involved in the biosynthesis of aspartate family amino acids; phytoene synthase; gibberellic acid biosynthesis (e.g., *ent*-kaurene synthases 1 and 2); and carotenoid biosynthesis (e.g., lycopene synthase).

[00392] In some embodiments, an alga is transformed with a nucleic acid which encodes a protein of interest, for example, a prenyl transferase, an isoprenoid synthase, or an enzyme capable of converting a precursor into a fuel product or a precursor of a fuel product (e.g., an isoprenoid or fatty acid).

[00393] In one embodiment, a transformation may introduce a nucleic acid into a plastid of the host alga (e.g., chloroplast). In another embodiment, a transformation may introduce a nucleic acid into the nuclear genome of the host alga. In still another embodiment, a transformation may introduce nucleic acids into both the nuclear genome and into a plastid.

[00394] Transformed cells can be plated on selective media following introduction of exogenous nucleic acids. This method may also comprise several steps for screening. A screen of primary transformants can be conducted to determine which clones have proper insertion of the exogenous nucleic acids. Clones which show the proper integration may be propagated and re-screened to ensure genetic stability. Such methodology ensures that the transformants contain the genes of interest. In many instances, such screening is performed by polymerase chain reaction (PCR); however, any other appropriate technique known in the art may be utilized. Many different methods of PCR are known in the art (e.g., nested PCR, real time PCR). For any given screen, one of skill in the art will recognize that PCR components may be varied to achieve optimal screening results. For example, magnesium concentration may

need to be adjusted upwards when PCR is performed on disrupted alga cells to which (which chelates magnesium) is added to chelate toxic metals. Following the screening for clones with the proper integration of exogenous nucleic acids, clones can be screened for the presence of the encoded protein(s) and/or products. Protein expression screening can be performed by Western blot analysis and/or enzyme activity assays. Transporter and/or product screening may be performed by any method known in the art, for example ATP turnover assay, substrate transport assay, HPLC or gas chromatography.

[00395] The expression of the protein or enzyme can be accomplished by inserting a polynucleotide sequence (gene) encoding the protein or enzyme into the chloroplast or nuclear genome of a microalgae. The modified strain of microalgae can be made homoplasmic to ensure that the polynucleotide will be stably maintained in the chloroplast genome of all descendants. A microalga is homoplasmic for a gene when the inserted gene is present in all copies of the chloroplast genome, for example. 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% or more of the total soluble plant protein. The process of determining the plasmic state of an organism of the present disclosure involves screening transformants for the presence of exogenous nucleic acids and the absence of wild-type nucleic acids at a given locus of interest.

[00396] Vectors

[00397] Construct, vector and plasmid are used interchangeably throughout the disclosure. Nucleic acids encoding the proteins described herein, can be contained in vectors, including cloning and expression vectors. A cloning vector is a self-replicating DNA molecule that serves to transfer a DNA segment into a host cell. Three common types of cloning vectors are bacterial plasmids, phages, and other viruses. An expression vector is a cloning vector designed so that a coding sequence inserted at a particular site will be transcribed and translated into a protein. Both cloning and expression vectors can contain nucleotide sequences that allow the vectors to replicate in one or more suitable host cells. In cloning vectors, this sequence is generally one that enables the vector to replicate independently of the host cell chromosomes, and also includes either origins of replication or autonomously replicating sequences.

[00398] In some embodiments, a polynucleotide of the present disclosure is cloned or inserted into an expression vector using cloning techniques known to one of skill in the art. The nucleotide sequences may be inserted into a vector by a variety of methods. In the most common method the sequences are inserted into an appropriate restriction endonuclease site(s) using procedures commonly known to those skilled in the art and detailed in, for example, Sambrook et al., *Molecular Cloning, A Laboratory Manual*, 2nd Ed., Cold Spring Harbor Press, (1989) and Ausubel et al., *Short Protocols in Molecular Biology*, 2nd Ed., John Wiley & Sons (1992).

[00399] Suitable expression vectors include, but are not limited to, baculovirus vectors, bacteriophage vectors, plasmids, phagemids, cosmids, fosmids, bacterial artificial chromosomes, viral vectors (e.g. viral vectors based on vaccinia virus, poliovirus, adenovirus, adeno-associated virus, SV40, and herpes simplex virus), PI-based artificial chromosomes, yeast plasmids, yeast artificial chromosomes, and any other vectors specific for specific hosts of interest (such as *E. coli* and yeast). Thus, for example, a polynucleotide encoding an FPP synthase, can be inserted into any one

of a variety of expression vectors that are capable of expressing the enzyme. Such vectors can include, for example, chromosomal, nonchromosomal and synthetic DNA sequences.

[00400] Suitable expression vectors include chromosomal, non-chromosomal and synthetic DNA sequences, for example, SV 40 derivatives; bacterial plasmids; phage DNA; baculovirus; yeast plasmids; vectors derived from combinations of plasmids and phage DNA; and viral DNA such as vaccinia, adenovirus, fowl pox virus, and pseudorabies. In addition, any other vector that is replicable and viable in the host may be used. For example, vectors such as Ble2A, Arg7/2A, and SEnuc357 can be used for the expression of a protein.

[00401] Numerous suitable expression vectors are known to those of skill in the art. The following vectors are provided by way of example; for bacterial host cells: pQE vectors (Qiagen), pBluescript plasmids, pNH vectors, lambda-ZAP vectors (Stratagene), pTre99a, pKK223-3, pDR540, and pRIT2T (Pharmacia); for eukaryotic host cells: pXT1, pSG5 (Stratagene), pSVK3, pBPV, pMSG, pET21a-d(+) vectors (Novagen), and pSVLSV40 (Pharmacia). However, any other plasmid or other vector may be used so long as it is compatible with the host cell.

[00402] The expression vector, or a linearized portion thereof, can encode one or more exogenous or endogenous nucleotide sequences. Examples of exogenous nucleotide sequences that can be transformed into a host include genes from bacteria, fungi, plants, photosynthetic bacteria or other algae. Examples of other types of nucleotide sequences that can be transformed into a host, include, but are not limited to, transporter genes, isoprenoid producing genes, genes which encode for proteins which produce isoprenoids with two phosphates (e.g., GPP synthase and/or FPP synthase), genes which encode for proteins which produce fatty acids, lipids, or triglycerides, for example, ACCases, endogenous promoters, and 5' UTRs from the psbA, atpA, or rbcL genes. In some instances, an exogenous sequence is flanked by two homologous sequences.

[00403] Homologous sequences are, for example, those that have at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, at least 98%, or at least at least 99% sequence identity to a reference amino acid sequence, for example, the amino acid sequence found naturally in the host cell. The first and second sequences enable recombination of the exogenous or endogenous sequence into the genome of the host organism. The first and second homologous sequences can be at least 100, at least 200, at least 300, at least 400, at least 500, or at least 1000, or at least 1500 nucleotides in length.

[00404] The polynucleotide sequence may comprise nucleotide sequences that are 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 in usage of nucleotide codons to specify a given amino acid. Without being bound by theory, by using a host cell's preferred codons, the rate of translation may be greater. Therefore, when synthesizing a gene for improved expression in a host cell, it may be desirable to design the gene such that its frequency of codon usage approaches the frequency of preferred codon usage of the host cell. In some organisms, codon bias differs between the nuclear genome and organelle genomes, thus, codon optimization or biasing may be performed for the target genome (e.g., nuclear codon biased or chloroplast codon biased). In some embodiments, codon biasing occurs before mutagenesis to generate a polypeptide. In other embodiments, codon biasing occurs after mutagenesis to generate a polynucleotide. In yet other embodiments, codon biasing occurs before mutagenesis as well as after mutagenesis. Codon bias is described in detail herein.

[00405] In some embodiments, a vector comprises a polynucleotide operably linked to one or more control elements, such as a promoter and/or a transcription terminator. A nucleic acid sequence is operably linked when it is placed into a

functional relationship with another nucleic acid sequence. For example, DNA for a presequence or secretory leader is operatively linked to DNA for a polypeptide if it is expressed as a preprotein which participates in the secretion of the polypeptide; a promoter is operably linked to a coding sequence if it affects the transcription of the sequence; or a ribosome binding site is operably linked to a coding sequence if it is positioned so as to facilitate translation. Generally, operably linked sequences are contiguous and, in the case of a secretory leader, contiguous and in reading phase. Linking is achieved by ligation at restriction enzyme sites. If suitable restriction sites are not available, then synthetic oligonucleotide adapters or linkers can be used as is known to those skilled in the art. Sambrook et al., *Molecular Cloning. A Laboratory Manual*, 2nd Ed., Cold Spring Harbor Press, (1989) and Ausubel et al., *Short Protocols in Molecular Biology*, 2nd Ed., John Wiley & Sons (1992).

[00406] A vector in some embodiments provides for amplification of the copy number of a polynucleotide. A vector can be, for example, an expression vector that provides for expression of an ACCase, a prenyl transferase, an isoprenoid synthase, or a mevalonate synthesis enzyme in a host cell, e.g., a prokaryotic host cell or a eukaryotic host cell.

[00407] A polynucleotide or polynucleotides can be contained in a vector or vectors. For example, where a second (or more) nucleic acid molecule is desired, the second nucleic acid molecule can be contained in a vector, which can, but need not be, the same vector as that containing the first nucleic acid molecule. For example, an algal host cell modified to express two endogenous or exogenous genes may be transformed with a single vector containing both sequences, or two vectors, each comprising one gene to be expressed. The vector can be any vector useful for introducing a polynucleotide into a genome and can include a nucleotide sequence of genomic DNA (e.g., nuclear or plastid) that is sufficient to undergo homologous recombination with genomic DNA, for example, a nucleotide sequence comprising about 400 to about 1500 or more substantially contiguous nucleotides of genomic DNA.

[00408] A regulatory or control element, as the term is used herein, broadly refers to a nucleotide sequence that regulates the transcription or translation of a polynucleotide or the localization of a polypeptide to which it is operatively linked. Examples include, but are not limited to, an RBS, a promoter, enhancer, transcription terminator, an initiation (start) codon, a splicing signal for intron excision and maintenance of a correct reading frame, a STOP codon, an amber or ochre codon, and an IRES. A regulatory element can include a promoter and transcriptional and translational stop signals. Elements may be provided with linkers for the purpose of introducing specific restriction sites facilitating ligation of the control sequences with the coding region of a nucleotide sequence encoding a polypeptide. Additionally, a sequence comprising a cell compartmentalization signal (i.e., a sequence that targets a polypeptide to the cytosol, nucleus, chloroplast membrane or cell membrane) can be attached to the polynucleotide encoding a protein of interest. Such signals are well known in the art and have been widely reported (see, e.g., U.S. Pat. No. 5,776,689).

[00409] Promoters are untranslated sequences located generally 100 to 1000 base pairs (bp) upstream from the start codon of a structural gene that regulate the transcription and translation of nucleic acid sequences under their control.

[00410] Promoters useful for the present disclosure may come from any source (e.g., viral, bacterial, fungal, protist, and animal). The promoters contemplated herein can be specific to photosynthetic organisms, non-vascular photosynthetic organisms, and vascular photosynthetic organisms (e.g., algae, flowering plants). In some instances, the nucleic acids above are inserted into a vector that comprises a promoter of a photosynthetic organism, e.g., algae. The promoter can be a constitutive promoter or an inducible promoter. A promoter typically includes necessary nucleic acid sequences near the start site of transcription, (e.g., a TATA element). Common promoters used in expression vectors include, but

are not limited to, LTR or SV40 promoter, the *E. coli* lac or trp promoters, and the phage lambda PL promoter. Other promoters known to control the expression of genes in prokaryotic or eukaryotic cells can be used and are known to those skilled in the art. Expression vectors may also contain a ribosome binding site for translation initiation, and a transcription terminator. The vector may also contain sequences useful for the amplification of gene expression.

[00411] A “constitutive” promoter is a promoter that is active under most environmental and developmental conditions. An “inducible” promoter is a promoter that is active under controllable environmental or developmental conditions. Examples of inducible promoters/regulatory elements include, for example, a nitrate-inducible promoter (for example, as described in Bock et al, *Plant Mol. Biol.* 17:9 (1991)), or a light-inducible promoter, (for example, as described in Feinbaum et al, *Mol. Gen. Genet.* 226:449 (1991), and Lam and Chua, *Science* 248:471 (1990)), or a heat responsive promoter (for example, as described in Muller et al., *Gene* 111: 165-73 (1992)).

[00412] In many embodiments, a polynucleotide of the present disclosure includes a nucleotide sequence encoding a protein or enzyme of the present disclosure, where the nucleotide sequence encoding the polypeptide is operably linked to an inducible promoter. Inducible promoters are well known in the art. Suitable inducible promoters include, but are not limited to, the pL of bacteriophage λ ; PlacO; P_{trp}; Ptac (P_{trp}-lac hybrid promoter); an isopropyl-beta-D-thiogalactopyranoside (IPTG)-inducible promoter, e.g., a lacZ promoter; a tetracycline-inducible promoter; an arabinose inducible promoter, e.g., P_{BAD} (for example, as described in Guzman et al. (1995) *J. Bacteriol.* 177:4121-4130); a xylose-inducible promoter, e.g., P_{xyl} (for example, as described in Kim et al. (1996) *Gene* 181:71-76); a GAL1 promoter; a tryptophan promoter; a lac promoter; an alcohol-inducible promoter, e.g., a methanol-inducible promoter, an ethanol-inducible promoter; a raffinose-inducible promoter; and a heat-inducible promoter, e.g., heat inducible lambda P_L promoter and a promoter controlled by a heat-sensitive repressor (e.g., C1857-repressed lambda-based expression vectors; for example, as described in Hoffmann et al. (1999) *FEMS Microbiol Lett.* 177(2):327-34).

[00413] In many embodiments, a polynucleotide of the present disclosure includes a nucleotide sequence encoding a protein or enzyme of the present disclosure, where the nucleotide sequence encoding the polypeptide is operably linked to a constitutive promoter. Suitable constitutive promoters for use in prokaryotic cells are known in the art and include, but are not limited to, a sigma70 promoter, and a consensus sigma70 promoter.

[00414] Suitable promoters for use in prokaryotic host cells include, but are not limited to, a bacteriophage T7 RNA polymerase promoter; a trp promoter; a lac operon promoter; a hybrid promoter, e.g., a lac/tac hybrid promoter, a tac/trc hybrid promoter, a trp/lac promoter, a T7/lac promoter; a trc promoter; a tac promoter; an araBAD promoter; in vivo regulated promoters, such as an ssaG promoter or a related promoter (for example, as described in U.S. Patent Publication No. 20040131637), a pagC promoter (for example, as described in Pulkkinen and Miller, *J. Bacteriol.*, 1991: 173(1): 86-93; and Alpuche-Aranda et al., *PNAS*, 1992; 89(21): 10079-83), a nirB promoter (for example, as described in Harborne et al. (1992) *Mol. Micro.* 6:2805-2813; Dunstan et al. (1999) *Infect. Immun.* 67:5133-5141; McKelvie et al. (2004) *Vaccine* 22:3243-3255; and Chatfield et al. (1992) *Biotechnol.* 10:888-892); a sigma70 promoter, e.g., a consensus sigma70 promoter (for example, GenBank Accession Nos. AX798980, AX798961, and AX798183); a stationary phase promoter, e.g., a dps promoter, an spv promoter; a promoter derived from the pathogenicity island SPI-2 (for example, as described in WO96/17951); an actA promoter (for example, as described in Shetron-Rama et al. (2002) *Infect. Immun.* 70:1087-1096); an rpsM promoter (for example, as described in Valdivia and Falkow (1996) *Mol. Microbiol.* 22:367-378); a tet promoter (for example, as described in Hillen, W. and Wissmann, A. (1989) In Saenger, W. and Heinemann, U. (eds), *Topics in Molecular and Structural Biology*, Protein-

Nucleic Acid Interaction. Macmillan, London, UK, Vol. 10, pp. 143-162); and an SP6 promoter (for example, as described in Melton et al. (1984) Nucl. Acids Res. 12:7035-7056).

[00415] In yeast, a number of vectors containing constitutive or inducible promoters may be used. For a review of such vectors see, Current Protocols in Molecular Biology, Vol. 2, 1988, Ed. Ausubel, et al., Greene Publish. Assoc. & Wiley Interscience, Ch. 13; Grant, et al., 1987, Expression and Secretion Vectors for Yeast, in Methods in Enzymology, Eds. Wu & Grossman, 31987, Acad. Press, N.Y., Vol. 153, pp. 516-544; Glover, 1986, DNA Cloning, Vol. II, IRL Press, Wash., D.C., Ch. 3; Bitter, 1987, Heterologous Gene Expression in Yeast, Methods in Enzymology, Eds. Berger & Kimmel, Acad. Press, N.Y., Vol. 152, pp. 673-684; and The Molecular Biology of the Yeast *Saccharomyces*, 1982, Eds. Strathern et al., Cold Spring Harbor Press, Vols. I and II. A constitutive yeast promoter such as ADH or LEU2 or an inducible promoter such as GAL may be used (for example, as described in Cloning in Yeast, Ch. 3, R. Rothstein In: DNA Cloning Vol. 11, A Practical Approach, Ed. DM Glover, 1986, IRL Press, Wash., D.C.). Alternatively, vectors may be used which promote integration of foreign DNA sequences into the yeast chromosome.

[00416] Non-limiting examples of suitable eukaryotic promoters include CMV immediate early, HSV thymidine kinase, early and late SV40, LTRs from retrovirus, and mouse metallothionein-I. Selection of the appropriate vector and promoter is well within the level of ordinary skill in the art. The expression vector may also contain a ribosome binding site for translation initiation and a transcription terminator. The expression vector may also include appropriate sequences for amplifying expression.

[00417] A vector utilized in the practice of the disclosure also can contain one or more additional nucleotide sequences that confer desirable characteristics on the vector, including, for example, sequences such as cloning sites that facilitate manipulation of the vector, regulatory elements that direct replication of the vector or transcription of nucleotide sequences contain therein, and sequences that encode a selectable marker. As such, the vector can contain, for example, one or more cloning sites such as a multiple cloning site, which can, but need not, be positioned such that a exogenous or endogenous polynucleotide can be inserted into the vector and operatively linked to a desired element.

[00418] The vector also can contain a prokaryote origin of replication (ori), for example, an *E. coli* ori or a cosmid ori, thus allowing passage of the vector into a prokaryote host cell, as well as into a plant chloroplast. Various bacterial and viral origins of replication are well known to those skilled in the art and include, but are not limited to the pBR322 plasmid origin, the 2u plasmid origin, and the SV40, polyoma, adenovirus, VSV, and BPV viral origins.

[00419] A regulatory or control element, as the term is used herein, broadly refers to a nucleotide sequence that regulates the transcription or translation of a polynucleotide or the localization of a polypeptide to which it is operatively linked. Examples include, but are not limited to, an RBS, a promoter, enhancer, transcription terminator, an initiation (start) codon, a splicing signal for intron excision and maintenance of a correct reading frame, a STOP codon, an amber or ochre codon, an IRES. Additionally, an element can be a cell compartmentalization signal (i.e., a sequence that targets a polypeptide to the cytosol, nucleus, chloroplast membrane or cell membrane). In some aspects of the present disclosure, a cell compartmentalization signal (e.g., a cell membrane targeting sequence) may be ligated to a gene and/or transcript, such that translation of the gene occurs in the chloroplast. In other aspects, a cell compartmentalization signal may be ligated to a gene such that, following translation of the gene, the protein is transported to the cell membrane. Cell compartmentalization signals are well known in the art and have been widely reported (see, e.g., U.S. Pat. No. 5,776,689).

[00420] A vector, or a linearized portion thereof, may include a nucleotide sequence encoding a reporter polypeptide or other selectable marker. The term "reporter" or "selectable marker" refers to a polynucleotide (or encoded polypeptide) that confers a detectable phenotype. A reporter 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 (for example, as described in Giacomini, *Plant Sci.* 116:59-72, 1996; Scikantha, *J. Bacteriol.* 178:121, 1996; Gerdes, *FEBS Lett.* 389:44-47, 1996; and Jefferson, *EMBO J.* 6:3901-3907, 1997, β -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.

[00421] A selectable marker can provide a means to obtain, for example, prokaryotic cells, eukaryotic cells, and/or plant cells that express the marker and, therefore, can be useful as a component of a vector of the disclosure. The selection gene or marker can encode for a protein necessary for the survival or growth of the host cell transformed with the vector. One class of selectable markers are native or modified genes which restore a biological or physiological function to a host cell (e.g., restores photosynthetic capability or restores a metabolic pathway). Other examples of selectable markers include, but are not limited to, those that confer antimetabolite resistance, for example, dihydrofolate reductase, which confers resistance to methotrexate (for example, as described in Reiss, *Plant Physiol. (Life Sci. Adv.)* 13:143-149, 1994); neomycin phosphotransferase, which confers resistance to the aminoglycosides neomycin, kanamycin and paromycin (for example, as described in Herrera-Estrella, *EMBO J.* 2:987-995, 1983), hygromycin, which confers resistance to hygromycin (for example, as described in 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 (for example, as described in Hartman, *Proc. Natl. Acad. Sci., USA* 85:8047, 1988); mannose-6-phosphate isomerase which allows cells to utilize mannose (for example, as described in PCT Publication Application No. WO 94/20627); ornithine decarboxylase, which confers resistance to the ornithine decarboxylase inhibitor, 2-(difluoromethyl)-DL-ornithine (DFMO; for example, as described in McConlogue, 1987, In: *Current Communications in Molecular Biology*, Cold Spring Harbor Laboratory ed.); and deaminase from *Aspergillus terreus*, which confers resistance to Blasticidin S (for example, as described in 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 (for example, as described in White et al., *Nucl. Acids Res.* 18:1062, 1990; and Spencer et al., *Theor. Appl. Genet.* 79:625-631, 1990), a mutant EPSPV-synthase, which confers glyphosate resistance (for example, as described in Hinchey et al., *BioTechnology* 91:915-922, 1998), a mutant acetolactate synthase, which confers imidazolinone or sulfonylurea resistance (for example, as described in Lee et al., *EMBO J.* 7:1241-1248, 1988), a mutant psbA, which confers resistance to atrazine (for example, as described in Smeda et al., *Plant Physiol.* 103:911-917, 1993), or a mutant protoporphyrinogen oxidase (for example, as described in 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; tetracycline or ampicillin resistance for prokaryotes such as *E. coli*; and bleomycin, gentamycin, glyphosate, hygromycin, kanamycin, methotrexate, phleomycin, phosphinotricin, spectinomycin, daptomycin, streptomycin, sulfonamide and sulfonylurea resistance in plants (for example, as described in Maliga et al., *Methods in Plant Molecular Biology*, Cold Spring Harbor Laboratory

Press, 1995, page 39). Additional selectable markers include a mutation in dichlorophenyl dimethylurea (DCMU) that results in resistance to DCMU. Selectable markers also include chloramphenicol acetyltransferase (CAT) and tetracycline. The selection marker can have its own promoter or its expression can be driven by a promoter driving the expression of a polypeptide of interest.

[00422] Reporter genes greatly enhance the ability to monitor gene expression in a number of biological organisms. Reporter genes have been successfully used in chloroplasts of higher plants, and high levels of recombinant protein expression have been reported. In addition, reporter genes have been used in the chloroplast of *C. reinhardtii*. In chloroplasts of higher plants, β -glucuronidase (*uidA*, for example, as described in Staub and Maliga, *EMBO J.* 12:601-606, 1993), neomycin phosphotransferase (*uptII*, for example, as described in Carrer et al., *Mol. Gen. Genet.* 241:49-56, 1993), adenosyl-3-adenyltransf-erase (*aadA*, for example, as described in Svab and Maliga, *Proc. Natl. Acad. Sci., USA* 90:913-917, 1993), and the *Aequorea victoria* GFP (for example, as described in Sidorov et al., *Plant J.* 19:209-216, 1999) have been used as reporter genes (for example, as described in Heifetz, *Biochemie* 82:655-666, 2000). Each of these genes has attributes that make them useful reporters of chloroplast gene expression, such as ease of analysis, sensitivity, or the ability to examine expression in situ. Based upon these studies, other exogenous proteins have been expressed in the chloroplasts of higher plants such as *Bacillus thuringiensis* Cry toxins, conferring resistance to insect herbivores (for example, as described in Kota et al., *Proc. Natl. Acad. Sci., USA* 96:1840-1845, 1999), or human somatotropin (for example, as described in Staub et al., *Nat. Biotechnol.* 18:333-338, 2000), a potential biopharmaceutical. Several reporter genes have been expressed in the chloroplast of the eukaryotic green alga, *C. reinhardtii*, including *aadA* (for example, as described in Goldschmidt-Clermont, *Nucl. Acids Res.* 19:4083-4089 1991; and Zerges and Rochaix, *Mol. Cell Biol.* 14:5268-5277, 1994), *uidA* (for example, as described in Sakamoto et al., *Proc. Natl. Acad. Sci., USA* 90:477-501, 1993; and Ishikura et al., *J. Biosci. Bioeng.* 87:307-314 1999), Renilla luciferase (for example, as described in Minko et al., *Mol. Gen. Genet.* 262:421-425, 1999) and the amino glycoside phosphotransferase from *Acinetobacter baumannii*, *aphA6* (for example, as described in Bateman and Purton, *Mol. Gen. Genet* 263:404-410, 2000). In one embodiment the protein described herein is modified by the addition of an N-terminal strep tag epitope to add in the detection of protein expression.

[00423] In some instances, the vectors of the present disclosure will contain elements such as an *E. coli* or *S. cerevisiae* origin of replication. Such features, combined with appropriate selectable markers, allows for the vector to be "shuttled" between the target host cell and a bacterial and/or yeast cell. The ability to passage a shuttle vector of the disclosure in a secondary host may allow for more convenient manipulation of the features of the vector. For example, a reaction mixture containing the vector and inserted polynucleotide(s) of interest can be transformed into prokaryote host cells such as *E. coli*, amplified and collected using routine methods, and examined to identify vectors containing an insert or construct of interest. If desired, the vector can be further manipulated, for example, by performing site directed mutagenesis of the inserted polynucleotide, then again amplifying and selecting vectors having a mutated polynucleotide of interest. A shuttle vector then can be introduced into plant cell chloroplasts, wherein a polypeptide of interest can be expressed and, if desired, isolated according to a method of the disclosure.

[00424] Knowledge of the chloroplast or nuclear genome of the host organism, for example, *C. reinhardtii*, is useful in the construction of vectors for use in the disclosed embodiments. Chloroplast vectors and methods for selecting regions of a chloroplast genome for use as a vector are well known (see, for example, Bock, *J. Mol. Biol.* 312:425-438, 2001; Staub and Maliga, *Plant Cell* 4:39-45, 1992; and Kavanagh et al., *Genetics* 152:1111-1122, 1999, each of which is

incorporated herein by reference). The entire chloroplast genome of *C. reinhardtii* is available to the public on the world wide web, at the URL "biology.duke.edu/chlamy_genome/- chloro.html" (see "view complete genome as text file" link and "maps of the chloroplast genome" link; J. Maul, J. W. Lilly, and D. B. Stern, unpublished results; revised Jan. 28, 2002; to be published as GenBank Acc. No. AF396929; and Maul, J. E., et al. (2002) *The Plant Cell*, Vol. 14 (2659-2679)). Generally, the nucleotide sequence of the chloroplast genomic DNA that is selected for use is not a portion of a gene, including a regulatory sequence or coding sequence. For example, the selected sequence is not a gene that if disrupted, due to the homologous recombination event, would produce a deleterious effect with respect to the chloroplast. For example, a deleterious effect on the replication of the chloroplast genome or to a plant cell containing the chloroplast. In this respect, the website containing the *C. reinhardtii* chloroplast genome sequence also provides maps showing coding and non-coding regions of the chloroplast genome, thus facilitating selection of a sequence useful for constructing a vector (also described in Maul, J. E., et al. (2002) *The Plant Cell*, Vol. 14 (2659-2679)). For example, the chloroplast vector, p322, is a clone extending from the Eco (Eco RI) site at about position 143.1 kb to the Xho (Xho I) site at about position 148.5 kb (see, world wide web, at the URL "biology.duke.edu/chlamy_genome/chloro.html", and clicking on "maps of the chloroplast genome" link, and "140-150 kb" link; also accessible directly on world wide web at URL "biology.duke.edu/chlam- y/chloro/chlorol40.html").

[00425] In addition, the entire nuclear genome of *C. reinhardtii* is described in Merchant, S. S., et al., *Science* (2007), 318(5848):245-250, thus facilitating one of skill in the art to select a sequence or sequences useful for constructing a vector.

[00426] For expression of the polypeptide in a host, an expression cassette or vector may be employed. The expression vector will provide a transcriptional and translational initiation region, which may be inducible or constitutive, where the coding region is operably linked under the transcriptional control of the transcriptional initiation region, and a transcriptional and translational termination region. These control regions may be native to the gene, or may be derived from an exogenous source. Expression vectors generally have convenient restriction sites located near the promoter sequence to provide for the insertion of nucleic acid sequences encoding exogenous or endogenous proteins. A selectable marker operative in the expression host may be present.

[00427] The nucleotide sequences may be inserted into a vector by a variety of methods. In the most common method the sequences are inserted into an appropriate restriction endonuclease site(s) using procedures commonly known to those skilled in the art and detailed in, for example, Sambrook et al., *Molecular Cloning, A Laboratory Manual*, 2nd Ed., Cold Spring Harbor Press, (1989) and Ausubel et al., *Short Protocols in Molecular Biology*, 2nd Ed., John Wiley & Sons (1992).

[00428] The description herein provides that host cells may be transformed with vectors. One of skill in the art will recognize that such transformation includes transformation with circular or linearized vectors, or linearized portions of a vector. Thus, a host cell comprising a vector may contain the entire vector in the cell (in either circular or linear form), or may contain a linearized portion of a vector of the present disclosure. In some instances 0.5 to 1.5 kb flanking nucleotide sequences of chloroplast genomic DNA may be used. In some instances 0.5 to 1.5 kb flanking nucleotide sequences of nuclear genomic DNA may be used, or 2.0 to 5.0 kb may be used.

[00429] **Compounds**

[00430] The modified or transformed host organism disclosed herein is useful in the production of a desired compound, composition, or product. The present disclosure provides methods of producing, for example, an isoprenoid or

isoprenoid precursor compound in a host cell. One such method involves, culturing a modified host cell in a suitable culture medium under conditions that promote synthesis of a product, for example, an isoprenoid compound or isoprenoid precursor compound, where the isoprenoid compound is generated by the expression of an enzyme of the present disclosure, wherein the enzyme uses a substrate present in the host cell. In some embodiments, a method further comprises isolating the isoprenoid compound from the cell and/or from the culture medium.

[00431] In some embodiments, the product (e.g. fuel molecule) is collected by harvesting the liquid medium. As some fuel molecules (e.g., monoterpenes) are immiscible in water, they would float to the surface of the liquid medium and could be extracted easily, for example by skimming. In other instances, the fuel molecules can be extracted from the liquid medium. In still other instances, the fuel molecules are volatile. In such instances, impermeable barriers can cover or otherwise surround the growth environment and can be extracted from the air within the barrier. For some fuel molecules, the product may be extracted from both the environment (e.g., liquid environment and/or air) and from the intact host cells. Typically, the organism would be harvested at an appropriate point and the product may then be extracted from the organism. The collection of cells may be by any means known in the art, including, but not limited to concentrating cells, mechanical or chemical disruption of cells, and purification of product(s) from cell cultures and/or cell lysates. Cells and/or organisms can be grown and then the product(s) collected by any means known to one of skill in the art. One method of extracting the product is by harvesting the host cell or a group of host cells and then drying the cell(s). The product(s) from the dried host cell(s) are then harvested by crushing the cells to expose the product. In some instances, the product may be produced without killing the organisms. Producing and/or expressing the product may not render the organism unviable.

[00432] In some embodiments, a genetically modified host cell is cultured in a suitable medium (e.g., Luria-Bertoni broth, optionally supplemented with one or more additional agents, such as an inducer (e.g., where the isoprenoid synthase is under the control of an inducible promoter); and the culture medium is overlaid with an organic solvent, e.g. dodecane, forming an organic layer. The compound produced by the genetically modified host partitions into the organic layer, from which it can then be purified. In some embodiments, where, for example, a prenyl transferase, isoprenoid synthase or mevalonate synthesis-encoding nucleotide sequence is operably linked to an inducible promoter, an inducer is added to the culture medium; and, after a suitable time, the compound is isolated from the organic layer overlaid on the culture medium.

[00433] In some embodiments, the compound or product, for example, an isoprenoid compound will be separated from other products which may be present in the organic layer. Separation of the compound from other products that may be present in the organic layer is readily achieved using, e.g., standard chromatographic techniques.

[00434] Methods of culturing the host cells, separating products, and isolating the desired product or products are known to one of skill in the art and are discussed further herein.

[00435] In some embodiments, the compound, for example, an isoprenoid or isoprenoid compound is produced in a genetically modified host cell at a level that is at least about 2-fold, at least about 5-fold, at least about 10-fold, at least about 25-fold, at least about 50-fold, at least about 100-fold, at least about 500-fold, at least about 1000-fold, at least about 2000-fold, at least about 3000-fold, at least about 4000-fold, at least about 5000-fold, or at least about 10,000-fold, or more, higher than the level of the isoprenoid or isoprenoid precursor compound produced in an unmodified host cell that produces the isoprenoid or isoprenoid precursor compound via the same biosynthetic pathway.

[00436] In some embodiments, the compound, for example, an isoprenoid compound is pure, e.g., at least about 40% pure, at least about 50% pure, at least about 60% pure, at least about 70% pure, at least about 80% pure, at least about 90% pure, at least about 95% pure, at least about 98%, or more than 98% pure. "Pure" in the context of an isoprenoid compound refers to an isoprenoid compound that is free from other isoprenoid compounds, portions of compounds, contaminants, and unwanted byproducts, for example.

[00437] 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 are not limited to, nitrogen, oxygen, sulfur, and phosphorus. Some products can be hydrocarbon-rich, wherein, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, or at least 95% of the product by weight is made up of carbon and hydrogen.

[00438] In one embodiment, the vector comprises one or more nucleic acid sequences involved in isoprenoid synthesis. The terms "isoprenoid," "isoprenoid compound," "terpene," "terpene compound," "terpenoid," and "terpenoid compound" are used interchangeably herein. Isoprenoid compounds include, but are not limited to, monoterpenes, sesquiterpenes, diterpenes, triterpenes, and polyterpenes.

[00439] One exemplary group of hydrocarbon products are isoprenoids. Isoprenoids (including terpenoids) are derived from isoprene subunits, but are modified, for example, by the addition of heteroatoms such as oxygen, by carbon skeleton rearrangement, and by alkylation. Isoprenoids generally have a number of carbon atoms which is evenly divisible by five, but this is not a requirement as "irregular" terpenoids are known to one of skill in the art. Carotenoids, such as carotenes and xanthophylls, are examples of isoprenoids that are useful products. A steroid is an example of a terpenoid. Examples of isoprenoids include, but are not limited to, hemiterpenes (C₅), monoterpenes (C₁₀), sesquiterpenes (C₁₅), diterpenes (C₂₀), triterpenes (C₃₀), tetraterpenes (C₄₀), polyterpenes (C_n, wherein "n" is equal to or greater than 45), and their derivatives. Other examples of isoprenoids include, but are not limited to, limonene, 1,8-cineole, α -pinene, camphene, (+)-sabinene, myrcene, abietadiene, taxadiene, farnesyl pyrophosphate, fusicoccadiene, amorphaadiene, (E)- α -bisabolene, zingiberene, or diapophytoene, and their derivatives.

[00440] Products, for example fuel products, comprising hydrocarbons, may be precursors or products conventionally derived from crude oil, or petroleum, such as, but not limited to, liquid petroleum gas, naphtha (ligroin), gasoline, kerosene, diesel, lubricating oil, heavy gas, coke, asphalt, tar, and waxes.

[00441] Useful products include, but are not limited to, terpenes and terpenoids as described above. An exemplary group of terpenes are diterpenes (C₂₀). Diterpenes are hydrocarbons that can be modified (e.g. oxidized, methyl groups removed, or cyclized); the carbon skeleton of a diterpene can be rearranged, to form, for example, terpenoids, such as fusicoccadiene. Fusicoccadiene may also be formed, for example, directly from the isoprene precursors, without being bound by the availability of diterpene or GGDP. Genetic modification of organisms, such as algae, by the methods described herein, can lead to the production of fusicoccadiene, for example, and other types of terpenes, such as limonene, for example. Genetic modification can also lead to the production of modified terpenes, such as methyl squalene or hydroxylated and/or conjugated terpenes such as paclitaxel.

[00442] Other useful products can be, for example, a product comprising a hydrocarbon obtained from an organism expressing a diterpene synthase. Such exemplary products include ent-kaurene, casbene, and fusicoccadiene, and may also include fuel additives.

[00443] In some embodiments, a product (such as a fuel product) contemplated herein comprises one or more carbons derived from an inorganic carbon source. In some embodiments, at least 10%, at least 20%, at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, or at least 99% of the carbons of a product as described herein are derived from an inorganic carbon source. Examples of inorganic carbon sources include, but are not limited to, carbon dioxide, carbonate, bicarbonate, and carbonic acid. The product can be, for example, an organic molecule with carbons from an inorganic carbon source that were fixed during photosynthesis.

[00444] The products produced by the present disclosure may be naturally, or non-naturally (e.g., as a result of transformation) produced by the host cell(s) and/or organism(s) transformed. For example, products not naturally produced by algae may include non-native terpenes/terpenoids such as fusicoccadiene or limonene. A product naturally produced in algae may be a terpene such as a carotenoid (for example, beta-carotene). The host cell may be genetically modified, for example, by transformation of the cell with a sequence encoding a protein, wherein expression of the protein results in the secretion of a naturally or a non-naturally produced product (e.g. limonene) or products. The product may be a molecule not found in nature.

[00445] Examples of products include petrochemical products, precursors of petrochemical products, fuel products, petroleum products, precursors of petroleum products, and all other substances that may be useful in the petrochemical industry. The product may be used for generating substances, or materials, useful in the petrochemical industry. The 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 other types of engines. In one embodiment, a method herein comprises combusting a refined or "upgraded" composition. For example, combusting a refined composition can comprise inserting the refined composition into a combustion engine, such as an automobile engine or a jet engine. Products described herein may also be used to produce plastics, resins, fibers, elastomers, pharmaceuticals, nutraceuticals, lubricants, and gels, for example.

[00446] Useful products can also include isoprenoid precursors. Isoprenoid precursors are generated by one of two pathways; the mevalonate pathway or the methylerythritol phosphate (MEP) pathway. Both pathways generate dimethylallyl pyrophosphate (DMAPP) and isopentyl pyrophosphate (IPP), the common C5 precursor for isoprenoids. The DMAPP and IPP are condensed to form geranyl-diphosphate (GPP), or other precursors, such as farnesyl-diphosphate (FPP) or geranylgeranyl-diphosphate (GGPP), from which higher isoprenoids are formed.

[00447] Useful products can also include small alkanes (for example, 1 to approximately 4 carbons) such as methane, ethane, propane, or butane, which may be used for heating (such as in cooking) or making plastics. Products may also include molecules with a carbon backbone of approximately 5 to approximately 9 carbon atoms, such as naphtha or ligroin, or their precursors. Other products may include molecules with a carbon background of 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 useful as products. Other products include lubricating oil, heavy gas oil, or fuel oil, or their precursors, and can contain alkanes, cycloalkanes, or aromatics of approximately 12 to approximately 70 carbons. Products also include other residuals that can be derived from or found in crude oil, such as coke, asphalt, tar, and waxes, generally containing multiple rings with about 70 or more carbons, and their precursors.

[00448] Examples of products, which can include the isoprenoids of the present disclosure, are fuel products, fragrance products, and insecticide products. In some instances, a product may be used directly. In other instances, the product

may be used as a "feedstock" to produce another product. For example, where the product is an isoprenoid, the isoprenoid may be hydrogenated and "cracked" to produce a shorter chain hydrocarbon (e.g., farnesene is hydrogenated to produce farnesene which is then cracked to produce propane, butane, octane, or other fuel products).

[00449] Modified organisms can be grown, in some embodiments in the presence of CO₂, to produce a desired polypeptide. In some embodiments, the products produced by the modified organism are isolated or collected. Collected products, such as terpenes and terpenoids, may then be further modified, for example, by refining and/or cracking to produce fuel molecules or components.

[00450] The various products may be further refined to a final product for an end user by a number of processes. Refining can, for example, occur by fractional distillation. For example, a mixture of products, such as a mix of different hydrocarbons with various chain lengths may be separated into various components by fractional distillation.

[00451] Refining may also include any one or more of the following steps, cracking, unifying, or altering the product. Large products, such as large hydrocarbons (e.g. $\geq C_{10}$), may be broken down into smaller fragments by cracking. Cracking may be performed by heat or high pressure, such as by steam, visbreaking, or coking. Products may also be refined by visbreaking, for example by thermally cracking large hydrocarbon molecules in the product by heating the product in a furnace. Refining may also include coking, wherein a heavy, almost pure carbon residue is produced. Cracking may also be performed by catalytic means to enhance the rate of the cracking reaction by using catalysts such as, but not limited to, zeolite, aluminum hydrosilicate, bauxite, or silica-alumina. Catalysis may be by fluid catalytic cracking, whereby a hot catalyst, such as zeolite, is used to catalyze cracking reactions. Catalysis may also be performed by hydrocracking, where lower temperatures are generally used in comparison to fluid catalytic cracking. Hydrocracking can occur in the presence of elevated partial pressure of hydrogen gas. Products may be refined by catalytic cracking to generate diesel, gasoline, and/or kerosene.

[00452] The products may also be refined by combining them in a unification step, for example by using catalysts, such as platinum or a platinum-rhenium mix. The unification process can produce hydrogen gas, a by-product, which may be used in cracking.

[00453] The products may also be refined by altering, rearranging, or restructuring hydrocarbons into smaller molecules. There are a number of chemical reactions that occur in catalytic reforming processes which are known to one of ordinary skill in the arts. Catalytic reforming can be performed in the presence of a catalyst and a high partial pressure of hydrogen. One common process is alkylation. For example, propylene and butylene are mixed with a catalyst such as hydrofluoric acid or sulfuric acid, and the resulting products are high octane hydrocarbons, which can be used to reduce knocking in gasoline blends.

[00454] The products may also be blended or combined into mixtures to obtain an end product. For example, the products may be blended to form gasoline of various grades, gasoline with or without additives, lubricating oils of various weights and grades, kerosene of various grades, jet fuel, diesel fuel, heating oil, and chemicals for making plastics and other polymers. Compositions of the products described herein may be combined or blended with fuel products produced by other means.

[00455] Some products produced from the host cells of the disclosure, especially after refining, will be identical to existing petrochemicals, i.e. contain the same chemical structure. For instance, crude oil contains the isoprenoid pristane, which is thought to be a breakdown product of phytol, which is a component of chlorophyll. Some of the products may not be the same as existing petrochemicals. However, although a molecule may not exist in conventional

petrochemicals or refining, it may still be useful in these industries. For example, a hydrocarbon could be produced that is in the boiling point range of gasoline, and that could be used as gasoline or an additive, even though the hydrocarbon does not normally occur in gasoline.

[00456] A product herein can be described by its Carbon Isotope Distribution (CID). At the molecular level, a CID is the statistical likelihood of a single carbon atom within a molecule to be one of the naturally occurring carbon isotopes (for example, ^{12}C , ^{13}C , or ^{14}C). At the bulk level of a product, a CID may be the relative abundance of naturally occurring carbon isotopes (for example, ^{12}C , ^{13}C , or ^{14}C) in a compound containing at least one carbon atom. It is noted that the CID of a fossil fuel may differ based on its source. For example, with CID(fos), the CID of carbon in a fossil fuel, such as petroleum, natural gas, and coal is distinguishable from the CID(atm), the CID of carbon in current atmospheric carbon dioxide. Additionally, the CID(photo-atm) refers to the CID of a carbon-based compound made by photosynthesis in recent history where the source of inorganic carbon was carbon dioxide in the atmosphere. Also, CID(photo-fos) refers to the CID of a carbon based compound made by photosynthesis in recent history where the source of substantially all of the inorganic carbon was carbon dioxide produced by the burning of fossil fuels (for example, coal, natural gas, and/or petroleum). The exact distribution is also a characteristic of 1) the type of photosynthetic organism that produced the molecule, and 2) the source of inorganic carbon. These isotope distributions can be used to define the composition of photosynthetically-derived fuel products. Carbon isotopes are unevenly distributed among and within different compounds and the isotopic distribution can reveal information about the physical, chemical, and metabolic processes involved in carbon transformation. The overall abundance of ^{13}C relative to ^{12}C in a photosynthetic organism is often less than the overall abundance of ^{13}C relative to ^{12}C in atmospheric carbon dioxide, indicating that carbon isotope discrimination occurs in the incorporation of carbon dioxide into photosynthetic biomass.

[00457] A product, either before or after refining, can be identical to an existing petrochemical. Some of the fuel products may not be the same as existing petrochemicals. In one embodiment, a fuel product is similar to an existing petrochemical, except for the carbon isotope distribution. For example, it is believed that no fossil fuel petrochemicals have a $\delta^{13}\text{C}$ distribution of less than -32%, whereas fuel products as described herein can have a $\delta^{13}\text{C}$ distribution of less than -32%, less than -35%, less than -40%, less than -45%, less than -50%, less than -55%, or less than -60%. In another embodiment, a fuel product or composition is similar but not the same as an existing fossil fuel petrochemical and has a $\delta^{13}\text{C}$ distribution of less than -32%, less than -35%, less than -40%, less than -45%, less than -50%, less than -55%, or less than -60%.

[00458] A fuel product can be a composition comprising, for example, hydrogen and carbon molecules, wherein the hydrogen and carbon molecules are at least about 80% of the atomic weight of the composition, and wherein the $\delta^{13}\text{C}$ distribution of the composition is less than about -32%. For some fuel products described herein, the hydrogen and carbon molecules are at least 90% of the atomic weight of the composition. For example, a biodiesel or fatty acid methyl ester (which has less than 90% hydrogen and carbon molecules by weight) may not be part of the composition. In still other compositions, the hydrogen and carbon molecules are at least 95 or at least 99% of the atomic weight of the composition. In yet other compositions, the hydrogen and carbon molecules are 100% of the atomic weight of the composition. In some embodiments, the composition is a liquid. In other embodiments, the composition is a fuel additive or a fuel product.

[00459] Also described herein is a fuel product comprising a composition comprising: hydrogen and carbon molecules, wherein the hydrogen and carbon molecules are at least 80% of the atomic weight of the composition, and wherein the $\delta^{13}\text{C}$ distribution of the composition is less than -32%; and a fuel component. In some embodiments, the $\delta^{13}\text{C}$ distribution of the composition is less than about -35%, less than about -40%, less than about -45%, less than about -50%, less than about -55%, or less than about -60%. In some embodiments, the fuel component of the composition is a blending fuel, for example, a fossil fuel, gasoline, diesel, ethanol, jet fuel, or any combination thereof. In still other embodiments, the blending fuel has a $\delta^{13}\text{C}$ distribution of greater than -32%. For some fuel products described herein, the fuel component is a fuel additive which may be MTBE, an anti-oxidant, an antistatic agent, a corrosion inhibitor, or any combination thereof. A fuel product as described herein may be a product generated by blending a fuel product as described and a fuel component. In some embodiments, the fuel product has a $\delta^{13}\text{C}$ distribution of greater than -32%. In other embodiments, the fuel product has a $\delta^{13}\text{C}$ distribution of less than -32%. For example, an oil composition extracted from an organism can be blended with a fuel component prior to refining (for example, cracking) in order to generate a fuel product as described herein. A fuel component, can be a fossil fuel, or a mixing blend for generating a fuel product. For example, a mixture for fuel blending may be a hydrocarbon mixture that is suitable for blending with another hydrocarbon mixture to generate a fuel product. For example, a mixture of light alkanes may not have a certain octane number to be suitable for a type of fuel, however, it can be blended with a high octane mixture to generate a fuel product. In another example, a composition with a $\delta^{13}\text{C}$ distribution of less than -32% is blended with a hydrocarbon mixture for fuel blending to create a fuel product. In some embodiments, the composition or fuel component alone are not suitable as a fuel product, however, when combined, they are useful as a fuel product. In other embodiments, either the composition or the fuel component or both individually are suitable as a fuel product. In yet another embodiment, the fuel component is an existing petroleum product, such as gasoline or jet fuel. In other embodiments, the fuel component is derived from a renewable resource, such as bioethanol, biodiesel, and biogasoline.

[00460] Oil compositions, derived from biomass obtained from a host cell, can be used for producing high-octane hydrocarbon products. Thus, one embodiment describes a method of forming a fuel product, comprising: obtaining an upgraded oil composition, cracking the oil composition, and blending the resulting one or more light hydrocarbons, having 4 to 12 carbons and an Octane number of 80 or higher, with a hydrocarbon having an Octane number of 80 or less. The hydrocarbons having an Octane number of 80 or less are, for example, fossil fuels derived from refining crude oil.

[00461] The biomass feedstock obtained from a host organism can be modified or tagged such that the light hydrocarbon products can be identified or traced back to their original feedstock. For example, carbon isotopes can be introduced into a biomass hydrocarbon in the course of its biosynthesis. The tagged hydrocarbon feedstock can be subjected to the refining processes described herein to produce a light hydrocarbon product tagged with a carbon isotope. The isotopes allow for the identification of the tagged products, either alone or in combination with other untagged products, such that the tagged products can be traced back to their original biomass feedstocks.

[00462] Table 1. Examples of Enzymes Involved in the Isoprenoid Pathway

[00463] The enzymes utilized may be encoded by nucleotide sequences derived from any organism, including bacteria, plants, fungi and animals. In some instances, the enzymes are isoprenoid producing enzymes. As used herein, an "isoprenoid producing enzyme" is a naturally or non-naturally occurring enzyme which produces or increases production of an isoprenoid. 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. Non-limiting examples of such enzymes and their sources are shown in **Table 1**.

<u>Synthase</u>	<u>Source</u>	<u>NCBI protein ID</u>
Limonene	<i>M. spicata</i>	2ONH A
Cineole	<i>S. officinalis</i>	AAC26016
Pinene	<i>A. grandis</i>	AAK83564
Camphene	<i>A. grandis</i>	AAB70707
Sabinene	<i>S. officinalis</i>	AAC26018
Myrcene	<i>A. grandis</i>	AAB71084
Abietadiene	<i>A. grandis</i>	Q38710
Taxadiene	<i>T. brevifolia</i>	AAK83566
FPP	<i>G. gallus</i>	P08836
Amorphadiene	<i>A. annua</i>	AAF61439
Bisabolene	<i>A. grandis</i>	O81086
Diapophytoene	<i>S. aureus</i>	
Diapophytoene desaturase	<i>S. aureus</i>	
GPPS-LSU	<i>M. spicata</i>	AAF08793
GPPS-SSU	<i>M. spicata</i>	AAF08792
GPPS	<i>A. thaliana</i>	CAC16849
GPPS	<i>C. reinhardtii</i>	EDP05515
FPP	<i>E. coli</i>	NP_414955
FPP	<i>A. thaliana</i>	NP_199588
FPP	<i>A. thaliana</i>	NP_193452
FPP	<i>C. reinhardtii</i>	EDP03194
IPP isomerase	<i>E. coli</i>	NP_417365
IPP isomerase	<i>H. pluvialis</i>	ABB80114
Limonene	<i>L. angustifolia</i>	ABB73044
Monoterpene	<i>S. lycopersicum</i>	AAX69064
Terpinolene	<i>O. basilicum</i>	AAV63792
Myrcene	<i>O. basilicum</i>	AAV63791
Zingiberene	<i>O. basilicum</i>	AAV63788
Myrcene	<i>Q. ilex</i>	CAC41012
Myrcene	<i>P. abies</i>	AAS47696
Myrcene, ocimene	<i>A. thaliana</i>	NP_179998
Myrcene, ocimene	<i>A. thaliana</i>	NP_567511
Sesquiterpene	<i>Z. mays</i> ; B73	AAS88571

Sesquiterpene	<i>A. thaliana</i>	NP_199276
Sesquiterpene	<i>A. thaliana</i>	NP_193064
Sesquiterpene	<i>A. thaliana</i>	NP_193066
Curcumene	<i>P. cablin</i>	AAS86319
Farnesene	<i>M. domestica</i>	AAX19772
Farnesene	<i>C. sativus</i>	AAU05951
Farnesene	<i>C. junos</i>	AAK54279
Farnesene	<i>P. abies</i>	AAS47697
Bisabolene	<i>P. abies</i>	AAS47689
Sesquiterpene	<i>A. thaliana</i>	NP_197784
Sesquiterpene	<i>A. thaliana</i>	NP_175313
GPP Chimera		
GPPS-LSU+SSU fusion		
Geranylgeranyl reductase	<i>A. thaliana</i>	NP_177587
Geranylgeranyl reductase	<i>C. reinhardtii</i>	EDP09986
Chlorophyllidohydrolase	<i>C. reinhardtii</i>	EDP01364
Chlorophyllidohydrolase	<i>A. thaliana</i>	NP_564094
Chlorophyllidohydrolase	<i>A. thaliana</i>	NP_199199
Phosphatase	<i>S. cerevisiae</i>	AAB64930
FPP A118W	<i>G. gallus</i>	

[00464] Codon Optimization

[00465] As discussed above, one or more codons of an encoding polynucleotide can be "biased" or "optimized" to reflect the codon usage of the host organism. For example, one or more codons of an encoding polynucleotide can be "biased" or "optimized" to reflect chloroplast codon usage (Table 2) or nuclear codon usage (Table 3). Most amino acids are encoded by two or more different (degenerate) codons, and it is well recognized that various organisms utilize certain codons in preference to others. "Biased" or codon "optimized" can be used interchangeably throughout the specification. Codon bias can be variously skewed in different plants, including, for example, in alga as compared to tobacco. Generally, the codon bias selected reflects codon usage of the plant (or organelle therein) which is being transformed with the nucleic acids of the present disclosure.

[00466] A polynucleotide that is biased for a particular codon usage can be synthesized de novo, or can be genetically modified using routine recombinant DNA techniques, for example, by a site directed mutagenesis method, to change one or more codons such that they are biased for chloroplast codon usage.

[00467] Such preferential codon usage, which is utilized in chloroplasts, is referred to herein as "chloroplast codon usage." Table 2 (below) shows the chloroplast codon usage for *C. reinhardtii* (see U.S. Patent Application Publication No.: 2004/0014174, published January 22, 2004).

[00468] Table 2

Chloroplast Codon Usage in <i>Chlamydomonas reinhardtii</i>
--

UUU 34.1*(348**)	UCU 19.4(198)	UAU 23.7(242)	UGU 8.5(87)
UUC 14.2(145)	UCC 4.9(50)	UAC 10.4(106)	UGC 2.6(27)
UUA 72.8(742)	UCA 20.4(208)	UAA 2.7(28)	UGA 0.1(1)
UUG 5.6(57)	UCG 5.2(53)	UAG 0.7(7)	UGG 13.7(140)
CUU 14.8(151)	CCU 14.9(152)	CAU 11.1(113)	CGU 25.5(260)
CUC 1.0(10)	CCC 5.4(55)	CAC 8.4(86)	CGC 5.1(52)
CUA 6.8(69)	CCA 19.3(197)	CAA 34.8(355)	CGA 3.8(39)
CUG 7.2(73)	CCG 3.0(31)	CAG 5.4(55)	CGG 0.5(5)
AUU 44.6(455)	ACU 23.3(237)	AAU 44.0(449)	AGU 16.9(172)
AUC 9.7(99)	ACC 7.8(80)	AAC 19.7(201)	AGC 6.7(68)
AUA 8.2(84)	ACA 29.3(299)	AAA 61.5(627)	AGA 5.0(51)
AUG 23.3(238)	ACG 4.2(43)	AAG 11.0(112)	AGG 1.5(15)
GUU 27.5(280)	GCU 30.6(312)	GAU 23.8(243)	GGU 40.0(408)
GUC 4.6(47)	GCC 11.1(113)	GAC 11.6(118)	GGC 8.7(89)
GUA 26.4(269)	GCA 19.9(203)	GAA 40.3(411)	GGA 9.6(98)
GUG 7.1(72)	GCG 4.3(44)	GAG 6.9(70)	GGG 4.3(44)

[00469] * -Frequency of codon usage per 1,000 codons. ** - Number of times observed in 36 chloroplast coding sequences (10,193 codons).

[00470] The chloroplast codon bias can, but need not, be selected based on a particular organism in which a synthetic polynucleotide is to be expressed. The manipulation can be a change to a codon, for example, by a method such as site directed mutagenesis, by a method such as PCR using a primer that is mismatched for the nucleotide(s) to be changed such that the amplification product is biased to reflect chloroplast codon usage, or can be the de novo synthesis of polynucleotide sequence such that the change (bias) is introduced as a consequence of the synthesis procedure.

[00471] In addition to utilizing chloroplast codon bias as a means to provide efficient translation of a polypeptide, it will be recognized that an alternative means for obtaining efficient translation of a polypeptide in a chloroplast is to re-engineer the chloroplast genome (e.g., a *C. reinhardtii* chloroplast genome) for the expression of tRNAs not otherwise expressed in the chloroplast genome. Such an engineered algae expressing one or more exogenous tRNA molecules provides the advantage that it would obviate a requirement to modify every polynucleotide of interest that is to be introduced into and expressed from a chloroplast genome; instead, algae such as *C. reinhardtii* that comprise a genetically modified chloroplast genome can be provided and utilized for efficient translation of a polypeptide according to any method of the disclosure. Correlations between tRNA abundance and codon usage in highly expressed genes is well known (for example, as described in Franklin et al., Plant J. 30:733-744, 2002; Dong et al., J. Mol. Biol. 260:649-663, 1996; Duret, Trends Genet. 16:287-289, 2000; Goldman et. al., J. Mol. Biol. 245:467-473, 1995; and Komar et. al., Biol. Chem. 379:1295-1300, 1998). In *E. coli*, for example, re-engineering of strains to express underutilized tRNAs resulted in enhanced expression of genes which utilize these codons (see Novy et al., in Novations 12:1-3, 2001). Utilizing endogenous tRNA genes, site directed mutagenesis can be used to make a synthetic tRNA

gene, which can be introduced into chloroplasts to complement rare or unused tRNA genes in a chloroplast genome, such as a *C. reinhardtii* chloroplast genome.

[00472] Generally, the chloroplast codon bias selected for purposes of the present disclosure, including, for example, in preparing a synthetic polynucleotide as disclosed herein reflects chloroplast codon usage of a plant chloroplast, and includes a codon bias that, with respect to the third position of a codon, is skewed towards A/T, for example, where the third position has greater than about 66% AT bias, or greater than about 70% AT bias. In one embodiment, the chloroplast codon usage is biased to reflect alga chloroplast codon usage, for example, *C. reinhardtii*, which has about 74.6% AT bias in the third codon position. Preferred codon usage in the chloroplasts of algae has been described in US 2004/0014174.

[00473] Table 3 exemplifies codons that are preferentially used in algal nuclear genes. The nuclear codon bias can, but need not, be selected based on a particular organism in which a synthetic polynucleotide is to be expressed. The manipulation can be a change to a codon, for example, by a method such as site directed mutagenesis, by a method such as PCR using a primer that is mismatched for the nucleotide(s) to be changed such that the amplification product is biased to reflect nuclear codon usage, or can be the de novo synthesis of polynucleotide sequence such that the change (bias) is introduced as a consequence of the synthesis procedure.

[00474] In addition to utilizing nuclear codon bias as a means to provide efficient translation of a polypeptide, it will be recognized that an alternative means for obtaining efficient translation of a polypeptide in a nucleus is to re-engineer the nuclear genome (e.g., a *C. reinhardtii* nuclear genome) for the expression of tRNAs not otherwise expressed in the nuclear genome. Such an engineered algae expressing one or more exogenous tRNA molecules provides the advantage that it would obviate a requirement to modify every polynucleotide of interest that is to be introduced into and expressed from a nuclear genome; instead, algae such as *C. reinhardtii* that comprise a genetically modified nuclear genome can be provided and utilized for efficient translation of a polypeptide according to any method of the disclosure.

Correlations between tRNA abundance and codon usage in highly expressed genes is well known (for example, as described in Franklin et al., Plant J. 30:733-744, 2002; Dong et al., J. Mol. Biol. 260:649-663, 1996; Duret, Trends Genet. 16:287-289, 2000; Goldman et. Al., J. Mol. Biol. 245:467-473, 1995; and Komar et. Al., Biol. Chem. 379:1295-1300, 1998). In *E. coli*, for example, re-engineering of strains to express underutilized tRNAs resulted in enhanced expression of genes which utilize these codons (see Novy et al., in Novations 12:1-3, 2001). Utilizing endogenous tRNA genes, site directed mutagenesis can be used to make a synthetic tRNA gene, which can be introduced into the nucleus to complement rare or unused tRNA genes in a nuclear genome, such as a *C. reinhardtii* nuclear genome.

[00475] Generally, the nuclear codon bias selected for purposes of the present disclosure, including, for example, in preparing a synthetic polynucleotide as disclosed herein, can reflect nuclear codon usage of an algal nucleus and includes a codon bias that results in the coding sequence containing greater than 60% G/C content.

[00476] **Table 3**

[00477] fields: [triplet] [frequency: per thousand] ([number]) Coding GC 66.30% 1st letter GC 64.80% 2nd letter GC 47.90% 3rd letter GC 86.21%

Nuclear Codon Usage in <i>Chlamydomonas reinhardtii</i>			
UUU 5.0 (2110)	UCU 4.7 (1992)	UAU 2.6 (1085)	UGU 1.4 (601)
UUC 27.1 (11411)	UCC 16.1 (6782)	UAC 22.8 (9579)	UGC 13.1 (5498)

UUA 0.6 (247)	UCA 3.2 (1348)	UAA 1.0 (441)	UGA 0.5 (227)
UUG 4.0 (1673)	UCG 16.1 (6763)	UAG 0.4 (183)	UGG 13.2 (5559)
CUU 4.4 (1869)	CCU 8.1 (3416)	CAU 2.2 (919)	CGU 4.9 (2071)
CUC 13.0 (5480)	CCC 29.5 (12409)	CAC 17.2 (7252)	CGC 34.9 (14676)
CUA 2.6 (1086)	CCA 5.1 (2124)	CAA 4.2 (1780)	CGA 2.0 (841)
CUG 65.2 (27420)	CCG 20.7 (8684)	CAG 36.3 (15283)	CGG 11.2 (4711)
AUU 8.0 (3360)	ACU 5.2 (2171)	AAU 2.8 (1157)	AGU 2.6 (1089)
AUC 26.6 (11200)	ACC 27.7 (11663)	AAC 28.5 (11977)	AGC 22.8 (9590)
AUA 1.1 (443)	ACA 4.1 (1713)	AAA 2.4 (1028)	AGA 0.7 (287)
0AUG 25.7 (10796)	ACG 15.9 (6684)	AAG 43.3 (18212)	AGG 2.7 (1150)
GUU 5.1 (2158)	GCU 16.7 (7030)	GAU 6.7 (2805)	GGU 9.5 (3984)
GUC 15.4 (6496)	GCC 54.6 (22960)	GAC 41.7 (17519)	GGC 62.0 (26064)
GUA 2.0 (857)	GCA 10.6 (4467)	GAA 2.8 (1172)	GGA 5.0 (2084)
GUG 46.5 (19558)	GCG 44.4 (18688)	GAG 53.5 (22486)	GGG 9.7 (4087)

[00478] Table 4

[00479] Table 4 lists the codon selected at each position for backtranslating the protein to a DNA sequence for synthesis. The selected codon is the sequence recognized by the tRNA encoded in the chloroplast genome when present; the stop codon (TAA) is the codon most frequently present in the chloroplast encoded genes. If an undesired restriction site is created, the next best choice according to the regular *Chlamydomonas* chloroplast usage table that eliminates the restriction site is selected.

[00480] Table 4

Amino acid	Codon utilized
F	TTC
L	TTA
I	ATC
V	GTA
S	TCA
P	CCA
T	ACA
A	GCA
Y	TAC
H	CAC
Q	CAA
N	AAC
K	AAA
D	GAC

E	GAA
C	TGC
R	CGT
G	GGC
W	TGG
M	ATG
STOP	TAA

[00481] Percent Sequence Identity

[00482] One example of an algorithm that is suitable for determining percent sequence identity or sequence similarity between nucleic acid or polypeptide sequences is the BLAST algorithm, which is described, e.g., in Altschul *et al.*, *J. Mol. Biol.* 215:403-410 (1990). Software for performing BLAST analysis is publicly available through the National Center for Biotechnology Information. The BLAST algorithm parameters W, T, and X determine the sensitivity and speed of the alignment. The BLASTN program (for nucleotide sequences) uses as defaults a word length (W) of 11, an expectation (E) of 10, a cutoff of 100, M=-5, N=-4, and a comparison of both strands. For amino acid sequences, the BLASTP program uses as defaults a word length (W) of 3, an expectation (E) of 10, and the BLOSUM62 scoring matrix (as described, for example, in Henikoff & Henikoff (1989) *Proc. Natl. Acad. Sci. USA*, 89:10915). In addition to calculating percent sequence identity, the BLAST algorithm also can perform a statistical analysis of the similarity between two sequences (for example, as described in Karlin & Altschul, *Proc. Nat'l. Acad. Sci. USA*, 90:5873-5787 (1993)). One measure of similarity provided by the BLAST algorithm is the smallest sum probability (P(N)), which provides an indication of the probability by which a match between two nucleotide or amino acid sequences would occur by chance. For example, a nucleic acid is considered similar to a reference sequence if the smallest sum probability in a comparison of the test nucleic acid to the reference nucleic acid is less than about 0.1, less than about 0.01, or less than about 0.001.

[00483] Fatty Acids and Glycerol Lipids

[00484] The present disclosure describes host cells capable of making polypeptides that contribute to the accumulation and/or secretion of fatty acids, glycerol lipids, or oils, by transforming host cells (e.g., alga cells such as *C. reinhardtii*, *D. salina*, *H. pluvialis*, and cyanobacterial cells) with nucleic acids encoding one or more different enzymes. Examples of such enzymes include acetyl-CoA carboxylase, ketoreductase, thioesterase, malonyltransferase, dehydratase, acyl-CoA ligase, ketoacylsynthase, enoylreductase, and desaturase. The enzymes can be, for example, catabolic or biodegrading enzymes.

[00485] In some instances, the host cell will naturally produce the fatty acid, glycerol lipid, triglyceride, or oil of interest. Therefore, transformation of the host cell with a polynucleotide encoding an enzyme, for example an ACCase, will allow for the increased activity of the enzyme and/or increased accumulation and/or secretion of a molecule of interest (e.g., a lipid) in the cell.

[00486] A change in the accumulation and/or secretion of a desired product, for example, fatty acids, glycerol lipids, or oils, by a transformed host cell can include, for example, a change in the total oil content over that normally present in the cell, or a change in the type of oil that is normally present in the cell.

[00487] A change in the accumulation and/or secretion of a desired product, for example, fatty acids, glycerol lipids, or oils, by a transformed host cell can include, for example, a change in the total lipid content over that normally present in the cell, or a change in the type of lipids that are normally present in the cell.

[00488] Increased malonyl CoA production is required for increased. Increased fatty acid biosynthesis is required for increased accumulation of fatty acid based lipids. An increase in fatty acid based lipids can be measured by methyl tert-butyl ether (MTBE) extraction.

[00489] 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 glycerolipid synthesis pathway. One example of a pathway might include genes encoding an acetyl CoA carboxylase, a malonyltransferase, a ketoacylsynthase, and a thioesterase. Cells transformed with an entire pathway and/or enzymes extracted from those cells, can synthesize, for example, complete fatty acids or intermediates of the fatty acid synthesis pathway. Constructs may contain, for example, multiple copies of the same gene, multiple genes encoding the same enzyme from different organisms, and/or multiple genes with one or more mutations in the coding sequence(s).

[00490] The enzyme(s) produced by the modified cells may result in the production of fatty acids, glycerol lipids, triglycerides, or oils that may be collected from the cells and/or the surrounding environment (e.g., bioreactor or growth medium). In some embodiments, the collection of the fatty acids, glycerol lipids, triglycerides, or oils is performed after the product is secreted from the cell via a cell membrane transporter.

[00491] Examples of candidate *Chlamydomonas* genes encoding enzymes of glycerolipid metabolism that can be used in the described embodiments are described in The Chlamydomonas Sourcebook Second Edition, Organellar and Metabolic Processes, Vol. 2, pp. 41-68, David B. Stern (Ed.), (2009), Elsevier Academic Press.

[00492] For example, enzymes involved in plastid, mitochondrial, and cytosolic pathways, along with plastidic and cytosolic isoforms of fatty acid desaturases, and triglyceride synthesis enzymes are described (and their accession numbers provided). An exemplary chart of some of the genes described is provided below:

Acyl-ACP thioesterase	FAT1	EDP08596
Long-chain acyl-CoA synthetase	LCS1	EDO96800
CDP-DAG: Inositol phosphotransferase	PIS1	EDP06395
Acyl-CoA: Diacylglycerol acyltransferase	DGA1	EDO96893
Phospholipid: Diacylglycerol acyltransferase	LRO1(LCA1)	EDP07444

[00493] Examples of the types of fatty acids and/or glycerol lipids that a host cell or organism can produce, are described below.

[00494] Lipids are a broad group of naturally occurring molecules which includes fats, waxes, sterols, fat-soluble vitamins (such as vitamins A, D, E and K), monoglycerides, diglycerides, phospholipids, and others. The main biological functions of lipids include energy storage, as structural components of cell membranes, and as important signaling molecules.

[00495] Lipids may be broadly defined as hydrophobic or amphiphilic small molecules; the amphiphilic nature of some lipids allows them to form structures such as vesicles, liposomes, or membranes in an aqueous environment. Biological lipids originate entirely or in part from two distinct types of biochemical subunits or "building blocks": ketoacyl and isoprene groups. Lipids may be divided into eight categories: fatty acyls, glycerolipids, glycerophospholipids,

sphingolipids, saccharolipids and polyketides (derived from condensation of ketoacyl subunits); and sterol lipids and prenol lipids (derived from condensation of isoprene subunits). For this disclosure, saccharolipids will not be discussed. [00496] Fats are a subgroup of lipids called triglycerides. Lipids also encompass molecules such as fatty acids and their derivatives (including tri-, di-, and monoglycerides and phospholipids), as well as other sterol-containing metabolites such as cholesterol. Humans and other mammals use various biosynthetic pathways to both break down and synthesize lipids.

[00497] Fatty Acyls

[00498] Fatty acyls, a generic term for describing fatty acids, their conjugates and derivatives, are a diverse group of molecules synthesized by chain-elongation of an acetyl-CoA primer with malonyl-CoA or methylmalonyl-CoA groups in a process called fatty acid synthesis. A fatty acid is any of the aliphatic monocarboxylic acids that can be liberated by hydrolysis from naturally occurring fats and oils. They are made of a hydrocarbon chain that terminates with a carboxylic acid group; this arrangement confers the molecule with a polar, hydrophilic end, and a nonpolar, hydrophobic end that is insoluble in water. The fatty acid structure is one of the most fundamental categories of biological lipids, and is commonly used as a building block of more structurally complex lipids. The carbon chain, typically between four to 24 carbons long, may be saturated or unsaturated, and may be attached to functional groups containing oxygen, halogens, nitrogen and sulfur; branched fatty acids and hydroxyl fatty acids also occur, and very long chain acids of over 30 carbons are found in waxes. Where a double bond exists, there is the possibility of either a cis or trans geometric isomerism, which significantly affects the molecule's molecular configuration. Cis-double bonds cause the fatty acid chain to bend, an effect that is more pronounced the more double bonds there are in a chain. This in turn plays an important role in the structure and function of cell membranes. Most naturally occurring fatty acids are of the cis configuration, although the trans form does exist in some natural and partially hydrogenated fats and oils.

[00499] Examples of biologically important fatty acids are the eicosanoids, derived primarily from arachidonic acid and eicosapentaenoic acid, which include prostaglandins, leukotrienes, and thromboxanes. Other major lipid classes in the fatty acid category are the fatty esters and fatty amides. Fatty esters include important biochemical intermediates such as wax esters, fatty acid thioester coenzyme A derivatives, fatty acid thioester ACP derivatives and fatty acid carnitines. The fatty amides include N-acyl ethanolamines.

[00500] Glycerolipids

[00501] Glycerolipids are composed mainly of mono-, di- and tri-substituted glycerols, the most well-known being the fatty acid esters of glycerol (triacylglycerols), also known as triglycerides. In these compounds, the three hydroxyl groups of glycerol are each esterified, usually by different fatty acids. Because they function as a food store, these lipids comprise the bulk of storage fat in animal tissues. The hydrolysis of the ester bonds of triacylglycerols and the release of glycerol and fatty acids from adipose tissue is called fat mobilization.

[00502] Additional subclasses of glycerolipids are represented by glycosylglycerols, which are characterized by the presence of one or more sugar residues attached to glycerol via a glycosidic linkage. An example of a structure in this category is the digalactosyldiacylglycerols found in plant membranes.

[00503] Exemplary *Chlamydomonas* glycerolipids include: DGDG, digalactosyldiacylglycerol; DGTS, diacylglycerol-N, N, N-trimethylhomoserine; MGDG, monogalactosyldiacylglycerol; PtdEtn, phosphatidylethanolamine; PtdGro, phosphatidylglycerol; PtdIns, phosphatidylinositol; SQDG, sulfoquinovosyldiacylglycerol; and TAG, triacylglycerol.

[00504] Glycerophospholipids

[00505] Glycerophospholipids are any derivative of glycerophosphoric acid that contains at least one *O*-acyl, *O*-alkyl, or *O*-alkenyl group attached to the glycerol residue. The common glycerophospholipids are named as derivatives of phosphatidic acid (phosphatidyl choline, phosphatidyl serine, and phosphatidyl ethanolamine).

[00506] Glycerophospholipids, also referred to as phospholipids, are ubiquitous in nature and are key components of the lipid bilayer of cells, as well as being involved in metabolism and cell signaling. Glycerophospholipids may be subdivided into distinct classes, based on the nature of the polar headgroup at the *sn*-3 position of the glycerol backbone in eukaryotes and eubacteria, or the *sn*-1 position in the case of archaeobacteria.

[00507] Examples of glycerophospholipids found in biological membranes are phosphatidylcholine (also known as PC, GPCho or lecithin), phosphatidylethanolamine (PE or GPEtn) and phosphatidylserine (PS or GPSer). In addition to serving as a primary component of cellular membranes and binding sites for intra- and intercellular proteins, some glycerophospholipids in eukaryotic cells, such as phosphatidylinositols and phosphatidic acids are either precursors of, or are themselves, membrane-derived second messengers. Typically, one or both of these hydroxyl groups are acylated with long-chain fatty acids, but there are also alkyl-linked and 1Z-alkenyl-linked (plasmalogen) glycerophospholipids, as well as dialkylether variants in archaeobacteria.

[00508] Sphingolipids

[00509] Sphingolipids are any of class of lipids containing the long-chain amino diol, sphingosine, or a closely related base (i.e. a sphingoid). A fatty acid is bound in an amide linkage to the amino group and the terminal hydroxyl may be linked to a number of residues such as a phosphate ester or a carbohydrate. The predominant base in animals is sphingosine while in plants it is phytosphingosine.

[00510] The main classes are: (1) phosphosphingolipids (also known as sphingophospholipids), of which the main representative is sphingomyelin; and (2) glycosphingolipids, which contain at least one monosaccharide and a sphingoid, and include the cerebrosides and gangliosides. Sphingolipids play an important structural role in cell membranes and may be involved in the regulation of protein kinase C.

[00511] As mentioned above, sphingolipids are a complex family of compounds that share a common structural feature, a sphingoid base backbone, and are synthesized *de novo* from the amino acid serine and a long-chain fatty acyl CoA, that are then converted into ceramides, phosphosphingolipids, glycosphingolipids and other compounds. The major sphingoid base of mammals is commonly referred to as sphingosine. Ceramides (N-acyl-sphingoid bases) are a major subclass of sphingoid base derivatives with an amide-linked fatty acid. The fatty acids are typically saturated or mono-unsaturated with chain lengths from 16 to 26 carbon atoms.

[00512] The major phosphosphingolipids of mammals are sphingomyelins (ceramide phosphocholines), whereas insects contain mainly ceramide phosphoethanolamines, and fungi have phytoceramide phosphoinositols and mannose-containing headgroups. The glycosphingolipids are a diverse family of molecules composed of one or more sugar residues linked via a glycosidic bond to the sphingoid base. Examples of these are the simple and complex glycosphingolipids such as cerebrosides and gangliosides.

[00513] Sterol Lipids

[00514] Sterol lipids, such as cholesterol and its derivatives, are an important component of membrane lipids, along with the glycerophospholipids and sphingomyelins. The steroids, all derived from the same fused four-ring core structure, have different biological roles as hormones and signaling molecules. The eighteen-carbon (C18) steroids include the estrogen family whereas the C19 steroids comprise the androgens such as testosterone and androsterone.

The C21 subclass includes the progestogens as well as the glucocorticoids and mineralocorticoids. The secosteroids, comprising various forms of vitamin D, are characterized by cleavage of the B ring of the core structure. Other examples of sterols are the bile acids and their conjugates, which in mammals are oxidized derivatives of cholesterol and are synthesized in the liver. The plant equivalents are the phytosterols, such as β -sitosterol, stigmasterol, and brassicasterol; the latter compound is also used as a biomarker for algal growth. The predominant sterol in fungal cell membranes is ergosterol.

[00515] Prenol Lipids

[00516] Prenol lipids are synthesized from the 5-carbon precursors isopentenyl diphosphate and dimethylallyl diphosphate that are produced mainly via the mevalonic acid (MVA) pathway. The simple isoprenoids (for example, linear alcohols and diphosphates) are formed by the successive addition of C5 units, and are classified according to the number of these terpene units. Structures containing greater than 40 carbons are known as polyterpenes. Carotenoids are important simple isoprenoids that function as antioxidants and as precursors of vitamin A. Another biologically important class of molecules is exemplified by the quinones and hydroquinones, which contain an isoprenoid tail attached to a quinonoid core of non-isoprenoid origin. Prokaryotes synthesize polyprenols (called bactoprenols) in which the terminal isoprenoid unit attached to oxygen remains unsaturated, whereas in animal polyprenols (dolichols) the terminal isoprenoid is reduced.

[00517] Polyketides

[00518] Polyketides or sometimes acetogenin are any of a diverse group of natural products synthesized via linear poly- β -ketones, which are themselves formed by repetitive head-to-tail addition of acetyl (or substituted acetyl) units indirectly derived from acetate (or a substituted acetate) by a mechanism similar to that for fatty-acid biosynthesis but without the intermediate reductive steps. In many case, acetyl-CoA functions as the starter unit and malonyl-CoA as the extending unit. Various molecules other than acetyl-CoA may be used as starter, often with methylmalonyl-CoA as the extending unit. The poly- β -ketones so formed may undergo a variety of further types of reactions, which include alkylation, cyclization, glycosylation, oxidation, and reduction. The classes of product formed – and their corresponding starter substances – comprise *inter alia*: coniferyl (of hemlock) and orsellinate (of lichens) – acetyl-CoA; flavanoids and stilbenes – cinnamoyl-CoA; tetracyclines – amide of malonyl-CoA; urushiols (of poison ivy) – palmitoleoyl-CoA; and erythronolides – propionyl-CoA and methyl-malonyl-CoA as extender.

[00519] Polyketides comprise a large number of secondary metabolites and natural products from animal, plant, bacterial, fungal and marine sources, and have great structural diversity. Many polyketides are cyclic molecules whose backbones are often further modified by glycosylation, methylation, hydroxylation, oxidation, and/or other processes. Many commonly used anti-microbial, anti-parasitic, and anti-cancer agents are polyketides or polyketide derivatives, such as erythromycins, tetracyclines, avermectins, and antitumor epothilones.

[00520] The following examples are intended to provide illustrations of the application of the present disclosure. The following examples are not intended to completely define or otherwise limit the scope of the disclosure. One of skill in the art will appreciate that many other methods known in the art may be substituted in lieu of the ones specifically described or referenced herein.

EXAMPLES

[00521] Example 1: Transformation and Screening Methods

[00522] In this example, a method for transformation of *Scenedesmus sp.* is described. Algae cells are grown to log phase (approximately $0.5-1.0 \times 10^7$ cells/mL) in TAP medium (Gorman and Levine, *Proc. Natl. Acad. Sci., USA* 54:1665-1669, 1965, which is incorporated herein by reference) at 23°C under constant illumination of 50-100uE on a rotary shaker set at 100 rpm. Cells are harvested at 1000 x g for 5 min. The supernatant is decanted and cells are resuspended in TAP media at 10^8 cells/mL. 5×10^7 cells are spread on selective agar medium and transformed by particle bombardment with 550 nm or 1000nm diameter gold particles carrying the transforming DNA @ 375-500psi with the Helios Gene Gun (Bio-Rad) from a shot distance of 2-4cm. Desired algae clones are those that grow on selective media.

[00523] PCR is used to identify transformed algae strains. For PCR analysis, colony lysates are prepared by suspending algae cells (from agar plate or liquid culture) in lysis buffer (0.5% SDS, 100 mM NaCl, 10 mM EDTA, 75 mM Tris-HCl, pH 7.5) and heating to 98°C for 10 minutes, followed by cooling to near 23°C. Lysates are diluted 50-fold in 100mM Tris-HCl pH 7.5 and 2uL is used as template in a 25uL reaction. Alternatively, total genomic DNA preparations may be substituted for colony lysates. A PCR cocktail consisting of reaction buffer, dNTPs, PCR primer pair(s) (indicated in each example below), DNA polymerase, and water is prepared. Algal DNA is added to provide template for the reaction. Annealing temperature gradients are employed to determine optimal annealing temperature for specific primer pairs. In many cases, algae transformants are analyzed by PCR in two ways. First, primers are used that are specific for the transgene being introduced into the chloroplast genome. Desired algae transformants are those that give rise to PCR product(s) of expected size(s). Second, two sets of primer pairs are used to determine the degree to which the transforming DNA was integrated into the chloroplast genome (heteroplasmic vs. homoplasmic). The first pair of primers amplifies a region spanning the site of integration. The second pair of primers amplifies a constant, or control region that is not targeted by the transforming DNA, so should produce a product of expected size in all cases. This reaction confirms that the absence of a PCR product from the region spanning the site of integration did not result from cellular and/or other contaminants that inhibited the PCR reaction. Concentrations of the primer pairs are varied so that both amplicons are amplified in the same reaction. The number of cycles used is >30 to increase sensitivity. The most desired clones are those that yield a product for the constant region but not for the region spanning the site of integration. Once identified, clones are analyzed for changes in phenotype.

[00524] One of skill in the art will appreciate that many other transformation methods known in the art may be substituted in lieu of the ones specifically described or referenced herein.

[00525] Example 2: Chloroplast Transformation of *S. dimorphus* using 3-(3,4-Dichlorophenyl)-1,1-dimethylurea (DCMU) selection

[00526] In this example, DCMU resistance was established as a selection method for transformation of *S. dimorphus*. Transforming DNA (SEQ ID NO: 30, S264A fragment) is shown graphically in **Figure 1**. In this instance, a DNA fragment encompassing the 3' end of the gene encoding psbA and its 3' UTR from *S. dimorphus* was amplified by PCR, subcloned into pUC18, and mutated via Quikchange PCR (Stratagene) to generate a S264A mutation along with a silent *Xba*I restriction site. Nucleotide 1913 of the fragment was mutated from a T to a G for the S264A mutation, and nucleotides 1928 to 1930 were mutated from CGT to AGA to generate the silent *Xba*I restriction site.

[00527] Transforming DNA was introduced into *S. dimorphus* via particle bombardment (as described in **EXAMPLE 1**) with DNA carried on 1000 nm gold particles, @ 375 psi and a shooting distance of 2cm. Transformants were selected by growth on HSM media + 0.5uM DCMU under constant light 100-200uE @ 23°C for approximately 3 weeks.

[00528] Transformants were verified by PCR screening (as described in **EXAMPLE 1**) using primers (SEQ ID NO: 17 and SEQ ID NO: 14) specific for a 2.1 kb region surrounding the bases changed for the S264A mutation. The PCR products were then digested with *XbaI* to distinguish transformants from spontaneous mutants that may arise as a result of plating cells onto media containing DCMU. **Figure 2** shows that DNA amplified from clones 3 and 4 is completely digested by *XbaI* (indicating that clones 3 and 4 are bonafide transformants while DNA amplified from wildtype cells (WT) is not. These data were confirmed by DNA sequencing of the PCR product.

[00529] Transformants were grown to saturation in TAP media, diluted 1:100 in HSM + various concentrations of DCMU and grown under constant light 50-100uE with CO₂ enrichment for 4 days. **Figure 3** shows that transformants with the psbA S264A mutation grow in up to 10uM DCMU or 10uM Atrazine whereas wild type *S. dimorphus* (wt) fails to grow in 0.5uM DCMU or 0.5uM Atrazine.

[00530] In order to determine if DCMU selection could result in incorporation of an expression cassette downstream of the psbA gene, A vector was constructed containing an expression cassette consisting of an endogenous promoter, a chloramphenicol acetyltransferase (CAT) gene, and an endogenous terminator cloned ~500bp downstream of the S264A/*XbaI* mutated psbA gene fragment from *S. dimorphus* and including the rpl20 gene. Transforming DNA is shown graphically in **Figure 4**. In this instance the DNA segment labeled "CAT" is the chloramphenicol acetyl transferase gene from *E. coli*, the segment labeled "tufA" is the promoter and 5' UTR sequence for the *tufA* gene from *S. dimorphus*, and the segment labeled "rbcL" is the 3' UTR for the *rbcL* gene from *S. dimorphus*. The selection marker cassette is targeted to the *S. dimorphus* chloroplast genome via the segments labeled "Homology A2" and "Homology B2" which are 1000 bp fragments homologous to sequences of DNA adjacent to nucleotide 065,353 and include an S264A/*XbaI* mutated partial psbA coding sequence, its 3'UTR, and the rpl20 coding sequence. This vector targets integration of the selection marker cassette approximately 400bp 3' of the stop codon of the *psbA* gene.

[00531] Transforming DNA was introduced into *S. dimorphus* via particle bombardment (as described in **EXAMPLE 1**) with DNA carried on 550 nm gold particles, @ 500 psi and a shooting distance of 4cm. Transformants were selected by growth on HSM media + 1uM DCMU under constant light 100-200uE @ RT for approximately 3 weeks.

[00532] To determine if the transformants were resistant to chloramphenicol (CAM), they were patched onto TAP agar medium containing 25ug/mL CAM. In all cases, the DCMU transformants were also resistant to CAM indicating that the CAT cassette was incorporated into the genome.

[00533] One of skill in the art will appreciate that many other methods known in the art may be substituted in lieu of the ones specifically described or referenced.

[00534] Example 3: Use of Chloramphenicol Acetyl Transferase as a Selection Marker in *S. dimorphus*

[00535] In this example, a nucleic acid encoding chloramphenicol acetyl transferase gene (CAT) from *E. coli* was introduced into *S. dimorphus*. Transforming DNA is shown graphically in **Figure 5**. In this instance the DNA segment labeled "CAT" is the chloramphenicol acetyl transferase gene (SEQ ID NO: 28), the segment labeled "tufA" is the promoter and 5' UTR sequence for the *psbD* (SEQ ID NO: 40) or *tufA* gene (SEQ ID NO: 42) from *S. dimorphus*, and the segment labeled "rbcL 3' " is the 3' UTR for the *rbcL* gene from *S. dimorphus* (SEQ ID NO: 57). The selection marker cassette is targeted to the *S. dimorphus* chloroplast genome via the segments labeled "Homology A" and "Homology B" which are approximately 1000 bp fragments homologous to sequences of DNA adjacent to nucleotide 035,138 (Site 2; nucleotide locations according to the sequence available from NCBI for *S. obliquus*, NC_008101) on the 5' and 3' sides, respectively. All DNA segments were subcloned into pUC 18. All DNA manipulations carried out in

the construction of this transforming DNA were essentially as described by Sambrook et al., Molecular Cloning: A Laboratory Manual (Cold Spring Harbor Laboratory Press 1989) and Cohen et al., *Meth. Enzymol.* 297, 192-208, 1998.

[00536] Transforming DNA was introduced into *S. dimorphus* via particle bombardment according to the method described in **EXAMPLE 1** with DNA carried on 550 nm gold particles @ 500psi and a shooting distance of 4cm. Transformants were selected by growth on TAP agar medium + 25 µg/mL chloramphenicol (TAP-CAM) under constant light 50-100uE @ RT for approximately 2 weeks. Transformants were patched onto TAP-CAM agar medium, grown for 4 days under constant light.

[00537] Cells from the patched transformants were analyzed by PCR screening (as described in **EXAMPLE 1**). The presence of the CAT selection marker was determined using primers that amplify the entire 660 bp gene (SEQ ID NO: 18 and SEQ ID NO: 19). **Figure 6** shows that a 660bp fragment (representing the CAT gene) is amplified from DNA of several transformants (all lanes except +, - and ladders) while it is not amplified from DNA of wild type cells (-). DNA ladder is a 1 kb ladder.

[00538] One of skill in the art will appreciate that many other methods known in the art may be substituted in lieu of the ones specifically described or referenced.

[00539] Example 4: Production of Endoxylanase in *S. dimorphus*

[00540] In this example a nucleic acid encoding endoxylanase from *T. reesei* was introduced into *S. dimorphus*. Transforming DNA (p04-31) is shown graphically in **Figure 7**. In this instance the DNA segment labeled "BD11" is the endoxylanase encoding gene (SEQ ID NO: 21, BD11), the segment labeled "psbD" is the promoter and 5' UTR for the *psbD* gene from *S. dimorphus*, the segment labeled "D1 3'" is the 3' UTR for the *psbA* gene from *S. dimorphus*, and the segment labeled "CAT" is the chloramphenicol acetyl transferase gene (CAT) from *E. coli*, which is regulated by the promoter and 5' UTR sequence for the *tufA* gene from *S. dimorphus* and the 3' UTR sequence for the *rbcL* gene from *S. dimorphus*. The transgene expression cassette and selection marker are targeted to the *S. dimorphus* chloroplast genome via the segments labeled "Homology A" and "Homology B" which are approximately 1000 bp fragments homologous to sequences of DNA adjacent to nucleotide 071,366 (Site 1; nucleotide locations according to the sequence available from NCBI for *S. obliquus*, NC_008101) on the 5' and 3' sides, respectively. All DNA segments were subcloned into pUC 18 (gutless pUC). All DNA manipulations carried out in the construction of this transforming DNA were essentially as described by Sambrook et al., Molecular Cloning: A Laboratory Manual (Cold Spring Harbor Laboratory Press 1989) and Cohen et al., *Meth. Enzymol.* 297, 192-208, 1998.

[00541] Transforming DNA was introduced into *S. dimorphus* via particle bombardment according to the method described in **EXAMPLE 1** with DNA carried on 550 nm gold particles @ 500psi and a shooting distance of 4cm. Transformants were selected by growth on TAP-CAM agar medium under constant light 50-100uE @ RT for approximately 2 weeks. Transformants were streaked onto TAP-CAM agar medium to ensure single colony isolation and grown for 4 days under constant light.

[00542] Transformants were analyzed by PCR screening (as described in **EXAMPLE 1**). The degree to which the transforming DNA was integrated into the chloroplast genome was determined using primers that amplify a 400 bp constant region (SEQ ID NO: 1 and SEQ ID NO: 2) and a 250 bp region spanning the integration site (SEQ ID NO: 3 and SEQ ID NO: 4). Integration occurs approximately 1000bp 5' of the start codon of the *psbA* gene. **Figure 8** shows that subclones from two independent transformants (parent 2 and 4) are homoplasmic, i.e., only the constant region

(400bp product) was amplified, while in the control reactions (wt) both the constant region and the region spanning the integration site (250bp) were amplified.

[00543] To ensure that the presence of the endoxylanase-encoding gene led to expression of the endoxylanase protein, a Western blot was performed. Briefly, approximately 1×10^8 to 2×10^8 algae cells were collected from TAP agar medium and resuspended in approximately 1mL BugBuster solution (Novagen) in a 1.5mL eppendorf tube. 1.0 mm Zirconia beads (BioSpec Products, Inc) were then added to fill the tube with minimal headspace, ~500uL of beads. Cells were lysed in a bead beating apparatus (Mini Beadbeater™, BioSpec Products, Inc.) by shaking for 3-5 minutes three times. Cell lysates were clarified by centrifugation for 15 minutes at 20,000g and the supernatants were normalized for total soluble protein (Coomassie Plus Protein Assay Kit, Thermo Scientific). Samples were mixed 1:4 with loading buffer (XT Sample Buffer with β -mercaptoethanol, Bio-Rad), heated to 98°C for 5 min, cooled to 23°C, and proteins were separated by SDS-PAGE, followed by transfer to PVDF membrane. The membrane was blocked with Starting Block T20 Blocking Buffer (Thermo Scientific) for 15 min, incubated with horseradish peroxidase-linked anti-FLAG antibody (diluted 1:2,500 in Starting Block T20 Blocking Buffer) at 23°C for 2 hours, washed three times with TBST. Proteins were visualized with chemiluminescent detection. Results from multiple clones (Figure 9, parent 2 and 4) show that expression of the endoxylanase gene in *S. dimorphus* cells resulted in production of the protein.

[00544] To determine if the endoxylanase produced by transformed algae cells was functional, endoxylanase activity was tested using an enzyme function assay. Briefly, algae cells were collected from TAP agar medium and suspended in BugBuster solution (Novagen). Cells were lysed by bead beating using zirconium beads. Cell lysates were clarified by centrifugation and the supernatants were normalized for total soluble protein (Coomassie Plus Protein Assay Kit, Thermo Scientific). 100 μ L of each sample was mixed with 10 μ L of 10X xylanase assay buffer (1M sodium acetate, pH=4.8) and 50 μ L of the sample mixture was added to one well in a black 96-well plate. EnzCheck Ultra Xylanase substrate (Invitrogen) was dissolved at a concentration of 50ug/ml in 100mM sodium acetate pH 4.8, and 50 μ L of substrate was added to each well of the microplate. The fluorescent signal was measured in a SpectraMax M2 microplate reader (Molecular Devices), with an excitation wavelength of 360 nm and an emission wavelength of 460 nm, without a cutoff filter and with the plate chamber set to 42 degrees Celsius. The fluorescence signal was measured for 15 minutes, and the enzyme velocity was calculated with Softmax Pro v5.2 (Molecular Devices). Enzyme velocities were recorded as RFU/minute. Enzyme specific activities were calculated as milliRFU per minute per μ g of total soluble protein. Figure 10 shows that endoxylanase activity is at least 4 fold higher in transformants than in wild type cells and similar in velocity to a positive control (*Chlamydomonas* expressing endoxylanase algae cells).

[00545] These data demonstrate that the chloroplast of *S. dimorphus* can be transformed with foreign DNA containing an expression cassette with a selectable marker and a separate expression cassette with a gene encoding an endoxylanase, and the expressed proteins are functional. One of skill in the art will appreciate that many other methods known in the art may be substituted in lieu of the ones specifically described or referenced.

[00546] Example 5: Production of FPP Synthase in *S. dimorphus*

[00547] In this example a nucleic acid encoding FPP synthase from *G. gallus* was introduced into *S. dimorphus*. Transforming DNA is shown graphically in Figure 11. In this instance the DNA segment labeled "Is09" is the FPP synthase encoding gene (SEQ ID NO: 23 Is09), the segment labeled "psbD" is the promoter and 5' UTR for the *psbD* gene from *S. dimorphus*, the segment labeled "D1 3'UTR" is the 3' UTR for the *psbA* gene from *S. dimorphus*, and the segment labeled "CAT" is the chloramphenicol acetyl transferase gene (CAT) from *E. coli*, which is regulated by the

promoter and 5' UTR sequence for the *tufA* gene from *S. dimorphus* and the 3' UTR sequence for the *rbcL* gene from *S. dimorphus*. The transgene expression cassette and selection marker are targeted to the *S. dimorphus* chloroplast genome via the segments labeled "Homology A" and "Homology B" which are approximately 1000 bp fragments homologous to sequences of DNA adjacent to nucleotide 071,366 (Site 1; nucleotide locations according to the sequence available from NCBI for *S. obliquus*, NC_008101) on the 5' and 3' sides, respectively. All DNA segments were subcloned into pUC 18 (gutless pUC). All DNA manipulations carried out in the construction of this transforming DNA were essentially as described by Sambrook et al., *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory Press 1989) and Cohen et al., *Meth. Enzymol.* 297, 192-208, 1998.

[00548] Transforming DNA was introduced into *S. dimorphus* via particle bombardment according to the method described in **EXAMPLE 1** with DNA carried on 550 nm gold particles @ 500psi and a shooting distance of 4cm. Transformants were selected by growth on TAP-CAM agar medium under constant light 50-100uE @ RT for approximately 2 weeks. Transformants were streaked onto TAP-CAM agar medium to ensure single colony isolation and grown for 4 days under constant light.

[00549] Transformants were analyzed by PCR screening (as described in **EXAMPLE 1**). The degree to which the transforming DNA was integrated into the chloroplast genome was determined using primers that amplify a 400 bp constant region (SEQ ID NO: 1 and SEQ ID NO: 2) and a 250 bp region spanning the integration site (SEQ ID NO: 3 and SEQ ID NO: 4). **Figure 12** shows that seven independent transformants are homoplasmic, i.e., only the constant region (400bp product) was amplified, while in the control reactions (WT) both the constant region and the region spanning the integration site (250bp) were amplified.

[00550] To ensure that the presence of the FPP synthase-encoding gene led to expression of the FPP synthase protein, a Western blot was performed (as described in **EXAMPLE 4**). Results from multiple clones (**Figure 13**) show that expression of the FPP synthase gene in *S. dimorphus* cells resulted in production of the protein.

