

GENETICALLY ENGINEERED MICROBIAL STRAINS INCLUDING *PROTOTHECA*
LIPID PATHWAY GENES

CROSS-REFERENCE TO RELATED APPLICATIONS

5 [0001] This application claims the benefit under 35 U.S.C. 119(e) of US Provisional Patent Application No. 61/564,247, filed November 28, 2011, US Provisional Patent Application No. 61/581,538, filed December 29, 2011, and US Provisional Patent Application No. 61/674,251, filed July 20, 2012. Each of these applications is incorporated herein by reference in its entirety for all purposes.

10 **REFERENCE TO A SEQUENCE LISTING**

[0002] This application includes a sequence listing.

FIELD OF THE INVENTION

[0003] In certain embodiments genetically engineered microbes, e.g., *Prototheca* strains are provided that show improved oil production and so find uses in the fields of chemistry, food production, fuel and industrial chemicals production, microbiology, and
15 molecular biology.

DESCRIPTION OF RELATED DISCLOSURES

[0004] Microalgae, including genetically engineered microalgae, have been identified as important new sources of oil for use in food and fuels. See PCT Pub. Nos. WO
20 2008/151149, WO 2009/126843, and WO 2010/045368, each of which is incorporated herein by reference in its entirety for all purposes. While *Chlorella* strains have been the focus of much effort in developing microbial oil production methods, more recently strains of the genus *Prototheca* have been identified as also having promise as a new source of microbial oils, including tailored oils for specific applications. See PCT Pub. Nos. WO
25 2010/063031, WO 2010/063032, and WO 2011/150410 and WO 2012/106560, each of which is incorporated herein by reference in its entirety for all purposes.

SUMMARY

[0005] In certain embodiments, the invention provides a recombinant (e.g., isolated) nucleic acid comprising a coding sequence that encodes a *Prototheca* lipid biosynthesis
30 protein or portion thereof. In some embodiments, provided is a recombinant nucleic acid comprising a coding sequence that encodes a *Prototheca* lipid biosynthesis protein, provided that the protein is not ketoacyl acyl carrier protein synthase II, stearoyl acyl carrier protein desaturase, or delta 12 fatty acid desaturase. In some embodiments the

protein is not a ketoacyl acyl carrier protein synthase II protein of SEQ ID NO: 270, a stearol acyl carrier protein desaturase of SEQ ID NO: 271, or a delta 12 fatty acid desaturase of SEQ ID NOs: 272 or 273. In some embodiments, the coding sequence is in operable linkage with a promoter, an untranslated control element, and/or a targeting sequence, such as a plastidial targeting sequence and mitochondrial targeting sequence. The recombinant nucleic acid may be, e.g., a DNA molecule. In certain embodiments, recombinant nucleic acid is an expression vector. The recombinant nucleic acid can, for example, include an expression cassette that encodes an mRNA that encodes a functional *Prototheca* lipid biosynthesis enzyme. Alternatively or in addition, the recombinant nucleic acid can include an expression cassette that encodes an inhibitory RNA that suppresses expression of a *Prototheca* lipid biosynthesis gene. In some embodiments, the lipid biosynthesis protein is a protein in Table 1. In some embodiments, the protein has at least 50%, 60%, 70%, 80%, 85%, 90% or 95% sequence identity to a protein provided herein or a protein encoded by a gene provided herein. In some embodiments the protein or the gene encoding the protein is listed in Table 1.

[0006] In some embodiments, the protein encoded by the coding sequence in the nucleic acid contains one or more point mutations, deletions, substitutions, or combinations thereof. In other embodiments, the protein has at least one point mutation in comparison to a protein in Table 1. In some embodiments, the protein encoded by the coding sequence is a functional protein. In some embodiments, the protein is diacylglycerol diacyltransferase (DGAT) having at least one point mutation. In other embodiments, the recombinant nucleic acid further encodes sucrose invertase.