[00551] To determine if the FPP synthase produced by transformed algae cells was functional, FPP synthase activity was tested using an enzyme function assay. Algae cells were harvested from TAP media, resuspended in assay buffer (35mM HEPES, pH 7.4, 10mM MgCl₂, 5mM DTT) and lysed using zirconium beads in a bead beater. Crude lysate was clarified by centrifugation at 15,000 rpm for 20min. Isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP) were added to clarified lysates and the reaction allowed to proceed at 30C overnight. Reactions were then CIP treated for 4-6hours @ 37C in glycine buffer, pH 10.6, 5mM ZnCl₂. The samples were then overlaid with heptane and analyzed via GC/MS (**Figures 14A to G**). Additionally, IPP, DMAPP and *E. coli* purified amorpho-4, 11-diene were added to clarified lysates, the reactions allowed to proceed at 30 °C overnight, overlaid with heptane and analyzed via GC/MS (**Figures 15A to G**). For both methods, the diagnostic ions at m/Z 204 and 189 were detected in the engineered *S. dimorphus*, but not in the wt samples.

[00552] These data demonstrate that the chloroplast of *S. dimorphus* can be transformed with foreign DNA containing an expression cassette with a selectable marker and a separate expression cassette with a gene encoding an FPP synthase, and the expressed proteins are functional. One of skill in the art will appreciate that many other methods known in the art may be substituted in lieu of the ones specifically described or referenced.

[00553] Example 6: Production of Fusicoccadiene Synthase in *S. dimorphus*

[00554] In this example a nucleic acid encoding fusicoccadiene synthase from *P. amygdali* was introduced into *S. dimorphus*. Transforming DNA is shown graphically in **Figure 16**. In this instance the DNA segment labeled "Is88" is

the fusicoccadiene synthase encoding gene (SEQ ID NO: 25, Is88), the segment labeled “psbD” is the promoter and 5’ UTR for the *psbD* gene from *S. dimorphus*, the segment labeled “D1 3’ ” is the 3’ UTR for the *psbA* gene from *S. dimorphus*, and the segment labeled “CAT” is the chloramphenicol acetyl transferase gene (CAT) from *E. coli*, which is regulated by the promoter and 5’ UTR sequence for the *tufA* gene from *S. dimorphus* and the 3’ UTR sequence for the *rbcL* gene from *S. dimorphus*. The transgene expression cassette and selection marker are targeted to the *S. dimorphus* chloroplast genome via the segments labeled “Homology A” and “Homology B” which are approximately 1000 bp fragments homologous to sequences of DNA adjacent to nucleotide 071,366 (Site 1; nucleotide locations according to the sequence available from NCBI for *S. obliquus*, NC_008101) on the 5’ and 3’ sides, respectively. All segments were subcloned into pUC19. All DNA manipulations carried out in the construction of this transforming DNA were essentially as described by Sambrook et al., *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory Press 1989) and Cohen et al., *Meth. Enzymol.* 297, 192-208, 1998.

[00555] Transforming DNA was introduced into *S. dimorphus* via particle bombardment according to the method described in **EXAMPLE 1** with DNA carried on 550 nm gold particles @ 500psi and a shooting distance of 4cm. Transformants were selected by growth on TAP-CAM agar medium under constant light 50-100uE @ RT for approximately 2 weeks. Transformants were streaked onto TAP-CAM agar medium to ensure single colony isolation and grown for 4 days under constant light.

[00556] To determine if functional fusicoccadiene synthase is produced by transformed algal cells, cultures (2ml) of gene positive, homoplasmic algae were collected by centrifugation, resuspended in 250µl of methanol, and 500µl of saturated NaCl in water and 500µl of petroleum ether were added. The solution was vortexed for three minutes, then centrifuged at 14,000g for five minutes at room temperature to separate the organic and aqueous layers. The organic layer (100µl) was transferred to a vial insert in a standard 2ml sample vial and analyzed using GC/MS. The mass spectrum at 7.617 minutes for the sample from the engineered *S. dimorphus* is obtained. The diagnostic ions at m/z =, 229, 135, and 122 are present in this spectrum, demonstrating the presence of fusicocca-2,10 (14)-diene and indole (**Figure 17** and **Figure 18**).

[00557] These data demonstrate that the chloroplast of *S. dimorphus* can be transformed with foreign DNA containing an expression cassette with a selectable marker and a separate expression cassette with a gene encoding a fusicoccadiene synthase that produces a novel hydrocarbon *in vivo*. One of skill in the art will appreciate that many other methods known in the art may be substituted in lieu of the ones specifically described or referenced.

[00558] Example 7: Production of Phytase in *S. dimorphus*

[00559] In this example a nucleic acid encoding phytase from *E. coli* was introduced into *S. dimorphus*. Transforming DNA is shown graphically in **Figure 19**. In this instance the DNA segment labeled “FD6” is the phytase encoding gene (SEQ ID NO: 26, FD6), the segment labeled “psbD” is the promoter and 5’ UTR for the *psbD* gene from *S. dimorphus*, the segment labeled “D1 3’ ” is the 3’ UTR for the *psbA* gene from *S. dimorphus*, and the segment labeled “CAT” is the chloramphenicol acetyl transferase gene (CAT) from *E. coli*, which is regulated by the promoter and 5’ UTR sequence for the *tufA* gene from *S. dimorphus* and the 3’ UTR sequence for the *rbcL* gene from *S. dimorphus*. The transgene expression cassette and selection marker are targeted to the *S. dimorphus* chloroplast genome via the segments labeled “Homology A” and “Homology B” which are approximately 1000 bp fragments homologous to sequences of DNA adjacent to nucleotide 071,366 (Site 1; nucleotide locations according to the sequence available from NCBI for *S. obliquus*, NC_008101) on the 5’ and 3’ sides, respectively. All DNA segments were cloned into pUC19. All DNA

manipulations carried out in the construction of this transforming DNA were essentially as described by Sambrook et al., *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory Press 1989) and Cohen et al., *Meth. Enzymol.* 297, 192-208, 1998.

[00560] Transforming DNA was introduced into *S. dimorphus* via particle bombardment according to the method described in **EXAMPLE 1** with DNA carried on 550 nm gold particles @ 500psi and a shooting distance of 4cm. Transformants were selected by growth on TAP-CAM agar medium under constant light 50-100uE @ RT for approximately 2 weeks. Transformants were streaked onto TAP-CAM agar medium to ensure single colony isolation and grown for 4 days under constant light.

[00561] Transformants were analyzed by PCR screening (as described in **EXAMPLE 1**) and homoplasmic clones were identified and subcultured for further studies.

[00562] To ensure that the presence of the phytase-encoding gene led to expression of the phytase protein, a Western blot was performed (as described in **EXAMPLE 4**). Results from multiple clones (**Figure 20**) show that expression of the phytase gene in *S. dimorphus* cells resulted in production of the protein.

[00563] One of skill in the art will appreciate that many other methods known in the art may be substituted in lieu of the ones specifically described or referenced.

[00564] Example 8: Use of Erythromycin Esterase as a Selection Marker in *S. dimorphus* and *S. obliquus*

[00565] In this example, a nucleic acid encoding erythromycin esterase gene (EreB) (SEQ ID NO: 29) from *E. coli* was introduced into *S. dimorphus*. Transforming DNA is shown graphically in **Figure 21**. In this instance the DNA segment labeled "EreB ec" is the erythromycin esterase gene (EreB) from *E. coli*, the segment labeled "psbD" is the promoter and 5' UTR sequence for the *psbD* gene from *S. dimorphus*, and the segment labeled "D1 3'" is the 3' UTR for the *psbA* gene from *S. dimorphus*. The selection marker cassette is targeted to the *S. dimorphus* chloroplast genome via the segments labeled "Homology A" and "Homology B" which are approximately 1000 bp fragments homologous to sequences of DNA adjacent to nucleotide 071,366 (Site 1; nucleotide locations according to the sequence available from NCBI for *S. obliquus*, NC_008101) on the 5' and 3' sides, respectively. All segments were cloned into pUC19. All DNA manipulations carried out in the construction of this transforming DNA were essentially as described by Sambrook et al., *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory Press 1989) and Cohen et al., *Meth. Enzymol.* 297, 192-208, 1998.

[00566] Transforming DNA was introduced into *S. dimorphus* via particle bombardment according to the method described in **EXAMPLE 1** with DNA carried on 550 nm gold particles @ 500psi and a shooting distance of 4cm. Transformants were selected by growth on TAP agar medium + 50 µg/mL erythromycin (TAP-ERM50) under constant light 50-100uE @ RT for approximately 2 weeks. Transformants were streaked onto TAP-ERM50 agar medium to ensure single colony isolation and grown for 4 days under constant light.

[00567] Transformants were analyzed by PCR screening (as described in **EXAMPLE 1**). The presence of the EreB selection marker was determined using primers that amplify a 555 bp region within the gene (SEQ ID NO: 7 and SEQ ID NO: 8). **Figure 22** shows that the EreB gene (SEQ ID NO: 29) was amplified from DNA from several transformants but not from wildtype DNA from *S. dimorphus*.

[00568] One of skill in the art will appreciate that many other methods known in the art may be substituted in lieu of the ones specifically described or referenced.

[00569] Example 9: Use of codA as a Selection Marker in *S. dimorphus*

[00570] In this example, a nucleic acid encoding cytosine deaminase gene (codA) from *E. coli* was introduced into *S. dimorphus*. Transforming DNA is shown graphically in Figure 23. In this instance the DNA segment labeled “codA cr” is the codA encoding gene (SEQ ID NO: 31, codA), the segment labeled “psbD” is the promoter and 5' UTR for the *psbD* gene from *S. dimorphus*, the segment labeled “D1 3'” is the 3' UTR for the *psbA* gene from *S. dimorphus*, and the segment labeled “CAT” is the chloramphenicol acetyl transferase gene (CAT) from *E. coli*, which is regulated by the promoter and 5' UTR sequence for the *tufA* gene from *S. dimorphus* and the 3' UTR sequence for the *rbcL* gene from *S. dimorphus*. The transgene expression cassette and selection marker are targeted to the *S. dimorphus* chloroplast genome via the segments labeled “Homology A” and “Homology B” which are approximately 1000 bp fragments homologous to sequences of DNA adjacent to nucleotide 071,366 (Site 1; nucleotide locations according to the sequence available from NCBI for *S. obliquus*, NC_008101) on the 5' and 3' sides, respectively. All DNA segments were cloned into pUC 19. All DNA manipulations carried out in the construction of this transforming DNA were essentially as described by Sambrook et al., *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory Press 1989) and Cohen et al., *Meth. Enzymol.* 297, 192-208, 1998.

[00571] Transforming DNA was introduced into *S. dimorphus* via particle bombardment according to the method described in EXAMPLE 1 with DNA carried on 550 nm gold particles @ 500psi and a shooting distance of 4cm. Transformants were selected by growth on TAP-CAM agar medium under constant light 50-100uE @ RT for approximately 2 weeks. Transformants were streaked onto TAP-CAM agar medium to ensure single colony isolation and grown for 4 days under constant light.

[00572] Transformants were analyzed by PCR screening (as described in EXAMPLE 1) and homoplasmic clones were identified and subcultured for further studies.

[00573] To determine if functional codA is produced by transformed algae cells, cells were grown in TAP media to log phase, pelleted and resuspended to 10^8 cells/mL and plated onto TAP agar medium containing 1mg/mL 5-fluorocytosine (5FC). Wildtype *S. dimorphus*, survives on TAP agar containing 1mg/mL 5FC, while transformants containing the transgene do not (Figure 24). These data demonstrate that the chloroplast of *S. dimorphus* can be transformed with foreign DNA containing an expression cassette with a selectable marker and a separate expression cassette with a gene encoding a cytosine deaminase producing a cell with a negatively selectable phenotype.

[00574] This *S. dimorphus* homoplasmic codA line can now be transformed with either 1) a vector containing a gene of interest cassette without a selection marker in site 1 (the same site that the codA cassette is located within the genome) and after a recovery period on nonselective medium, selected for on medium containing 5FC, or 2) a vector containing a gene of interest cassette linked with an EreB cassette at site 1 and selected on medium containing erythromycin. In this instance, transformants can be streaked onto TAP medium + 50ug/mL erythromycin for single colony isolation and subclones can be patched onto TAP + 1mg/mL 5FC to select for clones homoplasmic for the EreB cassette.

[00575] One of skill in the art will appreciate that many other methods known in the art may be substituted in lieu of the ones specifically described or referenced.

[00576] Example 10: Use of codA as a Selection Marker of *S. obliquus*

[00577] In this example, a nucleic acid encoding cytosine deaminase gene (codA) from *E. coli* was introduced into *S. obliquus*. Transforming DNA is shown graphically in Figure 23. In this instance the DNA segment labeled “codA cr” is the codA encoding gene (SEQ ID NO: 27, codA), the segment labeled “psbD” is the promoter and 5' UTR for the *psbD* gene from *S. dimorphus*, the segment labeled “D1 3'” is the 3' UTR for the *psbA* gene from *S. dimorphus*, and

the segment labeled "CAT" is the chloramphenicol acetyl transferase gene (CAT) from *E. coli*, which is regulated by the promoter and 5' UTR sequence for the *tufA* gene from *S. dimorphus* and the 3' UTR sequence for the *rbcL* gene from *S. dimorphus*. The transgene expression cassette and selection marker are targeted to the *S. obliquus* chloroplast genome via the segments labeled "Homology A" and "Homology B" which are approximately 1000 bp fragments homologous to sequences of DNA adjacent to nucleotide 071,366 (Site 1; nucleotide locations according to the sequence available from NCBI for *S. obliquus*, NC_008101) on the 5' and 3' sides, respectively. All DNA segments were cloned into pUC19. All DNA manipulations carried out in the construction of this transforming DNA were essentially as described by Sambrook et al., *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory Press 1989) and Cohen et al., *Meth. Enzymol.* 297, 192-208, 1998.

[00578] Transforming DNA was introduced into *S. obliquus* via particle bombardment according to the method described in **EXAMPLE 1** with DNA carried on 550 nm gold particles @ 500psi and a shooting distance of 4cm. Transformants were selected by growth on TAP-CAM agar medium under constant light 50-100uE @ RT for approximately 2 weeks. Transformants were streaked onto TAP-CAM agar medium to ensure single colony isolation and grown for 4 days under constant light.

[00579] Transformants were analyzed by PCR screening (as described in **EXAMPLE 1**) and homoplasmic clones were identified and subcultured for further studies.

[00580] To determine if functional *codA* is produced by transformed algae cells, cells were plated onto TAP agar medium containing 1mg/mL 5-fluorocytosine (5FC). Wild type *S. dimorphus* survived on TAP agar containing 5FC, while transformants containing the transgene did not (**Figure 24**).

[00581] These data demonstrate that the chloroplast of *S. obliquus* can be transformed with foreign DNA containing an expression cassette with a selectable marker and a separate expression cassette with a gene encoding a cytosine deaminase producing a cell with a negatively selectable phenotype

[00582] This *S. obliquus* homoplasmic *codA* line can now be transformed with either 1) a vector containing a gene of interest cassette without a selection marker in site 1 (the same site that the *codA* cassette is located within the genome) and after a recovery period on nonselective medium, selected for on medium containing 5FC or 2) a vector containing a gene of interest cassette linked with an *EreB* cassette at site 1 and selected on medium containing erythromycin. In this instance, transformants can be streaked onto TAP medium + 50ug/mL erythromycin for single colony isolation and subclones can be patched onto TAP + 1mg/mL 5FC to select for clones homoplasmic for the *EreB* cassette.

[00583] One of skill in the art will appreciate that many other methods known in the art may be substituted in lieu of the ones specifically described or referenced.

[00584] Example 11: Identification of Functional Promoters for Gene Expression in *S. dimorphus*

[00585] In this example, 8 promoters were amplified from *S. dimorphus* DNA and cloned upstream of the *E. coli* CAT gene. Transforming DNA (p04-151) is shown graphically in **Figure 89**. In this instance the DNA segment labeled "CAT" is the chloramphenicol acetyl transferase gene (CAT) from *E. coli* (SEQ ID NO: 28, CAT), the segment labeled "tufA" is the promoter consisting of ~500bp of the 5' UTR sequence for the *chlB* (SEQ ID NO: 51), *psbB* (SEQ ID NO: 39), *psbA* (SEQ ID NO: 37), *rpoA* (SEQ ID NO: 44), *rbcL* (SEQ ID NO: 49), *cemA* (SEQ ID NO: 45), *ftsH* (SEQ ID NO: 47), *petA* (SEQ ID NO: 53), *petB* (SEQ ID NO: 55) genes from *S. dimorphus*, and the segment labeled "D1 3" is the 3' UTR for the *psbA* gene from *S. dimorphus*. The selection marker cassette is targeted to the *S. dimorphus* chloroplast genome via the segments labeled "Homology A" and "Homology B" which are approximately 1000 bp

fragments homologous to sequences of DNA adjacent to nucleotide 071,366 (Site 1; nucleotide locations according to the sequence available from NCBI for *S. obliquus*, NC_008101) on the 5' and 3' sides, respectively. All DNA segments were subcloned into pUC19. All DNA manipulations carried out in the construction of this transforming DNA were essentially as described by Sambrook et al., *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory Press 1989) and Cohen et al., *Meth. Enzymol.* 297, 192-208, 1998.

[00586] Transforming DNA was introduced into *S. dimorphus* via particle bombardment according to the method described in **EXAMPLE 1** with DNA carried on 550 nm gold particles @ 500psi and a shooting distance of 4cm. Transformants were selected by growth on TAP agar medium + 25 µg/mL chloramphenicol (TAP-CAM) under constant light 50-100uE @ RT for approximately 2 weeks. Each promoter *chlB* (SEQ ID NO: 51), *psbB* (SEQ ID NO: 39), *psbA* (SEQ ID NO: 37), *rpoA* (SEQ ID NO: 44), *rbcL* (SEQ ID NO: 49), *cemA* (SEQ ID NO: 45), *ftsH* (SEQ ID NO: 47), *petA* (SEQ ID NO: 53), *petB* (SEQ ID NO: 55) gave rise to chloramphenicol resistant transformants indicating that these promoter/ 5'UTR fragments were able to drive expression of the CAT gene.

[00587] Example 12: Multiple Gene Expression in *S. dimorphus*

[00588] In this example a nucleic acid encoding FPP synthase from *G. gallus* and a nucleic acid encoding bisabolene synthase from *A. grandis* was introduced into *S. dimorphus*. Transforming DNA is shown graphically in **Figure 25**. In this instance the DNA segment labeled "Is09" is the FPP synthase encoding gene (SEQ ID NO: 23, Is09), the segment labeled "psbD" is the promoter and 5' UTR for the *psbD* gene from *S. dimorphus*, the segment labeled "D1 3' " is the 3' UTR for the *psbA* gene from *S. dimorphus*, the segment labeled "Is11" is the bisabolene synthase encoding gene (SEQ ID NO: 35, Is011), the segment labeled "tufA" is the promoter and 5'UTR for the *tufA* gene from *S. dimorphus*, the segment labeled "rbcL" is the 3'UTR for the *rbcL* gene from *S. dimorphus*, and the segment labeled "CAT" is the chloramphenicol acetyl transferase gene (CAT) from *E. coli*, which is regulated by the promoter and 5' UTR sequence for the *psbD* gene from *S. dimorphus* and the 3' UTR sequence for the *psaB* gene (SEQ ID NO: 59) from *S. dimorphus*. The transgene expression cassette and selection marker are targeted to the *S. dimorphus* chloroplast genome via the segments labeled "Homology A" and "Homology B" which are approximately 1000 bp fragments homologous to sequences of DNA adjacent to nucleotide 071,366 (Site 1; nucleotide locations according to the sequence available from NCBI for *S. obliquus*, NC_008101) on the 5' and 3' sides, respectively. All DNA segments were subcloned into pUC19. All DNA manipulations carried out in the construction of this transforming DNA were essentially as described by Sambrook et al., *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory Press 1989) and Cohen et al., *Meth. Enzymol.* 297, 192-208, 1998.

[00589] Transforming DNA was introduced into *S. dimorphus* via particle bombardment according to the method described in **EXAMPLE 1** with DNA carried on 550 nm gold particles @ 500psi and a shooting distance of 4cm. Transformants were selected by growth on TAP-CAM agar medium under constant light 50-100uE @ RT for approximately 2 weeks. Transformants were streaked onto TAP-CAM agar medium to ensure single colony isolation and grown for 4 days under constant light.

[00590] Transformants were analyzed by PCR screening (as described in **EXAMPLE 1**) and homoplasmic clones were identified and subcultured for further studies.

[00591] To ensure that the presence of the FPP synthase-encoding gene and the bisabolene-encoding gene led to expression of the FPP synthase and bisabolene synthase proteins, a Western blot was performed (as described in **EXAMPLE 4**). Proteins were visualized by a colorimetric assay as per manufacturers instructions (1-step TMB blotting,

Pierce). Results from multiple clones (267 3-9; 267 15-6; and 367 3-4) (Figure 26) show that expression of the FPP synthase gene (Is09) and bisabolene synthase (Is11) in *S. dimorphus* cells resulted in production of both proteins. WT is untransformed *S. dimorphus*. These data demonstrate that the chloroplast of *S. dimorphus* can be transformed with a vector of foreign DNA containing an expression cassette with a selectable marker and two separate expression cassette with a gene encoding an FPP synthase and an E-alpha-bisabolene synthase, and that both proteins are expressed.

[00592] Example 13: Multiple Gene Expression in *S. dimorphus*

[00593] In this example, a nucleic acid encoding endoxylanase from *T. reesei* and chloramphenicol acetyl transferase gene (CAT) from *E. coli* linked together by a ribosome binding sequence from *E. coli* was introduced into *S. dimorphus*. Transforming DNA (BD11-RBS-CAT) is shown graphically in Figure 27. In this instance the DNA segment labeled "BD11" is the endoxylanase encoding gene (SEQ ID NO: 21, BD11), the segment labeled "CAT" is the chloramphenicol acetyl transferase encoding gene (SEQ ID NO: 28, CAT), the segment labeled "RBS1" is the ribosome binding sequence (SEQ ID NO: 60, RBS1), the segment labeled "psbD" is the promoter and 5' UTR for the *psbD* gene from *S. dimorphus*, the segment labeled "psbA" is the 3' UTR for the *psbA* gene from *S. dimorphus*, and the segment labeled "CAT" is the chloramphenicol acetyl transferase gene (CAT) from *E. coli*, which is regulated by the promoter and 5' UTR sequence for the *tufA* gene from *S. dimorphus* and the 3' UTR sequence for the *rbcl* gene from *S. dimorphus*. The transgene expression cassette and selection marker are targeted to the *S. dimorphus* chloroplast genome via the segments labeled "Homology A" and "Homology B" which are approximately 1000 bp fragments homologous to sequences of DNA adjacent to nucleotide 071,366 (Site 1; nucleotide locations according to the sequence available from NCBI for *S. obliquus*, NC_008101) on the 5' and 3' sides, respectively. All DNA segments were subcloned into pUC19. All DNA manipulations carried out in the construction of this transforming DNA were essentially as described by Sambrook et al., Molecular Cloning: A Laboratory Manual (Cold Spring Harbor Laboratory Press 1989) and Cohen et al., *Meth. Enzymol.* 297, 192-208, 1998.

[00594] Transforming DNA was introduced into *S. dimorphus* via particle bombardment according to the method described in EXAMPLE 1 with DNA carried on 550 nm gold particles @ 500psi and a shooting distance of 4cm. Transformants were selected by growth on TAP agar medium + 25 µg/mL chloramphenicol. Transformants were streaked onto TAP-CAM agar medium to ensure single colony isolation and grown for 4 days under constant light.

[00595] Transformants were analyzed by PCR screening (as described in EXAMPLE 1) and homoplasmic clones were identified and subcultured for further studies.

[00596] To ensure that the presence of the endoxylanase-encoding gene led to expression of the endoxylanase protein, a Western blot was performed (as described in EXAMPLE 4). Results from multiple clones (Figure 28) show that expression of the endoxylanase gene in *S. dimorphus* cells resulted in production of the protein of expected molecular weight and not of an endoxylanase-CAT fusion protein.

[00597] To determine if the endoxylanase produced by transformed algae cells was functional, endoxylanase activity was tested using an enzyme function assay (as described in EXAMPLE 4). Figure 29 shows that endoxylanase activity is detected in clarified lysates of *S. dimorphus* engineered with the endoxylanase-RBS-CAT construct (operon 1_1, 2_1, 2_2, 2_3) and not in lysates of wt.

[00598] To determine whether both enzymes are produced from the same transcript, RNA was isolated from wildtype and engineered algae cells using the Concert Plant RNA Reagent kit (Invitrogen). RNA was DNase treated and cleaned using the RNeasy clean up kit (Qiagen). cDNA was synthesized from each of RNA using the iScript kit (Biorad) and –

reverse transcriptase (-RT) controls were included. cDNA (and -RT controls) was used as template in PCR with primers that hybridize to the endoxylanase gene and the CAT gene (**Figure 30A**) (SEQ ID NO: 11 and SEQ ID NO: 12, respectively) and amplify a product of 1.3kb. **Figure 30B** shows that a product of appropriate size was amplified from cDNA templates from 4 of the 5 transformants indicating that in these lines, the endoxylanase and the CAT gene are transcribed on a single transcript.

[00599] To further investigate variants of RBS1 (e.g., RBS2) and to understand the strength of these RBS sequences to recruit ribosomes, a nucleic acid encoding chloramphenicol acetyl transferase gene (CAT) from *E. coli* and endoxylanase from *T. reesei* linked together by two distinct ribosome binding sequences from *E. coli* were introduced into *S. dimorphus*. Transforming DNA (p04-231 or p04-232) is shown graphically in **Figure 31**. In this instance the DNA segment labeled "CAT" is the chloramphenicol acetyl transferase encoding gene (SEQ ID NO: 28, CAT), the segment labeled "BD11" is the endoxylanase encoding gene (SEQ ID NO: 21, BD11), the segment labeled "RBS1" is the ribosome binding sequence (SEQ ID NO: 60, RBS1), the segment labeled "RBS2" is the ribosome binding sequence (SEQ ID NO: 61, RBS2) the segment labeled "psbD" is the promoter and 5' UTR for the *psbD* gene from *S. dimorphus*, the segment labeled "D1 3' " is the 3' UTR for the *psbA* gene from *S. dimorphus*. The transgene expression cassette and selection marker are targeted to the *S. dimorphus* chloroplast genome via the segments labeled "Homology A" and "Homology B" which are approximately 1000 bp fragments homologous to sequences of DNA adjacent to nucleotide 071,366 (Site 1; nucleotide locations according to the sequence available from NCBI for *S. obliquus*, NC_008101) on the 5' and 3' sides, respectively. All DNA segments were subcloned into pUC19. All DNA manipulations carried out in the construction of this transforming DNA were essentially as described by Sambrook et al., *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory Press 1989) and Cohen et al., *Meth. Enzymol.* 297, 192-208, 1998.

[00600] Transforming DNA was introduced into *S. dimorphus* via particle bombardment according to the method described in **EXAMPLE 1** with DNA carried on 550 nm gold particles @ 500psi and a shooting distance of 4cm. Transformants were selected by growth on TAP agar medium + 25 µg/mL chloramphenicol. Transformants were streaked onto TAP-CAM agar medium to ensure single colony isolation and grown for 4 days under constant light.

[00601] Transformants were analyzed by PCR screening (as described in **EXAMPLE 1**) and homoplasmic clones were identified and subcultured for further studies.

[00602] To determine if the endoxylanase produced by transformed algae cells was functional, endoxylanase activity was tested using an enzyme function assay (as described in **EXAMPLE 4**). **Figure 32A** shows that RBS1 between the two genes produces xylanase activity, however RBS2 does not produce active xylanase (**Figure 32B**).

[00603] These data demonstrate that the chloroplast of *S. dimorphus* can be transformed with a vector of foreign DNA containing an expression cassette that consists of a gene of interest linked to a selectable marker by a nucleotide sequence, allowing for the expression of multiple genes from one transcript, in this case a gene encoding an endoxylanase and a gene encoding chloramphenicol acetyl transferase. One of skill in the art will appreciate that many other methods known in the art may be substituted in lieu of the ones specifically described or referenced.

[00604] Example 14: Use of Conserved Gene Cluster for an Integration Site in *S. dimorphus*

[00605] In this example, a nucleic acid encoding chloramphenicol acetyl transferase gene from *E. coli* was introduced into *S. dimorphus*. Transforming DNA is shown graphically in **Figure 33**. In this instance the DNA segment labeled "CAT" is the chloramphenicol acetyl transferase gene from *E. coli*, the segment labeled "tufA" is the promoter and 5'

UTR sequence for the *tufA* gene from *S. dimorphus*, and the segment labeled "rbcL" is the 3' UTR for the *rbcL* gene from *S. dimorphus*. The selection marker cassette is targeted to the *S. dimorphus* chloroplast genome via the segments labeled "Homology A1" and "Homology B1" which are approximately 1000 bp fragments homologous to sequences of DNA in the psbB-psbT-psbN-psbH cluster wherein the CAT cassette is inserted between psbT and psbN. All DNA segments were subcloned into pUC19. All DNA manipulations carried out in the construction of this transforming DNA were essentially as described by Sambrook et al., *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory Press 1989) and Cohen et al., *Meth. Enzymol.* 297, 192-208, 1998.

[00606] Transforming DNA was introduced into *S. dimorphus* via particle bombardment according to the method described in **EXAMPLE 1** with DNA carried on 550 nm gold particles @ 500psi and a shooting distance of 4cm. Transformants were selected by growth on TAP agar medium + 25 µg/mL chloramphenicol (TAP-CAM) under constant light 50-100uE @ RT for approximately 2 weeks. Transformants were streaked onto TAP-CAM agar medium to ensure single colony isolation and grown for 4 days under constant light.

[00607] Cells from the transformants were analyzed by PCR screening (as described in **EXAMPLE 1**). The degree to which the transforming DNA was integrated into the chloroplast genome was determined using primers that amplify a 250 bp constant region (SEQ ID NO: 3 and SEQ ID NO: 4) and a 400 bp region spanning the integration site (SEQ ID NO: 15 and SEQ ID NO: 16). The homology regions target the integration site, the region of the chloroplast genome between psbT and psbN, approximately nucleotide 059,687 (nucleotide locations according to the sequence available from NCBI for *S. obliquus*, NC_008101). **Figure 34** shows that subclones from clone 52 are homoplasmic, i.e., only the constant region (250bp product) was amplified, while in the control reactions (WT) both the constant region and the region spanning the integration site (400bp) were amplified. Clone 6 is another parental clone, however subclones from clone 6 are not completely homoplasmic as the spanning region is still amplified. These data indicate that the psbB-psbH cluster can be utilized as an integration site in engineering *S. dimorphus*.

[00608] One of skill in the art will appreciate that many other methods known in the art may be substituted in lieu of the ones specifically described or referenced.

[00609] Example 15: Strategy to Generate Markerless Transgenic *S. dimorphus*

[00610] In this example, the transgenic line generated in **EXAMPLE 12**, was used to inoculate nonselective media. A saturated culture was diluted 1:300 in nonselective media, allowed to grow to saturation and diluted 1:4 in nonselective media. Once saturated, the culture was plated onto nonselective TAP medium to ensure single colony formation. Single clones were then patched to 1) nonselective TAP medium and 2) TAP-CAM medium. Clones that failed to grow on TAP-CAM were further analyzed by PCR.

[00611] **Figure 35 A** is a graphical representation of the transforming DNA (top) and loopout product (bottom) that results from recombination at the identical D2 (psbD) promoter segments. HR-A & HR-B represent the homology regions. D1 3', psaB 3' and rbcL represent the psbA 3'UTR, psaB 3'UTR, and rbcL 3'UTR, respectively. D2 and *tufA* is the psbD and *tufA* promoter, respectively. Is09 is FPP synthase and Is011 is bisabolene synthase.

[00612] To confirm the absence of the CAT gene, two methods were employed. First, PCR was performed using primers that amplify a 2.5kb +CAT fragment and/or a 700bp -CAT fragment (SEQ ID NO: 9 and SEQ ID NO: 10). **Figure 35 B** is an agarose gel showing that in subclones of the #74 transformant only the 700bp -CAT product was amplified while in the plasmid DNA control, the 2.5kb +CAT fragment was amplified. The presence of the 700bp

product in the plasmid DNA control is likely the result of recombination in the *E. coli* host as it is *RecB*⁺. Primers 7117 & 7119 (SEQ ID NO: 9 and SEQ ID NO: 10) were used to amplify the products. The “markerless” transgenic *S. dimorphus* shows amplification of 700bp -CAT loopout fragment and failure to amplify the 2.5kb +CAT fragment in subclones of clone #74.

[00613] Second, PCR was performed using primers that amplify the 660bp CAT gene (SEQ ID NO: 18 and SEQ ID NO: 19), and either primers that amplify a 1.3kb constant region of the *psbA* gene (SEQ ID NO: 13 and SEQ ID NO: 14) or those that amplify a 400bp constant region of the *psbA* gene (SEQ ID NO: 1 and SEQ ID NO: 2). **Figure 36** shows that only the constant fragment was amplified in the #74 markerless line, while the CAT gene was amplified in the parental line that was always kept on CAT selection. **Panel A** shows multiplex PCR using primers that amplify a 660bp CAT fragment and primers that amplify a 1.3kb constant region of the endogenous *psbA* gene. Only the 1.3kb constant region is amplified in the #74 markerless potential. **Panel B** shows multiplex PCR using primers that amplify a 660bp CAT fragment and primers that amplify a 400bp constant region of the endogenous *psbA* gene. Only the 400bp constant region is amplified in the #74 markerless potential. The PCR reactions in both panel A and panel B had a 50-60 degree Celcius annealing gradient to rule out the possibility that the annealing of the primers was temperature sensitive.

[00614] These data demonstrate that *S. dimorphus* clones can be obtained consisting of a genetically engineered chloroplast and without an antibiotic resistance marker.

[00615] One of skill in the art will appreciate that many other methods known in the art may be substituted in lieu of the ones specifically described or referenced.

[00616] Example 16: Use of Betaine Aldehyde Dehydrogenase to Confer Salt Tolerance and/or as a Negative Selection Mechanism

[00617] In this example, a nucleic acid sequence encoding betaine aldehyde dehydrogenase from spinach or sugar beet was engineered into *S. dimorphus* (as described in **EXAMPLE 4**). Transforming DNA is shown graphically in **Figure 37**. In this instance the DNA segment labeled “BAD1 or BAD4” is the betaine aldehyde dehydrogenase encoding gene from spinach (BAD1) or sugar beet (BAD4) (SEQ ID NO: 32, BAD1 or SEQ ID NO: 34, BAD4), the segment labeled “*psbD*” is the promoter and 5’ UTR for the *psbD* gene from *S. dimorphus*, the segment labeled “*rbcL*” is the 3’ UTR for the *psbA* gene from *S. dimorphus*, and the segment labeled “CAT” is the chloramphenicol acetyl transferase gene from *E. coli*, which is regulated by the promoter and 5’ UTR sequence for the *tufA* gene from *S. dimorphus* and the 3’ UTR sequence for the *rbcL* gene from *S. dimorphus*. The transgene expression cassette and selection marker are targeted to the *S. dimorphus* chloroplast genome via the segments labeled “Homology A” and “Homology B” which are approximately 1000 bp fragments homologous to sequences of DNA adjacent to nucleotide 671,366 (Site 1; nucleotide locations according to the sequence available from NCBI for *S. obliquus*, NC_008101) on the 5’ and 3’ sides, respectively. All DNA segments are subcloned into pUC19. All DNA manipulations carried out in the construction of this transforming DNA were essentially as described by Sambrook et al., *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory Press 1989) and Cohen et al., *Meth. Enzymol.* 297, 192-208, 1998.

[00618] Transforming DNA was introduced into *S. dimorphus* via particle bombardment according to the method described in **EXAMPLE 1** with DNA carried on 550 nm gold particles @ 500psi and a shooting distance of 4cm. Transformants were selected by growth on TAP-CAM agar medium under constant light 50-100uE @ RT for approximately 2 weeks. Transformants were streaked onto TAP-CAM agar medium to ensure single colony isolation and grown for 4 days under constant light.

[00619] Transformants were analyzed by PCR screening (as described in EXAMPLE 1) and homoplasmic clones were identified and subcultured for further studies.

[00620] To ensure that the presence of the betaine aldehyde dehydrogenase encoding gene led to expression of the protein, a Western blot was performed (as described in EXAMPLE 4). In this instance, the BAD genes were tagged with an HA tag and the primary antibody was an anti-HA HRP conjugated antibody (clone 3F10, Roche) in which a 1:10,000 dilution of a 50U/mL stock was used as the antibody solution. Results from multiple clones (Figure 38) show that expression of the BAD gene from spinach and from sugar beet gene in *S. dimorphus* cells resulted in production of the protein.

[00621] To determine if this protein confers salt tolerance or causes the cells to become sensitive to betaine aldehyde (and therefore allows this strain to be used in negative selection experiments as proposed in examples 9 and 10), cells expressing the BAD genes can be grown side-by-side with wildtype cells and the media supplemented with increasing concentrations of salt and/or betaine aldehyde.

[00622] One of skill in the art will appreciate that many other methods known in the art may be substituted in lieu of the ones specifically described or referenced.

[00623] Example 17: Development of a Transformation System for *D. tertiolecta*

[00624] In this example, a method for transformation of *D. tertiolecta* is described. Algae cells are grown to log phase (approximately 5.0×10^5 cells/mL) in G32 medium (32g/L NaCl, 0.0476mM CaCl₂, 0.162mM H₃BO₃, 0.406mM Mg₂SO₄, 0.00021mM NaVO₃, 5g/L bicarbonate, 12.9mL/L each of F/2 A and B algae food (Aquatic Eco-systems, Inc.)) at 23°C under constant illumination of 50-100uE on a rotary shaker set at 100 rpm. Cells are harvested at 1000 x g for 5 min. The supernatant is decanted and cells are resuspended in G32 media at 10^8 cells/mL. 5×10^7 cells are spread on selective agar medium and transformed by particle bombardment with 550 nm diameter gold particles carrying the transforming DNA @ 300-400psi with the Helios Gene Gun (Bio-Rad) from a shot distance of 4cm. Desired algae clones are those that grow on selective media.

[00625] PCR is used to identify transformed algae strains. For PCR analysis, colony lysates are prepared by suspending algae cells (from agar plate or liquid culture) in lysis buffer (0.5% SDS, 100 mM NaCl, 10 mM EDTA, 75 mM Tris-HCl, pH 7.5) and heating to 98°C for 10 minutes, followed by cooling to near 23°C. Lysates are diluted 50-fold in 100mM Tris-HCl pH 7.5 and 2uL is used as template in a 25uL reaction. Alternatively, total genomic DNA preparations may be substituted for colony lysates. A PCR cocktail consisting of reaction buffer, dNTPs, PCR primer pair(s) (indicated in each example below), DNA polymerase, and water is prepared. Algae DNA is added to provide template for the reaction. Annealing temperature gradients are employed to determine optimal annealing temperature for specific primer pairs. In many cases, algae transformants are analyzed by PCR with primers that are specific for the transgene being introduced into the chloroplast genome. Desired algae transformants are those that give rise to PCR product(s) of expected size(s).

[00626] One of skill in the art will appreciate that many other transformation methods known in the art may be substituted in lieu of the ones specifically described or referenced herein.

[00627] Example 18: Use of Conserved Gene Cluster for an Integration site in *D. tertiolecta*

[00628] In this example, a nucleic acid encoding erythromycin esterase gene (EreB) (SEQ ID NO: 29) from *E. coli* was introduced into *D. tertiolecta*. Transforming DNA is shown graphically in Figure 39. In this instance the DNA segment labeled "EreB ec" is the erythromycin esterase gene (EreB) from *E. coli*, the segment labeled "psbDp" is the

promoter and 5' UTR sequence for the *psbD* or *tufA* gene from a *D. tertiolecta* (SEQ ID NO: 62, *psbD2*, SEQ ID NO: 63, *tufA2*), and the segment labeled "rbcL 3'" is the 3' UTR for the *rbcL* gene from *D. tertiolecta* (SEQ ID NO: 64, *2rbcL 3'*). The selection marker cassette is targeted to the *D. tertiolecta* chloroplast genome via the segments labeled "HA" and "HB" which are approximately 1000 bp fragments homologous to sequences of DNA in the *psbB-psbT-psbN-psbH* cluster (SEQ ID NO: 133) wherein the *EreB* cassette is inserted between *psbT* and *psbN* at approximately nucleotide 2383 of SEQ ID NO: 133. All DNA segments were subcloned into pUC19. All DNA manipulations carried out in the construction of this transforming DNA were essentially as described by Sambrook et al., *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory Press 1989) and Cohen et al., *Meth. Enzymol.* 297, 192-208, 1998.

[00629] Transforming DNA was introduced into *D. tertiolecta* via particle bombardment according to the method described in **EXAMPLE 17** with DNA carried on 550 nm gold particles @ 300psi and a shooting distance of 4cm. Transformants were selected by growth on G32 agar medium + 75 µg/mL erythromycin (G32-Erm) under constant light 50-100uE @ RT for approximately 4 weeks. Transformants were inoculated into nonselective G32 media and grown for ~1week under constant light (50-100uE).

[00630] Cells from the transformants were analyzed by PCR screening (as described in **EXAMPLE 17**). The presence of the *EreB* selection marker was determined using primers that amplify a 555 bp region within the gene (SEQ ID NO: 7 and SEQ ID NO: 8). **Figure 40** shows that the *EreB* gene was amplified from DNA from transformants 4, 5, and 6 but not from wildtype DNA from *D. tertiolecta*.

[00631] These data demonstrate that the chloroplast of *D. tertiolecta* can be transformed with foreign DNA containing an expression cassette with a selectable marker. One of skill in the art will appreciate that many other methods known in the art may be substituted in lieu of the ones specifically described or referenced.

[00632] Example 19: Production of Endoxylanase in *D. tertiolecta*

[00633] In this example a nucleic acid encoding endoxylanase from *T. reesei* was introduced into *D. tertiolecta*. Transforming DNA is shown graphically in **Figure 41**. In this instance the DNA segment labeled "BD11" is the endoxylanase encoding gene (SEQ ID NO: 21, BD11), the segment labeled "psbD" is the promoter and 5' UTR for the *psbD* gene from *D. tertiolecta*, the segment labeled "D1 3'" is the 3' UTR for the *psbA* gene from *D. viridis* (SEQ ID NO: 65, 3 *psbA 3'*), and the segment labeled "EreB ec" is the erythromycin esterase gene from *E. coli*, which is regulated by the promoter and 5' UTR sequence for the *tufA* gene from *D. tertiolecta* and the 3' UTR sequence for the *rbcL* gene from *D. tertiolecta*. The transgene expression cassette and selection marker are targeted to the *D. tertiolecta* chloroplast genome via the segments labeled "HA" and "HB" which are approximately 1000 bp fragments homologous to sequences of DNA in the *psbB-psbT-psbN-psbH* cluster wherein the transgene cassette is inserted between *psbT* and *psbN*. All DNA segments were subcloned into pUC19. All DNA manipulations carried out in the construction of this transforming DNA were essentially as described by Sambrook et al., *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory Press 1989) and Cohen et al., *Meth. Enzymol.* 297, 192-208, 1998.

[00634] Transforming DNA was introduced into *D. tertiolecta* via particle bombardment according to the method described in **EXAMPLE 17** with DNA carried on 550 nm gold particles @ 400 and a shooting distance of 4cm. Transformants were selected by growth on G32 agar medium + 75 µg/mL erythromycin (G32-Erm) under constant light 50-100uE @ RT for approximately 4 weeks. Transformants were inoculated into G32 media + 100ug/mL erythromycin and grown for ~1week under constant light (50-100uE).

[00635] Cells from the transformants were analyzed by PCR screening (as described in **EXAMPLE 17**). The presence of the *EreB* selection marker was determined using primers that amplify a 555bp region within the gene (SEQ ID NO: 7 and SEQ ID NO: 8). **Figure 42** shows that the *EreB* gene was amplified from DNA from transformant 12-3 but not from wildtype DNA from *D. tertiolecta*.

[00636] To ensure that the presence of the endoxylanase-encoding gene led to expression of the endoxylanase protein, a Western blot was performed (as described in **EXAMPLE 4**). Results from transformant 12-3 (**Figure 43**) show that expression of the endoxylanase gene in *D. tertiolecta* cells resulted in production of the protein.

[00637] To determine if the endoxylanase produced by transformed algae cells was functional, endoxylanase activity was tested using an enzyme function assay (as described in **EXAMPLE 4**). **Figure 44** shows that endoxylanase activity is detected in the 12-3 transformant but not in wildtype cells.

[00638] These data demonstrate that the chloroplast of *D. tertiolecta* can be transformed with foreign DNA containing an expression cassette with a selectable marker and a separate expression cassette with a gene encoding an endoxylanase, and that the proteins expressed were functional. One of skill in the art will appreciate that many other methods known in the art may be substituted in lieu of the ones specifically described or referenced.

[00639] Example 20: Overview of Genetic Engineering

[00640] To engineer the chloroplast of an algae three things are required: a cassette expressing a selectable marker; a delivery method to deliver the plasmid DNA into the chloroplast; and a vector containing regions of DNA homologous to the chloroplast genome to be used in targeted homologous recombination (a homologous integration vector or homologous recombination vector).

[00641] In strains of algae that have little or no known chloroplast sequence information available, the identification of homologous regions of DNA and the construction of a vector containing those regions, are significant and time consuming tasks. Current methods for obtaining unknown sequence information, such as Inverse PCR (PCR Cloning Protocols Series: Methods in Molecular Biology, Volume: 192, Pub. Date: Apr-01-2002, Page Range: 301-307, DOI: 10.1385/1-59259-177-9:301) and Adaptor Ligated PCR (*Nature Protocols* 2, pp. 2910 - 2917 (2007) Published online: 8 November 2007) are time consuming in that they take multiple iterations in order to generate a DNA sequence that is long enough to be used in a homologous integration vector.

[00642] A method that allows for the quick identification of a large piece of chloroplast DNA sequence, sufficient in size to build a homologous integration vector, would be very useful in the engineering of algal genomes. The methods described herein, can be applied to all strains of an algae, for example, a green algae, for which there is little or no known DNA sequence information available. The methods described herein, can also be applied to an algae, for which there is incomplete sequence information available.

[00643] Example 21: Use of a Conserved Gene Cluster to Generate Sequence Information

[00644] Across the chloroplast genomes sequenced to date, there are only a few clusters of genes that are consistently found adjacent to each other. Two examples of such gene clusters are *ycf3-ycf4* and *psbF-psbL*. However, these two clusters are too small in size to yield enough DNA sequence information to be useful for homologous recombination.

[00645] Another gene cluster, *psbB-psbT-psbN-psbH*, is found together in the same orientation in most algae and plants. Knowledge of the presence of this gene cluster allows one to amplify a large region of chloroplast DNA that provides enough DNA sequence information to construct a vector for homologous recombination. This vector can then be used to modify the chloroplast genome of algal strains and plants that have not yet been genetically engineered.

[00646] The gene cluster psbB-psbT-psbN-psbH is a region of chloroplast DNA that is highly conserved amongst algae and plants. However, this cluster may not be conserved at the nucleic acid level or in the spacing between the genes (the intergenic regions). In addition, the nucleic acid contents of the intergenic regions may vary. While at the nucleotide level there may be significant diversity, at the protein level this region is quite conserved. **Figure 88** is an alignment of 4 algae that have had their chloroplast genomes sequenced: *C. reinhardtii* (NCBI NC_005353), *C. vulgaris* (NCBI NC_001865), *S. obliquus* (NCBI NC_008101), and *P. purpurea* (NCBI NC_000925). This figure shows the high degree of conservation in terms of gene placement and orientation.

[00647] Although the gene cluster, psbB-psbT-psbN-psbH, may not be conserved at the nucleic acid level, the proteins on the terminal ends of this region, psbB and psbH, are highly conserved at the amino acid level and contain regions of high conservation at the nucleotide level. **Figure 86** is an alignment of the *psbB* gene from four algae that have had their chloroplast genomes sequenced: *C. reinhardtii*, *C. vulgaris*, *S. obliquus*, and *P. purpurea*, and **Figure 87** is an alignment of the *psbH* gene from the same algae. Both figures show regions of high nucleic acid homology. This allows for the design of degenerate primers that will anneal to regions within the nucleic acid sequences encoding for the proteins psbB and psbH, resulting in the amplification of the whole gene cluster in one step. This double stranded product can then be quickly sequenced directly from both ends, and enough sequence information can then be generated to construct a homologous recombination vector. The time it takes to generate the sequence data is much less than with other methods.

[00648] Two degenerate primers (forward primers 4099 and 4100) specific to the *psbB* gene (reverse primers SEQ ID NO: 129 and SEQ ID NO: 130) and 2 two degenerate primers (4101 and 4102) specific to the *psbH* gene (SEQ ID NO: 131 and SEQ ID NO: 132) were designed from the conserved nucleotide regions of the *psbB* and *psbH*. These primers have been used to generate the sequence of the psbB-psbT-psbN-psbH gene cluster from different species of algae that have little or no sequence information available in public databases including *D. tertiolecta* (SEQ ID NO: 133), an alga from the genus *Dunaliella* of unknown species (SEQ ID NO: 134), *N. abudans* (SEQ ID NO: 135), an isolate of *C. vulgaris* differing from the published genome (SEQ ID NO: 136), and *T. suecica* (SEQ ID NO: 137). **Figures 74 and 75** show the degenerate primers amplifying a large fragment from the *Dunaliella* isolate and *N. abudans*, respectively. **Figure 73** shows the amplification from *S. dimorphus*. In each of these figures the center lane is occupied by a 1kb Plus ladder (Invitrogen). In each figure four different combinations of primers were used. The top left panel shows amplification with primers 4099 and 4101. The bottom left panel shows amplification with primers 4099 and 4102. The top right panel shows amplification with primers 4100 and 4101. The bottom right panel shows amplification with primers 4100 and 4102. After amplification the desired fragments are gel purified using the Qiaquick Gel Extraction Kit (Qiagen) and sequenced. In each of **Figures 73 to 75**, Product 1 represents the full length psbB-psbH gene cluster.

[00649] An integration vector built from this region has been shown to transform *Dunaliella tertiolecta* (see **EXAMPLE 18**).

[00650] Example 22: Additional Vectors constructed for *Scenedesmus dimorphus*

[00651] Additional vectors were constructed for *Scenedesmus dimorphus* since the sequence of a closely related species *Scenedesmus obliquus* is publicly available (NCBI for *S. obliquus*, NC_008101). These vectors were made to test integration sites and homoplasmy along the entire region of psbB-psbT-psbN-psbH, as well as the next adjacent protein in *S. dimorphus*, psbK. This set of vectors targeted integration into the intergenic region between psbT and psbN, psbN and psbH, psbH and psbK, and the region 3' of psbK (p04-128, p04-129, p04-130, and p04-131).

respectively) (Figures 76 to 79 respectively). p04-128 targets integration at approximately nucleotide 059,587. p04-129 targets integration at approximately nucleotide 059,999. p04-130 targets integration at approximately nucleotide 060,429. p04-131 targets integration at approximately nucleotide 060,961 (nucleotide locations according to the sequence available from NCBI for *S. obliquus*, NC_008101).

[00652] All vectors have an expression cassette consisting of a chloramphenicol (CAT) selectable marker, an endogenous promoter, and an endogenous terminator cloned between the Homology A and Homology B fragments. p04-128 had *tufA*-CAT-*rbcL* cloned between the Homology A and Homology B fragments (p04-142) (Figure 80). p04-129 had *tufA*-CAT-*rbcL* cloned between the Homology A and Homology B fragments (p04-143) (Figure 81). p04-130 had *tufA*-CAT-*rbcL* cloned between the Homology A and Homology B fragments (p04-144) (Figure 82). p04-131 had *tufA*-CAT-*rbcL* cloned between the Homology A and Homology B fragments (p04-145) (Figure 83). Vectors are shown graphically in their corresponding figures. In this instance the DNA segment labeled "CAT" is the chloramphenicol acetyl transferase gene from *E. coli*, the segment labeled "tufA" is the promoter and 5' UTR sequence for the *tufA* gene from *S. dimorphus*, and the segment labeled "rbcL" is the 3' UTR for the *rbcL* gene from *S. dimorphus*.

[00653] Transforming DNA was introduced into *S. dimorphus* via particle bombardment according to the method described in EXAMPLE 1 with DNA carried on 550 nm gold particles @ 500psi and a shooting distance of 4cm. Transformants were selected by growth on TAP-CAM agar medium under constant light 50-100uE @ RT for approximately 2 weeks. Transformants were streaked onto TAP-CAM agar medium to ensure single colony isolation and grown for 4 days under constant light.

[00654] To test for integration of the CAT gene in between the *psbT* and *psbN* genes (p04-142), clones were screened for homoplasmy using primers 3160 and 3162 (that amplify a 200bp constant band from the genome), and primers 4682 and 4982 (that amplify a 400 bp band that spanning the integration site).

[00655] To test for integration of the CAT gene in between the *psbN* and *psbH* genes (p04-143), clones were screened for homoplasmy using primers 2922 and 2923 that amplify a 400bp constant band from the genome, and primers 4684 and 4685, that amplify a 200 bp band that spans the integration site.

[00656] To test for integration of the CAT gene in between the *psbH* and *psbK* genes (p04-144), clones were screened for homoplasmy using primers 2922 and 2923 that amplify a 400bp constant band from the genome, and primers 4686 and 4687, that amplify a 300 bp band that spans the integration site.

[00657] To test for integration of the CAT gene 3' of the *psbK* gene (p04-145), clones were screened for homoplasmy using primers 3160 and 3162 that amplify a 200bp constant band from the genome, and primers 4688 and 4689 amplify a 300 bp band that spans the integration site.

[00658] Primers used for each of the PCR screens are listed in Table 5.

[00659] Table 5

p04-142	3160 (SEQ ID NO: 3) GAACTACAACTAATTATTTTC
	3162 (SEQ ID NO: 4) TGAAACCAGTCTTTGTAAAGCTCA
	4682 (SEQ ID NO: 15) CCACCTCGTATGGTAAAATAATTG
	4982 (SEQ ID NO: 16) GAAAGAATTATGGACAGTCCTGCT
p04-143	2922 (SEQ ID NO: 1) AGAAGGAGCTTCTACAGATGC

	2923 (SEQ ID NO: 2) TCATTAGTTACTTCATCTTTAATCCG
	4684 (SEQ ID NO: 140) GAAGGAGGTCCAAAAC TCACA
	4685 (SEQ ID NO: 141) CCTGGTTCTTGAAGTGCATC
p04-144	2922 (see above)
	2923 (see above)
	4686 (SEQ ID NO: 142) TGAGTTGGGAAACTTTAGCTTCTT
	4787 (SEQ ID NO: 143) AAAAGATTGCCAAGACCAAA
p04-145	3160 (see above)
	3162 (see above)
	4688 (SEQ ID NO: 144) AAAAAGAATGAAATTTTATGTTCG
	4689 (SEQ ID NO: 145) ATGGATGTCGTCCTCCAAA

[00660] Figure 84 and Figure 85 shows that homoplasmic clones are recovered from integration between psbT and psbN (p04-142) and integration 3' of psbK (p04-145).

[00661] Example 23: Creation of a Yeast-bacteria Shuttle Vector

[00662] Heterologous (exogenous) gene introduction into the chloroplast by homologous recombination is efficient when a selectable marker and the gene of interest is flanked by 5' and 3' homology to a locus that can tolerate integration. To integrate more than one gene, one can target a separate locus and use a second selectable marker. Integration of two or more genes is problematic from a time and labor standpoint. In addition, availability of selectable markers becomes an issue. To contend with these issues, a yeast-based system was created wherein, in a single step, several exogenous genes can be assembled along with an algal selectable marker, and placed into a yeast-bacteria shuttle vector. Two versions of this vector were created. One version contains a 5.2 kb region from the *Scenedesmus obliquus* chloroplast (*Scenedesmus* chloroplast sequence NCBI reference sequence: NC_008101, 057,611-062,850bp) (SEQ ID NO: 125). This 5.2 kb region is highly conserved (at the amino acid level) amongst algae species, and spans a region comprising psbB to rbcL genes. The second version of this vector contains two 1,000 bp "homology A3" (070,433-071,342bp) (SEQ ID NO: 126) and "homology B3" (071,379-072,254bp) (SEQ ID NO: 127) regions which target a locus immediately downstream of the psbA gene. The two shuttle vectors (Figures 49 and 58) comprise the above-mentioned sequences from the chloroplast genome of *Scenedesmus obliquus*, bacterial replication/selection elements, and yeast replication, segregation, and selection/counter-selection elements.

[00663] There are at least four advantages of the yeast-based system over the existing technology: 1) each of the 1, 2, 3, 4, or more gene expression cassettes can be amplified with primers containing 5' and 3' homology to adjacent cassettes, thereby alleviating the requirement to clone flanking homology into the gene cassette design; 2) several gene cassettes (for example, 2, 3, 4, 5, 6, or 20) can be assembled together as a contig in a single step and require a single selectable marker for chloroplast introduction; 3) this technology can be applied to other algal species due to the conserved nature of the psbB-rbcL locus across algae species; and 4) the 5.2 kb of homology contained within the shuttle vector (Figure 58) and the 2 kb of homology as shown in Figure 49, ensures that homologous recombination is accurate and efficient within the chloroplast.

[00664] It should be noted that, for example, more than 2, more than 5, more than 10, more than 15, more than 20, or more than 25 gene cassettes can be assembled in the shuttle vector.

[00665] Example 24: Plasmid Construction

[00666] A derivative of plasmid vector pUC19 (New England Biolabs, U.S.A.; Yanisch-Perron, C., *et al.* (1985) *Gene*, 33, 103-119) lacking a multiple cloning site (herein referred to as gutless pUC) (**Figure 45**) was used to create the backbones for three gene expression cassettes. Three different gene expression cassettes comprising the promoter-terminator pairs: *petA-chlL*, *D2-D1*, and *tufA-psaB*, respectively were cloned into gutless pUC (**Figures 50, 51 and 52**).

[00667] To insert the genes of interest ("GOF") (CC90, SEQ ID NO: 115; CC91, SEQ ID NO: 116; CC92, SEQ ID NO: 117; CC93, SEQ ID NO: 109; CC94, SEQ ID NO: 110; CC97, SEQ ID NO: 113; IS57, SEQ ID NO: 121; IS61, SEQ ID NO: 124; IS62, SEQ ID NO: 123; IS116, SEQ ID NO: 122; BD11, SEQ ID NO: 146; and IS99, SEQ ID NO: 147), each of the three vectors (Gene Vector 1 (**Figure 50**), Gene Vector 2 (**Figure 51**), and Gene Vector 3 (**Figure 52**)), along with the genes of interest, were double-digested with the restriction enzymes *NdeI* and *XbaI*, and ligated together resulting in 36 different vectors. Several of the 36 vectors served as PCR templates for the gene amplifications used in the 2-, 3-, or 4-gene contig assemblies described below.

[00668] The genes of interest are as follows:

[00669] CC90 *glcD* - glycolate oxidase subunit, FAD-linked NP_417453;

[00670] CC91 *glcE* - glycolate oxidase FAD binding subunit YP_026191;

[00671] CC92 *glcF* - glycolate oxidase iron-sulfur subunit YP_026190;

[00672] CC93 glyoxylate carboligase NP_415040;

[00673] CC94 tartronate semialdehyde reductase NP_417594; and

[00674] CC97 tartronate semialdehyde reductase - NADH dependent NP_415042.

[00675] These genes are described in Kebeish, R., et al., *Nature Biotechnology* (2007) 25(5) 593-599. All six genes are codon-optimized for the chloroplast genome of *Chlamydomonas reinhardtii*.

[00676] Additional genes of interest are as follows:

[00677] BD11 is an endoxylanase from *T. reesei*; and

[00678] IS99 is a mevalonate pyrophosphate decarboxylase from *S. cerevisiae*, codon optimized according to the tRNA usage of the *C. reinhardtii* chloroplast.

[00679] Other genes of interest are as follows:

[00680] IS57 is 1-Deoxy-D-xylulose 5-phosphate reductoisomerase (DXR);

[00681] IS-61 is *Chlamydomonas* chlorophyll synthase;

[00682] IS-62 is the same protein as IS-9, the chicken FPP synthase; the difference is that the C-terminal tag has been removed, and replaced with an N-terminal FLAG tag; and

[00683] IS-116 is 4-diphosphocytidyl-2-C- methylerythritol synthetase (CDP-ME synthase, it is the *E. coli* version of the gene).

[00684] These above four genes were all codon biased for expression in the *Chlamydomonas* chloroplast genome.

[00685] Plasmid vectors pRS414 (Sikorski and Hieter, *Genetics*. 1989 May;122(1):19-27) (**Figure 53**) and pBeloBAC11 (NEB)(**Figure 54**) were used to construct transformation platform vectors. In all instances, pRS414, and Gene Vectors 1, 2, and 3 were selectively maintained in DH10B cells (Invitrogen, U.S.A.) by growth in Luria Bertani (LB) medium supplemented with 100µg/ml ampicillin. Similarly, the plasmid pBeloBAC11 was selectively maintained in its host bacterium, DH10B, by growth in LB medium supplemented with 12.5µg/ml chloramphenicol.

[00686] To construct the first of the two base platform vectors (Figure 49) that can be used for the introduction of two genes into the chloroplast of *Scenedesmus obliquus*, the homology region A3 (SEQ ID NO: 126) and the homology region B3 (SEQ ID NO: 127) were amplified from *Scenedesmus* chloroplast DNA using primers 34 (SEQ ID NO: 99) and 35 (SEQ ID NO: 100), and 36 (SEQ ID NO: 101) and 37 (SEQ ID NO: 102), respectively, digested with *NotI* and *SpeI*, and ligated into *NotI* digested gutless pUC (Figure 45). Plasmid p04-35 (Figure 46) was then linearized with *SpeI* and ligated to a PCR product comprising the nucleotide sequence encoding the yeast genes URA3-ADE2 (SEQ ID NO: 105). The nucleotide sequence encoding the yeast genes URA3-ADE2 was obtained by PCR using as a DNA template, plasmid pSS-007 (Figure 47), and primers 30 (SEQ ID NO: 95) and primer 31 (SEQ ID NO: 96), which both contain *SpeI* restriction sites at their 5' termini. The resulting vector comprising the homology regions flanking the yeast genes (pSS-013) is shown in Figure 48.

[00687] The URA3-ADE2 cassette allows for positive selection in yeast that are deficient for *URA3* or *ADE2* gene function, respectively. Similarly, expression of the URA3 gene can be negatively selected against in the presence of 5-fluoroorotic acid (5-FOA) as URA3 converts 5-FOA to 5-fluorouracil, which is toxic to the cell. In addition, the presence or absence of a functional *ADE2* gene results in white or red yeast colonies, respectively -- thereby allowing for another level of selection when picking colonies.

[00688] To create the yeast-bacterial shuttle vector for two-gene contig assembly, which targets the A3-B3 region, pSS-013 (Figure 48) was digested with *NotI*, liberating the fragment containing A3-URA3-ADE2-B3, which was then ligated into *NotI* digested pRS414 (Figure 53), resulting in the vector pSS-023 (Figure 49). pSS-023 was confirmed by sequencing and restriction digest mapping with *NdeI*, *PacI*, *PstI*, *ScaI*, *SnaBI*, and *SpeI* (Figure 65). Order of lanes from left to right: 1kb DNA plus ladder (Invitrogen), uncut pSS-023, *NdeI*, *PacI*, *PstI*, *ScaI*, *SnaBI*, *SpeI*, 1kb DNA plus ladder (Invitrogen, U.S.A.). Expected bands are as follows: *NdeI*, 2187 bp and 8135bp; *PacI*, 2051 bp, 2981 bp, and 5290bp; *PstI*, 493 bp, 1872 bp, and 7957 bp; *ScaI*, 1761 bp, 4050 bp, and 4511 bp; *SnaBI*, 2587 bp and 7735 bp; and *SpeI*, 950 bp, 3694 bp, and 5678 bp. pSS-023 was used in all two-gene contig assemblies that target homology A3 and homology B3 regions.

[00689] To construct the base platform vector used for the three-gene, four-gene, and the second two-gene contig assembly (which all target the *psbB-rbcL* locus in *Scenedesmus*), primer 1 (SEQ ID NO: 66) and primer 2 (SEQ ID NO: 67), both of which contain *NotI* restriction sites at their 5' termini, were used to amplify the 5.2 kb sequence (SEQ ID NO: 125) spanning from the *psbB* gene to the *rbcL* gene. The resultant 5.2 kb PCR product and plasmid vector pRS414 (Figure 53) were both digested with *NotI* and ligated together, resulting in pLW001 (Figure 55). pLW001 was confirmed by sequencing and restriction digest mapping with *EcoRV*, *NotI*, *PmlI*, *PvuI*, and *SnaBI* (Figure 66). Order of lanes from left to right: 1kb DNA plus ladder (Invitrogen, U.S.A.), *EcoRV*, *NotI*, *PmlI*, *PvuI*, *SnaBI*, uncut, and 1kb DNA plus ladder (Invitrogen, U.S.A.). Expected bands are as follows: *EcoRV*, 1182 bp and 8867 bp; *NotI*, 4784 bp and 5265 bp; *PmlI*, 995 bp, 2644 bp, and 2695 bp; *PvuI*, 2868 bp and 7181 bp; and *SnaBI*, 2526 bp and 7523bp.

[00690] To assemble contigs of two, three, and four genes in pLW001, using negative selection, a PCR product containing the *Saccharomyces cerevisiae* genes URA3-ADE2 (SEQ ID NO: 105) was amplified with primer 27 (SEQ ID NO: 92) and primer 28 (SEQ ID NO: 93), which contain 5' tails homologous to the locus in the chloroplast sequence between *psbT* and *psbN*. This PCR product, along with pLW001 (Figure 55), were simultaneously transformed into *S. cerevisiae*. Transformants were selected for on complete synthetic media (CSM) lacking tryptophan, uracil, and

adenine (CSM – TRP – URA – ADE) using a standard lithium acetate transformation protocol (for example, as described in Gietz, R.D. and Woods, R.A., *Methods Enzymol.* (2002) 350:87-96).

[00691] Resultant yeast colonies were patched to CSM – TRP – URA – ADE and PCR screened for the correct homologous insertion of the URA3-ADE2 construct. Plasmid DNA was then harvested from PCR positive yeast clones and electroporated into *E. coli* DH10B cells (Invitrogen). Bacterial colonies were PCR screened. PCR positive clones were then harvested for plasmid DNA (Qiagen miniprep protocol). Twelve independent plasmid isolates from the above-mentioned yeast colonies were sequence confirmed and restriction enzyme mapped with *PacI*, *PstI*, *ScaI*, and *XhoI* (Figures 67A-E). Figure 67A is uncut plasmid DNA. Figure 67B is the plasmid DNA digested with *ScaI*; the expected fragments are 1761 bp, 5646 bp, and 6330 bp. Figure 67C is the plasmid DNA digested with *PacI*; the expected fragments are 4847 bp and 8890 bp. Figure 67D is the plasmid DNA digested with *XhoI*; expected fragments are 5830 bp and 7907 bp. Figure 67E is the plasmid DNA digested with *PstI*; the resulting fragments are 493 bp, 3011 bp, and 10233 bp. The resulting platform construct was designated as pLW092 (Figure 56).

[00692] The size of the contig becomes an issue in assembling contigs of three or more genes as the *colE1* origin present in the pLW092 backbone (Figure 56) is unable to support faithful duplication of plasmids greater than 20kb. To contend with this issue, a platform vector was created that is capable of larger assemblies based on the BAC cloning vector, pBeloBAC11 (Figure 54), which contains the OriS origin capable of maintaining very large DNA fragments, for example, upwards of 300kb. Briefly, pBeloBAC11 was linearized using the restriction enzyme *XhoI*. The *TRP1-ARS1-CEN4* gene sequences (SEQ ID NO: 107) was PCR-amplified from pYAC4 (ATCC; GenBank number U01086; Burke, D.T., *et al.*, *Science* (1987) 236: 806-812) with primer 3 (SEQ ID NO: 68) and primer 4 (SEQ ID NO: 69), which both contain *XhoI* ends. The *XhoI*-digested BeloBAC11 and pYAC4 sequences were ligated together. Resultant bacterial colonies were PCR screened for the correct ligation event, restriction enzyme mapped, and sequence confirmed. The resultant plasmid was designated pBeloBAC-TRP (Figure 57).

[00693] pBeloBAC-TRP was further modified to incorporate the *Scenedesmus* *psbB-rbcL* locus (containing URA3-ADE2 between *psbT* and *psbN*, from pLW092). Briefly, the *Scenedesmus* *psbB-rbcL* locus was digested away from pLW092 (Figure 56) using *NotI* and ligated into pBeloBAC11-TRP (Figure 57) (also digested with *NotI*). Resultant bacterial clones were sequence confirmed and restriction enzyme mapped with *EcoRV*, *NdeI*, *NotI*, *PacI*, *PstI*, *ScaI* and *XhoI* (Figure 68). Order of lanes from left to right: 1kb DNA plus ladder (Invitrogen, U.S.A.), empty, *EcoRV*, *NdeI*, *NcoI*, *PacI*, *PstI*, *ScaI*, *XhoI*, and 1kb DNA plus ladder (Invitrogen, U.S.A.). Expected bands are as follows: *EcoRV*, 229 bp, 1290 bp, 1461 bp, 2261 bp, 6558 bp, and 7048 bp; *NdeI*, 2187 bp, 2470 bp, 6183bp, and 8007 bp; *NotI*, 8953 bp and 9894 bp; *PacI*, 4847 bp and 14000 bp; *PstI*, 493 bp, 1541 bp, 3179 bp, 5559bp, and 8075 bp; *ScaI*, 1761 bp, 3835 bp, 4704 bp, and 8547bp; and *XhoI*, 3017 bp, 4942 bp, and 10888 bp. The resultant platform construct was designated as pLW100 (Figure 58) and is used in all of the 3- and 4-gene contig assemblies.

[00694] In addition to the genes of interest assembled into the 2- 3- and 4- gene contigs, a yeast positive selection marker and a *Scenedesmus* positive selection marker were also included. The yeast auxotrophic marker, LEU2 (SEQ ID NO: 108), along with the chloramphenicol acetyltransferase (CAT) gene (SEQ ID NO: 148) driven by the *rbcL* promoter (which confers resistance to chloramphenicol in *Scenedesmus*) (Figure 59) were ligated into gutless-pUC. Homology regions flanking these two genes were also cloned, which correspond to the adjacent genes of interest in contig assembly. Briefly, the *Saccharomyces cerevisiae* gene LEU2 (SEQ ID NO: 108), was amplified from total genomic DNA with primer 5 (SEQ ID NO: 70), which contains a *PstI* restriction site, and primer 6 (SEQ ID NO: 71),

which contains a *NotI* restriction site (at the 5' terminus) and 80bp of DNA, which are homologous to adjacent genes in 2-, 3-, and 4-gene contig assembly. In addition, the *rbcL*-CAT-psbE gene (SEQ ID NO: 128) was amplified from vector p04-198 (Figure 59) using primer 7 (SEQ ID NO: 72), which contains a *NotI* restriction site (at the 5' terminus) and 80bp of DNA which are homologous to adjacent genes in 2-, 3-, and 4-gene contig and primer 8 (SEQ ID NO: 73), which contains a *PstI* restriction site. The LEU2 and *rbcL*-CAT-psbE fragments were digested with *PstI* and *NotI* and ligated to *NotI* digested gutless-pUC. Resultant bacterial clones were sequence confirmed and restriction enzyme mapped with *EcoRI*, *EcoRV*, *KpnI*, *NotI*, *PvuII*, and *ScaI* (Figure 69). The order of lanes is as follows: 1kb DNA plus ladder (Invitrogen, U.S.A.), uncut DNA, *EcoRI*, *EcoRV*, *KpnI*, *NotI*, *PvuII*, *ScaI*, and 1kb DNA plus ladder (Invitrogen, U.S.A.). Expected bands are as follows: *EcoRI*, 3033 bp and 3458 bp; *EcoRV*, 6491 bp; *KpnI*, 6491 bp; *NotI*, 2436 bp and 4055 bp; *PvuII*, 958 bp and 5533 bp; *ScaI* 3023 bp and 3468 bp. This construct was designated as pSS-035 (Figure 60) and is used in all of the gene contigs to promote proper assembly and also to provide for a positive selection element during *Scenedesmus* transformation.

[00695] Example 25: Contig Assemblies

[00696] The *Saccharomyces cerevisiae* strain, YPH858 (*MATa*, *ura3-52*, *lys2-801*, *ade2-101*, *trp1Δ63*, *his3Δ200*, *leu2Δ1*, *cyh2R*), was used in all contig assembly reactions.

[00697] For two-gene contig assemblies targeting the A3-B3 region, the following were combined:

- 1) 1 μg of pSS-023 (Figure 49) linearized between URA3 and ADE2 with *SphI*;
- 2) 500 ng of a gel purified fragment, obtained by digesting pSS-035 (Figure 60) with *NotI*, and comprising the *rbcL*-CAT-psbE/LEU2 construct;
- 3) 500ng of PCR amplified *petA*-CC94-*chlL* (gene vector 1) (Figure 50), amplified with a forward primer, primer 9 (SEQ ID NO: 74), which is comprised of 60bp of homology to the *NotI* digestion product from pSS-035, and a reverse primer, primer 32 (SEQ ID NO: 97), which is comprised of 60bp of homology to pSS-023 just downstream of the nucleotide sequence encoding for ADE2; and
- 4) 500ng of PCR amplified *tufA*-CC93-*psaB* (gene vector 3) (Figure 52), amplified with a forward primer, primer 33 (SEQ ID NO: 98), which comprises 60bp of homology to pSS-023 just upstream of the nucleotide sequence encoding for URA3, and a reverse primer, primer 12 (SEQ ID NO: 77), which comprises 60bp of homology to the *NotI* digestion product described in step 2 above.

[00698] Cells were transformed with the mixture of DNA described above, using a standard lithium acetate transformation protocol. Transformants were selected for on CSM –TRP–LEU +5-FOA plates. After two days at 30 °C, yeast colonies were picked and patched to a CSM–TRP–LEU plate. The next day, yeast patches were PCR screened for the correct gene assembly. Plasmid DNA was then harvested from PCR positive yeast clones and electroporated into *E. coli* DH10B cells (Invitrogen). Bacterial colonies were also PCR screened. Four PCR positive clones were then harvested for the preparation of plasmid DNA (Qiagen miniprep protocol), which were subsequently restriction enzyme mapped with *NdeI* (Figure 70; expected band sizes, 1097 bp, 3703 bp, and 10283 bp; and 1kb DNA plus ladder (Invitrogen, U.S.A.). One of the four clones was picked and the sequence of that clone was confirmed. The resulting two-gene contig assembly is shown in Figure 61. Another embodiment of this assembly is shown in Figure 62.

[00699] For two-gene contig assemblies targeting the 5.2kb *psbB*-*rbcL* region, the following were combined:

- 1) 1 μg of pLW092 (Figure 56) linearized between URA3 and ADE2 with *SphI*;

- 2) 500 ng of a gel purified fragment, obtained by digesting pSS-035 (Figure 60) with *NotI*, and comprising the *rbcl*-CAT-*psbE*/LEU2 construct;
- 3) 500ng of PCR amplified *petA*-BD11-*chlL* (gene vector 1) (Figure 50), amplified with a reverse primer, primer 1001 (SEQ ID NO: 150), which is comprised of 60bp of homology to the *NotI* digestion product from pSS-035, and a forward primer, primer 1000 (SEQ ID NO: 149), which is comprised of 60bp of homology to pLW092 just upstream of the nucleotide sequence encoding for URA3; and
- 4) 500ng of PCR amplified *tufA*-IS99-*psaB* (gene vector 3) (Figure 52), amplified with a reverse primer, primer 1002 (SEQ ID NO: 151), which comprises 60bp of homology to pLW092 just downstream of the nucleotide sequence encoding for ADE2, and a forward primer, primer 1003 (SEQ ID NO: 152), which comprises 60bp of homology to the *NotI* digestion product described in step 2 above.

[00700] Cells were transformed with the mixture of DNA described above, using a standard lithium acetate transformation protocol. Transformants were selected for on CSM –TRP –LEU +5-FOA plates. After two days at 30 °C, yeast colonies were picked and patched to a CSM-TRP-LEU plate. The next day, yeast patches were PCR screened for the correct gene assembly. Plasmid DNA was then harvested from PCR positive yeast clones and electroporated into *E. coli* DH10B cells (Invitrogen). Bacterial colonies were also PCR screened. Four PCR positive clones were then harvested for the preparation of plasmid DNA (Qiagen miniprep protocol), which were subsequently restriction enzyme mapped. Figure 90A-D depicts mapping of the two gene contig assembly with the restriction enzymes: *KpnI* (A), *MscI* (B), *PvuII* (C), and also uncut DNA (D). Expected band sizes are as follows: *KpnI*: 670 bp, 1791 bp, 2555bp, and 13163 bp; *MscI*: 2206bp and 15973bp; and *PvuII*: 21bp, 195bp, 1421bp, 3289bp, 3908bp, 4336bp, and 5009bp (note: the 21bp and 195bp bands have run off the gel in Figure 90C). One of the four clones was picked and the sequence of that clone was confirmed. The resulting two-gene contig assembly targeting the *psbB*-*rbcl* locus is shown in Figure 91.

[00701] For a three-gene contig assembly, the following were combined:

- 1) 1µg of pLW100 (Figure 58) linearized between URA3 and ADE2 with *SphI*;
- 2) 500 ng of a gel purified fragment, obtained by digesting pSS-035 (Figure 60) with *NotI*, and comprising the *rbcl*-CAT-*psbE*/LEU2 construct;
- 3) 500ng of PCR amplified *petA*-CC90-*chlL* (gene vector 1) (Figure 50), amplified with a forward primer, primer 13 (SEQ ID NO: 78), which comprises 60bp of homology to the *NotI* digestion product from pSS-035, and a reverse primer, primer 14 (SEQ ID NO: 79), which comprises 60bp of homology to pLW100 just upstream of the nucleotide sequence encoding for URA3;
- 4) 500ng of PCR amplified *tufA*-CC91-*psaB* (gene vector 3) (Figure 52), amplified with a forward primer, primer 15 (SEQ ID NO: 80), which comprises 60bp of homology to the *NotI* digestion product from step 2, and a reverse primer, primer 16 (SEQ ID NO: 81), which comprises 60bp of homology to the PCR amplified gene vector 2 (Figure 51); and
- 5) 500ng of PCR amplified D2-CC92-D1 (gene vector 2) (Figure 51), amplified with a forward primer, primer 29 (SEQ ID NO: 94), which comprises 60bp of homology to PCR amplified gene

vector 2, and a reverse primer, primer 17 (SEQ ID NO: 82), which comprises 60bp of homology to pLW100, just downstream of the nucleotide sequence encoding for ADE2.

[00702] Cells were transformed with the mixture of DNA described above, using a standard lithium acetate transformation protocol. Transformants were selected for on CSM –TRP –LEU +5-FOA plates. After two days at 30°C, yeast colonies were picked and patched to a CSM –TRP –LEU plate. The next day, yeast patches were PCR screened for the correct gene assembly. Plasmid DNA was then harvested from PCR positive yeast clones and electroporated into *E. coli* DH10B cells (Invitrogen). Bacterial colonies were also PCR screened. Two PCR positive clones were then harvested for plasmid DNA (Qiagen maxiprep protocol), which were subsequently restriction enzyme mapped with *NdeI* (Figure 71; expected bands, 2396 bp, 3873 bp, 5114 bp, 6929 bp, and 8007 bp; and 1kb DNA plus ladder (Invitrogen, U.S.A.)). One of the two clones was picked and the sequence of that clone was confirmed. The resulting three-gene contig assembly is shown in Figure 63.

[00703] To facilitate proper assembly of the 4-gene contig assembly, two positive selection yeast auxotrophic markers, HIS3 (SEQ ID NO: 118) and LYS2 (SEQ ID NO: 119), were added to the contig assembly.

[00704] For four-gene contig assemblies, the following were combined:

- 1) 1µg of pLW100 (Figure 58) linearized between URA3 and ADE2 with *SphI*;
- 2) 500 ng of a gel purified fragment, obtained by digesting pSS-035 (Figure 60) with *NotI*, and comprising the *rbcl*-CAT-*psbE*/LEU2 construct;
- 3) 500ng of PCR amplified *tufA*-IS57-*psaB* (gene vector 3) (Figure 52), amplified with a forward primer, primer 19 (SEQ ID NO: 84, which contains 60bp of homology to PCR amplified HIS3, and a reverse primer, primer 20 (SEQ ID NO: 85), which contains 60bp of homology to pLW100 just upstream of the nucleotide sequence encoding for URA3;
- 4) 500ng of PCR amplified HIS3, amplified with a forward primer, primer 21 (SEQ ID NO: 86), which contains 60bp of homology to PCR amplified gene vector 3 and a reverse primer, primer 22 (SEQ ID NO: 87), which contains 60bp of homology to PCR amplified gene vector 1 (Figure 50);
- 5) 500ng of PCR amplified *petA*-IS116-*chlL* (gene vector 1), amplified with a forward primer, primer 13 (SEQ ID NO: 78), which contains 60bp of homology to the *NotI* digestion product from step 2, and a reverse primer, primer 23 (SEQ ID NO: 88), which contains 60bp of homology to PCR amplified HIS3;
- 6) 500ng of PCR amplified *tufA*-IS62-*psaB* (gene vector 3), amplified with a forward primer, primer 16 (SEQ ID NO: 81), which contains 60bp of homology to the *NotI* digestion product from step 2 and a reverse primer, primer 15 (SEQ ID NO: 80), which contains 60bp of homology to PCR amplified LYS2;
- 7) 500ng of PCR amplified LYS2, amplified with a forward primer, primer 24 (SEQ ID NO: 89), which contains 60bp of homology to PCR amplified gene vector 3 and a reverse primer, primer 25 (SEQ ID NO: 90), which contains 60bp of homology to PCR amplified gene vector 2 (Figure 51); and
- 8) 500ng of PCR amplified D2-IS61-D1 (gene vector 2), amplified with a forward primer, primer 26 (SEQ ID NO: 91), which contains 60bp of homology to PCR amplified LYS2 and a reverse primer,

primer 18 (SEQ ID NO: 83), which contains 60bp of homology to pLW100 just downstream of ADE2.

[00705] Cells were transformed with this mixture of DNA using a standard lithium acetate transformation protocol. Transformants were selected for on CSM-TRP-LEU-HIS-LYS +5-FOA plates. After two days at 30°C, yeast colonies were picked and patched to a CSM-TRP-LEU-HIS-LYS plate. The next day, yeast patches were PCR screened for the correct gene assembly. Plasmid DNA was then harvested from PCR positive yeast clones and electroporated into *E. coli* DH10B cells (Invitrogen). Bacterial colonies were also PCR screened. Four PCR positive clones were then harvested for plasmid DNA (Qiagen maxiprep protocol), which were subsequently restriction enzyme mapped with *NdeI* (Figure 72; expected bands, 553 bp, 564 bp, 1570 bp, 1791 bp, 1824 bp, 1969, 2040 bp, 3858 bp, 5114 bp, 7219 bp, and 8007 bp; and 1kb DNA plus ladder (Invitrogen, U.S.A.)). One of the four clones was picked and the sequence of that clone was confirmed. The resulting four-gene contig assembly is shown in Figure 64.

[00706] Example 26: *Scenedesmus* chloroplast transformation

[00707] Once construct integrity was confirmed for each of the gene assemblies (2-, 3-, and 4-gene contigs), each of the gene assemblies were individually transformed into *Scenedesmus obliquus*. Briefly, cells were grown to mid-log phase and harvested. Approximately 5×10^7 cells were spread onto TAP plates containing 25 µg/ml chloramphenicol and allowed to dry in a sterile culture hood. While plates were drying, 10 µg of plasmid DNA (from each of the contig assemblies) was bound to gold beads and transformation was conducted using a biolistic gene gun (Bio-rad) at 500psi. 2 µg of DNA was loaded into each shot and each plate was shot five times. Plates were placed under constant light for about 10 days. After which, chloramphenicol resistant colonies were picked and patched to a TAP plate containing 25 µg/ml chloramphenicol. Three to four days later, algae patches were picked into 10mM EDTA, boiled for 10 minutes and then used in a standard PCR reaction to screen for the introduction of the genes into the chloroplast.

Chloramphenicol resistant transformants potentially containing the 2-gene contig, targeting the *psbB-rbcL* locus, were screened for the presence of BD11 and IS99. Primers 1004 (SEQ ID NO: 153) and 1005 (SEQ ID NO: 154) screen for the presence of BD11, while primers 1006 (SEQ ID NO: 155) and 1007 (SEQ ID NO: 156) screen for the presence of IS99. Figure 92A and 92B depict 4 clones that screen PCR positive for both IS99 and BD11, respectively.

Chloramphenicol resistant transformants potentially containing the 3-gene contig, targeting the *psbB-rbcL* locus, were screened for the presence of CC90, CC91, and CC92. Primers 1008 (SEQ ID NO: 157) and 1009 (SEQ ID NO: 158) screen for the presence of CC90, primers 1010 (SEQ ID NO: 159) and 1011 (SEQ ID NO: 160) screen for the presence of CC91, and primers 1012 (SEQ ID NO: 161) and 1013 (SEQ ID NO: 162) screen for the presence of CC92. Figures 93A-C depict 4 clones that screen PCR positive for CC90, CC91, and CC92. Chloramphenicol resistant transformants potentially containing the 4-gene contig, targeting the *psbB-rbcL* locus, were screened for the presence of IS61, IS62, IS57, and IS116. Primers 1014 (SEQ ID NO: 163) and 1015 (SEQ ID NO: 164) screen for the presence of IS61, primers 1016 (SEQ ID NO: 165) and 1017 (SEQ ID NO: 166) screen for the presence of IS62, primers 1018 (SEQ ID NO: 167) and 1019 (SEQ ID NO: 168) screen for the presence of IS57, and primers 1020 (SEQ ID NO: 169) and 1021 (SEQ ID NO: 170) screen for the presence of IS116. Figure 94A and 94B depict 2 clones that screen PCR positive for IS57, IS116 (A), and IS61, IS62 (B). Taken together these data demonstrate that one skilled in the art can integrate multiple gene contigs of varying sizes (2 gene: 8.1kb, 3 gene: 11.2kb, and 4 gene: 19.4kb) into the chloroplast genome of *Scenedesmus* in a single step.

[00708] One of skill in the art will appreciate that many other methods known in the art may be substituted in lieu of the ones specifically described or referenced.

[00709] While certain embodiments 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 disclosure. It should be understood that various alternatives to the embodiments of the disclosure described herein may be employed in practicing the disclosure. It is intended that the following claims define the scope of the disclosure and that methods and structures within the scope of these claims and their equivalents be covered thereby.

THE CLAIMS DEFINING THE INVENTION ARE AS FOLLOWS:

1. A method of transforming a chloroplast genome of a *Scenedesmus sp.* with at least one exogenous nucleotide sequence, comprising:
 - i) obtaining the exogenous nucleotide sequence, wherein the exogenous nucleotide sequence comprises a nucleic acid sequence encoding a protein;
 - ii) binding the exogenous nucleotide sequence onto a particle; and
 - iii) transforming the exogenous nucleotide sequence into the *Scenedesmus sp.* chloroplast by biolistic particle bombardment, wherein the particle is accelerated using a helium pressure of 300 psi to 500 psi.
2. The method of claim 1, wherein the exogenous nucleotide sequence is at least 0.5 kb, at least 1.0 kb, at least 2 kb, at least 3kb, at least 5 kb, at least 8 kb, at least 11 kb, or at least 19 kb in size.
3. The method of claim 1 or 2, wherein the nucleic acid is codon optimized for expression in the chloroplast genome of the *Scenedesmus sp.*
4. The method of any one of claims 1 to 3, wherein the particle is a gold particle or a tungsten particle.
5. The method of claim 4, wherein the gold particle is about 550 nm to about 1000 nm in diameter.
6. The method of any one of claims 1 to 5, wherein the particle is accelerated using a helium pressure of at least 350psi, at least 400psi, at least 425psi, or at least 450psi.
7. The method of any one of claims 1 to 6, wherein the exogenous nucleotide sequence bound to the particle is shot at a distance of 2 to 4 cm from the *Scenedesmus sp.*
8. The method of any one of claims 1 to 7, wherein the biolistic bombardment is achieved using a Helicos Gene Gun or an Accell Gene Gun.
9. The method of any one of claims 1 to 8, wherein the nucleic acid encodes for a protein involved in isoprenoid biosynthesis.
10. The method of any one of claims 1 to 8, wherein the nucleic acid encodes for a protein involved in fatty acid biosynthesis.
11. The method of any one of claim 1 to 10, wherein the *Scenedesmus sp.* is *S. dimorphus*.

12. The method of any one of claims 1 to 10, wherein the *Scenedesmus sp.* is *S. obliquus*.
13. A transformed chloroplast genome of a *Scenedesmus sp.*, transformed by the method of any one of claims 1 to 12.
14. A transformed *Scenedesmus sp.* containing at least one exogenous nucleotide sequence transformed by the method of any one of claims 1 to 12.
15. A method according to claim 1, substantially as hereinbefore described.
16. A transformed chloroplast genome of a *Scenedesmus sp.*, according to claim 13, substantially as hereinbefore described.
17. A transformed *Scenedesmus sp.* according to claim 14, substantially as hereinbefore described.

1/113

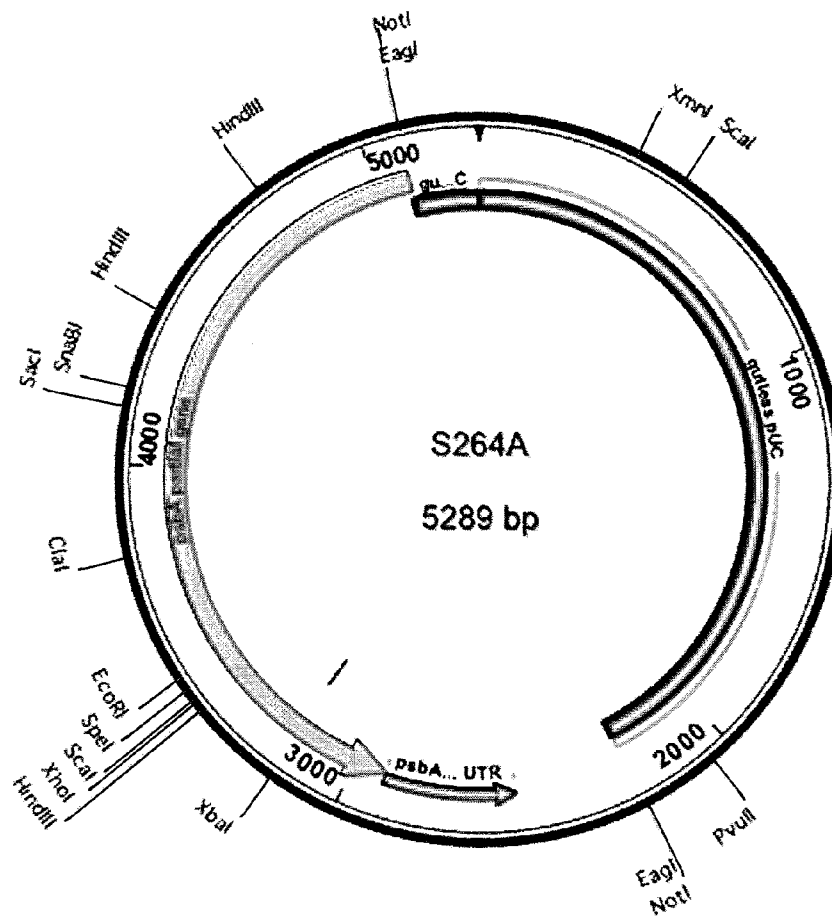


FIG. 1

2/113

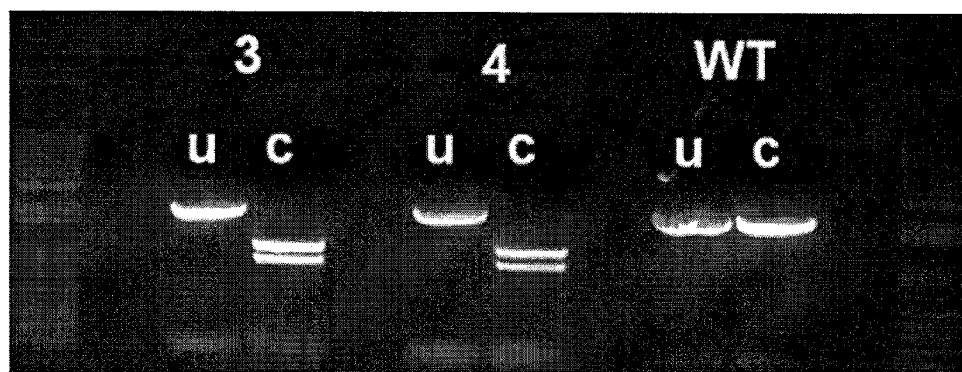


FIG. 2

3/113

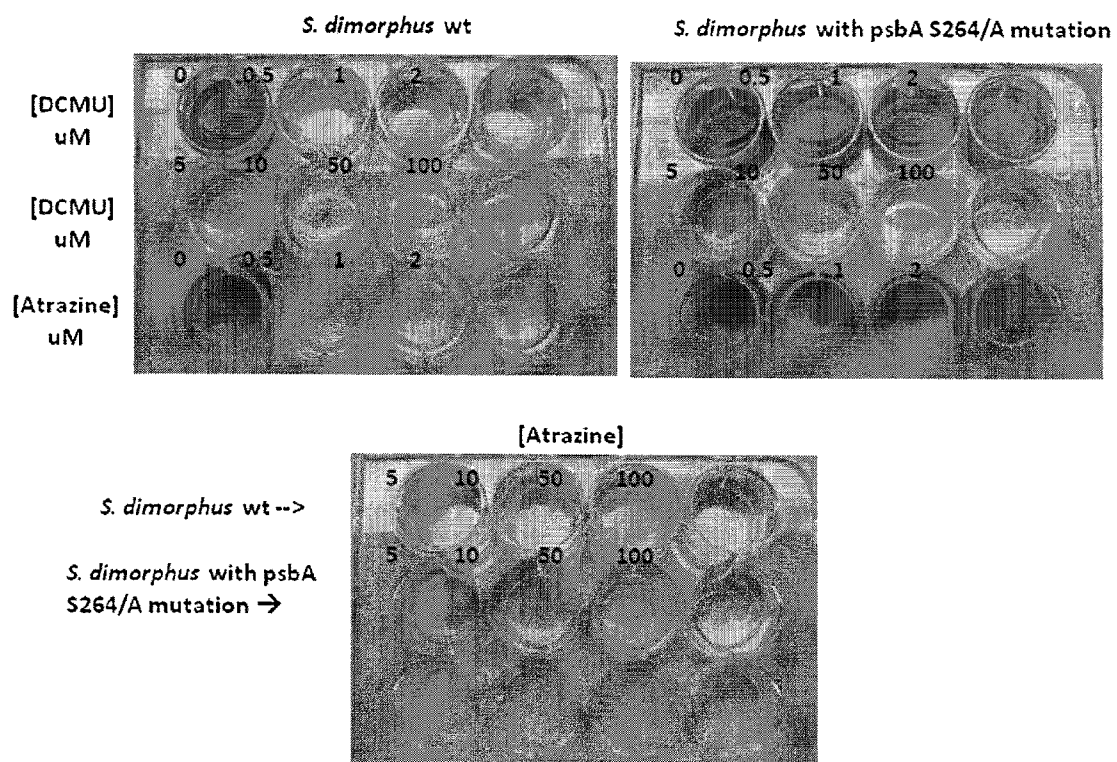


FIG. 3

4/113

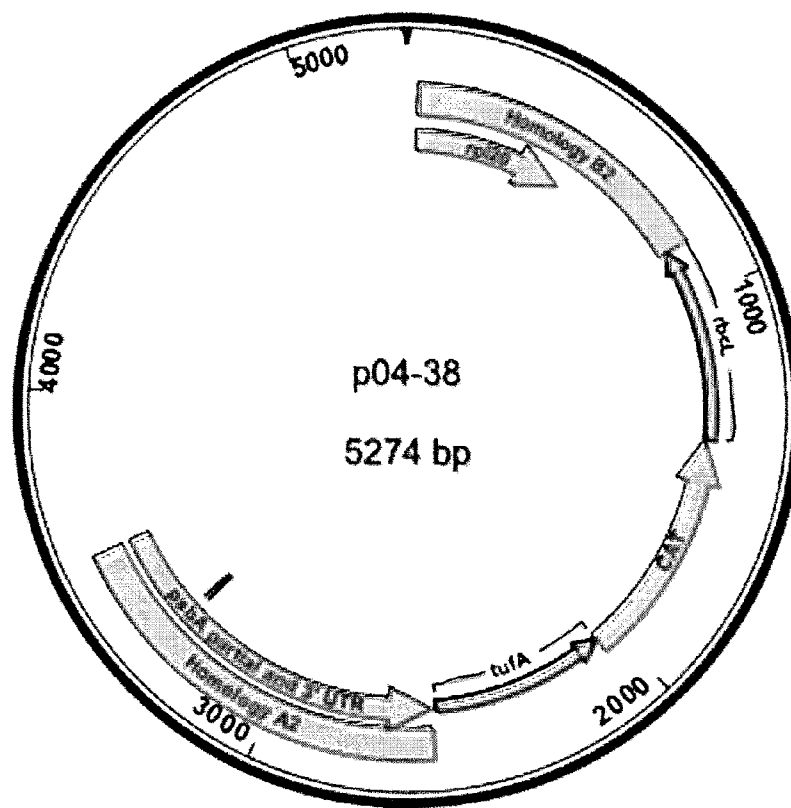


FIG. 4

5/113

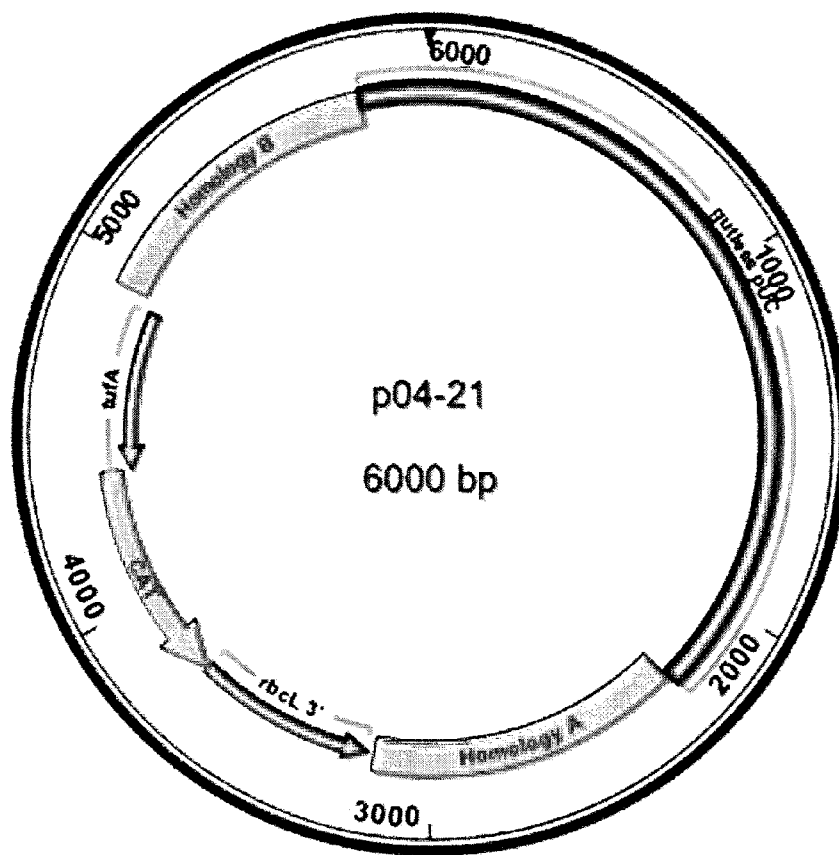


FIG. 5

6/113

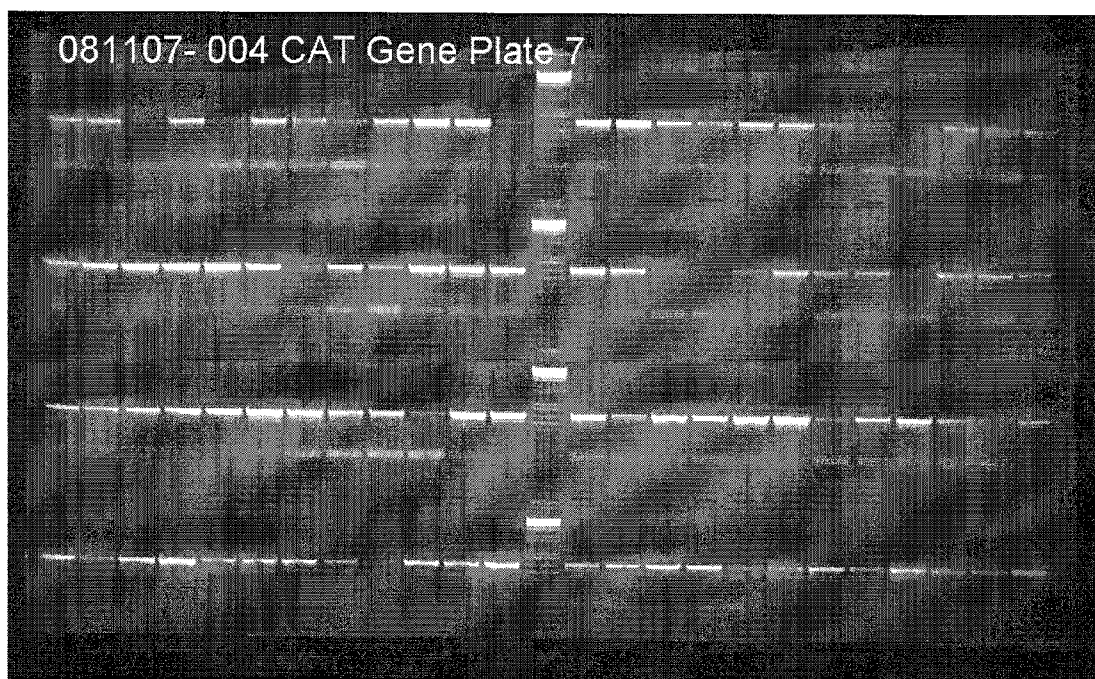


FIG. 6

7/113

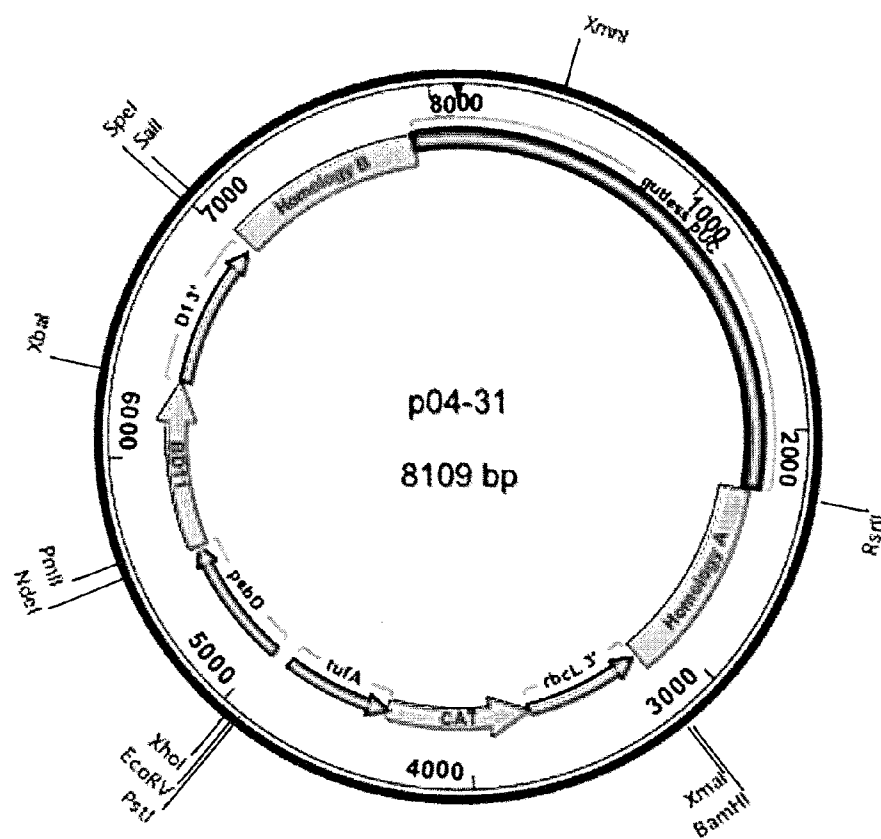


FIG. 7

8/113

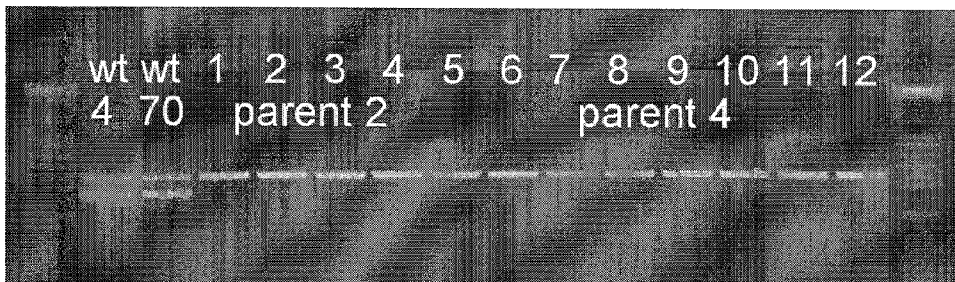


FIG. 8

9/113

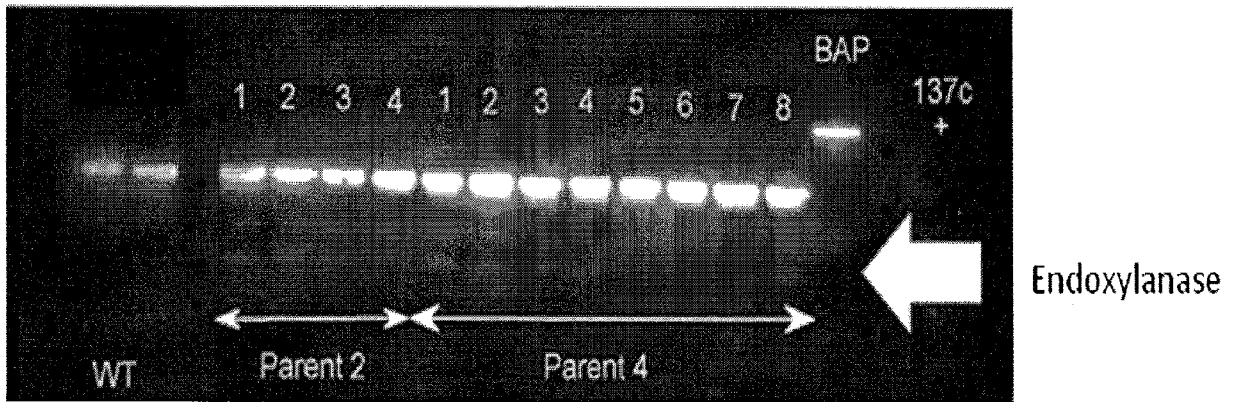


FIG. 9

10/113

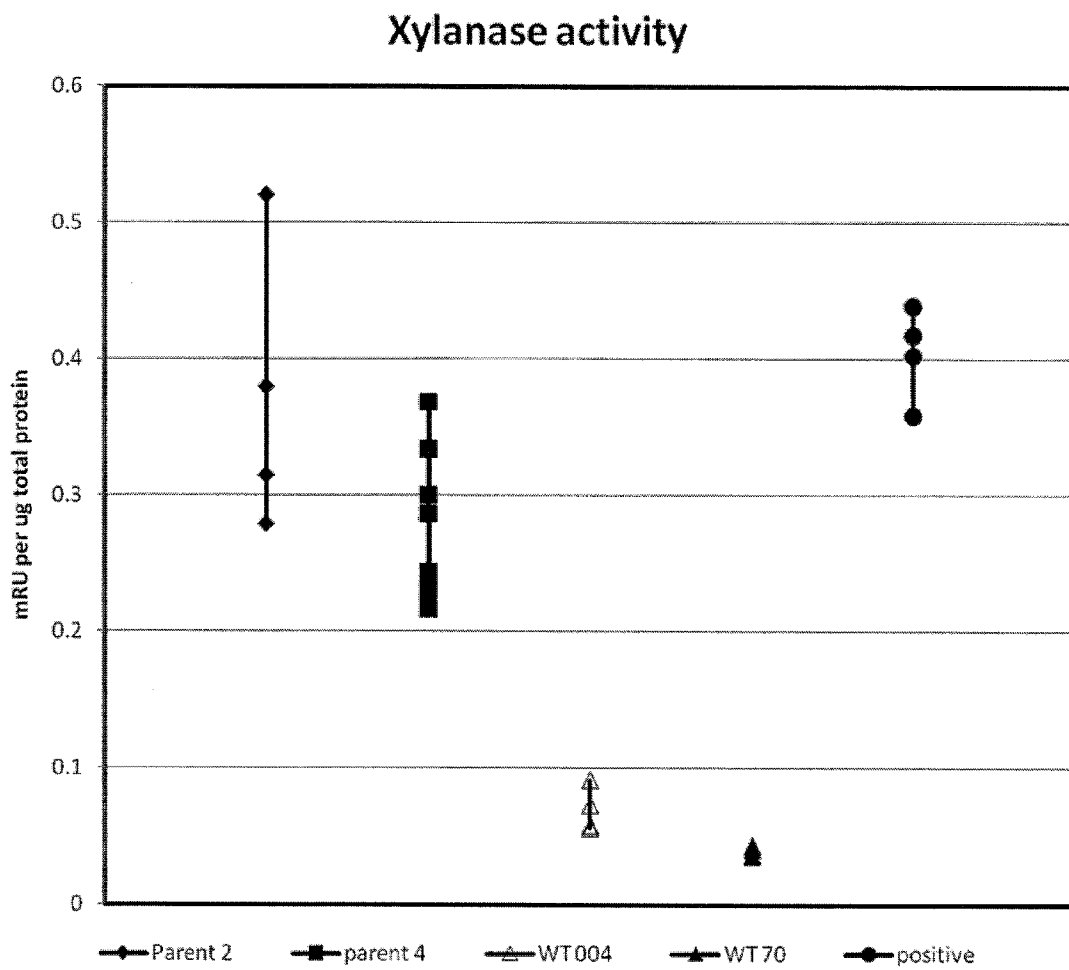


FIG. 10

11/113

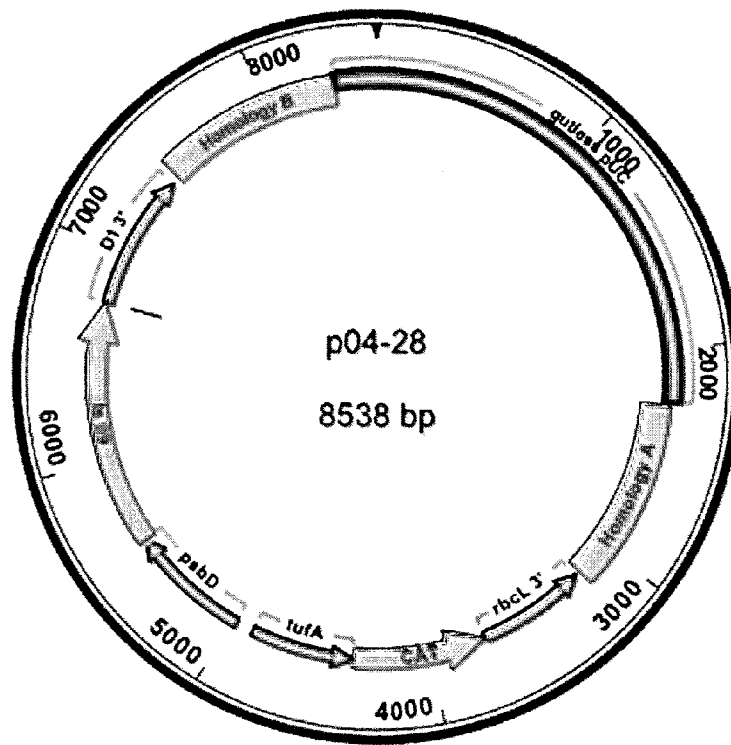


FIG. 11

12/113

WT 10 13 15 24 25 26 18 H₂O

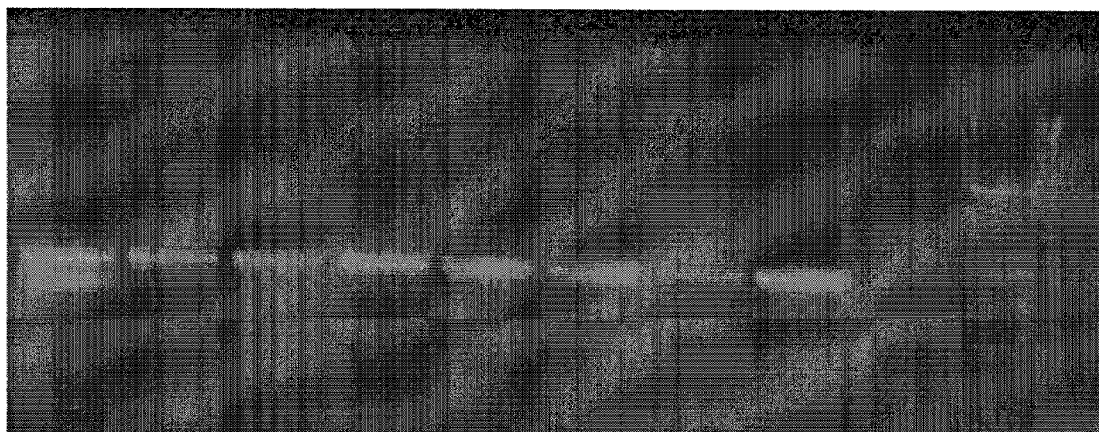


FIG. 12

13/113

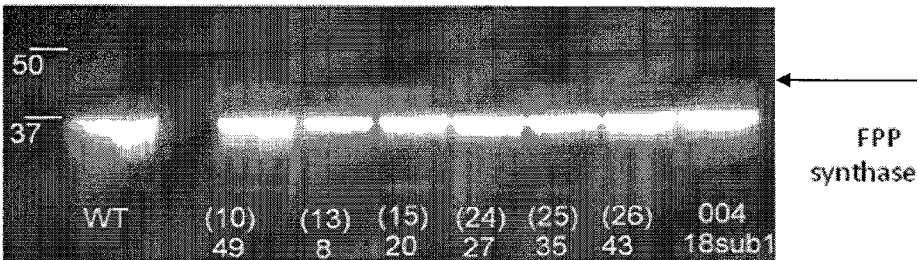


FIG. 13

14/113

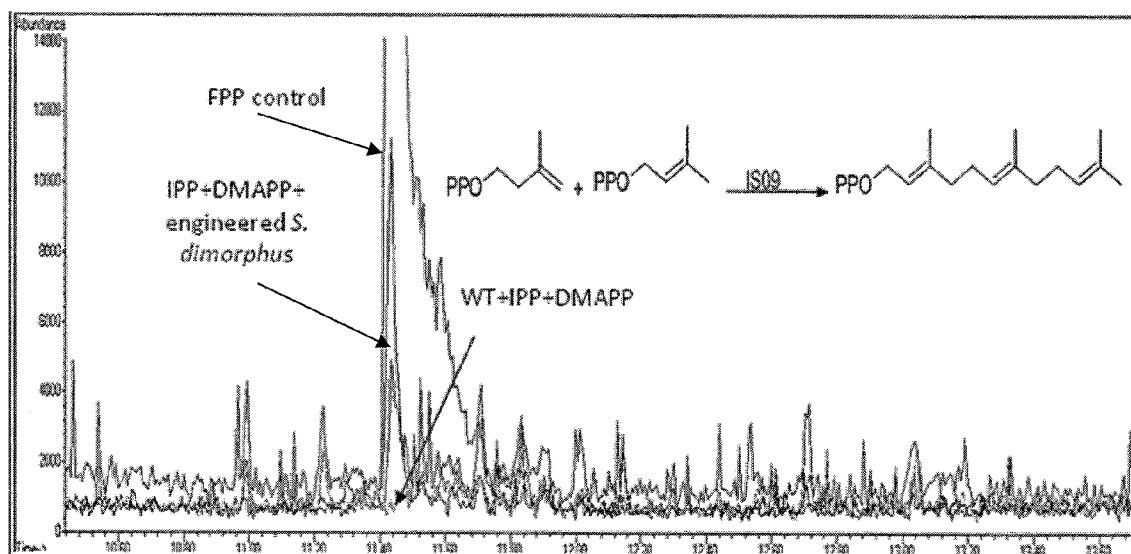
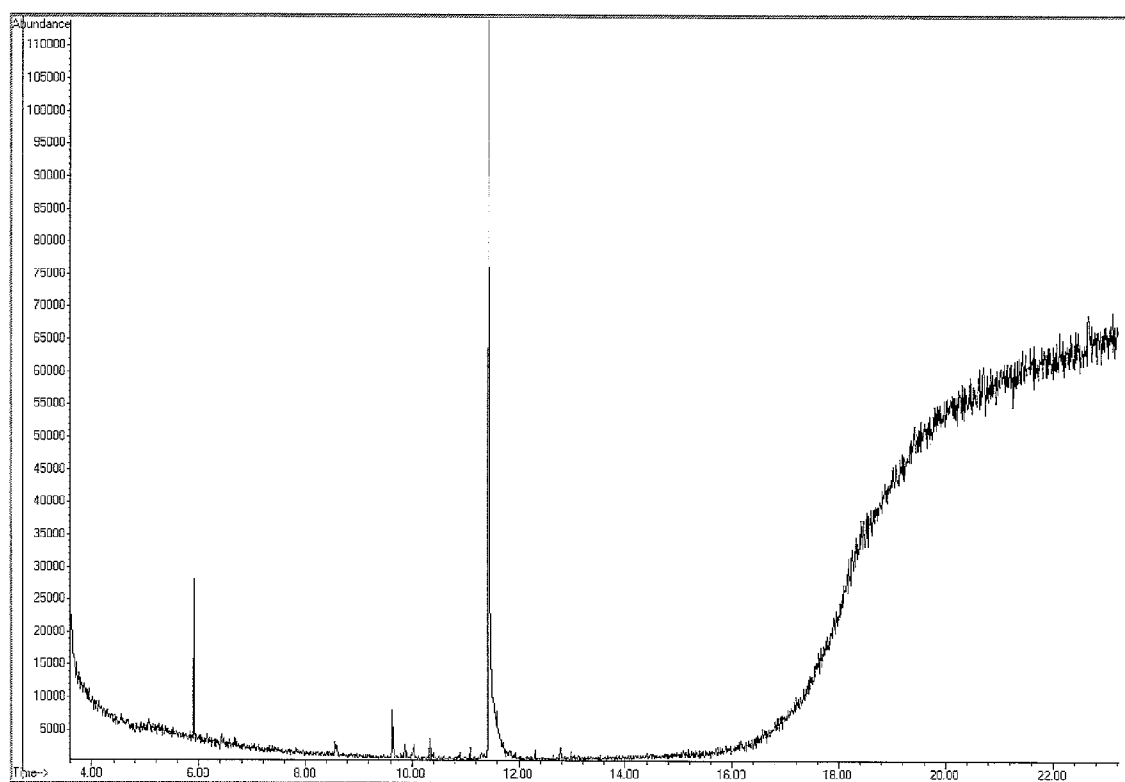
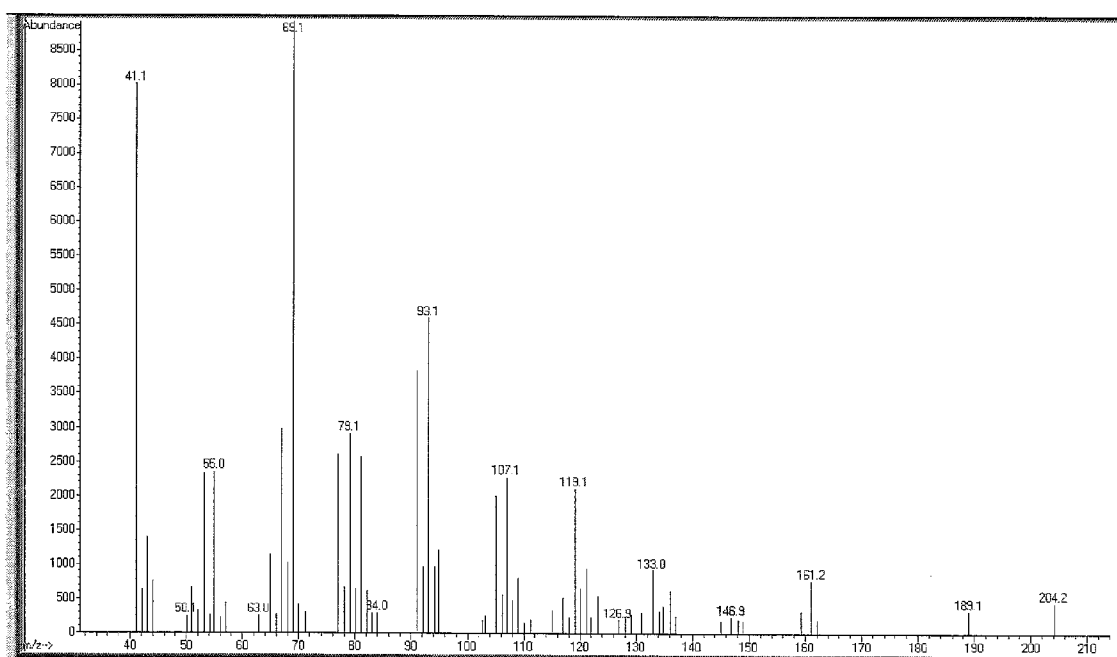
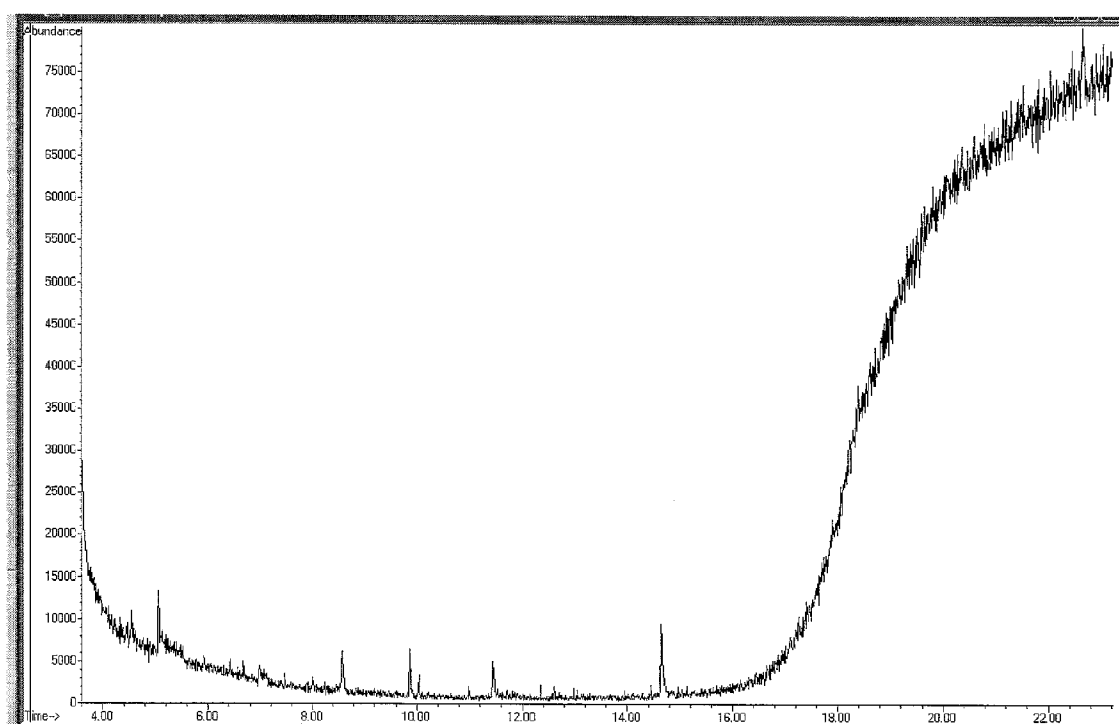
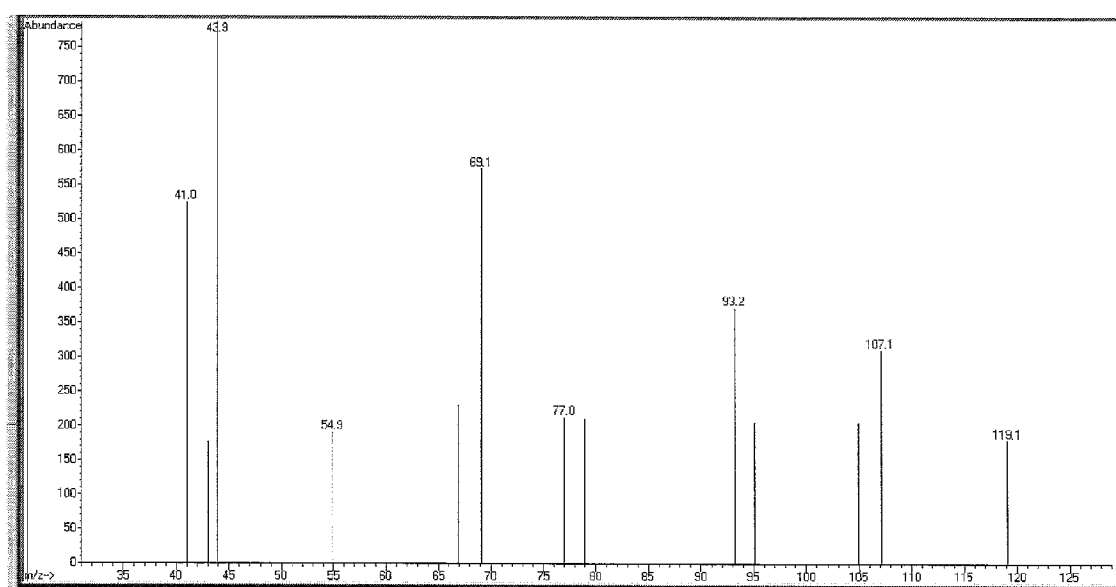


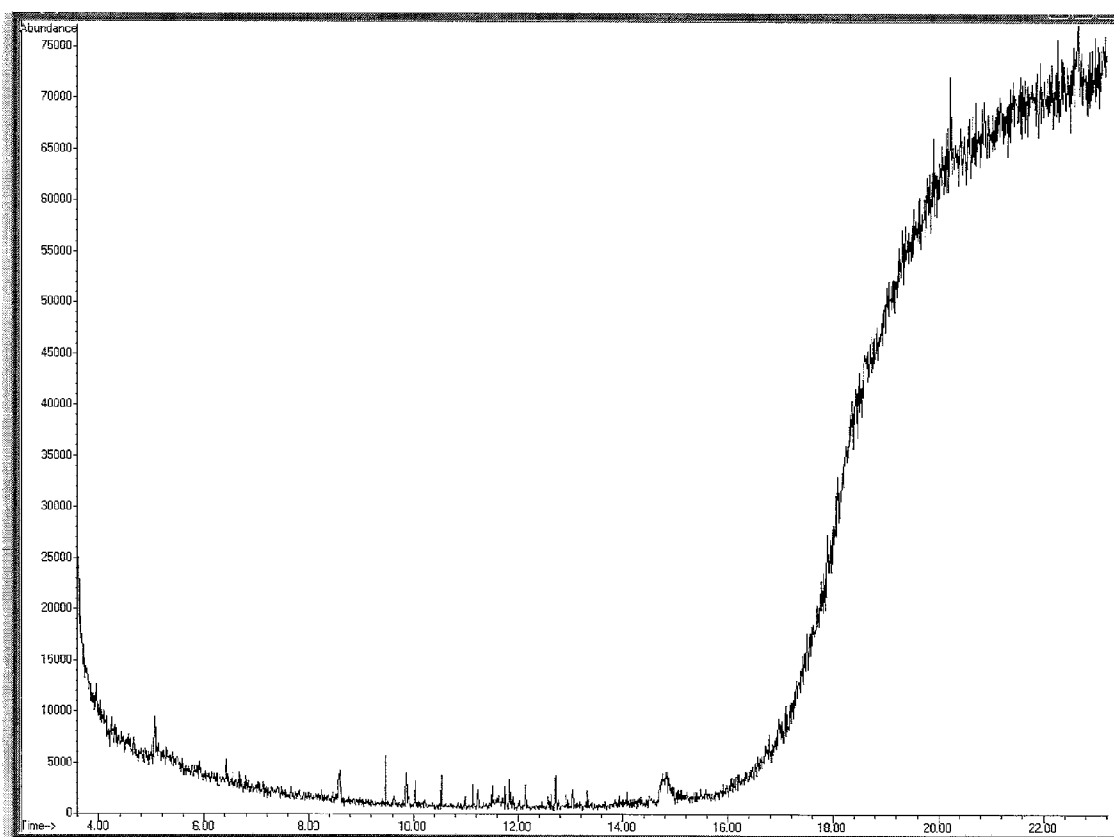
FIG. 14A

15/113*FIG. 14B*

16/113*FIG. 14C*

17/113*FIG. 14D*

18/113*FIG. 14E*

19/113*FIG. 14F*

20/113

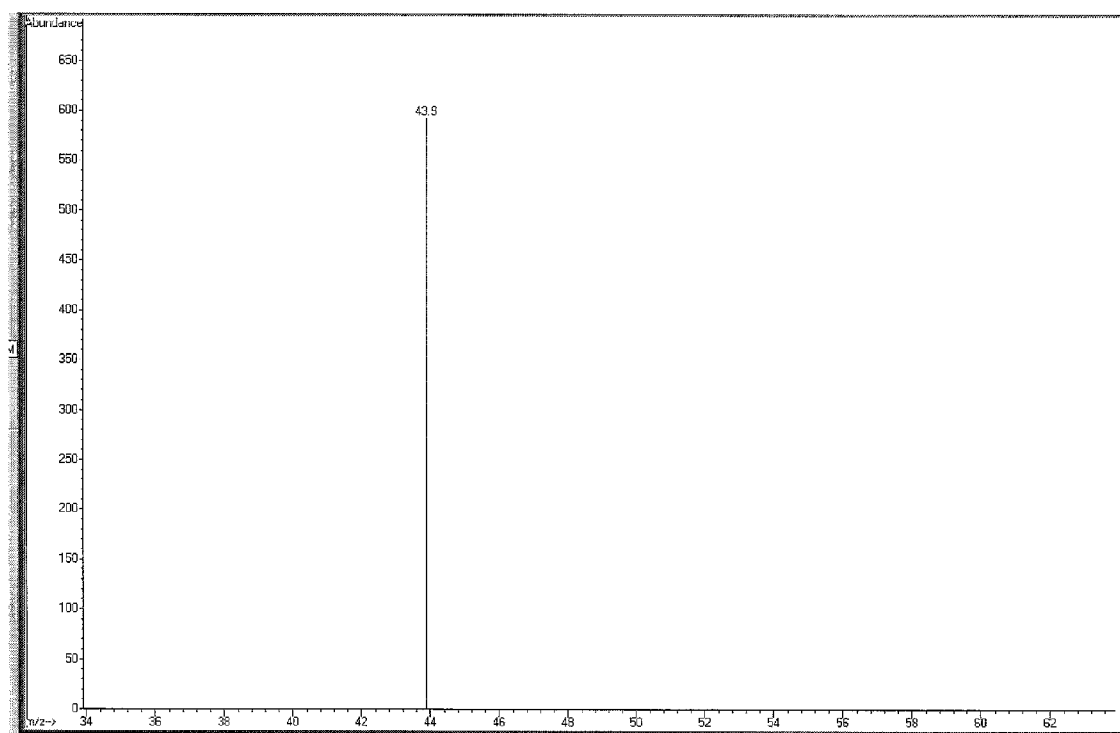


FIG. 14G

21/113

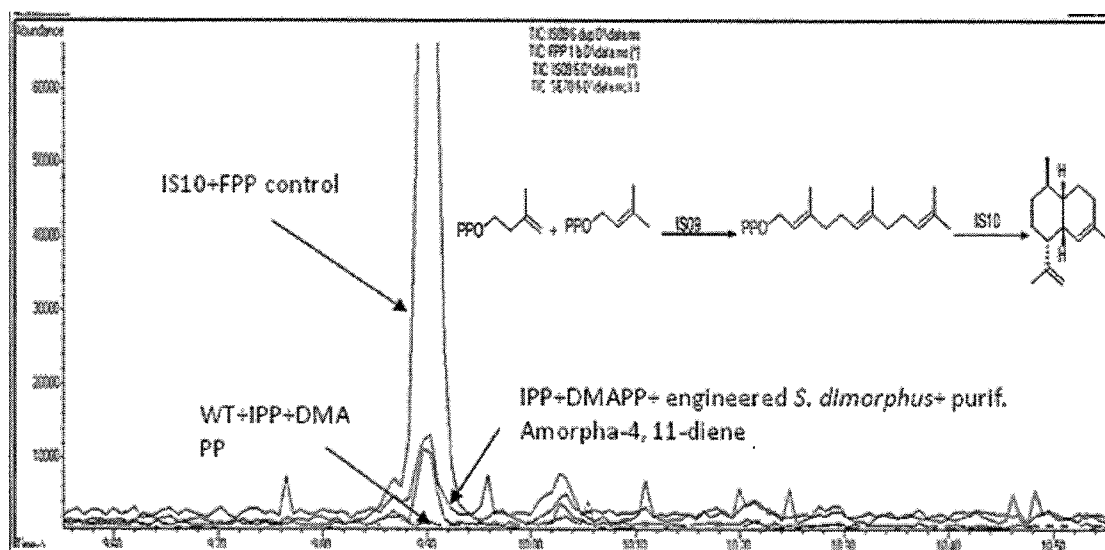


FIG. 15A

22/113

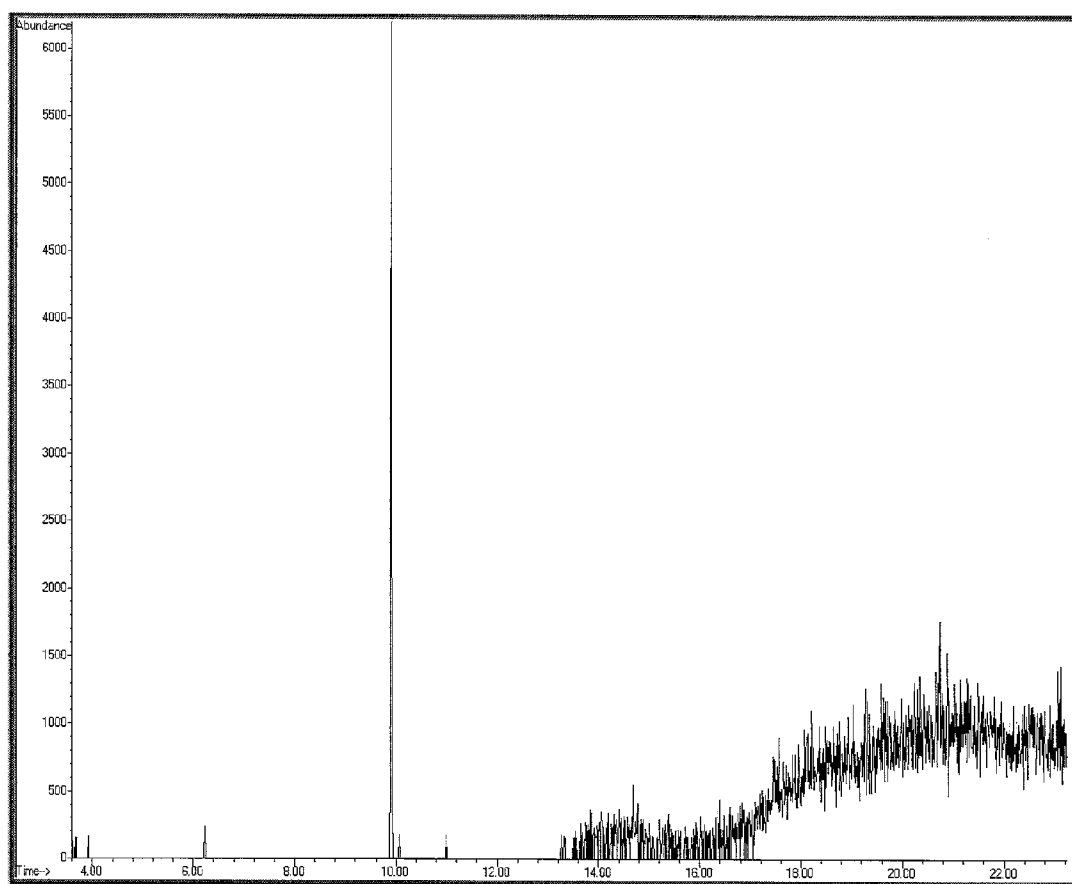
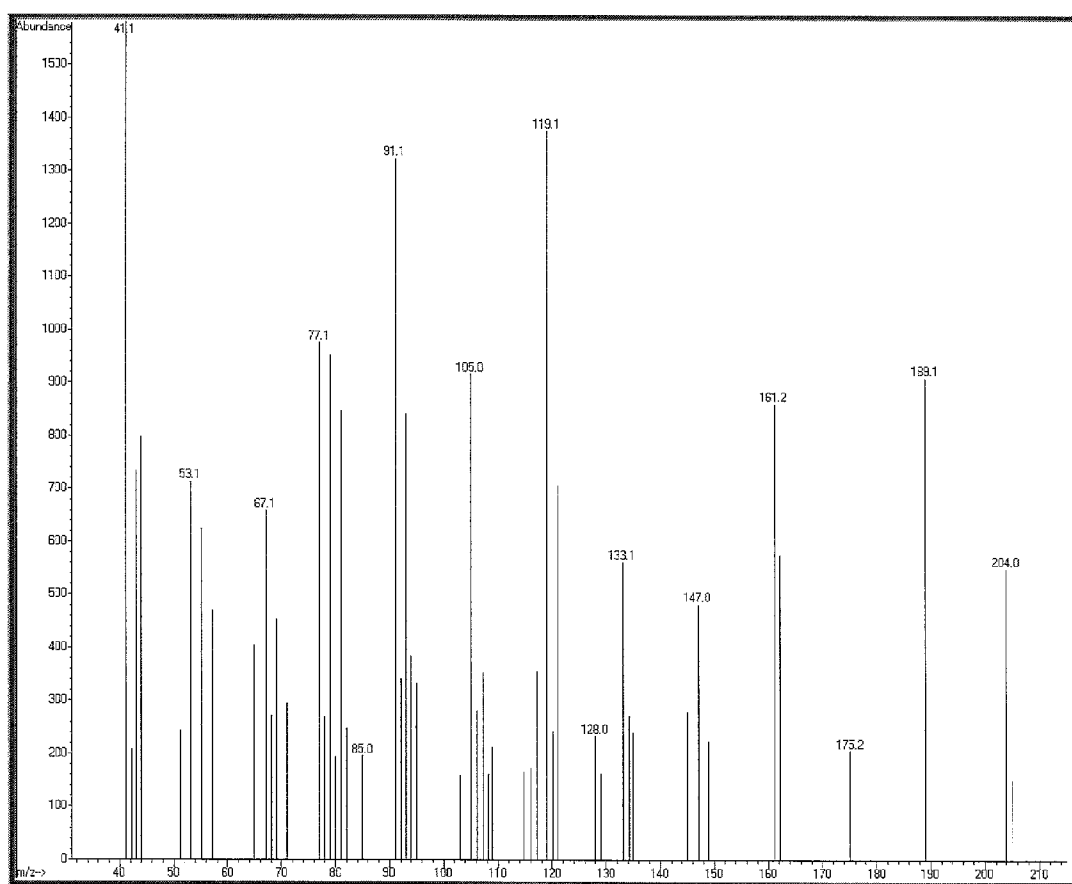


FIG. 15B

23/113*FIG. 15C*

24/113

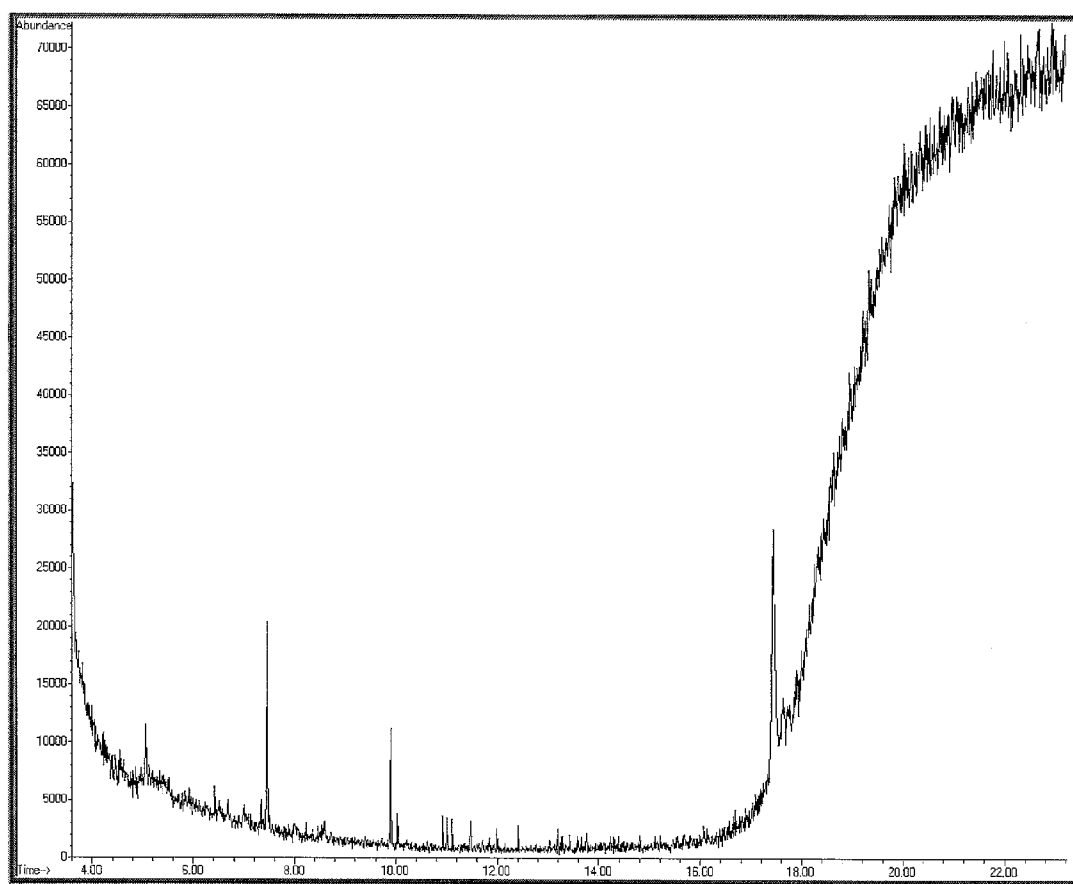


FIG. 15D

25/113

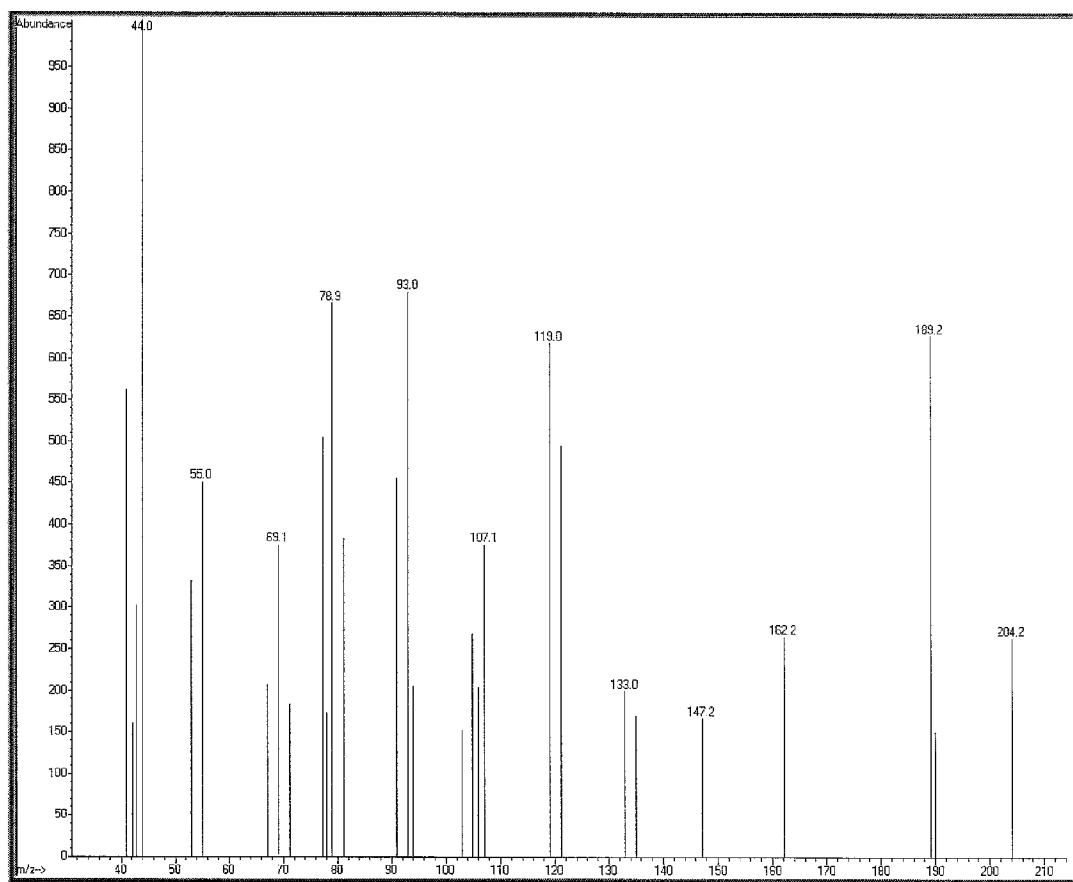


FIG. 15E

26/113

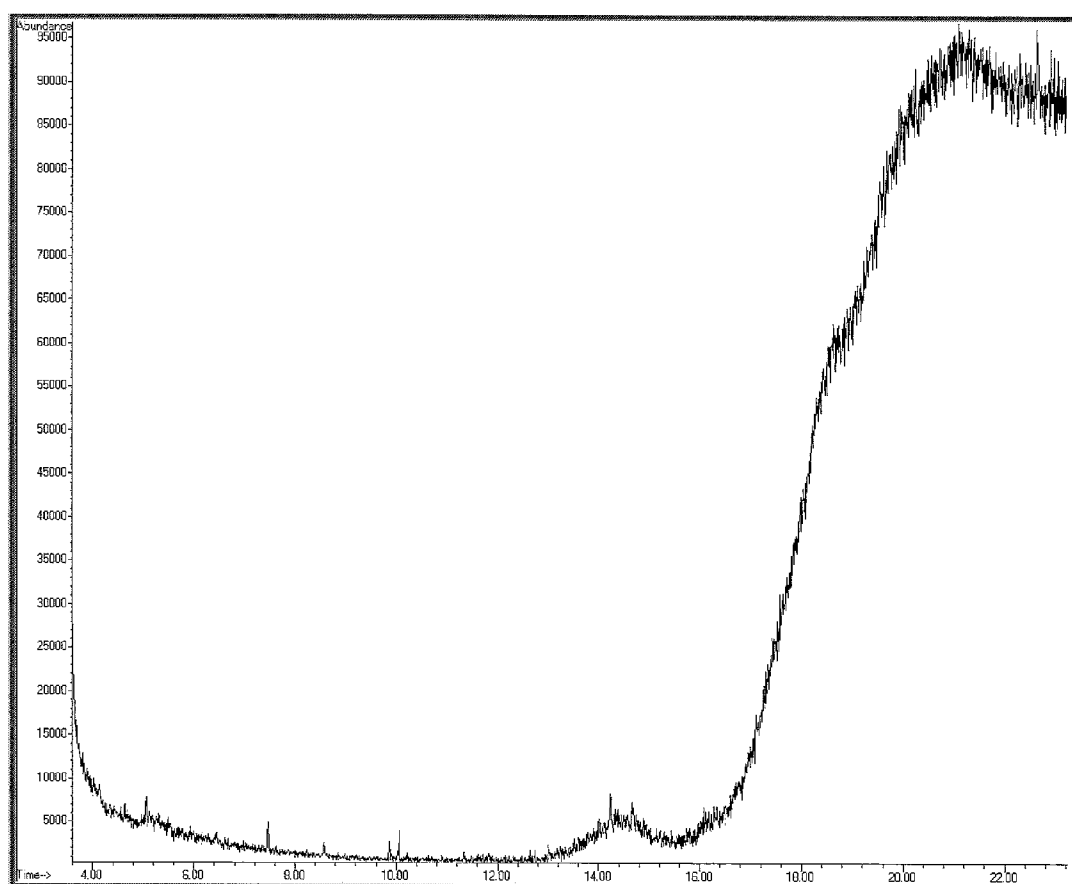


FIG. 15F

27/113

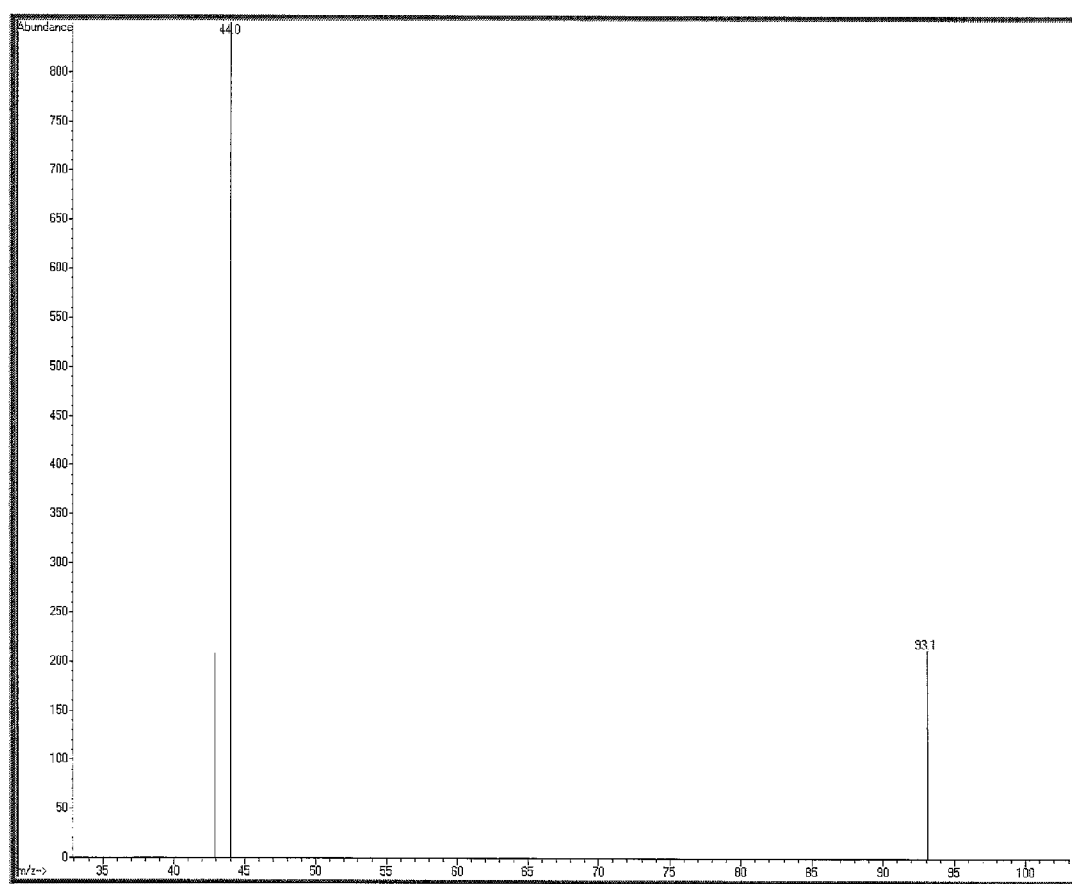


FIG. 15G

28/113

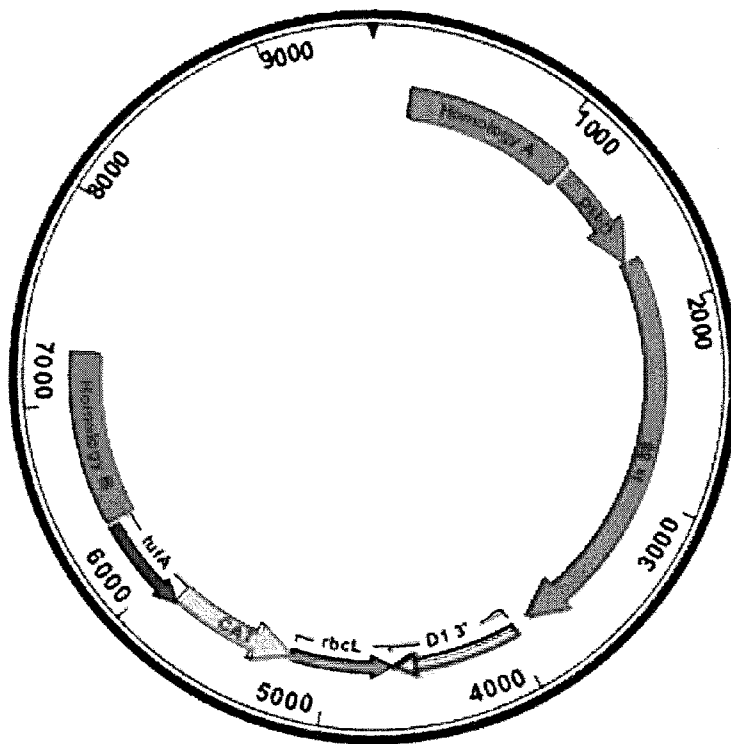
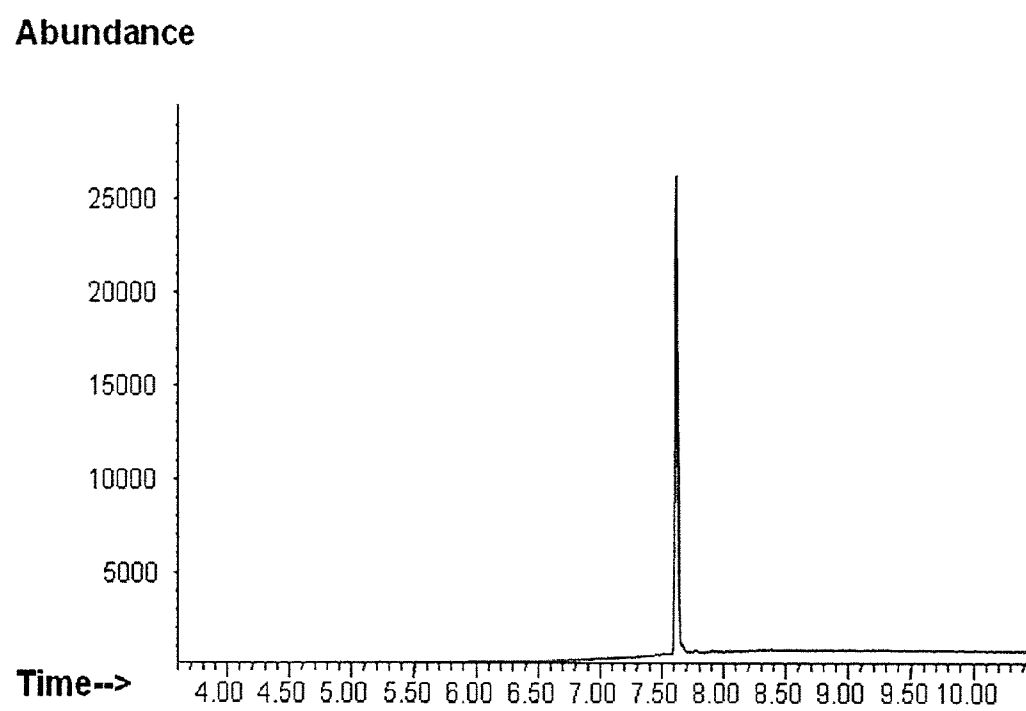
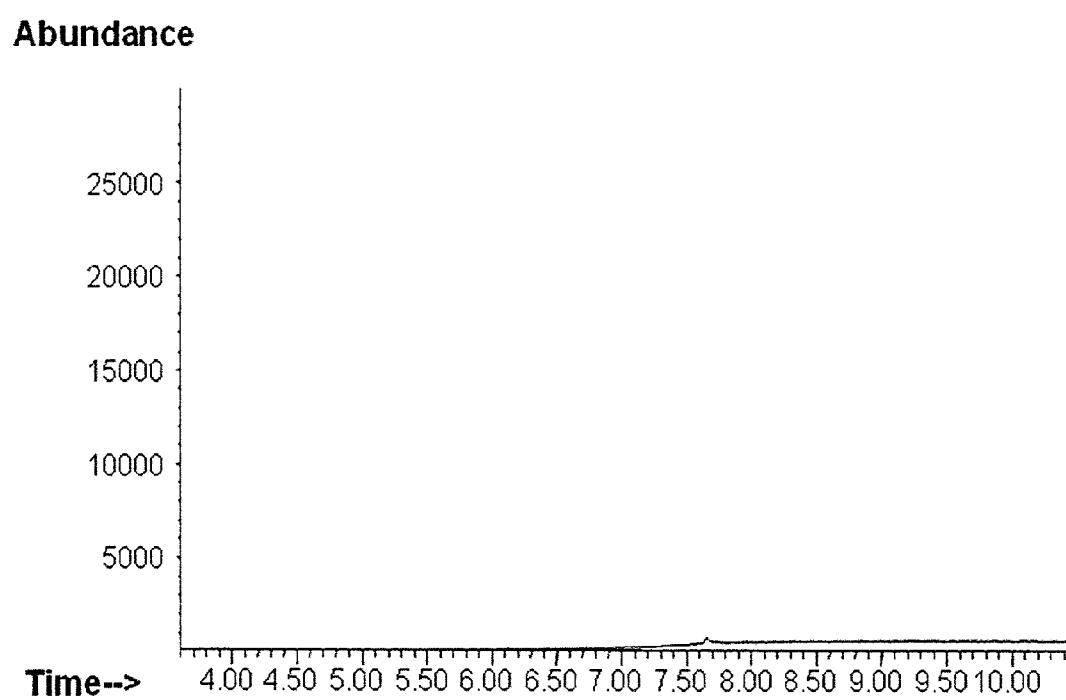