[0007] In certain embodiments, also provided is a genetically engineered microbial cell transformed with a recombinant nucleic acid provided herein. In some embodiments, the cell is a microbial, plant, or yeast cell. In particular embodiments, provided is a cell comprising one or more exogenous gene(s), wherein the exogenous gene is a *Prototheca* lipid biosynthesis gene selected from the genes listed in Table 1. The genetically engineered microbial cell can, for example, be a cell of the genus *Prototheca* or *Chlorella*. In particular embodiments, the cell comprises both an endogenous lipid biosynthesis gene and one or more exogenous *Prototheca* lipid biosynthesis gene(s) selected from the genes listed in Table 1. In certain embodiments, the exogenous gene can encode a lipid biosynthesis protein, wherein the amino acid sequence of the lipid biosynthesis protein is identical to the endogenous lipid biosynthesis protein. For example, the exogenous gene can include a nucleotide sequence in which the codons of the nucleotide sequence

encoding the amino acids of the lipid biosynthesis protein have been altered, as compared to the codons in the native nucleic acid. In various embodiments, the exogenous gene can encode a protein with at least 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% amino acid identity to the native *Prototheca* protein. The exogenous gene
5 can, in some embodiments, be in operable linkage with a promoter element that is not the native *Prototheca* promoter, an untranslated control element that is not the native *Prototheca* untranslated control element, and/or a nucleotide sequence encoding a transit peptide that is not the native *Prototheca* transit peptide. The transit peptide can, for example, be a plastidial targeting sequence or a mitochondrial targeting sequence. In
10 certain embodiments, the cell has a 23S rRNA sequence with at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% nucleotide identity to SEQ ID NO: 5. In particular embodiments, the cell is a *Prototheca* cell, wherein the cell has a fatty acid profile that is at least 10% C8-C14.

[0008] In certain embodiments, another aspect of the invention is a method for
15 obtaining microbial oil comprising culturing a genetically engineered cell, e.g., a *Prototheca* cell, described above under conditions such that oil is produced. In certain embodiments, the microbial oil thus produced has a fatty acid profile that is at least 10% C8-C14. The invention also includes a microbial oil produced by this method.

[0009] In particular embodiments, the invention provides genetically engineered cell,
20 e.g., of the genus *Prototheca*, wherein the activity of one or more endogenous lipid biosynthesis gene, selected from the genes listed in Table 1, has been attenuated. In various embodiments, the activity of the endogenous gene has been attenuated through chromosomal gene deletion, chromosomal gene insertion, frameshift mutation, point mutation, and/or inhibitory RNA. The genetically engineered cell can, in certain
25 embodiments, further comprise an exogenous *Prototheca* lipid biosynthesis pathway gene selected from the genes listed in Table 1. In particular embodiments, one or more allele(s) of an endogenous lipid biosynthesis gene in the genetically engineered cell is attenuated.

[0010] In certain embodiments, one allele of the endogenous lipid biosynthesis gene is replaced, in the genetically engineered cell, with a polynucleotide encoding, e.g., an
30 exogenous *Prototheca* lipid biosynthesis pathway gene selected from Table 1 and a selectable marker. In a variation of this embodiment, two or more alleles of the endogenous lipid biosynthesis gene are each replaced with a polynucleotide encoding an exogenous *Prototheca* lipid biosynthesis pathway gene selected from Table 1 and a

selectable marker. In certain embodiments, the genetically engineered cell has a 23S rRNA sequence with at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% nucleotide identity to SEQ ID NO: 5. In particular embodiments, the genetically engineered cell is a *Prototheca* cell, wherein the cell has a fatty acid profile that is at least 10% C8-C14. In some embodiments, the cell has a fatty acid profile that is at least 50%, 60%, or 70% C12:0. In some embodiments, the cell has a fatty acid profile that is at least 70%, 75%, 80%, 85%, or 90% C18:1. In certain embodiments, the invention also provides a method for obtaining microbial oil comprising culturing this genetically engineered cell, which may be, e.g., a *Prototheca* cell, under conditions such that oil is produced. In certain embodiments, the microbial oil thus produced has a fatty acid profile that is at least 10% C8-C14. The invention also includes a microbial oil produced by this method.

[0011] In another aspect, the present invention provides a genetically engineered microbial cell, e.g., *Prototheca* cell, in one or more lipid biosynthesis genes have been modified to increase or decrease expression of such one or more genes such that the fatty acid profile of the genetically engineered strain differs from that of the strain from which it was derived. In one embodiment, at least two genes have been modified. In various embodiments, the genetic modifications include one or more of the following modifications: (i) attenuation of a gene or its enzymatic product; and (ii) increased expression of a gene or its enzymatic product; (iii) altered activity of a gene or its enzymatic product.

[0012] In various embodiments, the genetically engineered cell has one or more attenuated genes, wherein the genes attenuated have been attenuated by a means selected from the group consisting of a homologous recombination event and introduction of an exogenous gene that codes for an interfering RNA. In various embodiments, one or more alleles of a gene are attenuated.