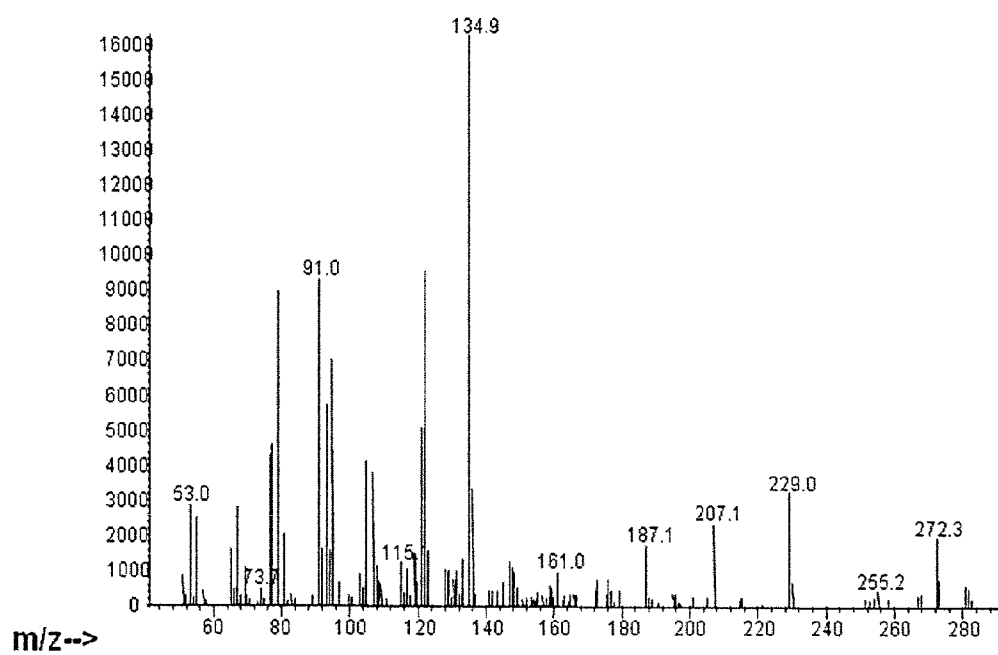
FIG. 16

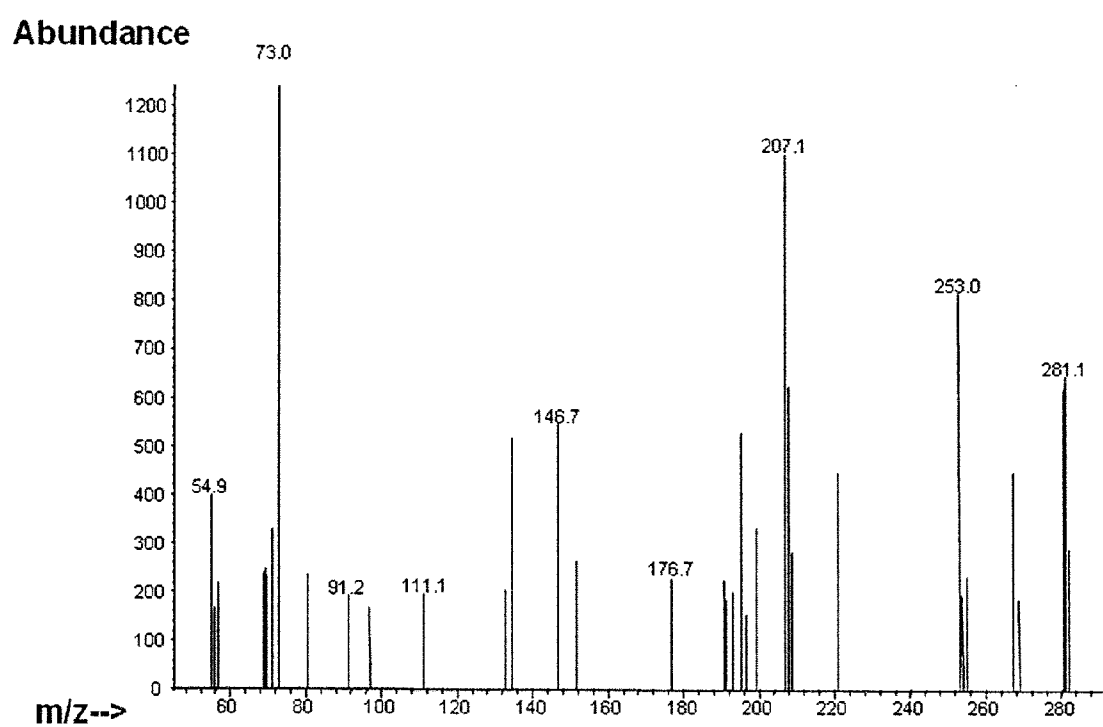
29/113*FIG. 17A*

30/113*FIG. 17B*

31/113

Abundance

*FIG. 18A*

32/113*FIG. 18B*

33/113

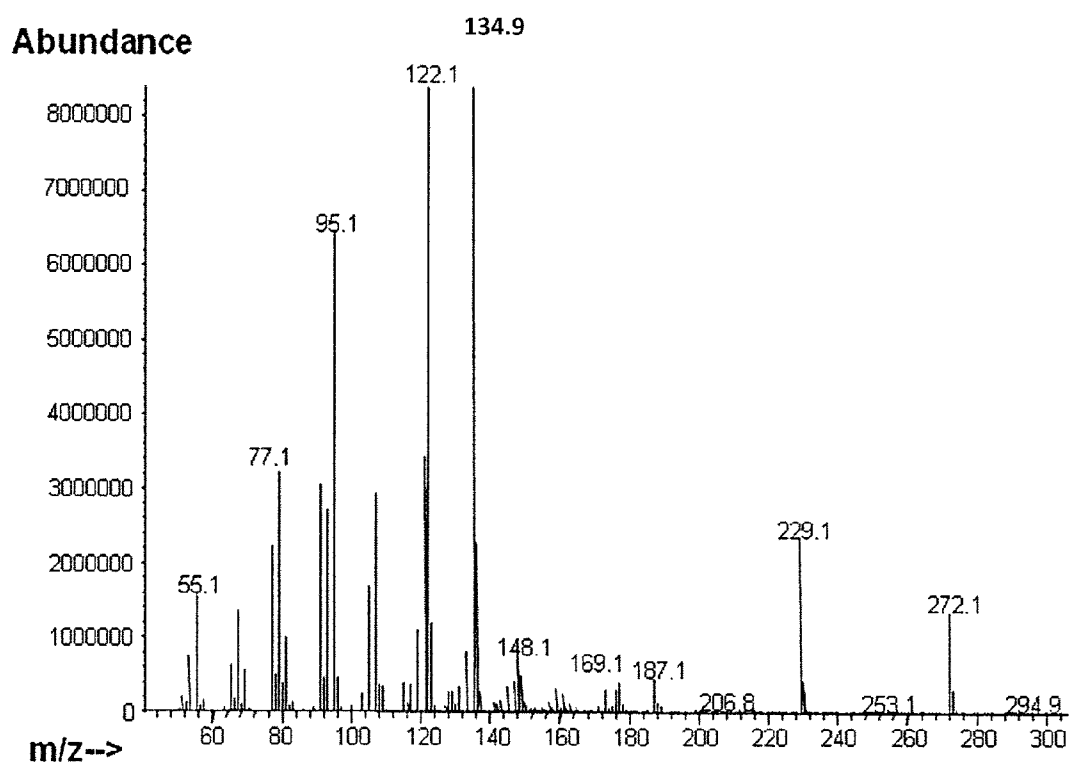


FIG. 18C

34/113

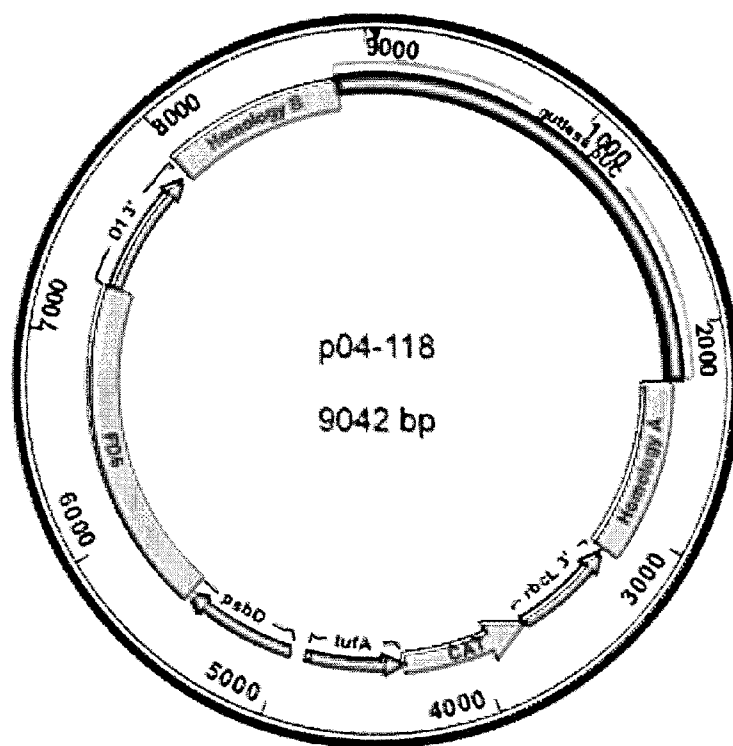


FIG. 19

35/113

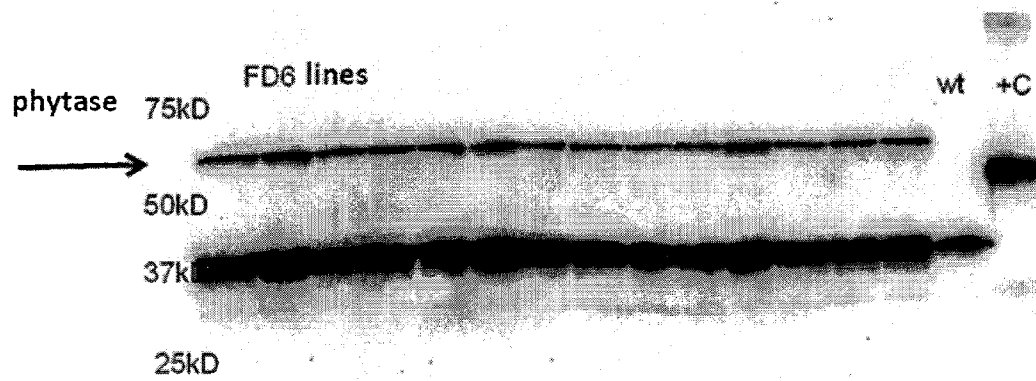


FIG. 20

36/113

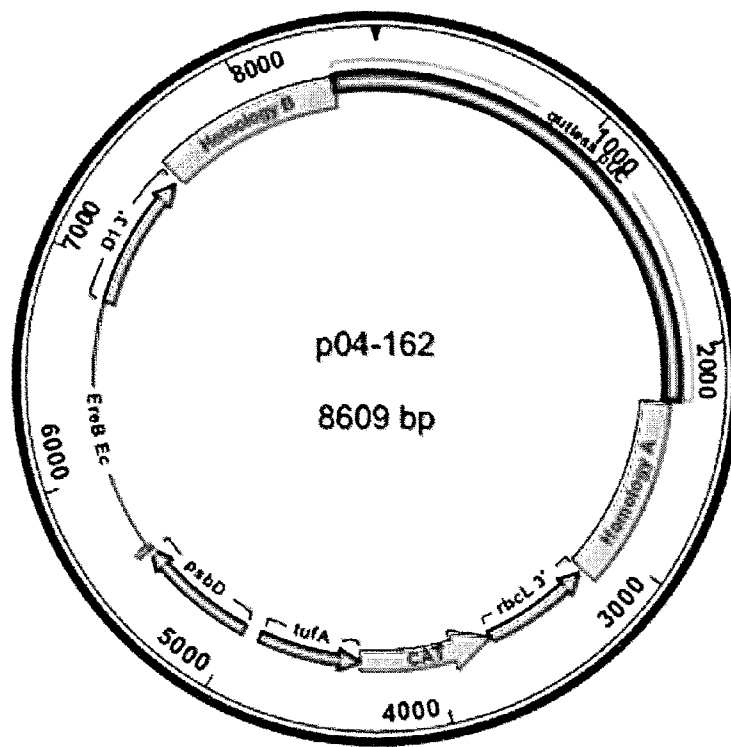


FIG. 21

37/113

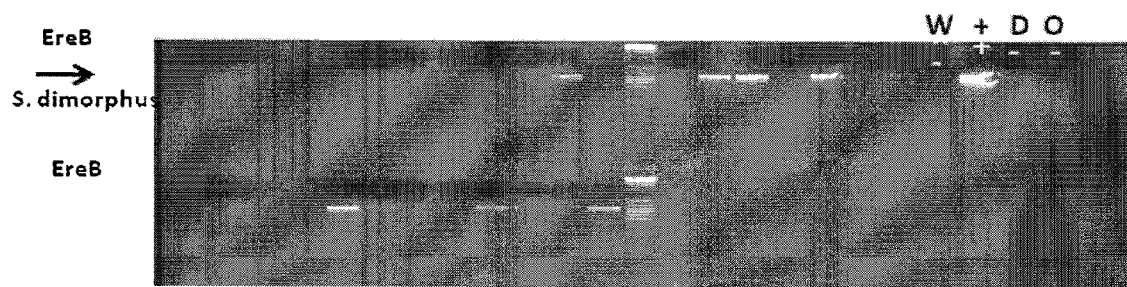


FIG. 22

38/113

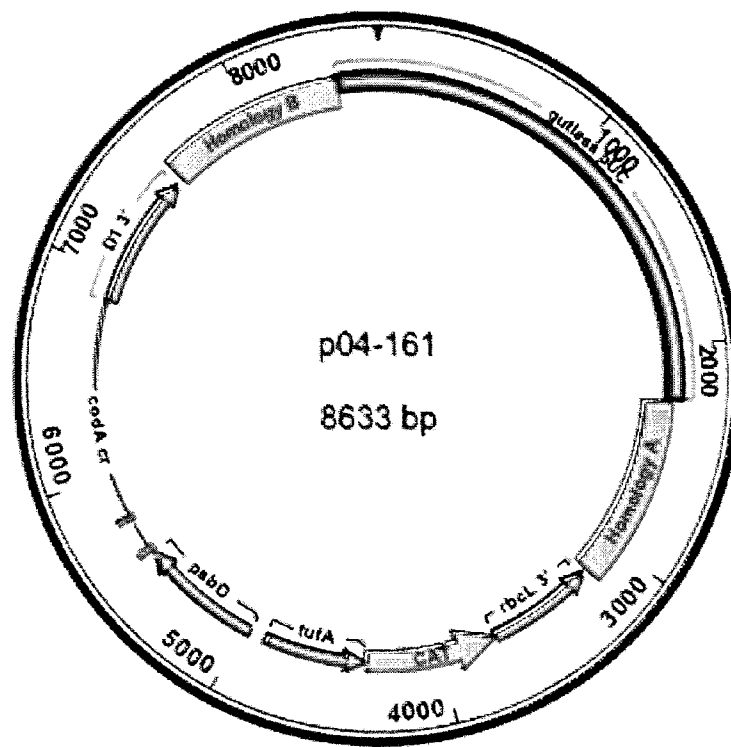


FIG. 23

39/113

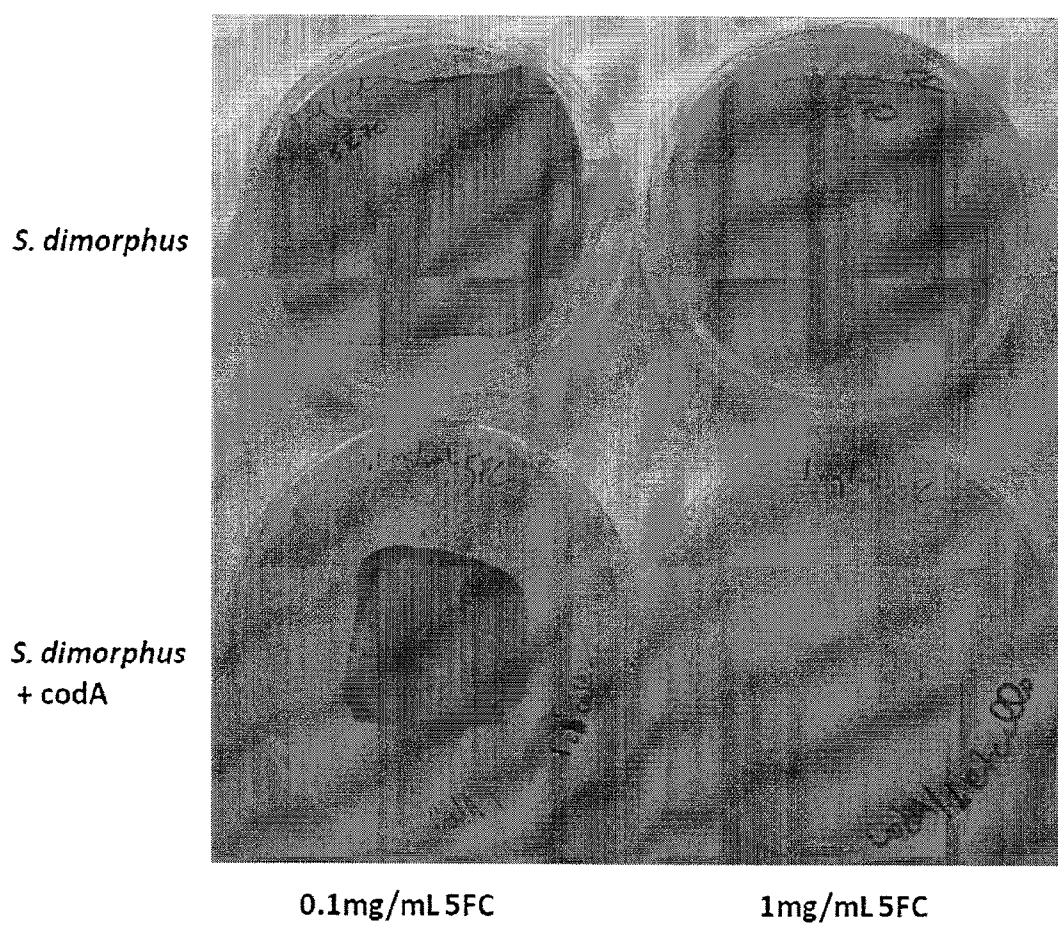


FIG. 24

41/113



1 2 3 1 2 3 1 2 3 1 2 3

FIG. 26

42/113

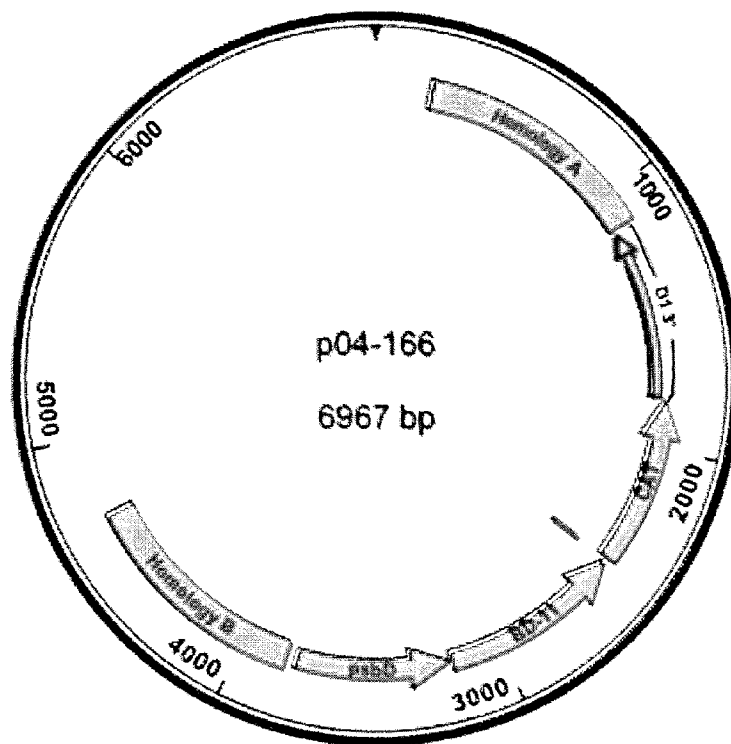


FIG. 27

43/113

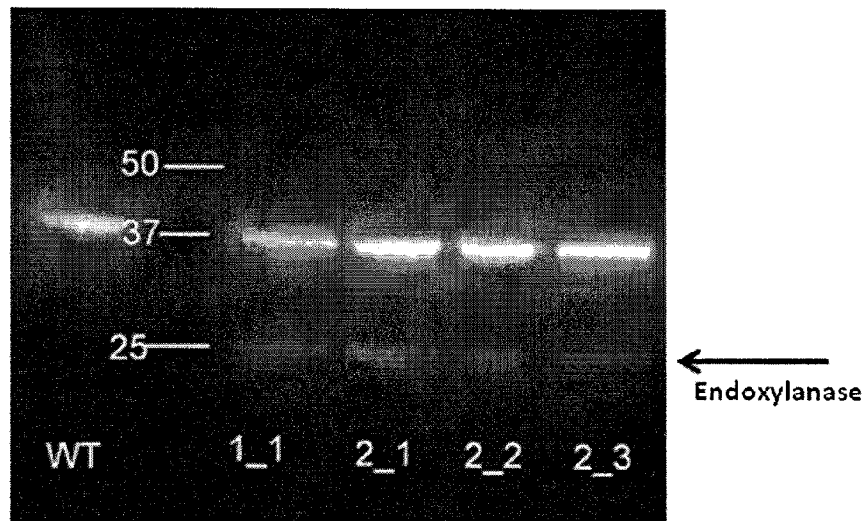


FIG. 28

44/113

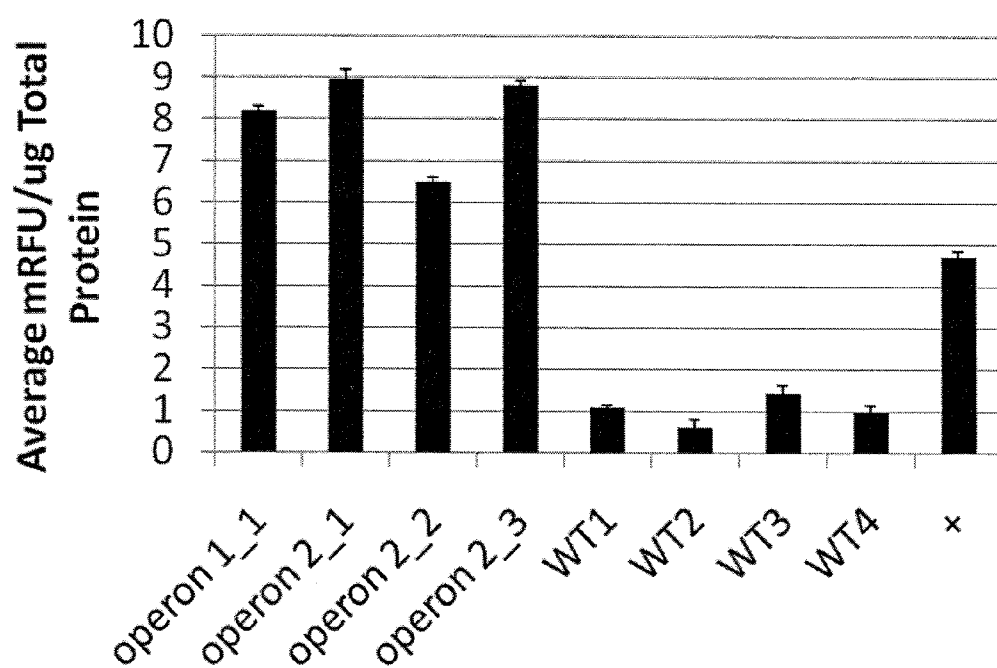


FIG. 29

45/113

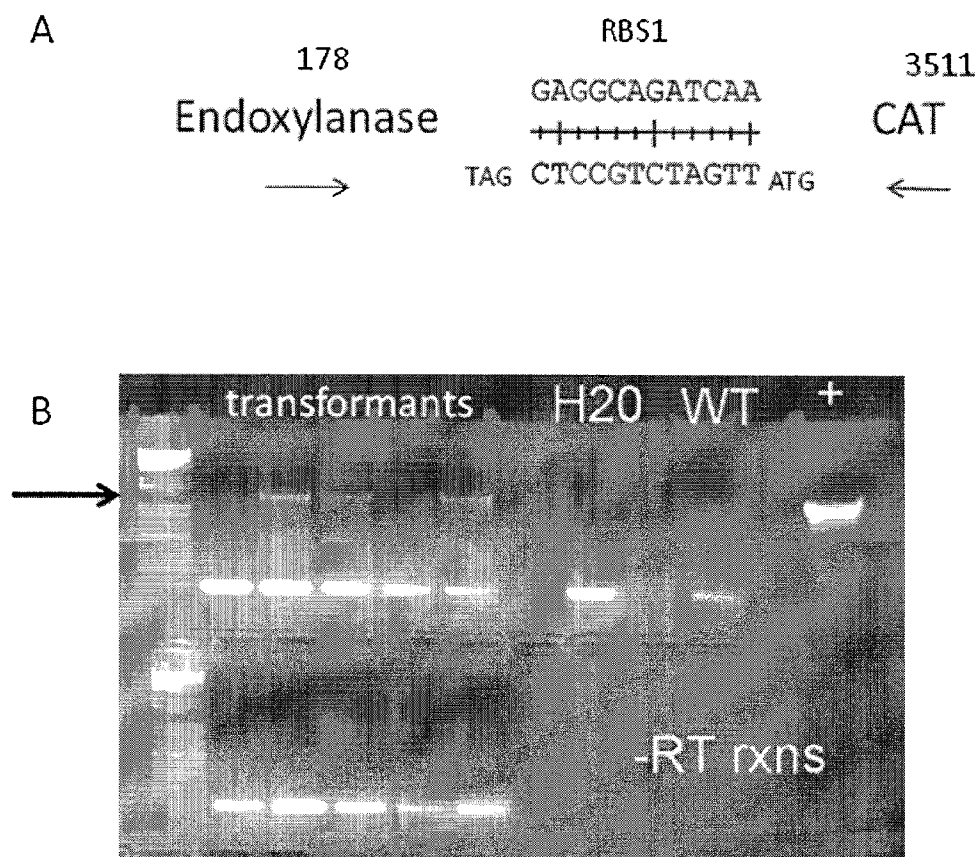


FIG. 30

46/113

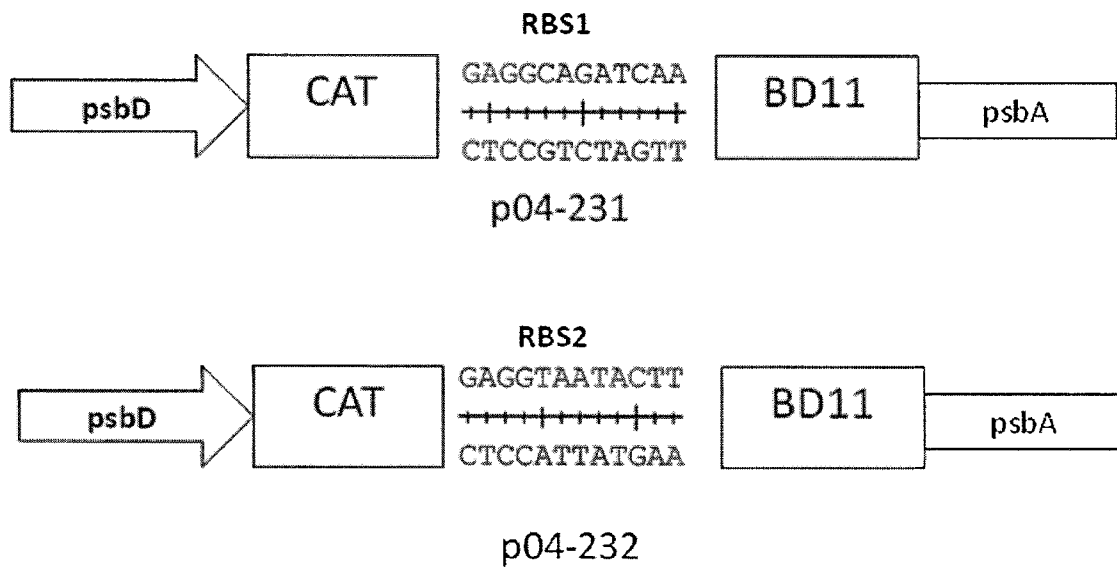


FIG. 31

47/113

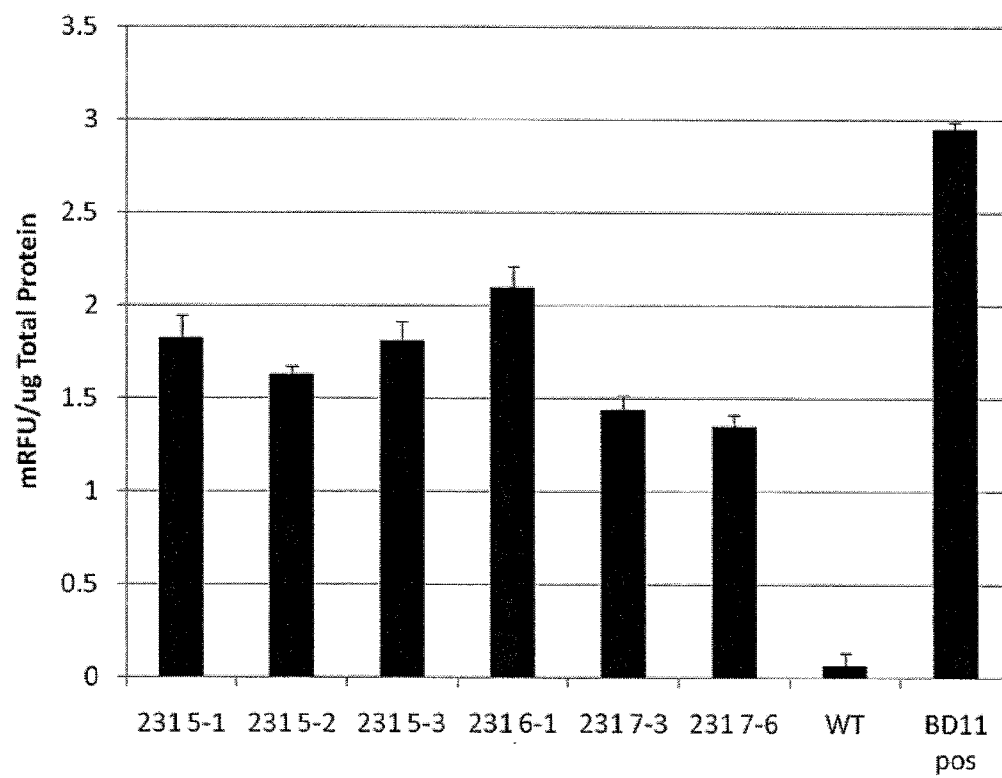


FIG. 32A

48/113

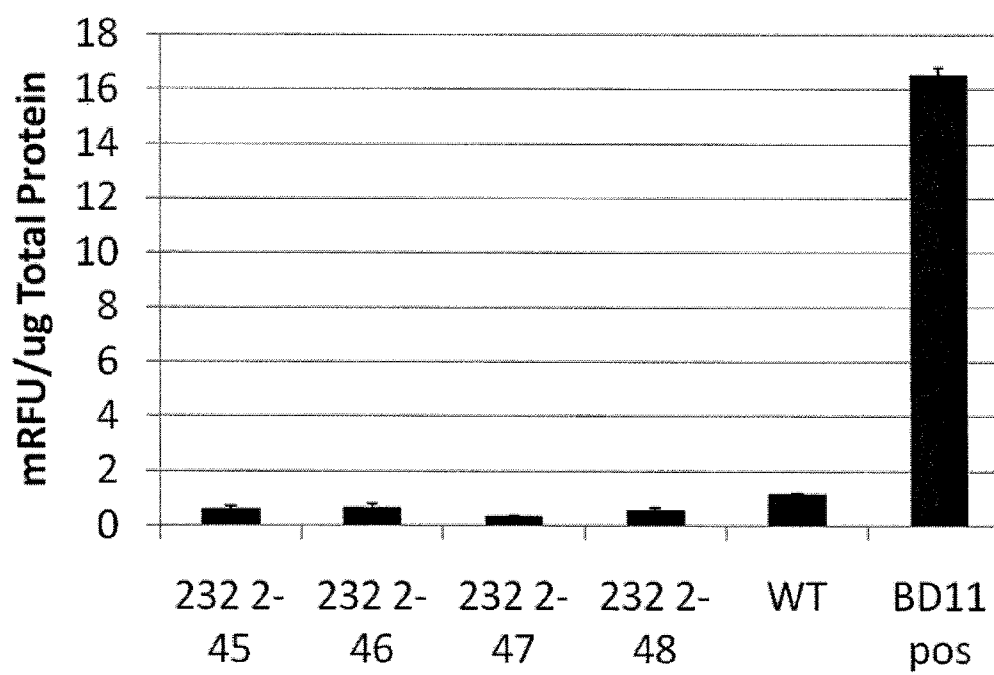


FIG. 32B

49/113



FIG. 33

50/113

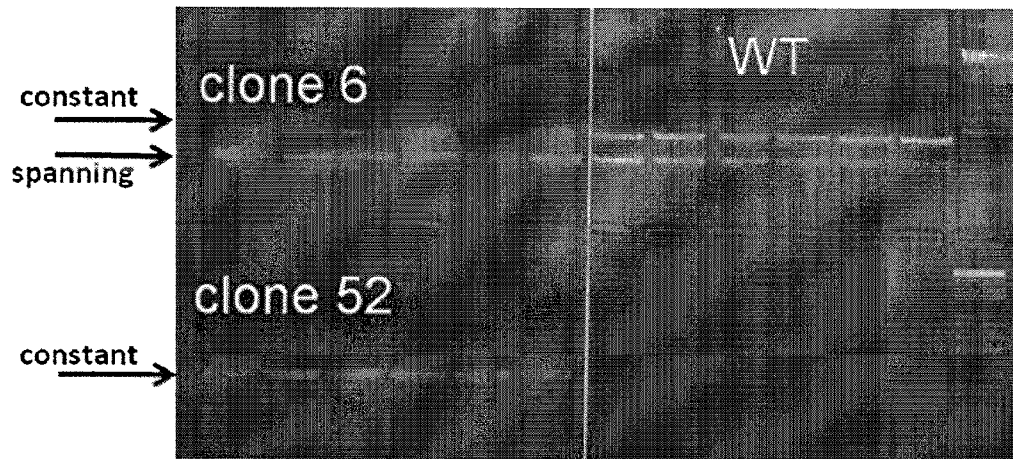


FIG. 34

51/113

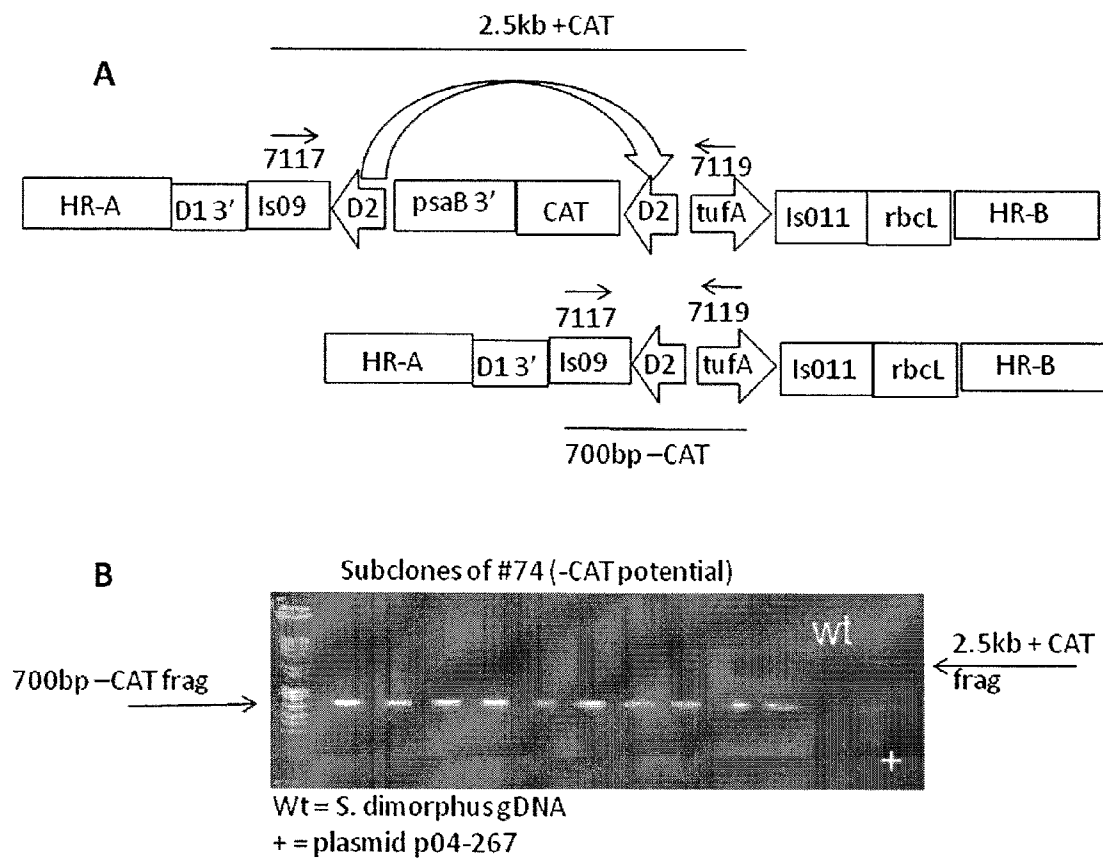


FIG. 35

52/113

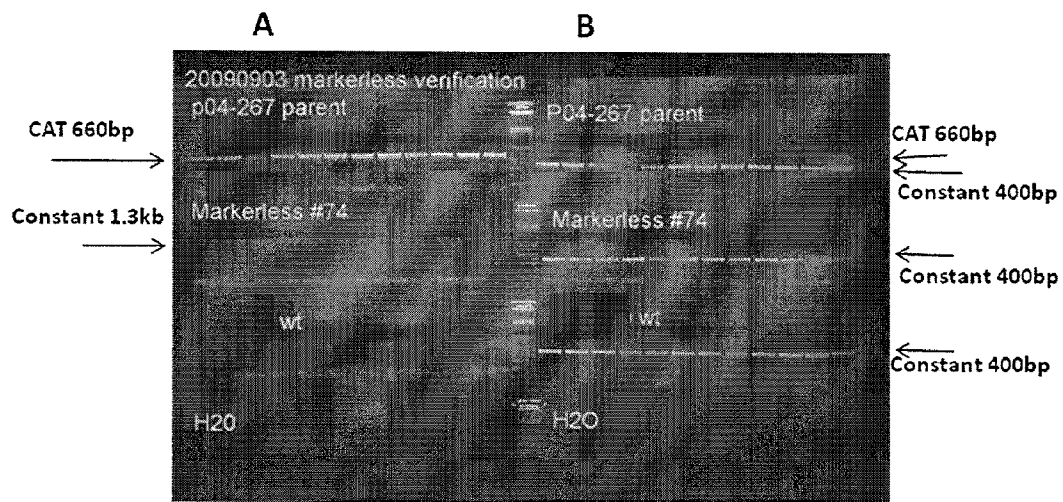


FIG. 36

53/113

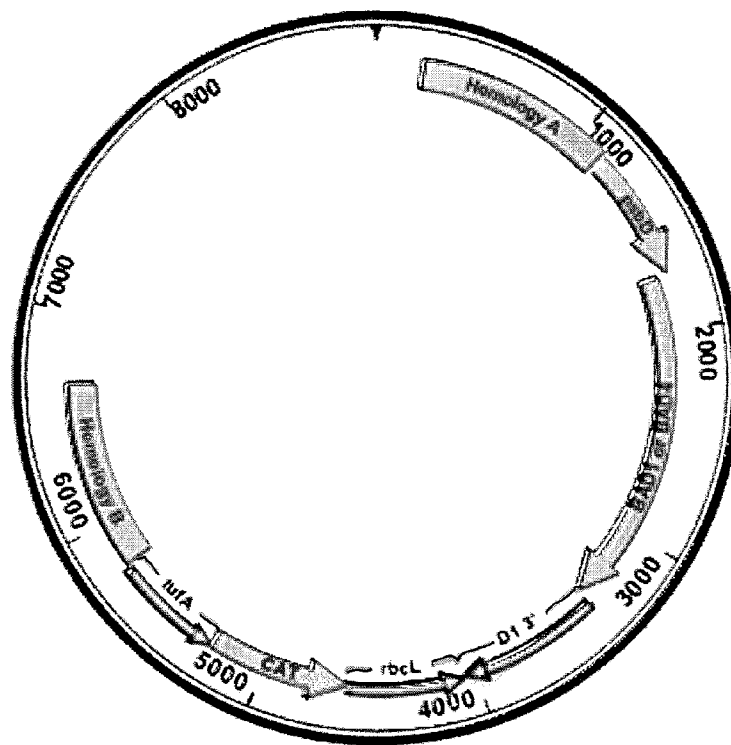


FIG. 37

54/113

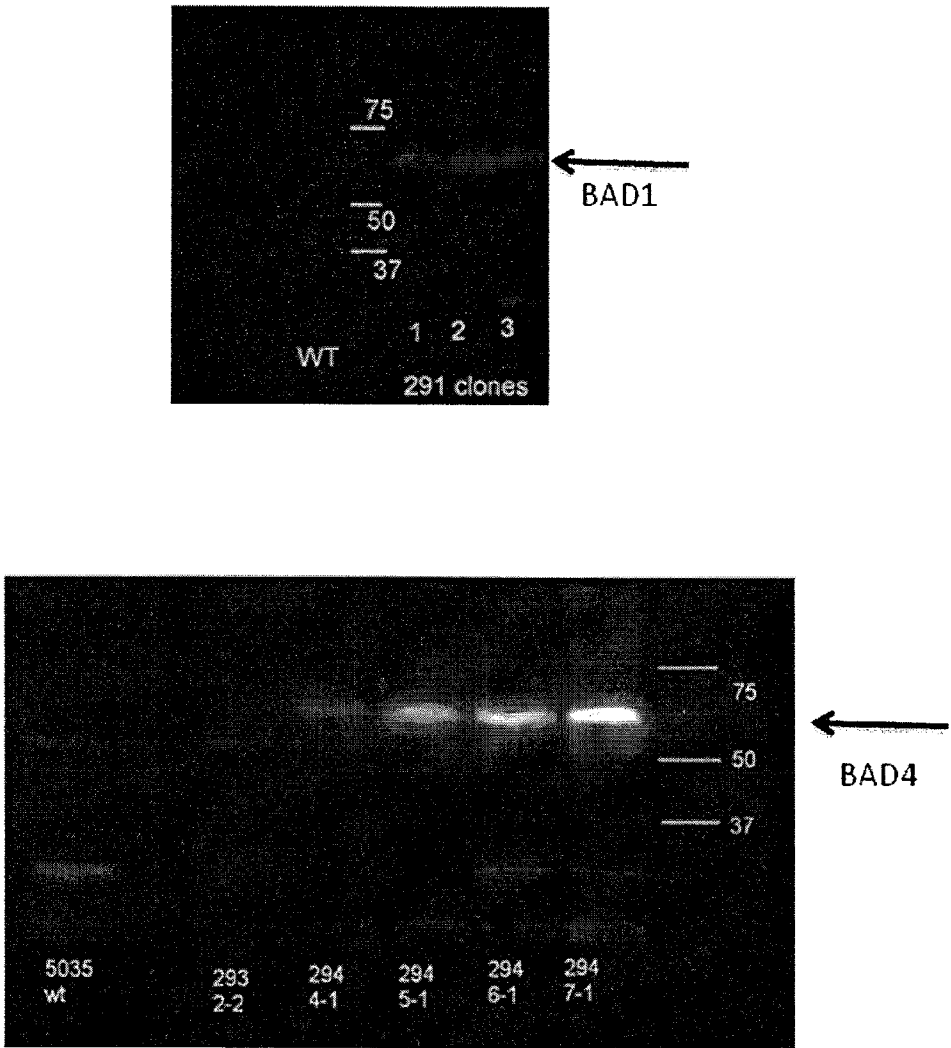


FIG. 38

55/113

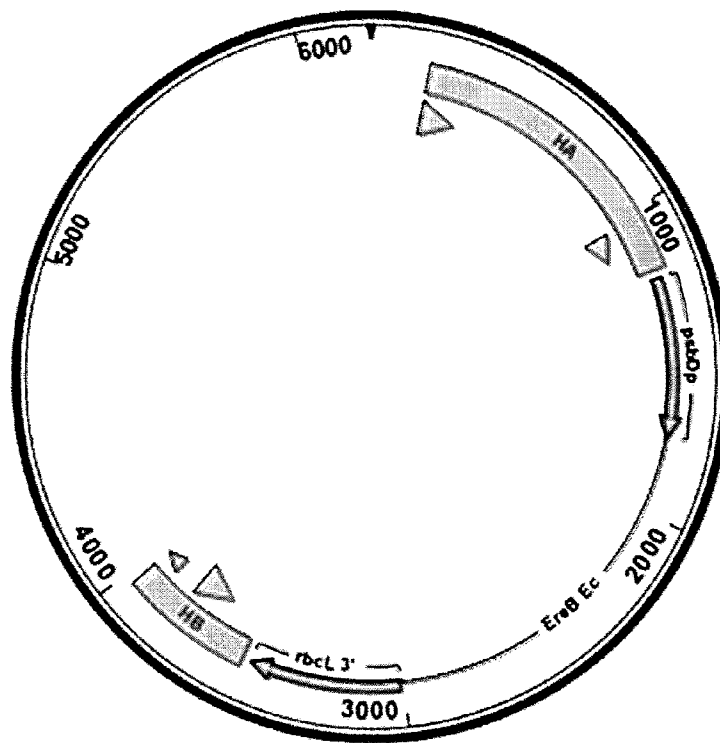


FIG. 39

56/113

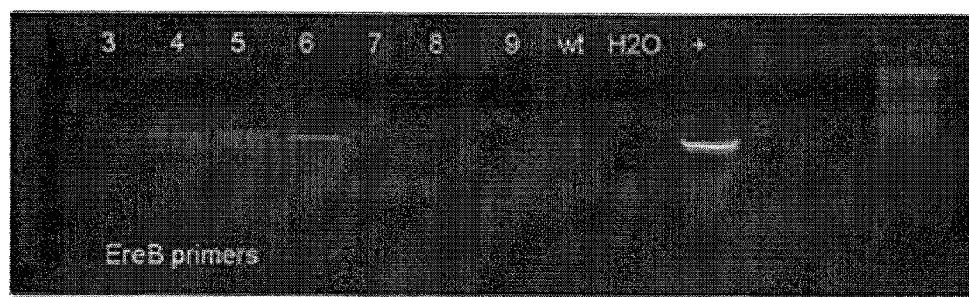


FIG. 40

57/113

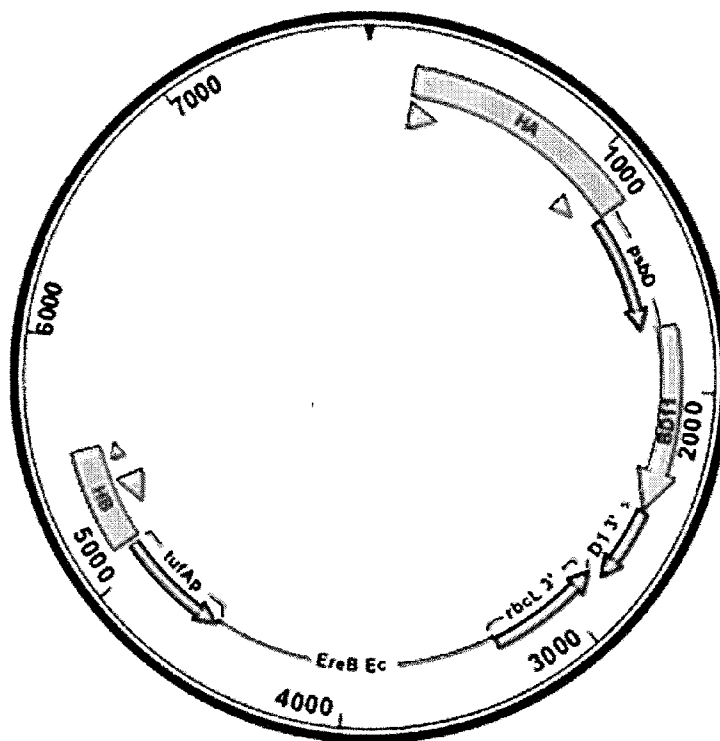


FIG. 41

58/113

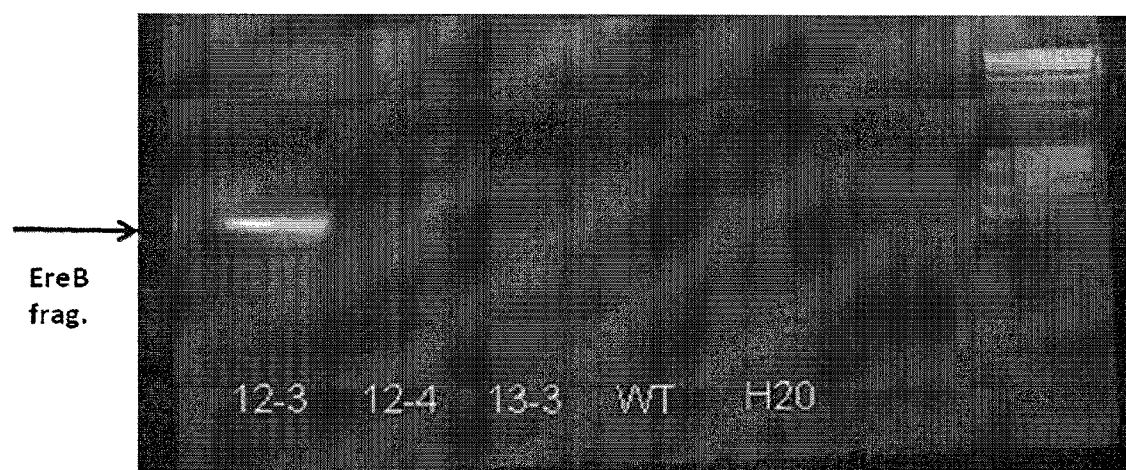


FIG. 42

59/113

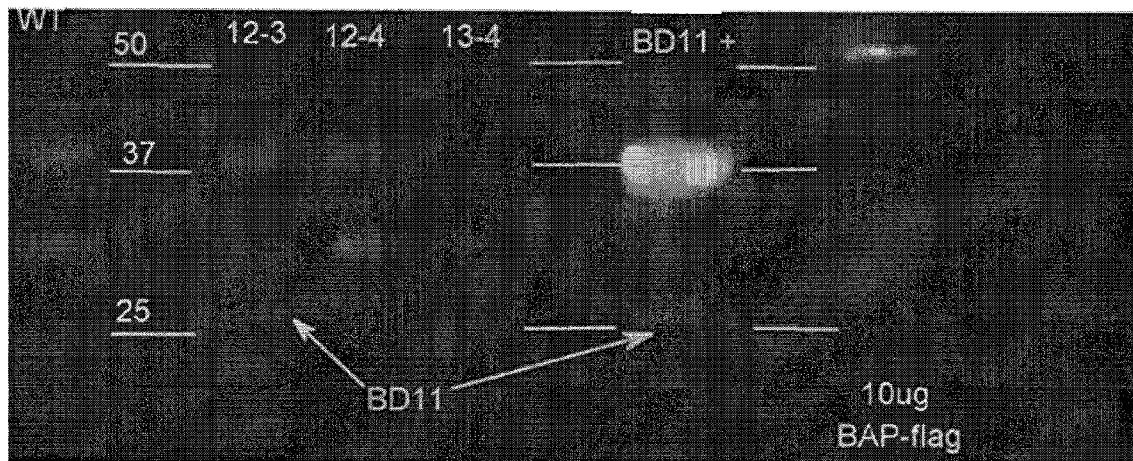
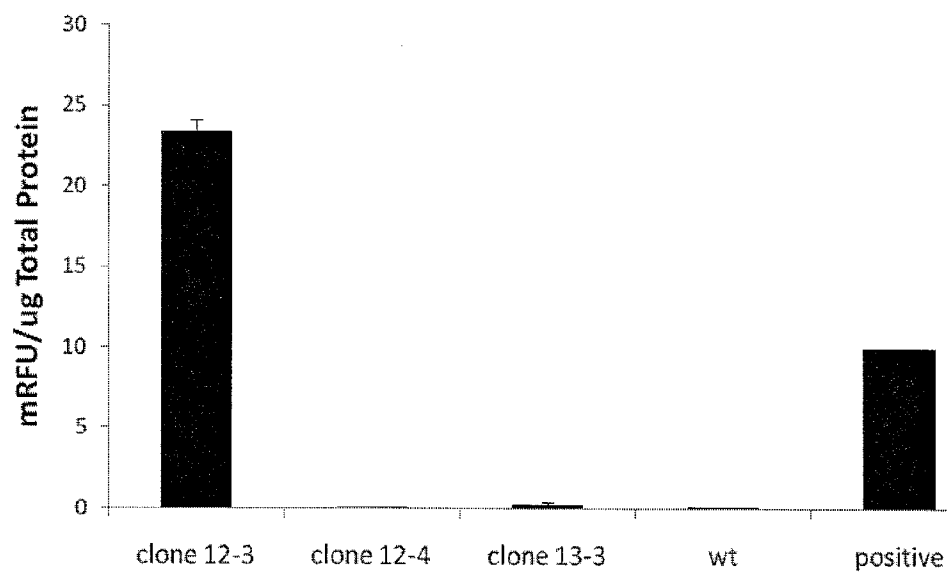


FIG. 43

60/113*FIG. 44*

61/113



FIG. 45

62/113

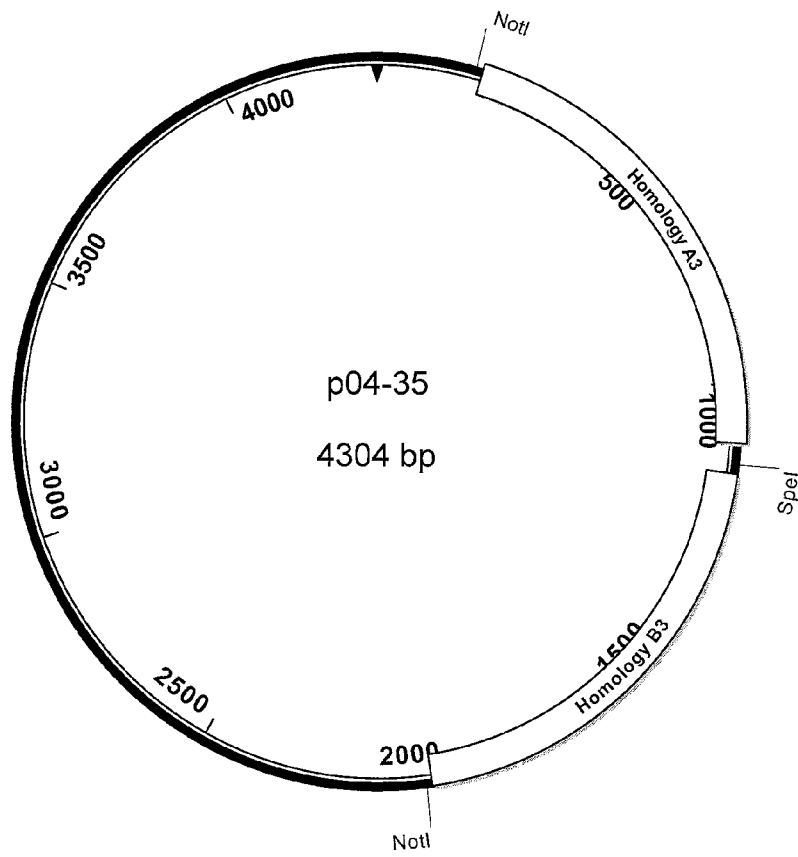


FIG. 46

63/113

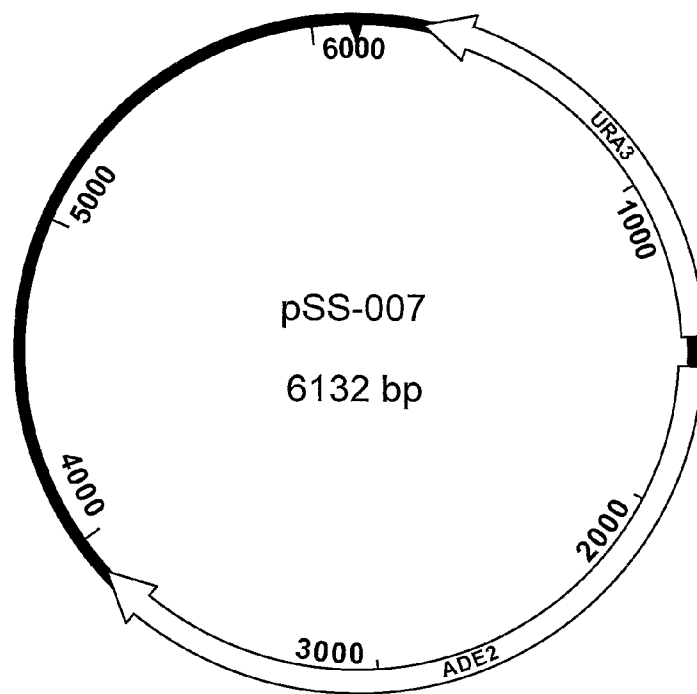


FIG. 47

64/113

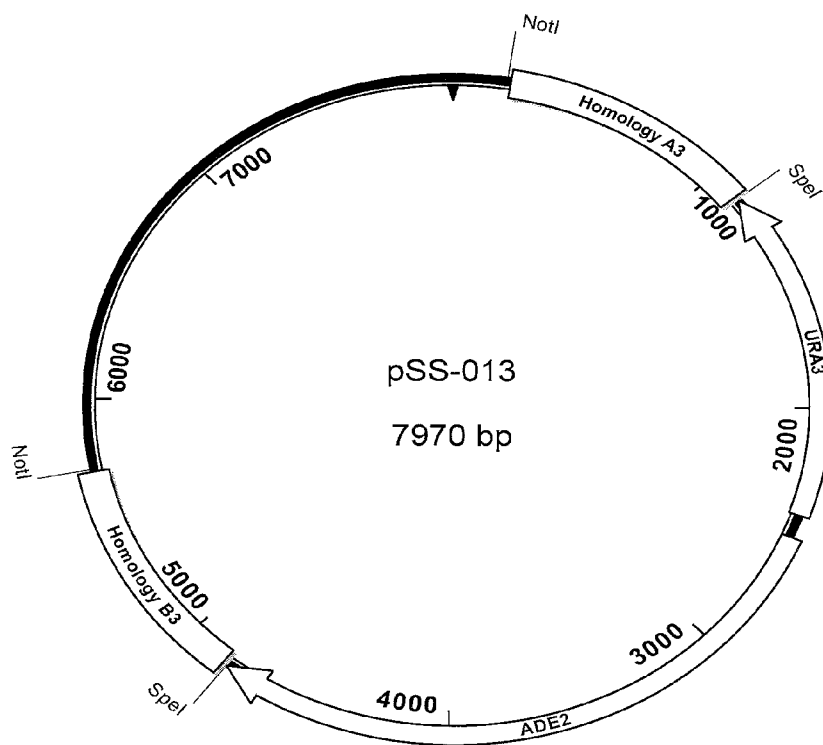


FIG. 48

65/113

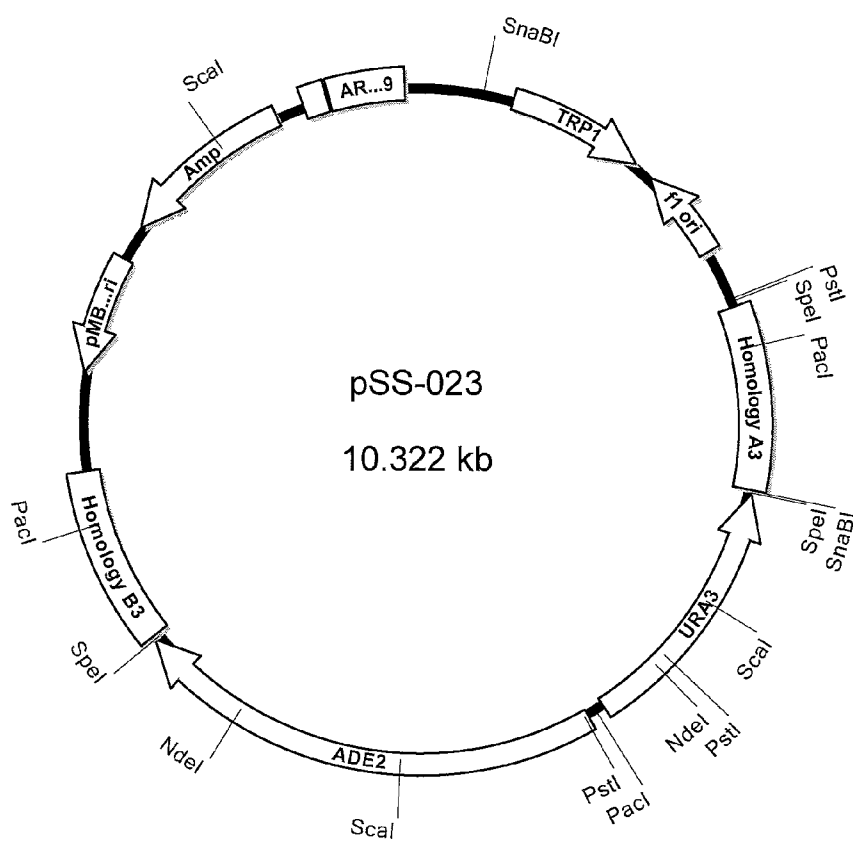


FIG. 49

66/113

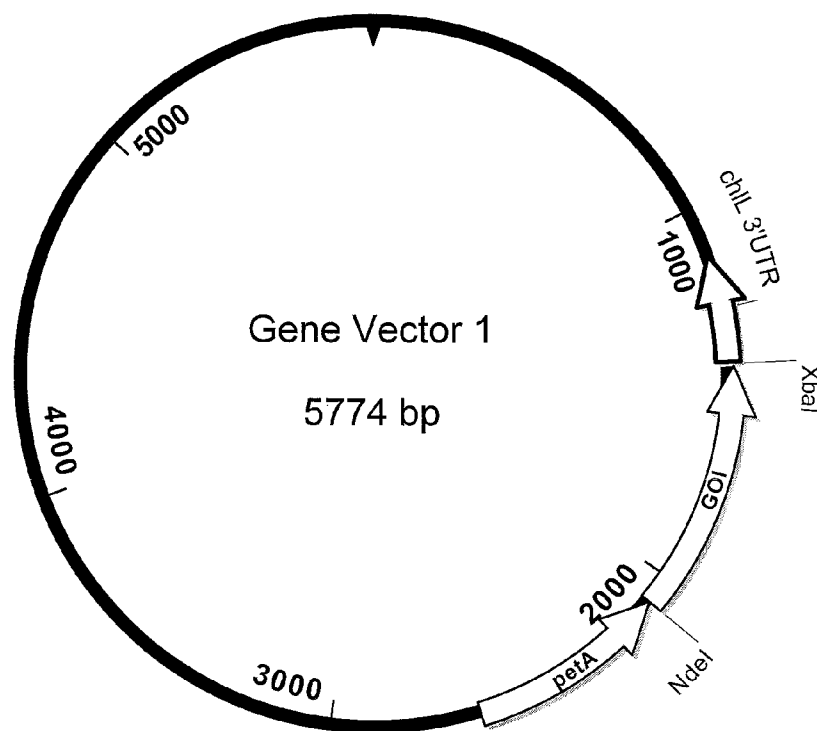


FIG. 50

67/113

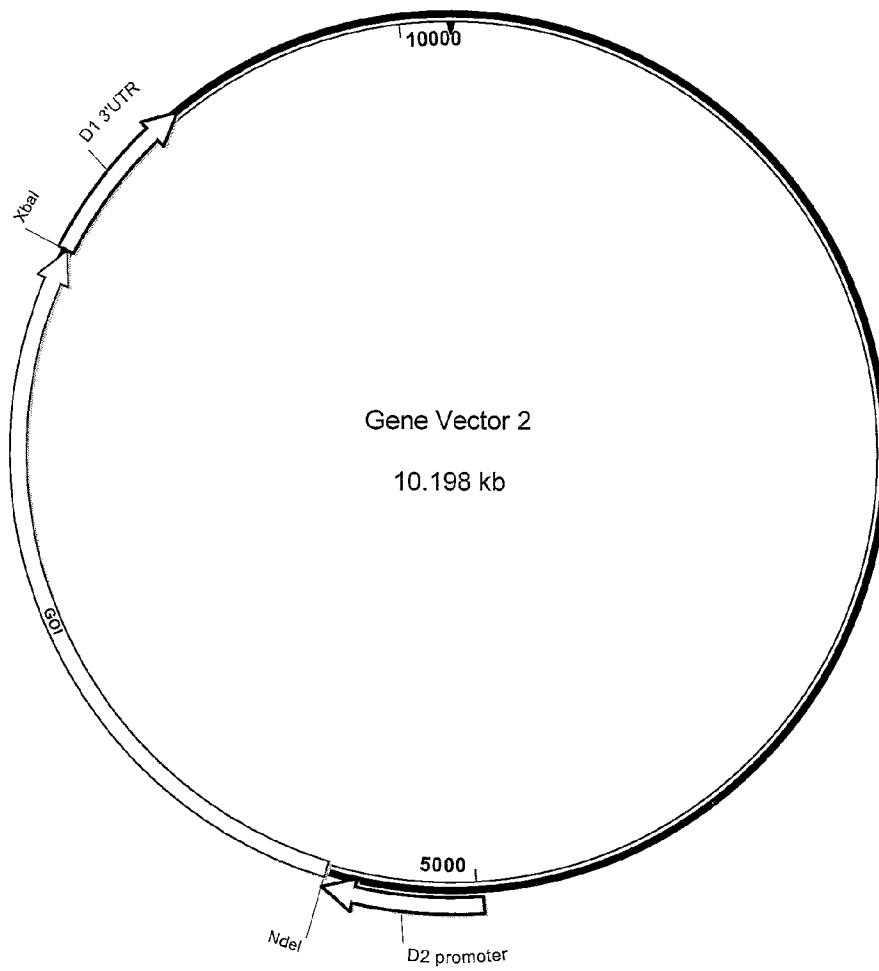


FIG. 51

68/113

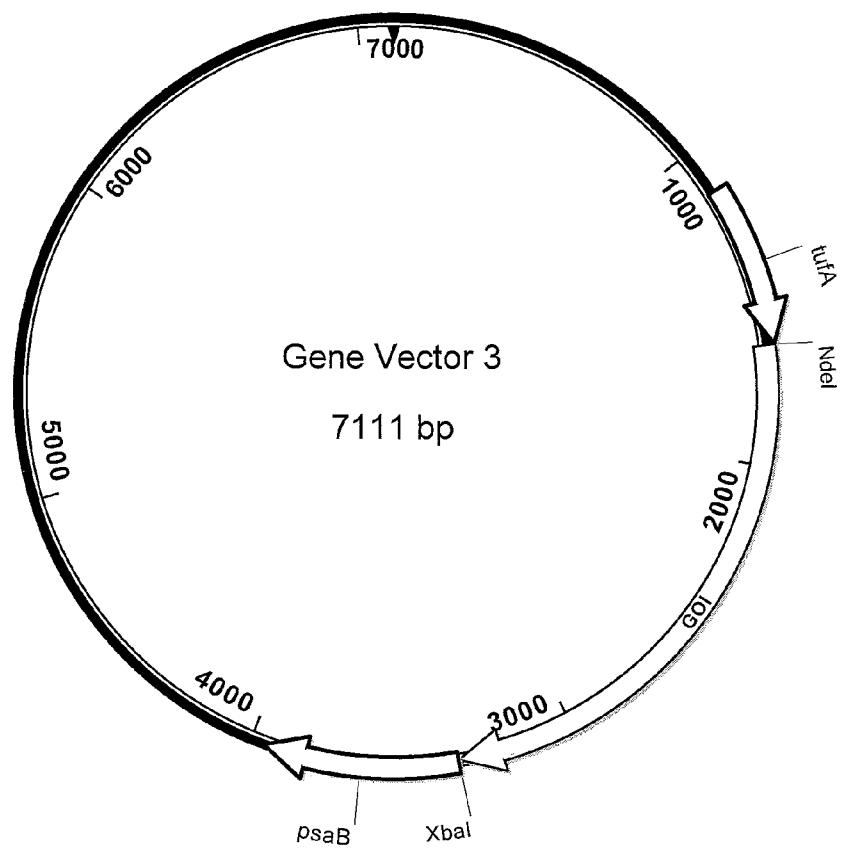


FIG. 52

69/113

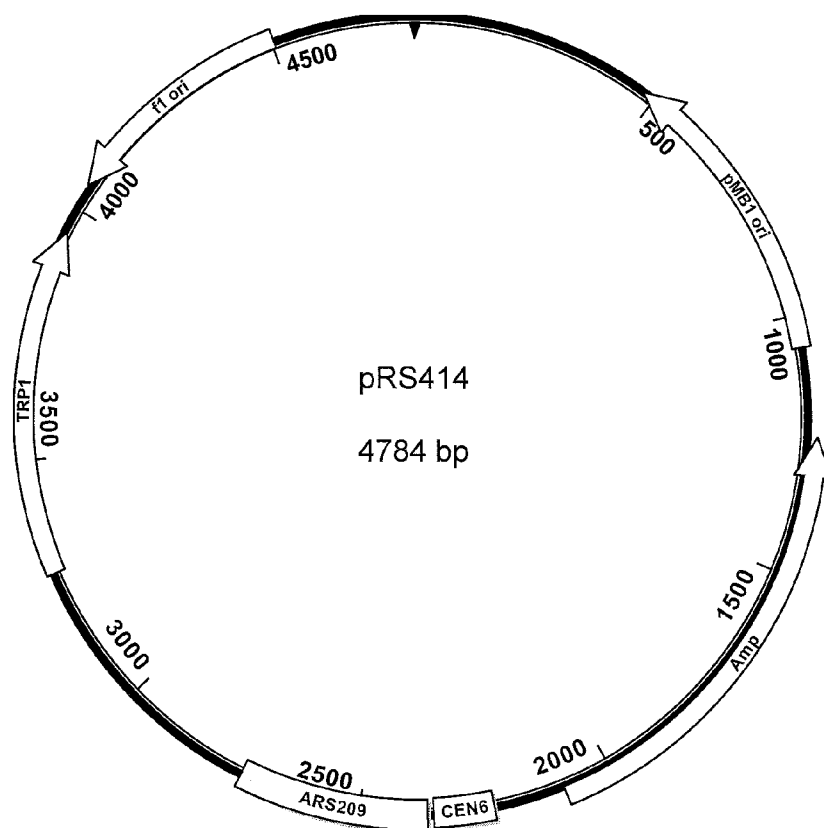


FIG. 53

70/113

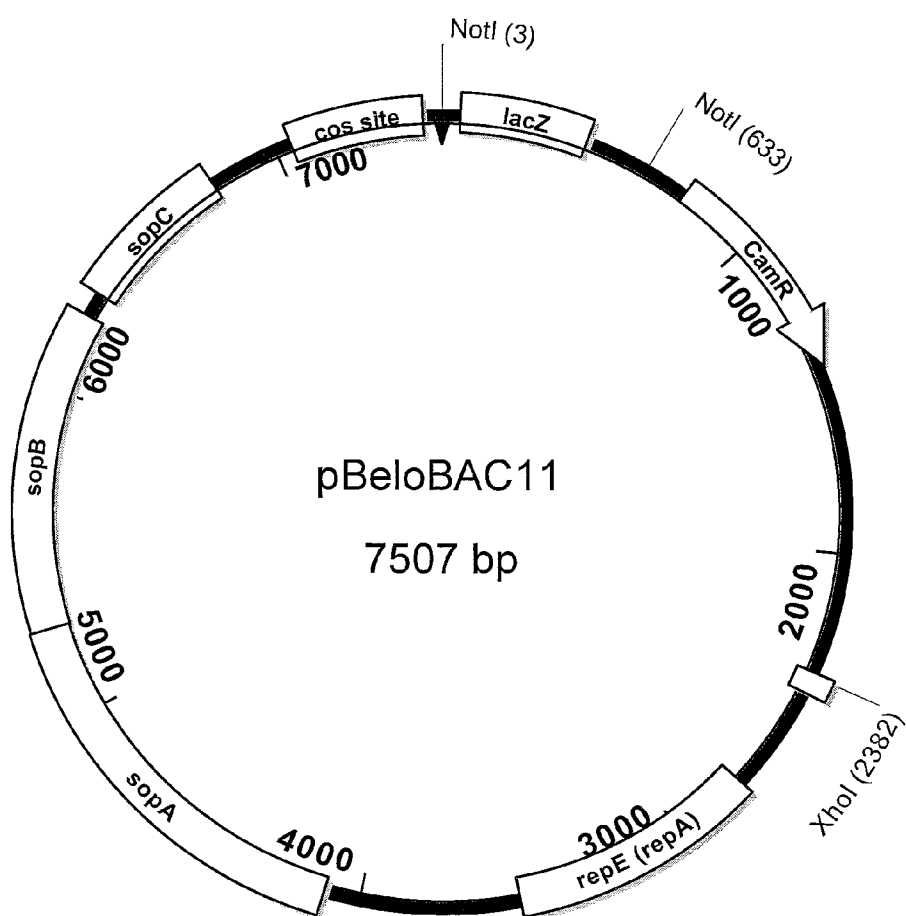


FIG. 54

71/113

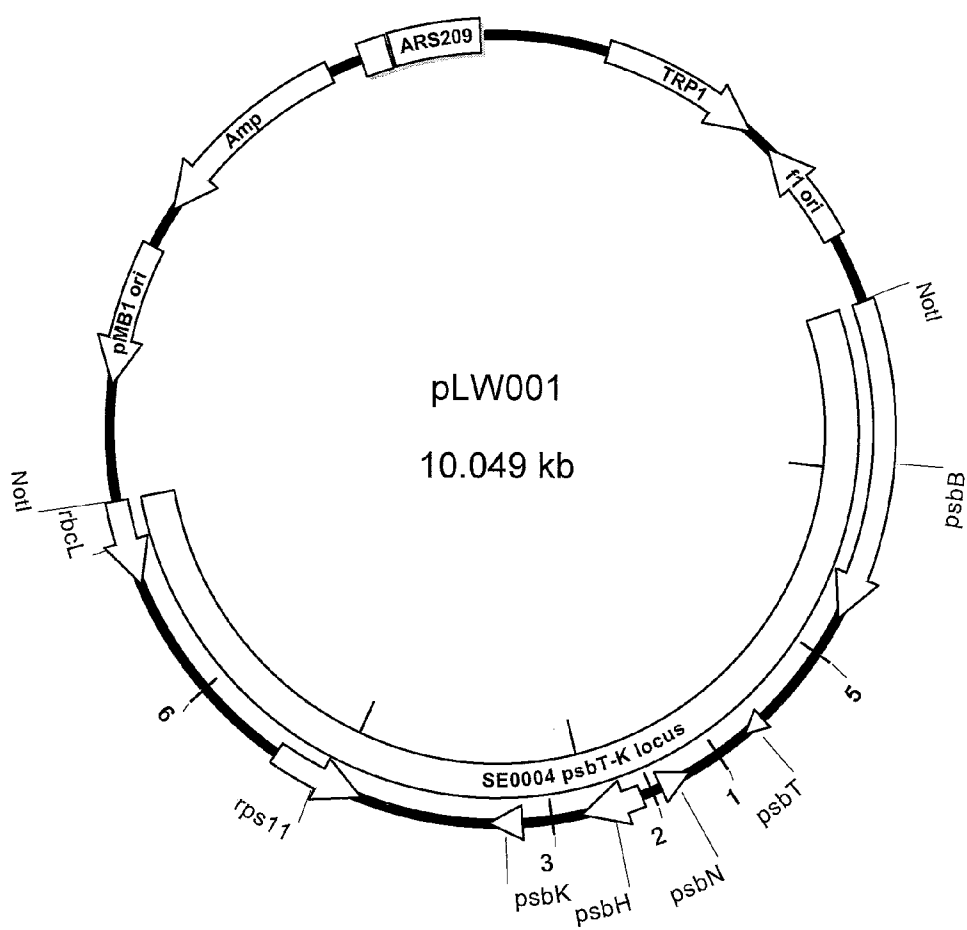


FIG. 55

72/113

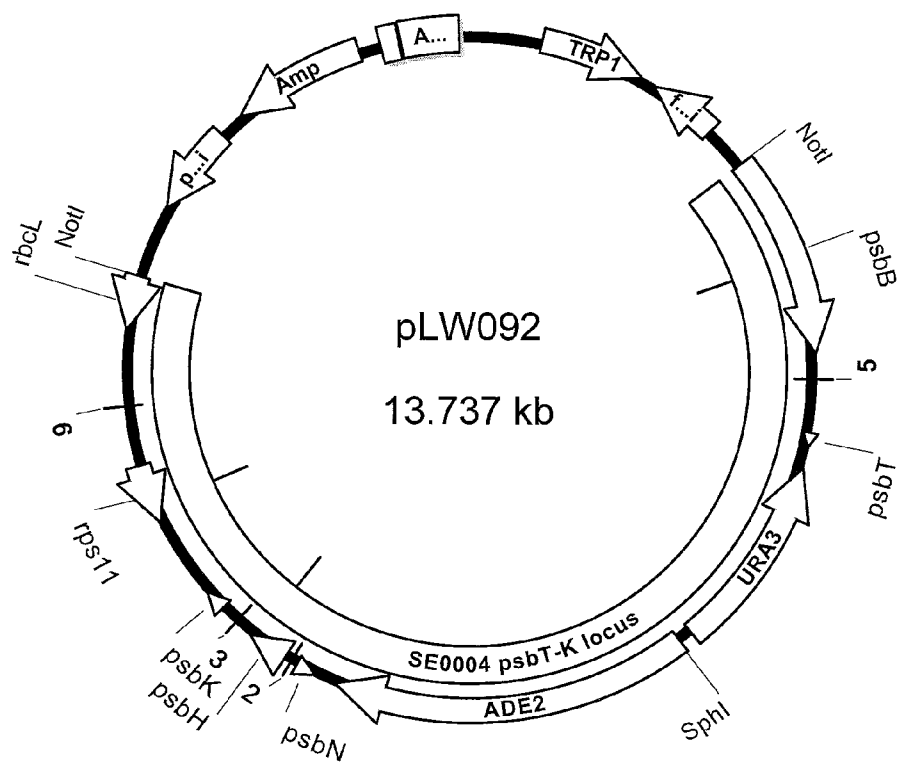


FIG. 56

73/113

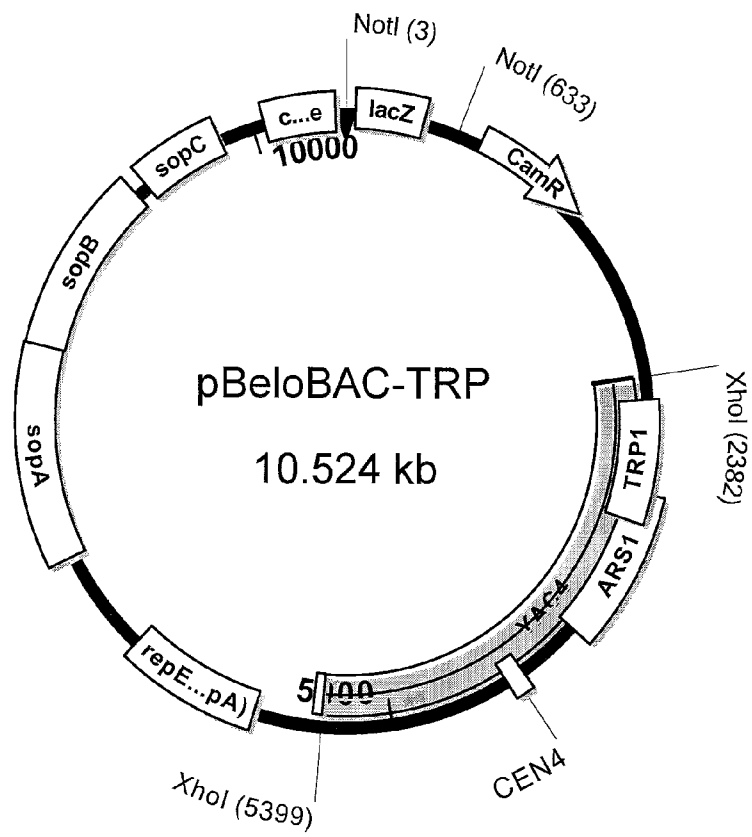


FIG. 57

74/113

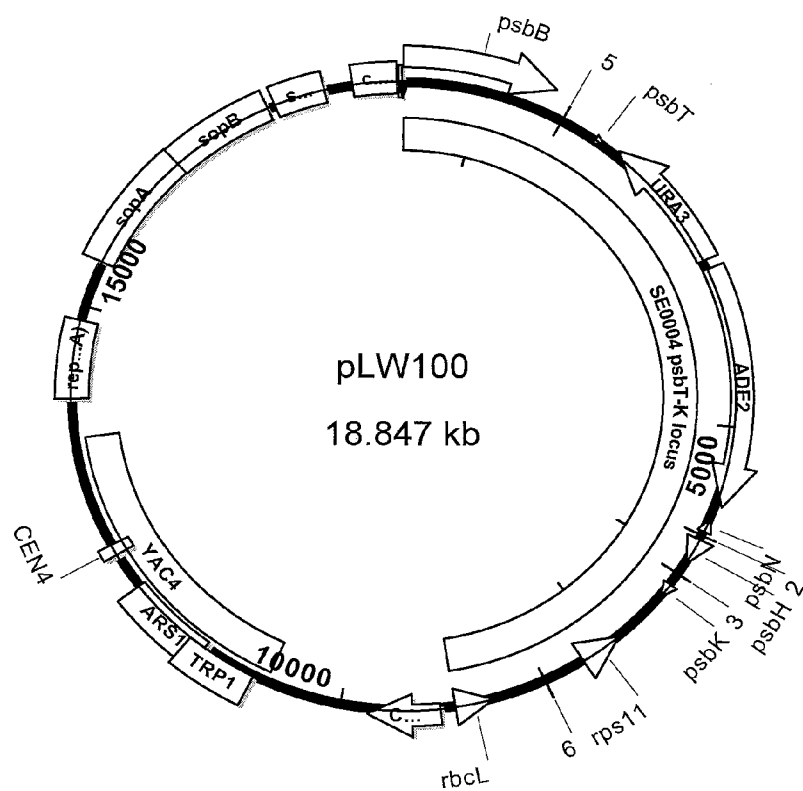


FIG. 58

75/113

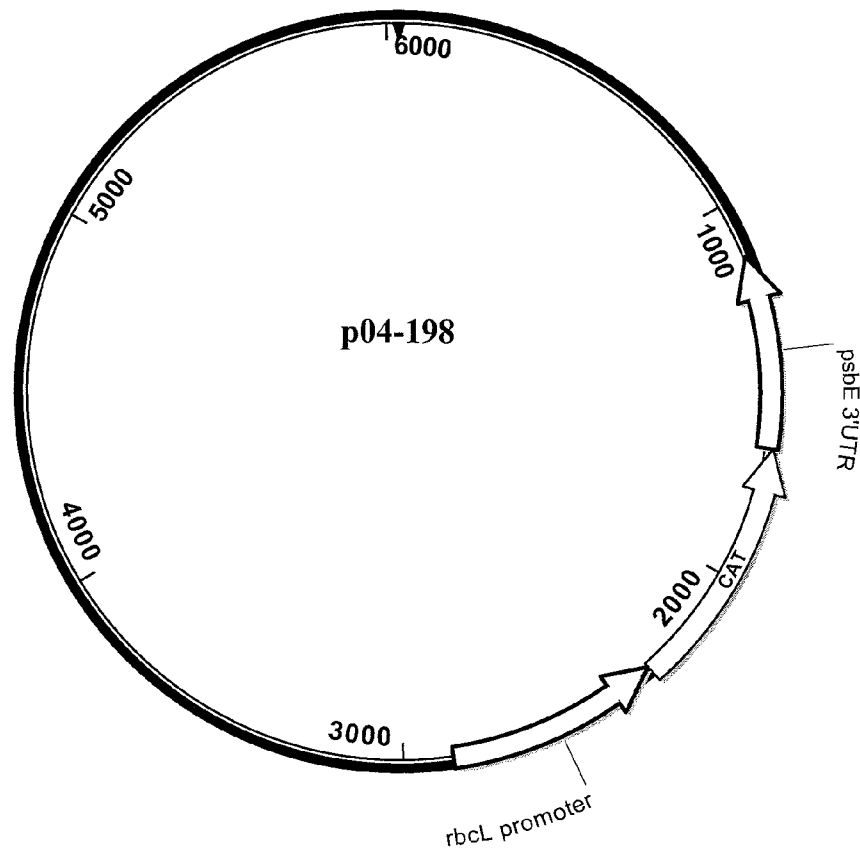


FIG. 59

76/113

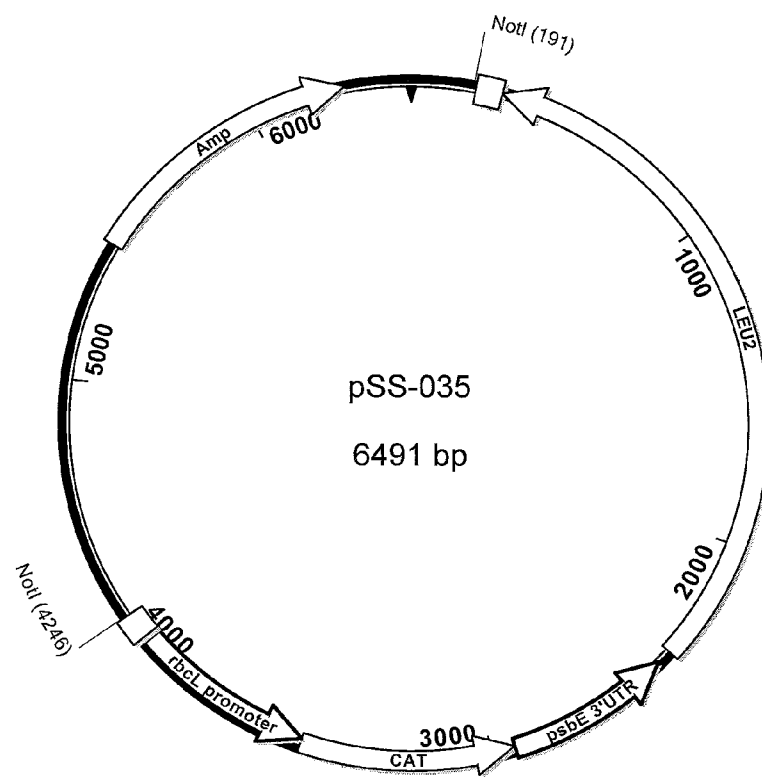


FIG. 60

77/113

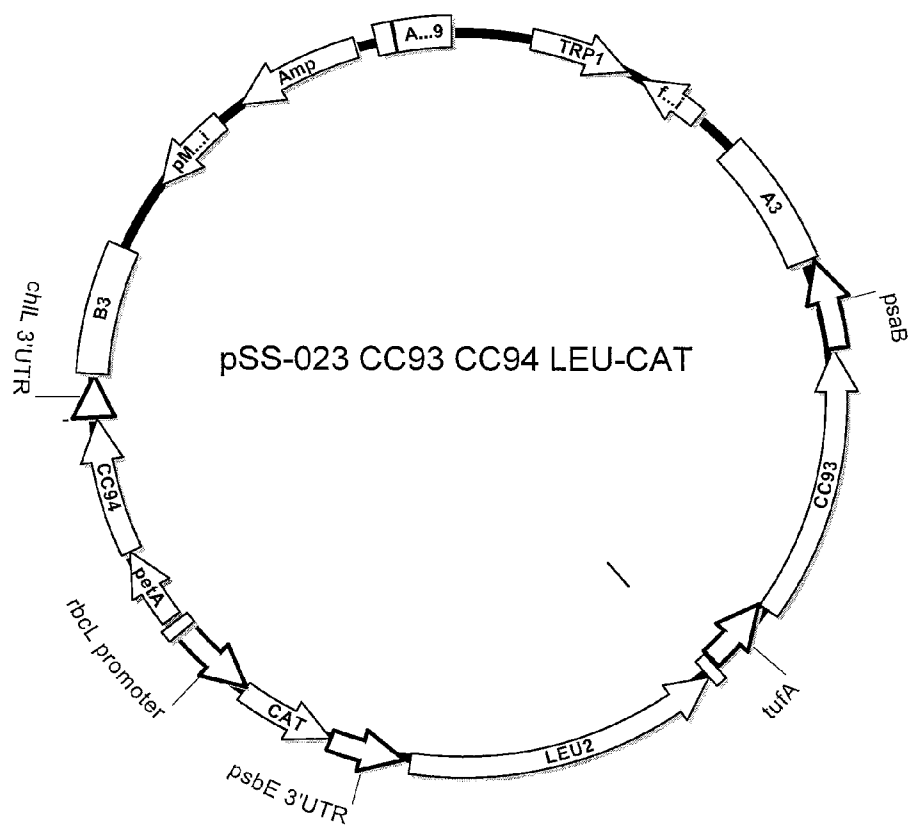


FIG. 61

78/113

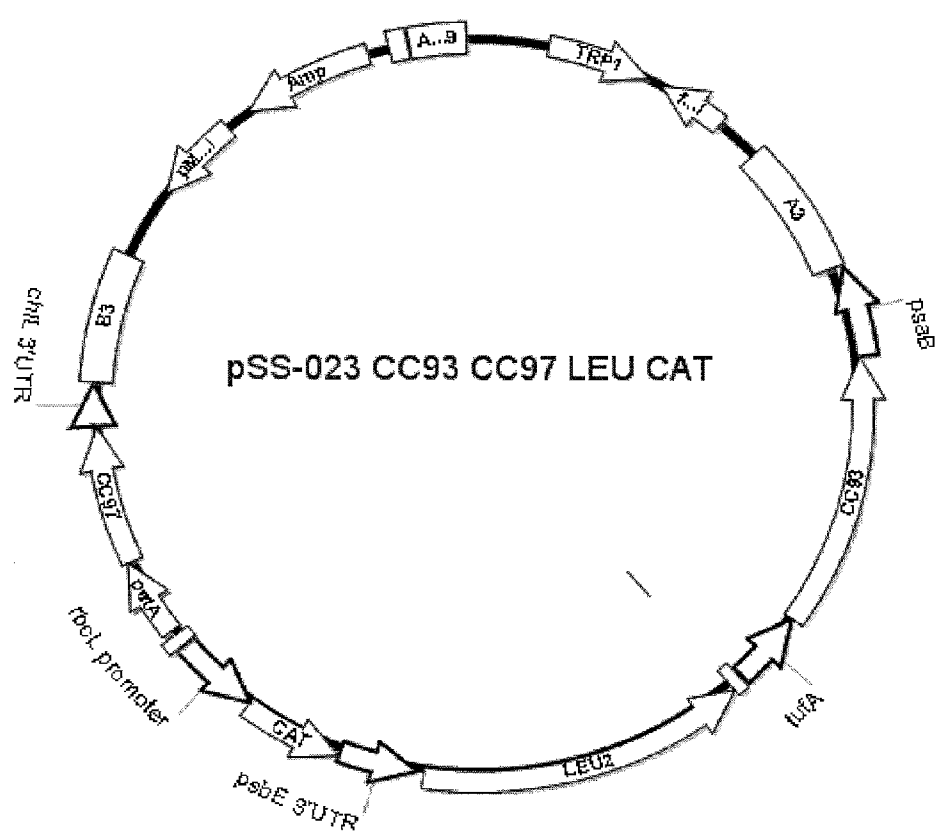


FIG. 62

79/113

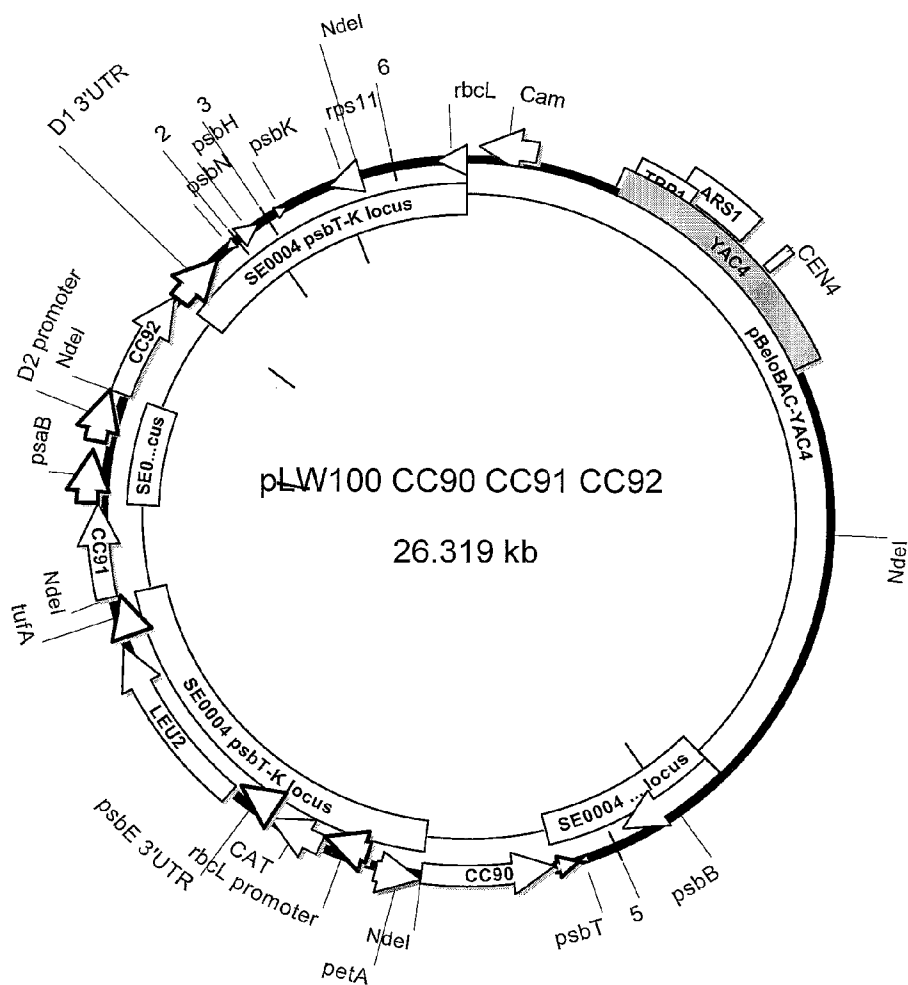


FIG. 63

80/113

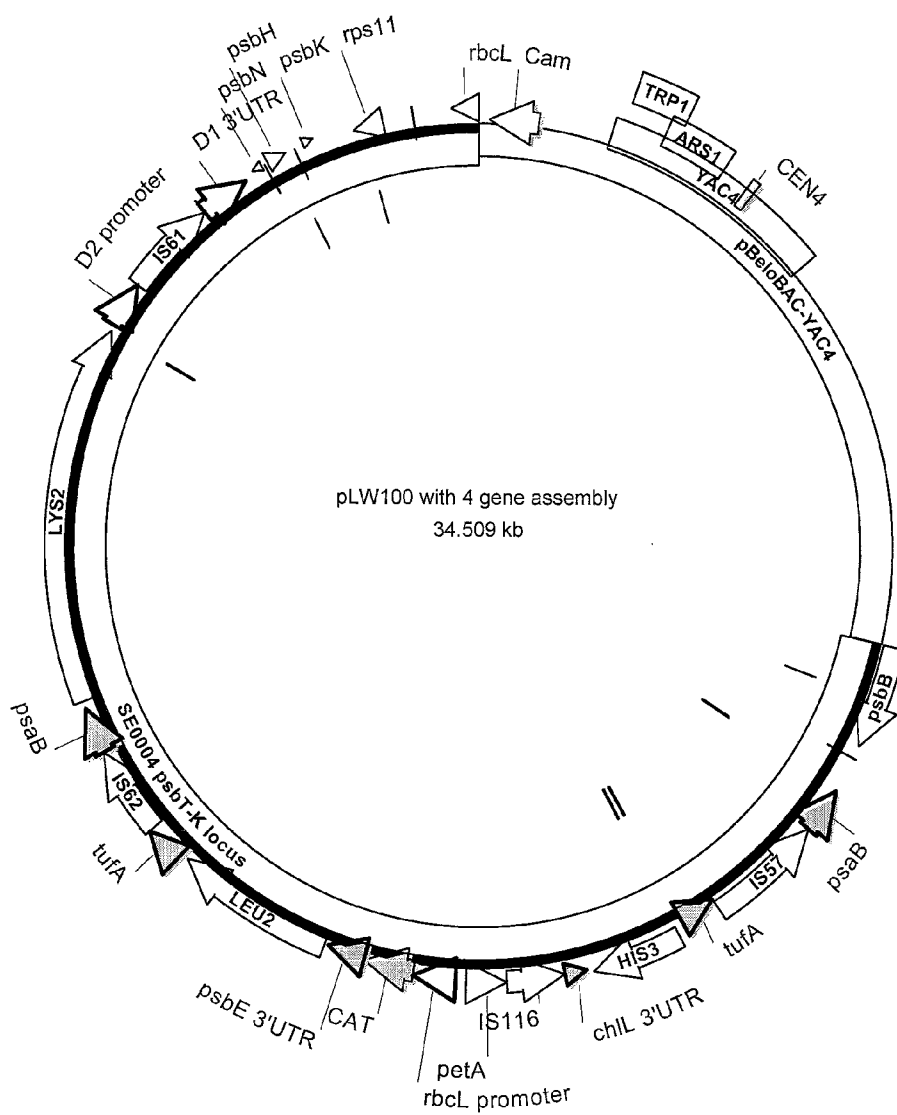


FIG. 64

81/113

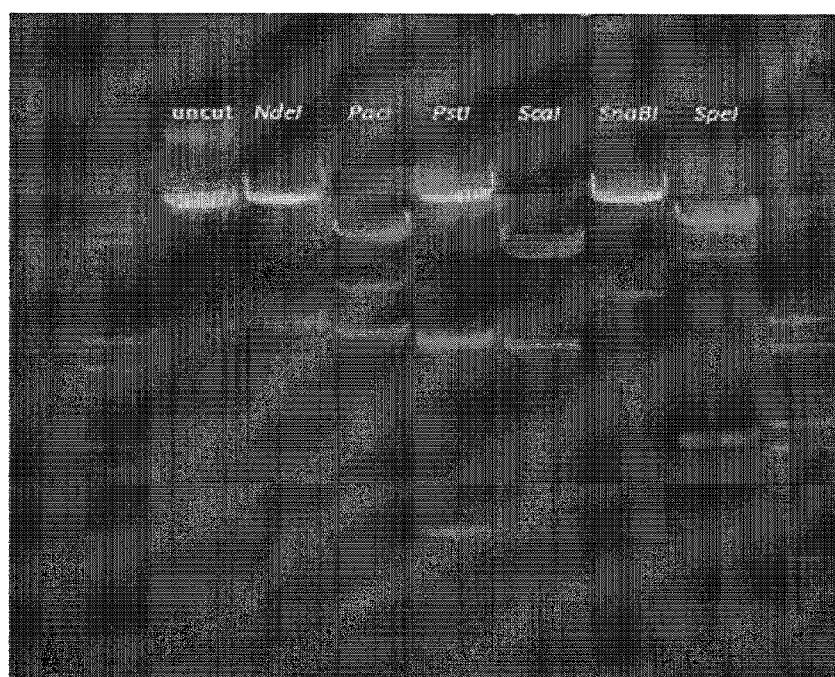


FIG. 65

82/113

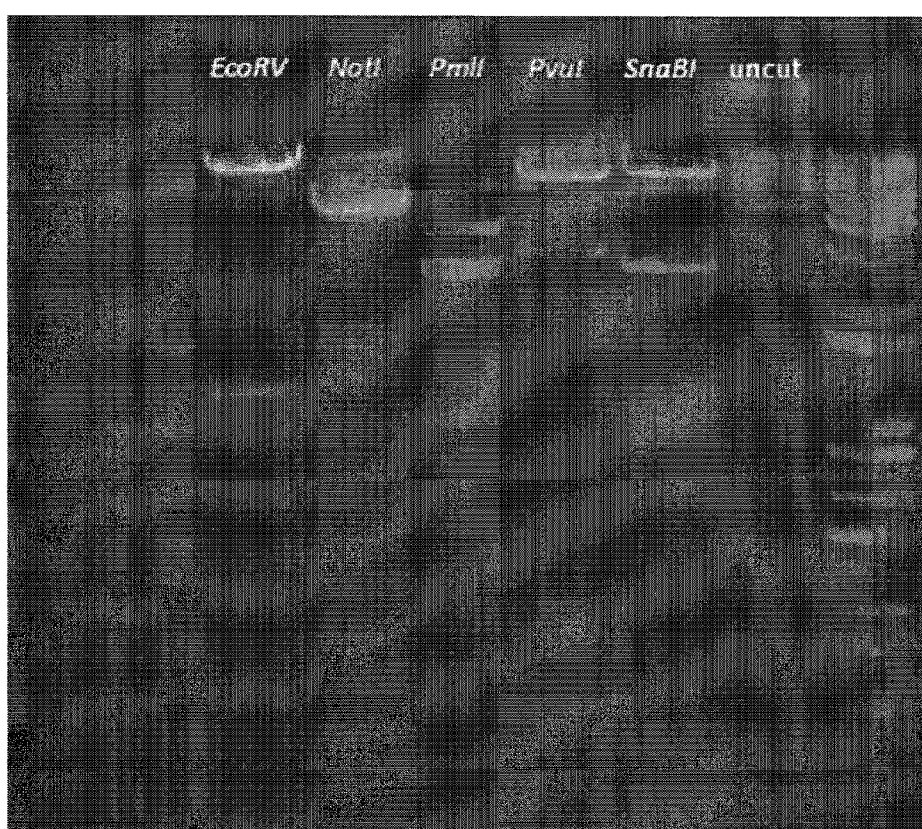


FIG. 66

83/113

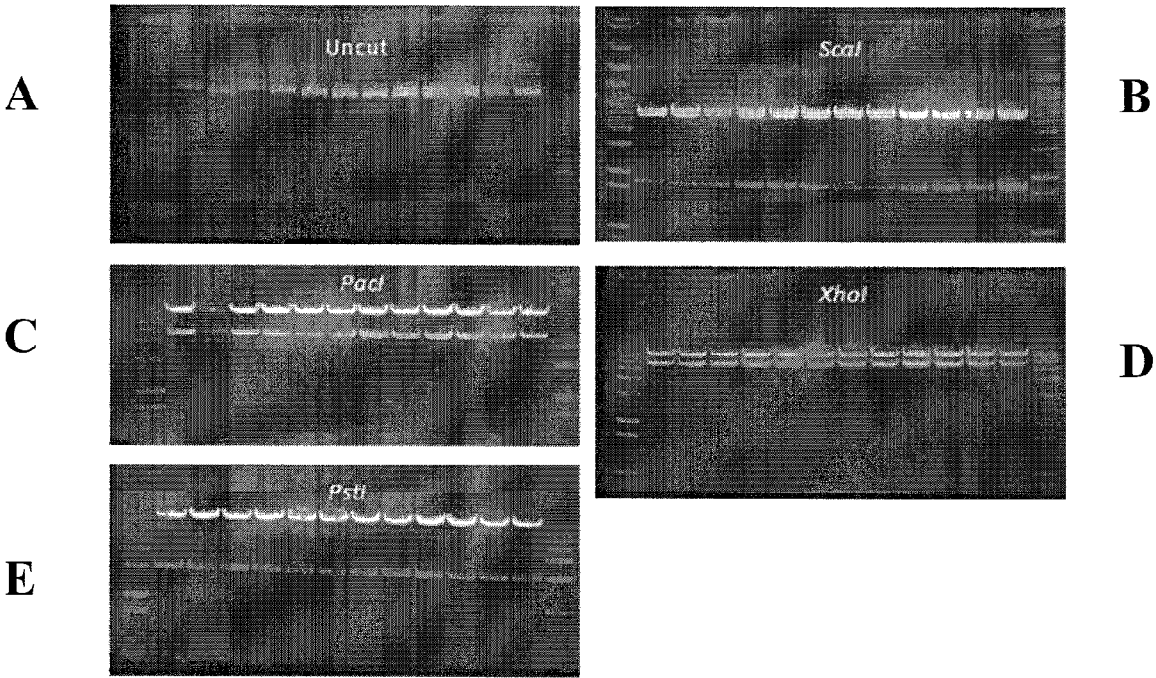


FIG. 67

84/113

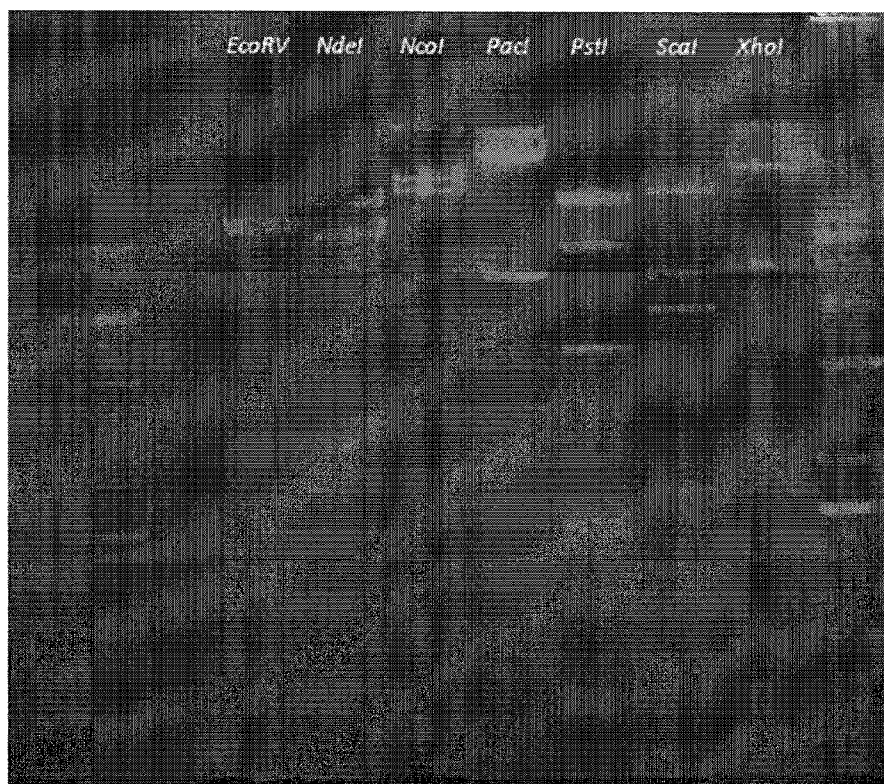


FIG. 68

85/113

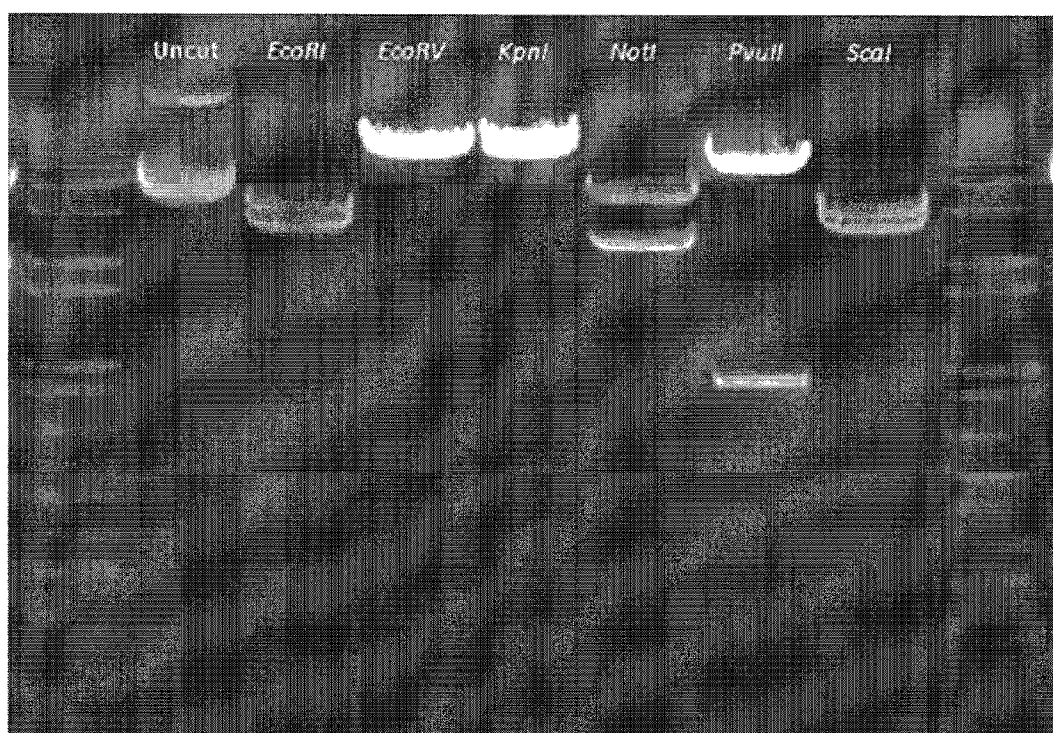


FIG. 69

86/113

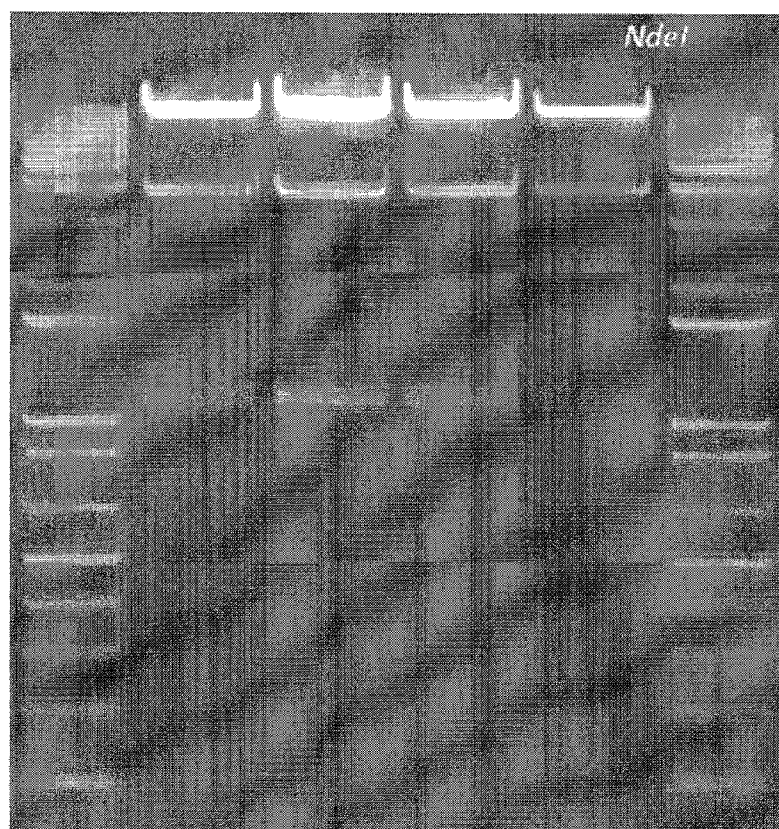


FIG. 70

87/113

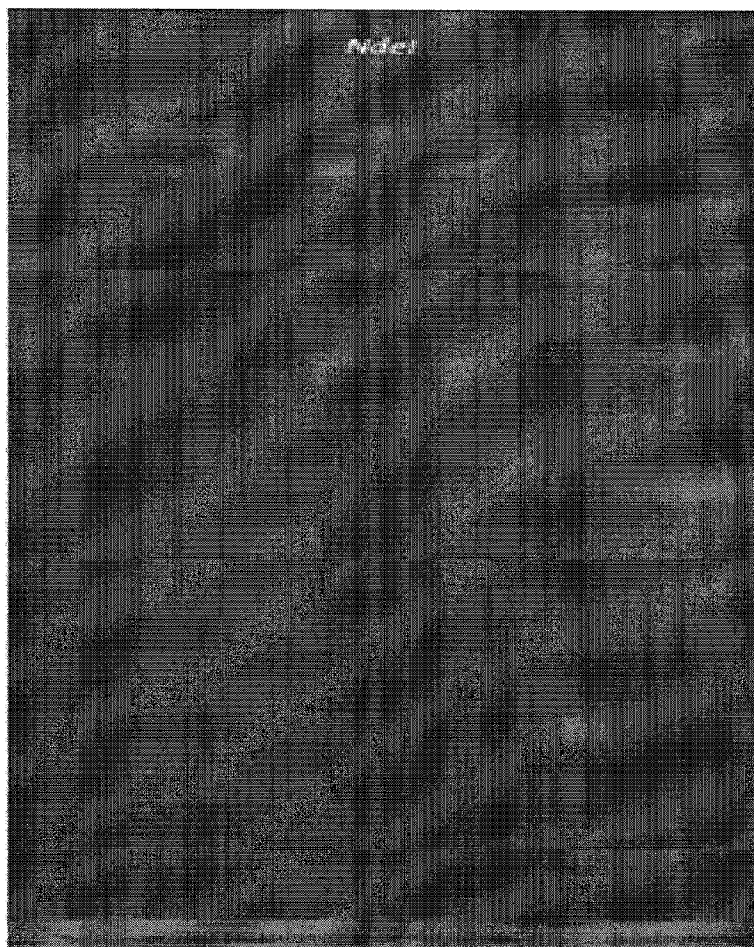


FIG. 71

88/113

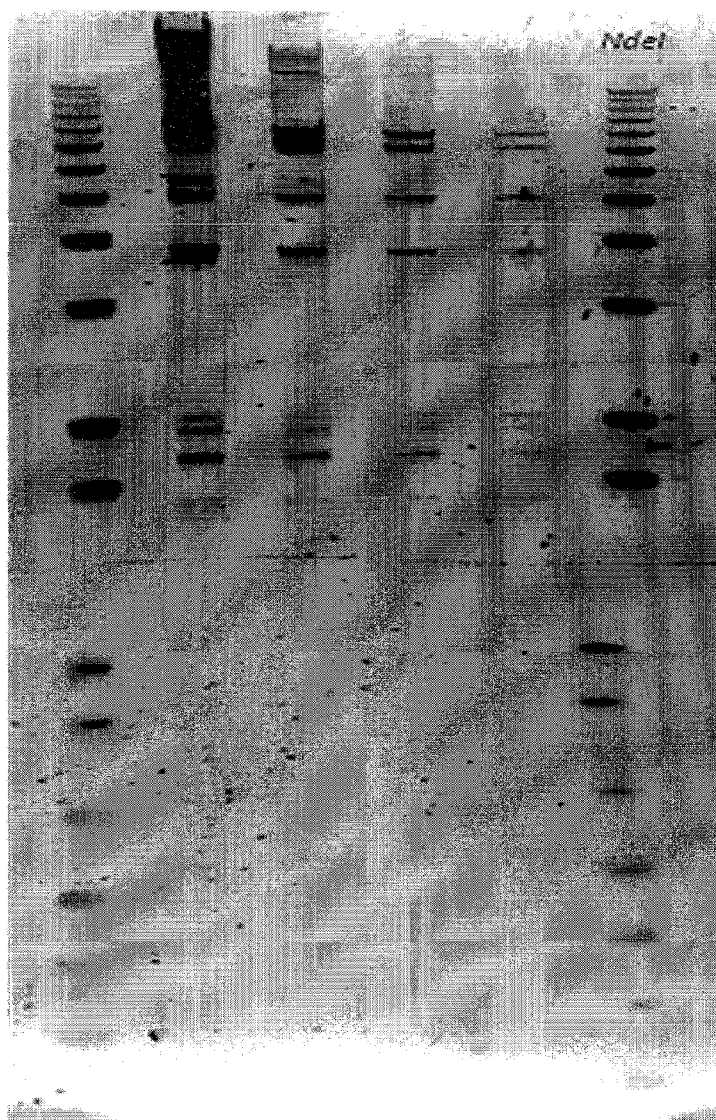


FIG. 72

89/113

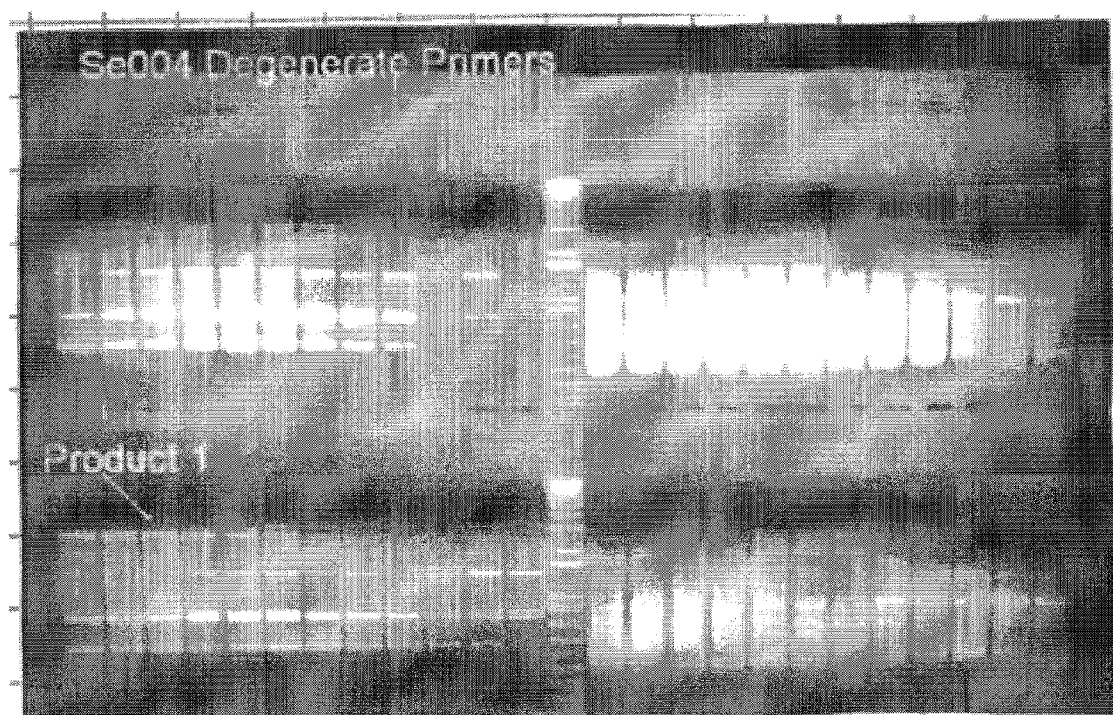


FIG. 73

90/113



FIG. 74

91/113

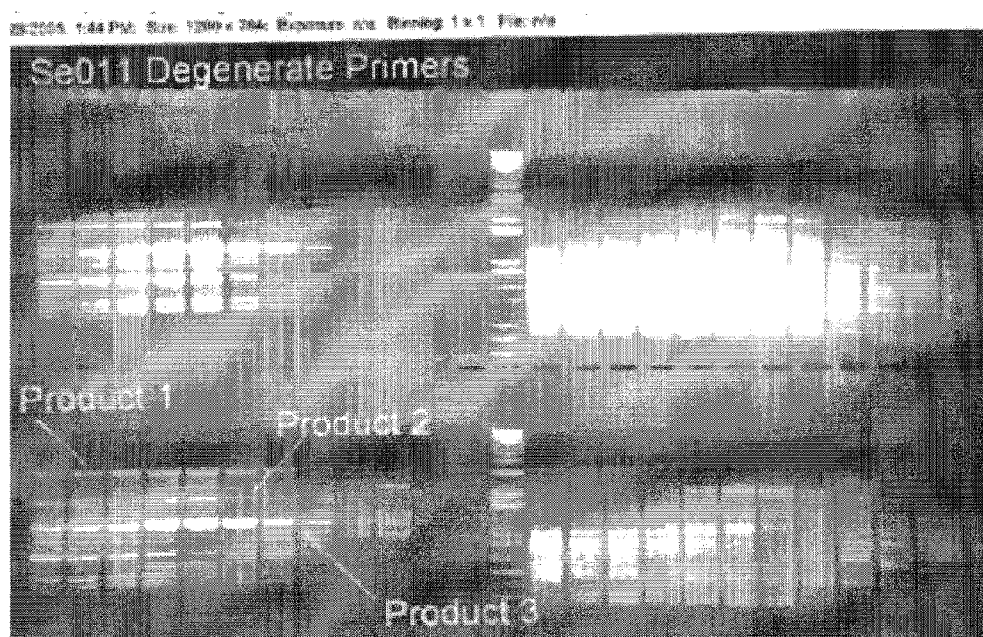


FIG. 75

92/113

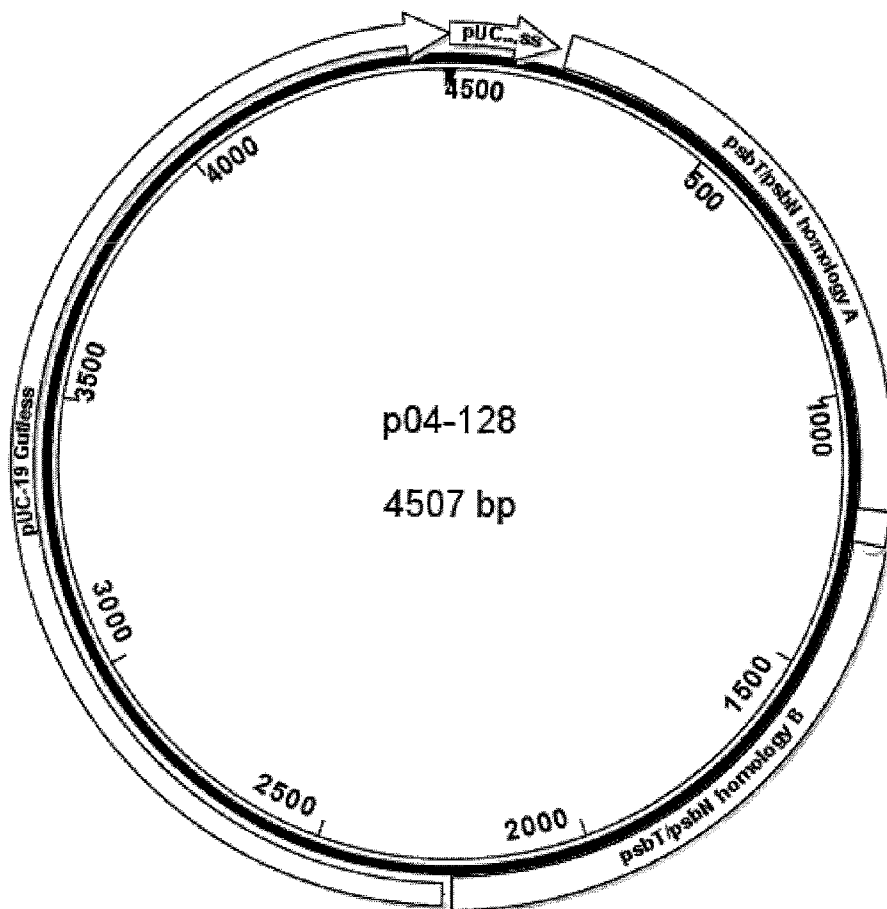


FIG. 76

93/113

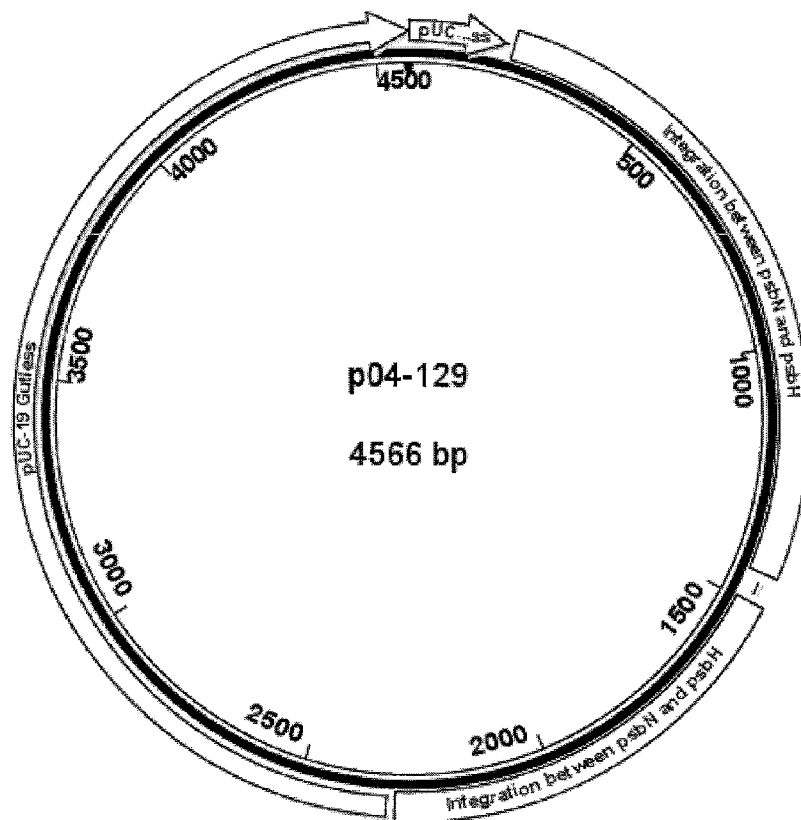


FIG. 77

94/113

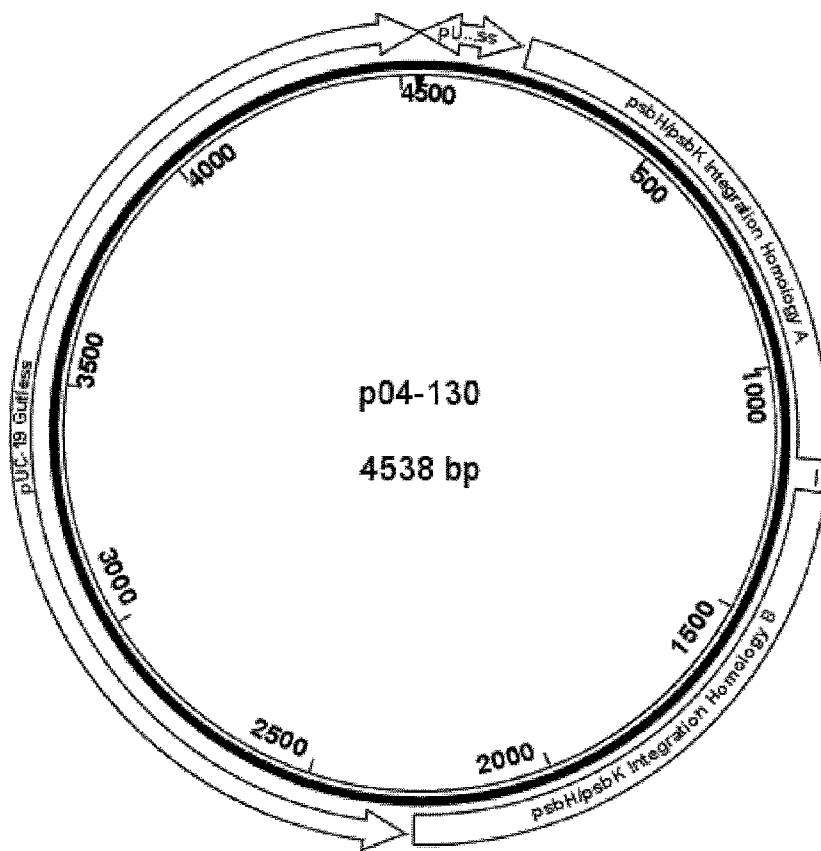


FIG. 78

95/113

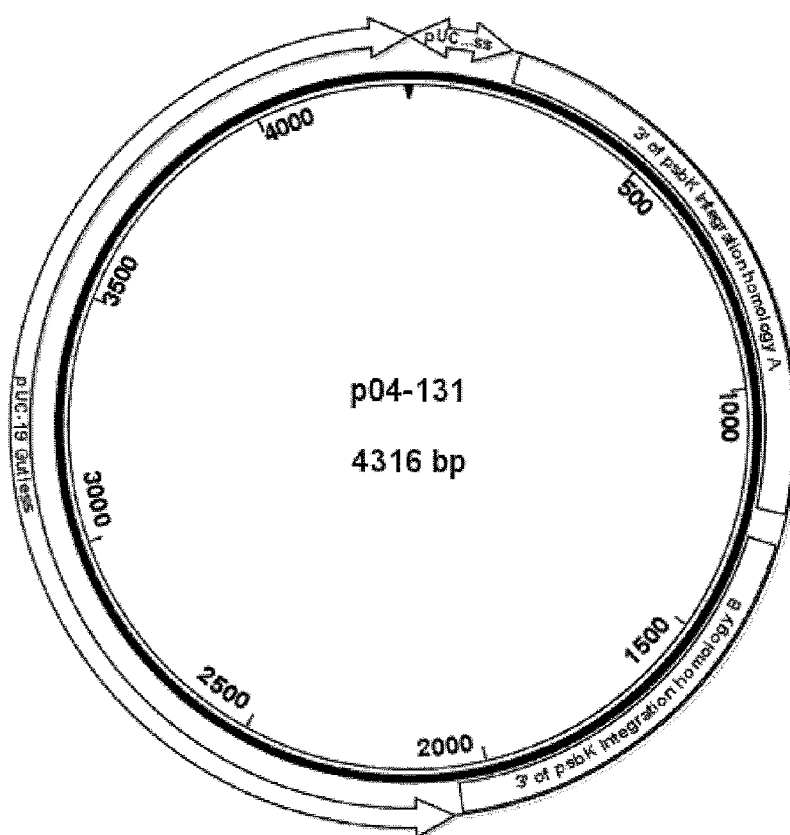


FIG. 79

96/113

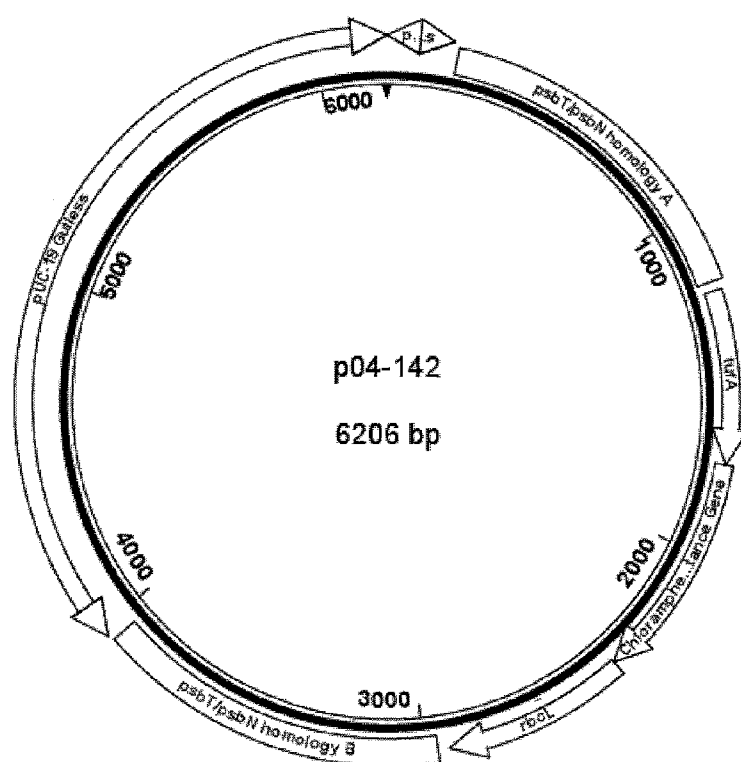


FIG. 80

97/113

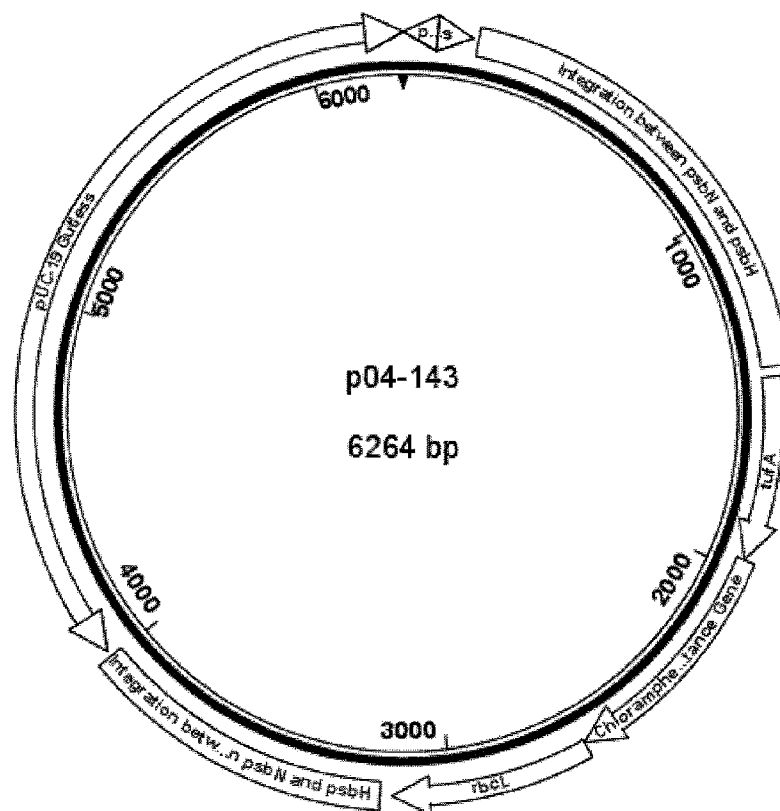


FIG. 81

98/113

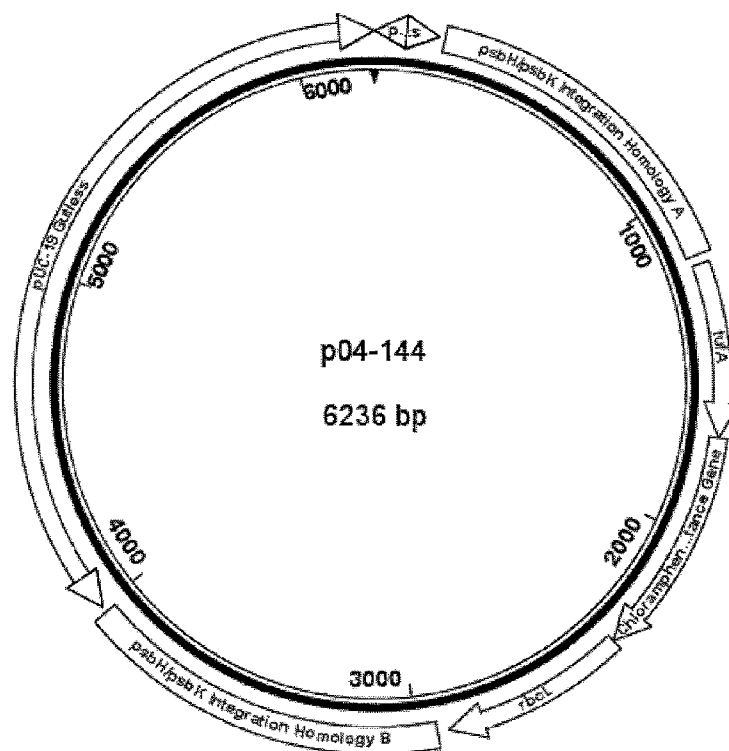


FIG. 82

99/113

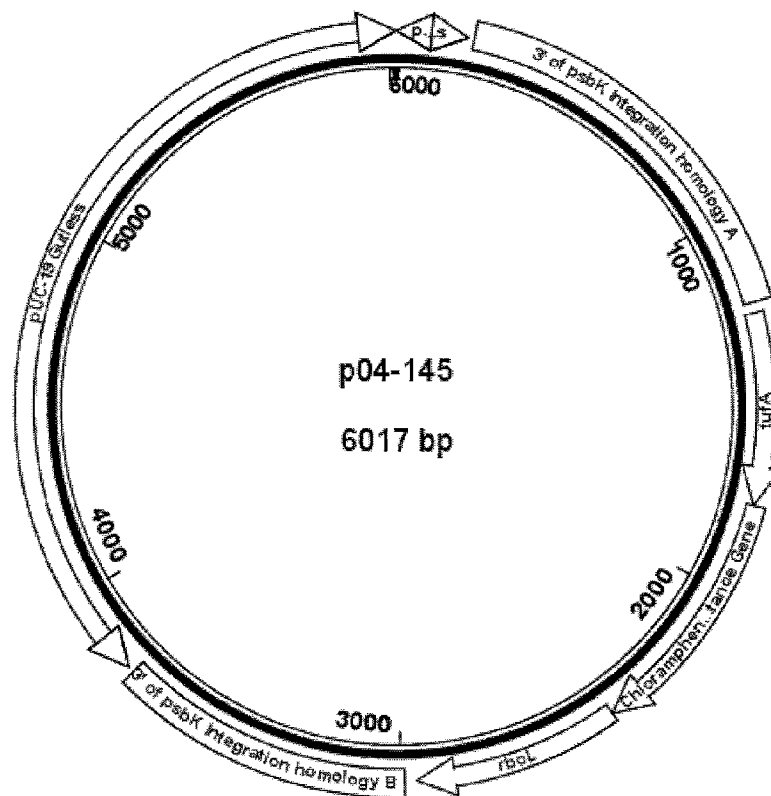
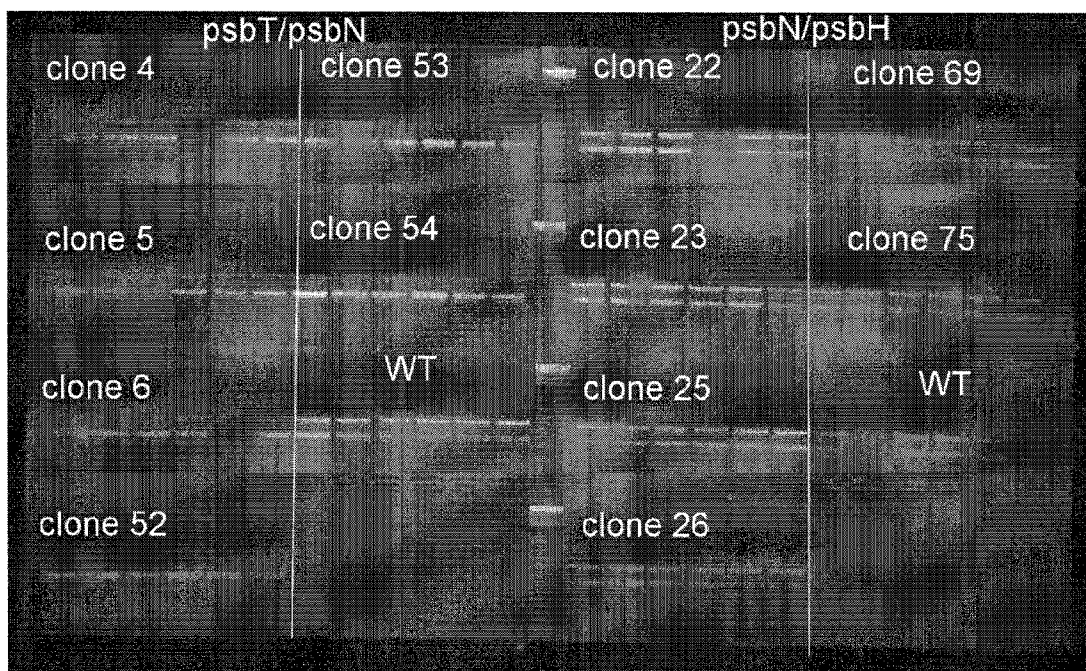


FIG. 83

100/113*FIG. 84*

101/113

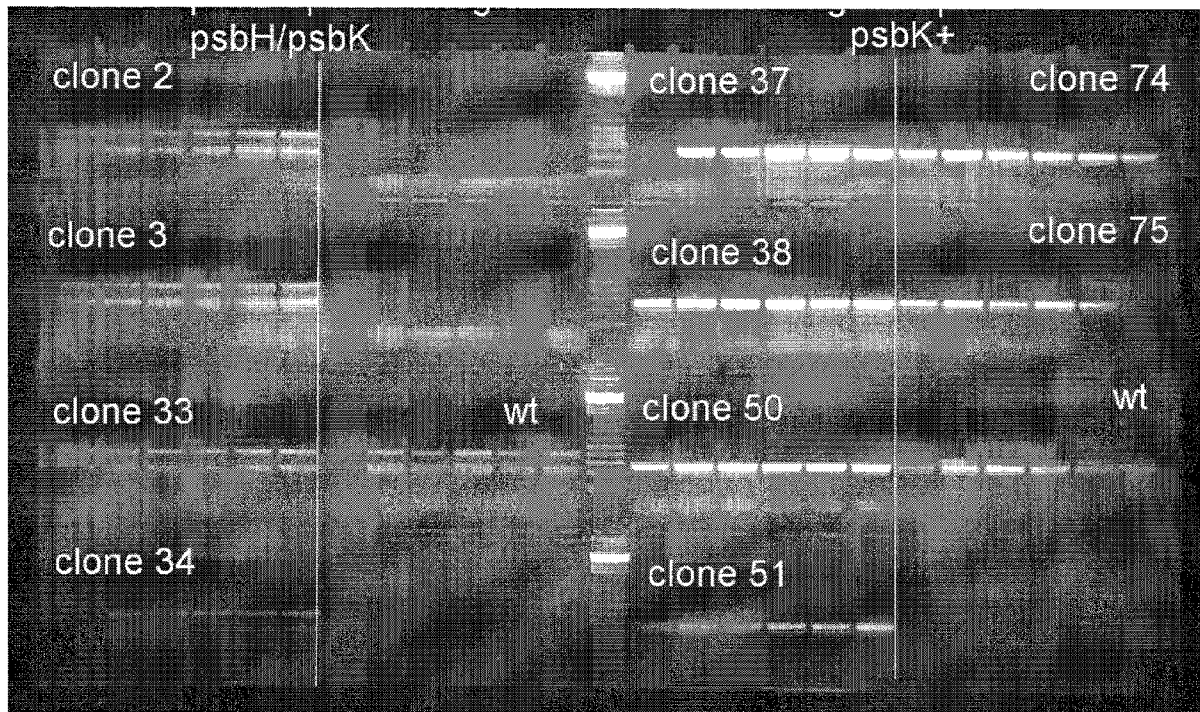


FIG. 85

102/113

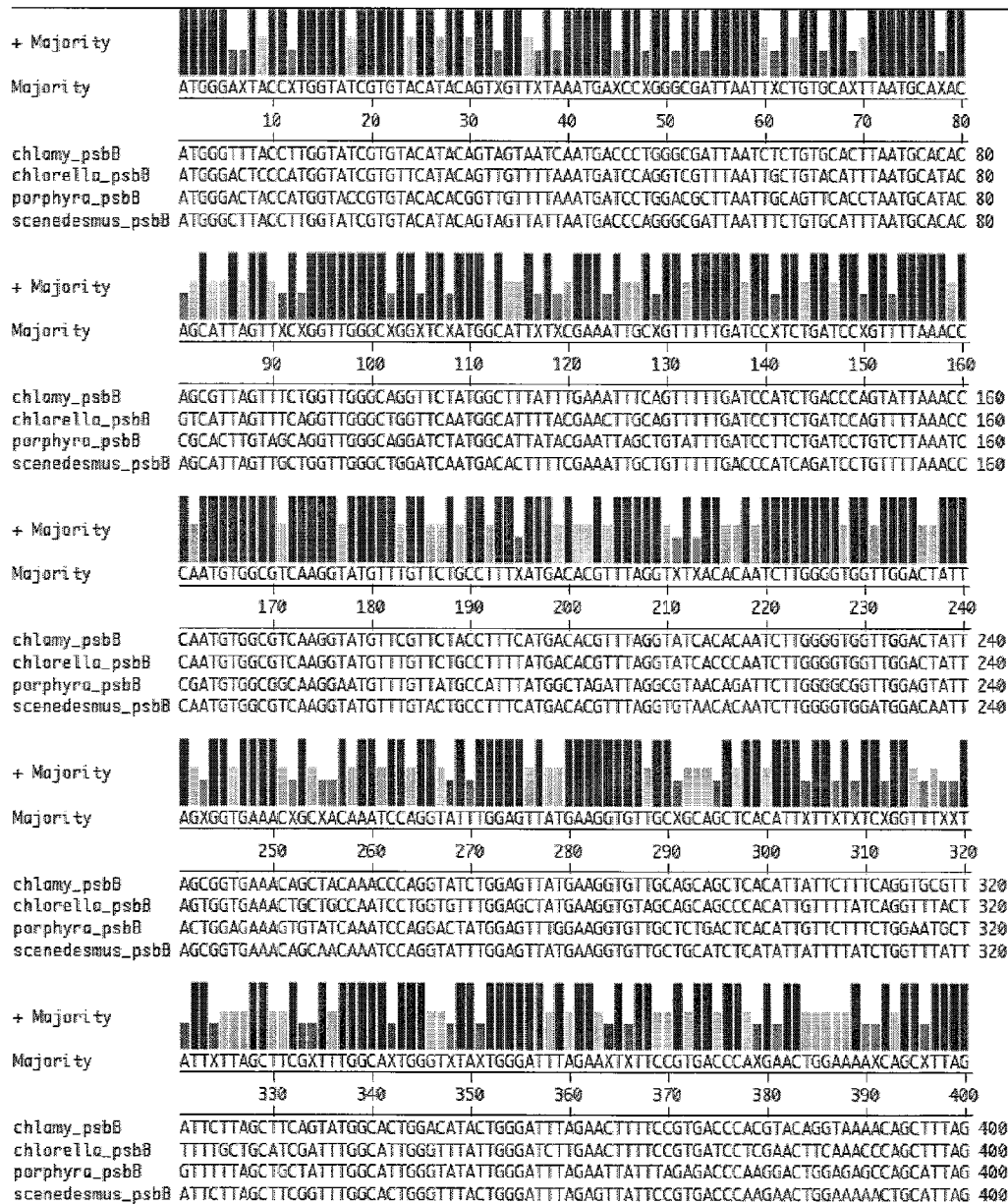


FIG. 86

103/113

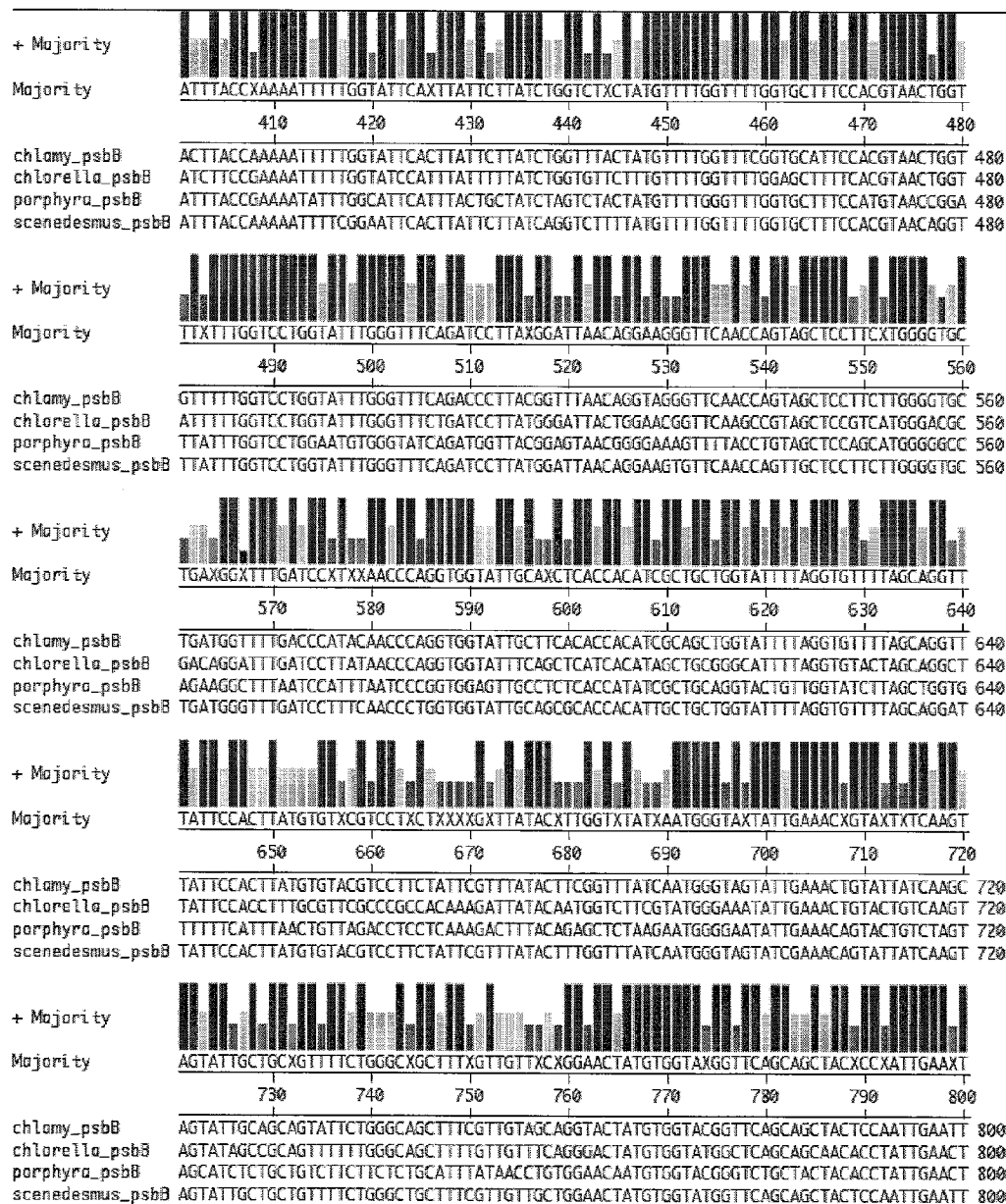


FIG. 86 (cont'd)

104/113

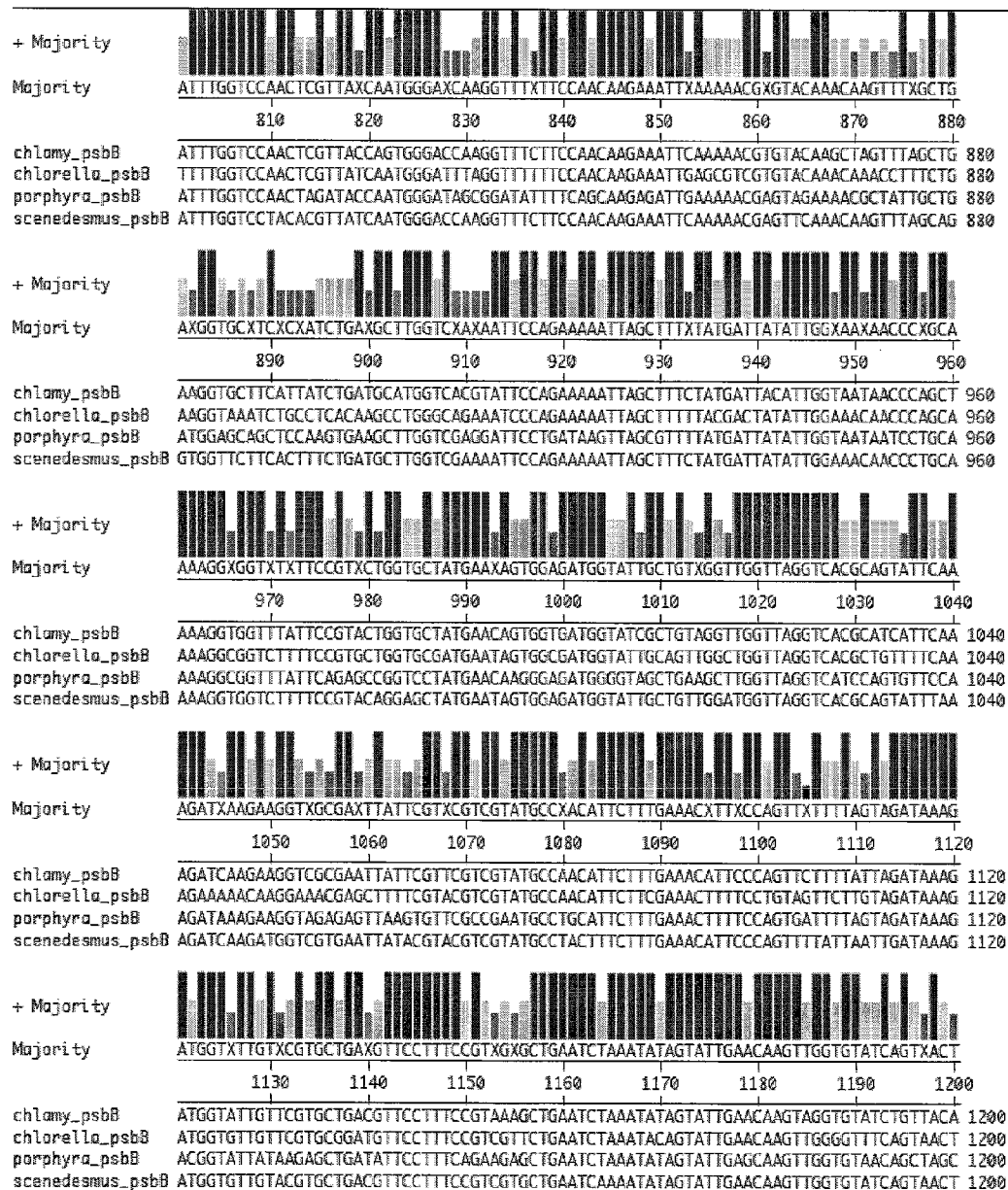


FIG. 86 (cont'd)

105/113

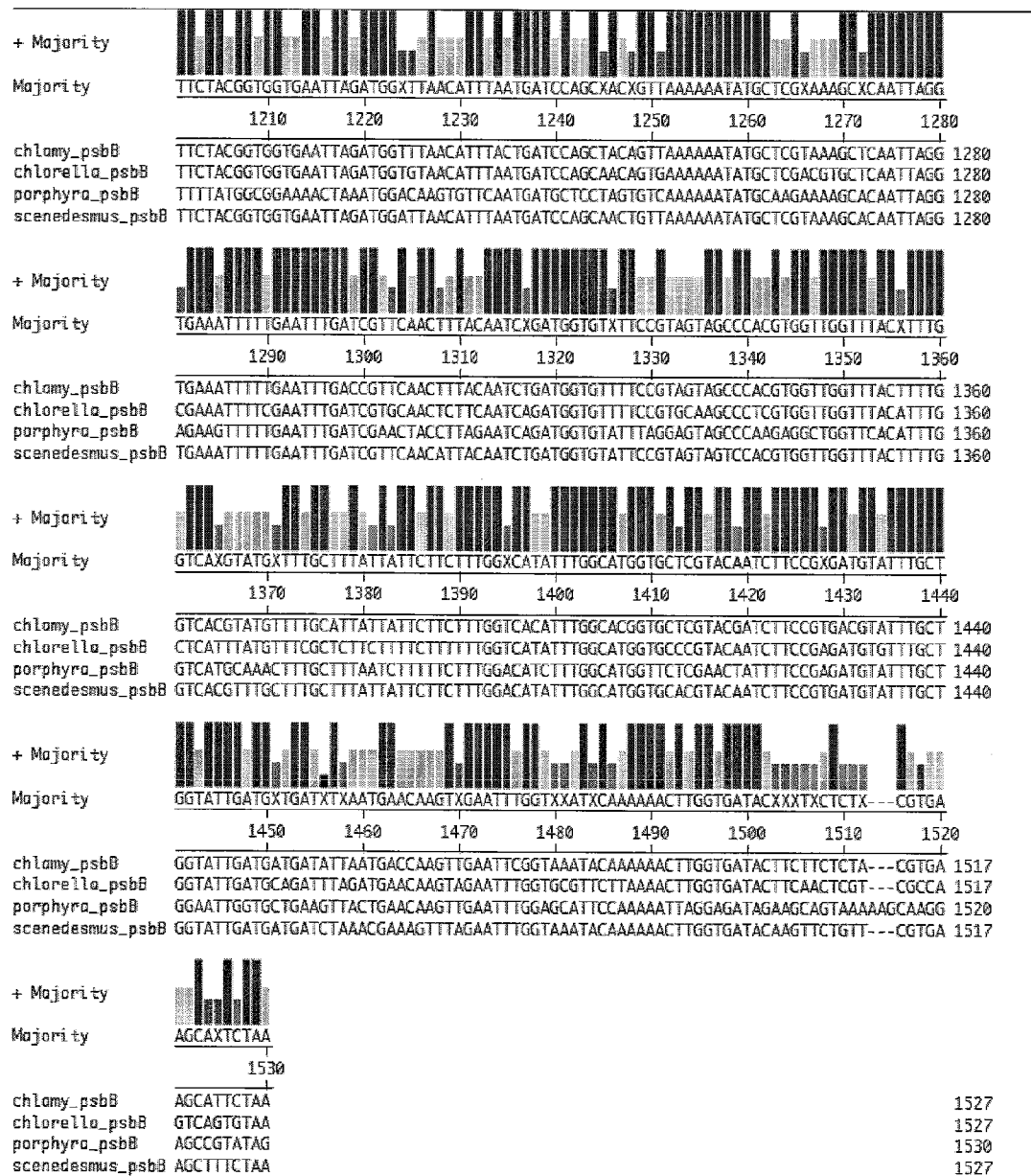


FIG. 86 (cont'd)

106/113

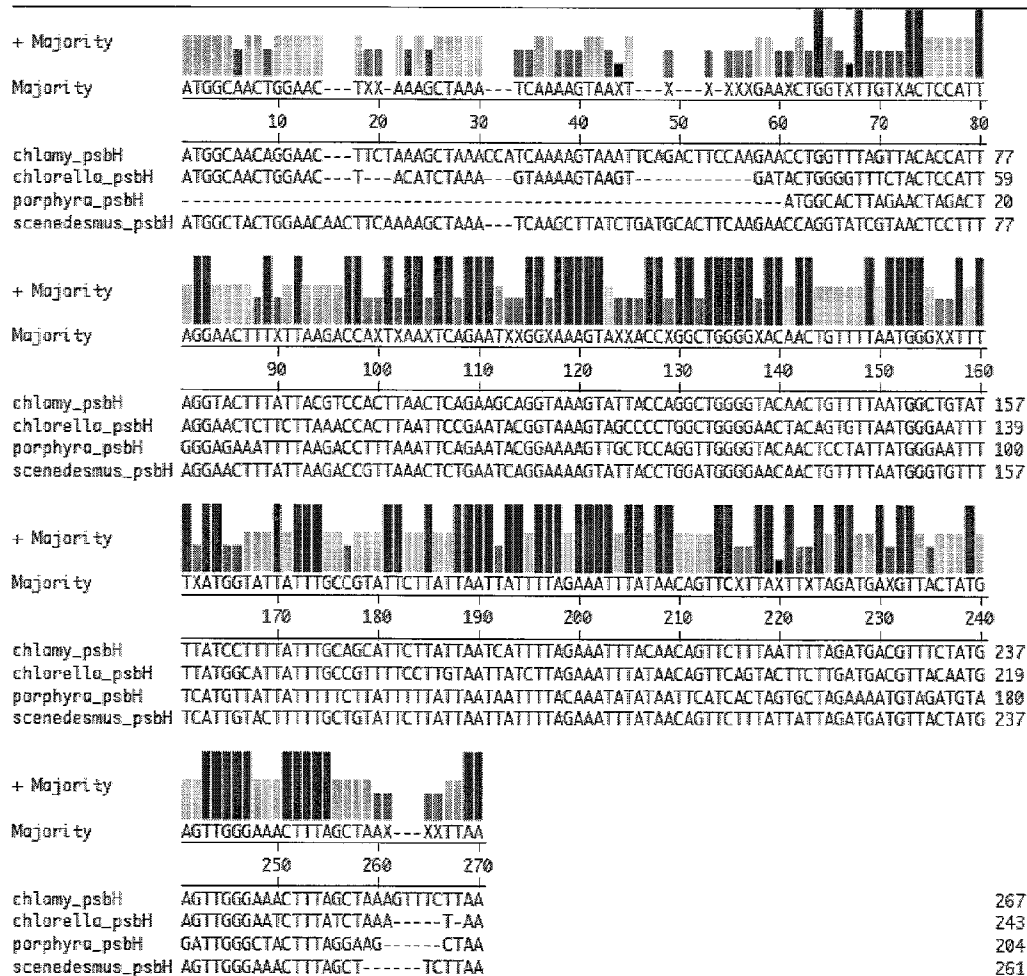
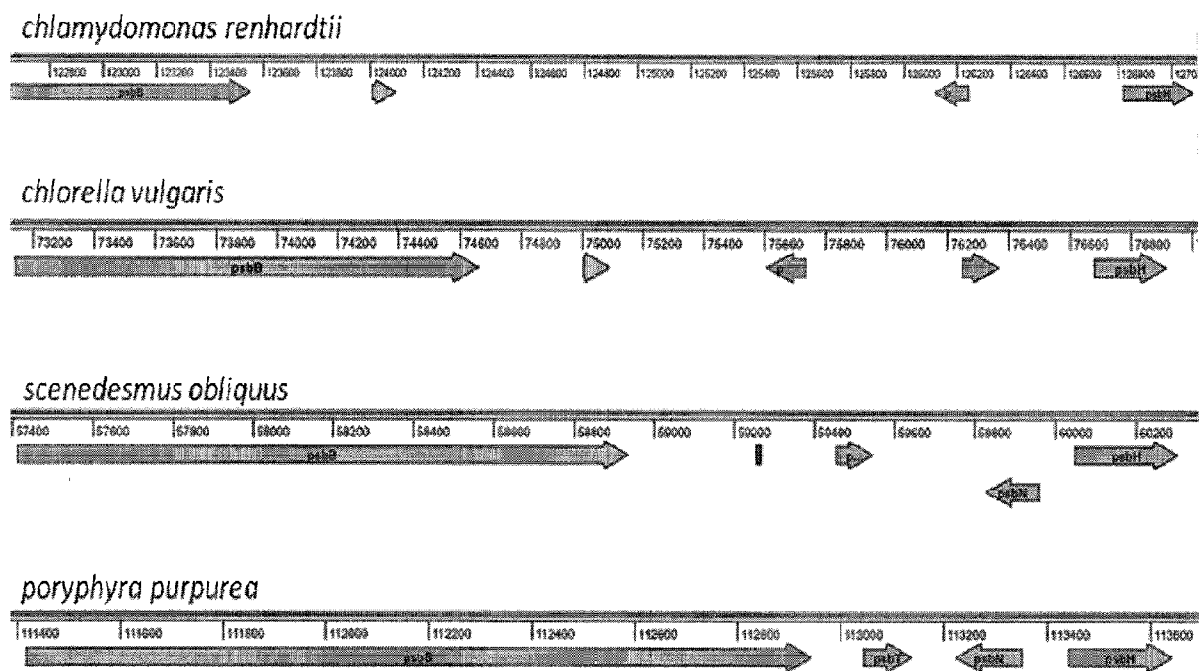


FIG. 87

107/113

FIG. 88



108/113

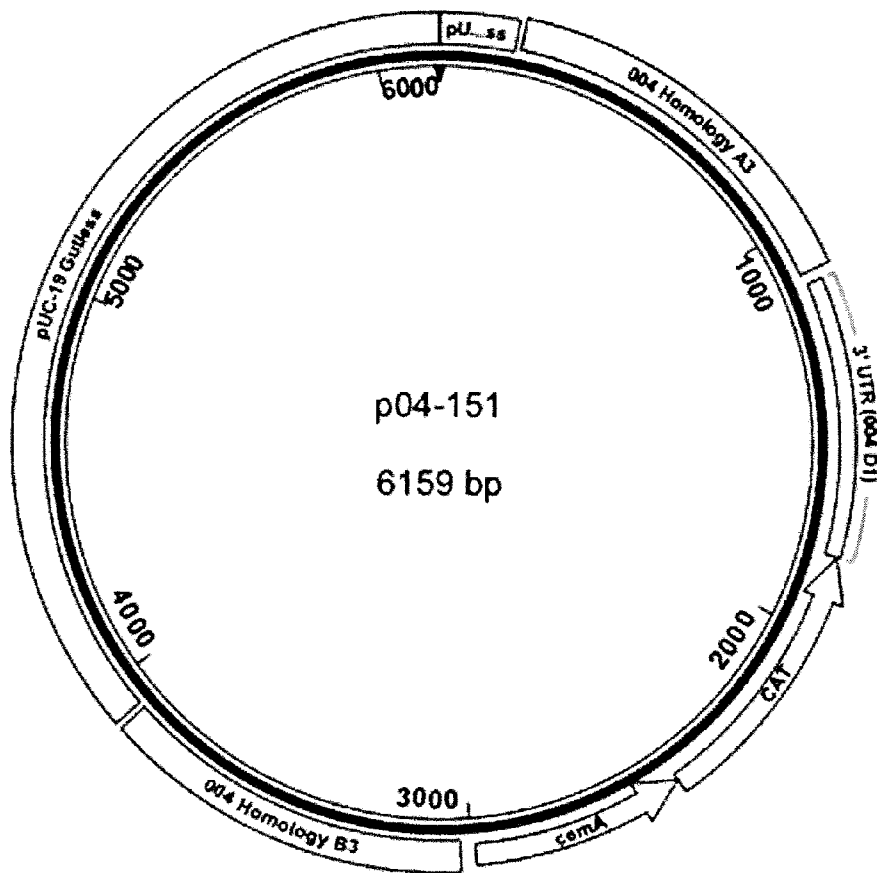


FIG. 89

109/113

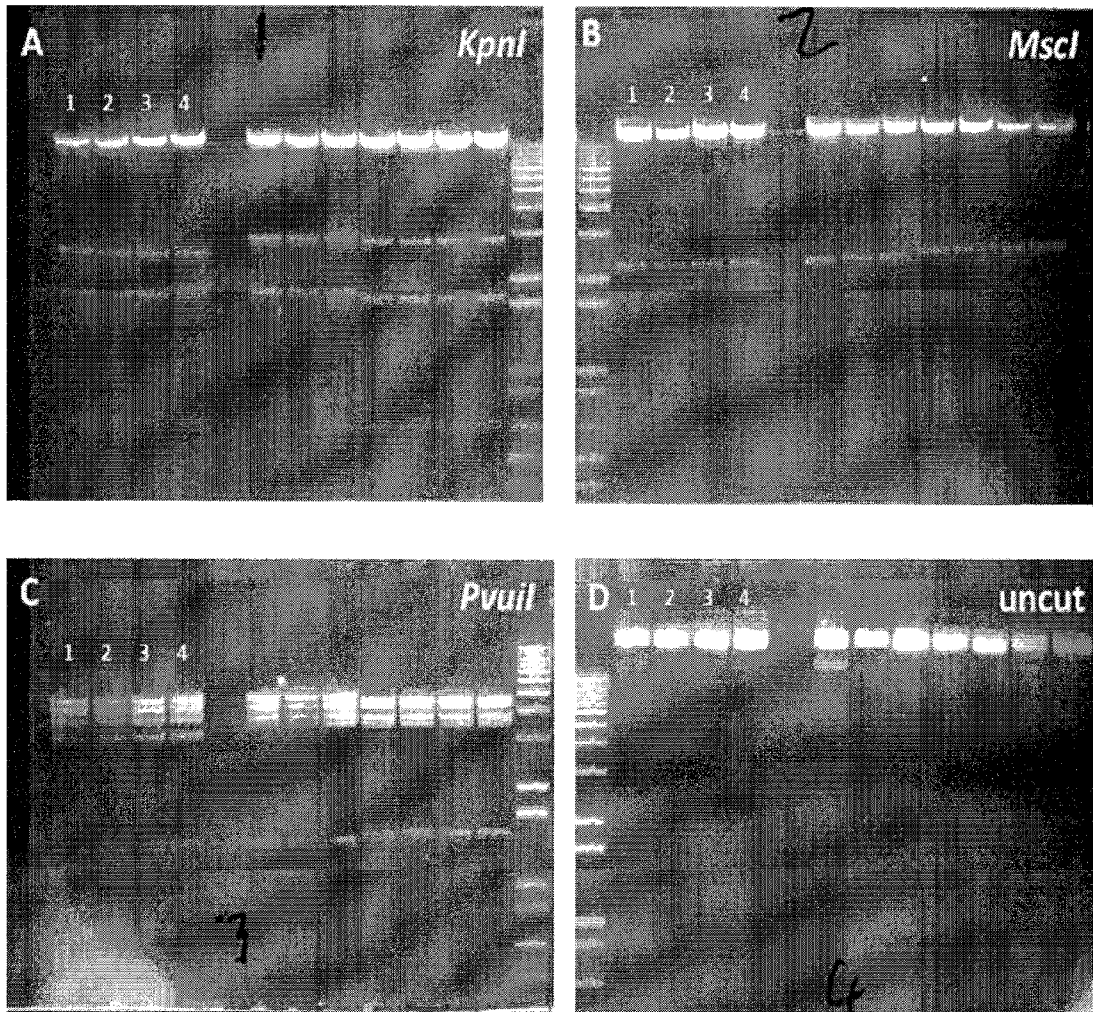


FIG. 90

110/113

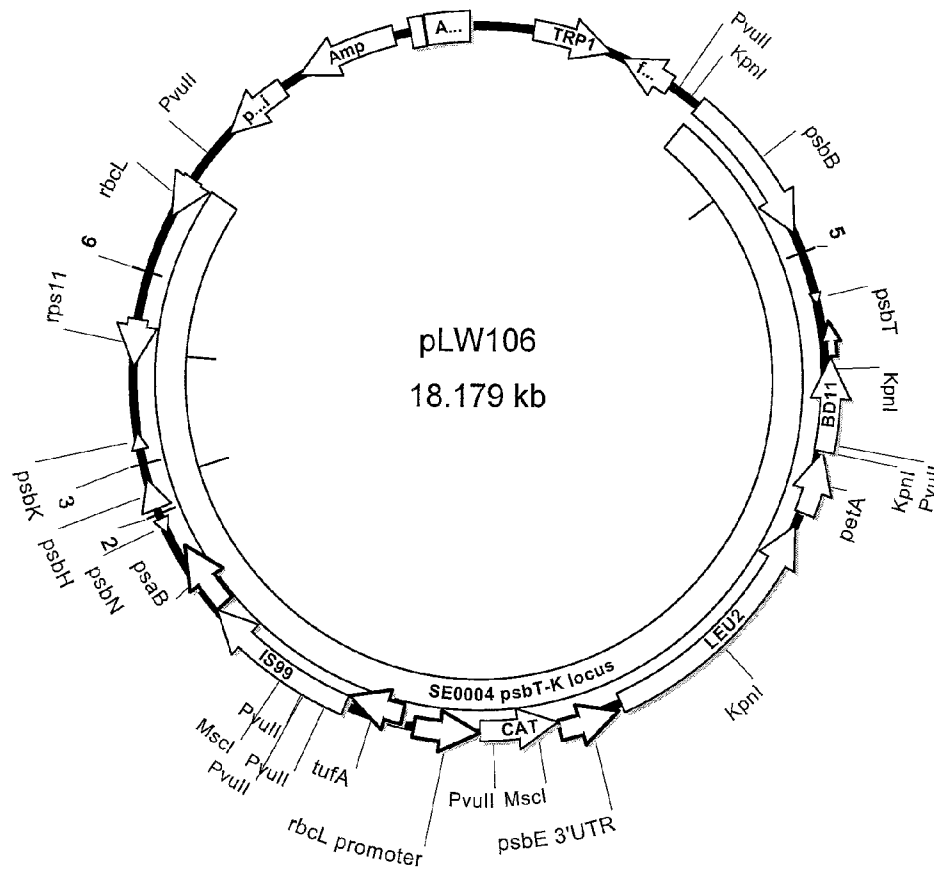


FIG. 91

111/113

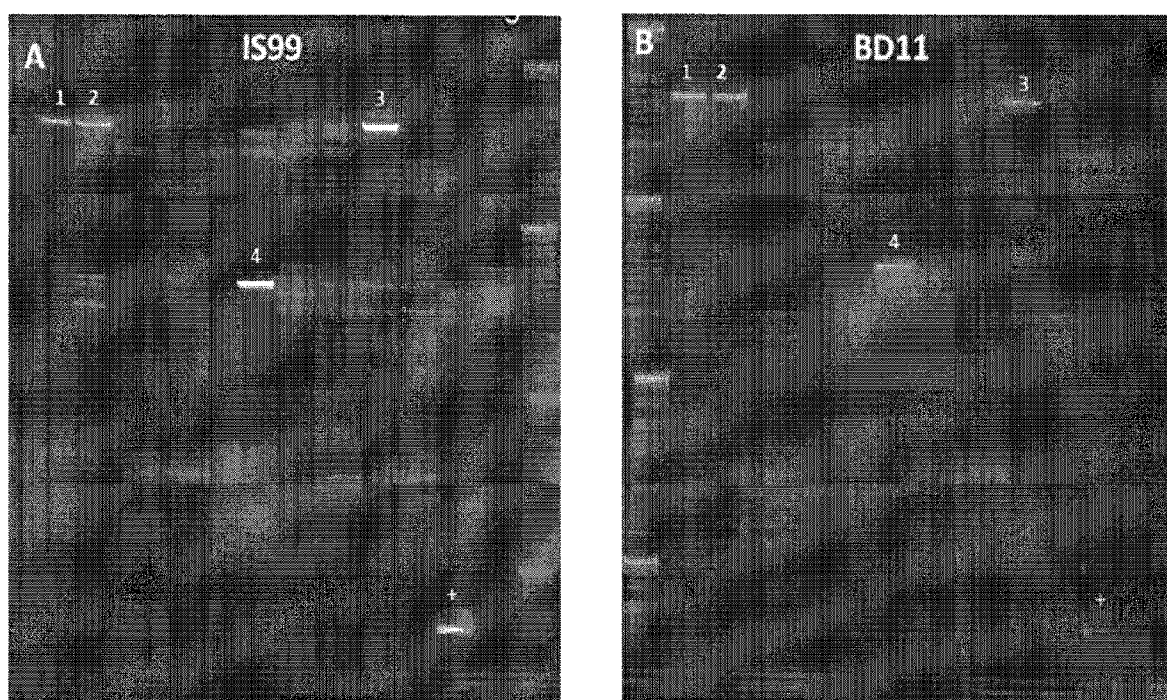
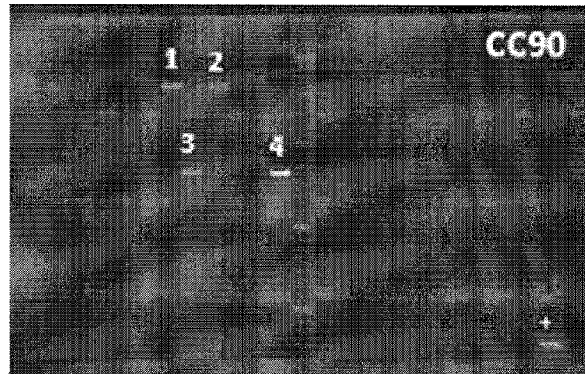
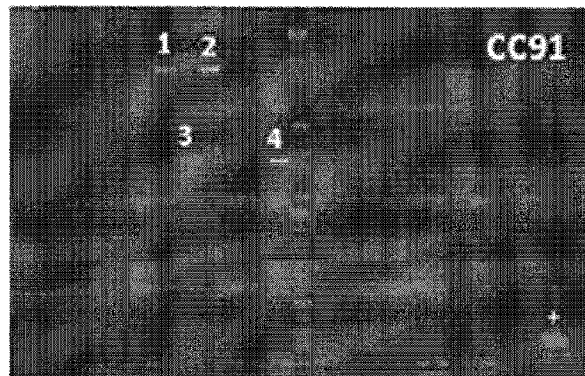


FIG. 92

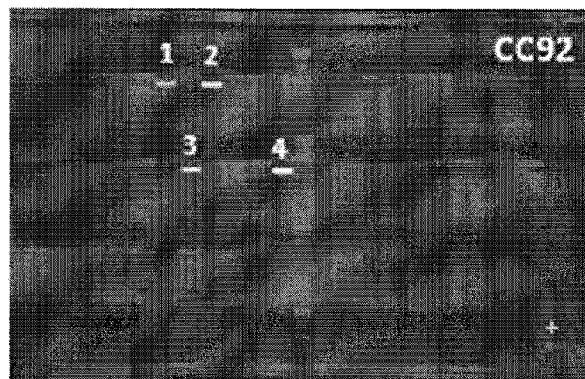
112/113



A



B



C

FIG. 93

113/113

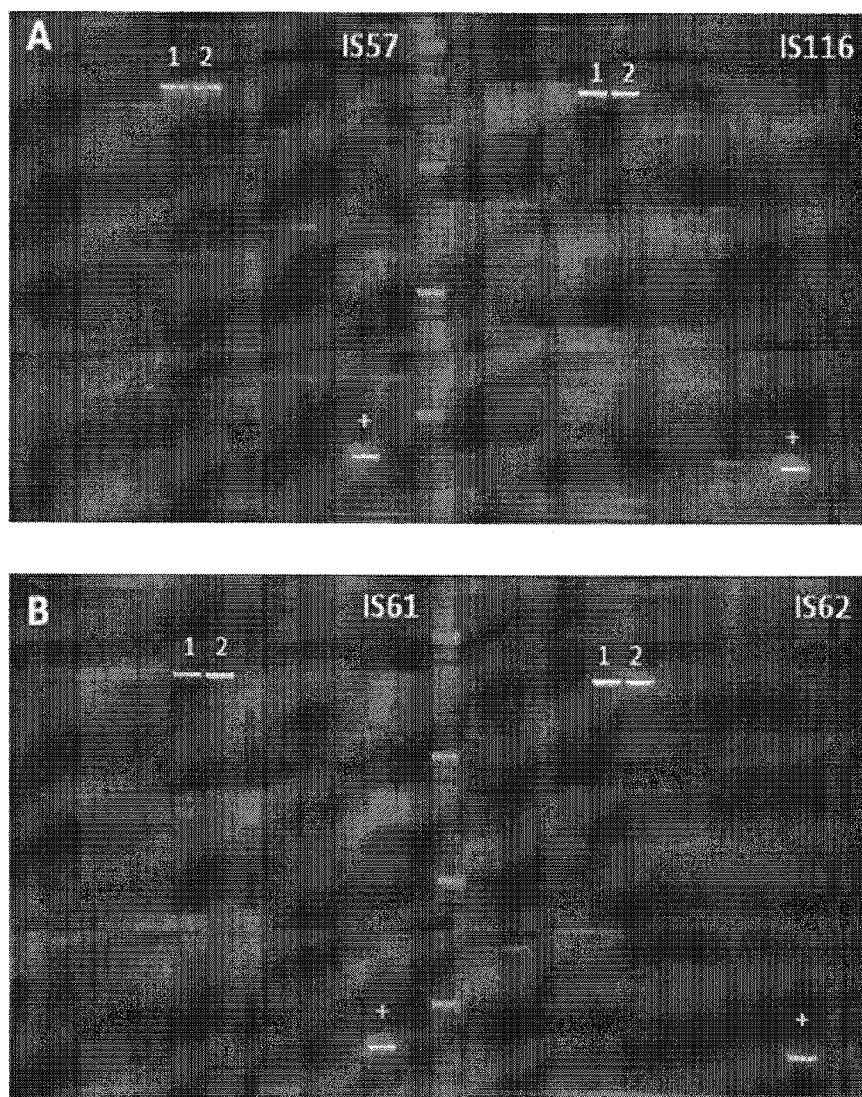


FIG. 94

0802WO1_ST25
SEQUENCE LISTING

<110> Sapphire Energy, Inc.
Botsch, Kyle
Szyjka, Shawn
LeVine, Wendy
Curran, Amy
O'Neill, Bryan
Mendez, Michael

<120> A SYSTEM FOR TRANSFORMATION OF THE CHLOROPLAST GENOME OF
SCENEDESMUS SP. AND DUNALIELLA SP.

<130> 0802WO1

<150> 61/242735

<151> 2009-09-15

<160> 170

<170> Patent In version 3.5

<210> 1

<211> 21

<212> DNA

<213> Artificial Sequence

<220>

<223> PCR primer

<400> 1

agaaggagct tctacagatg c

21

<210> 2

<211> 26

<212> DNA

<213> Artificial Sequence

<220>

<223> PCR primer

<400> 2

tcattagtta cttcatcttt aatccg

26

<210> 3

<211> 21

<212> DNA

<213> Artificial Sequence

<220>

<223> PCR primer

<400> 3

gaactacaac taattat ttt c

21

<210> 4

<211> 24

<212> DNA

<213> Artificial Sequence

<220>
 <223> PCR primer
 <400> 4
 t g a a c c a g t c t t t g t a a g c t c a 24

<210> 5
 <211> 22
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> PCR primer
 <400> 5
 c t a a t t c c a g c a a c c a g c a t t 22

<210> 6
 <211> 22
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> PCR primer
 <400> 6
 c g t t c t t c t g a g a a t g g c t t a 22

<210> 7
 <211> 27
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> PCR primer
 <400> 7
 t g t a a t t t a a g g c t g c c t g t g a t g t g 27

<210> 8
 <211> 28
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> PCR primer
 <400> 8
 g a g g t t c g a a g a a t g g g t c a a a g a t a a g 28

<210> 9
 <211> 24
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> PCR primer

<400> 9
caact accac t ggagat aaa t t t c 24

<210> 10
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> PCR primer

<400> 10
aat cact ct a ccaact gagt t at gg 25

<210> 11
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> PCR primer

<400> 11
gcact acct g at gaaaaat a acc 23

<210> 12
<211> 33
<212> DNA
<213> Artificial Sequence

<220>
<223> PCR primer

<400> 12
t t ggt t t ct a gatt acgccc cgccct gcc a ct c 33

<210> 13
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> PCR primer

<400> 13
gt gaat caac aact gat t gg 20

<210> 14
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> PCR primer

<400> 14
t gaat t gcat aaat t t acac at ac 24

<210> 15
 <211> 24
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> PCR primer

<400> 15
 ccacct cgt a t ggt aaaat a at t g

24

<210> 16
 <211> 24
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> PCR primer

<400> 16
 gaaagaatt a t ggacagt cc t gct

24

<210> 17
 <211> 42
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> PCR primer

<400> 17
 ttgtttgcggc cgcttttgaa gccgaaat ac tttattttta t g

42

<210> 18
 <211> 35
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> PCR primer

<400> 18
 cat cat t aat ggagaaaaaa at cact ggat at acc

35

<210> 19
 <211> 35
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> PCR primer

<400> 19
 aat t cct agg t t acgccccg ccct gccact cat cg

35

<210> 20
 <211> 87

0802W01_ST25

<212> DNA
 <213> Artificial Sequence
 <220>
 <223> FLAG epitope tag linked to a MAT epitope tag by a TEV protease site
 <400> 20
 ggt accggt g att acaaaga t gat gacgat aaaagt ggt g aaaccttta tttt caaggc 60
 cat aat cacc gt cacaaca caccgt 87

<210> 21
 <211> 762
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> codon optimized endoxylanase from T. reesei
 <220>
 <221> misc_feature
 <222> (673)..(759)
 <223> tag
 <400> 21
 at ggt accag t at cttt cac aagt cttt a gcagcat ct c cacct t cag t gcaagt t gc 60
 cgt ccagct g ct gaagt gga at cagt t gca gt agaaaaac gt caaacaat t caaccaggt 120
 acaggt t aca at aacggt t a ctttt at t ct t act ggaat g at ggacacgg t ggt gt t aca 180
 t at act aat g gacct ggt gg t caat t tagt gt aaat t gga gt aact cagg caat t t t gt t 240
 ggaggaaaag gt t ggcaacc t ggt acaaag aat aaggt aa t caat t t ct c t ggt agt t ac 300
 aaccct aat g gt aat t ct t a t t t aagt gt a t acggt t gga gccgt aaccc at t aat t gaa 360
 t at t at at t g t agagaact t t ggt acat ac aaccct t caa caggt gct ac t aaat t aggt 420
 gaagt t act t cagat ggat c agt t t at gat at t t at cgt a ct caacgcgt aaat caacca 480
 t ct at aat t g gaact gccac t t t ct accaa t act ggagt g t aagacgt aa t cat cgt t ca 540
 agt ggt agt g t t aat acagc aaaccact t t aat gcat ggg ct caacaagg t t t aacat t a 600
 ggt acaat gg act at caaat t gt agct gt t gaaggt t at t t t t cat cagg t agt gct t ct 660
 at cact gt t a gcggt accgg t gat t aaaa gat gat gacg at aaaagt gg t gaaaacct t 720
 t at t t t caag gccat aat ca ccgt caaaa cacaccggt t aa 762

<210> 22
 <211> 81
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> TEV protease site linked to a FLAG epitope tag

0802W01_ST25

<400> 22
 ggt accggt g aaaact t at a c t t t caaggc t caggt ggcg gt ggaagt ga t t acaaagat 60
 gat gat gat a aaggaaccgg t 81

<210> 23
 <211> 1191
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> codon optimized FPP synthase from G gallus

<220>
 <221> misc_feature
 <222> (1108)..(1188)
 <223> tag

<400> 23
 at ggt accac acaagt t cac aggt gt t aac gct aaat t cc agcaaccagc at t aagaat 60
 t t at ct ccag t ggt agt t ga gcgcgaacgt gaggaat t t g t aggat t c t t t ccacaaat t 120
 gt t cgt gact t aact gaaga t ggt at t ggt cat ccagaag t aggt gacgc t gt agct cgt 180
 c t t aaagaag t at t acaat a caacgcacct ggt ggt aaat gcaat agagg t t t aacagt t 240
 gt t gcagct t accgt gaact t t ct ggacca ggt caaaaag acgct gaaag t c t t cgt t gt 300
 gct t t agcag t aggat ggt g t at t gaat t a t t ccaagcct t t t t c t t agt t gct gacgat 360
 at aat ggacc agt catt aac t agacgt ggt caat t at gt t ggt acaagaa agaaggt gt t 420
 ggt t t agat g caat aaat ga t t c t t t c t t t t agaaaagct ct gt gt at cg cgt t c t t aaa 480
 aagt at t gcc gt caacgt cc at at t at gt a cat t t at t ag agct t t t t c t t caaacagct 540
 t accaaacag aat t aggaca aat gt t agat t t aat cact g ct cct gt at c t aaggt agat 600
 t t aagccat t t ct cagaaga acgt t acaaa gct at t gt t a agt at aaac t gct t t ct at 660
 t cat t ct at t t accagt t gc agcagct at g t at at ggt t g gt at agat t c t aaagaagaa 720
 cat gaaaacg caaaagct at t t t act t gag at ggt gaat act t ccaa t t caagat gat 780
 t at t t agat t gt t t t ggcga t cct gct t t a acaggt aaag t aggt act ga t at t caagat 840
 aacaaat gt t cat ggt t agt t gt gcaat gc t t acaaagag t aacaccaga acaacgt caa 900
 c t t t t agaag at aat t acgg t cgt aaagaa ccagaaaaag t t gct aaagt t aaagaat t a 960
 t at gaggct g t aggt at gag agccgcct t t caacaat acg aagaaagt ag t t accgt cgt 1020
 c t t caagagt t aat t gagaa acat t ct aat cgt t t accaa aagaaat t t t c t t aggt t t a 1080
 gct cagaaaa t at acaaacg t caaaaaggt accggt gaaa act t at act t t caaggct ca 1140
 ggt ggcggt g gaagt gat t a caaagat gat gat gat aaag gaaccggt t a a 1191

<210> 24

<211> 36
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> streptavidin epitope tag

<400> 24
 accggt agt g cttgggcaca ccctcaattt gagaaa

36

<210> 25
 <211> 2196
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> codon optimized fusi coccadi ene synthase from P. amygdali

<220>
 <221> misc_feature
 <222> (2158)..(2193)
 <223> tag

<400> 25
 atggaattta aatattcaga agttgttgaa ccatcaacat attat acaga aggtttatgt 60
 gaaggtattg atgttcgt aa atcaaaattt acaacattag aagatcgtgg tgcattcgt 120
 gctcatgaag attggaataa acatattgggt ccatgtggg g aat atcgtgg t acatt aggt 180
 ccacgttttt catttatttc agttgctgtt ccagaatgta ttccagaacg tttagaagt t 240
 atttcat acg ct aatgaatt t gctttttta catgatgat g ttacagat ca t gtt ggt cat 300
 gat acaggt g aagttgaaaa t gatgaaat g atgacagttt tttt agaagc t gct cat aca 360
 ggtgctattg atacatcaaa taaagt gat attcgtcgt g ctggtaaaaa acgtattcaa 420
 tcacaattat ttttagaaat gtt agctatt gatccagaat gtgctaaaac aacaatgaaa 480
 tcatgggctc gttttgttga agttgggtca tcacgtcaac atgaaacacg tttt gttgaa 540
 tt agctaaat atattccat a t cgtattat g gatgttggg g aaatgtttt g gttt ggttta 600
 gttacatttg gtttaggttt acatattcca gatcatgaat tagaattatg t cgt gaact t 660
 atggctaatg cttggattgc t gttggttta caaaat gat a ttt ggt cat g gccaaaagaa 720
 cgtgatgctg ctacattaca tggtaaatgat catgtt gtt a atgctatttg ggttttaatg 780
 caagaacatc aaacagatgt t gatgggct atgcaaat tt gt cgt aaact t at t gtt gaa 840
 tatgttgct a aatattt aga agttattgaa gctacaaaaa at gat gaat c aat t t cat t a 900
 gat t t acgt a aatattt aga t gct at gtt a t att caattt caggt aat gt t gtt t ggt ca 960
 t t agaatgt c cacgttat aa tccagatgtt t catttaata aaacacaatt agaatggat g 1020
 cgt caaggt t t accatcat t agaatcatgt ccagtttt ag ct cgt t cacc agaaat t gat 1080
 t cagat gaat cagcagtttc accaactgct gatgaatcag attcaacaga agattcat t a 1140

0802W01_ST25

```

gggttcaggtt  cacgtcaaga  ttcatcattat  tcaacaggtt  tatcattatc  accagttcat  1200
tcaaatgaag  gtaagatttt  acaacgtgtt  gatacagatc  atatTTTTTT  tgaaaaagct  1260
gttttagaag  ctccatacga  ttatatgtct  tcaatgccat  caaaagggtgt  tcgtgaccaa  1320
tttattgatg  ctttaaatga  ttgggttacgt  gtccagatg  ttaaagttgg  taaaattaaa  1380
gatgctgttc  gtgttttaca  taattcatca  ttattattag  atgattttca  agataattca  1440
ccattacgtc  gtggtaaacc  atcaacacat  aatatttttg  gttcagctca  aacagttaat  1500
acagctacat  attcaattat  taaagctatt  ggtcaaatta  tggaaatttc  tgcgtggtgag  1560
tcagttcaag  aagttatgaa  ctcaattatg  attttatttc  aaggccaagc  tatggattta  1620
ttttggacat  ataatggtca  tgttccatca  gaagaagaat  attatcgtat  gattgaccaa  1680
aaaacaggtc  aattattttc  aattgctaca  tcatatttat  taaatgctgc  tgaataatgaa  1740
attccacgta  caaaaattca  atcatgttta  catcgtttaa  cacgtttatt  aggtcgttgt  1800
tttcaaattc  gtgatgattat  tcaaaattta  gtctctgctg  attacactaa  aaaaaagga  1860
ttctgtgaag  atttagatga  aggtaaatgg  tcattagctt  taattcacat  gattcataaa  1920
caacgttcac  acatggcttt  attaaatgtt  ttatcaacag  gtcgtaaaca  tgggggtatg  1980
acattagaac  aaaaacaatt  tgttttagat  attattgaag  aagaaaaatc  attagattat  2040
acacgttcag  ttatgatgga  tcttcatgtt  caattacgtg  ctgaaattgg  tcgtattgaa  2100
attttattag  attcaccaaa  tccagctatg  cgtttattat  tagaattatt  acgtgttacc  2160
ggtagtgctt  ggtcacaccc  tcaatttgag  aaataa  2196

```

```

<210> 26
<211> 1704
<212> DNA
<213> Artificial Sequence

```

```

<220>
<223> codon optimized phytase from E. coli

```

```

<220>
<221> misc_feature
<222> (1615)..(1701)
<223> tag

```

```

<400> 26
atggtagccaa  tgcgtatctc  tcttaaaaaa  tcaggaatgt  taaaacttgg  tttatctctt  60
gtagctatga  cagt agctgc  ttctgtacaa  gctaaaacat  tagtatattg  ttcagaaggc  120
tctccagaag  gttttaatcc  tcaacttttt  acttcaggtat  caacttatga  cgcttcttca  180
gtacctttat  acaaccgttt  agtagagttc  aaaatcggtat  caacagaagt  tattccaggt  240
ttagctgaaa  aatgggaagt  atcagaagac  ggtaaaactt  acacattcca  ttt aagaaaa  300

```

0802W01_ST25

ggt g t t a a t g g c a c g a t a a t a a g a g t t t a a c c a a c a a g a g a a t t a a t g c t g a t g a t 360
 g t a g t t t t c t c a t t t g a c c g t c a g a a a a t g c t c a a a a t c c a t a t c a t a a a g t t t c a g g a 420
 g g a t c t t a c g a a t a t t t c g a a g g c a t g g g a t t a c c a g a a c t t a t t t c t g a a g t t a a a a a 480
 g t t g a t g a t a a c a c a g t t c a a t t t g t t t t a a c t a g a c c a g a a g c t c c a t t t t t a g c t g a t 540
 t t a g c t a t g g a t t t c g c t a g t a t t t t a t c a a a g a a t a t g c a g a t g c t a t g a t g a a g c t 600
 g g t a c a c c t g a a a a a c t t g a t c t t a a t c c a a t t g g t a c t g g t c c a t t c c a a t t a c a a c a a 660
 t a t c a g a a a g a t t c a c g t a t t c g t t a c a a a g c a t t t g a c g g c t a t t g g g g t a c a a a a c c t 720
 c a a a t t g a t a c t t t a g t a t t t t c a a t t a c a c c a g a c g c a t c a g t t c g t t a c g c a a a a t t a 780
 c a g a a a a a t g a a t g c c a a g t a a t g c c a t a t c c a a a t c c a g c t g a t a t t g c a c g t a t g a a a 840
 c a a g a t a a a t c t a t c a a t t t a a t g g a a a t g c c t g g t t t a a a t g t t g g t t a t t t a t c a t a t 900
 a a c g t t c a a a a a a a c c a t t a g a t g a t g t a a a a g t t c g t c a a g c a t t a a c t t a t g c a g t t 960
 a a t a a g a c g c a a t c a t t a a a g c a g t a t a t c a a g g t g c t g g a g t t a g t g c t a a a a a t c t t 1020
 a t t c c a c c a a c a a t g t g g g g t t a c a a c g a t g a c g t t c a a g a t t a t a c t t a c g a c c c t g a g 1080
 a a a g c t a a a g c a t t a c t t a a a g a a g c a g g t t t a g a a a a a g g t t t c t c a a t t g a t t t a t g g 1140
 g c a a t g c c a g t a c a a c g t c c t t a c a a t c c a a a t g c t a g a c g t a t g g c a g a a a t g a t t c a a 1200
 g c a g a c t g g g c t a a a g t a g g t g t t c a a g c a a a a a t t g t t a c a t a c g a a t g g g g t g a a t a c 1260
 t t a a a a c g t g c t a a a g a t g g t g a a c a c c a a a c t g t t a t g a t g g g t t g g a c a g g t g a t a a t 1320
 g g t g a t c c t g a c a a t t t c t t t g c a a c a t t a t t t t c a t g t g c t g c t t c a g a a c a a g g t t c a 1380
 a a t t a t t c a a a a t g g t g t t a t a a a c c t t t t g a a g a c t t a a t t c a a c c t g c t c g t g c t a c a 1440
 g a c g a t c a c a a t a a a c g t g t t g a a t t a t a c a a a c a a g c a c a a g t t g t a a t g c a c g a c c a a 1500
 g c t c c a g c t c t t a t t a t t g c t c a t t c a a c a g t a t t c g a a c c a g t t a g a a a a g a a g t t a a a 1560
 g g t t a t g t a g t a g a t c c a t t a g g t a a a c a t c a c t t t g a a a a c g t a t c a a t t g a a g g t a c c 1620
 g g t g a c t a t a a a g a t g a t g a t g a c a a a a g t g g a g a g a a c t t a t a c t t t c a a g g t c a t a a t 1680
 c a c c g t c a c a a a c a c a c c g g t t a a 1704

<210> 27
 <211> 87
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> FLAG epitope tag linked to a MAT epitope tag by a TEV protease site

<400> 27
 g g t a c c g g t g a c t a t a a a g a t g a t g a t g a c a a a a g t g g a g a g a a c t t a t a c t t t c a a g g t 60
 c a t a a t c a c c g t c a c a a a c a c a c c g g t 87

0802W01_ST25

<210> 28
 <211> 660
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> modified chloramphenicol acetyl transferase gene from E. coli

<400> 28
 atggagaaaa aaatcactgg atat accacc gttagat atat cccaatggca t cgt aaagaa 60
 cat t t t t gagg cat t t cagtc agt t gct caa t gtacct at a accagaccgt t cagct ggat 120
 att acggcct t t t t aaagac cgt aaagaaa aat aagcaca agt t t t at cc ggcct t t at t 180
 cacattcttg cccgcctgat gaatgctcat ccggagt t cc gtat ggcaat gaaagacggt 240
 gagct ggt ga t at gggat ag t g t t caccct t g t t acaccg t t t t ccat ga gcaaact gaa 300
 acgt t t t cat cgct ct ggag t gaat accac gacgat t t cc ggcagt t t ct acacat at at 360
 t cgcaagat g t ggcgt g t t a cgggtaaaac ct ggcct at t t cctt aaagg gt t t at t gag 420
 aat at g t t t t t cgt cagcgc caat ccct gg gtgagt t t ca ccagt t t t ga t t t aaacgt g 480
 gccaat at gg acaact t ct t cgccccgt t t t cact at gg gcaaat at t a t acgcaaggc 540
 gacaagg t gc t gat gccgct ggcgat t cag gt t cat cat g ccgt t t gt ga t ggct t ccat 600
 gt cggcagaa t gct t aat ga at t acaacag t act gcgat g agt ggcaggg cggggcgt aa 660

<210> 29
 <211> 1260
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> modified erythromycin esterase gene from E. coli

<400> 29
 atgaggt t cg aagaatgggt caaagataag catat t cct t t caaact gaa t caccct gat 60
 gat aat t acg atgatt t t aa gccatt aaga aaaat aat t g gagat acccg agt t gt agca 120
 t t aggt gaaa at t ct cat t t cat aaaagaa t t t t t t t gt t acgacat ac gct t t t gcgt 180
 t t t t t at cg aagacct cgg t t t t act acg t t t gct t t t g aat t t ggt t t t gct gaggg 240
 caaat cat ca at aact ggat acat ggacaa ggaact gacg at gaaat agg cagat t ct t a 300
 aaacact t ct att at ccaga agagct caaa accacat t t c t at ggct aag ggagt acaat 360
 aaagcagcaa aagaaaaaat cacatt t ct t ggcat t gat a t acccagaaa t ggaggt t ca 420
 t act t accaa at at ggagat agt gcat gac t t t t t agaa cagcggat aa agaagcact a 480
 cacatt at cg at gat gcat t t aat at t gca aaaaagat t g at t act t ct c cacat cacag 540
 gcagcct t aa at t t acat ga gct aacagat t ct gagaaaat gccgt t t aac t agccaat t a 600
 gcacgagt aa aagt t gcct t gaagct at g gct ccaat t c acat t gaaaa at at gggat t 660

0802W01_ST25

gat aaat at g agacaat t ct gcat t at gcc aacggt at ga t at act t gga ct at aacat t 720
 caagct at gt cgggct t t at t t caggaggc ggaat gcagg gcgat at ggg t gcaaaagac 780
 aaat acat gg cagat t ct gt gct gt ggcat t t aaaaaacc cacaaagt ga gcagaaagt g 840
 at agt agt ag cacat aat gc acat at t caa aaaacaccca t t ct gt at ga t ggat t t ct a 900
 agt t gcct ac caat gggcca aagact t aaa aat gccat t g gt gat gat t a t at gt ct t t a 960
 ggt at t act t ct t at agt gg gcat act gca gccct ct at c cggaagt t ga t acaaaat at 1020
 ggt t t t cgag t t gat aact t ccaact gcag gaaccaaat g aaggt t ct gt cgagaaagct 1080
 at t t ct ggt t gt ggagt t ac t aat t ct t t t gt ct t t t t t a gaaat at t cc t gaagat t t a 1140
 caat ccat cc cgaacat gat t cgat t t gat t ct at t t aca t gaaagcaga act t gagaaa 1200
 gcat t cgat g gaat at t t ca aat t gaaaag t cat ct gt at ct gaggt cgt t t at gaat aa 1260

<210> 30

<211> 2844

<212> DNA

<213> Artificial Sequence

<220>

<223> modified genomic DNA from *S. dimorphus*

<400> 30

gcct t t aggt at t t caggaa ct t t caact t cat gaaaat a aacgt gacaa at aaaaacct 60
 t t t t at t t g aaaat t gggg aaat t gat ga aaaat t gaaa aaaat aaat a aaat at aaaa 120
 t act t t at ga caaaaaaat g aacaaaat t t gagat agaac aaat t gcaat t ct t t t t t c 180
 aaaaat t gtt t t t t caat aa t cat caaccg agct t at t t t t gt aagcagg t t cagagact 240
 aaaaaaat ag agaaat t t t t t t t gtt ccca aaaacagaaa aagaacaaa agat t t ct t c 300
 acct gt t t gt gagat agt cc gat ct t t t t t gaaaaaaaa gagaaact t a agct t t cat a 360
 acaaaat gga t cgt t t t cca agct gaacac aacat t t t aa t gcaccct t t ccacat gt t a 420
 ggt gt agct g gt gt at t t gg t ggt t cat t a t t ct ct gcaa t gcat aggt a ct gagaaact 480
 gt gccct t t g t t t t t t t gt g t t t t t t c t t ct t t t t t ct t t gt gcct gct ct ct t t t 540
 t t t ct acaaa caagt aaaaa aaaagcaaaa acgcat aaaa at gcacgcaa aaacgt at cg 600
 t t gt gcat ct aaaaagt at a aaaaagaaac aaaaacgct t t gct t cgct a ct t t t t t ct 660
 t ct t cat t t t t t gcaaagca aaaaat gat a aaaagt agcg aagcgt agca aaaaaagct t 720
 ct aaaaacaa t t gaaaat cg agt gaat t gc at aaat t t ac acat acaaaa aaat aaaaaa 780
 caagaat aaa aaat at gaaa acaacaaaag cact t gacga t t t t acaaaa aacgcgt caa 840
 acaagt at t t agat t t t at t gct ct ct gca aacaaaaagc agaagagt at t gaccagaac 900
 aacgcggaac ct gt acgt at caccat at t g t gcct cgcca t cact acaag gcaaat gaggc 960

0802W01-ST25

t c t c t t g g g a	a a a c t t t g a g	a g t c c a g a a a	a t t t a g t t t t	g g t g a c c t t t	g a a g a c c a c a	1020
t t a a g c a c a	c g a a c t t c g t	t t t g a a g t t t	a t g g t g a a a a	c g g g g a c a a t	a t t g c t g c t c	1080
t t c g a a t g g c	t g g t c a c a a g	g a a g a g a g c a	t g c g a g c t a t	g c a a c a a g c a	g g t g g t c a a g	1140
c a g t a a a c a a	a a t t t t c a a a	g a a a a a a a a c	a a c t a a t g c a	t a a t c a a g a a	t t t c a a a a a g	1200
a a a t g g c t c g	c c g c t c a a t g	g c c c g t c c a g	a t g c t c g t c a	a a t t c g c a g c	g a a g g a g g a a	1260
a g a a g g g t g c	g c a t g t t c g t	c a c c a a a a c c	g a a c t g t t c g	a g c a a c a g a t	c g a t t t c t t t	1320
g g c g t t t t a a	a a g a g a a g a c	t t t t t a t g t a	c c t t t g g t t t	t g a c a a t g c t	g g t g a t c t t t	1380
g t t g a g a a t t	g a a t t t g g c a	a a g c c a g a a a	t t t t t g t t g g	t c g t t t g a g c	g g t t t t c t t a	1440
c t g g a a a c g	a a a g a c g a a c	a g g g g t t g g t	c g t g t g a a a a	a a t t g a a g a a	t c t t c c g a a a	1500
a c g a a g a a c a	a a c a t g a c c c	a a t c c a c a c t	a a a t t t t t a c	a a a a a a g t g t	a c g c t c t t t c	1560
c t t t t t g t t t	t t g t g c a a a g	t t c t t t g t t t	t t a t t t t t t t	t t g t g a a a a a	t a t g c a a c g a	1620
a t t c c a g t t g	t a a a a g a c t g	g a c t c g t t c a	g a g a c t a g t t	t t a t t t t t a t	g a a a c a a a g t	1680
a c t c g a g a g a	a a a a a t a a g c	t t t a t t t t t t	c t t a t g a t a t	a g t c c t c a a g	a a a c t t t t t t	1740
c t t t t a a a a a	a a a a g a a a g t	t c t t g a a g g g	t t c a t t a g t t	a c t t c a t c t t	t a a t c c g t g a	1800
a a c a a c t g a a	a a c g a a t c t g	c t a a c g a a g g	a t a c a a a t t c	g g t c a a g a a g	a a g a a a c t t a	1860
c a a c a t c g t a	g c t g c a c a c g	g t t a c t t t g g	t c g t t t a a t c	t t c c a a t a c g	c t g c t t t c a a	1920
c a a c t c t a g a	t c a t t a c a c t	t c t t c t t a g c	t g c t t g g c c a	g t t g t a g g t a	t t t g g t t c a c	1980
t g c t t t a g g t	a t t t c a a c t a	t g g c t t t c a a	c c t t a a c g g t	t t c a a c t t c a	a c c a a t c a g t	2040
t g t t g a t t c a	c a a g g t c g t g	t t t t a a a c a c	a t g g g c t g a c	a t c a t c a a c c	g t g c t a a c t t	2100
a g g t a t g g a a	g t t a t g c a c g	a a c g t a a c g c	t c a c a a c t t c	c c a t t a g a c t	t a g c a t c t g t	2160
a g a a g c t c c t	t c t g t a a a c g	c t t a g t t t t a	t t t t t t a t g a	a a a a c t c a g g	c t t a a t t t a g	2220
g c t t g a g t t t	t t c a t t c t t t	t t g a a g c t c t	g a a a t t t t a a	a a t t t c t a g t	c t t c t t t a a t	2280
g t t t t t a a a t	t t t a a a a a a t	a a a t t t c t t c	t c t g c t g t g t	t t t t c t t t t t	t t t t g a a a a a	2340
a c a a g a a a a a	a a a a t t t t t t	t g t t t t c t t c	t t t g t t t t t t	t a t t t c t t t t	t g t t t t g t t t	2400
a t t t t t t a g t	t t c a g a a t c t	t t g a t t c a a a	a a a a a t t t a	g t c c g a t t a c	t c c a t a g g a g	2460
c a a g c a g t a a	a a a a t a a a a a	c t g t a a t a a a	a a a t a a a a c a	a a a a t t t t a t	t t c t t t t t g t	2520
t t t g c t t g a a	c t t t t c a a a a	a a a a a t t g a a	a a a t t c a a g c	a a a a c a a a a a	g a a a c a a a t a	2580
a a a a a t t t a t	g a a t t t t c t a	c t t t t t c a g g	a g t t g a a a t t	t c t c c t t t a c	t t a a a a c a t a	2640
t t t t g c t a a a	a a a a g c g c t t	g t g t t g c t t t	t t t t g c t a c t	t t t t g t t t c c	a a g c a t t t t t	2700
t c g a a t a t t t	t t t t t t g a t t	t t g a t g t g c g	t t t t t g t t a a	c c t a a a a t c t	t g a a a a g a t t	2760
t a c t c t t t t c	a a a t t t t t a t	g t t t t t a t t t	t t t t t a t t c a	t a a a a a a a a a	c a a t a c a t a a	2820
a a a t a a a g t a	t t t c g g c t t c	a a a a				2844

0802W01_ST25

<210> 31
 <211> 1284
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> codon optimized cytosine deaminase from E. coli

<400> 31
 atgtct aaca acgctttaca aactattatc aacgcaagat taccaggt aa agaaggttta 60
 tggcaaatcc accttcaaga tggtaaaatt tctgctattg acgcacaatc aggtgtaatg 120
 ccaataacag aaaattcatt agatgctgaa caaggcttag ttattcctcc ttctgttgag 180
 ccacatat ac accttgacac aactcaaaca gctggtcaac caaattggaa tcaatcaggc 240
 acactttttg aaggtattga acgttgggct gaggctaaag cattattaac acatgacgat 300
 gttaaacaac gtgcttggca aacattaaaa tggcaaattg caaatggtat acaacatgtt 360
 cgtactcatgtttgacgtttc agacgcaact ttaacagcat taaaagctat gttagaagtt 420
 aaacaagaag tagctccatg gatagactta caaattgt ag cttttccaca agaggggtatt 480
 ttatcatatc caaatggcga agcattatta gaagaggctt tacgtcttgg t gctgacgt a 540
 gt aggtgct a tt cctcactt cgaatttact cgtgaat at g gtgt agaaag ttt acat aaa 600
 actttcgcac ttgctcaaaa atatgatcgt cttattgatg tt cactgtga tgaaattgat 660
 gacgaacaat cacgtttcgt tgaaacagtt gctgcat tag ct caccgtga aggtatgggt 720
 gctcgtgtta cagctagtca tactacagct atgcatctt ataatggagc atacacttct 780
 cgtcttttta gat t actt aa aatgtcaggt att aacttcg ttgcaaatcc tttagt aaac 840
 attcacttac aaggtagatt tgat acat at cctaaacgt a gaggtattac acgtgtt aaa 900
 gaaatgttag aatctggtat caatgtatgt tttggtcacg atgatgtatt cgatccatgg 960
 tatccattag gcacagctaa tatgtt acaa gt tttacaca tgggtttaca tgtttgtcaa 1020
 cttatgggtt acgggtcaaat caacgatggc ctt aactt aa ttacacat ca ttctgcacgt 1080
 actttaaatc ttcaagatt a tggat t gca gctggt aatt cagctaacct t attattctt 1140
 ccagcagaaa atggttttga t gctcttcgt cgtcaagt ac ctgttcgtta ttctgttcgt 1200
 ggcggcaaag ttattgctag tacacaacct gcacaaacaa ctgtatattt agaacagcca 1260
 gaagctattg attacaaacg ttaa 1284

<210> 32
 <211> 1653
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> codon optimized betaine aldehyde dehydrogenase from S. oleracea

<220>

<221> mi sc_f eat ur e

<222> (1510)..(1650)

<223> t ag

<400> 32

at gat ggct t t cccaat t cc agct cgt caa t t at t cat t g acggt gaat g gcgt gaacca	60
t t at t aaaaa accgt at t cc aat t at t aac ccat caacag aagaaat t at t ggt gacat t	120
ccagct gct a cagct gaaga cgt agaagt a gct gt agt ag ct gct cgt aa agct t t caaa	180
cgt aacaaag gt cgt gact g ggct gct t t a t ggt cacacc gt gct aaat a ct t acgt gct	240
at t gct gct a aaat t acaga aaaaaagac cact t cgt aa aat t agaaac at t agact ca	300
ggg aaaccac gt gacgaagc t gt at t agac at t gacgacg t agct acat g t t t cgaat ac	360
t t cgaat act t cgct ggt ca agct gaagct t t agacgct a aacaaaaagc t ccagt aaca	420
t t accaat gg aacgt t t caa at cacacgt a t t acgt caac caat t ggt gt agt aggt t t a	480
at t t caccat ggaact accc at t at t aat g gacacat gga aaat t gct cc agct t t agct	540
gct ggt t gt a caacagt at t aaaaccat ca gaat t agct t cagt aacat g t t t agaat t c	600
ggg gaagt at gt aacgaagt aggt t t acca ccaggt gt at t aaacat t t t aacaggt t t a	660
ggg ccagacg ct ggt gct cc aat t gt at ca caccagaca t t gacaaagt agct t t caca	720
ggg t cat cag ct acaggt t c aaaaaat t at g gct t cagct g ct caat t agt aaaaccagt a	780
acat t agaat t aggt ggt aa at caccagt a at t at gt t cg aagacat t ga cat t gaaaca	840
gct gt agaat ggacat t at t cggg gt at t c t ggacaaacg gt caaat t t g t t cagct aca	900
t cagct t t at t agt acacga at caat t gct gct gaat t cg t agaccgt at ggt aaaaat gg	960
acaaaaaaca t t aaaaat t t c agacccat t c gaagaaggt t gt cgt t t agg t ccagt aat t	1020
t caaaaggt c aat acgacaa aat t at gaaa t t cat t t caa cagct aaat c agaaggt gct	1080
acaat t t t at gt ggt ggt t c acgt ccagaa cact t aaaaa aaggt t act a cat t gaacca	1140
acaat t at t a cagacat t ac aacat caat g caaat t t gga aagaagaagt at t cggg cca	1200
gt aat t t gt g t aaaaacat t caaaacagaa gacgaagct a t t gaat t agc t aacgacaca	1260
gaat acggt t t agct ggt gc t gt at t ct ca aaagact t ag aacgt t gt ga acgt gt aaca	1320
aaagct t t ag aagt aggt gc t gt at gggg a aact gt t cac aaccat gt t t cgt acacgct	1380
ccat ggggt g gt gt aaaacg t t caggt t t c ggt cgt gaat t aggt gaat g gggg at t gaa	1440
aact act t aa acat t aaaca agt aacat ca gacat t t cag acgaaccat g gggg t ggt ac	1500
aaat caccaa cgggt t accc at acgacgt a cct gact at g ct t accct t a cgacgt acca	1560
gact at gct t at ccat acga cgt accagac t acgct gaaa act t at act t ccaaggt cac	1620
caccaccacc accat cacca cccaccaggt t aa	1653

0802W01_ST25

<210> 33
 <211> 141
 <212> DNA
 <213> Artificial Sequence

<220>

<223> 3xHA tag linked to a 6xHIS tag by a TEV protease site

<400> 33
 accggttacc cat acgacgt acctgactat gcttaccctt acgacgtacc agactatgct 60
 tatccat acg acgt accaga ctacgctgaa aacttat act tccaaggtca ccaccaccac 120
 caccatcacc acccaccagg t 141

<210> 34
 <211> 1647
 <212> DNA
 <213> Artificial Sequence

<220>

<223> codon optimized betaine aldehyde dehydrogenase from B. vulgaris

<220>
 <221> misc_feature
 <222> (1504)..(1644)
 <223> tag

<400> 34
 atgatgtcaa tgccaat tcc atcacgtcaa ttattcat t g acggtgaatg gcgtgaacca 60
 attaaaaaaa accgtattcc aattattaac ccatcaaacg aagaaattat tgggtgacatt 120
 ccagctggtt catcagaaga cat tgaagt a gctgt agct g ctgctcgtcg t gctttaaaa 180
 cgt aacaaag gt cgt gaat g ggctgctaca t caggtgctc accgtgctcg t t actt acgt 240
 gctattgctg ct aaagt aac agaacgt aaa gaccact t cg taaaatt aga aacaatt gac 300
 t caggt aaac cat t cgacga agctgtatt a gacatt gacg acgt agct ac atgtttcgaa 360
 t actt cgct g gt caagct ga agct at ggac gct aaacaaa aagct ccagt aacatt acca 420
 at ggaacgt t t caaat caca cgt att acgt caaccaatt g gtgt agt agg ttt aatt aca 480
 ccat ggaact acccatt att aat ggctaca t ggaaaatt g ct ccagcttt agct gct ggt 540
 t gt acagct g t att aaaacc atcagaatt a gctt caatt a catgttt aga att cggt gaa 600
 gt atgt aacg aagt aggttt accaccaggt gt att aaaca t t gt aacagg ttt aggt cca 660
 gacgct ggt g ctccatt agc t gct caccca gacgt agaca aagt agcttt cacaggt t ca 720
 t cagct acag gtt caaaagt aat ggctt ca gct gct caat t agt aaaacc agt aacatt a 780
 gaatt aggt g gt aaat cacc aattattgt a t t cgaagacg tagacatt ga ccaagt agt a 840
 gaat ggacaa t gtt cggtt g ttt ctggaca aacggt caaa ttt gtt cagc t acat cacgt 900

0802W01_ST25

```

t t a t t a g t a c a c g a a t c a a t t g c t g c t g a a t t c a t t g a c c g t t t a g t a a a a t g g a c a a a a 960
a a c a t t a a a a t t t c a g a c c c a t t c g a a g a a g g t t g t c g t t t a g g t c c a g t a a t t t c a a a a 1020
g g t c a a t a c g a c a a a a t t a t g a a t t c a t t t c a a c a g c t a a a t c a g a a g g t g c t a c a a t t 1080
t t a t g t g g t g g t t c a c g t c c a g a a c a c t t a a a a a a g g t t a c t t c a t t g a a c c a a c a a t t 1140
a t t t c a g a c a t t t c a a c a t c a a t g c a a a t t t g g c g t g a a g a a g t a t t c g g t c c a g t a t t a 1200
t g t g t a a a a a c a t t c t c a t c a g a a g a c g a a g c t t t a g a c t t a g c t a a c g a c a c a g a a t a c 1260
g g t t t a g c t t c a g c t g t a t t c t c a a a g a c t t a g a a c g t t g t g a a c g t g t a t c a a a a t t a 1320
t t a g a a t c a g g t g c t g t a t g g g t a a a c t g t t c a c a a c c a t g t t t c g t a c a c g c t c c a t g g 1380
g g t g g t a t t a a a c g t t c a g g t t t c g g t c g t g a a t t a g g t g a a t g g g g t a t t g a a a a c t a c 1440
t t a a a c a t t a a a c a a g t a a c a t c a g a c a t t t c a a a c g a a c c a t g g g g t t g g t a c a a a t c a 1500
c c a a c c g g t t a c c c a t a c g a c g t a c c t g a c t a t g c t t a c c c t t a c g a c g t a c c a g a c t a t 1560
g c t t a t c c a t a c g a c g t a c c a g a c t a c g c t g a a a a c t t a t a c t t c c a a g g t c a c c a c c a c 1620
c a c c a c c a t c a c c a c c c a c c a g g t t a a 1647

```

```

<210> 35
<211> 2541
<212> DNA
<213> Artificial Sequence

```

```

<220>
<223> codon optimized E-alpha-biabolene synthase from A. grandis

```

```

<220>
<221> misc_feature
<222> (2458)..(2538)
<223> tag

```

```

<400> 35
a t g g t a c c a g c a g g t g t a t c a g c t g t g t c a a a a g t t t c t t c a t t a g t a t g t g a c t t a a g t 60
a g t a c t a g c g g c t t a a t t c g t a g a a c t g c a a a t c c t c a c c c t a a t g t a t g g g g t t a t g a c 120
t t a g t t c a t t c t t t a a a a t c t c c a t a t a t t g a t a g t a g c t a t c g t g a a c g t g c t g a a g t g 180
c t t g t a a g t g a a a t a a a a g c t a t g t t a a a t c c a g c a a t t a c t g g a g a t g g t g a a t c a a t g 240
a t t a c a c c t t c a g c t t a t g a c a c t g c t t g g g t t g c a c g t g t a c c a g c a a t t g a t g g t a g c 300
g c a c g t c c a c a a t t t c c a c a a a c a g t a g a t t g g a t t t t a a a g a a t c a a t t a a a a g a t g g t 360
t c t t g g g g t a t t c a a t c a c a c t t t t a c t t t c a g a c c g t t t a t t a g c t a c t c t t a g c t g t 420
g t t t t a g t t t t a c t t a a a t g g a a t g t t g g t g a t t t a c a g g t t g a g c a a g g t a t t g a g t t t 480
a t t a a g t c a a a c c t t g a a t t a g t a a a a g a t g a a a c t g a t c a a g a t t c t t t a g t g a c t g a t 540
t t t g a g a t t a t t t t c c c t a g c t t a c t t c g t g a g g c c c a a a g t t t a c g t t t a g g t c t t c c a 600
t a c g a t t t a c c t t a c a t c c a c t t a t t a c a a a c a a a a c g t c a g g a a c g t t t a g c a a a a t t a 660

```

0802W01_ST25

agccgt gaag aaat at at gc agt tccaagt ccact t t t at at t c t t t aga ggg t at t caa	720
gat at t g t t g agt gggaacg t at t at ggaa gt acaat c t c aggat ggat c at t t t t aagt	780
t c t ccagcat caaccgcat g t g t t t t at g cat acaggt g acgct aagt g t t t agaat t t	840
c t t aacagt g t aat gat t aa g t t t ggt aat t t t gt accat gcct t t at cc t gt agat t t a	900
t t agaacgt t t act t at agt agat aat at a g t t cgt c t t g gt at t t accg t cact t cgaa	960
aaagaaat t a aagaagcat t agat t at gt a t at cgccat t ggaat gaacg t ggt at t ggt	1020
t ggggt cgt t t aaat ccaat t gct gact t a gaaacaact g c t t t aggt t t t cgt t t at t a	1080
cgt t t acacc g t t at aat gt at c t ccagca at c t t t gat a at t t caaaga t gccaat ggc	1140
aaat t cat t t gt agcact gg t cagt t t aat aaggat gt gg c t t caat gt t aaact t at ac	1200
cgt gcat cac aat t agcat t cccaggcgaa aacat t t t ag at gaagct aa at c t t t t gcc	1260
accaaat act t acgt gaagc cct t gaaaaa t c t gaaact t cat cagct t g gaacaat aaa	1320
cagaat t t aa gt caagaaat caagt at gca t t aaaaaact t cat ggcacgc t t c t gt acca	1380
cgt gt t gaag caaaacgt t a t t gt caagt t t at cgt cct g at t acgt cg t at t gct aag	1440
t gt gt at aca aat t accat a cgt t aacaac gaaaaat t c t t agaat t agg t aaat t agat	1500
t t t aacat ca t t caat caat t cat caagaa gaaat gaaaa at gt gacaag t t ggt t t cgt	1560
gat t c t ggct t accat t at t t act t t cgct cgcgaaacgt c c t t t agaat t t t act t c t t a	1620
gt t gct gct g gt act t at ga acct caat at gct aaat gt c gt t t c t t at t cacaaaagt a	1680
gct t gt c t t c aaacagt at t agacgat at g t acgat act t acggt act t t agacgaat t a	1740
aaact t t t t a ccgaggct gt gcgt cgt t gg gat t t at c t t t t acagaaaa t t t acct gac	1800
t at at gaaat t at gt t at ca aat c t at t at gacat cgt t c at gaagt ggc t t gggaagct	1860
gaaaaagaac aaggt agaga at t agt gt ca t t c t t ccgt a aaggct ggga agact act t a	1920
t t aggt t act at gaagaagc agaat ggt t a gcagcagaat acgt t ccaac at t agat gaa	1980
t acat t aaaa acggt at t ac at caat cggc caacgt at c t t at t act t t c aggt gt gt t a	2040
at t at ggat g gccaaact t t t at cacaagaa gcat t agaaa aagt t gat t a ccct ggt cgt	2100
cgt gt t t t aa c t gagt t aaa c t cact t at t agccgt t t ag c t gacgacac t aaaact t at	2160
aaagcagaaa aagct cgt gg agaat t agcc t cat caat t g aat gct acat gaaagat cat	2220
cct gaat gt a cagaagaaga agcct t agac cacat t t at t c t at t c t t ga accagccgt a	2280
aaagaat t aa c t cgt gaat t t c t t aaacca gacgacgt t c cat t t gct t g t aaaaagat g	2340
t t at t cgaag aaact cgt gt t acaat ggt g at c t t t aaag at ggt gat gg t t t t ggt gt a	2400
t c t aagt t ag aagt t aaaga t cacat caaa gaat gct t aa t t gaaccat t accat t aggt	2460
accggt gaaa act t at act t t caaggct ca ggt ggcggt g gaagt gat t a caaagat gat	2520

gat gat aaag gaaccggtt a a

2541

<210> 36
 <211> 542
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> modified endogenous promoter from the psbA gene of *S. dimorphus*

<400> 36
 tcgttgagta gtttttcaga ttaattgcta tgcaccccat gtgaattaaa aatataacat 60
 aaatttcaaa aatgtcaatt tttagtctaa aaaatattat tcatgtttt tttatgacaa 120
 ttttttttaa atttttcaat aaaaacaaaa atataattat aaagaataca taaaaaatc 180
 aaaaattcat aataaatca ccaaaaaatt tattttttaa ttttgattt caatattttt 240
 atttgaattt aaaaattaat tttttaaatt tttatttat atatttgaat ttatatttca 300
 gttttattt aacttaagtt ttcaaatcat aaatttaata gttaatattt ttttaactc 360
 taaattattt atctttaaaa ttcaaatct ttaacacttg aaattataaa ctctattgtt 420
 tttgtttgaa ttttttttta agtttgaaaa ctttataaaa ttaataaac taaatagaat 480
 tttgaattgt ataaaaatta aaatgaaaag tttgtgtttt tcagattaaa tgtagcaacc 540
 aa 542

<210> 37
 <211> 559
 <212> DNA
 <213> *Scenedesmus dimorphus*

<400> 37
 atttaattctg aaaaacacaa acttttcatt ttaattttta tacaattcaa aattctattt 60
 agtttattta attttataaa gttttcaaac ttaaaaaaaaa attcaaaca aaacaatgaa 120
 gtttatatt tcaagtgtta aagatttgaa attttaaga ttaataattt agagttttaa 180
 aaaatattaa ctattaaatt tatgatttga aaacttaagt ttaataaaaa ctgaagtata 240
 aattcaaat atataaat aaa aattttaaat aattaaattt ttaattcaaa taaaaaatat 300
 gaaatcaaat attaaaaaat aaattttttg gtgatttatt tatgaatttt tgattttttt 360
 atgtattctt taaataata tttttgtttt tattgaaaaa tttaaaaaaa attgtcataa 420
 aaaaacatcg aataatattt tttagactaa aaattgacat ttttgaaatt tatgttatat 480
 ttttaattaa caagagttca attgaattca tgaaaaatct ttttaagatt cggagaattt 540
 taaattttta tttttttat 559

<210> 38
 <211> 716
 <212> DNA

<213> Artificial Sequence

<220>

<223> modified endogenous promoter for the psbB gene of *S. dimorphus*

<400> 38

```

ttaaattact ttatttatgg aaaaatgtat tttttcgatt tctagagttt ttttacactt      60
tttgaattgt gtgttttttt cttttctaaa agaaaaaaga caattgaaag tcttgtttct      120
tcatttttat tattttatca taaatctaga atttttaaaa tattttttat tttttaccgc      180
ggagcggact aaattttttt taaacaatat tttttaattg aaaacatttt ttttcttcaa      240
aatataat aa taaaatttga aaaaaaagaa aaacaaaaaa taaaagcttt cctctgctta      300
agttgtat tt tttatgttt ttattttttt tgaaagtttt gaaaaaagta caaaaaagtt      360
taaatttttt ttcatttttt tcacattttc tttatcatat tatcgtctta agtttttttc      420
ttttttttca gtttttttta agaaactgaa aaaaaagaga aaaacttaag acaataaaaa      480
agattttttt aacagaaaaa agtttctttt ttaaataaaa tgaaaacaaa agttgtaggg      540
aaccaaakat ttatttaatt cataaagaga gtacaactta agcagaggaa agcttttatt      600
ttttgttttt cttttttttc aatttttatt aatataat tt gaagaaaaaa aatgtttttc      660
aattaaaaaa tatgttttaa aaaaaaattt aatcccgctc cgcggtaac agt aaa      716

```

<210> 39

<211> 550

<212> DNA

<213> *Scenedesmus dimorphus*

<400> 39

```

aacttttgtt ttcattttat ttaaaaaaga aacttttttc tgttaaaaaa atctttttta      60
ttgtcttaag tttttctctt ttttttcagt ttcttaaaaa aaactgaaaa aaaagaaaaa      120
aacttaagac gataatatga taaagaaaat gtgaaaaaaa tgaaaaaaa tttaaacttt      180
tttgtacttt tttcaaaact ttcaaaaaaa ataaaaacat aaaaaaat ac aacttaagca      240
gaggaaagct tttatttttt gttttcttt tttttcaaat tttattatta tttttgaag      300
aaaaaaaatg ttttcaatta aaaaatattg tttaaaaaaa atttagaaaa aaataaaaaa      360
tattttaaaa attctagatt tatgataaaa taataaaaat gaagaaacaa gactttcaat      420
tgtctttttt cttttagaaa agaaaaaac acacaattca aaaagtgt aa aaaaactcta      480
gaaatcgaaa aaatacattt ttccat aaat aaagt aatt a agt aaacaaa aattcttttt      540
tcattaat tt      550

```

<210> 40

<211> 597

<212> DNA

<213> *Scenedesmus dimorphus*

0802W01_ST25

<400> 40
 tttttgaatt agat aaat ga gtgttctcaa ttttttttc tttgcat ttt ttgtttgtgt 60
 t gat t t acaa aaacaat aga aaaaagaaaa caat at t t t c t t t c t aaaaa aaaacaaaat 120
 t gat gaaaaa t agacat gaa caaaaaat t t gaaagt t ga c t t t t t a a a a a t t t t g g 180
 t at aat acaa aaaaagaat t t t t g g a a a g g t g g c a g a g t g g t t g a a t g c t c t g g t t t t g a 240
 aaaccagcgt ggct t t t a c g g t c a c c g g g g g t t c g a a t c c c t c c c t t t c c g at aat at at a 300
 caaaaat t t t t a a a g t t t t t t g t t t a t t t t g t at agat aa aaaaat c t g c a at aaaaat t t 360
 c g t t t t t t a t t t a t t c a a a a a t t c t g t t t t t t t g a a a a g a a a t a a a a a a a a t g c c a a a a 420
 g t g a g t t t t t t a t t c a a a t a t t a g a a a a a g t t t t t g a a a a a t t t a a a a a a a t a g a a a a a a 480
 t t t t t t t a t t t t t t c a t a a t t t a a a a a t t a t g t t a t a a t t t a a a t t a c a a a t a g g t t t 540
 t a t t a a a a a a t t t t t a c g t a c a g a t g a a t t c t a t a a a a t t a t t t t g g a g a t c a c c a t 597

<210> 41
 <211> 537
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> modified endogenous promoter for the tufA gene of *S. dimorphus*

<400> 41
 t t a a a c c a a t t t t c c a a g t a a c t t t a c t t t a t c a a a a a t t a a a a a a t t a a a a a c t t t t a 60
 t t g a a c t t a a a a t a a a t t t t t a a c a a a a t t a t t t t t a a a a a a g a a a a a a t t t t t t t a 120
 t t t t g g t t t t a t t t a t t t c t t t t t t t a c a a c a a a a a t t t t t t a a a c a g a a t a a t a a 180
 a a a a a t t t t a t t t a a g a a t g g t t t t t t a a t a t t t t g c t c a t g a c a a a t g a t t t t t a c 240
 t a c t t t t a t g c t t t t t t t c a a a a a a g c a g c a a a g c a a a a a g t t a t a a a a a g t g t a t g g 300
 a g c a a a c a g t t a a a t t g a c a c t t t t a a a a g t a t t a t a g g c c c a a c c g g a c t t g a a c c g 360
 a t g a c c t a t t g c t t g t a a g g c a a t c a c t c t a c c a a c t g a g t t a t g g g c c t a a a a a t a t t 420
 a t t t a t a t t t t a t a a t a g a a t a t a a a a t c t a a c a a c t t c t t t a t c c t c c t c t t t t c t c a c 480
 c t a t c t t t t t t g g t t g g g g g t a a g t g a a a a a g g a a a t g g t t g c t c g c a a c g a a a c c c 537

<210> 42
 <211> 492
 <212> DNA
 <213> *Scenedesmus dimorphus*

<400> 42
 t a a a g a a g t t g t t a g a t t t t a t a t t c t a t t a t a a a a t a t a a a t a a t a t t t t t t a g g c c c a 60
 t a a c t c a g t t g g t a g a g t g a t t g c c t t a c a a g c a a t a g g t c a t c g g t t c a a g t c c g g t t g 120
 g g c c t a t a a a t a c t t t t a a a a a g t g t c a a t t t a a c c g c t t g c t c c a t a c a c t t t t t a t a a 180
 c t t t t t t g c t t t g c t g c t t t t t t a a a a a a a a g c a t a a a a g t a g t a a a a a t c a t t t g t c 240