[0013] In various embodiments, the genetically engineered cell has one or more over-expressed genes, wherein the genes over-expressed have been up-regulated by a means selected from the group consisting of introduction of additional copies of said gene into said cell; introduction of new expression control elements for said gene; and alteration of the protein-coding sequence of the gene. In various embodiments, one or more alleles of a gene are over-expressed.

[0014] In various embodiments, the modified genes of the genetically engineered cell are selected from the group consisting of *Prototheca* lipid biosynthesis genes presented in Table 1. In various embodiments, the genetically engineered cell comprises an exogenous gene selected from the group consisting of *Prototheca* lipid biosynthesis genes presented in Table 1. In various embodiments, the genetically engineered cell comprises one or more over-expressed alleles of a gene, the gene selected from the group consisting of *Prototheca* lipid biosynthesis genes presented in Table 1. In various embodiments, the genetically engineered cell has an attenuated gene selected from the group consisting of *Prototheca* lipid biosynthesis genes presented in Table 1. In various embodiments, the genetically engineered cell has one more attenuated alleles of a gene, the gene selected from the group consisting of *Prototheca* lipid biosynthesis genes presented in Table 1.

[0015] In various embodiments, the genetically engineered cell has a fatty acid profile selected from the group consisting of: 3% to 60% C8:0, 3% to 60% C10:0, 3% to 70% C12:0, 3% to 95% C14:0, 3% to 95% C16:0, 3% to 95% C18:0, 3% to 95% C18:1, 0% to 60% C18:2, 0% to 60% C18:3 or combinations thereof. In various embodiments, the ratio of C10:0 to C12:0 is at least 3:1. In some cases, the ratio of C10:0 to C14:0 is at least 10:1. In various embodiments, the ratio of C12:0 to C14:0 is at least 3:1. In various embodiments, the genetically engineered cell has a fatty acid profile of at least 40% saturated fatty acids, of at least 60% saturated fatty acids, or at least 85% saturated fatty acids.

[0016] In another aspect, the present invention provides methods for obtaining microbial oil comprising culturing a genetically engineered *Prototheca* cell of the invention under conditions such that oil is produced. In various embodiments, the microbial oil has a fatty acid profile selected from the group consisting of: 3% to 40% C8:0, 3% to 60% C10:0, 3% to 70% C12:0, 3% to 95% C14:0, 3% to 95% C16:0, 3% to 95% C18:0, 3% to 95% C18:1, 0% to 60% C18:2, 0% to 60% C18:3 or combinations thereof. In various embodiments, the ratio of C10:0 to C12:0 is at least 3:1. In some cases, the ratio of C10:0 to C14:0 is at least 10:1. In various embodiments, the ratio of C12:0 to C14:0 is at least 3:1. In various embodiments, the genetically engineered cell has a fatty acid profile of at least 40% saturated fatty acids, of at least 60% saturated fatty acids, or at least 85% saturated fatty acids.

[0017] In an additional aspect, the present invention provides microbial oils and foods, fuels, and chemicals containing said oil or a chemical derived therefrom.

[0018] In another aspect, the present invention provides recombinant nucleic acids useful in methods for making genetically modified *Prototheca* and other cells. The nucleic acids of the invention comprise all or some portion of a *Prototheca* lipid biosynthesis gene.

5 [0019] In various embodiments, these nucleic acids include expression cassettes, which consist of a coding sequence and control sequences that regulate expression of the coding sequence, which may code for an mRNA that encodes a lipid biosynthesis protein, enzyme, or for an RNAi that acts to suppress expression of a lipid biosynthesis gene.

10 [0020] In other embodiments, these nucleic acids are expression vectors that include one or more expression cassettes and stably replicate in a *Prototheca* or other host cell, either by integration into chromosomal DNA of the host cell or as freely replicating vectors.

15 [0021] In other embodiments, these nucleic acids comprise only a portion of a *Prototheca* lipid biosynthesis gene, which portion may be a portion of a coding sequence, an exon, or a control element. Such nucleic acids are useful in the construction of expression cassettes for *Prototheca* and non-*Prototheca* host cells, for integration of exogenous DNA into *Prototheca* host cells, and for construction of nucleic acids useful for attenuating *Prototheca* lipid biosynthetic genes by homologous recombination.

20 [0022] In some embodiments, provided are sequences, compositions, host cells, and methods for overexpression of a lipid biosynthesis gene. In some aspects, the overexpressed lipid biosynthesis gene is one or more of LEC2, DGAT, ATP:citrate lyase (ACL), malic enzyme, lipase, fatty acyl-CoA reductase, Acyl-CoA Binding Proteins (ACBPs), or Lipoic Acid Synthase (LS1).