V01

atgagcaaaa	tattaaaaaa	ccattcttta	aataaaattt	tttttattat	tctgtttaaa	300
aaaatttttg	tttgtaaaaa	aaaagaaata	aataaaacca	aaataaaaaa	attttttctt	360
ttttttaaaa	taaattttgt	taaaaatttt	attttaagtt	caataaaagt	tttttaattt	420
tttaattttt	gataaagtaa	agttacttgg	aaaattgggt	tatacagaaa	aaattataaa	480
ttattttattc	at					492

<210>	43
<211>	557
<212>	DNA
<213>	Artificial Sequence

<220>
<223> modified endogenous promoter of the rpoA of *S. di morphus*

<400> 43	
cttgtttttt tttctaaatt tacccttaac ttgaaagatc taaaatgaca aaaatttttt	60
ttagagagca agttttactt ttgttgtcta aaaaactcaa attacacttt gtcttaagaa	120
ccttttttta ttaaaaaaca aaattttttt attttgtttc tttttattca tttgaaaaaa	180
aagttttcta aacttaactg ataattcaag aaaaaaatat tgcctgctac tagattcgaa	240
ctagtacttt tccgcgtatg aaacggatgc tctggccaac tgagctaac aggcttaat	300
tcatatatta taaaaaattt ttttcaacaa aaacaagttt ttttttcgaa ttttgaatt	360
ttttgattat ttttgtttgt ttgaatattc caaacaaaaa taaagttttt ctctttttt	420
ttactttttt cctgcttcca aatactaaaa aagtttaata gaaaaactct ttttttgaaa	480
gttgtcaaat ttttcttttc aaaaaattat ttttttgttg ggctattttt gaaaataaat	540
gtgagctcgg aaaaaaaa	557

<210>	44
<211>	532
<212>	DNA
<213>	Scenedesmus dimorphus

[illegible]

0802W01_ST25

aat t t t t g t c a t t t t a g a t c t t t c a a g t t a a a g g t a a a t t t a g a a a a a a a a a 532

<210> 45

<211> 483

<212> DNA

<213> *Scenedesmus dimorphus*

<400> 45

aaaagaaatt t t t t t t t t t t c c t t t a t t t t t t g t a t t c c a a a a a a t a t t t t g t t t t t c t 60

t c a a t a c a g t t t t t t t g c t a t t g c t t t t t c t c a c t t t t t t g c t t t g c t g c t t t t t t t t t 120

a a g c g t a a a a g t g a a a a a a a a g t t t g t a a a a c a a a a a a t t g a a a a a a a a g a t t g a t t 180

t t t t a t a c a g t a a a a a t t t t t t g t t t t t t a a a t a t t t a t t t a a a t t c a a c t c a a t t t c a a 240

a a a a a a a t g t c t t t g t t g a a t g c a a a t t t t t t g a a c a a a c a c a a t t c a a a a t a a a a c a 300

a t t t t t g g t t t a a a a t t c a a a a t t t t a a a t t c a t c t a c a g a a a a t t c t t c c a t g a g t a t t 360

t t t a a a a c t t t a a a g g g a g t t t g t t c t g g a a a a g a a a t a t a c t t t g c a a a a g t a c a c c t 420

c g t c c t a t g a a t t c t g a t t t t t a a a a a a t a a t a g t t c t a t t c c t t c t c a a a a c a a a a a 480

g g a 483

<210> 46

<211> 669

<212> DNA

<213> Artificial Sequence

<220>

<223> modified endogenous promoter of the *ftsH* gene in *S. dimorphus*

<400> 46

a g t a g a a a a a a t g a t g t a c t t g t a t a a a t a a a a a a a a t t a t t a t a c a a a a a t t t t t t 60

t t t a t t t t g t t t t a c t t t t t t c a t t t t t t t g c t t t g c t g t t t t t a t g a a t t t t t t a c 120

t t t a c t g c t t t t t t c t t c t t a c t t t t t a c g c t t t t t a t c a c t t t t t t g a t t t g c t g c t t 180

t t t t t c t t c t t c a t t t t t t g c t t g c t g c t t t t t t t c t t c t t c a t t t t t t g c t t t g c a a a 240

a a a t g a t a a a a g c g t a a a a t g a t a a a a a g c g t a a a a a a g c g t a a a a t g a t a a 300

a a a g c g t a a a a g t a a a a a a g a a a a a a g c a g c a a a g t a a a a a g t g a t a a a a g t t t a a a 360

a g t g c t a a a a a g c g t a a g t a a a c a a a a a a a g c a a a a g a c a a a a a a a g t t t t c a a a a a a a 420

a a t t t t t a t t t t t g t t t t t t a t a a a a a t t t t a a c a a g t t t t a t t t t c t a a a t a a a a c a a 480

a a a a t t g t g a g t t t t a a a a a a a a c g a a a a a g t t t c a a a a a t t t t t t t t g a t g g a c a a a a a g 540

a a a g a a a t t t t a a a t t t a t t a a t t t t t c a a a t a a c t a a t a a a t t t t g t t c a a c a g g a g a g 600

a a c t a t t g a a a a a a g c c t a c t a t c g g a g t t g a a c c g a t a a c c t t c g a t t t a g c t a g c a a 660

c c a a c a a a c 669

0802W01_ST25

<210> 47
 <211> 557
 <212> DNA
 <213> Scenedesmus dimorphus

<400> 47
 ctgtttcaca acgtttgcat ccaacgcaat cttcagtacg tggtgctgaa gccat tt gat 60
 ttgctttaca accattccat ggaaccattt ctaaaacatc taaatggacaa gcacgtacac 120
 attgtgtaca accaatacaa gtatcataaa tttttacaat atgtgacatt ttttattttt 180
 aatagaacta ttcaaaagaa gattgagaat ttcaaaaaag aattttttta catgaaaaac 240
 gattttttag aaattttact ataaaaaagt aaaaaaaaaat attaaat ttt gt agt ttttt 300
 gcaggagcag gat ttgaacc tgcgacattg ggattat gag cccacagagc taccagactg 360
 ctctatcctg cgttttttgt ttaaaaaagt tttattcgt a aaaaaaat ttt cttttatata 420
 tattacttta ttgattttta tttgtcaaat tttttttgt ttgaattttt atttcaaatt 480
 ttttaaatat agaaaaaatt tttttttatg ataatataaa actacagttt gaattataaa 540
 taaaaaaaaa aaatgaa 557

<210> 48
 <211> 580
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> modified endogenous promoter of the rbcL gene in S. dimorphus

<400> 48
 tttttttggg gttggttacg gtttttttta cttttgcttt ttttgctttt gttcaaagaa 60
 aaaaaaat ac aaaataaaa aaactaaaat gaaaaaaciaa agaattctaa aattcataaa 120
 aaaaattaaa acccaatttt ttttttggaa acttttccaa ataataaaa aatcaaaaaa 180
 aaatttttct agtatttttt tcatattttg aaactttttt tgagtttata aaaaaataga 240
 aaaaacaaat agatgaaaat tttagaaaaat tataaaccaa taaaaatgaa gttttgcgta 300
 gaaaaaaaat ttagtttact tgttcccaa gagcaagtgg taactttgaa aaaaatattt 360
 aaacttaaaa atttgctaaa gttttgaatt tatgttaaaa ttttaaaaaa ataaaaattt 420
 ttaaaactatt tttttatgtt aaaaaaatag tttttattat tttctataat atagtttagt 480
 tttttatttt tttcaatttc tttttttttt ttc aaagaaa aaagt tttcc acggatagat 540
 ttttatagga tgcacaaaat gttctatgaa ctttaaaaaa 580

<210> 49
 <211> 555
 <212> DNA
 <213> Scenedesmus dimorphus

<400> 49

0802W01_ST25

```

ggttttttttt acttttgctt tttttgcttt tgttcaaaga aaaaaaat a caaaat aaaa      60
aaaact aaaa tgaaaaaaca aagaattcta aaattcat aa aaaaaatt aa aacccaattt      120
tttttttggg aacttttcca aat aat aaaa aaat caaaaa aaaaatttttc tagt attttt      180
ttcat atttt gaaacttttt ttgagtttat aaaaaaat ag aaaaaacaaa tagat gaaaa      240
tttagaaaaa ttat aaacca ataaaaat ga agttttgcgt agaaaaaaaaa tttagtttac      300
ttgttcccca agagcaagt gtaactttga aaaaaat att taaactt aaa aatttgctaa      360
agttttgaat ttatgtt aaa atttaaaaaa aat aaaaatt ttt aaact at tttttatgt      420
taaaaaaat a gtttttatta ttttctataa tat agtttag tttttat tt ttttcaattt      480
cttttttttt ticaagaaa aaagttttcc acggat agat ttttat agga t cgacaaaat      540
gttctatgaa ctttt                                     555

```

<210> 50
 <211> 532
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> modified endogenous promoter for the chl B gene from S. dimorphus

```

<400> 50
cccagtcttc agt acgt ggt gct gaagcca tttgatttgc tttacaacca ttccatggaa      60
ccatttctaa aacatctaat ggacaagcac gtacacattg tgtacaacca atacaagtat      120
cataaat ttt tacaatatgt gacatttttt atttttaata gaactattca aaagaagatt      180
gagaatttca aaaaagaatt tttttttaca tgaaaaacga ttttttagaa attttactat      240
aaaaaagt aa aaaaaaat at taaattttgt agttttttgc aggagcagga ttigaacctg      300
cgacattggg attatgagcc ccacgagcta ccagactgct ctatcctgcg tttttgttt      360
taaaaagt ttt tattcgt aaa aaaaatttct tttat atata ttactttatt gat ttttatt      420
tgtcaaat ttt tttttgttt gaatttttat ticaaat ttt taaat at ag agaaaatttt      480
ttttatgat aatataaac tacagtttga attataaat aaaaaaaaaa at                                     532

```

<210> 51
 <211> 615
 <212> DNA
 <213> Scenedesmus dimorphus

```

<400> 51
taaaccgaag gttatcgggt caactccgat agtaggcttt ttttcaat ag ttctctcctg      60
tgaacaaaat ttattagtta ttgacaaaa ttaataaatt taaaattctt tccttttgtc      120
catcaaaaaa aaaatttttg aactttttcg tttttttaa actcacaatt ttttgtttta      180
ttagaaaat a aaacttgttt aaatttttat aaaaaacaa aaat aaaaat tttttttga      240

```

0802W01_ST25

aaactttttt ttgtcttttg cttttttgt ttacttacgc tttttagcac ttttaaat 300
 tttatcactt ttttactttg ctgctttttt tcttctttac ttttgcgtt tttatcattt 360
 tacgcttttt atcattttac gctttgcaaa aaatgaagaa gaaaaaaagc agcaaagcaa 420
 aaaatgaaga agaaaaaaag cagcaaatca aaaaagt gat aaaaagcgt a aaagt aaaga 480
 agaaaaaaag cagt aaagt a aaaaat tcat aaaaaacagc aaagcaaaaa aatgaaaaaa 540
 agt aaaaaca aaat aaaaaa aaat tttttg tat aaat aat tttttttat ttat acaagt 600
 acatcatttt tttcta 615

<210> 52
 <211> 533
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> modified endogenous promoter of the pet A gene in S. di morphus

<400> 52
 tgcacaaaag aaatcaaatg ttttcaaatc gtgcaaacat caaat tgcac aaaat aattt 60
 ttaaaattca ttttatgaaa gttcgttctt cagttaaaaa aatttgtact aaatgtcgtt 120
 taattcgtcg aaaaggtaca gtaatggtta tttgtacaaa tcctaacat aaacaacgtc 180
 aaggataatt ttttttaatg gaagaaatgt tttttacttc ttccattaaa aaagaaaaaa 240
 gaaaagt gca gaactttttt ttgatcaaa aaaaagat ac aaaaaat ttt ttgtttta 300
 ttttatgata taattctatt tcagagaaga aaaaaaat ttaaacaaaac aaaaaaat ac 360
 agagaagttt aaatttagac ct aaaggt ag atttccat aa attcattttt ctcat tttgtt 420
 tttttctttt ttttgcattg cgaaaaaaa aagaaaaaag caaaaagt ca ttttaagag 480
 araatggag aaagat aaaa gtttttaact ttttttgaac aattccat gg ggg 533

<210> 53
 <211> 521
 <212> DNA
 <213> Scenedesmus di morphus

<400> 53
 acaaaagaaa tcaaatgttt tcaaatcgtg caaacat caa attgcacaaa ataat tttta 60
 aaattcattt tatgaaagtt cgttcttcag ttaaaaaaat ttgtactaaa tgtcgtttaa 120
 ttctgcgaaa aggtacagta atggttattt gtacaaatcc taaacataaa caacgtcaag 180
 gat aattttt tttaatggaa gaaatgtttt ttacttcttc cattaaaaaa gaaaaagaa 240
 aagtgcagaa cttttttttt gatcaaaaaa aagat acaaa aaat tttttt gt ttt aattt 300
 tatgat agaa ttctatttca gagaagaaaa aaaaat ttaa acaaaacaaa aaaat acaga 360
 gaagttt aaa tttagacct a aggt agatt tccat aaatt catttttctc atttgtttt 420

0802W01_ST25

ttcttttttt tgcattgcaa aaaaaaaga aaaaagcaaa aagtcatttt aaagagaaaa	480
atggagaaag ataaaagttt ttaacttttt ttgaacaatt c	521

<210> 54
 <211> 503
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> modified endogenous promoter for the petB gene from S. dimorphus

<400> 54 aagaaagttt aaaaaaatTT acataagaag aagatacaaa aacaaattat ttccaatttt	60
tgttttgt aa aaaaaaatT tCGttttta gaatttctat gttttttttt attttttgtt	120
ttctgaaaaa aaaagttttt tcagaaaaca agaagaaat taaaggtcta caaaataaaa	180
attattttgt tacaacgac caatgtttat tattttttgt ttttttcatt ttctatttct	240
cacttatctt tttaaagtca aaagaaaatt tagaatgaaa gt aaaaaaaa gaaaaaaca	300
t aaaaaaaca aaaagcaaaa aaaattcttt gcaatttttg tttaaaaatt ttggtttca	360
at tttaaaat ttatgaaaa tgtacaccat aagctaaaaa ttacaaaaa ttatctaaaa	420
aaacttcttt cttttgtaac aaaaggcaat aagcccat aa ttcgagcact tttaatgtt	480
tttgtaaaca aaccaaaaaa aaa	503

<210> 55
 <211> 513
 <212> DNA
 <213> Scenedesmus dimorphus

<400> 55 aaaaacaatt aaaagtgtc gaattatggg ctatttgcct ttgtttacaa aagaagaag	60
tttttttaga taatttttgt aaattttaga ctatgggtgt acatttttca taatttttaa	120
aattgaaacc aaaaattttt aaacaaaaat tgc aaagaat tttttttgct tttttgtttt	180
tttattgttt tttctttttt ttaactttcat tctaaatttt cttttgactt taaaaagata	240
agt gagaaat agaaaatgaa aaaaacaaaa aataat aaac attgggtcgtt tgt aacaaaa	300
taatttttat tttgtagacc tttaattttc ttcttgtttt ctgaaaaaac tttttttttc	360
agaaaacaaa aaat aaaaaa aaacatagaa attctaaaaa acaaattttt tttttacaaa	420
acaaaattga aaaaatttgt ttttgaatct tattcctatg aattttttta actttttcgt	480
ttaatgggat tat aaaaaat tttctttttc tat	513

<210> 56
 <211> 537
 <212> DNA
 <213> Artificial Sequence

0802W01_ST25

<220>

<223> modified endogenous terminator region for the rbcL gene from *S. di morphus*

<400> 56

ttttttaa	aat	acttcctctt	t	aaaggggaa	gtat	tttttt	cttgatttta	gaactctaaa	60
aacacaaatt	gtttat	tttt	gtttttcatt	ttcatttatt	tattgat	aaa	aacaaaagaa	120	
gcagcaaagc	aaaaaa	acaa	aaaaaactaa	aaaaatttgt	tttgttcatt	tcaat	agaat	180	
aaaaaaaaca	aat	ttttca	aaaaaattta	t	aaaattatt	gt	aagttttc	240	
at	ttttttaa	aat	at	tttttt	aat	tg	aaaaaaaa	300	
aaagaatttt	aaaaaa	atag	aagagttttt	aatt	aaaaaa	aatt	gaaatt	360	
tttgtta	aat	a	t	aat	cttag	t	at	420	
agcat	gggcc	gat	gccc	gag	t	gg	gggcggattg	480	
acgt	t	gg	t	c	gaat	ccgact	cggcccaaaa	537	
at	acaaaat	a	cat	t	gt	t	ga		
aagact	a								

<210> 57

<211> 538

<212> DNA

<213> *Scenedesmus di morphus*

<400> 57

ttttttttta	aat	acttcc	t	ctttaaagg	ggaagt	at	tt	tttcttgat	60
t	aaaaacaca	aatt	gtttat	ttttgttttt	cat	tttcatt	t	atttattga	120
agaagcagca	aagcaaaaaa	acaaaaaaa	ct	aaaaaat	ttgt	tttgtt	cat	ttcaata	180
gaat	aaaaaa	aacaaaattt	ttcaaaaaa	ttt	at	aaaat	t	attgt	240
aaaaat	tttt	t	aaaaat	att	ttttaat	atg	tttttttttt	aaatt	300
aaacaaagaa	tttt	aaaaaa	at	agaagagt	tttt	aatt	aa	aaaaaat	360
at	ttttttgtt	aat	at	aat	ct	tt	agt	at	420
aaaaaagcat	gggccgat	gc	ccgagt	ggt	t	aat	gggggcg	gat	480
cgcct	acgt	t	ggt	t	cgaat	c	cgact	cggcc	538
caaaaat	aca	aat	acattg	ttt	gaaag				

<210> 58

<211> 669

<212> DNA

<213> *Scenedesmus di morphus*

<400> 58

ctagatttta	ttttttatga	aaaactcagg	cttaatttag	gcttgagttt	ttcattcttt	60
ttgaagctct	gaaatttta	aat	ttctagt	cttctttaat	gtttttaaat	120
aaatttcttc	tctgctgtgt	ttttcttttt	ttttgaaaaa	acaaagaaaa	aaaatttttt	180
tgttttcttc	tttgtttttt	tatttctttt	tgttttgttt	at	tttttagt	240
ttcagaat	ct					

0802W01_ST25

ttgattcaaa aaaaaattta gtccgattac tccataggag caagcagt aa aaaat aaaaa	300
ctgtaat aaa aaat aaaaca aaaattttat ttctttttgt tttgcttgaa cttttcaaaa	360
aaaaattgaa aaattcaagc aaaacaaaaa gaaacaaat a aaaaaattat gaattttcta	420
ctttttcagg agttgaaatt tctcctttac ttaaaacata ttttgctaaa aaaagcgctt	480
gtgttgcttt ttttgctact ttttgtttcc aagcatTTTT tcgaatattt ttttttgatt	540
ttgatgtgcg tttttgttaa cctaaaatct tgaaaagatt tactcttttc aaatttttat	600
gtttttattt tttttattca taaaaaaaa caat acat aa aaat aaagta tttcggttc	660
aaaaact ag	669

<210> 59
 <211> 597
 <212> DNA
 <213> Scenedesmus dimorphus

<400> 59 aatttccatt tttttcattt ttttttagaa aagtgtgtatt ttctttgatg aaatgaaat	60
ttcaaaaaag aaaaat acaa tttttctcta cttttttttg attctttact ttttttgaa	120
ttttttttgt gcttcgtttt tgaaaaaaac tttttatttg aaaaattttg ttgaattaaa	180
aaaaatcaac tcaat acatt tttttgaac ttctttttac ttttttgctt tgttgctttt	240
cttttcattt ttatcatttt ttgcttcgct gctttttatc attttttgct ttgcaaaaaa	300
tgaagaagaa aaaaagcgt a aaatgaaaa gaaaaaaagt ttcaaagaa aaaaagcgaa	360
aaagcaagag aagat aaaaa acaagaacaa aaaaattgtt caaaaacaca atagaaaat	420
ctttaaaaat tttttgattt ttatagaat ttgtgagaaa tagt aaaaaa agtcaaaaaa	480
tcacaacatt ataat agaa aaaatcttgt ttgtgaaaaa ttat aattt ttatattat	540
ttcaatttta aatagatct ttggagctg aaaaaat atg agaat agtt tggacaa	597

<210> 60
 <211> 12
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> nucleic acid linker

<400> 60 gaggcagatc aa	12
---------------------------	----

<210> 61
 <211> 12
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> nucleic acid linker

12

0802W01_ST25

t t a c g t a g t a a a t t t t c t t t g c t g g c c c t g c c a a a a a c a a c a c c a t t a a c c t a t a a g t a g	120
t a a t a a t t c t t t a g t a t t a c t t t t a g g t t a t t t a t a a a t t t g a g a a g t a t a g a a g a a t c t	180
a t a g a t t t t g c t t a t g t g t t t a t c t a t a g a t t c t t c t a t a c t t c t c a t t t t t a a c a a a t t	240
t t t a t t a a g a t t t t t t t a a a c a a a a a a a a g t t t t c a a c t t a t a t a a t t a a a c c t a a a c a	300
a c g t t g t a t a t t t t t t a t t t t a a g t t t t g g t a a a g t a t g t a t a c c a g t a a a c c t t t a g t a	360
a a t t t t t t t a c c g c t t a g g c t a g g a c c t a t a a a a t t t a g c g c g g c g c c c t a a g a g c t a g g	420
a t a a a g a t c c g t t g t c t g t c c g c t g c g c t a a a t t t c t t t a g a c g a a g c g a a g c g t c g a g	480
t c t c a c g c a c c c a c c	495

<210> 65
 <211> 311
 <212> DNA
 <213> D u n a l i e l l a

<400> 65 t t c t a a t c t c t t a a c g a g a t t a g a t t a a t c c c t t a a t g t c t a a g g g a t t a a t a c t t a t c c	60
c a t c c t g t t a a a c c c a g g a t g g g a t a a a a a a a t t t a c a a g a c a a g g t a t t a a a a t g a a t	120
a a a a t a g a t t a t a t g t a t t t a a a c g t a t a t a a t a a a a t t a t g a g t a a a a t a a c a g a g t a	180
a a c c a a a a c a a t a a g a t t t t t t t t t t t t t t c a a t t c t a a a t a t c a a t a t a g a a g g t t	240
a g g a a a t t t c c c c g g a a a a a a t c c t t t g g a t t t t t t c c t a a g g g c g a a a a a a t a t t c t c t	300
g a a t a t t t t t t	311

<210> 66
 <211> 40
 <212> DNA
 <213> A r t i f i c i a l S e q u e n c e

<220>
 <223> P C R p r i m e r

<400> 66 t t g g t t g g g c g g c c g c c g t t t a g g t g t a a c a c a a t c t t g g	40
---	----

<210> 67
 <211> 40
 <212> DNA
 <213> A r t i f i c i a l S e q u e n c e

<220>
 <223> P C R p r i m e r

<400> 67 t t g g t t g g g c g g c c g c t a g c c g t g g t a t t t a c t t c a c t c a	40
---	----

<210> 68
 <211> 35
 <212> DNA

<213> Artificial Sequence

<220>

<223> PCR primer

<400> 68

ttgggttggct cgaggccctt t cgtcttcaa gaaat

35

<210> 69

<211> 34

<212> DNA

<213> Artificial Sequence

<220>

<223> PCR primer

<400> 69

ttgggttggct cgagaagtct tgcgccttaa acca

34

<210> 70

<211> 77

<212> DNA

<213> Artificial Sequence

<220>

<223> PCR primer

<400> 70

ttgggttggct gcagtttaatt aaggatccac tagtatTTAA attcctgatg cggtattttc

60

tctttacgca tctgtgc

77

<210> 71

<211> 116

<212> DNA

<213> Artificial Sequence

<220>

<223> PCR primer

<400> 71

ttgggttgggc ggccgctcac agatgagaag atatttgctc gatataaat actctaggca

60

tctaactttt cccattgtct taaaccgact taccttagga atcatagttt catgat

116

<210> 72

<211> 116

<212> DNA

<213> Artificial Sequence

<220>

<223> PCR primer

<400> 72

ttgggttgggc ggccgctgct ggttgtcat tgcctctggat aatttttctc gaactatgcc

60

tgcgcgttga taccatcca atggatctac aggcagaacg gcctctagcg gttttt

116

0802W01_ST25

<210> 73
 <211> 62
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> PCR primer

<400> 73
 ttgggttggat ttaaat acta gtggat cctt aattaa ctgc agggccggcc tct agtttta 60
 tt 62

<210> 74
 <211> 100
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> PCR primer

<400> 74
 ttctgcctgt agatccattg gat tggatc aacgcgcagg cat agttcga gaaaaat tat 60
 ccagaggcaa tgacaaccag cacgggctcg agact agtgg 100

<210> 75
 <211> 100
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> PCR primer

<400> 75
 gaatttttca aaaaattcat actttgtttt ttatttttt ctgagttttt aatcaaaaaa 60
 ctttttgtat aaaattcgta agatctccta ggaaaatgaa 100

<210> 76
 <211> 100
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> PCR primer

<400> 76
 tttttttgaa aatagttttt ctcaattttt ttttttttt gttttttctc taaaaatcaa 60
 aaattcaatt ttgagaaaag ggctcgagac tagtttgtcc 100

<210> 77
 <211> 100
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> PCR primer

0802W01_ST25

<400> 77
 ggt aagt cgg t t t aagacaa t gggaaaagt t agat gcct a gagt att gat t at cgagcaa 60
 at at ct t ct c at ct gt gact agt gct agct aaagaagt t g 100

<210> 78
 <211> 80
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> PCR primer

<400> 78
 gat t ggt at c aacgcgcagg cat agt t cga gaaaaat t at ccagaggcaa t gacaaccag 60
 cacgggct cg agact agt gg 80

<210> 79
 <211> 83
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> PCR primer

<400> 79
 catt aggaat t at c t t t t t c gcaat t t t c t t tagagaacc acct cgt at g gt aaaat aac 60
 gt aagat ct c ct aggaaaat gaa 83

<210> 80
 <211> 80
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> PCR primer

<400> 80
 aaaagt t aga t gcct agagt att gat t at c gagcaa at at ct t ct cat ct gt gat ct cct 60
 agt gct agct aaagaagt t g 80

<210> 81
 <211> 80
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> PCR primer

<400> 81
 t caacgcct a t ggcaact ct gt aga at t cat cagcgt a acgcct t aga at at cat acg 60
 ggct cgagac t agt t t gt cc 80

0802W01_ST25

<210> 82
 <211> 80
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> PCR primer

<400> 82
 acgctgatga atattctaca gaggcttgcac aggcgttgaa cgctacacgg acgatacgaa 60
 tttttgaatt agataaatga 80

<210> 83
 <211> 80
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> PCR primer

<400> 83
 tcatactttgtttttttatt ttttctgagt ttttaataca aaaacttttt gtataaat 60
 ttttgaagcc gaaatacttt 80

<210> 84
 <211> 80
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> PCR primer

<400> 84
 cattaggaat tatctttttc gcaattttct ttagagaacc acctcgtatg gtaaaat aag 60
 ggctcgagac tagtttgtcc 80

<210> 85
 <211> 80
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> PCR primer

<400> 85
 cattaccaga gtgctccgca gacgagtatg gcacatggct ccgcgaggct attttctcct 60
 agtgctagct aaagaagttg 80

<210> 86
 <211> 80
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> PCR primer

0802W01_ST25

<400> 86
act ct ggt aa t gcat at ggt ccacaggaca t t cgt cgct t ccgggt at gc gct ct at gaa 60
t t cccgt t t t aagagct t gg 80

<210> 87
<211> 80
<212> DNA
<213> Artificial Sequence

<220>
<223> PCR primer

<400> 87
act gat t gac acggt t t agc agaacgt t t g aggact aggt caaat t gagt ggt agt t caa 60
gagaaaaaaa aagaaaaagc 80

<210> 88
<211> 98
<212> DNA
<213> Artificial Sequence

<220>
<223> PCR primer

<400> 88
accact caat t t gacct agt cct caaacgt t ct gct aaac cgt gt caat c agt gt ct gct 60
t cct gagt ga aaccctg aag at ct cct agg aaaat gaa 98

<210> 89
<211> 80
<212> DNA
<213> Artificial Sequence

<220>
<223> PCR primer

<400> 89
acgct gat ga at at t ct aca gagt t gccat aggcgt t gaa cgct acacgg acgat acgaa 60
agcagt t gct t t ct cct at g 80

<210> 90
<211> 80
<212> DNA
<213> Artificial Sequence

<220>
<223> PCR primer

<400> 90
t c t t t ggcg t t gt gccggg agcgt cat g t caat act t t t ct cct ct ag aagt t gaaag 60
t acaat aaac caagat gaag 80

0802W01_ST25

<210> 91
 <211> 80
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> PCR primer

<400> 91
 aaagtattga catgagcgct cccggcaciaa cggccaaaga agtctccaat ttcttatttc 60
 tttttgaatt agataaatga 80

<210> 92
 <211> 102
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> PCR primer

<400> 92
 ccattttttt gaaaatagtt ttctcaatt tttattttt ttgtttttt ctctaaaaat 60
 caaaaattca atttgagaa aagggtata actgatataa tt 102

<210> 93
 <211> 102
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> PCR primer

<400> 93
 aggtatgaat ttttcaaaaa attcatactt tgtttttta ttttttctga gtttttaatc 60
 aaaaaacttt ttgtataaaa ttgatcctcg agagatctta tg 102

<210> 94
 <211> 80
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> PCR primer

<400> 94
 acgctgatga atattctaca gagggttccat aggcgttgaa cgctacacgg acgatacgaa 60
 tttttgaatt agataaatga 80

<210> 95
 <211> 34
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> PCR primer

<400> 95
 ttgggttggac t agt gggtaa t aact gat at aatt 34

<210> 96
 <211> 34
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> PCR primer

<400> 96
 ttgggttggac t agt gat cct cgagagat ct tatg 34

<210> 97
 <211> 83
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> PCR primer

<400> 97
 ttgtattttt aaattttttag tttgaactac aactaattat tttctacgt a agatctact a 60

gtgggctcga gactagtttg tcc 83

<210> 98
 <211> 83
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> PCR primer

<400> 98
 ttgtattttt aaattttttag tttgaactac aactaattat tttctacgt a agatctact a 60

gtgggctcga gactagtttg tcc 83

<210> 99
 <211> 43
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> PCR primer

<400> 99
 ggttggttgc ggccgccggt ccggt agcag ttaataatgt agg 43

<210> 100
 <211> 59
 <212> DNA
 <213> Artificial Sequence

<220>

0802W01_ST25

<223> PCR primer

<400> 100

ccggcctct a gaggccggcc cctaggagat cttacgtaga aaataattag ttgtagttc

59

<210> 101

<211> 57

<212> DNA

<213> Artificial Sequence

<220>

<223> PCR primer

<400> 101

ccggcctct a gaggccggcc actagctctg agcccgggtt gaactatcaa gtttagg

57

<210> 102

<211> 44

<212> DNA

<213> Artificial Sequence

<220>

<223> PCR primer

<400> 102

ccaaccaagc ggccgcggcg cgcctttatt atggcggaac gacg

44

<210> 103

<211> 1291

<212> DNA

<213> Artificial Sequence

<220>

<223> codon optimized sequence comprising URA3

<400> 103

tcgcgcgttt cggatgatgac ggtgaaaacc tctgacacat gcagctcccg gagacgggtca 60

cagcttgtct gt aagcggat gccgggagca gacaagcccg tcagggcgcg tcagcgggtg 120

ttggcgggtg tcggggctgg cttaactatg cggcatcaga gcagattgta ctgagagtgc 180

accataccac cttttcaatt catcattttt tttttattct tttttttgat ttcggtttcc 240

ttgaaat tttgattcgg taatctccga acagaaggaa gaacgaagga aggagcacag 300

acttagattg gtatatat ac gcatatgt ag tgttgaagaa acatgaaatt gcccagtat 360

cttaacccaa ctgcacagaa caaaaacctg caggaaacga agataaatca tgtcgaaagc 420

t acatataag gaacgtgctg ctactcatcc tagtcctgtt gctgccaaagc tttttaatat 480

catgcacgaa aagcaaacaa acttgtgtgc ttcat tggat gt t cgtacca ccaaggaatt 540

actggagtta gttgaagcat taggtcccaa aatttgttta ctaaaacac atgtggat at 600

cttgactgat ttttccatgg agggcacagt taagccgcta aaggcatat ccgccaaagta 660

caat ttttta ctcttcgaag acagaaaatt t gctgacatt ggt aat acag t caaat t gca 720

0802W01_ST25

gt act ct gcg ggt gt at aca gaat agcaga at gggcagac at t acgaat g cacacggt gt 780
 ggt gggccca ggt at t gt t a gcggt t t gaa gcaggcggca gaagaagt aa caaaggaacc 840
 t agaggcct t t t gat gt t ag cagaat t gt c at gcaagggc t cct at ct a ct ggagaat a 900
 t act aagggt act gt t gaca t t gcgaagag cgacaaagat t t t gt t at cg gct t t at t gc 960
 t caaagagac at ggggt ggaa gagat gaagg t t acgat t gg t t gat t at ga cacccggt gt 1020
 ggg t t agat gacaagggag acgcat t ggg t caacagt at agaaccgt gg at gat gt ggt 1080
 ct ct acagga t ct gacatt a t t at t gt t gg aagaggact a t t t gcaaagg gaagggat gc 1140
 t aagggt agag ggt gaacgt t acagaaaagc aggc t gggaa gcat at t t ga gaagat gcgg 1200
 ccagcaaaac t aaaaaact g t at t at aagt aaat gcat gt at act aaact cacaat t ag 1260
 agct t caat t t aat t at at c agt t at t acc c 1291

<210> 104
 <211> 2309
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> codon optimized sequence comprising ADE2

<400> 104
 aagct t gcat gcct gcaggt cgat cgact c t agaaat cga t agat ct gaa t t aat t ct t g 60
 aat aat acat aact t t t ct t aaaagaat ca aagacagat a aaat t t aaga gat at t aaat 120
 at t agt gaga agccgagaat t t t gt aacac caacat aaca ct gacat ct t t aacaact t t 180
 t aat t at gat acat t t ct t a cgt cat gat t gat t at t aca gct at gct ga caaat gact c 240
 t t gt t gcat g gct acgaacc ggg t aat act aagt gat t ga ct ct t gct ga cct t t t at t a 300
 agaact aaat ggacaat at t at ggagcat t t cat gt at aa at t ggt gcgt aaaat cgt t g 360
 gat ct ct ct t ct aagt acat cct act at aa caat caagaa aaacaagaaa at cggacaaa 420
 acaat caagt at ggat t ct a gaacagt t gg t at at t agga gggggacaat t gggacgt at 480
 gat t gt t gag gcagcaaaca ggct caacat t aagacggt a at act agat g ct gaaaat t c 540
 t cct gccaaa caaat aagca act ccaat ga ccacgt t aat ggct cct t t t ccaat cct ct 600
 t gat at cgaa aaact agct g aaaaat gt ga t gt gct aacg at t gagat t g agcat gt t ga 660
 t gt t cct aca ct aaagaat c t t caagt aaa acat cccaaa t t aaaaat t t accct t ct cc 720
 agaaacaat c agat t gat ac aagacaaat a t at t caaaaa gagcat t t aa t caaaaat gg 780
 t at agcagt t acccaaagt g t t cct gt gga acaagccagt gagacgt ccc t at t gaat gt 840
 t ggaagagat t t ggg t t t t c cat t cgt ct t gaagt cgagg act t t ggcat acgat ggaag 900
 aggt aact t c gt t gt aaaga at aaggaaat gat t ccggaa gct t t ggaag t act gaagga 960
 t cgt cct t t g t acgccgaaa aat gggcacc at t t act aaa gaat t agcag t cat gat t gt 1020

0802W01_ST25

gaggt ct gt t aacgggt t t ag t g t t t t c t t a c c c a a t t g t a g a g a c t a t c c a c a a g g a c a a 1080
 t a t t t g t g a c t t a t g t t a t g c g c c t g c t a g a g t t c c g g a c t c c g t t c a a c t t a a g g c g a a 1140
 g t t g t t g g c a g a a a t g c a a t c a a t c t t t t c c c g g t t g t g g t a t a t t t g g t g t g g a a t 1200
 g t t c t a t t t a g a a c a g g g g a a t t g c t t a t t a a c g a a a t t g c c c c a a g g c c t c a c a a c t c 1260
 t g g a c a t t a t a c c a t t g a t g c t t g c g t c a c t t c t c a a t t t g a a g c t c a t t t g a g a t c a a t 1320
 a t t g g a t t t g c c a a t g c c a a a g a a t t t c a c a t c t t t c t c c a c c a t t a c a a c g a a c g c c a t 1380
 t a t g c t a a a t g t t c t t g g a g a c a a a c a t a c a a a g a t a a a g a g c t a g a a a c t t g c g a a a g 1440
 a g c a t t g g c g a c t c c a g g t t c c t c a g t g t a c t t a t a t g g a a a g a g t c t a g a c c t a a c a g 1500
 a a a a g t a g g t c a c a t a a a t a t t a t t g c c t c c a g t a t g g c g g a a t g t g a a c a a a g g c t g a a 1560
 c t a c a t t a c a g g t a g a a c t g a t a t t c c a a t c a a a t c t c t g t c g c t c a a a a g t t g g a c t t 1620
 g g a a g c a a t g g t c a a a c c a t t g g t t g g a a t c a t c a t g g g a t c a g a c t c t g a c t t g c c g g t 1680
 a a t g t c t g c c g c a t g t g c g g t t t t t a a a a g a t t t t g g c g t t c c a t t t g a a g t g a c a a t a g t 1740
 c t c t g c t c a t a g a a c t c c a c a t a g g a t g t c a g c a t a t g c t a t t t c c g c a a g c a a g c g t g g 1800
 a a t t a a a a c a a t t a t c g c t g g a g c t g g t g g g g c t g c t c a c t t g c c a g g t a t g g t g g c t g c 1860
 a a t g a c a c c a c t t c c t g t c a t c g g t g t g c c c g t a a a a g g t t c t t g t c t a g a t g g a g t a g a 1920
 t t c t t t a c a t t c a a t t g t g c a a a t g c c t a g a g g t g t t c c a g t a g c t a c c g t c g c t a t t a a 1980
 t a a t a g t a c g a a c g c t g c g c t g t t g g c t g t c a g a c t g c t t g g c g c t t a t g a t t c a a g t t a 2040
 t a c a a c g a a a a t g g a a c a g t t t t t a t t a a a g c a a g a a g a a g t t c t t g t c a a a g c a c a 2100
 a a a g t t a g a a a c t g t c g g t t a c g a a g c t t a t c t a g a a a a c a a g t a a t a t a t a a g t t t a t t 2160
 g a t a t a c t t g t a c a g c a a a t a a t t a t a a a a t g a t a t a c c t a t t t t t t a g g c t t t g t t a t g 2220
 a t t a c a t c a a a t g t g g a c t t c a t a c a t a g a a a t c a a c g c t t a c a g g t g t c c t t t t t t a a g 2280
 a a t t t c a t a c a t a a g a t c t c t c g a g g a t c 2309

<210> 105
 <211> 3688
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> codon optimized sequence comprising URA3-ADE2

<220>
 <221> misc_feature
 <222> (1292)..(1379)
 <223> nucleic acid linker

<400> 105
 g g g t a a t a a c t g a t a t a a t t a a a t t g a a g c t c t a a t t t g t g a g t t t a g t a t a c a t g c a t t 60

0802W01_ST25

t act t at aat acagt t t t t t agt t t t gct g gccgcat c t t ct caaat at g ct t cccagcc 120
 t gct t t t ct g t aacgt t cac cct ct acct t agcat ccct t ccct t t gcaa at agt cct ct 180
 t ccaacaat a at aat gt cag at cct gt aga gaccacat ca t ccacgggt t c t at act gt t g 240
 acccaat gcg t ct ccct t gt cat ct aaacc cacaccgggt gt cat aat ca accaat cgt a 300
 acct t cat ct ct t ccacca t gt ct ct t t g agcaat aaag ccgat aacaa aat ct t t gt c 360
 gct ct t cgca at gt caacag t accct t agt at at t ct cca gt agat aggg agccct t gca 420
 t gacaat t ct gct aacat ca aaaggcct ct aggt t cct t t gt t act t ct t ct gccgcct g 480
 ct t caaacg ct aacaat ac ct gggccac cacaccgt gt gcat t cgt aa t gt ct gccca 540
 t t ct gct at t ct gt at acac ccgcagagt a ct gcaat t t g act gt at t ac caat gt cagc 600
 aaat t t t ct g t ct t cgaaga gt aaaaaat t gt act t ggcg gat aat gcct t t agcggct t 660
 aact gt gccc t ccat ggaaa aat cagt caa gat at ccaca t gt gt t t t a gt aaacaaat 720
 t t t gggacct aat gct t caa ct aact ccag t aat t cct t g gt ggt acgaa cat ccaat ga 780
 agcacacaag t t t gt t t gct t t t cgt gcat gat at t aaat agct t ggcag caacaggact 840
 aggat gagt a gcagcacgt t cct t at at gt agct t t cgac at gat t t at c t t cgt t t cct 900
 gcagggt t t t t gt t ct gt gca gt t ggggt t aa gaat act ggg caat t t cat g t t t ct t caac 960
 act acat at g cgt at at at a ccaat ct aag t ct gt gct cc t t cct t cgt t ct t cct t ct g 1020
 t t cggagat t accgaat caa aaaaaat t t ca aggaaaccga aat caaaaaa aagaat aaaa 1080
 aaaaaat gat gaat t gaaaa ggt ggt at gg t gact ct ca gt acaat ct g ct ct gat gcc 1140
 gcat agt t aa gccagccccg acaccgcca acaccgct g acgcgccct g acgggct t gt 1200
 ct gct cccgg cat ccgct t a cagacaagct gt gaccgt ct ccgggagct g cat gt gt cag 1260
 aggt t t t cac cgt cat cacc gaaacgcgcg acgt aact at aacgggt cct a aggt agcgaa 1320
 ggccggcct t aat t aaat t t aaat cgccga gt t acgct ag ggat aacagg gt aat at aga 1380
 agct t gcat g cct gcagggt c gat cgact ct agaaat cgat agat ct gaat t aat t ct t ga 1440
 at aat acat a act t t t ct t a aaagaat caa agacagat aa aat t t aagag at at t aaat a 1500
 t t agt gagaa gccgagaat t t t gt aacacc aacat aacac t gacat ct t t aacaact t t t 1560
 aat t at gat a cat t t ct t ac gt cat gat t g at t at t acag ct at gct gac aaat gact ct 1620
 t gt t gcat gg ct acgaaccg ggt aat act a agt gat t gac t ct t gct gac ct t t t at t aa 1680
 gaact aaat g gacaat at t a t ggagcat t t cat gt at aaa t t ggt gcgt a aaat cgt t gg 1740
 at ct ct ct t c t aagt acat c ct act at aac aat caagaaa aacaagaaaa t cggacaaaa 1800
 caat caagt a t ggat t ct ag aacagt t ggt at at t aggag ggggacaat t gggacgt at g 1860
 at t gt t gagg cagcaaacag gct caacat t aagacgggt aa t act agat gc t gaaaat t ct 1920
 cct gccaaac aaat aagcaa ct ccaat gac cacgt t aat g gct cct t t t c caat cct ct t 1980

0802W01_ST25

gat at cgaaa aact agct ga aaaat gt gat gt gct aacga tt gagat t ga gcat gt t gat 2040
gt t cct acac t aaagaat ct t caagt aaaa cat cccaaat t aaaaat t t a ccct t ct cca 2100
gaaacaat ca gat t gat aca agacaaat at at t caaaaag agcat t t aat caaaaat ggt 2160
at agcagtt t a cccaaagt gt t cct gt ggaa caagccagt g agacgt ccct at t gaat gt t 2220
ggaagagat t t ggggt t t t cc at t cgt ct t g aagt cgagga ct t t ggcat a cgat ggaaga 2280
ggt aact t cg t t gt aaagaa t aaggaaat g at t ccggaag ct t t ggaagt act gaaggat 2340
cgt cct t t gt acgccgaaaa at gggcacca t t t act aaag aat t agcagt cat gat t gt g 2400
aggt ct gt t a acgggt t t agt gt t t t ct t ac ccaat t gt ag agact at cca caaggacaat 2460
at t t gt gact t at gt t at gc gcct gct aga gt t ccggact ccgt t caact t aaggcgaag 2520
t t gt t ggcag aaaat gcaat caaat ct t t t cccggt t gt g gt at at t t gg t gt ggaat g 2580
t t ct at t t ag aaacagggga at t gct t at t aacgaaat t g ccccaaggcc t cacaact ct 2640
ggacat t at a ccat t gat gc t t gcgt cact t ct caat t t g aagct cat t t gagat caat a 2700
t t ggat t t gc caat gccaaa gaat t t caca t ct t t ct cca ccat t acaac gaacgccat t 2760
at gct aaat g t t ct t ggaga caaacat aca aaagat aaag agct agaaac t t gcgaaaga 2820
gcat t ggcga ct ccaggt t c ct cagt gt ac t t at at ggaa aagagt ct ag acct aacaga 2880
aaagt aggt c acat aaat at t at t gcct cc agt at ggcgg aat gt gaaca aaggct gaac 2940
t acat t acag gt agaact ga t at t ccaat c aaaat ct ct g t cgct caaaa gt t ggact t g 3000
gaagcaat gg t caaacat t ggt t ggaat c at cat gggat cagact ct ga ct t gccggt a 3060
at gt ct gccg cat gt gcggt t t t aaaagat t t t ggcgt t c cat t t gaagt gacaat agt c 3120
t ct gct cat a gaact ccaca t aggat gt ca gcat at gct a t t t ccgcaag caagcgt gga 3180
at t aaaacaa t t at cgct gg agct ggt ggg gct gct cact t gccaggt at ggt ggct gca 3240
at gacaccac t t cct gt cat cgggt gt gccc gt aaaaggt t ct t gt ct aga t ggagt agat 3300
t ct t t acat t caat t gt gca aat gcct aga ggt gt t ccag t agct accgt cgct at t aat 3360
aat agt acga acgct gcgct gt t ggct gt c agact gct t g gcgct t at ga t t caagt t at 3420
acaacgaaaa t ggaacagt t t t t at t aaag caagaagaag aagt t ct t gt caaagcacia 3480
aagt t agaaa ct gt cgggt t a cgaagct t at ct agaaaaca agt aat at at aagt t t at t g 3540
at at act t gt acagcaaat a at t at aaaat gat at acct a t t t t t aggc t t t gt t at ga 3600
t t acat caaa t gt ggact t c at acat agaa at caacgct t acaggt gt cc t t t t t aaga 3660
at t t cat aca t aagat ct ct cgaggat c 3688

<210> 106
<211> 88
<212> DNA

<213> Artificial Sequence

<220>

<223> nucleic acid linker

<400> 106

cgt aact at a acggt cct aa ggt agcgaag gccggcctt a attaaattt a aat cgccgag 60

ttacgctagg gat aacaggg t aat at ag 88

<210> 107

<211> 3011

<212> DNA

<213> Artificial Sequence

<220>

<223> codon optimized comprising TRP1-ARS1-CEN4

<400> 107

gccctttcgt cttcaagaaa ttcggtcgaa aaaagaaaag gagagggcca agagggaggg 60

cat t ggt gac t att gagcac gtgagt at ac gtgatt aagc acacaaaggc agct t ggagt 120

atgtctgtt a ttaattt cac aggt agt t ct ggtccatt gg t gaaagt t t g cggct t gcag 180

agcacagagg ccgcagaat g t gct ct agat t ccgat gct g act t gct ggg t att at at gt 240

gtgcccaat a gaaagagaac aat t gacccg gtt att gcaa ggaaaattt c aagt ctt gt a 300

aaagcat at a aaaat agt t c aggcact ccg aaat act t gg t t ggctgt t t cgt aat caa 360

cct aaggagg atgtttt ggc t ct ggt caat gat t acggca tt gat at cgt ccaact gcat 420

ggagat gagt cgt ggcaaga at accaagag tt cct cgggt t t gccagt t at taaaagact c 480

gtattt ccaa aagact gcaa cat act act c agt gcagct t cacagaaacc t cat t cgt t t 540

attccctt gt t t gat t caga agcaggt ggg acaggt gaac tttt ggatt g gaact cgatt 600

tct gact ggg t t ggaaggca agagagcccc gaaagct t ac atttt atgt t agct ggt gga 660

ctgacgccag aaaatgtt gg t gat gcgct t agatt aaat g gcgt t att gg t gtt gat gt a 720

agcggaggt g t ggagacaaa t ggt gt aaaa gact ct aaca aaat agcaaa ttt cgt caaa 780

aat gct aaga aat aggt t at t act gagt ag t attt attt a agt attgttt gtgcact t gc 840

ctgcaggcct ttt gaaaagc aagcat aaaa gat ct aaaca taaaat ct gt aaaat aacaa 900

gatgt aaaga t aat gct aaa t cat t t ggct tttt gat t ga t t gt acagga aaat at acat 960

cgcaggggggt t gact ttt ac cat t t caccg caat ggaat c aaact t gtt g aagagaat gt 1020

t cacaggcgc at acgct aca at gacccgat t ctt gct agc ctttt ct cgg t ctt gcaaac 1080

aaccgccggc agct t agt at at aaat acac at gt acat ac ct ct ct ccgt at cct cgt aa 1140

t cat ttt t ctt gt attt at cg t ctttt cgct gt aaaaact t t at cacact t at ct caaat a 1200

cactt attt aa ccgct ttt ac t att at ctt c t acgct gaca gt aat at caa acagt gacac 1260

at att aaaca cagt ggt t t c ttt gcat aaa caccat cagc ct caagt cgt caagt aaaga 1320

0802W01_ST25

t t t c g t g t t c a t g c a g a t a g a t a a c a a t c t a t a t g t t g a t a a t t a g c g t t g c c t c a t c a a 1380
 t g c g a g a t c c g t t t a a c c g g a c c c t a g t g c a c t t a c c c c a c g t t c g g t c c a c t g t g t g c c 1440
 g a a c a t g c t c c t t c a c t a t t t t a a c a t g t g g a a t t a a t t c t a a a t c c t c t t t a t a t g a t c 1500
 t g c c g a t a g a t a g t t c t a a g t c a t t g a g g t t c a t c a a c a a t t g g a t t t t c t g t t t a c t c g 1560
 a c t t c a g g t a a t g a a a t g a g a t g a t a c t t g c t t a t c t c a t a g t t a a c t g g c a t a a a t t t t 1620
 a g t a t a g g t t a a c t c t a a g a g g t g a t a c t t a t t t a c t g t a a a a c t g t g a c g a t a a a a c c g 1680
 g a a g g a a g a a t a a g a a a a c t c g a a c t g a t c t a t a a t g c c t a t t t t c t g t a a a g a g t t t a a 1740
 g c t a t g a a a g c c t c g g c a t t t t g g c c g c t c c t a g g t a g t g c t t t t t t c c a a g g a c a a a a 1800
 c a g t t t c t t t t t c t t g a g c a g g t t t t a t g t t c g g t a a t c a t a a a c a a t a a a t a a a t t a t 1860
 t t c a t t t a t g t t t a a a a t a a a a a t a a a a a a g t a t t t t a a a t t t t a a a a a g t t g a t t 1920
 a t a a g c a t g t g a c c t t t t g c a a g c a a t t a a a t t t t g c a a t t t g t g a t t t a g g c a a a a g t t 1980
 a c t a t t t c t g g c t c g t g t a a t a t a t g t a t g c t a a t g t g a a c t t t t a c a a a g t c g a t a t g g 2040
 a c t t a g t c a a a a g a a a t t t t c t t a a a a t a t a t a g c a c t a g c c a a t t t a g c a c t t c t t t a 2100
 t g a g a t a t a t t a t a g a c t t t a t t a a g c c a g a t t t g t g t a t t a t a t g t a t t t a c c c g g c g a 2160
 a t c a t g g a c a t a c a t t c t g a a a t a g g t a a t a t t c t c t a t g g t g a g a c a g c a t a g a t a a c c 2220
 t a g g a t a c a a g t t a a a a g c t a g t a c t g t t t t g c a g t a a t t t t t t c t t t t t t a t a a g a a t 2280
 g t t a c c a c c t a a a t a a g t t a t a a a g t c a a t a g t t a a g t t t g a t a t t t g a t t g t a a a a t a c 2340
 c g t a a t a t a t t t g c a t g a t c a a a a g g c t c a a t g t t g a c t a g c c a g c a t g t c a a c c a c t a t 2400
 a t t g a t c a c c g a t a t t a g g a c t t c c a c a c c a a c t a g t a a t a t g a c a a t a a a t t c a a g a t a 2460
 t t c t t c a t g a g a a t g g c c c a g c t c a t g t t t g a c a g c t t a t c a t c g a t a a g c t t t a a t g c g 2520
 g t a g t t t a t c a c a g t t a a a t t g c t a a c g c a g t c a g g c a c c g t g t a t g a a a t c t a a c a a t g 2580
 c g c t c a t c g t c a t c c t c g g c a c c g t c a c c c t g g a t g c t g t a g g c a t a g g c t t g g t t a t g c 2640
 c g g t a c t g c c g g g c c t c t t g c g g g a t a t c g t c c a t t c c g a c a g c a t c g c c a g t c a c t a t g 2700
 g c g t g c t g c t a g c g c t a t a t g c g t t g a t g c a a t t t c t a t g c g c a c c c g t t c t c g g a g c a c 2760
 t g t c c g a c c g c t t t g g c c g c c g c c c a g t c c t g c t c g c t t c g c t a c t t g g a g c c a c t a t c g 2820
 a c t a c g c g a t c a t g g c g a c c a c a c c c g t c c t g t g g a t c a a t t c t t t a g t a t a a a t t t c a c 2880
 t c t g a a c c a t c t t g g a a g g a c c g g a t a a t t a t t t g a a a t c t c t t t t t c a a t t g t a t a t g t 2940
 g t t a t g t a g t a t a c t c t t t c t t c a a c a a t t a a a t a c t c t c g g t a g c c a a g t t g g t t t a a g 3000
 g c g c a a g a c t t 3011

<210> 108
 <211> 2103
 <212> DNA

<213> Artificial Sequence

<220>

<223> codon optimized sequence comprising LEU2

<400> 108

```

cct gat gcgg t at t t t c t c c t t a c g c a t c t g t g c g g t a t t t c a c a c c g c a t a t c g a c c g g      60
t c g a g g a g a a c t t c t a g t a t a t c t a c a t a c c t a a t a t t a t t g c c t t a t t a a a a t g g a a t      120
c c c a a c a a t t a c a t c a a a a t c c a c a t t c t c t t c a a a a t c a a t t g t c c t g t a c t t c c t t g t      180
t c a t g t g t g t t c a a a a a c g t t a t a t t t a t a g g a t a a t t a t a c t c t a t t t c t c a a c a g t a      240
a t t g g t t g t t t g g c c g a g c g g t c t a a g g c g c c t g a t t c a a g a a t a t c t t g a c c g c a g t t      300
a a c t g t g g g a a t a c t c a g g t a t c g t a a g a t g c a a g a g t t c g a a t c t c t t a g c a a c c a t t a      360
t t t t t t t c c t c a a c a t a a c g a g a a c a c a c a g g g g c g c t a t c g c a c a g a a t c a a a t t c g a t      420
g a c t g g a a a t t t t t t g t t a a t t t c a g a g g t c g c c t g a c g c a t a t a c c t t t t t c a a c t g a a      480
a a a t t g g g a g a a a a a g g a a g g a a g g t g a g a g c g c c g g a a c c g g c t t t t c a t a t a g a a t a g a g a      540
a g c g t t c a t g a c t a a a t g c t t g c a t c a c a a t a c t t g a a g t t g a c a a t a t t a t t t a a g g a c      600
c t a t t g t t t t t t c c a a t a g g t g g t t a g c a a t c g t c t t a c t t t c t a a c t t t t c t t a c c t t t      660
t a c a t t t c a g c a a t a t a t a t a t a t a t t c a a g g a t a t a c c a t t c t a a t g t c t g c c c c t      720
a a g a a g a t c g t c g t t t t g c c a g g t g a c c a c g t t g g t c a a g a a a t c a c a g c c g a a g c c a t t      780
a a g g t t c t t a a a g c t a t t t c t g a t g t t c g t t c c a a t g t c a a g t t c g a t t t c g a a a a t c a t      840
t t a a t t g g t g g t g c t g c t a t c g a t g c t a c a g g t g t t c c a c t t c c a g a t g a g g c g c t g g a a      900
g c c t c c a a g a a g g c t g a t g c c g t t t t g t t a g g t g c t g t g g g t g g t c c t a a a t g g g g t a c c      960
g g t a g t g t t a g a c c t g a a c a a g g t t t a c t a a a a t c c g t a a a g a a c t t c a a t t g t a c g c c      1020
a a c t t a a g a c c a t g t a a c t t t g c a t c c g a c t c t c t t t t a g a c t t a t c t c c a a t c a a g c c a      1080
c a a t t t g c t a a a g g t a c t g a c t t c g t t g t t g t c a g a g a a t t a g t g g g a g g t a t t t a c t t t      1140
g g t a a g a g a a a g g a a g a c g a t g g t g a t g g t g t c g c t t g g g a t a g t g a a c a a t a c a c c g t t      1200
c c a g a a g t g c a a a g a a t c a c a a g a a t g g c c g c t t t c a t g g c c c t a c a a c a t g a g c c a c c a      1260
t t g c c t a t t t g g t c c t t g g a t a a a g c t a a t g t t t t g g c c t c t t c a a g a t t a t g g a g a a a a      1320
a c t g t g g a g g a a a c c a t c a a g a a c g a a t t c c c t a c a t t g a a g g t t c a a c a t c a a t t g a t t      1380
g a t t c t g c c g c c a t g a t c c t a g t t a a g a a c c c a a c c c a c c t a a a t g g t a t t a t a a t c a c c      1440
a g c a a c a t g t t t g g t g a t a t c a t c t c c g a t g a a g c c t c g t t a t c c c a g g t t c c t t g g g t      1500
t t g t t g c c a t c t g c g t c c t t g g c c t c t t t g c c a g a c a a g a a c a c c c g c a t t t g g t t t g t a c      1560
g a a c c a t g c c a c g g t t c t g c t c c a g a t t t g c c a a a g a a t a a g g t c a a c c c t a t c g c c a c t      1620
a t c t t g t c t g c t g c a a t g a t g t t g a a a t t g t c a t t g a a c t t g c c t g a a g a a g g t a a g g c c      1680
a t t g a a g a t g c a g t t a a a a a g g t t t t g g a t g c a g g t a t c a g a a c t g g t g a t t t a g g t g g t      1740

```

0802W01_ST25

t ccaacagt a ccaccgaagt cggt gat gct gt cgccgaag aagt t aagaa aat cct t gct	1800
t aaaaagat t ctcttttttt at gat att t g t acat aaact t t at aaat ga aat t cat aat	1860
agaaacgaca cgaaatt aca aaat ggaat a t gt t cat agg gt agacgaaa ct at at acgc	1920
aat ct acat a cat t t at caa gaaggagaaa aaggaggat g t aaaggaat a cagg t aagca	1980
aat t gat act aat ggct caa cgt gat aagg aaaaagaat t gcact t t aac at t aat at t g	2040
acaaggagga gggcaccaca caaaaagt t a ggt gt aacag aaaat cat ga aact at gat t	2100
cct	2103

<210> 109
 <211> 1833
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> codon optimized sequence comprising CC-93

<400> 109	
at ggcaaaaa t gcgt gct gt t gat gcagca at gt at gt at t agagaaaga aggt at t act	60
acagct t t cg gt gt accagg t gcagcaat c aat ccat t t t at t ct gct at gcgt aaacat	120
ggg ggcatt c gt cacat t t t agct cgt cat gt agaagg t g ct agt cacat ggcagaagg t	180
t at acacgt g ct acagcagg t aat at t ggt gt t t gt t t ag gt acat ct gg cccagct ggt	240
acagat at ga t t act gcat t at at agt gct t cagct gat a gt at t cct at ct t at gt at t	300
act ggt caag cacct cgt gc t cgt ct t cat aaagaagat t t t caagcagt t gat at t gaa	360
gct at t gct a aacct gt aag t aaaat ggct gt aact gt t c gt gaggct gc t t t agt acca	420
cgt gt at t ac aacaggct t t t cact t aat g cgt t cagg t c gt cct ggt cc t gt at t agt t	480
gat t t acct t t cgat gt t ca agt agct gaa at t gaat t cg at ccagacat gt at gaacct	540
t t acct gt at at aaacct gc t gct t ct cgt at gcaaat t g aaaaagcagt agaaat gt t a	600
at t caagct g aacgt cct gt aat cgt t gct ggt ggt ggt g t t at t aat gc t gat gcagct	660
gct ct t t t ac agcaat t t gc agaat t aact t cagt acct g t aat t cct ac t t t aat gggc	720
t ggggt t gt a t t ccagat ga t cacgaat t a at ggcaggca t ggt aggt t t acaaacagct	780
caccgt t acg gt aat gct ac at t at t agca t ct gat at gg t at t cgg t at t ggaaat cgt	840
t t cgct aat c gt cat act gg t t cagt agag aaat at act g aaggacgt aa aat t gt t cac	900
at t gacat cg aacct acaca aat t ggt cgt gt at t at gcc ct gat t t agg cat t gt aagt	960
gat gcaaaag ct gct t t aac t t t act t gt t gaggt t gct c aagaaat gca aaaagcaggt	1020
cgt t t acct t gt agaaaaga at ggg t agct gact gccaac aacgt aaacg t act t t at t a	1080
agaaaaact c at t t t gat aa t gt t cct gt t aaacct caac gt gt at acga agaaat gaat	1140

0802W01_ST25

```

aaagcatttg gt cgt gat gt ttgct atgtt actacaattg gttaaagt ca aat agct gct 1200
gcacaaatgc ttcatgtatt taaagaccgt cattggattt actgtggcca agctggccct 1260
ttagggttgg caattcctgc agctttagggt gtatgtgcag ctgatccaaa acgtaatgta 1320
gttgctat tt cagggtgattt tgactttcaa ttctttattg aggaattagc tgttggg gca 1380
caatttaaca ttccat atat ccatgtttta gtaaat aacg ctatattagg tttaattcgt 1440
caatcacaac gtgcttttga tatggattac tgtgttcaat tagcatttga aaatataaac 1500
agttcagaag tt aacggttt tgggtgt agat cacgtaaaag ttgctgaagg cttaggatgt 1560
aaagctattc gtgtttttaa accagaagat atagctccag cttttgaaca ggctaaagca 1620
ttaatggctc aat atcgtgt tctgt agta gtgaagttt ttttagaaag agttacaaac 1680
at t t caatgg gttcagaatt agataatgtt atggagtttg aagatattgc agataatgct 1740
gctgacgcac ctactgaaac ttgctttatg cattatgaag attat aaaga t gat gat gat 1800
aaaggacaca accaccgtca caaacattaa tct 1833

```

<210> 110
 <211> 934
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> codon optimized sequence comprising CC-94

```

<400> 110
atgaaagt ag gat t t at agg attagggtatt atgggcaaac caatgtctaa aaacttactt 60
aaagctgggtt actcattagt tgttgctgat cgt aaccag aagcaatcgc tgacgtaat 120
gcagctggag ctgaaacagc ttcaactgct aaagctattg cagaacaatg t gatgttatt 180
attactatgt tacctaat t c acctcacgt a aaagaagt ag ctttaggcga aaatggaatt 240
attgaagggt g ct aaaccagg cacagtatt a atagat atgt catcaattgc tccactt gca 300
tcacgtgaaa tt t ct gaagc attaaaagca aaaggtattg atatgtt aga t gct ccagt t 360
agtgggtggg aacccaaaagc tatcgatggg acactt agt g ttatggtagg cggtgataaa 420
gcaatttttg acaaat acta cgat t t aat g aaagct atgg ctgggttctgt agt acacact 480
gggtgaaatcg gtgcaggt aa cgt aactaaa tt agct aacc aggttattgt tgcattaat 540
atagctgcaa tgtcagaagc tcttacttta gctacaaaag cagggtgtaaa t cct gat t t a 600
gtat at cagg caattcgtgg cggtttagca ggcagtactg t attagacgc aaaagctcca 660
atgggtt atgg atcgtaat t t caaacagggt tttagaattg attt acat at taaagacct 720
gctaatgctt t agat acat c acacgggtgt a ggagctcaat t accattaac tgcagctgtt 780
atggaaatga t gcaagcat t acgtgctgat ggttt aggt a cagcagat ca ct cagct t t a 840
gcttgttatt atgagaaatt agcaaaagt t gaagttacac gtgat t at aa agat gat gat 900

```

gat aaaggac acaaccaccg tcacaaacat t aat 934

<210> 111
 <211> 8765
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> codon optimized sequence comprising CC93-CC94

<400> 111
 tccaaactat tttctcatatt ttttcagctc caaaagatct tat t taaaat tgaaat aat a 60
 taaaaaatt a taaat t t t t t a caaaacaaga t t t t t t c t a t t a t aat g t t g t g a t t t t t 120
 gact t t t t t t t act a t t t c t c acaaat t c t a t aaaaaat ca aaaaa t t t t t aaagaat t t t 180
 ct a t t g t g t t t t t g a c a a t t t t t t g t t c t t g t t t t t a t c t t c t c t t g c t t t t c g c t 240
 t t t t t t c t t t g a a a c t t t t t t c t t t t t t c a t t t a c g c t t t t t t c t t c t t c a t t t t t 300
 g c a a g c a a a a a t g a t a a a a g c a g c g a a g c a a a a a t g a t a a a a t g a a a g a a a g c 360
 a a c a a g c a a a a a g t a a a a a g a a g t t c a a a a a a t g t a t t g a g t t g a t t t t t t a a t 420
 t c a c a a a a t t t t t c a a a t a a a a g t t t t t t t c a a a a a c g a a g c a c a a a a a a a t t c a a a 480
 a a a a g t a a a g a a t c a a a a a a a a g t a g a g a a a a t t g t a t t t t t c t t t t t t g a a a t t t t c 540
 a t t t c a t c a a g a a a a t a c a a c t t t t c t a a a a a a a t g a a a a a t g g a a a t t t c t a g a 600
 t t a a t g t t t g t g a c g g t g g t t g t g t c c t t t a t c a t c a t c a t c t t t a t a a t c t t c a t a a t g 660
 c a t a a g c a a g t t t c a g t a g g t g c g t c a g c a g c a t t a t c t g c a a t a t c t t c a a a c t c c a t 720
 a a c a t t a t c t a a t t c t g a a c c c a t t g a a a t g t t t g t a a c t c t t t c t a a a a t a a c t t c a a c 780
 t a c t a c a g g a a c a c g a t a t t g a g c c a t t a a t g c t t t a g c c t g t t c a a a a g c t g g a g c t a t 840
 a t c t t c t g g t t t a a a a c a c g a a t a g c t t t a c a t c c t a a g c c t t c a g c a a c t t t t a c g t g 900
 a t c t a c a c c a t a a c c g t t a a c t t c t g a a c t g t t t a t a t t t t c a a a t g c t a a t t g a a c a c a 960
 g t a a t c c a t a t c a a a a g c a c g t t g t g a t t g a c g a a t t a a a c c t a a a t a a g c g t t a t t t a c 1020
 t a a a c a t g g a t a t a t g g a a t g t t a a a t t g t g c a c c a a c a g c t a a t t c c t c a a t a a g a a a 1080
 t t g a a a g t c a a a t c a c c t g a a a t a g c a a c t a c a t t a c g t t t t g g a t c a g c t g c a c a t a c 1140
 a c c t a a a g c t g c a g g a a t t g t c c a a c c t a a a g g g c c a g c t t g g c c a c a g t t a a t c c a a t g 1200
 a c g g t c t t t a a a t a c a t g a a g c a t t t g t g c a g c a g c t a t t t g a c t t a a a c c a a t t g t a g t 1260
 a a c a t a g c a a a c a t c a c g a c c a a a t g c t t t a t t c a t t t c t t c g t a t a c a c g t t g a g g t t t 1320
 a a c a g g a a c a t t a t c a a a a t g a g t t t t t c t t a a t a a a g t a c g t t t a c g t t g t t g g c a g t c 1380
 a g c t a c c c a t t c t t t t c t a c a a g g t a a a c g a c c t g c t t t t t g c a t t t c t t g a g c a a c c t c 1440
 a a c a a g t a a a g t t a a a g c a g c t t t t g c a t c a c t t a c a a t g c c t a a a t c a g g g c a t a a t a c 1500

0802W01_ST25

acgaccaat t t gt gt aggt t cgat gt caat gt gaacaat t t t acgt cct t cagt at at t t 1560
 ct ct act gaa ccagt at gac gat t agcgaa acgat t t cca at accgaat a ccat at caga 1620
 t gct aat aat gt agcat t ac cgt aacggg g agct gt t t gt aaacct acca t gcct gccat 1680
 t aat t cgt ga t cat ct ggaa t acaacccca gcccatt aaa gt aggaat t a cagggt act ga 1740
 agt t aat t ct gcaaat t gct gt aaaagagc agct gcat ca gcat t aat aa caccaccacc 1800
 agcaacgat t acaggacgt t cagct t gaat t aacat t t ct act gct t t t t caat t t gcat 1860
 acgagaagca gcagggt t t at at acagggt aa aggt t cat ac at gt ct ggat cgaat t caat 1920
 t t cagct act t gaacat cga aagggt aaat c aact aat aca ggaccaggac gacct gaacg 1980
 cat t aagt ga aaagcct gt t gt aat acacg t ggt act aaa gcagcct cac gaacagt t ac 2040
 agccat t t t a ct t acagggt t t agcaat agc t t caat at ca act gct t gaa aat ct t ct t t 2100
 at gaagacga gcacgaggt g ct t gaccagt aat acat aag at aggaat ac t at cagct ga 2160
 agcact at at aat gcagt aa t cat at ct gt accagct ggg ccagat gt ac ct aaacaaac 2220
 accaat at t a cct gct gt ag cacgt gt at a acct t ct gcc at gt gact ag cacct t ct ac 2280
 at gacgagct aaaat gt gac gaat gccacc at gt t t acgc at agcagaat aaaat ggat t 2340
 gat t gct gca cct ggt acac cgaaagct gt agt aat acct t ct t t ct ct a at acat acat 2400
 t gct gcat ca acagcacgca t t t t t gccat at gaat aaat aat t t at aat t t t t ct gt a 2460
 t aaaccaat t t t ccaagt aa ct t t act t t a t caaaaaat t a aaaaaat t aaa aaact t t t at 2520
 t gaact t aaa at aaaat t t t t aacaaaaat t t at t t t aaaa aaaagaaaaa at t t t t t at 2580
 t t t ggt t t t a t t t at t t ct t t t t t t t aca aacaaaaat t t t t t aaaca gaat aat aaa 2640
 aaaaaat t t t a t t t aaagaat ggt t t t t t aa t at t t t gct c at gacaaat g at t t t t t act 2700
 act t t t at gc t t t t t t t t aa aaaaagcagc aaagcaaaaa agt t at aaaa agt gt at gga 2760
 gcaagcgggt t aaat t gacac t t t t t aaaag t at t t at agg cccaaccgga ct t gaaccga 2820
 t gacct at t g ct t gt aaggc aat cact ct a ccaact gagt t at gggcct a aaaaaat at t a 2880
 t t t at at t t t at aat agaat at aaaat ct a acaact t ct t t agct agcac t aggagat ca 2940
 cagat gagaa gat at t t gct cgat aat caa t act ct aggc at ct aact t t t cccat t gt c 3000
 t t aaaccgac t t acct t agg aat cat agt t t cat gat t t t ct gt t acacc t aact t t t t g 3060
 t gt ggt gccc t cct cct t gt caat at t aat gt t aaagt gc aat t ct t t t t cct t at cacg 3120
 t t gagccat t agt at caat t t gct t acct g t at t cct t t a cat cct cct t t t t ct cct t c 3180
 t t gat aaat g t at gt agat t gcgt at at ag t t t cgt ct ac cct at gaaca t at t ccat t t 3240
 t gt aat t t cg t gt cgt t t ct at t at gaat t t cat t t at aa agt t t at gt a caaat at cat 3300
 aaaaaaagag aat ct t t t t a agcaaggat t t t ct t aact t ct t cggcgac agcat caccg 3360
 act t cgggt gg t act gt t gga accacct aaa t caccagt t c t gat acct gc at ccaaaacc 3420

0802W01_ST25

t t t t t a a c t g	c a t c t t c a a t	g g c c t t a c c t	t c t t c a g g c a	a g t t c a a t g a	c a a t t t c a a c	3480
a t c a t t g c a g	c a g a c a a g a t	a g t g g c g a t a	g g g t t g a c c t	t a t t c t t t g g	c a a a t c t g g a	3540
g c a g a a c c g t	g g c a t g g t t c	g t a c a a a c c a	a a t g c g g t g t	t c t t g t c t g g	c a a a g a g g c c	3600
a a g g a c g c a g	a t g g c a a c a a	a c c c a a g g a a	c c t g g g a t a a	c g g a g g c t t c	a t c g g a g a t g	3660
a t a t c a c c a a	a c a t g t t g c t	g g t g a t t a t a	a t a c c a t t t a	g g t g g g t t g g	g t t c t t a a c t	3720
a g g a t c a t g g	c g g c a g a a t c	a a t c a a t t g a	t g t t g a a c c t	t c a a t g t a g g	g a a t t c g t t c	3780
t t g a t g g t t t	c c t c c a c a g t	t t t t c t c c a t	a a t c t t g a a g	a g g c c a a a a c	a t t a g c t t t a	3840
t c c a a g g a c c	a a a t a g g c a a	t g g t g g c t c a	t g t t g t a g g g	c c a t g a a a g c	g g c c a t t c t t	3900
g t g a t t c t t t	g c a c t t c t g g	a a c g g t g t a t	t g t t c a c t a t	c c c a a g c g a c	a c c a t c a c c a	3960
t c g t c t t c c t	t t c t c t t a c c	a a a g t a a a t a	c c t c c c a c t a	a t t c t c t g a c	a a c a a c g a a g	4020
t c a g t a c c t t	t a g c a a a t t g	t g g c t t g a t t	g g a g a t a a g t	c t a a a a g a g a	g t c g g a t g c a	4080
a a g t t a c a t g	g t c t t a a g t t	g g c g t a c a a t	t g a a g t t c t t	t a c g g a t t t t	t a g t a a a c c t	4140
t g t t c a g g t c	t a a c a c t a c c	g g t a c c c c a t	t t a g g a c c a c	c c a c a g c a c c	t a a c a a a a c g	4200
g c a t c a g c c t	t c t t g g a g g c	t t c c a g c g c c	t c a t c t g g a a	g t g g a a c a c c	t g t a g c a t c g	4260
a t a g c a g c a c	c a c c a a t t a a	a t g a t t t t c g	a a a t c g a a c t	t g a c a t t g g a	a c g a a c a t c a	4320
g a a a t a g c t t	t a a g a a c c t t	a a t g g c t t c g	g c t g t g a t t t	c t t g a c c a a c	g t g g t c a c c t	4380
g g c a a a a c g a	c g a t c t t c t t	a g g g g c a g a c	a t t a g a a t g g	t a t a t c c t t g	a a a t a t a t a t	4440
a t a t a t a t t g	c t g a a a t g t a	a a a g g t a a g a	a a a g t t a g a a	a g t a a g a c g a	t t g c t a a c c a	4500
c c t a t t g g a a	a a a a c a a t a g	g t c c t t a a a t	a a t a t t g t c a	a c t t c a a g t a	t t g t g a t g c a	4560
a g c a t t t a g t	c a t g a a c g c t	t c t c t a t t c t	a t a t g a a a a g	c c g g t t c c g g	c g c t c t c a c c	4620
t t t c c t t t t t	c t c c c a a t t t	t t c a g t t g a a	a a a g g t a t a t	g c g t c a g g c g	a c c t c t g a a a	4680
t t a a c a a a a a	a t t t c c a g t c	a t c g a a t t t g	a t t c t g t g c g	a t a g c g c c c c	t g t g t g t t c t	4740
c g t t a t g t t g	a g g a a a a a a a	t a a t g g t t g c	t a a g a g a t t c	g a a c t c t t g c	a t c t t a c g a t	4800
a c c t g a g t a t	t c c c a c a g t t	a a c t g c g g t c	a a g a t a t t t c	t t g a a t c a g g	c g c c t t a g a c	4860
c g c t c g g c c a	a a c a a c c a a t	t a c t t g t t g a	g a a a t a g a g t	a t a a t t a t c c	t a t a a a t a t a	4920
a c g t t t t t g a	a c a c a c a t g a	a c a a g g a a g t	a c a g g a c a a t	t g a t t t t g a a	g a g a a t g t g g	4980
a t t t t g a t g t	a a t t g t t g g g	a t t c c a t t t t	t a a t a a g g c a	a t a a t a t t a g	g t a t g t a g a t	5040
a t a c t a g a a g	t t c t c c t c g a	c c g g t c g a t a	t g c g g t g t g a	a a t a c c g c a c	a g a t g c g t a a	5100
g g a g a a a a t a	c c g c a t c a g g	a a t t t a a a t a	c t a g t g g a t c	c t t a a t t a a c	t g c a g g g c c g	5160
g c c t c t a g t t	t t a t t t g t a a	a c c a a a a a a a	a t g a a a a g c c	a a a a a t t t a a	g a a a t a a a a a	5220
g t c a a a g t t a	t t t a a c a a a a	a t g a a t t t c c	a a a a c t t g c a	c g a g a t a a a a	a a g a t a a c t c	5280

0802W01_ST25

ttt aaatgaa aaaagaatgc ttttttcaaa	aaaagttttta aacaat acgt aaaaacttta	5340
tttttttttaa actttttttt gaaaaaaggc	attctttttt ttttaagaaa ttttaagt aa	5400
tactttcata tttttttt agt atttttttat	tgaat aaaaaa aaaacttt aa agt aaaaaat	5460
tggtcactttt gaaagtccca gctttttttt	aattcactttt tttctttatt ttttttctg	5520
ttt aaaagaa aataaaattt ttaaaaattt	taaaaaat aa aagaacaaac tttcttgaga	5580
taacactaaa tcatctcaag aaagttta at	tttttgaaa agagttcgtt caaaaatttt	5640
tcagttttta atacaaattc tcaaat atct	aggttacgcc ccgccctgcc actcatcgca	5700
gtactgttgt aattcattaa gcatctgcc	gacatggaag ccatcacaaa cggcatgatg	5760
aacctgaatc gccagcggca tcagcacctt	gtcgccttgc gtataatatt tgcccatagt	5820
gaaaacgggg gcgaagaagt tgtccatatt	ggccacgttt aaatcaaaac tggtgaaact	5880
caccagggga ttggcgctga cgaaaaacat	attctcaata aacccttt ag ggaaataggc	5940
caggttttca ccgtaacacg ccacatcttg	cgaatatatg tgtagaaact gccggaatcg	6000
tcgtggtatt cactccagag cgatgaaaac	gtttcagttt gctcatggaa aacggtgt aa	6060
caagggtgaa cactatcca tatcaccagc	tcaccgtctt tcattgccat acggaactcc	6120
ggatgagcat tcatcaggcg ggcaagaatg	tgaat aaagg ccggat aaaa cttgtgctta	6180
tttttcttta cggctcttta aaaggccgt a	atatccagct gaacggtctg gttataggt a	6240
cattgagcaa ctgactgaaa tgcctcaaaa	tgttctttac gatgccattg ggat atatca	6300
acggtggtat atccagtgat ttttttctcc	attatgaaaa gtatcatagaa cattttgtcg	6360
atcctataaa aatctatccg tggaaaactt	ttttctttga aaaaaaaaaa gaaattgaaa	6420
aaaat aaaaa actaaactat attatagaaa	ataat aaaaa ctattttttt aacat aaaaa	6480
aat agttta aaatttttat tttttttaa	ttttaacata aattcaaaac tttagcaaat	6540
ttttaagttt aaatatttt ttcaaagtta	ccacttgctc ttggggaaca agt aaact aa	6600
attttttttc tacgcaaaac ttcattttta	ttgggtttat aattttctaa attttcatct	6660
atttgttttt tctatttttt tataaactca	aaaaaagttt caaaat atga aaaaaat act	6720
agaaaaattt ttttttgatt tttttattat	ttggaaaagt ttccaaaaa aaaattgggt	6780
tttaattttt tttatgaatt ttagaattct	ttgttttttc attttagttt tttttatttt	6840
gtattttttt ttctttgaac aaaagcaaaa	aaagcaaaag taaaaaaact agaggccgtt	6900
ctgcctgt ag atccattgga ttgggtatcaa	cgcgcaggca tagttcgaga aaaattatcc	6960
agaggcaatg acaaccagca cgggctcgag	act agtggcc ggcctctagt gct agcacia	7020
aagaaatcaa atgttttcaa atcgtgcaaa	catcaaat tg cacaaaat aa tttttaaaat	7080
tcattttatg aaagttcgtt cttcagttaa	aaaaatttgt actaaatgtc gtttaattcg	7140
tcgaaaaggt acagtaatgg ttatttgtac	aaatcctaaa cataaacaac gtcaaggat a	7200

0802W01_ST25

```

at t t t t t t t a at ggaagaaa t g t t t t t t a c t t c t t c c a t t a a a a a a g a a a a a g a a a g t 7260
gcagagct t t t t t t t g a t c a a a a a a a g a t a c a a a a a a t t t t t t t g t t t t a a t t t t a t g 7320
at a t a a t t c t a t t t c a g a g a g a a a a a a a a t t t a a a c a a a a c a a a a a a t a t a g a g a a g 7380
t t t a a a t t t a g a c c t a a a g g t a g a t t t c c a t a a a t t c a t t t t c t c a t t t g t t t t t t c t 7440
t t t t t t t g c a t t g c a a a a a a a a a a g a a a a a a g c a a a a a a g t c a t t t t a a a g a g a a a a t g g 7500
a g a a g a t a a a a g t t t t t a a c t t t t t t g a a c a a t t c c a t a t g a a a g t a g g a t t t a t a g g 7560
a t t a g g t a t t a t g g g c a a a c c a a t g t c t a a a a a c t t a c t t a a a g c t g g t t a c t c a t t a g t 7620
t g t t g c t g a t c g t a a c c c a g a a g c a a t c g c t g a c g t a a t t g c a g c t g g a g c t g a a a c a g c 7680
t t c a a c t g c t a a a g c t a t t g c a g a c a a t g t g a t g t t a t t a t t a c t a t g t t a c c t a a t t c 7740
a c c t c a c g t a a a g a a g t a g c t t t a g g c g a a a t g g a a t t a t t g a a g g t g c t a a a c c a g g 7800
c a c a g t a t t a a t a g a t a t g t c a t c a a t t g c t c c a c t t g c a t c a c g t g a a a t t t c t g a a g c 7860
a t t a a a g c a a a a g g t a t t g a t a t g t t a g a t g c t c c a g t t a g t g g t g g t g a a c c a a a a g c 7920
t a t c g a t g g t a c a c t t a g t g t t a t g g t a g g c g g t g a t a a a g c a a t t t t t g a c a a a t a c t a 7980
c g a t t t a a t g a a a g c t a t g g c t g g t t c t g t a g t a c a c a c t g g t g a a a t c g g t g c a g g t a a 8040
c g t a a c t a a a t t a g c t a a c c a g g t a t t g t t g c a t t a a a t a t a g c t g c a a t g t c a g a a g c 8100
t c t t a c t t t a g c t a c a a a a g c a g g t g t a a a t c c t g a t t t a g t a t a t c a g g c a a t t c g t g g 8160
c g g t t t a g c a g g c a g t a c t g t a t t a g a c g c a a a a g c t c c a a t g g t t a t g g a t c g t a a t t t 8220
c a a a c c a g g t t t t a g a a t t g a t t a c a t a t t a a a g a c c t t g c t a a t g c t t t a g a t a c a t c 8280
a c a c g g t g t a g g a g c t c a a t t a c c a t t a a c t g c a g c t g t t a t g g a a a t g a t g c a a g c a t t 8340
a c g t g c t g a t g g t t t a g g t a c a g c a g a t c a c t c a g c t t t a g c t t g t t a t t a t g a g a a a t t 8400
a g c a a a a g t t g a a g t t a c a c g t g a t t a t a a a g a t g a t g a t g a t a a a g g a c a c a a c c a c c g 8460
t c a c a a c a t t a a t c t a g a a a a a a a g c a t c t t t c a a a a t t a a a c t t t a a g t t t t t t t c t 8520
t t t t t t t a t t t t t t c c t t t t c t t t t a t t t a t t t a a c a a a a g a a a a g g a a a a a t a a a 8580
a a a a a t t t t a a g c g a c t t a t g t t t t t a a g t t t c a t t t t t t t a t t a t t t t t a t t a t t t t t 8640
a t t t c t t t t t t t a c a a a a c t t a a a a a a a g t t t a a a a a t a a a a a t t t t t g a c a a a a g a a a 8700
t c a a a t g t t t t c a a a t c g t g c a a c a t c a a a t t g c a c a a a a t a a t t t t t a a a a t t c a t t t 8760
t c c t a 8765

```

<210> 112
 <211> 8759
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> codon optimized sequence comprising CC93-CC97
 Page 52

0802W01_ST25

<400> 112
tccaaactat tttctcatatt ttttcagctc caaaagatct tat t t t t t aaaaat t gaaat aat a 60
t aaaaaatt a t a a a t t t t t a c a a a c a a g a t t t t t t c t a t t t a t a a t g t t g t g a t t t t t t 120
g a c t t t t t t t a c t a t t t c t c a c a a a t t c t a t a a a a a a t c a a a a a t t t t t a a g a a t t t t 180
c t a t t g t g t t t t t g a c a a t t t t t t g t t c t t g t t t t t a t c t t c t c t t g c t t t t t c g c t 240
t t t t t t t c t t t g a a a c t t t t t t t c t t t t t t c a t t t t a c g c t t t t t t c t t c t t c a t t t t t t 300
g c a a a g c a a a a a t g a t a a a a g c a g c g a a g c a a a a a a t g a t a a a a a t g a a a g a a a a g c 360
a a c a a a g c a a a a a g t a a a a a g a g a g t t c a a a a a a a t g t a t t g a g t t g a t t t t t t t a a t 420
t c a c a a a a t t t t t c a a a t a a a a g t t t t t t t c a a a a a c g a a g c a c a a a a a a a t t c a a a 480
a a a a g t a a a g a a t c a a a a a a a g t a g a g a a a a t t g t a t t t t t c t t t t t t g a a a t t t t c 540
a t t t c a t c a a g a a a a t a c a a c t t t t c t a a a a a a a a t g a a a a a a t g g a a a t t t c t a g a 600
t t a a t g t t t g t g a c g g t g g t t g t g t c c t t t a t c a t c a t c a t c t t t a t a a t c t t c a t a a t g 660
c a t a a g c a a g t t t c a g t a g g t g c g t c a g c a g c a t t a t c t g c a a t a t c t t c a a a c t c c a t 720
a a c a t t a t c t a a t t c t g a a c c c a t t g a a a t g t t t g t a a c t c t t t c t a a a a t a a c t t c a a c 780
t a c t a c a g g a a c a c g a t a t t g a g c c a t t a a t g c t t t a g c c t g t t c a a a a g c t g g a g c t a t 840
a t c t t c t g g t t t a a a a a c a c g a a t a g c t t t a c a t c c t a a g c c t t c a g c a a c t t t t a c g t g 900
a t c t a c a c c a t a a c c g t t a a c t t c t g a a c t g t t t a t a t t t t c a a a t g c t a a t t g a a c a c a 960
g t a a t c c a t a t c a a a a g c a c g t t g t g a t t g a c g a a t t a a a c c t a a a t a a g c g t t a t t t a c 1020
t a a a a c a t g g a t a t a t g g a a t g t t a a a t t g t g c a c c a a c a g c t a a t t c c t c a a t a a g a a a 1080
t t g a a a g t c a a a a t c a c c t g a a a t a g c a a c t a c a t t a c g t t t t g g a t c a g c t g c a c a t a c 1140
a c c t a a a g c t g c a g g a a t t g t c c a a c c t a a a g g g c c a g c t t g g c c a c a g t t a a t c c a a t g 1200
a c g g t c t t t a a a t a c a t g a a g c a t t t g t g c a g c a g c t a t t t g a c t t a a a c c a a t t g t a g t 1260
a a c a t a g c a a a c a t c a c g a c c a a a t g c t t t a t t c a t t t c t t c g t a t a c a c g t t g a g g t t t 1320
a a c a g g a a c a t t a t c a a a a t g a g t t t t t c t t a a t a a a g t a c g t t t a c g t t g t t g g c a g t c 1380
a g c t a c c c a t t c t t t t t c t a c a a g g t a a a c g a c c t g c t t t t t g c a t t t c t t g a g c a a c c t c 1440
a a c a a g t a a a g t t a a a g c a g c t t t t g c a t c a c t t a c a a t g c c t a a a t c a g g g c a t a a t a c 1500
a c g a c c a a t t t g t g t a g g t t c g a t g t c a a t g t g a a c a a t t t t a c g t c c t t c a g t a t a t t t 1560
c t c t a c t g a a c c a g t a t g a c g a t t a g c g a a a c g a t t t c c a a t a c c g a a t a c c a t a t c a g a 1620
t g c t a a t a a t g t a g c a t t a c c g t a a c g g t g a g c t g t t t g t a a a c c t a c c a t g c c t g c c a t 1680
t a a t t c g t g a t c a t c t g g a a t a c a a c c c c a g c c c a t t a a a g t a g g a a t t a c a g g t a c t g a 1740
a g t t a a t t c t g c a a a t t g c t g t a a a a g a g c a g c t g c a t c a g c a t t a a t a a c a c c a c c a c c 1800
a g c a a c g a t t a c a g g a c g t t c a g c t t g a a t t a a c a t t t c t a c t g c t t t t t c a a t t t g c a t 1860

0802W01_ST25

acgagaagca gcagggtttat atacaggtaa aggttcat ac atgtctggat cgaattcaat 1920
 tt cagct act tgaacat cga aaggtaaat c aactaat aca ggaccaggac gacctgaacg 1980
 catt aagt ga aaagcctgtt gt aat acacg tggtaact aaa gcagcctcac gaacagttac 2040
 agccatttt a cttacagggtt tagcaat agc ttcaat at ca actgcttgaa aatcttcttt 2100
 atgaagacga gcacgagggt g cttgaccagt aat acat aag ataggaat ac tatcagctga 2160
 agcact at at aatgcagt aa tcat at ctgt accagctggg ccagatgt ac ct aaacaaac 2220
 accaat at t a cctgctgt ag cacgtgt at a accttctgcc atgtgact ag caccttctac 2280
 atgacgagct aaaatgtgac gaatgccacc atgtttacgc atagcagaat aaaatggatt 2340
 gat t gct gca cctggtagac cgaaagctgt agtaaat acct tcttctct a at acat acat 2400
 t gct gcat ca acagcacgca tttttgccat atgaat aaat aat t t at aat tttttctgt a 2460
 taaaccaatt t tccaagtaaa ctttacttta tcaaaaatt a aaaaattaaa aaacttttat 2520
 tgaacttaaa ataaaatttt taacaaaatt tat t t t aaaa aaaagaaaaa at t t t t t at 2580
 tttggtttt a tttat t t ctt tttttttaca aacaaaaatt ttttt aaaca gaat aat aaa 2640
 aaaaatttt a tttaaagaat ggttttttaaa t at t t t gct c at gacaaat g at t t t t t act 2700
 acttttatgc tttttttttaa aaaaagcagc aaagcaaaaa agttat aaaa agtgtat gga 2760
 gcaagcgggtt aaattgacac tttttaaaag tat t t at agg cccaaccgga cttgaaccga 2820
 tgacct at t g cttgt aaggc aat cactcta ccaactgagt tatgggccta aaaaat at t a 2880
 tttat at t t t ataatagaat ataaaatcta acaacttctt tagct agcac taggagat ca 2940
 cagat gagaa gat at t t gct cgat aat caa t act ct aggc at ct aacttt tcccat t gtc 3000
 ttaaacggac ttaccttagg aat cat agtt tcat gat ttt ctgtt acacc taactttttg 3060
 tgt ggt gccc t cct cct t gt caat at t aat gtt aaagt gc aat t c t t t t cct t at cacg 3120
 ttgagccatt agt at caatt t gct t acct g tat t cct t t a cat cct cct t tttctccttc 3180
 ttgat aaat g tatgt agatt gcgt at at ag tttcgtctac cct at gaaca tat t ccat t t 3240
 tgt aat t t cg tgt cgt t t ct at t at gaatt t cat t t at aa agttt at gta caaat at cat 3300
 aaaaaaagag aatcttttta agcaaggatt t t c t t aact t c t t cggcgac agcat caccg 3360
 act t cgggt gg t act g t t gga accacctaaa t caccagttc t gat acctgc at ccaaaacc 3420
 ttttt aact g catcttcaat ggccttacct tctt caggca agtt caat ga caat t t caac 3480
 at catt gcag cagacaagat agtggcgat a ggggtt gacct tat t c t t t gg caaat ct gga 3540
 gcagaaccgt ggcat ggt t c gt acaaacca aatgcgggtgt tcttgtctgg caaagaggcc 3600
 aaggacgcag atggcaacaa acccaaggaa cctgggat aa cggaggctt c atcggagat g 3660
 at at caccaa acatgtt gct ggtgat t at a at accat t t a ggtgggtt gg gttctt aact 3720

0802W01_ST25

aggat cat gg cggcagaat c aat caat t ga t gt t gaacct t caat gt agg gaat t cgt t c 3780
 t t gat ggt t t cct ccacagt t t t t ct ccat aat ct t gaag aggccaaaac at t agct t t a 3840
 t ccaaggacc aaat aggcaa t ggt ggct ca t gt t gt aggg ccat gaaagc ggccat t ct t 3900
 gt gat t ct t t gcact t ct gg aacggt gt at t gt t cact at cccaagcgac accat cacca 3960
 t cgt ct t cct t t ct ct t acc aaagt aaat a cct cccact a at t ct ct gac aacaacgaag 4020
 t cagt acct t t agcaaatt g t ggct t gat t ggagat aagt ct aaaagaga gt cggat gca 4080
 aagt t acat g gt ct t aagt t ggcgt acaat t gaagt t ct t t acggat t t t t agt aaacct 4140
 t gt t caggt c t aacact acc ggt accccat t t aggaccac ccacagcacc t aacaaaacg 4200
 gcat cagcct t ct t ggaggc t t ccagcgcc t cat ct ggaa gt ggaacacc t gt agcat cg 4260
 at agcagcac caccaat t aa at gat t t t cg aaat cgaact t gacat t gga acgaacat ca 4320
 gaaat agct t t aagaacct t aat ggct t cg gct gt gat t t ct t gaccaac gt ggt cacct 4380
 ggcaaaaacga cgat ct t ct t aggggcagac at t agaatt gg t at at cct t g aaat at at at 4440
 at at at at t g ct gaaat gt a aaaggt aaga aaagt t agaa agt aagacga t t gct aacca 4500
 cct at t ggaa aaaacaat ag gt cct t aaat aat at t gt ca act t caagt a t t gt gat gca 4560
 agcat t t agt cat gaacgct t ct ct at t ct at at gaaaag ccggt t ccgg cgct ct cacc 4620
 t t t cct t t t t ct cccaatt t t t cagt t gaa aaaggt at at gcgt caggcg acct ct gaaa 4680
 t t aacaaaaa at t t ccagt c at cgaat t t g at t ct gt gcg at agcgcccc t gt gt gt t ct 4740
 cgt t at gt t g aggaaaaaa t aat ggt t gc t aagagat t c gaact ct t gc at ct t acgat 4800
 acct gagt at t cccacagt t aact gcggt c aagat at t t c t t gaat cagg cgcct t agac 4860
 cgct cggcca aacaaccaat t act t gt t ga gaaat agagt at aat t at cc t at aaat at a 4920
 acgt t t t t ga acacacat ga acaaggaagt acaggacaat t gat t t t gaa gagaat gt gg 4980
 at t t t gat gt aat t gt t ggg at t ccat t t t t aat aaggca at aat at t ag gt at gt agat 5040
 at act agaag t t ct cct cga ccggt cgat a t gcggt gt ga aat accgcac agat gcgt aa 5100
 ggagaaaat a ccgcat cagg aat t t aaat a ct agt ggat c ct t aat t aac t gcagggccg 5160
 gcct ct agt t t t at t t gt aa accaaaaaaa at gaaaagcc aaaaat t t aa gaaat aaaaa 5220
 gt caaagt t a t t t aacaaaa at gaat t t cc aaaact t gca cgagat aaaa aagat aact c 5280
 t t t aaat gaa aaaagaat gc t t t t t caaa aaaagt t t t a aacaat acgt aaaaact t t a 5340
 t t t t t t t t aa act t t t t t t gaaaaaggc at t ct t t t t t t t t t aagaaa t t t t aagt aa 5400
 t act t t cat a t t t t t t agt at t t t t t at t gaat aaaaa aaaact t t t aa agt aaaaaat 5460
 t ggt cact t t gaaagt ccca gct t t t t t t aat t cact t t t t ct t t at t at t t t cct g 5520
 t t t aaaagaa aat aaaat t t t t aaaaat t t t aaaaaat aa aagaacaaac t t t ct t gaga 5580
 t aacact aaa t cat ct caag aaagt t t aat at t t t t gaaa agagt t cgt t caaaaat t t t 5640

0802W01_ST25

t cagt t t t a a at a c a a a t t c t c a a a t a t c t a g g t t a c g c c c c g c c c c t g c c a c t c a t c g c a 5700
 g t a c t g t t g t a a t t c a t t a a g c a t t c t g c c g a c a t g g a a g c c a t c a c a a a c g g c a t g a t g 5760
 a a c c t g a a t c g c c a g c g g c a t c a g c a c c t t g t c g c c t t g c g t a t a a t a t t t g c c c a t a g t 5820
 g a a a a c g g g g g c g a a g a a g t t g t c c a t a t t g g c c a c g t t t a a a t c a a a a c t g g t g a a a c t 5880
 c a c c c a g g g a t t g g c g c t g a c g a a a a c a t a t t c t c a a t a a a c c c t t t a g g g a a t a g g c 5940
 c a g g t t t t c a c c g t a a c a c g c c a c a t c t t g c g a a t a t a t g t g t a g a a a c t g c c g g a a t c g 6000
 t c g t g g t a t t c a c t c c a g a g c g a t g a a a a c g t t t c a g t t t g c t c a t g g a a a a c g g t g t a a 6060
 c a a g g g t g a a c a c t a t c c c a t a t c a c c a g c t c a c c g t c t t t c a t t g c c a t a c g g a a c t c c 6120
 g g a t g a g c a t t c a t c a g g c g g g c a a g a a t g t g a a t a a a g g c c g g a t a a a a c t t g t g c t t a 6180
 t t t t t c t t t a c g g t c t t t a a a a g g c c g t a a t a t c c a g c t g a a c g g t c t g g t t a t a g g t a 6240
 c a t t g a g c a a c t g a c t g a a a t g c c t c a a a a t g t t c t t t a c g a t g c c a t t g g g a t a t a t c a 6300
 a c g g t g g t a t a t c c a g t g a t t t t t t c t c c a t t a t g a a a a g t t c a t a g a a c a t t t t g t c g 6360
 a t c c t a t a a a a a t c t a t c c g t g g a a a a c t t t t t c t t t g a a a a a a a a a a g a a a t t g a a a 6420
 a a a a t a a a a a a c t a a a c t a t a t t a t a g a a a a a a t a a a a a a c t a t t t t t t t a a c a t a a a a a 6480
 a a t a g t t t a a a a a t t t t t a t t t t t t t a a a t t t t a a c a t a a a t t c a a a a c t t t a g c a a a t 6540
 t t t t a a g t t t a a a t a t t t t t t t c a a a g t t a c c a c t t g c t c t t g g g a a c a a g t a a a c t a a 6600
 a t t t t t t t t c t a c g c a a a a c t t c a t t t t t a t t g g t t t a t a a t t t t c t a a a t t t c a t c t 6660
 a t t t g t t t t t t c t a t t t t t t t a t a a a c t c a a a a a a g t t t c a a a a t a t g a a a a a a t a c t 6720
 a g a a a a a t t t t t t t t g a t t t t t t a t t a t t t g g a a a a g t t t c c a a a a a a a a a a t t g g g t 6780
 t t t a a t t t t t t t t a t g a a t t t t a g a a t t c t t t g t t t t t t c a t t t t a g t t t t t t t a t t t t 6840
 g t a t t t t t t t t t c t t t g a a c a a a a g c a a a a a a a g c a a a a g t a a a a a a a c t a g a g g c c g t t 6900
 c t g c c t g t a g a t c c a t t g g a t t g g t a t c a a c g c g c a g g c a t a g t t c g a g a a a a t t a t c c 6960
 a g a g g c a a t g a c a a c c a g c a c g g g c t c g a g a c t a g t g g c c g g c c t c t a g t g c t a g c a c a a 7020
 a a g a a a t c a a a t g t t t t c a a a t c g t g c a a a c a t c a a a t t g c a c a a a a t a a t t t t t a a a a t 7080
 t c a t t t t a t g a a a g t t c g t t c t t c a g t t a a a a a a t t t g t a c t a a a t g t c g t t t a a t t c g 7140
 t c g a a a a g g t a c a g t a a t g g t t a t t t g t a c a a a t c c t a a a c a t a a a c a a c g t c a a g g a t a 7200
 a t t t t t t t t a a t g g a a g a a a t g t t t t t t a c t t c t t c c a t t a a a a a g a a a a a g a a a g a a a g t 7260
 g c a g a g c t t t t t t t t g a t c a a a a a a a a g a t a c a a a a a a t t t t t t g t t t t a a t t t t a t g 7320
 a t a t a a t t c t a t t t c a g a g a a g a a a a a a a a t t t a a c a a a a c a a a a a a a t a t a g a g a a g 7380
 t t t a a a t t t a g a c c t a a a g g t a g a t t t c c a t a a a t t c a t t t t t c t c a t t t g t t t t t t t c t 7440
 t t t t t t t g c a t t g c a a a a a a a a a a g a a a a a a g c a a a a a g t c a t t t t a a a g a g a a a a t g g 7500

0802W01_ST25

agaaagat aa aagt t t t t aa c t t t t t t t ga acaat t ccat at gaaat t ag g t t t cat cgg 7560
 t t t aggt at t at ggg t act c caat ggcaat t aat c t t gct agagct ggt c at caat t aca 7620
 t g t t act act at t ggt cct g t t gcagacga act t t t at ct t t aggt gct g t at cagt t ga 7680
 aacagcacgt caagt t acag aagcat caga cat aat at t t at t at ggt t c cagat act cc 7740
 acaagt agaa gaagt t t t at t t ggt gaaaa t ggt t gcact aaagcat c t t t aaaaggt aa 7800
 aact at t g t t gat at g t c t t ct at ct cacc t at t gaaact aaacgt t t t g caagacaagt 7860
 aaat gaat t a ggt ggt gact at c t t gacgc t ccagt t t ca ggt ggt gaaa t t ggt gct cg 7920
 t gaaggt aca t t at ct at ca t ggt aggt gg agat gaagct g t t t t cgaac g t g t aaaacc 7980
 t t t at t t gaa t t act t ggca aaaat at cac at t agt t gga ggt aat ggt g at ggacaaac 8040
 at g t aaagt t gcaaat caaa t t at t g t agc act t aat at t gaagcagt t t ct gaagct ct 8100
 t t t at t t gct t caaaagct g g t gcagat cc t g t t cgt g t t cgt caagct c t t at gggcgg 8160
 t t t t gct t ca t caagaat t t t agaagt aca t ggt gagcgt at gat t aaac g t acat t caa 8220
 t cct ggt t t c aaaat t gct t t acat caaaa agat t t aaac t t agct t t ac agt cagct aa 8280
 agcat t agct t t aaact t ac ct aat act gc t act t gccaa gaact t t t ca at acat gcgc 8340
 t gct aat ggt ggcagt cagt t agat cat t c agct t t agt t caagcat t ag aact t at ggc 8400
 t aaccat aaa t t agcagat t at aaagat ga cgacgat aaa ggt cacaat c accgt cat aa 8460
 acaact aat ct agaaaaaaa gcat c t t t ca aaat t aaact t t aagt t t t t t t c t t t t t t 8520
 t at t t t t t cc t t t t c t t t t t at t t t t t a acaaaaagaa aaggaaaaaa t aaaaaaaa 8580
 t t t t aagcga c t t at g t t t t aag t t t cat t t t t t t at t t t t at t t at t t t t ct 8640
 t t t t t t acaa aact t aaaaa aagt t t aaaa at aaaaaat t t t t gacaaaa gaaat caaat 8700
 g t t t t caaat cgt gcaaaca t caaat t gca caaaat aat t t t t aaaat t c at t t t cct a 8759

<210> 113

<211> 927

<212> DNA

<213> Artificial Sequence

<220>

<223> codon optimized sequence comprising CC-97

<400> 113

at gaaat t ag g t t t cat cgg t t t aggt at t at ggg t act c caat ggcaat t aat c t t gct 60
 agagct ggt c at caat t aca t g t t act act at t ggt cct g t t gcagacga act t t t at ct 120
 t t aggt gct g t at cagt t ga aacagcacgt caagt t acag aagcat caga cat aat at t t 180
 at t at ggt t c cagat act cc acaagt agaa gaagt t t t at t t ggt gaaaa t ggt t gcact 240
 aaagcat c t t t aaaaggt aa aact at t g t t gat at g t c t t ct at ct cacc t at t gaaact 300
 aaacgt t t t g caagacaagt aaat gaat t a ggt ggt gact at c t t gacgc t ccagt t t ca 360

0802W01_ST25

ggg ggt gaaa t t ggt gct cg t gaaggt aca t t at ct at ca t ggt aggt gg agat gaagct	420
gt t t t cgaac gt gt aaaacc t t t at t t gaa t t act t t ggca aaaat at cac at t agt t gga	480
ggg aat ggt g at ggacaaac at gt aaagt t gcaaat caaa t t at t gt agc act t aat at t	540
gaagcagt t t ct gaagct ct t t t at t t gct t caaaagct g gt gcagat cc t gt t cgt gt t	600
cgt caagct c t t at gggcgg t t t t gct t ca t caagaat t t t agaagt aca t ggt gagcgt	660
at gat t aaac gt acat t caa t cct ggt t t c aaaat t gct t t acat caaaa agat t t aaac	720
t t agct t t ac agt cagct aa agcat t agct t t aaact t ac ct aat act gc t act t gccaa	780
gaact t t t ca at acat gcgc t gct aat ggt ggcagt cagt t agat cat t c agct t t agt t	840
caagcatt ag aact t at ggc t aaccat aaa t t agcagat t at aaagat ga cgacgat aaa	900
ggg cacaat c accgt cat aa acact aa	927

<210> 114
 <211> 16417
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> codon optimized sequence comprising CC90-CC91-CC92

<400> 114	
cgt t t aggt g t aacacaat c t t ggggt gga t ggacaat t a gcggt gaaac agcaacaat	60
ccaggt at t t ggagt t at ga aggt gt t gct gcat ct cat a t t at t t t at c t ggt t t at t a	120
t t c t t agct t cgg t t t ggca ct ggg t t t ac t ggg at t t ag agt t at t ccg t gacccaaga	180
act ggaaaaa ct gcatt aga t t t accaaaa at t t t cggaa t t cact t at t ct t at caggt	240
ct t t t at gt t t t ggt t t t gg t gct t t ccac gt aacagg t t t at t t ggt cc t ggt at t t gg	300
gt t t cagat c ct t at ggatt aacaggaagt gt t caaccag t t gct cct t c t t ggggt gct	360
gat ggg t t t g at cct t t caa ccct ggt ggt at t gcagcgc accacat t gc t gct ggt at t	420
t t aggt gt t t t agcaggat t at t ccact t a t gt gt acgt c ct t ct at t cg t t t at act t t	480
gg t t t at caa t ggg t agt at cgaaacagt a t t at caagt a gt at t gct gc t gt t t t ct gg	540
gct gct t t cg t t gt t gct gg aact at gt gg t at ggt t cag cagct act cc aat t gaat t a	600
t t t ggt cct a cacgt t at ca at gggacca ggt t t ct t cc aacaagaaat t caaaaacga	660
gt t caaaca gt t t agcagg t ggt t ct t ca ct t t ct gat g ct t gggcgaa aat t ccagaa	720
aaat t agct t t ct at gatt a t at t ggaaac aaccct gcaa aagg t ggt ct t t t ccgt aca	780
ggagct at ga at agt ggaga t ggt at t gct gt t ggat ggt t aggt cacgc agt at t t aaa	840
gat caagat g gt cgt gaat t at acgt acgt cgt at gcct a ct t t ct t t ga aacat t ccca	900
gt t t t at t aa t t gat aaaga t ggt gt t gt a cgt gct gacg t t cct t t ccg t cgt gct gaa	960

0802W01_ST25

t caaaat at a	gt at t gaaca	agt t ggt gt a	t cagt aact t	t ct acggt gg	t gaat t agat	1020
ggat t aacat	t t aat gat cc	agcaact gt t	aaaaaat at g	ct cgt aaagc	acaat t aggt	1080
gaaat t t t t g	aat t t gat cg	t t caacat t a	caat ct gat g	gt gt at t ccg	t agt agt cca	1140
cgt ggt t ggt	t t act t t t gg	t cacgt t t gc	t t t gct t t at	t at t ct t ct t	t ggacat at t	1200
t ggcat ggt g	cacgt acaat	ct t ccgt gat	gt at t t gct g	gt at t gat ga	t gat ct aaac	1260
gaaagt t t ag	aat t t ggt aa	at acaaaaaa	ct t ggt gat a	caagt t ct gt	t cgt gaagct	1320
t t ct aat t cg	t t t t t t t ct c	t t t t t t t ct	t t t t t ct ct t	t ggaaaaaga	aaaaacat gt	1380
t t at t t t gaa	t t t t t t g t t t	agaact t t ac	t gt t ct t t t t	t t at t t t aaa	gt gt t t ct gt	1440
t t t t t t t aa	t acaaaaaact	t t t t t aaaat	gaat t t aaaa	aacacaaaaa	aagagt t at t	1500
gct at t caaa	at aaacaaga	gt t t aaaaac	aaagt t t t t t	t ct t t agaaa	aaaact t ct t	1560
cat t t t t t t	gaat t g t t t t	t gaact t t t t	t ct t ct ct t g	ct t t t at cgt	t t t t t ct t c	1620
act t t t t gca	aaaaagt gag	aaaaaacagc	aaagcaaaaa	agt gaaaaaa	agt t caaaaa	1680
caat t caaaa	aagacaaaac	ct aaaaaaat	at cact t gag	at ggggt ct gg	at t t t t t cca	1740
agcaaaagaa	t t t t gt at t t	t gt t gaaagt	t t t t cat aaa	aat acaaat t	t gcaat t at t	1800
at t ct t aaaa	t caaaat at t	t gt t aaccac	at t t cat t ct	at ggaagcat	t agt t t at ac	1860
t t t t t at t a	at cggaacat	t aggaat t at	ct t t t t cgca	at t t t ct t t a	gagaaccacc	1920
t cgt at ggt a	aaat aacgt a	agat ct cct a	ggaaaaat gaa	t t t t aaaaat	t at t t t gt gc	1980
aat t t gat gt	t t gcacgat t	t gaaaacat t	t gat t t ct t t	t gt caaaaat	t t t t t at t t t	2040
t aaact t t t t	t t aagt t t t g	t aaaaaaaga	aat aaaaaat	aat aaaaaat	aaaaaaaat g	2100
aaact t aaaa	acat aagt cg	ct t aaaat t t	t t t t t at t t	t t t cct t t t c	t t t t t gt t aa	2160
at aaaat aaa	aagaaaagga	aaaaat aaaa	aaaagaaaaa	aact t aaagt	t t aat t t t ga	2220
aagat gct t t	t t t t t ct aga	t t aat gt t t g	t gacggt ggt	t gt gt cct t t	at cat cat ca	2280
t ct t t at aat	caaaacgt t c	aagct ct gga	aaaggt aat	ggccgt gat g	t acat gcat a	2340
gcaccaaaact	ct gcacaacg	at gt aaagt t	ggaat at t t t	t accaggat t	aagt aaaccg	2400
t ct ggat caa	aagcagct t t	t acagcgt ga	aat gt agt aa	t t t cat cact	at t gaat t ga	2460
gcacacat t t	gat t gat t t t	t t cacggccg	at t ccat gct	caccact aat	agaaccacct	2520
act t ct acac	at aact ct aa	gat t t t t cca	ccaagt t ct t	cagcacgagc	aaact cacca	2580
gggt t cgt t ag	cat cgaat aa	gat t aat gga	t gcat at t ac	cat cgccagc	at gaaat aca	2640
t t agct acac	gt aaat cgt a	t t gt t gagaa	agacgt gcaa	t acct t ct aa	aacacct ggt	2700
aat gcacgac	gt ggaat t gt	accat ccat a	caat aat agt	ct ggt gaaat	acgacct act	2760
gcagggaaaag	cat t t t t acg	acct gcccaa	aat ct t acac	gt t cagct t c	at ct t gt gct	2820
aaacgaacat	cagt agcacc	agct t t aagt	aat at gt cgt	t t acacgct c	acagt ct t cc	2880

WO1_ST25

t gaacat cac	t ct ct act cc	at ct aat t ca	caaagt aaaa	t t gct t cagc	at ct act gga	2940
t aaccagcgt	gt at aaaat c	t t cagct gca	cgaat t gat a	agt t at ccat	cat ct ct aat	3000
cct cct ggaa	t t at accat t	t gcaat gat a	t cgcct acag	ct aat cct gc	t t t t t caact	3060
gaat caaaag	at gct aat aa	aacacgagca	acaggt ggt t	t t ggaagt aa	t t t aacagt t	3120
act t ct gt t g	t aacacct aa	cat t cct t ct	gat cct gt ga	at aaagct aa	t aaat caaaa	3180
ccaggagaaat	ct aat gcat c	act acct aaa	gt aagagcct	caccgt ct aa	agt t t gt act	3240
t caat t t t t a	at aagt t gt g	t act gt t aaa	ccat at t t t a	aacagt gt ac	gccaccagca	3300
t t ct ct gct a	cat t accacc	aat t gaacaa	gcaat t t gac	t act aggat c	t ggagcgt aa	3360
t at aagt t at	gaggagcaac	t gct t gt gaa	at t gct aaat	t acgaacacc	aggt t gaaca	3420
cgt gcacgac	gaccaact gg	gt t aat at ct	aaaat t t ct t	t aaaacgt gc	cat aact aat	3480
aat acacct t	t t t ct aaagg	aagt gcaccg	ccact t aagc	ct gt acct gc	accacgt gt a	3540
acaact ggaa	cacgt aat ct	at gacat act	gct aat at ag	cagt aact t g	t t ccat t t gt	3600
t t t ggt aaaa	ct act aat aa	aggt ct t gt t	ct gt aagct g	at aat ccat c	acat t cgt at	3660
ggaat aat t t	ct t cat ct gt	at gt aaaat c	t ct aagccag	gt acgt gt t c	acgaagt gcc	3720
at t aaaact g	aagt acgat c	aacat ct ggt	aaagcaccat	ct aaacgt t c	t t cat at aat	3780
at act cat at	ggaat t gt t c	aaaaaaagt t	aaaaact t t t	at ct t t ct cc	at t t t t ct ct	3840
t t aaaat gac	t t t t t gct t t	t t t ct t t t t t	t t t t gcaat g	caaaaaaaag	aaaaaaacaa	3900
at gagaaaaa	t gaat t t at g	gaaat ct acc	t t t aggt ct a	aat t t aaact	t ct ct at at t	3960
t t t t t gt t t t	gt t t aaat t t	t t t t t t ct t c	t ct gaaat ag	aat t at at ca	t aaaat t aaa	4020
acaaaaaaat	t t t t t gt at c	t t t t t t t t ga	t caaaaaaaa	agct ct gcac	t t t t ct t t t t	4080
t ct t t t t t aa	t ggaagaagt	aaaaaacat t	t ct t ccat t a	aaaaaaat t a	t cct t gacgt	4140
t gt t t at gt t	t aggat t t gt	acaaat aacc	at t act gt ac	ct t t t cgacg	aat t aaacga	4200
cat t t agt ac	aaat t t t t t t	aact gaagaa	cgaact t t ca	t aaaat gaat	t t t aaaaat t	4260
at t t t gt gca	at t t gat gt t	t gcacgat t t	gaaaacat t t	gat t t ct t t t	gt gct agcac	4320
t agaggccgg	ccact agt ct	cgagcccgt g	ct ggt t gt ca	t t gcct ct gg	at aat t t t t c	4380
t cgaact at g	cct gcgcgt t	gat accaat c	caat ggat ct	acaggcagaa	cggcct ct ag	4440
cgg t t t t t t t	t act t t t gct	t t t t t t gct t	t t gt t caaag	aaaaaaaat	acaaaat aaa	4500
aaaaact aaa	at gaaaaaac	aaagaat t ct	aaaat t cat a	aaaaaaat t a	aaaccaat t	4560
t t t t t t t t gg	aaact t t t cc	aaat aat aaa	aaaat caaaa	aaaaat t t t t	ct agt at t t t	4620
t t t cat at t t	t gaaact t t t	t t t gagt t t a	t aaaaaaat a	gaaaaaacaa	at agat gaaa	4680
at t t agaaaa	at t at aaacc	aat aaaaat g	aagt t t t gcg	t agaaaaaaa	at t t agt t t a	4740

0802W01_ST25

ct t g t t c c c c a a g a g c a a g t g g t a a c t t t g a a a a a a a t a t t t a a a c t t a a a a t t t g c t a 4800
 a a g t t t t g a a t t t a t g t t a a a t t t a a a a a a a t t t t t a a a c t a t t t t t t t a t g 4860
 t t a a a a a a a t a g t t t t t a t t a t t t t c t a t a a t a t a g t t t a g t t t t t a t t t t t t c a a t t 4920
 t c t t t t t t t t t t c a a g a a a a a g t t t t c c a c g g a t a g a t t t t t a t a g g a t c g a c a a a a 4980
 t g t t c t a t g a a c t t t t c a t a a t g g a g a a a a a a t c a c t g g a t a t a c c a c c g t t g a t a t a t 5040
 c c c a a t g g c a t c g t a a g a a c a t t t t g a g g c a t t t c a g t c a g t t g c t c a a t g t a c c t a t a 5100
 a c c a g a c c g t t c a g c t g g a t a t t a c g g c c t t t t t a a g a c c g t a a g a a a a a t a a g c a c a 5160
 a g t t t t a t c c g g c c t t t a t t c a c a t t c t t g c c c g c c t g a t g a a t g c t c a t c c g g a g t t c c 5220
 g t a t g g c a a t g a a a g a c g g t g a g c t g g t g a t a t g g g a t a g t g t t c a c c c t t g t t a c a c c g 5280
 t t t t c c a t g a g c a a a c t g a a a c g t t t c a t c g c t c t g g a g t g a a t a c c a c g a c g a t t c c g 5340
 g c a g t t t c t a c a c a t a t a t t c g c a a g a t g t g g c g t g t t a c g g t g a a a a c c t g g c c t a t t t 5400
 c c c t a a a g g g t t t a t t g a g a a t a t g t t t t t c g t c a g c g c c a a t c c c t g g g t g a g t t t c a c 5460
 c a g t t t t g a t t t a a a c g t g g c c a a t a t g g a c a a c t t c t t c g c c c c g t t t t c a c t a t g g g 5520
 c a a a t a t t a t a c g c a a g g c g a c a a g g t g c t g a t g c c g c t g g c g a t t c a g g t t c a t c a t g c 5580
 c g t t t g t g a t g g c t t c c a t g t c g g c a g a a t g c t t a a t g a a t t a c a a c a g t a c t g c g a t g a 5640
 g t g g c a g g g c g g g g c g t a a c c t a g a t a t t t g a g a a t t t g t a t t t a a a a c t g a a a a a t t t t 5700
 t g a a c g a a c t c t t t t c a a a a a t a t t a a a a c t t t c t t g a g a t g a t t t a g t g t t a t c t c a a g a 5760
 a a g t t t g t t c t t t t a t t t t t t a a a a t t t t t a a a a a t t t t a t t t t c t t t t a a a c a g g a a a a 5820
 t a a a t a a a g a a a a a a g t g a a t t a a a a a a a a g c t g g g a c t t t c a a a g t g a c c a a t t t t t t a 5880
 c t t t a a a g t t t t t t t t t a t t c a a t a a a a a a t a c t a a a a a a a t a t g a a a g t a t t a c t t a a 5940
 a a t t t c t t a a a a a a a a a g a a t g c c t t t t t t c a a a a a a a a g t t a a a a a a a t a a a g t t t 6000
 t t a c g t a t t g t t t a a a a c t t t t t t g a a a a a a g c a t t c t t t t t c a t t t a a a g a g t t a t c 6060
 t t t t t t a t c t c g t g c a a g t t t t g g a a a t t c a t t t t t g t t a a a t a a c t t t g a c t t t t t a t t 6120
 t c t t a a a t t t t t g g c t t t t c a t t t t t t t g g t t a c a a a t a a a a c t a g a g g c c g g c c c t g 6180
 c a g t t a a t t a a g g a t c c a c t a g t a t t t a a a t t c c t g a t g c g g t a t t t t c t c c t t a c g c a t 6240
 c t g t g c g g t a t t t c a c a c c g c a t a t c g a c c g g t c g a g g a g a a c t t c t a g t a t a t c t a c a t 6300
 a c c t a a t a t t a t t g c c t t a t t a a a a a t g g a a t c c c a a c a a t t a c a t c a a a a t c c a c a t t c 6360
 t c t t c a a a a t c a a t t g t c c t g t a c t t c c t t g t t c a t g t g t g t t c a a a a a c g t t a t a t t t a 6420
 t a g g a t a a t t a t a c t c t a t t t c t c a a c a a g t a a t t g g t t g t t t g g c c g a g c g g t c t a a g g 6480
 c g c c t g a t t c a a g a a a t a t c t t g a c c g c a g t t a a c t g t g g g a a t a c t c a g g t a t c g t a a g 6540
 a t g c a a g a g t t c g a a t c t c t t a g c a a c c a t t a t t t t t t t c c t c a a c a t a a c g a g a a c a c a 6600
 c a g g g g c g c t a t c g c a c a g a a t c a a a t t c g a t g a c t g g a a a t t t t t t g t t a a t t t c a g a g 6660

0802W01_ST25

gt cgcct gac gcat at acct t t t t caact g aaaaaat t ggg agaaaaagga aagggt gagag	6720
cgccggaacc ggct t t t cat at agaata gaagcgt t ca t gact aaat g ct t gcat cac	6780
aat act t gaa gt t gacaat a t t at t t aagg acct at t gt t t t t ccaat a ggt ggt t agc	6840
aat cgt ct t a ct t t ct aact t t t ct t acct t t t acat t t c agcaat at at at at at at at	6900
t t caaggat a t accat t ct a at gt ct gccc ct aagaagat cgt cgt t t t g ccagggt gacc	6960
acgt t ggt ca agaaat caca gccgaagcca t t aagggt t ct t aaagct at t t ct gat gt t c	7020
gt t ccaat gt caagt t cgat t t cgaaaat c at t t aat t gg t ggt gct gct at cgat gct a	7080
cagggt gt t cc act t ccagat gaggcgct gg aagcct ccaa gaaggct gat gccgt t t t gt	7140
t aggt gct gt ggggt ggt cct aaat ggggt a ccggt agt gt t agacct gaa caagggt t t ac	7200
t aaaaat ccg t aaagaact t caat t gt acg ccaact t aag accat gt aac t t t gcat ccg	7260
act ct ct t t t agact t at ct ccaat caagc cacaat t t gc t aaagggt act gact t cgt t g	7320
t t gt cagaga at t agt ggga ggt at t t act t t ggt aagag aaaggaagac gat ggt gat g	7380
gt gt cgct t g ggat agt gaa caat acaccg t t ccagaagt gcaaagaat c acaagaat gg	7440
ccgct t t cat ggccct acaa cat gagccac cat t gcct at t t ggt cct t g gat aaagct a	7500
at gt t t t ggc ct ct t caaga t t at ggagaa aaact gt gga ggaaaccat c aagaacgaat	7560
t ccct acat t gaagggt t caa cat caat t ga t t gat t ct gc cgccat gat c ct agt t aaga	7620
acccaaccca cct aaat ggt at t at aat ca ccagcaacat gt t t ggt gat at cat ct ccg	7680
at gaagcct c cgt t at ccca ggt t cct t gg gt t t gt t gcc at ct gcgt cc t t ggcct ct t	7740
t gccagacaa gaacaccgca t t t ggt t t gt acgaaccat g ccacggt t ct gct ccagat t	7800
t gccaaagaa t aagggt caac cct at cgcca ct at ct t gt c t gct gcaat g at gt t gaaat	7860
t gt cat t gaa ct t gcct gaa gaagggt aagg ccat t gaaga t gcagt t aaa aagggt t t t gg	7920
at gcagggt at cagaact ggt gat t t aggt g gt t ccaacag t accaccgaa gt cgggt gat g	7980
ct gt cgccga agaagt t aag aaaat cct t g ct t aaaaaga t t ct ct t t t t t t at gat at t	8040
t gt acat aaa ct t t at aaat gaaat t cat a at agaaacga cacgaaat t a caaaat ggaa	8100
t at gt t cat a ggggt agacga aact at at ac gcaat ct aca t acat t t at c aagaaggaga	8160
aaaaggagga t gt aaaggaa t acagggt aag caaat t gat a ct aat ggct c aacgt gat aa	8220
ggaaaaagaa t t gcact t t a acat t aat at t gacaaggag gagggcacca cacaaaaagt	8280
t aggt gt aac agaaaaat cat gaaact at ga t t cct aagggt aagt cgggt t t aagacaat gg	8340
gaaaagt t ag at gcct agag t at t gat t at cgagcaaat a t ct t ct cat c t gt gat ct cc	8400
t agt gct agc t aaagaagt t gt t agat t t t at at t ct at t at aaaaat at a aat aat at t t	8460
t t t aggccca t aact cagt t ggt agagt ga t t gcct t aca agcaat aggt cat cgggt t ca	8520

0802W01_ST25

agt ccggt t g ggcct at aaa t act t t t aaa aagt gt caat t t aaccgct t gct ccat aca 8580
 ct t t t t at aa ct t t t t gct t t gct gct t t t t t aaaaaa aagcat aaaa gt agt aaaaa 8640
 at cat t t gt c at gagcaaaa t at t aaaaaa ccat t ct t t a aat aaaat t t t t t at t at 8700
 t ct gt t t aaa aaaat t t t t g t t t gt aaaaa aaaagaaat a aat aaaacca aaat aaaaaa 8760
 at t t t t t ct t t t t t t aaaa t aaat t t t gt t aaaaaat t t t at t t t aagt t caat aaaagt 8820
 t t t t t aat t t t t aat t t t t gat aaagt aa agt t act t gg aaaat t ggt t t at acagaaa 8880
 aaat t at aaa t t at t t at t c at at gt t acg t gaat gt gat t at agt caag ct ct t t t aga 8940
 acaagt t aat caagct at ct ct gacaaaac t ccat t agt t at t caaggct caaact caaa 9000
 agcat t ct t a ggacgt ccag t aacaggcca aact t t agac gt t cgt t gt c at cgt ggt at 9060
 t gt aaat t at gat ccaact g agct t gt aat t act gcacgt gt t ggt acac ct t t agt aac 9120
 aat t gaagct gct t t agaaa gt gct ggt ca aat gt t acca t gt gaaccac ct cat t acgg 9180
 t gaagaagct act t ggggt g gt at ggt agc t t gt ggt t t a gct ggt ccac gt agacct t g 9240
 gagt ggt t ct gt t cgt gat t t t gt at t agg t act cgt at a at t acagggt g ct ggt aaaca 9300
 t t t acgt t t c ggt ggt gaag t t at gaaaaa t gt agct ggt t at gat t t at cacgt t t aat 9360
 ggt aggt agt t acggt t gct t aggcgt t t t aacagaaat t t caat gaaag t t t t accacg 9420
 t ccaagagct t ct t t at cat t acgt cgt ga aat at cat t a caagaagcaa t gt cagaaat 9480
 t gcagaat gg caat t acaac ct t t accaat aagt ggt t t a t gt t at t t t g at aat gct t t 9540
 at ggat caga t t agaaggag gcgaaggt ag t gt t aaagct gct cgt gaat t at t aggt gg 9600
 t gaggaagt a gcaggt cagt t t t ggcaaca at t acgt gaa caacagct t c cat t t t t ct c 9660
 at t accagggt acat t at ggc gt at t agt t t accat ct gat gcaccaat ga t ggact t acc 9720
 aggagaacaa ct t at t gat t ggggaggt gc t ct t cgt t gg t t aaaat caa ct gct gaaga 9780
 t aat caaat c cat cgt at t g cacgt aat gc t ggcggt cac gcaact cgt t t ct cagcagg 9840
 t gat ggt ggt t t cgcacct t t at ct gct cc at t at t ccgt t at cat caac aact t aaaca 9900
 acaat t agat cct t gt ggt g t t t t t aat cc aggacgt at g t acgct gagt t agat t at aa 9960
 agat gat gat gat aaaggac acaaccaccg t cacaacat t agt ct agaa at t t ccat t t 10020
 t t t t cat t t t t t t t agaaa agt t gt at t t t t ct t gat ga aat gaaaat t t caaaaaaga 10080
 aaaat acaat t t t t ct ct ac t t t t t t t ga t t ct t t act t t t t t t t gaat t t t t t t t gt g 10140
 ct t cgt t t t t gaaaaaaact t t t t at t t ga aaaat t t t gt t gaat t aaaa aaaat caact 10200
 caat acat t t t t t t gaact t ct t t t t act t t t t t gct t t gt t gct t t t c t t t t cat t t t 10260
 t at cat t t t t t gct t c gct g ct t t t t at ca t t t t t gct t t t gcaaaaaat gaagaagaaa 10320
 aaaagcgt aa aat gaaaaag aaaaaaagt t t caaagaaaa aaaagcgaaa aagcaagaga 10380
 agat aaaaaa caagaacaaa aaaat t gt t c aaaaacacaa t agaaaat t c t t t aaaaat t 10440

0802W01_ST25

t t t t g a t t t t	t t a t a g a a t t	t g t g a g a a t	a g t a a a a a a	g t c a a a a a t	c a c a a c a t t a	10500
t a a t a g a a a	a a a t c t t g t t	t t g t a a a a t	t t a t a a t t t t	t t a t a t t a t t	t c a a t t t t a a	10560
a t a a g a t c t t	t t g g a g c t g a	a a a a a t a t g a	g a a a t a g t t t	g g a c a a a c t a	g t c t c g a g c c	10620
c g t a t g a t a t	t c t a a g g c g t	t a c g c t g a t g	a a t a t t c t a c	a g a g t t g c c a	t a g g c g t t g a	10680
a c g c t a c a c g	g a c g a t a c g a	a t t t t t g a a t	t a g a t a a a t g	a g t g t t c t c a	a t t t t t t t t t	10740
c t t t g c a t t t	t t t g t t t g t g	t t g a t t t a c a	a a a c a a t a g	a a a a a g a a a	a c a a t a t t t t	10800
c t t t c t a a a a	a a a a a c a a a a	t t g a t g a a a a	a t a g a c a t g a	a c a a a a a a t t	t t g a a a g t t g	10860
a c t t t t t t a a	a a a a t t t t t g	g t a t a a t a c a	a a a a a g a a t	t t t t g g a a a g	g t g g c a g a g t	10920
g g t t g a a t g c	t c t g g t t t t g	a a a a c c a g c g	t g g c t t t a c g	g t c a c c g g g g	g t t c g a a t c c	10980
c t c c c t t t c c	g a t a a t a t a t	a c a a a a a t t t	t t a a a g t t t t	t t g t t t a t t t	t g t a t a g a t a	11040
a a a a a t c t g c	a a t a a a a t t	t c g t t t t t a	t t t a t t c a a a	a a t t c t g t t t	t t t t g a a a a g	11100
a a a a t a a a a a	a a a t g c c a a a	a g t g a g t t t t	t t a t t c a a a t	a t t a g a a a a a	g t t t t t g a a a	11160
a a t t t a a a a a	a a t a g a a a a a	a t t t t t t a t	t t t t t t c a t a	a t t t a a a a a a	t t a t g t t a t a	11220
a t t t a a a t t a	c a a a t a g g t t	t t a t t a a a a a	a t t t t t a c g t	a c a g a t g a a t	t c t a t a a a a t	11280
t a t t t t g g a g	a t c a c c a t a t	g c a a a c a c a g	t t a a c a g a a g	a a a t g c g t c a	a a a t g c t c g t	11340
g c t t t a g a a g	c a g a c a g t a t	c t t a c g t g c t	t g t g t a c a t t	g t g g c t t t t g	t a c t g c t a c a	11400
t g c c c t a c a t	a c c a a c t t c t	t g g t g a t g a a	t t a g a t g g a c	c a a g a g g t c g	t a t t t a t t t a	11460
a t a a a c a a g	t a t t a g a a g g	t a a t g a a g t t	a c t t t a a a a a	c t c a a g a a c a	c t t a g a c c g t	11520
t g t t t a a c a t	g c c g t a a t t g	t g a a a c t a c t	t g t c c a a g t g	g a g t a c g t t a	t c a t a a t t t a	11580
c t t g a t a t a g	g t c g t g a t a t	c g t a g a a c a a	a a a g t t a a a c	g t c c t t t a c c	a g a a c g t a t t	11640
t t a c g t g a a g	g a c t t c g t c a	a g t a g t t c c a	a g a c c a g c a g	t t t t c c g t g c	t t t a a c t c a a	11700
g t a g g t t t a g	t t t t a c g t c c	a t t t t t a c c t	g a a c a a g t a c	g t g c t a a a t t	a c c t g c a g a a	11760
a c t g t t a a a g	c a a a a c c a c g	t c c t c c t t t a	c g t c a t a a a c	g t c g t g t t t t	a a t g t t a g a a	11820
g g t t g c g c t c	a a c c a a c a t t	a t c t c c a a a t	a c a a a t g c a g	c a a c t g c t c g	t g t a t t a g a t	11880
c g t t t a g g t a	t t t c a g t t a t	g c c a g c a a a t	g a a g c a g g t t	g t t g t g g t g c	t g t t g a t t a t	11940
c a c t t a a a t g	c t c a a g a a a a	a g g t t t a g c t	a g a g c t c g t a	a c a a c a t t g a	c g c t t g g t g g	12000
c c a g c a a t c g	a a g c a g g t g c	t g a a g c t a t t	t t a c a a a c t g	c a t c a g g t t g	c g g t g c a t t t	12060
g t t a a a g a a t	a t g g c c a a a t	g t t a a a a a c	g a c g c t t t a t	a t g c t g a t a a	a g c a c g t c a a	12120
g t a a g t g a a c	t t g c t g t t g a	c t t a g t a g a a	t t a t t a c g t g	a a g a a c c t c t	t g a a a a a c t t	12180
g c t a t t c g t g	g t g a t a a a a a	a c t t g c t t t c	c a c t g t c c a t	g t a c t t t a c a	a c a c g c t c a a	12240
a a a c t t a a t g	g t g a a g t a g a	a a a a g t t c t t	t t a a g a t t a g	g t t t t a c t t t	a a c a g a t g t t	12300

0802W01_ST25

cctgattcac acttatgttg tgggtcagct ggtacat acg ctctt acaca cccagactta 12360
 gctcgtcaat tacgtgacaa caaaatgaat gcacttgaaa gtggaaaacc agaaatgatt 12420
 gttacagcta atattggctg ccaaactcac tttagcttctg ctggtcgtac aagtgttcgt 12480
 cattggattg aaattgtaga acaagcattt gaaaaagaag attat aaaga t gat gat gat 12540
 aaaggacaca accaccgtca caaacattaa tctagat ttt at tttttatg aaaaactcag 12600
 gcttaatttta ggcttgagtt tttcattctt tttgaagctc tgaaatttta aaatttctag 12660
 tcttcttttaa tgtttttaaa ttttaaaaaa taaatttctt ctctgctgtg tttttctttt 12720
 tttttgaaaa aacaaagaaa aaaaattttt ttgttttctt ctttgttttt ttatttcttt 12780
 ttgttttgtt ttttttttag ttccagaatc tttagattcaa aaaaaaat tt agtccgattt 12840
 ctccatagga gcaagcagta aaaaat aaaa actgtaat aa aaaat aaaac aaaaatttta 12900
 tttctttttg ttttgcttga acttttcaaa aaaaaattga aaaattcaag caaaacaaaa 12960
 agaaacaaat aaaaaattta tgaattttct actttttcag gagttgaaat ttctccttta 13020
 cttaaaacat attttgctaa aaaaagcgct tgtgttgctt tttttgctac tttttgttc 13080
 caagcatttt ttcgaaattt tttttttgat tttagatgtgc gtttttgtta acctaaaatc 13140
 ttgaaaagat tttactcttt caaattttta tgtttttatt tttttttattc ataaaaaaaa 13200
 acaatacata aaaaat aaagt atttcggctt caaaaaattt tat acaaaaa gttttttgat 13260
 taaaaactca gaaaaaat aa aaaaacaaag tatgaatttt ttgaaaaatt catacctttt 13320
 atttttttgt aatttttagc ctttcaaaaa atttttgaag gcattttttt tttaatcctc 13380
 atgttcttca aaaggatctc tcaatttttt tgaaggagggt ccaaaactca catagattga 13440
 atatcctgtt gcacttaata atagaaacca taaaaaaaaag gt aaagaaaa aagcaggact 13500
 gtccat aatt ctctcatgtt ttttgttcaa atttattctc caataattat attacgacaa 13560
 aaagt aaaaa aaat caaaat ttattcaaaa aaatggctac tggacaact tcaaaagcta 13620
 aatcaagctt atctgatgca ctccaagaac caggatcgt aactccttta ggaactttat 13680
 taagaccgtt aaactctgaa tcaggaaaag tattacctgg atggggaaca actgttttaa 13740
 tgggtgttt cattgtactt tttgctgtat tcttattaat ttttttagaa atttataaca 13800
 gttctttatt attagataat gttactatga gttgggaaac tttagcttct taattcaata 13860
 gaatagtttt attgcttttt ttatttttta ttttatcaaa aatttttttt gcaaaaat aa 13920
 agaataaat a aaattcaaaa aaattataga attagataaa attagtttca agttgaacta 13980
 agttgtcaat aaactttcaa atttgttttc tttttactgt tcatt aagag caat aaaaaa 14040
 aacttttggc ctggcaatc ttttaaaaaa gt cagaatca attctatttt aagaatccta 14100
 tggaaatctat gtatttaatt tttagcaaaat taccagaagc ttatgcacct tttagatccta 14160
 ttgtagatgt tttaaccaatt attcctattt tcttcttatt attagccttt gtatggcaag 14220

0802W01_ST25

cat ct gt aag t t t t agat aa aaaat t t a a a	agt t t t t t t t gat act t t t t g t a a a a a t at	14280
caaaaaaaac t t t t aaat t t t t caat t t t t	t cat t agcaa c t t t agct t t t aat at t agct	14340
aaagt t gct c t caaaaat at aat t t t t t t t	t gact t t t t a t t t t t t at t t t g t t t c t t t	14400
t t t aaaagt t acaacat aaa gaaaat gaaa at agaaaat t	t gt gaaacat aaaaaaaaag	14460
aat gaaat t t t at gt t cgt t t t t g t t t t	at c t t t t cca act aaagt cg gcct ct agct	14520
agaggccggc caaat t t t t t t ccaaaaat t c t at aaaaaat	caaaaaat t a aaaaaaaaag	14580
aaaaaaact t t g t t t t g t gca aaacaaaaat c t t gaat t ca	aaacaaaaat a aagaat t caa	14640
aaagat t t t t t t t aagcaaa aggt aaaat g gaaaaaaat g	t t t t t aaaaa aat t t t t t c t	14700
t t t t t t t a a a g c t t t g c t t t t t t cat gaaa	aaaacaaaagc t t t aaaaaaa agaaaaaat t	14760
t t aaagcaaa aaaaaagaat t aaacacgt ct t t t t t t gga	ggacgacat c cat t at gt gg	14820
aat t cct gt t t t t t c t gaa t aacat t t ac t t t aat t cca	g c t t t aaaaa t t t cacgaat	14880
agct gt t t c g c gac c t t g t c c t g gaccagt t act aaaat t	t t t g c t t cat t t aat gcaaa	14940
t t c acgt gat t t t t t t gcca caact t cagc agct t t t t t t	g c t gcaaat g t t g t t g c t t t	15000
t c t t t t t cca cggaaaccac aagct ccagc agaact ccaa	caaaggact t caccacgaag	15060
at t t gct aat gt aat aat ag t at t at gat g t ccagct t ga	at at aaacaa t t cct cgat a	15120
t gt acgt t t t t t aat t t t t t g taggt gat ac t t t t c t agt t	t gt c t agcca t at gt aaaaa	15180
t t t aact at a aat t t t c t t t t at at t t t t aa at cat t t gat	t t act at c t a aaaaaaat aa	15240
gat t t t gaat c t t t g gaaag aaccat t at t t g gaaagat t	t at t t g t t ca at t c t t t t g t	15300
at t t t t t t t c aaaaaaat t t t t t t t t t t t at c t aat	t t t t t aat gt t t t g t t caaa	15360
cat aaaaaat t t t c t t t t ca aaaaaggaaa aaaaat t t c t	t t t t t t t gaa aaagt aaaaa	15420
t aaaaaaagc t gt agct t t t t t g t t gaat c aaaaaaaaat	aagaaat t gt cct at t t t t a	15480
t gacaaaaga t t caaaaaaa t gaaat aaaa aagaaaaat t	caaact t t ca aacaat gt at	15540
t t t g t at t t t t g g g c c g a g t c g g a t t c g a a	ccaacgt agg cgt aaccagc ggat t t acaa	15600
t c c g c c c c c a t t a a c c a c t c g g g c a t c g g c	c c a t g c t t t t t t t g a c t c a t t t g a t c a t t	15660
t at t t t g c a t g a a t t at act aaagat t at a t t a a c a a a a	at t t t t g a a a t t t c a a t t t t	15720
t t t t a a t t a a a a a c t c t t c t a t t t t t t a a	a a t t c t t t g t t t t g a a t t t t t t t t c a a t	15780
t t a a a a a a a a a c a t a t t a a a a a t a t t t t	t a a a a a t t t t t t g t a t t g a a a a c t t a c a a	15840
t a a t t t t a t a a a t t t t t t t g a a a a t t t t g	t t t t t t t a t t c t a t t g a a a t g a a c a a a a c	15900
aaat t t t t t t t agt t t t t t t t g t t t t t t t g c	t t t g c t g c t t c t t t t g t t t t t a t c a a t a a a	15960
t aaat gaaaa t gaaaaacaa aaat aaacaa t t t g t g t t t t	t agagt t c t a aaat caagaa	16020
aaaaat act t c c c c t t t a a a g a g g a a g t a t	t t t a a a a a a a a a t t a t a g t t t g t c a a t a g t	16080

0802W01_ST25

t t c a a a t t c a a a t t t g a t t t c t t t c c a a a c t t c a c a a g c a g c a g c t a a t t c t g g a g a c c a	16140
t t t a c a a g c t g a g c g g a t a a c a t c a c c a c c t t c a c g a g c t a a a t c a c g a c c t t c g t t a c g	16200
a g c t t g a g t a c a a g c t t c t a a a g c a a c a c g g t t a g c a a c a g c a c c a g g a g c g t t a c c c c a	16260
a g g g t g t c c t a a a g t a c c t c c a c c g a a t t g t a a c a a g c g t c a t c a c c a a a g a t t t c a a c	16320
t a a a g c t g g c a t g t g c c a t a c g t g a a t a c c a c c a g a a g c a a c t g g c a t a g t a c c a c c c a t	16380
a g a a c a c c a g t c t t g a g t g a a g t a a a t a c c a c g g c t a	16417

<210> 115
 <211> 1553
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> codon optimized sequence comprising CC-90

<400> 115	
a t g a g t a t a t t a t a t g a a g a a c g t t t a g a t g g t g c t t t a c c a g a t g t t g a t c g t a c t t c a	60
g t t t t a a t g g c a c t t c g t g a a c a c g t a c c t g g c t t a g a g a t t t t a c a t a c a g a t g a a g a a	120
a t t a t t c c a t a c g a a t g t g a t g g a t t a t c a g c t t a c a g a a c a a g a c c t t t a t t a g t a g t t	180
t t a c c a a a a c a a a t g g a a c a a g t t a c t g c t a t a t t a g c a g t a t g t c a t a g a t t a c g t g t t	240
c c a g t t g t t a c a c g t g g t g c a g g t a c a g g c t t a a g t g g c g g t g c a c t t c c t t t a g a a a a a	300
g g t g t a t t a t t a g t t a t g g c a c g t t t t a a a g a a t t t t a g a t a t t a a c c c a g t t g g t c g t	360
c g t g c a c g t g t t c a a c c t g g t g t t c g t a a t t t a g c a a t t t c a c a a g c a g t t g c t c c t c a t	420
a a c t t a t a t t a c g c t c c a g a t c c t a g t a g t c a a a t t g c t t g t t c a a t t g g t g g t a a t g t a	480
g c a g a g a a t g c t g g t g g c g t a c a c t g t t t a a a a t a t g g t t t a a c a g t a c a c a a c t t a t t a	540
a a a a t t g a a g t a c a a a c t t t a g a c g g t g a g g c t c t t a c t t t a g g t a g t g a t g c a t t a g a t	600
t c t c c t g g t t t t g a t t t a t t a g c t t t a t t c a c a g g a t c a g a a g g a a t g t t a g g t g t t a c a	660
a c a g a a g t a a c t g t t a a a t t a c t t c c a a a a c c a c c t g t t g c t c g t g t t t t a t t a g c a t c t	720
t t t g a t t c a g t t g a a a a a g c a g g a t t a g c t g t a g g c g a t a t c a t t g c a a a t g g t a t a a t t	780
c c a g g a g g a t t a g a g a t g a t g g a t a a c t t a t c a a t t c g t g c a g c t g a a g a t t t t a t a c a c	840
g c t g g t t a t c c a g t a g a t g c t g a a g c a a t t t t a c t t t g t g a a t t a g a t g g a g t a g a g a g t	900
g a t g t t c a g g a a g a c t g t g a g c g t g t a a a c g a c a t a t t a c t t a a a g c t g g t g c t a c t g a t	960
g t t c g t t t a g c a c a a g a t g a a g c t g a a c g t g t a a g a t t t t g g g c a g g t c g t a a a a a t g c t	1020
t t t c c t g c a g t a g g t c g t a t t t c a c c a g a c t a t t a t t g t a t g g a t g g t a c a a t t c c a c g t	1080
c g t g c a t t a c c a g g t g t t t t a g a a g g t a t t g c a c g t c t t t c t c a a c a a t a c g a t t t a c g t	1140
g t a g c t a a t g t a t t t c a t g c t g g c g a t g g t a a t a t g c a t c c a t t a a t c t t a t t c g a t g c t	1200
a a c g a a c c t g g t g a g t t t g c t c g t g c t g a a g a a c t t g g t g g a a a a a t c t t a g a g t t a t g t	1260

0802W01_ST25

gt agaagt ag gt ggt t ct at t agt ggt gag cat ggaat cg gccgt gaaaa aat caat caa 1320
at gt gt gct c aat t caat ag t gat gaaat t act acat t t c acgt gt aaa agct gct t t t 1380
gat ccagacg g t t t act t aa t cct ggt aaa aat at t ccaa c t t t acat cg t t gt gcagag 1440
t t t ggt gct a t gcat gt aca t cacggccat t t acct t t t c cagagct t ga acgt t t t gat 1500
t at aaagat g at gat gat aa aggacacaac caccgt caca aacat t aat c t ag 1553

<210> 116
<211> 1102
<212> DNA
<213> Artificial Sequence

<220>
<223> codon optimized sequence comprising CC-91

<400> 116
at gt t acgt g aat gt gatt a t agt caagct c t t t t agaac aagt t aat ca agct at ct ct 60
gacaaaact c catt agt t at t caaggct ca aact caaaag cat t ct t agg acgt ccagt a 120
acaggccaaa c t t t agacgt t cgt t gt cat cgt ggt at t g t aaat t at ga t ccaact gag 180
ct t gt aat t a ct gcacgt gt t ggt acacct t t agt aacaa t t gaagct gc t t t agaaagt 240
gct ggt caaa t gt t accat g t gaaccacct cat t acggt g aagaagct ac t t ggggt ggt 300
at ggt agct t gt ggt t t agc t ggt ccacgt agacct t gga gt ggt t ct gt t cgt gat t t t 360
gt at t aggt a ct cgt at aat t acaggt gct ggt aaacat t t acgt t t cgg t ggt gaagt t 420
at gaaaaat g t agct ggt t a t gat t t at ca cgt t t aat gg t aggt agt t a cgg t t gct t a 480
ggcgt t t t aa cagaaat t t c aat gaaagt t t t accacgt c caagagct t c t t t at cat t a 540
cgt cgt gaaa t at cat t aca agaagcaat g t cagaaat t g cagaat ggca at t acaacct 600
t t accaat aa gt ggt t t at g t t at t t t gat aat gct t t at ggat cagat t agaaggaggc 660
gaaggt agt g t t aaagct gc t cgt gaat t a t t aggt ggt g aggaagt agc aggt cagt t t 720
t ggcaacaat t acgt gaaca acagct t cca t t t t t ct cat t accaggt ac at t at ggcgt 780
at t agt t t ac cat ct gat gc accaat gat g gact t accag gagaacaact t at t gat t gg 840
ggaggt gct c t t cgt t ggt t aaaat caact gct gaagat a at caaat cca t cgt at t gca 900
cgt aat gct g gcggt cacgc aact cgt t t c t cagcaggt g at ggt ggt t t cgcacct t t a 960
t ct gct ccat t at t ccgt t a t cat caacaa ct t aaacaac aat t agat cc t t gt ggt gt t 1020
t t t aat ccag gacgt at gt a cgct gagt t a gat t at aaag at gat gat ga t aaaggacac 1080
aaccaccgt c acaaacatt a gt 1102

<210> 117
<211> 1272
<212> DNA

<213> Artificial Sequence

<220>

<223> codon optimized sequence comprising CC-92

<400> 117

```

atgcaaacac agttaacaga agaaatgcgt caaaatgctc gtgctttaga agcagacagt      60
atcttacctg ctgtgtgtaca ttgtggcttt tgtactgcta catgccctac ataccaactt      120
cttggatgat gaattagatgg accaagaggt cgtatttatt taataaaaca agtattagaa      180
ggtaatgaag tttactttaaa aactcaagaa cacttagacc gtgttttaac atgccgtaat      240
tgtgaaacta ctgtgtccaag tggagtacgt tatcataat tttacttgat ataggctgat      300
atcgtagaac aaaaagttaa acgtccttta ccagaacgt a ttttacctga aggacttcgt      360
caagt agttc caagaccagc agttttccgt gctttaactc aagt aggttt agttttacct      420
ccat ttttac ctgaacaagt acgtgctaaa ttacctgcag aaactgttaa agcaaaacca      480
cgtcctcctt tacctcat aa acgtcgtgtt ttatgttag aaggttgcgc tcaaccaaca      540
ttatctccaa atacaaatgc agcaactgct cgtgtattag atcgtttagg ttttcagtt      600
atgccagcaa atgaagcagg ttgttgtgtt gctgttgatt atcacttaa t gctcaagaa      660
aaaggtttag ctagagctcg taacaacatt gacgcttggg ggccagcaat cgaagcaggt      720
gctgaagct a tttt acaaac tgcacaggt tgcggtgcat ttgttaaaga atatggccaa      780
atgttaaaaa acgacgcttt atatgctgat aaagcacgtc aagt aagtga acttgctgtt      840
gacttagt ag aattattacg tgaagaacct cttgaaaaac ttgctattcg tggatgat aaa      900
aaacttgctt tccactgtcc atgtacttta caacacgctc aaaaacttaa tggagaagt a      960
gaaaaagttc ttttaagatt aggttttact ttaacagatg ttcctgattc acacttatgt      1020
tgtggttcag ctggtacata cgctcttaca caccagact tagctcgtca attacctgac      1080
aacaaaatga atgcacttga aagtggaaaa ccagaaatga ttgttacagc taatattggc      1140
tgccaaactc acttagcttc tgcgtggtcgt acaagtgttc gtcatggat tgaaattgt a      1200
gaacaagcat tagaaaaaga agattataaa gatgatgatg ataaaggaca caaccaccgt      1260
cacaaacatt aa                                     1272

```

<210> 118

<211> 1179

<212> DNA

<213> Artificial Sequence

<220>

<223> codon optimized sequence comprising H1S3

<400> 118

```

aat tcccgtt ttaagagctt ggtgagcgt aggagtcact gccaggtatc gtttgaacac      60
ggcattagt c agggaagtca taacacagtc ctttcccgca atttctttt tctattactc      120

```

0802W01_ST25

ttggcctcct ct agt acact ct at at t t t t t t at gcct cg gt aat gat t t t cat t t t t t t	180
t t t t ccacct agcggat gac t c t t t t t t t t c t t agc gat t ggcat t at c acat aat gaa	240
t t at acat t a t at aaagt aa t gt gat t t c t t c gaagaat a t act aaaaaa t gagcaggca	300
agat aaacga aggcaaagat gacagagcag aaagccct ag t aaagcgt at t acaaat gaa	360
accaagatt c agat t gcgat ct c t t t aaag ggt ggt cccc t agc gat aga gcact cgat c	420
t t cccagaaa aagaggcaga agcagt agca gaacaggcca cacaat cgca agt gat t aac	480
gt ccacacag gt at aggg t t t c t ggacat at gat acat g ct ct ggccaa gcatt ccggc	540
t ggt cgct aa t cgt t gagt g cat t ggt gac t t acacat ag acgacat ca caccact gaa	600
gact gcggga t t gct ct cgg t caagc t t t t aaagaggccc t aggggccgt gcgt ggagt a	660
aaaagg t t t g gat caggat t t gcgcct t t g gat gaggcac t t t ccagagc ggt ggt agat	720
ct t t cgaaca ggccgt acgc agt t gt cgaa ct t ggt t t gc aaagggagaa agt aggagat	780
ct ct ct t gcg agat gat ccc gcatt t t ct t gaaagct t t g cagaggct ag cagaat t acc	840
ct ccacgt t g at t gt ct gcg aggcaagaat gat cat cacc gt agt gagag t gcgt t caag	900
gct ct t gcgg t t gccat aag agaagccacc t cgcccaat g gt accaacga t gt t ccct cc	960
accaaagg t g t t ct t at gt a gt gacaccga t t at t t aaag ct gcagcat a cgat at at at	1020
acat gt gt at at at gt at ac ct at gaat gt cagt aagt at gt at acgaac agt at gat ac	1080
t gaagat gac aagg t aat gc at cat t ct at acgt gt cat t ct gaacgagg cgcgct t t cc	1140
t t t t t t c t t t t t gct t t t t c t t t t t t t c t c t t gaact	1179

<210> 119
 <211> 4879
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> codon optimized sequence comprising LYS2

<400> 119	
agcagttgct t t ct cct at g ggaagagct t t ct aagt ct g aagaagt aaa cagt t ct t t g	60
ct at t t caca ct t cct ggt t gat ggt cact t gct gcct ga aat at at at a t at gt at gac	120
at at gt act t g t t t t c t t t t t gt gcct t t gt t acgt ct a t at t cat t ga aact gat t at	180
t cgatt t t t c t c t t gct gac cgct t t ct aga ggcat cgcac agt t t t agcg aggaaaact c	240
t t caat agt t t t gccagcgg aat t ccact t gcaat t acat aaaaaat t cc ggcggt t t t t	300
cgcgt gt gac t caat gt cga aat acct gcc t aat gaacat gaacat cgcc caaat gt at t	360
t gaagacccg ct gggagaag t t caagat at at aagt aaca agcagccaat agt at aaaaa	420
aaaat ct gag t t t at t acct t t cct ggaat t t cagt gaaa aact gct aat t at agagaga	480

0802W01_ST25

t at cacagag	t t act cact a	at gact aacg	aaaagggt ct g	gat agagaag	t t ggat aat c	540
caact ct t t c	agt gtt acca	cat gact t t t	t acgcccaca	acaagaacct	t at acgaaac	600
aagct acat a	t t cgt t acag	ct acct cagc	t cgat gt gcc	t cat gat agt	t t t t ct aaca	660
aat acgct gt	cgct t t gagt	gt at gggct g	cat t gat at a	t agagt aacc	ggg gacgat g	720
at at t gt t ct	t t at at t gcg	aat aacaaaa	t ct t aagat t	caat at t caa	ccaacgt ggt	780
cat t t aat ga	gct gt at t ct	acaat t aaca	at gagt t gaa	caagct caat	t ct at t gagg	840
ccaat t t t t c	ct t t gacgag	ct agct gaaa	aaat t caaag	t t gccaaagt	ct ggaaagga	900
cccct cagt t	gt t ccgt t t g	gcct t t t t gg	aaaaccaaga	t t t caaat t a	gacgagt t ca	960
agcat cat t t	agt ggact t t	gct t t gaat t	t ggat accag	t aat aat gcg	cat gt t t t ga	1020
act t aat t t a	t aacagct t a	ct gt at t cga	at gaaagagt	aacct at t gt t	gcggaccaat	1080
t t act caat a	t t t gact gct	gcgct aagcg	at ccat ccaa	t t gcat aact	aaaat ct ct c	1140
t gat caccgc	at cat ccaag	gat agt t t ac	ct gat ccaac	t aagaact t g	ggct ggt gcg	1200
at t t cgt ggg	gt gt at t cac	gacat t t t cc	aggacaat gc	t gaagcct t c	ccagagagaa	1260
cct gt gt t gt	ggagact cca	acact aaat t	ccgacaagt c	ccgt t ct t t c	act t at cgcg	1320
acat caaccg	cact t ct aac	at agt t gccc	at t at t t gat	t aaaacaggt	at caaaagag	1380
gt gat gt agt	gat gat ct at	t ct t ct aggg	gt gt ggat t t	gat ggt at gt	gt gat ggggt g	1440
t ct t gaaagc	cggcgcaacc	t t t t cagt t a	t cgaccct gc	at at ccccca	gccagacaaa	1500
ccat t t act t	aggt gt t gct	aaaccacgt g	gggt t gat t gt	t at t agagct	gct ggacaat	1560
t ggat caact	agt agaagat	t acat caat g	at gaat t gga	gat t gt t t ca	agaat caat t	1620
ccat cgct at	t caagaaaat	ggg accat t g	aaggt ggcaa	at t ggacaat	ggcgaggat g	1680
t t t t ggct cc	at at gat cac	t acaaagaca	ccagaacagg	t gt t gt agt t	ggaccagat t	1740
ccaacccaac	cct at ct t t c	acat ct ggt t	ccgaaggt at	t cct aagggt	gt t ct t ggt a	1800
gacat t t t t c	ct t ggct t at	t at t t caat t	ggat gt ccaa	aaggt t caac	t t aacagaaa	1860
at gat aaat t	cacaat gct g	agcgggt at t g	cacat gat cc	aat t caaaga	gat at gt t t a	1920
caccat t at t	t t t aggt gcc	caat t gt at g	t ccct act ca	agat gat at t	ggg acaccgg	1980
gccgt t t agc	ggaat ggat g	agt aagt at g	gt t gcacagt	t acccat t t a	acacct gcca	2040
t ggggt caat t	act t act gcc	caagct act a	caccat t ccc	t aagt t acat	cat gcgt t ct	2100
t t gt ggggt ga	cat t t t aaca	aaacgt gat t	gt ct gaggt t	acaaacct t g	gcagaaaat t	2160
gccgt at t gt	t aat at gt ac	ggg accact g	aaacacagcg	t gcagt t t ct	t at t t cgaag	2220
t t aaat caaa	aaat gacgat	ccaaact t t t	t gaaaaaat t	gaaagat gt c	at gcct gct g	2280
gt aaagggt at	gt t gaacgt t	cagct act ag	t t gt t aacag	gaacgat cgt	act caaat at	2340
gt ggt at t gg	cgaaat aggt	gagat t t at g	t t cgt gcagg	t ggt t t ggcc	gaagggt t at a	2400

0802W01_ST25

gaggat t acc agaatt gaat aaagaaaaat ttgt gaacaa ct ggt ttgtt gaaaaagat c 2460
 act ggaatt a ttt ggat aag gat aat ggt g aacct ttggag acaat tct gg tt aggt ccaa 2520
 gagat agat t gt acagaacg ggt gat tt ag gt cgt t at ct accaaacggt gact gt gaat 2580
 gt t gcggt ag ggct gat gat caagt t aaaa tt cgt ggggt t cagaat cgaa tt aggagaaa 2640
 t agat acgca cat t t cccaa cat ccat tt gg t aagagaaaa cat t act t t a gt t cgcaaaa 2700
 at gccgacaa t gagccaaca tt gat cacat tt at ggt ccc aagat tt gac aagccagat g 2760
 act t gt ct aa gt t ccaaagt gat gt t ccaa aggaggt t ga aact gaccct at agt t aagg 2820
 gct t aat cgg tt accat ct t tt at ccaagg acat caggac t t t ct t aaag aaaagat t gg 2880
 ct agct at gc t at gcct t cc tt gat t gt gg tt at ggat aa act accat t g aat ccaa t g 2940
 gt aaagt t ga t aagcct aaa ct t caat t cc caact cccaa gcaat t aaat tt ggt agct g 3000
 aaaat acagt tt ct gaaact gacgact ct c agt t t accaa t gt t gagcgc gaggt t agag 3060
 act t at ggt t aagt at at t a cct accaagc cagcat ct gt at caccagat gat t cgt t t t 3120
 t cgat t t agg t ggt cat t ct at ct t ggct a ccaaaat gat t t t t acct t a aagaaaaagc 3180
 t gcaagt t ga t t t accat t g ggcacaat t t t caagt at cc aacgat aaag gcct t t gccg 3240
 cggaaat t ga cagaat t aaa t cat cgggt g gat cat ct ca aggt gaggt c gt cgaaaat g 3300
 t cact gcaaa tt at gcggaa gacgccaaga aat t ggt t ga gacgct acca agt t cgt acc 3360
 cct ct cgaga at at t t t gt t gaacct aat a gt gccgaagg aaaaacaaca at t aat gt gt 3420
 ttgt t accgg t gt cacagga t t t ct gggct cct acat cct t gcagat t t g tt aggacgt t 3480
 ct ccaaagaa ct acagt t t c aaagt gt t t g cccacgt cag ggccaaggat gaagaagct g 3540
 cat t t gcaag at t acaaaag gcaggt at ca cct at ggt ac tt ggaacgaa aaat t t gcct 3600
 caaat at t aa agt t gt at t a ggcgat t t at ct aaaagcca at t t ggt ct t t cagat gaga 3660
 agt ggat gga t t t ggcaaac acagt t gat a t aat t at cca t aat ggt gcg tt agt t cact 3720
 ggggt t t at cc at at gccaaa tt gagggat c caaat gt t at tt caact at c aat gt t at ga 3780
 gct t agccgc cgt cggcaag ccaaagt t ct tt gact t t gt tt cct ccact t ct act ct t g 3840
 acact gaat a ct act t t aat tt gt cagat a aact t gt t ag cgaagggaag ccaggcat t t 3900
 t agaat caga cgat t t aat g aact ct gcaa gcgggct cac t ggt ggat at ggt cagt cca 3960
 aat gggct gc t gagt acat c at t agacgt g caggt gaaag gggcct acgt ggggt gt at t g 4020
 t cagaccagg tt acgt aaca ggt gcct ct g ccaat ggt t c tt caaacaca gat gat t t ct 4080
 t at t gagat t t t t gaaaggt t cagt ccaat t aggt aagat t ccagat at c gaaaat t ccg 4140
 t gaat at ggt t ccagt agat cat gt t gct c gt gt t gt t gt t gct acgt ct tt gaat cct c 4200
 ccaaagaaaa t gaat t ggcc gt t gct caag t aacgggt ca cccaagaat a tt at t caaag 4260

0802W01_ST25

act act t gt a t act t t acac gat t at ggt t acgat gt cga aat cgaaagc t at t ct aaat 4320
 ggaagaaat c at t ggaggcg t ct gt t at t g acaggaat ga agaaaat gcg t t gt at cct t 4380
 t gct acacat ggt ct t agac aact t acct g aaagt accaa agct ccggaa ct agacgat a 4440
 ggaacgccgt ggcat ct t t a aagaaagaca ccgcat ggac aggt gt t gat t ggt ct aat g 4500
 gaat aggt gt t act ccagaa gaggt t ggt a t at at at t gc at t t t t aaac aaggt t ggat 4560
 t t t t acct cc accaact cat aat gacaaac t t ccact gcc aagt at agaa ct aact caag 4620
 cgcaaat aag t ct agt t gct t caggt gct g gt gct cgt gg aagct ccgca gcagct t aag 4680
 gt t gagcat t acgt at gat a t gt ccat gt a caat aat t aa at at gaat t a ggagaaagac 4740
 t t agct t ct t t t cgggt gat gt cact t aaa aact ccgaga at aat at at a at aagagaat 4800
 aaaat at t ag t t at t gaat a agaact gt aa at cagct ggc gt t agt ct gc t aat ggcagc 4860
 t t cat ct t gg t t t at t gt a 4879

<210> 120
 <211> 24607
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> codon optimized sequence comprising | S57-| S116-| S62-| S61

<400> 120
 cgt t t aggt g t aacacaat c t t ggggt gga t ggacaat t a gcggt gaaac agcaacaaat 60
 ccaggt at t t ggagt t at ga aggt gt t gct gcat ct cat a t t at t t t at c t ggt t t at t a 120
 t t ct t agct t cgg t t t ggca ct ggg t t t ac t gggat t t ag agt t at t ccg t gacccaaga 180
 act ggaaaaa ct gcatt aga t t t accaaaa at t t t cggaa t t cact t at t ct t at caggt 240
 ct t t t at gt t t t ggt t t t gg t gct t t ccac gt aacaggt t t at t t ggt cc t ggt at t t gg 300
 gt t t cagat c ct t at ggat t aacaggaagt gt t caaccag t t gct cct t c t t ggggt gct 360
 gat ggg t t t g at cct t t caa ccct ggt ggt at t gcagcgc accacat t gc t gct ggt at t 420
 t t aggt gt t t t agcaggat t at t ccact t a t gt gt acgt c ct t ct at t cg t t t at act t t 480
 ggt t t at caa t ggg t agt at cgaaacagt a t t at caagt a gt at t gct gc t gt t t t ct gg 540
 gct gct t t cg t t gt t gct gg aact at gt gg t at ggt t cag cagct act cc aat t gaat t a 600
 t t t ggt cct a cacgt t at ca at gggaccaaa ggt t t ct t cc aacaagaaat t caaaaacga 660
 gt t caaaciaa gt t t agcagg t ggt t ct t ca ct t t ct gat g ct t gggcgaa aat t ccagaa 720
 aaat t agct t t ct at gatt a t at t ggaaac aaccct gcaa aaggt ggt ct t t t ccgt aca 780
 ggagct at ga at agt ggaga t ggt at t gct gt t ggat ggt t aggt cacgc agt at t t aaa 840
 gat caagat g gt cgt gaat t at acgt acgt cgt at gcct a ct t t ct t t ga aacat t ccca 900
 gt t t t at t aa t t gat aaaga t ggt gt t gt a cgt gct gacg t t cct t t ccg t cgt gct gaa 960

WO1_ST25

t caaaaat at a	gt at t gaaca	agt t ggt gt a	t cagt aact t	t ct acggt gg	t gaat t agat	1020
ggat t aacat	t t aat gat cc	agcaact gt t	aaaaaat at g	ct cgt aaagc	acaat t aggt	1080
gaaat t t t t g	aat t t gat cg	t t caacat t a	caat ct gat g	gt gt at t ccg	t agt agt cca	1140
cgt ggt t ggt	t t act t t t gg	t cacgt t t gc	t t t gct t t at	t at t ct t ct t	t ggacat at t	1200
t ggcat ggt g	cacgt acaat	ct t ccgt gat	gt at t t gct g	gt at t gat ga	t gat ct aaac	1260
gaaagt t t ag	aat t t ggt aa	at acaaaaaa	ct t ggt gat a	caagt t ct gt	t cgt gaagct	1320
t t ct aat t cg	t t t t t t ct c	t t t t t t t ct	t t t t ct ct t	t ggaaaaaga	aaaaacat gt	1380
t t at t t t gaa	t t t t t t gt t t	agaact t t ac	t gt t ct t t t t	t t at t t t aaa	gt gt t t ct gt	1440
t t t t t t t t aa	t acaaaaaact	t t t t t t aaaat	gaat t t aaaa	aacacaaaaa	aagagt t at t	1500
gct at t caaa	at aaacaaga	gt t t aaaaac	aaagt t t t t t	t ct t t agaaa	aaaact t ct t	1560
cat t t t t t t t	gaat t gt t t t	t gaact t t t t	t ct t ct ct t g	ct t t t at cgt	t t t t t ct t c	1620
act t t t t gca	aaaaagt gag	aaaaaacagc	aaagcaaaaa	agt gaaaaaa	agt t caaaaa	1680
caat t caaaa	aagacaaaaac	ct aaaaaaat	at cact t gag	at ggggt ct gg	at t t t t t cca	1740
agcaaaagaa	t t t t gt at t t	t gt t gaaagt	t t t t cat aaa	aat acaaat t	t gcaat t at t	1800
at t ct t aaaa	t caaaat at t	t gt t aaccac	at t t cat t ct	at ggaagcat	t agt t t at ac	1860
t t t t t t at t a	at cggaacat	t aggaat t at	ct t t t t cgca	at t t t ct t t a	gagaaccacc	1920
t cgt at ggt a	aaat aagggc	t cgagact ag	t t t gt ccaaa	ct at t t ct ca	t at t t t t t ca	1980
gct ccaaaaag	at ct t at t t a	aaat t gaaat	aat at aaaaa	at t at aaat t	t t t acaaaac	2040
aagat t t t t t	ct at t t at aa	t gt t gt gat t	t t t t gact t t	t t t t act at t	t ct cacaaat	2100
t ct at aaaaa	at caaaaaat	t t t t aaagaa	t t t t ct at t g	t gt t t t t gaa	caat t t t t t t	2160
gt t ct t gt t t	t t t at ct t ct	ct t gct t t t t	cgct t t t t t t	t ct t t gaaac	t t t t t t t ct t	2220
t t t cat t t t a	cgct t t t t t t	ct t ct t cat t	t t t t gcaaag	caaaaaat ga	t aaaaagcag	2280
cgaagcaaaaa	aat gat aaaa	at gaaaagaa	aagcaacaaa	gcaaaaaagt	aaaaagaagt	2340
t caaaaaaaa	t gt at t gagt	t gat t t t t t t	t aat t caaca	aaat t t t t ca	aat aaaaagt	2400
t t t t t t caaa	aacgaagcac	aaaaaaaat t	caaaaaaaag	t aaagaat ca	aaaaaaagt a	2460
gagaaaaat t	gt at t t t t ct	t t t t t gaaat	t t t cat t t ca	t caagaaaaa	t acaact t t t	2520
ct aaaaaaaa	at gaaaaaaa	t ggaaat t t c	t agat t aacc	gggt gt gt t t a	t gacgat gat	2580
t gt gacct t g	aaagt at aag	t t ct caccac	t t t t at cat c	at cat ct t t g	t aat caccgg	2640
t accagcat g	aact ggacga	gct ccact t g	at aat t gaac	at t t gcagca	t act cacgag	2700
cccat aagt c	at aat gt aca	at t t ct t caa	gact t ggaga	agt t act aat	t cat t acggt	2760
gt t t at caca	t gt t aat t ca	act act t t ga	agat gt ct aa	gt at gaaat t	t t ct cgt caa	2820

0802W01_ST25

t gaacattttc t acagctttc t cattagcag cagataaaac accagt cat a gt accacct g	2880
cacgaccagc t gcat aagca agatccattg atggatattt t acgttgt ca ggttttttga	2940
atgt aagaga t cct aattt a cat aatct a aacgtggcca agt aactt ca ct gcaaggaa	3000
cacggtctgg ccatgacatt gt at at aaaa taggt aaacg catgt caggc cagcct aat t	3060
gagct aat ac t gaagaat ct tgtgtttcaa t cat act at g aat aat t gat tgtggat gaa	3120
t aacaattttc aat at cgt cg tat t cagcac cgaaaagat a atgt gctt ca at aactt ct a	3180
aacctttgtt gaat aaagtt gcact at ct a cagt aat ttt tttgcccat a t tccagtttg	3240
gggtgtttt aa t gcat cagct acttttacct ct t t aagttt t t ct acaggc caat cacgaa	3300
atgcaccacc t gaagctgtt aat at gat ttt t t t aat gc acctt cagggt aagcctt gaa	3360
t acatt gaaa aat t gct gaa tgt t cact gt ct gct ggaag aat ttttaca t tgtgtttgt	3420
t t gcaagt gg t aat acgaat ggaccacct g cgatt aat gt t t cttt at t t gct aaagcaa	3480
t at cttt acc agcct caat a gct gcaact g taggtttt aa t ccagcacaa ccaacaat ac	3540
ctgt aact ac tgt aact gct tctggatgac gagcaactt c aat t acacct tgtt ct cct g	3600
gaat aat ct c t aat t t gt ag t ct aat ct g ct aat gctt c tttt aat t ca t t aat t aaag	3660
at t cat t acg aact gcaact aaagcaggt t taaaacgacg t acct gat ca gct aat aaag	3720
t aacgt t aga accagcagct aat gcaact a cacgaaattt at ct ggattt t ct gcaacaa	3780
tgt ct aaagt t t gagt accg at t gaaccag t t gat cct ac aat act aat a ggt t t aggac	3840
cat cccat ga t t gt ct agga gct t ct ggaa cagcagcacc t ggccaagca ggt ggt ggt t	3900
gt t gct gt t g ct gt act ttt acagagcat t t t act cct t t accaaaacca cgacctt gat	3960
t acgacgacg t aaagaaaaa ccaccagat a at t t aggaat t ggatt gaaa cgagaagt at	4020
ct aaaaat ga aat t gct ttt gat t ct gct g gagat aat ga at t t aat gt t ggt accat at	4080
gaat aaat aa ttt at aat ttt ttt ct gt at a aaccaatttt ccaagt aact t t act t t at c	4140
aaaaat t aaa aaat t aaaaa act ttt at t g aact t aaaaat aaaaat tttt a caaaaat t t a	4200
tttt aaaaaa aagaaaaaat tttttt at t t t ggt ttt at t t at t t ctttt ttttt acaaa	4260
caaaaat ttt ttt aaacaga at aat aaaaa aaat ttt at t t aaagaat gg tttttt aat a	4320
tttt gct cat gacaaat gat ttttt act ac tttt at gctt tttttt aaaa aaagcagcaa	4380
agcaaaaaag t t at aaaaag tgt at ggagc aagcggtt aa at t gacact t ttt aaaagt a	4440
ttt at aggcc caaccggact tgaaccgat g acct at t gct tgt aaggcaa t cact ct acc	4500
aact gagt t a t gggcct aaa aaat at t at t t at at tttt at aat aga at aaaaat ct aac	4560
aact t cttt a gct agcact a ggagaaaat a gcct cgcgga gccat gt gcc at act cgt ct	4620
gcggagcact ct ggt aat gc at at ggt cca caggacatt c gt cgct t ccg ggt at gcgct	4680
ct at gaat t c ccgt ttt aag agct t ggt ga gcgct aggag t cact gccag gt at cgt t t g	4740

0802W01_ST25

aacacggcat t agt cagggga agt cat aaca cagt cct t t c ccgcaat t t t c t t t t t ct at 4800
 t act ct t ggc ct cct ct agt acact ct at a t t t t t t at g cct cgg t aat gat t t t cat t 4860
 t t t t t t t t c cacct agcgg at gact ct t t t t t t ct t a gcgat t ggca t t at cacat a 4920
 at gaat t at a cat t at at aa agt aat gt ga t t t ct t cgaa gaat at act a aaaaat gaggc 4980
 aggcaagat a aacgaaggca aagat gacag agcagaaaagc cct agt aaag cgt at t acaa 5040
 at gaaaccaa gat t cagat t gcgat ct ct t t aaaggg t gg t cccct agcg at agagcact 5100
 cgat ct t ccc agaaaaagag gcagaagcag t agcagaaca ggccacacaa t cgcaagt ga 5160
 t t aacgt cca cacaggt at a ggg t t t ct gg accat at gat acat gct ct g gccaaagcat t 5220
 ccggct ggt c gct aat cgt t gagt gcat t g gt gact t aca cat agacgac cat cacacca 5280
 ct gaagact g cgggat t gct ct cgg t caag ct t t t aaaga ggccct aggg gccgt gcgt g 5340
 gagt aaaaag gt t t ggat ca ggat t t gcgc ct t t ggat ga ggcaact t t cc agagcgg t gg 5400
 t agat ct t t c gaacaggccg t acgcagt t g t cgaact t gg t t t gcaaagg gagaaagt ag 5460
 gagat ct ct c t t gcgagat g at cccgcat t t t ct t gaaag ct t t gcagag gct agcagaa 5520
 t t accct cca cgt t gat t gt ct gcgaggca agaat gat ca t caccgt agt gagagt gcgt 5580
 t caaggct ct t gcgg t t gcc at aagagaag ccacct cgcc caat ggt acc aacgat gt t c 5640
 cct ccaccaa aggt gt t ct t at gt agt gac accgat t at t t aaagct gca gcat acgat a 5700
 t at at acat g t gt at at at g t at acct at g aat gt cagt a agt at gt at a cgaacagt at 5760
 gat act gaag at gacaaggt aat gcat cat t ct at acgt g t cat t ct gaa cgaggcgcg 5820
 t t t cct t t t t t ct t t t t gct t t t t ct t t t t t t t ct ct t g aact accact caat t t gacc 5880
 t agt cct caa acgt t ct gct aaaccgt gt c aat cagt gt c t gct t cct ga gt gaaacccg 5940
 t aagat ct cc t aggaaaat g aat t t t aaaa at t at t t t gt gcaat t t gat gt t t gcacga 6000
 t t t gaaaaca t t t gat t t ct t t t gt caaaa at t t t t at t t t t aaact t t t t t aagt t t 6060
 t gt aaaaaaa gaaat aaaaa at aaat aaaa at aaaaaaaa t gaaact t aa aaacat aagt 6120
 cgct t aaaaat t t t t t t t at t t t t cct t t t t ct t t t t gt t aaat aaaaat a aaaagaaaag 6180
 gaaaaaat aa aaaaaagaaa aaaact t aaa gt t t aat t t t gaaagat gct t t t t t t ct a 6240
 gat t aaccgg t t gt at t t t c t t gat gaat t gt acgt gt t a aat aaaaat t c agct aaagct 6300
 aaat ct t ct g gacgt gt aac t t t aat gt t a t cagcacgac ct t caact aa t t gt ggat ga 6360
 aaaccacagt at t ct aaagc t gat gct t ca t ct gt aat t g t agcacct t c at t t aaagca 6420
 cgt gt t aaac aat cat gt aa t aat t cacgt ggaaaaat t gt ggt gt t aa agcgt gccat 6480
 aaaccat t ac gat caact gt at gagcaat a gcat t t t t ac ct ggt t cagc acgt t t cat t 6540
 gt at cacgaa ct ggagcagc t aaaaat acca cct gt acgt g at gt t t ct ga t aaagct aat 6600

0802W01_ST25						
aaacgagct a	aat cat ct t g	at gt aaacat	ggacgagcag	cat cat gaac t	aaaacccat	6660
t gagcat cac	cagcagct t t	t aaaccagct	aaaact gaat	cagcacgt t c	at caccacca	6720
t caacaact g	t aat t t gt gg	at gat t agct	aat ggt aat t	gagcaaaacg	t gaat cacct	6780
ggt gaaat ag	caat aacaac	acgt t t aaca	cgt ggat gag	ct aat aaagc	at gaact gaa	6840
t gt t ct aaaa	t t gt t t gat t	accaat t gat	aagt at t gt t	t t ggacat t c	t gt t t gcat a	6900
cgacgaccaa	aaccagcagc	aggaacaaca	gcacaaacat	ct aaat gt gt	t gt agcacct	6960
t t gt cgt cat	cgt ct t t at a	at ccat at gg	aat t gt t caa	aaaaagt t aa	aaact t t t at	7020
ct t t ct ccat	t t t t ct ct t t	aaaat gact t	t t t gct t t t t	t ct t t t t t t t	t t gcaat gca	7080
aaaaaaagaa	aaaaacaaat	gagaaaaat g	aat t t at gga	aat ct acct t	t aggt ct aaa	7140
t t t aaact t c	t ct at at t t t	t t t gt t t t gt	t t aaat t t t t	t t t t ct t ct c	t gaaat agaa	7200
t t at at cat a	aaat t aaaac	aaaaaaat t t	t t t gt at ct t	t t t t t t gat c	aaaaaaaaag	7260
ct ct gcact t	t t ct t t t t t c	t t t t t t aat g	gaagaagt aa	aaaacat t t c	t t ccat t aaa	7320
aaaaat t at c	ct t gacgt t g	t t t at gt t t a	ggat t t gt ac	aat aaccat	t act gt acct	7380
t t t cgacgaa	t t aaacgaca	t t t agt acaa	at t t t t t t aa	ct gaagaacg	aact t t cat a	7440
aaat gaat t t	t aaaaaat t at	t t t gt gcaat	t t gat gt t t g	cacgat t t ga	aaacat t t ga	7500
t t t ct t t t gt	gct agcact a	gaggccggcc	act agt ct cg	agcccgt gct	ggt t gt cat t	7560
gcct ct ggat	aat t t t t ct c	gaact at gcc	t gcgcgt t ga	t accaat cca	at ggat ct ac	7620
aggcagaacg	gcct ct agcg	gt t t t t t t t a	ct t t t gct t t	t t t t gct t t t	gt t caaagaa	7680
aaaaaaat ac	aaaaat aaaaa	aaact aaaat	gaaaaaacaa	agaat t ct aa	aat t cat aaa	7740
aaaaat t aaa	accaat t t t	t t t t t t ggaa	act t t t ccaa	at aat aaaaa	aat caaaaaa	7800
aaat t t t t ct	agt at t t t t t	t cat at t t t g	aaact t t t t t	t gagt t t at a	aaaaaat aga	7860
aaaaacaaat	agat gaaaat	t t agaaaaat	t at aaacca	t aaaaat gaa	gt t t t gcgt a	7920
gaaaaaaaaat	t t agt t t act	t gt t ccccaa	gagcaagt gg	t aact t t gaa	aaaaat at t t	7980
aaact t aaaa	at t t gct aaa	gt t t t gaat t	t at gt t aaaa	t t t aaaaaaa	at aaaaat t t	8040
t t aaact at t	t t t t t at gt t	aaaaaaat ag	t t t t t at t at	t t t ct at aat	at agt t t agt	8100
t t t t at t t t	t t t caat t t c	t t t t t t t t t	t caaagaaaa	aagt t t t cca	cggat agat t	8160
t t t at aggat	cgacaaaat g	t t ct at gaac	t t t t cat aat	ggagaaaaaa	at cact ggat	8220
at accaccgt	t gat at at cc	caat ggcat c	gt aaagaaca	t t t t gaggca	t t t cagt cag	8280
t t gct caat g	t acct at aac	cagaccgt t c	agct ggat at	t acggcct t t	t t aaagaccg	8340
t aaagaaaaa	t agcacaaag	t t t t at ccgg	cct t t at t ca	cat t ct t gcc	cgcct gat ga	8400
at gct cat cc	ggagt t ccgt	at ggcaat ga	aagacggg ga	gct ggt gat a	t gggat agt g	8460
t t caccct t g	t t acaccgt t	t t ccat gagc	aaact gaaac	gt t t t cat cg	ct ct ggagt g	8520

0802W01_ST25

aat accacga cgat t ccggc agt t t ct aca cat at at t cg caagat gt gg cgt gt t acgg 8580
t gaaaacct g gcct at t t cc ct aaaggg t t at t gagaat at gt t t t t cg t cagcgccaa 8640
t ccct ggg t g agt t t cacca gt t t t gat t t aaacgt ggcc aat at ggaca act t ct t cgc 8700
ccccgt t t t c act at gggca aat at t at ac gcaaggcgac aagg t gct ga t gccgct ggc 8760
gat t caggt t cat cat gccg t t t gt gat gg ct t ccat gt c ggcagaat gc t t aat gaat t 8820
acaacagt ac t gcgat gagt ggcagggcgg ggcgt aacct agat at t t ga gaat t t gt at 8880
t t aaaact ga aaaat t t t t g aacgaact ct t t t caaaaat at t aaact t t ct t gagat ga 8940
t t t agt gt t a t ct caagaaa gt t t gt t ct t t t at t t t t t a aaat t t t t aa aaat t t t at t 9000
t t ct t t t t aaa caggaaaat a aat aaagaaa aaagt gaat t aaaaaaaagc t gggact t t c 9060
aaagt gacca at t t t t t act t t aaagt t t t t t t at t ca at aaaaaaat act aaaaaaa 9120
t at gaaagt a t t act t t aaaa t t t ct t aaaa aaaaaagaat gcct t t t t t c aaaaaaagt 9180
t t aaaaaaaa t aaagt t t t t acgt at t gt t t aaaact t t t t t gaaaaaa gcat t ct t t t 9240
t t cat t t t aaa gagt t at ct t t t t t at ct cg t gcaagt t t t ggaaat t cat t t t t gt t aaa 9300
t aact t t gac t t t t t at t t c t t aaat t t t t ggct t t t cat t t t t t t t ggt t t acaaat aa 9360
aact agaggc cggccct gca gt t aat t aag gat ccact ag t at t t aaat t cct gat gcgg 9420
t at t t t ct cc t t acgcat ct gt gcgg t at t t cacaccgca t at cgaccgg t cgaggagaa 9480
ct t ct agt at at ct acat ac ct aat at t at t gcct t at t a aaaat ggaat cccaacaat t 9540
acat caaaat ccacat t ct c t t caaaat ca at t gt cct gt act t cct t gt t cat gt gt gt 9600
t caaaaacgt t at at t t at a ggat aat t at act ct at t t c t caacaagt a at t ggt t gt t 9660
t ggccgagcg gt ct aaggcg cct gat t caa gaaat at ct t gaccgcagt t aact gt ggga 9720
at act caggt at cgt aagat gcaagagt t c gaat ct ct t a gcaaccat t a t t t t t t cct 9780
caacat aacg agaacacaca ggggcgct at cgcacagaat caaat t cgat gact ggaaat 9840
t t t t t gt t aa t t t cagaggt cgcct gacgc at at acct t t t t caact gaa aaat t gggag 9900
aaaaaggaaa ggt gagagcg ccggaaccgg ct t t t cat at agaat agaga agcgt t cat g 9960
act aaat gct t gcat caca t act t gaagt t gacaat at t at t t aaggac ct at t gt t t t 10020
t t ccaat agg t ggt t agcaa t cgt ct t act t t ct aact t t t ct t acct t t t acat t t cag 10080
caat at at at at at at at t t caaggat at a ccat t ct aat gt ct gcccct aagaagat cg 10140
t cgt t t t gcc aggt gaccac gt t ggt caag aaat cacagc cgaagccat t aagg t t ct t a 10200
aagct at t t c t gat gt t cgt t ccaat gt ca agt t cgat t t cgaaaat cat t t aat t ggt g 10260
gt gct gct at cgat gct aca ggt gt t ccac t t ccagat ga ggcgct ggaa gcct ccaaga 10320
aggct gat gc cgt t t t gt t a ggt gct gt gg gt ggt cct aa at ggggt acc ggt agt gt t a 10380

0802W01_ST25

gacctgaaca aggtttacta aaaatccgt a aagaacttca attgt acgcc aactt aagac 10440
catgt aactt t gcatccgac tctcttttag acttatctcc aatcaagcca caatttgct a 10500
aaggtagtga cttcgttgtt gt cagagaat t agtgggagg t attt acttt ggt aagagaa 10560
aggaagacga t ggt gat ggt gt cgct tggg at agt gaaca at acaccgt t ccagaagt gc 10620
aaagaatcac aagaatggcc gctttcatgg ccctacaaca tgagccacca ttgcctat tt 10680
ggtcctt gga t aaagctaat gt tttggcct cttcaagatt atggagaaaa actgt ggagg 10740
aaaccatcaa gaacgaattc cctacattga aggttcaaca tcaattgat t gat tctgccg 10800
ccatgatcct agtt aagaac ccaaccacc taaatggtat t ataatcacc agcaacatgt 10860
ttggtagtat catctccgat gaagcctccg ttatcccagg ttcttgggt ttgttgccat 10920
ctgcgtcctt ggcctctttg ccagacaaga acaccgcat t tggtttgt ac gaaccatgcc 10980
acggttctgc tccagatttg ccaaagaat a aggtcaaccc tatcgccact atcttgtctg 11040
ctgcaatgat gttgaaattg tcat tgaact tgcctgaaga aggt aaggcc at tgaagatg 11100
cagttaaaaa ggttttggat gcaggtatca gaactggtga ttt aggtggt tccaacagt a 11160
ccaccgaagt cggtagtgct gtcgccgaag aagtt aagaa aatccttgct taaaaagatt 11220
ctcttttttt atgatatttg t acataaact ttat aaatga aattcat aat agaaacgaca 11280
cgaaattaca aaatggaat a tgttcatagg gt agacgaaa ctat atacgc aatctacata 11340
catttatcaa gaaggagaaa aaggaggatg taaaggaat a caggtaagca aattgat act 11400
aatggctcaa cgtgat aagg aaaaagaatt gcacttt aac at t aat at t g acaaggagga 11460
gggcaccaca caaaaagtt a ggtgt aacag aaaatcatga aactatgat t cct aaggta 11520
gtcggttt aa gacaatggga aaagttagat gcctagagt a ttgat tatcg agcaaatatc 11580
ttctcatctg tgatctccta gtgctagcta aagaagttgt tagattttat at tctat t at 11640
aaaatataa t aatattttt t aggcccat a actcagttgg tagagtgat t gccttacaag 11700
caataggta t cggttcaag tccggttggg cctataaat a cttttaaaaa gtgtcaattt 11760
aaccgcttgc tccatacact ttttat aact tttttgcttt gctgcttttt tt aaaaaaaa 11820
gcataaaagt agt aaaaaat cat tttgtcat gagcaaaat a tt aaaaaacc at tctttaaa 11880
t aaaaat tttt tttattattc tgtttaaaaa aat ttttgtt tgt aaaaaaa aagaaat aaa 11940
t aaaaacaaa at aaaaaaat tttttctttt tttt aaaaat a aat tttgtta aaaaat tttat 12000
tttaagttca ataaaagttt ttt aat tttt t aat ttttga taaagt aaag tt act tggaa 12060
aat tggttt a t acagaaaaa at t at aat t at t t at t cat at ggattat a aagat gat ga 12120
cgacaaagggt atgcacaagt t cacagggtgt t aacgctaaa t t ccagcaac cagcat t aag 12180
aaat t t at ct ccagtggt ag ttgagcgca acgtgaggaa tttgt aggat tctttccaca 12240
aat tgt t cgt gact t aact g aagatggtat tgggtcatcca gaagt aggt g acgctgt agc 12300

0802W01_ST25

t cgt ct t aaa gaagt at t ac aat acaacgc acct ggt ggt aaat gcaat a gaggt t t aac	12360
agt t gt t gca gct t accgt g aact t t ct gg accaggt caa aaagacgct g aaagt ct t cg	12420
t t gt gct t t a gcagt aggat ggt gt at t ga at t at t ccaa gcct t t t t ct t agt t gct ga	12480
cgat at aat g gaccagt cat t aact agacg t ggt caat t a t gt t ggt aca agaaagaagg	12540
t gt t ggt t t a gat gcaat aa at gat t ct t t t ct t t t agaa agct ct gt gt at cgcgt t ct	12600
t aaaaagt at t gccgt caac gt ccat at t a t gt acat t t a t t agagct t t t t ct t caaac	12660
agct t accaa acagaat t ag gacaaat gt t agat t t aat c act gct cct g t at ct aaggt	12720
agat t t aagc cat t t ct cag aagaacgt t a caaagct at t gt t aagt at a aaact gct t t	12780
ct at t cat t c t at t t accag t t gcagcagc t at gt at at g gt t ggt at ag at t ct aaaga	12840
agaacat gaa aacgcaaaag ct at t t t act t gagat ggt gaat act t cc aaat t caaga	12900
t gat t at t t a gat t gt t t t g gcgat cct gc t t t aacaggt aaagt aggt a ct gat at t ca	12960
agat aacaaa t gt t cat ggt t agt t gt gca at gct t acaa agagt aacac cagaacaacg	13020
t caact t t t a gaagat aat t acggt cgt aa agaaccagaa aaagt t gct a aagt t aaaga	13080
at t at at gag gct gt aggt a t gagagccgc ct t t caacaa t acgaagaaa gt agt t accg	13140
t cgt ct t caa gagt t aat t g agaaacat t c t aat cgt t t a ccaaaagaaa t t t t ct t agg	13200
t t t agct cag aaaat at aca aacgt caaaa at caggt cca agat ct t aat ct agaaat t t	13260
ccat t t t t t t cat t t t t t t t t agaaaagt t gt at t t t t ct t gat gaaat g aaaat t t caa	13320
aaaagaaaaa t acaat t t t t ct ct act t t t t t t t gat t ct t t act t t t t t t t t gaat t t t t	13380
t t t gt gct t c gt t t t t gaaa aaaact t t t t at t t gaaaaa t t t t gt t gaa t t aaaaaaaa	13440
t caact caat acat t t t t t t t t gaact t ct t	13500
cat t t t t at c at t t t t t gct t cgct gct t t t t at cat t t t t t gct t t gca aaaaat gaag	13560
aagaaaaaaa gcgt aaaat g aaaaagaaaa aaagt t t caa agaaaaaaa gcgaaaaagc	13620
aagagaagat aaaaaacaag aacaaaaaaa t t gt t caaaa acacaat aga aaat t ct t t a	13680
aaaat t t t t t gat t t t t t at agaat t t gt g agaaat agt a aaaaaagt ca aaaaat caca	13740
acat t at aaa t agaaaaaat ct t gt t t t gt aaaaaat t t at aat t t t t t at at t at t t caa	13800
t t t t aaat aa gat ct t t t t gg agct gaaaaa at at gagaaa t agt t t ggac aaact agt ct	13860
cgagcccgt a t gat at t ct a aggcgt t acg ct gat gaat a t t ct acagag t t gccat agg	13920
cgt t gaacgc t acacggacg at acgaaagc agt t gct t t c t cct at ggga agagct t t ct	13980
aagt ct gaag aagt aaacag t t ct t t gct a t t t cact t t cct ggt t gat ggt cact t gc	14040
t gcct gaaat at at at at at gt at gacat a t gt act t gt t t t ct t t t t t t t t gcct t t gt t	14100
acgt ct at at t cat t gaaac t gat t at t cg at t t t ct t ct t gct gaccgc t t ct agaggc	14160

0802W01_ST25

at cgcacagt t t t agcgagg aaaact ct t c aat agt t t t g ccagcggaat t ccaact t gca 14220
 at t acat aaa aaat t ccggc ggt t t t t cgc gt gt gact ca at gt cgaaat acct gcct aa 14280
 t gaacat gaa cat cgcccaa at gt at t t ga agacccgct g ggagaagt t c aagat at at a 14340
 agt aacaagc agccaat agt at aaaaaaaaa at ct gagt t t at t acct t t c ct ggaat t t c 14400
 agt gaaaaac t gct aat t at agagagat at cacagagt t a ct cact aat g act aacgaaa 14460
 aggt ct ggat agagaagt t g gat aat ccaa ct ct t t cagt gt t accacat gact t t t t ac 14520
 gcccaacaaca agaacct t at acgaaacaag ct acat at t c gt t acagct a cct cagct cg 14580
 at gt gcct ca t gat agt t t t t ct aacaaat acgct gt cgc t t t gagt gt a t gggct gcat 14640
 t gat at at ag agt aaccggg gacgat gat a t t gt t ct t t a t at t gcgaat aacaaaat ct 14700
 t aagat t caa t at t caacca acgt ggt cat t t aat gagct gt at t ct aca at t aacaat g 14760
 agt t gaacaa gct caat t ct at t gaggcca at t t t t cct t t gacgagct a gct gaaaaaa 14820
 t t caaagt t g ccaagat ct g gaaaggaccc ct cagt t gt t ccgt t t ggcc t t t t t ggaaa 14880
 accaagat t t caaat t agac gagt t caagc at cat t t agt ggact t t gct t t gaat t t gg 14940
 at accagt aa t aat gcgcat gt t t t gaact t aat t t at aa cagct t act g t at t cgaat g 15000
 aaagagt aac cat t gt t gcg gaccaat t t a ct caat at t t gact gct gcg ct aagcgat c 15060
 cat ccaat t g cat aact aaa at ct ct ct ga t caccgcat c at ccaaggat agt t t acct g 15120
 at ccaact aa gaact t gggc t ggt gcgat t t cgt ggggt g t at t cacgac at t t t ccagg 15180
 acaat gct ga agcct t ccca gagagaacct gt gt t gt gga gact ccaaca ct aaat t ccg 15240
 acaagt cccg t t ct t t cact t at cgcgaca t caaccgcac t t ct aacat a gt t gcccat t 15300
 at t t gat t aa aacaggt at c aaaagaggt g at gt agt gat gat ct at t ct t ct aggggt g 15360
 t ggat t t gat ggt at gt gt g at ggggt gt ct t gaaagccgg cgcaacct t t t cagt t at cg 15420
 accct gcat a t cccccagcc agacaaacca t t t act t agg t gt t gct aaa ccacgt ggggt 15480
 t gat t gt t at t agagct gct ggacaat t gg at caact agt agaagat t ac at caat gat g 15540
 aat t ggagat t gt t t caaga at caat t cca t cgct at t ca agaaaat ggt accat t gaag 15600
 gt ggcaaat t ggacaat ggc gaggat gt t t t ggct ccat a t gat cact ac aaagacacca 15660
 gaacaggt gt t gt agt t gga ccagat t cca acccaacct at ct t t caca t ct ggt t ccg 15720
 aaggt at t cc t aagggt gt t ct t ggt agac at t t t t cct t ggct t at t at t t caat t gga 15780
 t gt ccaaaag gt t caact t a acagaaaat g at aaat t cac aat gct gagc ggt at t gcac 15840
 at gat ccaat t caaagagat at gt t t acac cat t at t t t t aggt gcccaa t t gt at gt cc 15900
 ct act caaga t gat at t ggt acaccggggt gt t t agcgga at ggat gagt aagt at ggt t 15960
 gcacagt t ac ccat t t aaca cct gccat gg gt caat t act t act gcccaa gct act acac 16020
 cat t ccct aa gt t acat cat gcgt t ct t t g t ggggt gacat t t t aacaaaa cgt gat t gt c 16080

0802W01_ST25

t gaggt t aca aacct t ggca gaaaat t gcc gt at t gt t aa t at gt acggt accact gaaa	16140
cacagcgt gc agt t t ct t at t t cgaagt t a aat caaaaaa t gacgat cca aact t t t t ga	16200
aaaaat t gaa agat gt cat g cct gct ggt a aaggt at gt t gaacgt t cag ct act agt t g	16260
t t aacaggaa cgat cgt act caaat at gt g gt at t ggcga aat aggt gag at t t at gt t c	16320
gt gcaggt gg t t t ggccgaa ggt t at agag gat t accaga at t gaat aaa gaaaaat t t g	16380
t gaacaact g gt t t gt t gaa aaagat cact ggaat t at t t ggat aaggat aat ggt gaac	16440
ct t ggagaca at t ct ggt t a ggt ccaagag at agat t gt a cagaacgggt gat t t aggt c	16500
gt t at ct acc aaacggt gac t gt gaat gt t gcggt agggc t gat gat caa gt t aaaat t c	16560
gt ggg t cag aat cgaat t a ggagaaat ag at acgcacat t t cccaacat ccat t ggt aa	16620
gagaaaacat t act t t agt t cgcaaaaat g ccgacaat ga gccaacat t g at cacat t t a	16680
t ggt cccaag at t t gacaag ccagat gact t gt ct aagt t ccaaagt gat gt t ccaaagg	16740
aggt t gaaac t gaccct at a gt t aagggt t aat cggt t a ccat ct t t t a t ccaaggaca	16800
t caggact t t ct t aaagaaa agat t ggct a gct at gct at gcct t cct t g at t gt ggt t a	16860
t ggat aaact accat t gaat ccaaat ggt a aagt t gat aa gcct aaact t caat t cccaa	16920
ct cccaagca at t aaat t t g gt agct gaaa at acagt t t c t gaaact gac gact ct cagt	16980
t t accaat gt t gagcgcgag gt t agagact t at ggt t aag t at at t acct accaagccag	17040
cat ct gt at c accagat gat t cgt t t t t cg at t t aggt gg t cat t ct at c t t ggct acca	17100
aaat gat t t t t acct t aaag aaaaagct gc aagt t gat t t accat t gggc acaat t t t ca	17160
agt at ccaac gat aaaggcc t t t gccgcgg aaat t gacag aat t aaat ca t cgggt ggat	17220
cat ct caagg t gaggt cgt c gaaaat gt ca ct gcaaat t a t gcggaagac gccaaagaaat	17280
t ggt t gagac gct accaagt t cgt acccct ct cgagaat a t t t t gt t gaa cct aat agt g	17340
ccgaaggaaa aacaacaat t aat gt gt t t g t t accggt gt cacaggat t t ct gggct cct	17400
acat cct t gc agat t t gt t a ggacgt t ct c caaagaact a cagt t t caaa gt gt t t gccc	17460
acgt cagggc caaggat gaa gaagct gcat t t gcaagat t acaaaaaggca ggt at cacct	17520
at ggt act t g gaacgaaaaa t t t gcct caa at at t aaagt t gt at t aggc gat t t at ct a	17580
aaagccaat t t ggt ct t t ca gat gagaagt ggat ggat t t ggcaaacaca gt t gat at aa	17640
t t at ccat aa t ggt gcgt t a gt t cact ggg t t t at ccat a t gccaaat t g agggat ccaa	17700
at gt t at t t c aact at caat gt t at gagct t agccgccgt cggcaagcca aagt t ct t t g	17760
act t t gt t t c ct ccact t ct act ct t gaca ct gaat act a ct t t aat t t g t cagat aaac	17820
t t gt t agcga agggaagcca ggcat t t t ag aat cagacga t t t aat gaac t ct gcaagcg	17880
ggct cact gg t ggat at ggt cagt ccaaat gggct gct ga gt acat cat t agacgt gcag	17940

0802W01_ST25

gt gaaagggg cct acgt ggg t gt at t gt ca gaccagggt t a cgt aacaggt gcct ct gcca 18000
 at ggt t ct t c aaacacagat gat t t ct t at t gagat t t t t gaaaggt t ca gt ccaat t ag 18060
 gt aagat t cc agat at cgaa aat t ccgt ga at at ggt t cc agt agat cat gt t gct cgt g 18120
 t t gt t gt t gc t acgt ct t t g aat cct ccca aagaaaat ga at t ggccgt t gct caagt aa 18180
 cgggt caccc aagaat at t a t t caaagact act t gt at ac t t t acacgat t at ggt t acg 18240
 at gt cgaaat cgaaagct at t ct aaat gga agaaat cat t ggaggcgt ct gt t at t gaca 18300
 ggaat gaaga aaat gcgt t g t at cct t t gc t acacat ggt ct t agacaac t t acct gaaa 18360
 gt accaaagc t ccggaact a gacgat agga acgccgt ggc at ct t t aaag aaagacaccg 18420
 cat ggacagg t gt t gat t gg t ct aat ggaa t aggt gt t ac t ccagaagag gt t ggt at at 18480
 at at t gcat t t t t aaacaag gt t ggat t t t t acct ccacc aact cat aat gacaaact t c 18540
 cact gccaaag t at agaact a act caagcgc aaat aagt ct agt t gct t ca ggt gct ggt g 18600
 ct cgt ggaag ct ccgcagca gct t aaggt t gagcat t acg t at gat at gt ccat gt acaa 18660
 t aat t aaat a t gaat t agga gaaagact t a gct t ct t t t c ggggt gat gt c act t aaaaac 18720
 t ccgagaat a at at at aat a agagaat aaa at at t agt t a t t gaat aaga act gt aaat c 18780
 agct ggcgt t agt ct gct aa t ggcagct t c at ct t ggt t t at t gt act t t caact t ct ag 18840
 aggagaaaag t at t gacat g agcgt cccg gcacaacggc caaagaagt c t ccaat t t ct 18900
 t at t t ct t t t t gaat t agat aaat gagt gt t ct caat t t t t t t t ct t t g cat t t t t t gt 18960
 t t gt gt t gat t t acaaaaaac aat agaaaaa agaaaacaat at t t t ct t t c t aaaaaaaa 19020
 caaaat t gat gaaaaat aga cat gaacaaa aaat t t t gaa agt t gact t t t t aaaaaat 19080
 t t t t ggt at a at acaaaaaa agaat t t t t g gaaaggt ggc agagt ggt t g aat gct ct gg 19140
 t t t t gaaaac cagcgt ggct t t acggt cac cgggggt t cg aat ccct ccc t t t ccgat aa 19200
 t at at acaaa aat t t t t aaa gt t t t t t gt t t at t t t gt at agat aaaaaa t ct gcaat aa 19260
 aaat t t cgt t t t t at t t at t caaaaat t c t gt t t t t t g aaaagaaaat aaaaaaat g 19320
 ccaaaagt ga gt t t t t t at t caaat at t ag aaaaagt t t t t gaaaaat t t aaaaaat ag 19380
 aaaaaat t t t t t t t t t t cat aat t t a aaaaat t at g t t at aat t t a aat t acaaat 19440
 aggt t t t at t aaaaaat t t t t acgt acaga t gaat t ct at aaaaat t at t t t ggagat cac 19500
 cat at ggt ac caacat caat t ct t aat act gt at caact a t t cacagt t c t cgt gt aact 19560
 t ct gt t gat c gt gt t ggt gt t t t aagt t t a agaaat t ct g at t cagt t ga at t t acacgt 19620
 cgt cgt agt g gat t t agt ac t ct t at t t ac gaat cacct g gt agacgt t t t gt agt acgt 19680
 gcagct gaaa ct gat acaga t aaagt aaaa t ct caaact c ct gat aaagc t cct gcaggt 19740
 ggt t ct t caa t t aaccaact t t t aggaat a aaaggt gct a gt caagaaac aaacaaat gg 19800
 aaaat acgt t t acaact t ac t aaaccagt a acat ggccac ct t t agt at g ggggt gt agt t 19860

0802W01_ST25

t gt ggt gct g ct gcat cagg caact t ccat t ggacaccag aagat gt agc t aaaagt at t	19920
t t at gt at ga t gat gt ct gg t cct t gt t t a acagg t t at a cacaact at t aat gat t gg	19980
t at gacagag acat t gat gc aat caat gaa cct t accgt c ct at accaag t ggt gct at t	20040
t cagaaccag aagt t at t ac acaagt t t gg gt t ct t t t ac t t ggt ggt ct t ggaat t gct	20100
ggg at ct t ag at gt at gggc t ggt cat aca acacct acag t at t ct at ct t gct t t agga	20160
ggg agt t t ac t t t ct t acat t t act cagct ccacct t t aa aact t aaaca aaat ggt t gg	20220
gt t ggt aat t t cgct t t agg t gct t cat at at at cat t ac ct t ggt gggc aggt caagct	20280
ct t t t t ggca cat t aacacc agacgt t gt t gt act t act c t t ct t t acag t at agct ggc	20340
t t aggt at cg caat t gt aaa t gact t t aaa t cagt t gaag gt gat agagc act t ggt t t a	20400
cagagt t t ac cagt t gct t t t ggaacagaa act gct aaat ggat at gt gt t ggcgct at t	20460
gacat aacac aat t aagt gt agcagg t t ac t t act t gct a gt ggt aaacc at at t acgct	20520
t t agct t t ag t agct t t aat cat t cct caa at t gt t t t t c agt t caaat a ct t t ct t aaa	20580
gat cct gt aa aat acgacgt aaaat at caa gcat ct gct c aaccat t t t t agt t t t aggt	20640
at t t t t cgt aa ct gct ct t gc aagt caacac ggt accggg t g at t at aaaga cgat gat gac	20700
aaat cagg t g aaaact t at a ct t t caaggt cacaat cat c gt cacaacaa caccgg t t aa	20760
t ct agat t t t at t t t t at g aaaaact cag gct t aat t t a ggct t gagt t t t t cat t ct t	20820
t t t gaagct c t gaaat t t t a aaat t t ct ag t ct t ct t t aa t gt t t t t aaa t t t t aaaaaa	20880
t aaat t t ct t ct ct gct gt g t t t t t ct t t t t t t gaaaa aacaaagaaa aaaaat t t t t	20940
t t gt t t t t ct t ct t t gt t t t t t at t t t t t t t gt t t t t t ag t t t cagaat c	21000
t t t gat t caa aaaaaaat t t agt ccgat t a ct ccat agga gcaagcagt a aaaaat aaaa	21060
act gt aat aa aaaat aaac aaaaat t t t a t t t ct t t t t g t t t t gct t ga act t t t caaa	21120
aaaaaat t ga aaaat t caag caaaacaaaa agaaacaaat aaaaaat t t a t gaat t t t ct	21180
act t t t t cag gagt t gaaat t t ct cct t t a ct t aaaaacat at t t t gct aa aaaaagcgct	21240
t gt gt t gct t t t t t gct ac t t t t gt t t c caagcat t t t t t cgaat at t t t t t t gat	21300
t t t gat gt gc gt t t t t gt t a acct aaaaat c t t gaaaagat t t act ct t t t caaat t t t t a	21360
t gt t t t t at t t t t t t at t c at aaaaaaaa acaat acat a aaaat aaagt at t t cggct t	21420
caaaaaat t t t at acaaaaa gt t t t t t gat t aaaaact ca gaaaaaat aa aaaaacaaag	21480
t at gaat t t t t t gaaaaat t cat acct t t t at t t t t t gt aat t t t t agc ct t t caaaaa	21540
at t t t t gaag gcat t t t t t t t t t aat cct c at gt t ct t ca aaaggat ct c t caat t t t t t	21600
t gaaggagg t ccaaaact ca cat agat t ga at at cct gt t gcaact t aat a at agaaacca	21660
t aaaaaaaag gt aaagaaaa aagcaggact gt ccat aat t ct t t cat gt t t t t gt t caa	21720

0802W01_ST25

atttatctc caataattat attacgacaa aaagt aaaaa aaat caaat ttattcaaaa 21780
 aaatggctac tggacaact tcaaaagcta aatcaagctt atctgatgca ctccaagaac 21840
 caggtatcgt aactccttta ggaactttat taagaccgtt aaactctgaa tcaggaaaag 21900
 tattacctgg atggggaaca actgttttaa tgggtgtttt cattgtactt tttgctgtat 21960
 tcttattaat ttttttagaa attataaca gtctttatt attagataat gtactatga 22020
 gttgggaaac tttagcttct taattcaata gaatagtttt attgcttttt ttatttttta 22080
 ttttatcaaa aatttttttt gcaaaaat aaagaat aaata aaattcaaaa aaattataga 22140
 attagataaa attagtttca agttgaacta agttgtcaat aaactttcaa atttgttttc 22200
 tttttactgt tcatlaagag caataaaaaa aacttttggc ctggcaatc ttttaaaaaa 22260
 gtcagaatca attctatttt aagaatccta tggaatctat gtatttaatt tttagcaaaat 22320
 taccagaagc ttatgcacct tttagatccta ttgtagatgt tttaccaatt attcctattt 22380
 tcttcttatt attagccttt gtatggcaag catctgt aag ttttagat aa aaaattttaa 22440
 agtttttttt gatacttttg taaaaaat at caaaaaaac ttttaaat tt ttttcaattt 22500
 tcatlagcaa ctttagcttt aatattagct aaagt t gctc tcaaaaat at aatttttttt 22560
 tgacttttta tttttttatt ttgtttcttt tttaaaagt tacaacataaa gaaaatgaaa 22620
 atagaaaatt tgtgaaacat aaaaaaaaag aatgaaattt ttatgttcgt tttttgtttt 22680
 atcttttcca actaaagtcg gcctctagct agaggccggc caaatttttt tccaaaattc 22740
 tataaaaaat caaaaaatta aaaaaaaaag aaaaaacttt gttttgtgca aaacaaaaat 22800
 cttegaattca aaacaaaata aagaattcaa aaagattttt ttttagcaaa aggtaaaatg 22860
 gaaaaaaatg tttttaaaaa aattttttct tttttttaaa gctttgcttt ttcatgaaa 22920
 aaaacaaagc tttaaaaaaa agaaaaaat tttaagcaaa aaaaagaatt aaacacgtct 22980
 tttttttgga ggacgacatc cattatgtgg aattcctgtt ttttctcgaa taacatttac 23040
 tttaatcca gctttaaaaa ttcacgaat agctgtttcg cgacctgtc ctggaccagt 23100
 tactaaaatt tttgcttcat ttaatgcaaa ttcacgtgat ttttttgcca caacttcagc 23160
 agcttttttt gctgcaaatg ttgttgcttt tctttttcca cggaaccac aagctccagc 23220
 agaactccaa caaaggactt caccacgaag atttgtaat gtataatag tattatgatg 23280
 tccagcttga atataaaca ttctcgata tgtacgtttt ttatttttg taggtgatac 23340
 ttttctagt tgtctagcca tatgtaaaaa tttaactata aatttctttt atatttttaa 23400
 atcatttgat ttaactatcta aaaaaaat aa gat tt tgaat ctttggaaag aaccattatt 23460
 tggaaagatt tatttgttca attcttttgt atttttttc aaaaaaat tt ttagaattt 23520
 tttatctaatt ttttaaatgt ttgttcaaa cataaaaat tttcttttca aaaaaggaaa 23580
 aaaaatttct tttttttgaa aaagt aaaaa taaaaaagc tgtagctttt ttgttgaatc 23640

0802W01_ST25

aaaaaaaaat aagaaattgt cctattttta tgacaaaaga ttcaaaaaaa t gaaat aaaa	23700
aagaaaaat t caaactttca aacaatgtat ttigtat ttt tgggccgagt cggattcgaa	23760
ccaacgtagg cgt aaccagc ggattt acaa tccgccccca tt aaccact c gggcat cggc	23820
ccatgctttt ttttgactca ttgatcat t attttgc at gaattat act aaagattat a	23880
tt aacaaaaa atttttgaaa tttcaatttt ttttaattaa aaactcttct attttttt aa	23940
aattctttgt tttgaatttt ttttttcaat tt aaaaaaaa aacatattaa aaaatatttt	24000
t aaaaaattt tttgtattga aaactt acaa taattttat a aatttttttg aaaaattttg	24060
ttttttttat tctattgaaa tgaacaaaac aaattttttt agtttttttt gtttttttgc	24120
tttgctgctt cttttgtttt tatcaataaa taaatgaaaa t gaaaaacaa aaat aaacaa	24180
tttgtgtttt tagagttcta aaatcaagaa aaaaat actt cccctttaaa gaggaagt at	24240
ttt aaaaaa aattat agtt tgtcaatagt ttcaaattca aatttgattt ctttccaac	24300
ttcacaagca gcagct aatt ctggagacca tttacaagct gagcggat aa catcaccacc	24360
ttcacgagct aaatcacgac cttcgttacg agcttgagta caagcttcta aagcaacacg	24420
gttagcaaca gcaccaggag cgttacccca aggggtgtcct aaagtacct c caccgaattg	24480
t aaacaagcg tcatcaccaa agatttcaac taaagctggc atgtgccat a cgtgaat acc	24540
accagaagca actggcat ag taccacccat agaacaccag tcttgagtga agt aaat acc	24600
acggct a	24607

<210> 121
 <211> 1529
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> codon optimized sequence comprising IS57

<400> 121	
atggtagcaaa cattaaattc attatctcca gcagaatcaa aagcaatttc atttttagat	60
acttctcgtt tcaatccaat tcctaaattat tctgggtggtt tttctttacg t cgt cgt aat	120
caaggtcgtg gttttggt aa aggagt aaaa t gctctgt aa aagt acagca acagcaacaa	180
ccaccacctg cttggccagg t cgtgctgtt ccagaagctc ct agacaatc atgggat ggt	240
cctaaacctat ttagtattgt aggatcaact ggttcaatcg gtactcaaac tttagacatt	300
gttgcagaaa atccagat aa atttcgtgta gttgcat tag ctgctggttc taacgttact	360
ttattagctg atcaggtacg t cgttttaa cctgctttag ttgcagttcg taatgaatct	420
tt aattaatg aattaaaaga agcattagca gat tt agact acaaattaga gattattcca	480
ggagaacaag gtgt aattga agttgctcgt catccagaag cagttacagt agttacaggt	540

0802W01_ST25

at t g t t g g t t g t g c t g g a t t a a a c c t a c a g t t g c a g c t a t t g a g g c t g g t a a a g a t a t t 600
g c t t t a g c a a a t a a a g a a a c a t t a a t c g c a g g t g g t c c a t t c g t a t t a c c a c t t g c a a a c 660
a a a c a c a a t g t a a a a a t t c t t c c a g c a g a c a g t g a a c a t t c a g c a a t t t t t c a a t g t a t t 720
c a a g g c t t a c c t g a a g g t g c a t t a a g a a a a a t c a t a t t a a c a g c t t c a g g t g g t g c a t t t 780
c g t g a t t g g c c t g t a g a a a a a c t t a a a g a g g t a a a a g t a g c t g a t g c a t t a a a c a c c c a 840
a a c t g g a a t a t g g g c a a a a a a a t t a c t g t a g a t a g t g c a a c t t t a t t c a a c a a g g t t t a 900
g a a g t t a t t g a a g c a c a t t a t c t t t t c g g t g c t g a a t a c g a c g a t a t t g a a a t t g t t a t t 960
c a t c c a c a a t c a a t t a t t c a t a g t a t g a t t g a a a c a c a a g a t t c t t c a g t a t t a g c t c a a 1020
t t a g g c t g g c c t g a c a t g c g t t t a c c t a t t t t a t a t a c a a t g t c a t g g c c a g a c c g t g t t 1080
c c t t g c a g t g a a g t t a c t t g g c c a g t t t a g a t t t a t g t a a a t t a g g a t c t c t t a c a t t c 1140
a a a a a c c t g a c a a c g t a a a a t a t c c a t c a a t g g a t c t t g c t t a t g c a g c t g g t c g t g c a 1200
g g t g g t a c t a t g a c t g g t g t t t t a t c t g c t g c t a a t g a g a a a g c t g t a g a a a t g t t c a t t 1260
g a c g a g a a a a t t t c a t a c t t a g a c a t c t t c a a a g t a g t t g a a t t a a c a t g t g a t a a a c a c 1320
c g t a a t g a a t t a g t a a c t t c t c c a a g t c t t g a a g a a a t t g t a c a t t a t g a c t t a t g g g c t 1380
c g t g a g t a t g c t g c a a a t g t t c a a t t a t c a a g t g g a g c t c g t c c a g t t c a t g c t g g t a c c 1440
g g t g a t t a c a a a g a t g a t g a t g a t a a a a g t g g t g a g a a c t t a t a c t t t c a a g g t c a c a a t 1500
c a t c g t c a t a a a c a c a c c g g t t a a t c t a g 1529

<210> 122
<211> 749
<212> DNA
<213> Artificial Sequence

<220>
<223> codon optimized sequence comprising IS116

<400> 122
a t g g a t t a t a a a g a c g a t g a c g a c a a a g g t g c t a c a a c a c a t t t a g a t g t t t g t g c t g t t 60
g t t c c t g c t g c t g g t t t t g g t c g t c g t a t g c a a a c a g a a t g t c c a a a a c a a t a c t t a t c a 120
a t t g g t a a t c a a a c a a t t t t a g a a c a t t c a g t t c a t g c t t t a t t a g c t c a t c c a c g t g t t 180
a a a c g t g t t g t t a t t g c t a t t t c a c c a g g t g a t t c a c g t t t t g c t c a a t t a c c a t t a g c t 240
a a t c a t c c a c a a a t t a c a g t t g t t g a t g g t g g t g a t g a a c g t g c t g a t t c a g t t t t a g c t 300
g g t t t a a a a g c t g c t g g t g a t g c t c a a t g g g t t t t a g t t c a t g a t g c t g c t c g t c c a t g t 360
t t a c a t c a a g a t g a t t t a g c t c g t t t a t t a g c t t t a t c a g a a a c a t c a c g t a c a g g t g g t 420
a t t t t a g c t g c t c c a g t t c g t g a t a c a a t g a a a c g t g c t g a a c c a g g t a a a a a t g c t a t t 480
g c t c a t a c a g t t g a t c g t a a t g g t t t a t g g c a c g c t t t a a c a c c a c a a t t t t t t c c a c g t 540
g a a t t a t t a c a t g a t t g t t t a a c a c g t g c t t t a a a t g a a g g t g c t a c a a t t a c a g a t g a a 600

0802W01_ST25

gcat cagctt t agaat act g tggttttcat ccacaatt ag tt gaagggt cg t gct gat aac 660
 att aaagtt a cacgt ccaga agatttagct tt agct gaat tttat ttaac acgt acaatt 720
 cat caagaaa at acaaccgg ttaat ct ag 749

<210> 123
 <211> 1151
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> codon optimized sequence comprising IS62

<400> 123
 at at ggatt a t aaagat gat gacgacaaag gt at gcacaa gtt cacagggt gtt aacgct a 60
 aat t ccagca accagcatt a agaaat t t at ct ccagt ggt agt t gagcgc gaacgt gagg 120
 aat t t gt agg att ct t t cca caaat t gt t c gt gact t aac t gaagat ggt att ggt cat c 180
 cagaagt agg t gacgct gt a gct cgt ct t a aagaagt att acaat acaac gcacct ggt g 240
 gt aaat gcaa t agagggt t t a acagt t gt t g cagct t accg t gaact t t ct ggaccagggt c 300
 aaaaagacgc t gaaagt ct t cgt t gt gct t t agcagt agg at ggt gt att gaatt at t cc 360
 aagcct t t t t ct t agt t gct gacgat at aa t ggaccagt c att aact aga cgt ggt caat 420
 t at gt t ggt a caagaaagaa ggt gt t ggt t t agat gcaat aaat gat t ct t t t ct t t ag 480
 aaagct ct gt gt at cgcgt t ct t aaaaagt att gccgt ca acgt ccat at t at gt acat t 540
 t att agagct t t t t ct t caa acagct t acc aaacagaatt aggacaaat g t t agat t t aa 600
 t cact gct cc t gt at ct aag gt agat t t aa gccat t t ct c agaagaacgt t acaaagct a 660
 t t gt t aagt a t aaaact gct t t ct at t cat t ct at t t acc agt t gcagca gct at gt at a 720
 t ggt t ggt at agat t ct aaa gaagaacat g aaaacgcaaa agct at t t t a ct t gagat gg 780
 gt gaat act t ccaaatt caa gat gat t at t t agat t gt t t t ggcgat cct gct t t aacag 840
 gt aaagt agg t act gat att caagat aaca aat gt t cat g gt t agt t gt g caat gct t ac 900
 aaagagt aac accagaacaa cgt caact t t t agaagat aa t t acgggt cgt aaagaaccag 960
 aaaaagt t gc t aaagt t aaa gaatt at at g aggct gt agg t at gagagcc gcct t t caac 1020
 aat acgaaga aagt agt t ac cgt cgt ct t c aagagt t aat t gagaaacat t ct aat cgt t 1080
 t accaaaaga aat t t t ct t a ggt t t agct c agaaaat at a caaacgt caa aaat cagggt c 1140
 caagat ct t a a 1151

<210> 124
 <211> 1257
 <212> DNA
 <213> Artificial Sequence

0802W01_ST25

<220>

<223> codon optimized sequence comprising IS61

<400> 124

```

at ggt accaa cat caat t ct t aat act gt a t caact att c acagt t ct cg t gt aact t ct      60
gt t gat cgt g t t ggt gt t t t aagt t t aaga aat t ct gat t cagt t gaat t t acacgt cgt    120
cgt agt g gat t t agt act ct t at t t acgaa t cacct ggt a gacgt t t t gt agt acgt gca      180
gct gaaact g at acagat aa agt aaaat ct caaact cct g at aaagct cc t gcaggt ggt          240
t ct t caat t a accaact t t t aggaat aaaa ggt gct agt c aagaacaaaa caaat ggaaa        300
at acgt t t ac aact t act aa accagt aaca t ggccacct t t agt at gggg t gt agt t t gt      360
ggg gct gct g cat caggcaa ct t ccat t gg acaccagaag at gt agct aa aagt at t t t a      420
t gt at gat ga t gt ct ggt cc t t gt t t aaca ggt t at acac aaact at t aa t gat t ggt at    480
gacagagaca t t gat gcaat caat gaacct t accgt cct a t accaagt gg t gct at t t ca      540
gaaccagaag t t at t acaca agt t t ggg t t ct t t t act t g gt ggt ct t gg aat t gct ggt    600
at ct t agat g t at gggct gg t cat acaaca cct acagt at t ct at ct t gc t t t aggaggt    660
agt t t act t t ct t acat t t a ct cagct cca cct t t aaaac t t aaacaaaa t ggt t ggg t t    720
ggg aat t t cg ct t t aggt gc t t cat at at a t cat t acct t ggt gggcagg t caagct ct t    780
t t t ggcacat t aacaccaga cgt t gt t gt a ct t act ct t c t t t acagt at agct ggct t a    840
ggg at cgcaa t t gt aaat ga ct t t aaat ca gt t gaagg t g at agagcact t ggt t t acag    900
agt t t accag t t gct t t t gg aacagaaact gct aaat gga t at gt gt t gg cgct at t gac    960
at aacacaat t aagt gt agc aggt t act t a ct t gct agt g gt aaaccat a t t acgct t t a    1020
gct t t agt ag ct t t aat cat t cct caaat t gt t t t cagt t caaat act t t ct t aaagat    1080
cct gt aaaat acgacgt aaa at at caagca t ct gct caac cat t t t t agt t t t aggt at t    1140
t t cgt aact g ct ct t gcaag t caacacggt accggt gat t at aaagacga t gat gacaaa    1200
t caggt gaaa act t at act t t caaggt cac aat cat cgt c acaaacacac cggt t aa          1257

```

<210> 125

<211> 5240

<212> DNA

<213> Scenedesmus dimorphus

<400> 125

```

cgt t t aggt g t aacacaat c t t ggggt gga t ggacaat t a gcggt gaaac agcaacaaat      60
ccaggt at t t ggagt t at ga aggt gt t gct gcat ct cat a t t at t t t at c t ggt t t at t a    120
t t ct t agct t cggt t t ggca ct ggg t t t ac t gggat t t ag agt t at t ccg t gaccaaga    180
act ggaaaaa ct gcatt aga t t t accaaaa at t t t cgga t t cact t at t ct t at caggt      240
ct t t t at gt t t t ggt t t t gg t gct t t ccac gt aacaggt t t at t t ggt cc t ggt at t t gg    300

```

0802W01_ST25

gt t t c a g a t c c t t a t g g a t t a a c a g g a a g t g t t c a a c c a g t t t g c t c c t t c t t g g g g t g c t 360
 g a t g g g t t t g a t c c t t t c a a c c c t g g t g g t a t t g c a g c g c a c c a c a t t g c t g c t g g t a t t 420
 t t a g g t g t t t t a g c a g g a t t a t t c c a c t t a t g t g t a c g t c c t t c t a t t c g t t t a t a c t t t 480
 g g t t t a t c a a t g g g t a g t a t c g a a a c a g t a t t a t c a a g t a g t a t t g c t g c t g t t t t c t g g 540
 g c t g c t t t c g t t g t t g c t g g a a c t a t g t g g t a t g g t t c a g c a g c t a c t c c a a t t g a a t t a 600
 t t t g g t c c t a c a c g t t a t c a a t g g g a c c a a g g t t t c t t c c a a c a g a a a t t c a a a a c g a 660
 g t t c a a a c a a g t t t a g c a g g t g g t t c t t c a c t t t c t g a t g c t t g g t c g a a a a t t c c a g a a 720
 a a a t t a g c t t t c t a t g a t t a t a t t g g a a a c a a c c c t g c a a a a g g t g g t c t t t t c c g t a c a 780
 g g a g c t a t g a a t a g t g g a g a t g g t a t t g c t g t t g g a t g g t t a g g t c a c g c a g t a t t t a a a 840
 g a t c a a g a t g g t c g t g a a t t a t a c g t a c g t c g t a t g c c t a c t t t c t t t g a a a c a t t c c c a 900
 g t t t t a t t a a t t g a t a a a g a t g g t g t t g t a c g t g c t g a c g t t c c t t t c c g t c g t g c t g a a 960
 t c a a a a t a t a g t a t t g a a c a a g t t g g t g t a t c a g t a a c t t t c t a c g g t g g t g a a t t a g a t 1020
 g g a t t a a c a t t t a a t g a t c c a g c a a c t g t t a a a a a t a t g c t c g t a a a g c a c a a t t a g g t 1080
 g a a a t t t t t g a a t t t g a t c g t t c a a c a t t a c a a t c t g a t g g t g t a t t c c g t a g t a g t c c a 1140
 c g t g g t t g g t t t a c t t t t g g t c a c g t t t g c t t t g c t t t a t t a t t c t t c t t t g g a c a t a t t 1200
 t g g c a t g g t g c a c g t a c a a t c t t c c g t g a t g t a t t t g c t g g t a t t g a t g a t g a t c t a a a c 1260
 g a a a g t t t a g a a t t t g g t a a a t a c a a a a a a c t t g g t g a t a c a a g t t c t g t t c g t g a a g c t 1320
 t t c t a a t t c g t t t t t t t c t c t t t t t t t c t t t t t c t c t t t g g a a a a g a a a a a c a t g t 1380
 t t a t t t t g a a t t t t t t g t t t a g a a c t t t a c t g t t c t t t t t t a t t t t t a a a g t g t t t t t g t 1440
 t t t t t t t t a a t a c a a a a a c t t t t t t a a a a t g a a t t t a a a a a c a c a a a a a a a g a g t t a t 1500
 t g c t a t t c a a a a t a a c a a g a g t t t a a a a c a a a g t t t t t t t c t a a a g a a a a a a c t t c t 1560
 t c a t t t t t t t t g a a t t g t t t t t a a a c t t t t t t c t t c t t t t t a g a g t t t t t t t c t t 1620
 c a c t t t t t t g c a a a a a a g t g a g a a a a a c a g c a a a g a a g t g a a a a a a g t t c a a a a 1680
 a c a a t t c a a a a a a g a c a a a a c c t a a a a a a t a t c a c t t g a g a t g g g t c t g g a t t t t t t c c 1740
 a a g c a a a a g a a t t t t g t a t t t t g t t g a a a g t t t t t c a t a a a a t a c a a a t t t g c a a t t a t 1800
 t a t t c t t a a a a t c a a a a t a t t t g t t a a c c a c a t t t c a t t c t a t g g a a g c a t t a g t t t a t a 1860
 c t t t t t t a t t a a t c g g a a c a t t a g g a a t t a t c t t t t t c g c a a t t t t c t t t a g a g a a c c a c 1920
 c t c g t a t g g t a a a a t a a t t g a a a t t t t g c t t t t t t t t a t g g a t a a a g a a a t g a t t t c a 1980
 a t t c a t t t c t t t t a t c c a t t t t t t g a a a a t a g t t t t t c t c a a t t t t t t a t t t t t t t g t 2040
 t t t t t c t c t a a a a t c a a a a a t t c a a t t t t g a g a a a a a a t t t t a t a c a a a a a g t t t t t t g 2100
 a t t a a a a a c t c a g a a a a a a g a a a a a a c a a a g t a t g a a t t t t t t g a a a a a t t c a t a c c t t 2160
 t t a t t t t t t t t g t a a t t t t t a g c c t t t c a a a a a t t t t t g a a g g c a t t t t t t t t a a t c 2220

0802W01_ST25

ct cat gtt ct t caaaaggat ct ct caattt tttt gaagga ggt ccaaaac t cacat agat	2280
t gaat at cct gtt gcactt a ataat agaaa ccat aaaaaa aaggt aaaga aaaaagcagg	2340
act gt ccat a attcttt cat gttttt gtt caaat ttt att ct ccaat aat t at att acga	2400
cagaaagt aa aaaaaat caa aat ttt att ca aaaaaat ggc t act ggaaca act t caaaag	2460
ct aaat caag ctt at ct gat gcact t caag aaccagggt at cgt aact cct tt aggaact t	2520
t att aagacc gtt aaact ct gaat caggaa aagt att acc t ggat gggga acaact gttt	2580
t aat ggggt gt ttt cattt gta cttttt gct g t att ct t att aatt at ttt a gaaat ttt at a	2640
acagt t cttt att att agat gat gtt act a t gagt t ggga aact t t agct t ctt aat t ca	2700
at agaat agt ttt attt gctt tttttt attt ttt att ttt at c aaaaaattttt ttt gcaaaaa	2760
t aaagaat aa at aaaat t ca aaaaaat t at agaatt agat aaaaat t agtt t caagt t gaa	2820
ct aagt t gtc aat aaacttt caaat t t gtt tttttt ac t gtt catt aa gagcaat aaa	2880
aaaaact tttt ggt ctt ggca at cttttt aaa aaagt cagaa t caat t ct at ttt aagaat c	2940
ct at ggaat c t at gt attt a atttt agcaa aatt accaga agct t at gca cctttt gat c	3000
ct attt gtaga t gttttt acca att attt cct a tttt ctt ctt att attt agcc ttt gt at ggc	3060
aagcat ct gt aagt tttt aga t aaaaaat t g aaaagt tttt ttt gat actt tttt aaaaaa	3120
t at caaaaaa aact tttt caa tttttt caa tttt catt cg caact t t agc ttt aat att a	3180
gct aaagt t g ct ct caaaaa t at aat tttt tttt gacttt tttttttttt atttt gtttc	3240
ttttttt aaaa gtt acaacat aaagaaaat g aaaat agaaa attt gt gaaa cat aaaaaaa	3300
aagaat gaaa tttttt at gtt cgtttttt gt ttt at ctttt ccaact aaag t aaat ttttt	3360
tt ccaaaat t ct at aaaaaa t caaaaaat t aaaaaaaaag aaaaaacttt gtttt gt gca	3420
aaacaaaaat ctt gaatt ca aaacaaaat a aagaatt caa aaagattttt ttt aagcaaa	3480
aggt aaaat g gaaaaaaat g tttttt gctt taaaattttt tctttttttt aaagctttt gc	3540
ttttttt cat g aaaaaaaciaa agctttt aaaa aaaagaaaaa atttt aaagc aaaaaaaga	3600
att aaacacg tctttttttt ggaggacgac at ccat t at g t ggaatt cct gttttttt ctc	3660
gaat aacatt t actttt aatt ccagctttt aa aaattt cacg aat agct gtt t cgcgccctt	3720
gt cct ggacc agtt act aaa attttt gctt cattt aat gc aaattt cacgt gat tttttt g	3780
ccacaactt c agcagctttt ttt gct gcaa at gt t gt t gc tttt cttttt ccacggaaac	3840
cacaagct cc agcagaact c caacaaagga ctt caccacg aagattt gct aat gt aat aa	3900
t agt att at g at gt ccagct t gaat at aaa caat t cct cg at at gt acgt ttttt aattt	3960
ttgt aggt ga t actttt cta gttt gt ct ag ccat at gt aa aaattt aact at aaattt ct	4020
ttt at atttt t aaat cattt gat t t act at ct aaaaaaaa taaaattttt g aat cttt gga	4080

0802W01_ST25

aagaaccattt atttggaaag atttatttgt tcaattcttt tgtatttttt ttcaaaaaaa 4140
ttttttagaa ttttttatct aattttttaa tgtttgttc aaacataaaa aattttcttt 4200
tcaaaaaaga aaaaaaaatt tctttttttt gaaaaaggaa aaataaaaaa agctgt agct 4260
tttttgttga atcaaaaaaa aat aagaaat tgtcctattt ttatgacaaa agattcaaaa 4320
aaatgaaat aaaaaagaaaa attcaaactt tcaacaatg tattttgtat ttttgggccg 4380
agt cggattc gaaccaacgt aggcgt aacc agcggattta caatccgccc ccat taacca 4440
ctcgggcattc ggcccatgct tttttttgac tcatttaatc atttattttg catgaattat 4500
actaaagatt atattaacaa aaaatttttg aaatttcaat tttttttaat taaaaactct 4560
tctatttttt aaaattcttt gtittgaatt tttttttcaa tttaaaaaaa aaaaacat at 4620
taaaaaat at ttttaaaaaa ttttttgtat tgaaaactta caataatttt ataaattttt 4680
ttgaaaaatt ttgttttttt tatctattg aaatgaacaa aacaaatttt tttagttttt 4740
tttgtttttt tgttttgctg ctcttttgt tttatcaat aaataaatga aaatgaaaaa 4800
caaaaataaa caatttatgt ttttagagtt ctaaaatcaa gaaaaaaat a cttccccttt 4860
aaagaggaag tattttaaaa aaaaattata gtttgtcaat agtttcaaat tcaaat ttga 4920
tttctttcca aacttcacaa gcagcagcta attctggaga ccatttacia gctgagcgga 4980
taacatcacc accttcacga gctaaatcac gaccttcgtt acgagcttga gtacaagctt 5040
ctaaagcaac acggttagca acagcaccag gagcgttacc ccaagggtgt cctaaagtac 5100
ctccaccgaa ttgt aaacaa gcgtcatcac caaagatttc aactaaagct ggcatgtgcc 5160
atcgtgaat accaccagaa gcaactggca tagtaccacc catagaacac cagtcttgag 5220
tgaagt aaat accacggcta 5240

<210> 126

<211> 5240

<212> DNA

<213> Scenedesmus dimorphus

<400> 126

cgtttagggtg taacacaatc ttgggggtgga tggacaattta gcggtgaaac agcaacaaat 60
ccagggtattt ggagttatga aggtgttgct gcatctcata ttattttatc tggttttat 120
ttcttagctt cggtttggca ctgggtttac tgggatttag agttattccg tgaccaaga 180
actggaaaaa ctgcattaga tttaaaaaa attttcggaa ttactttatt ctatcagggt 240
cttttatgtt ttggttttgg tgcittccac gt aacagggtt tatttgggtcc tgggtatttgg 300
gtttcagatc ctatggatt aacaggaagt gttaaccag ttgctccttc ttgggggtgct 360
gatgggtttg atcctttcaa ccctgggtgt attgcagcgc accacattgc tgcgtggtatt 420
ttagggtgtt tagcaggatt attccactta tgtgtacgtc ctctattcgt ttatacttt 480

WD1

Page 93

0802W01_ST25

cagaaagt aa aaaaaat caa aat t t a t t c a aaaaaat ggc t a c t g g a a c a a c t t c a a a a g 2460
 c t a a a t c a a g c t t a t c t g a t g c a c t t c a a g a a c c a g g t a t c g t a a c t c c t t t a g g a a c t t 2520
 t a t t a a g a c c g t t a a a c t c t g a a t c a g g a a a a g t a t t a c c t g g a t g g g g a a c a a c t g t t t 2580
 t a a t g g g t g t t t t c a t t g t a c t t t t t g c t g t a t t c t t a t t a a t t a t t t t a g a a t t t a t a 2640
 a c a g t t c t t t a t t a t t a g a t g a t g t t a c t a t g a g t t g g g a a a c t t t a g c t t c t t a a t t c a 2700
 a t a g a a t a g t t t t a t t g c t t t t t t a t t t t t a t t a t t t t a t c a a a a t t t t t t t t g c a a a a a 2760
 t a a a g a a t a a a t a a a a t t c a a a a a a t t a t a g a a t t a g a t a a a a t t a g t t t c a a g t t g a a 2820
 c t a a g t t g t c a a t a a a c t t t c a a a t t t g t t t t c t t t t t a c t g t t c a t t a a g a g c a a t a a a 2880
 a a a a c t t t t g g t c t t g g c a a t c t t t t a a a a a g t c a g a a t c a a t t c t a t t t t a a g a a t c 2940
 c t a t g g a a t c t a t g t a t t t a a t t t a g c a a a a t a c c a g a a g c t t a t g c a c c t t t t g a t c 3000
 c t a t t g t a g a t g t t t t a c c a a t t a t t c c t a t t t t c t t c t t a t t a t t a g c c t t t g t a t g g c 3060
 a a g c a t c t g t a a g t t t t a g a t a a a a a a t t g a a a g t t t t t t t t g a t a c t t t t g t a a a a a 3120
 t a t c a a a a a a a c t t t t c a a t t t t t t c a a t t t t c a t t c g c a a c t t t a g c t t t a a t a t t a 3180
 g c t a a a g t t g c t c t c a a a a a t a a t a a t t t t t t t t g a c t t t t a t t t t t t t a t t t t g t t t c 3240
 t t t t t t a a a a g t t a c a a c a t a a a g a a a a t g a a a a t a g a a a a t t t g t g a a a c a t a a a a a a 3300
 a a g a a t g a a a t t t t t a t g t t c g t t t t t t g t t t t a t c t t t t c c a a c t a a a g t a a a t t t t t t 3360
 t t c c a a a a t t c t a t a a a a a t c a a a a a a t t a a a a a a a a g a a a a a c t t t g t t t t g t g c a 3420
 a a a c a a a a a t c t t g a a t t c a a a c a a a a t a a a g a a t t c a a a a a g a t t t t t t t t a a g c a a a 3480
 a g g t a a a a t g g a a a a a a t g t t t t t t g c t t t a a a a t t t t t t c t t t t t t t t t a a a g c t t t g c 3540
 t t t t t t c a t g a a a a a a c a a a g c t t t a a a a a a a g a a a a a a a a t t t t a a a g c a a a a a a a g a 3600
 a t t a a a c a c g t c t t t t t t t t t t g g a g g a c g a c a t c c a t t a t g t g g a a t t c c t g t t t t t t c t c 3660
 g a a t a a c a t t t a c t t t t a a t t c c a g c t t t a a a a a t t t c a c g a a t a g c t g t t t c g c g c c c t t 3720
 g t c c t g g a c c a g t t a c t a a a a t t t t t g c t t c a t t t a a t g c a a a t t c a c g t g a t t t t t t t g 3780
 c c a c a a c t t c a g c a g c t t t t t t t g c t g c a a a t g t t g t t g c t t t t c t t t t t c c a c g g a a a c 3840
 c a c a a g c t c c a g c a g a a c t c c a c a a a g g a c t t c a c c a c g a a g a t t t g c t a a t g t a a t a a 3900
 t a g t a t t a t g a t g t c c a g c t t g a a t a t a a a c a a t t c c t c g a t a t g t a c g t t t t t a a t t t 3960
 t t g t a g g t g a t a c t t t t c t a g t t t g t c t a g c c a t a t g t a a a a a t t t a a c t a t a a a t t t c t 4020
 t t t a t a t t t t t a a a t c a t t t g a t t t a c t a t c t a a a a a a a a t a a a a t t t t g a a t c t t t g g a 4080
 a a g a a c c a t t a t t t g g a a a g a t t t a t t t g t t c a a t t c t t t t g t a t t t t t t t t c a a a a a a a 4140
 t t t t t t a g a a t t t t t t a t c t a a t t t t t t a a t g t t t t g t t c a a a c a t a a a a a a t t t t t c t t t 4200
 t c a a a a a a g a a a a a a a a t t t c t t t t t t t t t g a a a a a g g a a a a t a a a a a a a g c t g t a g c t 4260

0802W01_ST25

t t t t t g t t g a a t c a a a a a a a a t a a g a a a t t g t c c t a t t t t t a t g a c a a a a g a t t c a a a a 4320
 a a a t g a a a t a a a a a g a a a a a t t c a a a c t t t c a a a c a a t g t a t t t t g t a t t t t g g g c c g 4380
 a g t c g g a t t c g a a c c a a c g t a g g c g t a a c c a g c g g a t t t a c a a t c c g c c c c c a t t a a c c a 4440
 c t c g g g c a t c g g c c c a t g c t t t t t t t g a c t c a t t t a a t c a t t t a t t t g c a t g a a t t a t 4500
 a c t a a g a t t a t a t t a a c a a a a a t t t t t g a a t t t c a a t t t t t t a a t t a a a a c t c t 4560
 t c t a t t t t t t a a a a t t c t t t g t t t g a a t t t t t t t c a a t t t a a a a a a a a a a a c a t a t 4620
 t a a a a a t a t t t t t a a a a a t t t t t t g t a t t g a a a a c t t a c a a t a a t t t t a t a a a t t t t t 4680
 t t g a a a a a t t t t g t t t t t t t t a t t c t a t t g a a t g a a c a a a a c a a a t t t t t t a g t t t t t 4740
 t t t g t t t t t t t g t t t t g c t g c t t c t t t t g t t t t a t c a a t a a a t a a a t g a a a a a a a a t g a a a a a 4800
 c a a a a a t a a a c a a t t a t g t t t t a g a g t t c t a a a a t c a a g a a a a a a a t a c t t c c c c t t t 4860
 a a a g a g g a a g t a t t t t a a a a a a a a t t a t a g t t t g t c a a t a g t t t c a a a t t c a a a t t t g a 4920
 t t t c t t t c c a a a c t t c a c a a g c a g c a g c t a a t t c t g g a g a c c a t t t a c a a g c t g a g c g g a 4980
 t a a c a t c a c c a c c t t c a c g a g c t a a a t c a c g a c c t t c g t t a c g a g c t t g a g t a c a a g c t t 5040
 c t a a a g c a a c a c g g t t a g c a a c a g c a c c a g g a g c g t t a c c c c a a g g g t g t c c t a a a g t a c 5100
 c t c c a c c g a a t t g t a a a c a a g c g t c a t c a c c a a a g a t t t c a a c t a a a g c t g g c a t g t g c c 5160
 a t a c g t g a a t a c c a c c a g a a g c a a c t g g c a t a g t a c c a c c a t a g a a c a c c a g t c t t g a g 5220
 t g a a g t a a a t a c c a c g g c t a 5240

<210> 127

<211> 876

<212> DNA

<213> Scenedesmus dimorphus

<400> 127

t t g a a c t a t c a a g t t t a g g t t t t a a a a t c t t a t t t a t t t a c t t t a t t t t t a a t t t g a a 60
 a a c t c t g c g a g c t t t g c g a g c a c t g a t t t c a a a a t c t t a g t t t c a a g t a a a a c t t a t t t t 120
 c a a t c t t t a t t t a t t t g t a t t t t c a a a c t a a a a g t t t g a a t t a t c t a a t t t g a a a c t t t a 180
 t g a g c t t t a c a a a g a c t g g t t t c a a a t t t t t c t t t g t t t a g t t t g t t t t t c a a a c t a t 240
 t a g t t t a a a c t a t c t a a t t t g a a t t a t a a g t t t g t a t t t t c a a a c t c t t t a t t t g t t a a c 300
 t t t g t t t t t t a g t t t g a a a t t c t a t g a t t t t g g a c a t t a g a a g g c t t t g c c t t a a t g a t t 360
 t t c c c c g a g c c c c t c t t g g g a t t c t t t t t t t t t a t t t t c c t t t c a g g a c t t a t t a t a a t a 420
 t a t c a a a a a t a a t t t t t t t g t c a a t t t t t t t a a t a t t t t a a a a a t t t t t a a a t t a a a a 480
 a t a a a t t t t t t t t a t t t t t t t a g a t t t t a a t t t t a t t t g t a a g t t t a a t a t t t c t a a a a a 540
 a t t t g a a t t t a a g a a t t t t t t a a t t t g a t g a a a a a a a t t g t t t t t g a a t t t t t t t t t t t 600
 a c t t t a a t t a a c t t t t t t a a a a a t g a t t a a a a t t t g a a g t t t t t a a a a a c t a t c t t t 660

0802W01_ST25

ttttttgt aa aaat ggt att attttgtgt a aaat acaaca aaaat aacaa tttt cacct t	720
at ttt aagtt t aat tttt ca at gcaaaat t t at ttt caac aaaat gaaaa at at tttttt	780
agaat at att t t at ggcggg cgt agccaag t ggt aaggca at ggat t gt g act ccat cat	840
t cgcgggt t c gaaccccg t c gtt cgcccat aat aaa	876

<210> 128
 <211> 1724
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> codon optimized sequence comprising rb1L-CAT-psbE

<400> 128 cggttttttt t acttttgct ttttttgctt ttgttcaaag aaaaaaaaaat acaaaat aaa	60
aaaaactaaa at gaaaaaac aaagaattct aaaattcata aaaaaaat t a aaaccaatt	120
ttttttttgg aaacttttcc aaat aat aaa aaaat caaaa aaaaaat tttt ct agt at ttt	180
tttcat at ttt t gaaactttt tttgagttta t aaaaaaat a gaaaaaacaa at agat gaaa	240
at ttagaaaa attat aaacc aat aaaaat g aagttttgcg tagaaaaaaa at ttagttta	300
cttgttcccc aagagcaagt ggt aactttg aaaaaaat at ttaactt aa aaat ttgct a	360
aagttttgaa tttatgttaa aat ttaaaaa aaat aaaaat ttttaacta ttttttatg	420
ttaaaaaaat agtttttatt attttctata at at agttta gttttttatt tttttcaatt	480
tctttttttt tttcaaagaa aaaagttttc cacggat aga tttttatagg atcgacaaaa	540
tgttctatga acttttcata atggagaaaa aaat cactgg at at accacc gttgat at at	600
cccaatggca t cgt aaagaa cat tttgagg cat ttcagtc agttgctcaa tgtacct at a	660
accagaccgt t cagctggat attacggcct ttttaagac cgt aaagaaa aat aagcaca	720
agttttatcc ggcctttatt cacattcttg cccgcctgat gaatgctcat ccggagt tcc	780
gtatggcaat gaaagacggt gagctggtga tatgggat ag tgttcaccct tgttacaccg	840
ttttccatga gcaaactgaa acgttttcat cgctctggag tgaat accac gacgat tccg	900
gcagtttcta cacat at att cgcaagatgt ggcgtgttac ggt gaaaacc tggcct at ttt	960
ccctaaaggg tttattgaga at atgttttt cgt cagcgcc aatccctggg tgagt tttcac	1020
cagttttgat ttaaacgtgg ccaat at gga caacttcttc gcccccg ttt t cact at ggg	1080
caaat at tat acgcaaggcg acaagggtgt gatgccgtg gcgat t cagg t tcat cat gc	1140
cgtttgtgat ggcttccatg t cggcagaat gctt aat gaa t tacaacagt actgcgat ga	1200
gtggcagggc ggggcgt aac ct agat at ttt gagaat tttgt at t taaaact gaaaaat ttt	1260
tgaacgaact cttttcaaaa at at taaact t tcttgagat gat ttagt gt tatct caaga	1320
aagtttgttc ttttat tttt taaaat tttt aaaaat tttt ttttct tttt aacaggaaaa	1380

0802W01_ST25

t aaat aaaga aaaaagt gaa ttaaaaaaaa gctgggactt tcaaagt gac caat t t t t t a	1440
ct t t aaagt t t t t t t t att caat aaaaaa at act aaaaaa aat at gaaag t att act t aa	1500
aat t t c t t aa aaaaaaaaga atgcct t t t t t t caaaaaaaa gt t t aaaaaa aat aaagt t t	1560
t t acgt att g t t t aaaaact t t t t t gaaaa aagcat t c t t t t t c a t t t a aagagt t at c	1620
t t t t t t at c t cgt gcaagt t t t ggaaat t c at t t t t g t t a aat aact t t g act t t t t att	1680
t c t t aaat t t t t ggct t t t c at t t t t t t t g gt t t acaaat aaaa	1724

<210> 129
 <211> 25
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> PCR primer

<400> 129	
at ggghyt mc cwt ggt aycg t gt hc	25

<210> 130
 <211> 25
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> PCR primer

<400> 130	
ccr at gt ggc gr caaggdat gt t yg	25

<210> 131
 <211> 25
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> PCR primer

<400> 131	
t t t gwarr at r at havdar r aawrg	25

<210> 132
 <211> 24
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> PCR primer

<400> 132	
kccwggdr mh act t t wccdk m t c	24

<210> 133

0802W01_ST25

<211> 2805

<212> DNA

<213> Dunal i e l l a a e r t i o l e c t a

<400> 133

t c t g t g c a t t t a a t g c a c a c a g c t c t a g t a g c t g g t t g g g c t g g t g c t a t g a c a t t a t t t	60
g a a a t t g c a g t t t t t g a t c c a t c a g a t c c a g t a t t a a a c c c t a t g t g g c g t c a a g g g a t g	120
t t c g t t c t t c c t t t c c t t a c a c g t t t a g g t g t a a c a c a a t c a t g g g g t g g t t g g a c a a t t	180
a g t g g t g a a a c a t c t t c a a a c c c a g g t a t c t g g a g t t a t g a a g g t g c t g c a g c t t c g c a c	240
a t t g t t c t t t c a g g t t t a t t a t t c c t t g c t t c a g t t t g g c a c t g g g t t t a c t g g g a t t t a	300
g a a t t a t t c c g t g a t c c a c g t a c a g g t a a a a c t g c a t t a g a t t t a c c a a a a a t t t t t g g t	360
a t t c a t c t t t t c t t a g c a g g a c t t c t t t g t t t t g g t t t t g g t g c t t t c c a c g t a a c a g g t	420
g t t t t t g g a c c t g g t a t t t g g g t a t c a g a t c c a t a t g g a t t a a c a g g t a g t g t a c a a c c a	480
g t a g c t c c t t c t t g g g g t g c t g a a g g t t t t g a c c c t t a c a a c c c a g g t g g t g t a c c a g c t	540
c a c c a t a t t g c t g c t g g t a t t t t a g g t g t a t t a g c a g g t t t a t t c c a c c t t t g t g t t c g t	600
c c a t c a a t t c g t t t a t a t t t t g g t t t a t c a a t g g g t t c t a t t g a a t c a g t t t t a t c a a g t	660
a g t a t t g c a g c c g t a t t c t g g g c a g c t t t c g t a g t a g c a g g t a c t a t g t g g t a t g g t t c t	720
g c a g c a a c t c c a a t t g a a t t a t t t g g t c c a a c a c g t t a c c a a t g g g a t c a a g g t t t c t t c	780
c a a c a a g a a a t t c a a a a a c g a g t a g c a c a a a g t a c a t c t g a a a g g t t t a t c t g t t t c g t a	840
g c a c a a a g t a c a t c t g a a g g t t t a t c t g t t t c a g a a g c t t g g g c a a a a a t t c c t g a a a a a	900
t t a g c t t t c t a t g a t t a c a t t g g t a a t a a c c c a g c t a a a g g t g g a t t a t t c c g t a c a g g t	960
g c t a t g a a c a g t g g t g a t g g t a t c g c t g t a g g t t g g t t a g g a c a c g c t a g t t t t a a a g a t	1020
c a a g a a g g t c g t g a a c t t t t t g t t c g t c g t a t g c c t a c t t t c t t t g a a a c t t t c c c t g t t	1080
g t t t t a a t t g a t a a a g a c g g t g t t g t t c g t g c t g a c g t a c c a t t c c g t a a a g c t g a a t c a	1140
a a a t a c t c a a t t g a a c a a g t t g g t g t t t c a g t t a c a t t c t a t g g t g g t g a a t t a a a t g g t	1200
t t a a c a t t t a c t g a c c c t t c a a c t g t t a a a a a a t a t g c a c g t a a a g c t c a a t t a g g t g a a	1260
a t c t t t g a a t t t g a c c g t t c g a c t t t a c a a t c t g a c g g t g t a t t c c g t a g t a g c c c a c g t	1320
g g t t g g t t c a c t t t t g g a c a c t t a t c t t t t g c c t t a t t a t t c t t c t t t g g t c a t a t t t g g	1380
c a t g g t t c a a g a a c t a t t t t c c g t g a c g t t t t c g c t g g t a t t g a t g a a g a c a t t a a t g a t	1440
c a a t t a g a a t t c g g t a a a t a t a a g a a a c t t g g t g a t a c t t c a t c t g t t c g t g a a g c t t t c	1500
t a a t c a c t a t a t t a a g t t t t a t t c c a a t a t t c t a c c a a g a a t a t t t a g a a a t t c t t g t t t	1560
t t a a a t t c a t a t a g a a a a a a t c t a t t a g a a t t t c c g t c g a g a a a t t c t a a t a g a t t t t t t	1620
c t c t t t t t g g c g g a c t a a a c a c t t t t c t t t g t t t t t t g t t t t c c g a c t a g a a a a t a g a a	1680
g a t t c a t t t a a c a a a g t t c a c a t a t t t t g t a a t g t g a t t t t t t t g t a a t g t g a t t t t t c t	1740

0802W01_ST25

tatctttatt agatttttcta tcttaccgta gattttaataa cggatataatt gttttttttt 1800
 ttttagatgg gcttagattt tctctgagca aaagaatttg agttagtgt aacaaatttg 1860
 atcaaagttt atttcataaa aagaattttt ttttataaat acggaagaaa atatcagagc 1920
 taaattttat gttcttccgt tttattttta taaaatgtta atattttatt tttgttaaaa 1980
 aaatctaaga taatattttt tttaaactcct tattggaatt aaattttatc ttttat aact 2040
 tt aggt aagt cgttagatat tttaaattta tcttctgacc gcttcgttta ttttcttttc 2100
 gtaattcgga agaaaattct ttttgtttta cactaatgta aaatttggtat ttatgttctt 2160
 atttgaaaag actagtcttt tcaataaat tattaaacac ctttcatgga agctttagt 2220
 tacactttct tattaattgg aacgttaggt attatctttt tctctatctt ttttagagag 2280
 cctccacgta ttgcaaaata gtatagaat aattttttat taaaattact tgtcggacca 2340
 aacaataaag ttgttttttc cgatcgaata atctacgaga aattctat at ttatcgtaaa 2400
 tacgtttcaa atgaagtatt atagggttg gatttcatia acattaatia gtgaagtcca 2460
 acccttagaa tacttaaaat ttttacttaa ct aagattat taatttaaatc tttcatgttct 2520
 tcaaaaggat ctctcaattt tcttgaagga ggtccaaaac taacataaac tgaataacct 2580
 gt agcactta ataaaagaaa ccataaaaag aaggtaaaga aaaaagccgg actttccata 2640
 aatttaaaat ttttgacaag acaaaattat tctccatata tattat acaa tattacagaa 2700
 aggaaaaaaaa caaaaagatt ttattaaacta tcaattatgg ctacaggaaa aaatagtaca 2760
 caaacttcaa catcacaaga accaggaatt gttaacccat tagaa 2805

<210> 134
 <211> 3561
 <212> DNA
 <213> Dunaliella

<400> 134
 catttaatgc acactgctct agtagcaggt tgggctggtt caatgacatt attcgaaatc 60
 gccgtatttg acccttcaga ccctgtatta aacctatgt ggcgtcaagg gatgttcgta 120
 ctctctttct taacacgtct aggtgt aaca caatcatggg gtggttggac aattactggt 180
 gaatcagcat caaacccagg tatctggagt tatgaagggtg ctgcagcttc tcatattgtt 240
 ctctcaggtt tattatttct tgcatacgtt tggcactggg tttactggga tttagaattta 300
 ttccgtgacc cacgtacagg taaaacagca tt agatttac caaaaatttt tggattcac 360
 ctattcttat caggacttct ttgttttggg tttggtgctt tccatactac aggtgttttt 420
 ggacctggtat ttgggtatc agatccatatt ggtttaacag gt agt gta aca accagt agct 480
 ccttcttggg gtgcagaagg ttttgatcct tacaaccag gtggtgtacc t gct caccat 540
 attgcagcag gtattttagg tgtattagct ggtttattcc acctatgtgt t cgt ccatct 600

0802W01_ST25

attcgtttat actttggttt atcaatgggt tctattgaat cagttttatc aagt agt att 660
 gcagcagt at tctgggcagc attcgttgtt tcaggtacta tgtggtacgg tt cagcaaca 720
 actccaattg aattatttgg tccaactcgt taccaatggg atcaaggttt ctccaacaa 780
 gaaattcaaa aacgagt agc acaaagtaca tctgacggtt tatctctttc tgaagcttgg 840
 tcaaaaattc ctgaaaaatt agctttctat gat t acattg gt aacaaccc t gct aaaggt 900
 ggtttattcc gtacaggtgc tatgaatagt ggtgatggt a ttgccgtagg ttggttaggc 960
 cacgctagt t tcaaagatca agaaggacgt gaactttttg tt cgt cgt at gcct act ttc 1020
 ttcgaaactt tccctgttgt tttaat t gat aaagatggt a ttgtt cgt gc t gacgt tcca 1080
 ttccgt aaag cggaatctaa atactctatt gaacaagt ag gtgtttcagt t acattctac 1140
 ggtggtgaat taaatggttt aacatttact gatccttcta cagttaaaaa atacgctcgt 1200
 aaagctcaat taggtgaaat cttcgaattt gaccgttcta ct t tacaatc t gat ggt gta 1260
 ttccgt agt a gtccacgtgg ttggtttact tt cggacact tatcttttgc tttattattc 1320
 ttctttggtc acatttggca tgggtcaaga actattttcc gtgatgtttt cgctggtatt 1380
 gacgaagat a tt aat gacca attagaattc ggt aaat aca agaaacttgg t gat act tca 1440
 tctgtt cgt g aagcttttcta atcaattttt t gat t t aatc atttcgccac agagaacttt 1500
 taaaaaat a attttataaa actctctgaa attattttctc cacactttga ccttgggtctc 1560
 tagttcgtct ggagaccaag gtcaaagtaa agttcttgt a t gat atttct agaagaaaaa 1620
 atttgatttt cttcgataaa tactcatata aaaataaaaa atcgaaggga agaaaacttc 1680
 tt cgaagttt tcttttttta agat accgcc agttt aaat ct aacaattt t attcct aag 1740
 gagaaaattc ctataggaat tttctctgcc gcgtttcgaa gaaacgcgt a ggacttctta 1800
 gaagtccgcc tattaatttt ttaaaagaaa atttttattg aatgtttttc tt agacaaaa 1860
 gaaaaaacat tttttcgtgt aggattttaa agaaattcaa gat t tctatc ct agt ct taa 1920
 ttt aagaaaa agaattttt ttctaaggat ttacaaaatt attattcaat tgttgattta 1980
 aaatct taaa attcgaaaat atttttgatt ttt aaagaaa gagattaaaa aaaaaaaaaa 2040
 cgatttttcta aatttaactt tagattctat ttttacttac ct aactaat a aagtttttta 2100
 gaaattttaa taaaaaattt tatgtggttt ttttgacgaa aaccctaat ct tgaacact 2160
 ttgtaaaat a tctttt aaag attttttaa ct tctat at g taaaatttt a ttttatttat 2220
 aaaagtttt tctcctttaa agttttattt atacttttaa agtattattt ttcgataaaa 2280
 aaaaacaaat atttcaat aa aatataat t aattttat aa aat aaat aaa agatgggctt 2340
 agaatttctc tgagcaaaaag aattt aaact aatgt gaaaa acgt aagtt a attttat aaa 2400
 aaaaatctaa tttt gt aaat tatatgcgtt tttt acatt a agttt aaatt t ggtttt at g 2460
 tt cct atttg aaaagcaaca gcttttcaat aaatttatt a aacatttcgc atggaagctt 2520

0802W01_ST25

tagtttacac tttcttatta gtggacat taggtattat tttcttctct attttcttta 2580
 gagaacctcc acgtattgct aaataataaaa aaaacatttc tcaacgtaaa ttacaaat aa 2640
 tttaat aatt actttgggat ttgtgt aaaa ctgcct aag aaatctgttg tatctttaat 2700
 cttcaggctt tgcacggttt aaattttcgt aggtcgcctc actgaggcga cctcgtgcct 2760
 ctcaaaaggt gctcaaagaa aaaaaaaaaa aaaaaagaaa gacaaaaaac ctttgggttt 2820
 ttgtacctta gaaaaaaact ttaattttt tttcctacta aaaattat tttaggcggtgt 2880
 cgcgagcaaa cataggaaaa aaattccgc tagggaagaa aaatttcctt aggaaatttt 2940
 ctattttttt aattctagtt tttttgtcg gaaaaaaacc ttagggtttt ttcct aaggg 3000
 tctttttgcc gaaggcgaaa agcccgctt attataaaca gtttattaat aagggttgat 3060
 ttaactataa attttttaa aaatgattat tttgttaaa tcctaatat tttatgtatg 3120
 aaattttcat tgttagt aaa tcaattatt attaacatga aaaatttgaa gttaaaaaa 3180
 tttcaagga ttttttaaat cttctatatt tctcgaagaa aatcaaagat tttctccgag 3240
 aaatcacagat atatcttgaa aatcaaagat tttcaagat a ctttatcgtg aataaagatt 3300
 attttttaaat cttcatgttc ttcaaaagga tctctcaatt tttctgaagg tggccaaaa 3360
 ctatataaaa ctgaataacc tgtagcactt aataaaagaa accataaaa gaaggtaag 3420
 aaaaaagccg gactttccat aaatttaaaa ttttgacaa gacaaaatta tttccat aa 3480
 attattat ac aatattacag aaaggaaaaa atataagat ttaatttatt ttttattatg 3540
 gctacaggt a aaaataat ac a 3561

<210> 135
 <211> 2730
 <212> DNA
 <213> N. abundans

<220>
 <221> misc feature
 <222> (1477)..(1488)
 <223> N is A, C, T or G

<400> 135
 gctgtacatt taatgcat ac ttcat tagtt tctggttggg ccggttcaat ggctttttat 60
 gagcttgctg tttttgatcc ttctgatcca gttttaaatc caatgtggcg tcaaggatg 120
 tttgttttac cttttatgac acgtttaggt atcactcaat cctgggggtgg ttggactatc 180
 agtggtgaaa cggcttcaaa tccgggtatc tggagttatg aggggtgt agc cgcagctcac 240
 atcgttttat cagggtttact ctttgctgcg tctatctggc actgggttta ttgggatctt 300
 gaactttttc gtgatccaag aacttcaaat ccagctttag atcttccaaa aatttttgggt 360
 atccatttat ttttatctgg tgttctttgt tttgggttttg gagctttcca cgtaacaggt 420

0802W01_ST25

at t t t t g g t c c t g g t a t t t g g g t t t c t g a t c c t t a t g g a a t t a c a g g a a c a g t t c a a g c a 480
 g t t g c g c c t t c t t g g g a t g c t a c a g g g t t t g a t c c c t a t a a t c c g g g t g g a a t t t c a g c a 540
 c a t c a t a t t g c t g c c g g c a t t t t a g g t g t a t t a g c t g g t t t a t t c c a c c t t t g t g t t c g t 600
 c c g c c a c a a c g a t t a t a c a a t g g t c t c c g t a t g g g g a a t a t t g a a c a g t a c t t t c t a g c 660
 a g t a t t g c a g c a g t t t t t t g g g c a g c t t t t g t t g t t t c t g g t a c t a t g t g g t a t g g t t c g 720
 g c g g c a a c a c c a a t t g a a c t t t t t g g t c c t a c t c g t t a t c a a t g g g a t t t a g g c t t c t t c 780
 c a a c a a g a a a t t g a a c g t c g t g t a c a a a c a a g t c t t t c t g a g g g c a a a t c t g c t t c g c a a 840
 g c g t g g g c a g a a a t t c c a g a a a a t t a g c t t t t t a c g a t t a c a t t g g a a a t a a t c c a g c a 900
 a a a g g t g g t c t t t t c c g t g c g g g t g c t a t g a a c a g t g g a g a t g g t a t t g c a g t g g g c t g g 960
 t t a g g t c a t g c t g t t t t c a a a g a t a a a c a a g g t a a c g a a c t t t t t g t a c g t c g t a t g c c a 1020
 a c t t t c t t t g a a c c t t c c c t g t c g t t c t t g t a g a t a a a g a t g g t g t t g t t c g t g c a g a t 1080
 g t t c c t t t c c g t c g t t c t g a a t c a a a g t a c a g t a t c g a a c a a g t t g g t g t t t c t g t a a c t 1140
 t t c t a t g g t g g a g a g t t a g a t a g t g t a a c t t t t a a t g a t c c a g c a a c t g t g a a a a a t a t 1200
 g c t c g a c g t g c t c a a t t a g g a g a a t t t t c g a a t t t g a t c g t g c a a c t c t t c a a t c a g a c 1260
 g g t g t t t t c c g t a c a a g t c c t c g t g g t t g g t t t a c a t t t g c t c a t t t a t g t t t t g c t c t t 1320
 c t t t t c t t c t t t g g t c a t a t t t g g c a t g g t g c t c g c a c a a t c t t c c g a g a t g t a t t t g c t 1380
 g g t a t c g a t g c a g a c t t a g a t g a a c a a g t a g a g t t t g g t g c a t t c t t a a a a c t t g g t g a t 1440
 a c t t c t a c t c g t c g t c a a t c g g t t t a a a g g t t c g g g n n n n n n n n n n n n a g t t a g t a c a a a 1500
 a c t a c t a t t t t t g g a a a a g g a a a t a c c t t t c a c a a a c t a g a c t c c a a g g t t t c t a t t c c a 1560
 g a a g c g g a g a a a g a a a t a g a a c c a g t g a g t t t t a a a a c t c a c t g g t t c t t a a a g g c t c g 1620
 t g a g a t a g a t t t t t t g a a a a t t t t a t c t a a t c t a c a a a a a t t t t t t g a a a c t a a c t a a g c 1680
 c a g t t t c a t t t t t t t a g a a a a a t g a a g a a a g a a g t a t c c a c t c a a a t t a t c c a a c a g a a 1740
 c a a a a a a a g a a g c t t a a t c c t c a t g t t c t t c a a a t g g a t c a c g a a g t t g t t g t g a t g g t g 1800
 g a c c a a a a c c t a c g t a a a t a g a a t a a c c g g t a a t a c t t a a t a a t a a a c a c c a t a a a a a a 1860
 t g g t a a a g a a a a a g c t g g g c t t t c c a t a c t g a t t t a a t t t t a g t t a c t t a a t a c t g a g 1920
 a c a a a a c g t t t t t a a a a c a g a g t t g t t c c a c a a a a a g c c c t t t t g g g c c t a a a t g g a a t 1980
 t g g t a c c a a t t t t c t t t t c g g a a g g a a a t a c a t t g a c c g c t t a t t t g t a t a t t t t c c c g a 2040
 a a t t t t a t a a a a g t t t a g t c t t c g g c c c a a a a t a g a c t c a a a a c c t a a t t t t t a g t c g 2100
 a a c a c g a a a a t a a g a t c t t t a c a t t t c c c a a a a t a a a a t t t t t g g g a a c a a c a a g t t t t 2160
 t g g g a c c t a a a a t c t t t t a g t t t a a t t t t a g c a a a a a c c t t g t t t t c t t t c t t t a a g t a a 2220
 a a c a c a t t t t g t a a t t a a a t g a g a a t t a g a t g a a t t g a a a t c a t t a a a a a g t c a a t t a g g 2280

0802W01_ST25

t t c g a g t t t a t t t t t t t g t a g a a a a t t t a g c t a a a a g t t t t t t c t a a t t t a a a g t t c	2340
c t a t a t a a t a g a t t t a g a a a a a g t c t t g a a a t t t t t g a a a g c c a a t t a c a a a a a a a t t	2400
t g g g g t c t t t c t g t t t t t t g t t a a c a a a c a a t g g a a c g g t t a c t c a g t c c t a a a a t c	2460
t t a g t t t t g a g a c a a a a a c a t t g t t t t t a a t t t t t a t t t t g c c t c t t t a a t g t c a a a t	2520
a g a a g t t a a t t c a t a c t a g t t t g g t t c a t c t c g g c c c t t t c c t t t a g a c a a g a t g t t g a g	2580
g c c c a a a c t g t g t t t t t g a g c g c a c t a a a a a t t t t g t t a a t a a a t t t t t g t t t t c g g t	2640
t t t c t t t t t a t t a c a g a a a g t a a a a a a a a a t t a t a t c a t t g t a a a a c t a t g g c a a c t g	2700
g a a c t a c a t c a a a a g t a a a a t c t c t g a c g a c a	2730

<210> 136
 <211> 2996
 <212> DNA
 <213> *Chlamydomonas vulgaris*

<400> 136	
t a c g a t c g g g t c g t t t a t t g c t g t g c a t t t a a t g c a t a c t t c t t t a g t t t c t g g t t g g g c	60
t g g t t c a a t g g c t t t t a t g a a c t t g c t g t t t t g a t c c t t c t g a t c c a g t t t t a a a c c c	120
a a t g t g g c g t c a a g g g a t g t t t g t t c t t c c t t t t a t g a c a c g t t t a g g c a t c a c t c a a t c	180
t t g g g g t g g t t g g a c t a t c a g t g g t g a a a c a g c a t c a a a t c c a g g t a t t t g g a g t t a t g a	240
g g g g g t t g c t g c t g c t c a c a t c a t t t t a t c t g g a t t a c t t t t g c t g c t t c t a t t t g g c a	300
c t g g g t t t a t t g g g a c c t t g a g c t t t t c c g t g a c c c a c g t a c g t c a a a t c c a g c a t t a g a	360
c c t t c c a a a a a t t t t t g g t a t t c a c t t a t t t t t a t c a g g t c t t c t t t g c t t t g g t t t t g g	420
a g c t t t c c a t g t a a c t g g a t t a t t t g g t c c t g g t a t t t g g t t t c a g a t c c t t a t g g t a t	480
t a c a g g a a c t g t t c a a g c a g t t g c t c c t t c t t g g g a t g c t a c a g g a t t t g a t c c t t a c a a	540
c c c a g g t g g a a t t t c a g c a c a t c a c a t t g c t g c a g g t a t t t t a g g c g t c t t a g c t g g t t t	600
a t t c c a c c t t t g t g t t c g t c c a c c a a a a g a t t a t a c a a t g g a c t t c g c a t g g g t a a c a t	660
t g a a a c a g t t c t t t c t a g c a g t a t t g c a g c a g t t t t c t g g g c a g c a t t t g t t g t a t c g g g	720
a a c t a t g t g g t a t g g t t c t g c t g c a a c a c c a a t t g a a c t t t t t g g t c c a a c t c g t t a c c a	780
a t g g g a t t t a g g t t t c t t c c a a c a g a a a t t g a a c g t c g t g t a c a a a c a a g t c t t g c t g a	840
a g g a a a a t c a g c t t c a c a a g c t t g g g c a g a a a t t c c a g a a a a t t a g c t t t t t a c g a t t a	900
t a t t g g a a a c a a c c c a g c a a a a g g t g g t c t t t t c c g t g c c g g t g c t a t g a a t a g t g g c g a	960
t g g t a t t g c a g t a g g t t g g t t a g g c c a t g c t g t t t t c a a a g a t a a a c a a g g t a a t g a a c t	1020
t t t t g t t c g t c g t a t g c c a a c t t t c t t t g a a a c a t t c c c t g t a g t t c t t c t t g a c a a a g a	1080
c g g t g t t g t t c g t g c a g a c g t g c c t t t c c g t c g t t c t g a a t c t a a a t a c a g t a t t g a a c a	1140
a g t g g g t g t t t c t g t t a c a t t c t a c g g t g g t g a a t t a g a t g g t g t a a c a t t t a g t g a t c c	1200

0802W01_ST25

agcaacagt g aaaaaat at g ct cgacgcgc gcaatt agga gaaat ttt cg aatt t gat cg 1260
 t gccact ct a caat ct gat g ggg ttt ccg tacaagccct cgt ggt t ggt t t act ttt gc 1320
 t cat t t at gt ttt g cact cc ttt t ctt ctt t ggt cat att t tggcacggg g ct cgt acaat 1380
 ctt ccgagat gt at ttt gcag gt at t gat gc agat t t agac gaacaagt ag aat ttt ggt gc 1440
 att ctt aaaa ctt ggt gat a ctt caact cg t cgt caat ca gt at aat t ca ttt ttt ctt t 1500
 t act t cct ct ct caaat ttt t caaat ttt gg gagaat t t ac caaaact gaa at aat ttt gca 1560
 agcct ct at c at ctt aaaaa t at t gtt t gt aagaaaaaaa agagg t cgg t t at t t gt t ct 1620
 ccaaact ttt ttt cgaagt t ttt agagaaa aaagt ttt ggt cccaaaaacg t agt ttt t gg 1680
 ggaaaggaag t at ttt ttt cg ttt t ctt ccc ttt t t ctt gt a ttt ttt ttt ac ctt t gct ttt c 1740
 caacaaaaaca gt at t gtt t gg agaaat cgaa t t agt cccaa aat t t aaat t t t gggact ag 1800
 gt at t agcat gaaaaaaaaat gat agaaaaa ggt aaacagt gcact ttt ct cccaaact aa 1860
 at gat ttt t c at ct agt t gg t t act aacaa acagaact ttt ttt ct at gga agcgt t agt t 1920
 t at act ttt t t at t agt t gg aacat t aggt att at ctt ttt ttt gccat ttt ctt t agagaa 1980
 ccaccacgt a t t gt aaaat a aaagt accat ttt t ggt ttt cgt t gaaaaa aact ttt gat 2040
 at ttt t catt t at ttt caaaa gtt at aaaat t t ggaat aaa ggg t t aat ttt t cagaccaga 2100
 ttt ttt ccca aaaact ttt gt ttt t gggat t gggagaatt a t t act gt ttt agaaacacca 2160
 aagct t at t c t aaaaacaat a at agt ttt ag caccaaggaa aat t cacat t cccgaaaat t 2220
 cgaat ggt ct t aat at ttt ac t gact ttt gt g agt ttt aaaa ct cacaaggt t at t t aaacg 2280
 aat t aaaaag ggt t aat at a ttt cct t ctt att t ct at ttt agt agt ttt aa t gt aaaacag 2340
 t agt gaagaa ct agt ttt ggg aat t t at ttt aat ctt ttt c t t at t aaagg act ttt gt ga 2400
 at agt t ct aa at t t agt t gg t cact ttt ga aaaaaagt aa t aat aat t aa agaaaact aa 2460
 t gaaat ttt ga ttt act agag catt aat ctt cat gt t ctt c aaaaggat ca cgaagt ttt t 2520
 gt gacggg gg cccaaagcca acat agat t g aat aaccagt aat act t aat aaaagacacc 2580
 at aaaaaaat ggt aaagaaa aaagct ggac ttt ccat agg t aaaaat t ac aaaact aaga 2640
 aat ttt ctt t a gaagat aaaa ctt ttt ttt aa ct gt t t agt t ttt at gaaaa cagaaaaaaa 2700
 t t at ctt cga t ct at t ctt g ttt t catt ttt accaaaaaat aagcaaaaag t t at gt t aaa 2760
 gt at t ct act t t at agaaaa gaact ttt ttt act ttt gt aa aat gaaaat a t ctt ttt ttt gt 2820
 att ct aaaac ttt t gagaaa ctt aaact t a ct at at aat a aagat aagag ggagaaaaaa 2880
 gct t aaaagc ttt t t ctt ttt t acagaa agt aaaaaaa at t ctt ttt at t t caaaaat t 2940
 at ggcaact g gaact acat c aaaagt aaaa t cagaagat a ct ggaat t ca act cca 2996

<210> 137
 <211> 1982

<212> DNA

<213> Tetrasel mis sueci ca

<220>

<221> misc_feature

<222> (1050)..(1052)

<223> N is A, C, T or G

<400> 137

gtcgtttatt gcggtgcatt taaatgcacac atcattagt tctggttggg ctggttcaat	60
ggccttttat gaacttgctg tttttgatcc atcagatcca gtttt aaacc caatgtggcg	120
tcaaggatg tttgttttac ctttcatgac tctgtt aggt atcacccaat cttgggggtgg	180
ttggacaatt agtgggaaa cggcttcaaa cccaggatc tggagttatg aagggtgttgc	240
tgcggctcac atcgttttat ctggagctct ttttggggca gctatttggc attgggtttt	300
ttgggattta gaattattcc gtgacccaag aacaggtaac cctgcattag atttaccaaa	360
aatttttggg attcacttat tcttatcagg gttattatgt tttggttttg gagcattcca	420
tgtaacagggt ttatttggac ctgggatttg ggtttctgat ccttatggat taacaggaag	480
tgttcaacca gtatctccat catggggagc cgatgggttt gatccataca atccgggtgg	540
tattgcatct caccatattg ccgcaggat ttt aggtatt attgctggtt tattccattt	600
atgtgttcgt cctccacaac gtttatataa tggcttctgt atggggaaca ttgaaacagt	660
tcttctagt agtatgctg ctgttttctg ggcagctttt gttgttgcag gaacaatgtg	720
gtatggttgt gcggaacac caattgaatt atttggccca actcgttacc aatgggatca	780
agggtatttc caagaggaaa ttacaaaacg tgtagaaaaa tctttatctg aagggaatc	840
tttatcagaa gcttggctctc aaattcctga aaaattagct tttctatgatt acattggaaa	900
caaccagct aagggtgggt ttattccgt a ctgggctat gaacagtggc gatggtattg	960
ctgttgggtt ggttagtcat gcagttttcc cagattt aga tggatattgag tttatcagtt	1020
tctgctatg cccacgttct tgaacttttn nnagttaaat ttacggatcc cagcaacccg	1080
tt aagaaat a t gctcgtcgt gctcaattta ggagaaaatt ttgaaatttg accgtgccac	1140
attacaaatc agatggtgtt ttccgaagca gtccacgtgg ttggtttact ttgggcattt	1200
atcattt gct ttattat ttt tctttggtca tatttggcac ggagctcgt a caatcttccg	1260
t gatgttttt gcagggattg atccagattt agatgagcaa gtagaatttg gggcattcca	1320
aaaatt agga gatacaacga ctcgtcgtca atctgttt ag tttttcattt tgaattcatt	1380
cctcggattc aattat atcc gctt aaatca attattcttt taaaaattta ttatggaagc	1440
tcttgtttac acatttttac ttgt aggaac ttt aggtatc atctttttt caatcttttt	1500
tagagaacca cctcgtattg caaaat aaat agtttaactt caaacttatt atcaaaattg	1560
ttgtgaattg gggatt aacc caattcacia caattcaaat attaatcaaa aatagtttgg	1620

0802W01_ST25

t t a a t c t t c a t g c t c t t c a a a c g g a t c a c g a a g a t c c t t t g a a g g t g g a c c a a c c c a a c	1680
a t a a a c t g a a t a c c c t g t a a t a c t t a a a g t a a a c a c c a t a a g a a a a c a g t a a a a a a a a	1740
a g c a g g a c t t t c c a t a a g t a a a t t t t t t a a t c t t t t a a g t t a a a t t c a a t a g c a a t a a c	1800
t a a c t t a t t g a a t t t a t a a a a c t c t a t t a t t g t t a t a t c a t a a g t g a a a g a a c t t t t g a c	1860
t g a a a a a t t c a a g a a a g t a a a a a t a a t t c c t a c g t t t a t a t a t t a t g g c a a c a g g a a c a	1920
t c a a a a g c a t c t a a a a a t a c a c c t g t a a a t a c a g c a a t g g c g a c a c a a t t a g a a c g c t a t	1980
a a	1982

<210> 138
 <211> 24
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> PCR primer

<400> 138	
ccacct cgt a t ggt aaaat a at t g	24

<210> 139
 <211> 24
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> PCR primer

<400> 139	
gaaagaat t a t ggacagt cc t gct	24

<210> 140
 <211> 21
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> PCR primer

<400> 140	
gaaggaggt c caaaact cac a	21

<210> 141
 <211> 20
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> PCR primer

<400> 141	
cct ggt t c t t g a a g t g c a t c	20

<210> 142
 <211> 24
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> PCR primer

<400> 142
 tgagttggga aactttagct tctt 24

<210> 143
 <211> 20
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> PCR primer

<400> 143
 aaaagattgc caagaccaa 20

<210> 144
 <211> 25
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> PCR primer

<400> 144
 aaaaagaatg aaatttttat gttcg 25

<210> 145
 <211> 20
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> PCR primer

<400> 145
 atggatgtcg tcctccaaa 20

<210> 146
 <211> 762
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> codon optimized sequence comprising BD11

<400> 146
 atggtagcag tatctttcac aagcttttta gcagcatctc caccttcacg tgaagttgc 60

cgtccagctg ctgaagtgga atcagttgca gt agaaaaac gt caaacaat tcaaccaggt 120

acaggttaca ataacggtta cttttattct tactggaatg atggacacgg tgggtgttaca 180

0802W01_ST25

t at act aat g gacct ggt gg t caat t t agt gt aaat t gga gt aact cagg caat t t t gt t 240
 ggaggaaaag gt t ggcaacc t ggt acaaag aat aagggt aa t caat t t ct c t ggt agt t ac 300
 aaccct aat g gt aat t ct t a t t t aagt gt a t acggt t gga gccgt aaccc at t aat t gaa 360
 t at t at at t g t agagaact t t ggt acat ac aaccct t caa cagggt gct ac t aaat t aggt 420
 gaagt t act t cagat ggat c agt t t at gat at t t at cgt a ct caacgcgt aaat caacca 480
 t ct at aat t g gaact gccac t t t ct accaa t act ggagt g t aagacgt aa t cat cgt t ca 540
 agt ggt agt g t t aat acagc aaaccact t t aat gcat ggg ct caacaagg t t t aacat t a 600
 ggt acaat gg act at caaat t gt agct gt t gaagggt t at t t t t cat cagg t agt gct t ct 660
 at cact gt t a gcggt accgg t gat t acaaa gat gat gacg at aaaagt gg t gaaaacct t 720
 t at t t t caag gccat aat ca ccgt caaaa cacact ggt t aa 762

<210> 147
 <211> 1332
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> codon optimized sequence comprising IS99

<400> 147
 at gacagt at acacagct t c agt aacagct ccagt aaaca t t gct acat t aaaat act gg 60
 ggt aaacgt g acacaaaat t aaact t acca acaaact cat caat t t cagt aacat t at ca 120
 caagacgact t acgt acat t aacat cagct gct acagct c cagaat t cga acgt gacaca 180
 t t at ggt t aa acggt gaacc acact caat t gacaacgaac gt acacaaaa ct gt t t acgt 240
 gact t acgt c aat t acgt aa agaaat ggaa t caaaagacg ct t cat t acc aacat t at ca 300
 caat ggaaat t acacat t gt at cagaaaac aact t cccaa cagct gct gg t t t agct t ca 360
 t cagct gct g gt t t cgct gc t t t agt at ca gct at t gct a aat t at acca at t accacaa 420
 t caacat cag aaat t t cacg t at t gct cgt aaagggt t cag gt t cagct t g t cgt t cat t a 480
 t t cggt ggt t acgt agct t g ggaaat ggggt aaagct gaag acggt cacga ct caat ggct 540
 gt acaaat t g ct gact cat c agact ggcca caaat gaaag ct t gt gt at t agt agt at ca 600
 gacat t aaaa aagacgt at c at caacacaa ggt at gcaat t aacagt agc t acat cagaa 660
 t t at t caaag aacgt at t ga acacgt agt a ccaaaacgt t t cgaagt aat gcgt aaagct 720
 at t gt agaaa aagact t cgc t acat t cgct aaagaaacaa t gat ggact c aaact cat t c 780
 cacgct acat gt t t agact c at t cccacca at t t t ct aca t gaacgacac at caaaacgt 840
 at t at t t cat ggt gt cacac aat t aaccaa t t ct acggt g aaacaat t gt agct t acaca 900
 t t cgacgct g gt ccaaacgc t gt at t at ac t act t agct g aaaacgaat c aaaat t at t c 960

0802W01_ST25

gctttcattt acaaattatt cgggttcagta ccagggttggg acaaaaaatt cacaacagaa 1020
 caattagaag ctttcaacca ccaattcgaa tcatcaaact t cacagctcg tgaattagac 1080
 ttagaattac aaaaagacgt agctcgtgta attttaacac aagt aggttc aggtccacaa 1140
 gaaacaaacg aatcattaat tgacgctaaa acagggtttac caaaagaaac cgggttaccba 1200
 t acgacgtac ctgactatgc ttacccttac gacgtaccag actatgctta tccat acgac 1260
 gtaccagact acgctgaaaa ctat acttc caaggtcacc accaccacca ccatcaccac 1320
 ccaccaggtt aa 1332

<210> 148
 <211> 660
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> codon optimized sequence comprising CAT

<400> 148
 atggagaaaa aaatcactgg atat accacc gttagat atat cccaatggca t cgt aaagaa 60
 cat t t t gagg cat t t cagt c agt t gct caa t gt acct at a accagaccgt t cagct ggat 120
 att acggcct t t t t aaagac cgt aaagaaa aat aagcaca agt t t t at cc ggcct t t at t 180
 cacattcttg cccgcctgat gaatgctcat ccggaattcc gtat ggcaat gaaagacggt 240
 gagctggtaga tatgggatag tgttcaccct tgttacaccg t t t t ccat ga gcaaact gaa 300
 acgt t t t cat cgctctggag tgaat accac gacgat t t cc ggcagt t t ct acacat at at 360
 t cgcaagat g tggcgtgtta cggtgaaaac ctggcctatt tccctaaagg gt t t at t gag 420
 aat atgt t t t t cgtctcagc caatccctgg gtgag t t t ca ccagt t t t ga t t t aaacgt g 480
 gccaat at gg acaact t ct t cgccccgt t t t caccat gg gcaaat at t a t acgcaaggc 540
 gacaaggtagc t gat gccgct ggcgat t cag gt t cat cat g ccgt t t gt ga t ggct t ccat 600
 gt cggcagaa t gct t aat ga at t acaacag t act gcgat g agt ggcaggg cggggcgt aa 660

<210> 149
 <211> 90
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> PCR primer

<400> 149
 agt t t t t ct c aat t t t t at t t t t t gt t t t t ct ct aa aaat caaaaa t t caat t t t g 60
 agaaaacgt a agat ct cct a ggaaaat gaa 90

<210> 150
 <211> 90

<212> DNA
 <213> Artificial Sequence
 <220>
 <223> PCR primer
 <400> 150
 t c g g t t t a a g a c a a t g g g a a a a g t t a g a t g c c t a g a g t a t t g a t t a t c g a g c a a a t a t c t 60
 t c t c a t c t g t g a c g g g c t c g a g a c t a g t g g 90

<210> 151
 <211> 90
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> PCR primer
 <400> 151
 t g g a t t g g t a t c a a c g c g c a g g c a t a g t t c g a g a a a a t t a t c c a g a g g c a a t g a c a a c c 60
 a g c a t c t c c t a g t g c t a g c t a a a g a a g t t g 90

<210> 152
 <211> 90
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> PCR primer
 <400> 152
 t c a a a a a a t t c a t a c t t t g t t t t t t a t t t t t c t g a g t t t t a a t c a a a a a a c t t t t t g 60
 t a t a a a a t t g g g c t c g a g a c t a g t t t g t c c 90

<210> 153
 <211> 22
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> PCR primer
 <400> 153
 t c t t a c t g g a a t g a t g g a c a c g 22

<210> 154
 <211> 22
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> PCR primer
 <400> 154
 g t g t t t g t g a c g g t g a t t a t g g 22

<210> 155
 <211> 22
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> PCR primer

<400> 155
 tgtggacctg aacctacttg tg

22

<210> 156
 <211> 22
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> PCR primer

<400> 156
 gaaatgggt a aagctgaaga cg

22

<210> 157
 <211> 20
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> PCR primer

<400> 157
 cttccaaaac cacctgttgc

20

<210> 158
 <211> 20
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> PCR primer

<400> 158
 accgtctgga tcaaaagcag

20

<210> 159
 <211> 20
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> PCR primer

<400> 159
 ttggagtgg tctgttcgtg

20

<210> 160
 <211> 21

<212> DNA
 <213> Artificial Sequence

<220>
 <223> PCR primer

<400> 160
 cagcgt acat acgt cct gga t 21

<210> 161
 <211> 20
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> PCR primer

<400> 161
 gtt gcgct ca accaacatt a 20

<210> 162
 <211> 20
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> PCR primer

<400> 162
 gt gacggt gg ttgt gt cct t 20

<210> 163
 <211> 20
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> PCR primer

<400> 163
 cct gcaggt g gtt ctt caat 20

<210> 164
 <211> 20
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> PCR primer

<400> 164
 at gt caat ag cgccaacaca 20

<210> 165
 <211> 25
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> PCR primer
 <400> 165
 t ggat t at aa agat gat gac gacaa 25

<210> 166
 <211> 22
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> PCR primer
 <400> 166
 gct gct gcaa ct ggt aaat a ga 22

<210> 167
 <211> 20
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> PCR primer
 <400> 167
 t ccagcagaa t caaaagcaa 20

<210> 168
 <211> 20
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> PCR primer
 <400> 168
 gcacct t cag gt aagcct t g 20

<210> 169
 <211> 21
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> PCR primer
 <400> 169
 aagacgat ga cgacaaaggt g 21

<210> 170
 <211> 21
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> PCR primer

0802W01_ST25

<400> 170
t g t t a t c a g c a c g a c c t t c a a

21