25 [0023] These and other aspects and embodiments of the invention are described in the accompanying drawings, a brief description of which immediately follows, the detailed description of the invention below, and are exemplified in the examples below. Any or all of the features discussed above and throughout the application can be combined in various embodiments of the present invention.

BRIEF DESCRIPTION OF THE DRAWINGS

30 [0024] Figure 1 shows PstI restriction maps of *Prototheca moriformis* FADc alleles with and without a targeted gene disruption, as described in Example 37.

[0025] Figure 2 shows the results of the Southern blot described in Example 37.

DETAILED DESCRIPTION

[0026] For the convenience of the reader, this detailed description of the invention is divided into sections. Section I provides definitions of terms used herein. Section II provides an overview of the *Prototheca* lipid biosynthesis pathway. Section III describes culturing methods for *Prototheca* cells of the invention. Section IV describes genetic engineering methods for *Prototheca* and genetically engineered cells of the invention. Section V describes microbial oils provided by the invention. Section VI describes nucleic acids of the invention.

SECTION I. DEFINITIONS

[0027] Unless defined otherwise, all technical and scientific terms used herein have the meaning commonly understood by a person skilled in the art to which this invention belongs. The following references provide one of skill with a general definition of many of the terms used in this invention: Singleton *et al.*, *Dictionary of Microbiology and Molecular Biology* (2nd ed. 1994); *The Cambridge Dictionary of Science and Technology* (Walker ed., 1988); *The Glossary of Genetics*, 5th Ed., R. Rieger *et al.* (eds.), Springer Verlag (1991); and Hale & Marham, *The Harper Collins Dictionary of Biology* (1991). As used herein, the following terms have the meanings ascribed to them unless specified otherwise.

[0028] A nucleic acid “active in microalgae” refers to a nucleic acid that is functional in microalgae. For example, a promoter that has been used to drive an antibiotic resistance gene to impart antibiotic resistance to a transgenic microalgae is active in microalgae.

[0029] “Acyl carrier protein” or “ACP” is a protein that binds a growing acyl chain during fatty acid synthesis as a thiol ester at the distal thiol of the 4'-phosphopantetheine moiety and comprises a component of the fatty acid synthase complex.

[0030] “Acyl-CoA molecule” or “acyl-CoA” is a molecule comprising an acyl moiety covalently attached to coenzyme A through a thiol ester linkage at the distal thiol of the 4'-phosphopantetheine moiety of coenzyme A.

[0031] “Allele” refers to one or two or more forms of a gene or genetic locus. Alleles of a gene may share 100% or less nucleotide sequence identity. Gene products encoded by alleles of a gene may share 100% or less amino acid sequence identity.

Overexpression of different alleles of a gene and/or the gene products encoded therein may confer different phenotypes to a genetically engineered organism. Attenuation of

different alleles of a gene and/or the gene products encoded therein may confer different phenotypes to a genetically engineered organism.

[0032] "Attenuation of a Gene" refers to (i) genetically engineering a gene so that it has, relative to a wild-type gene, different control sequences that result in decreased amounts of a gene product (RNA including mRNA, inhibitory RNA molecules, and other RNAs, polypeptides); (ii) genetically engineering a cell so that it has, relative to a wild-type cell, fewer or no detectable copies of a gene and decreased amounts of the corresponding gene product; and/or (iii) genetically engineering the coding sequence of a gene to either decrease the stability and/or activity of the gene product (i.e., if the increase the stability of an RNA gene product, increase translation of an mRNA gene product, and/or decrease the level of enzymatic activity of a protein encoded by the mRNA gene product, i.e., by making the protein less stable or less active (which may also be referred to as "attenuation of an Enzymatic Product"). An "Attenuated Gene Product" is the gene product of attenuation of a gene by any of the foregoing methods. An "Attenuated Gene" is a gene that has been genetically engineered by one or more of the methods described herein that results in decreased amounts of gene product. Attenuation of a gene thus results in "Decreased Expression of a Gene", "down-regulation of the gene", or "inactivation of the gene".

[0033] "Axenic" is a culture of an organism substantially free from contamination by other living organisms.

[0034] "Biomass" is material produced by growth and/or propagation of cells. Biomass may contain cells and/or intracellular contents as well as extracellular material, includes, but is not limited to, compounds secreted by a cell.

[0035] "Catalyst" is an agent, such as a molecule or macromolecular complex, capable of facilitating or promoting a chemical reaction of a reactant to a product without becoming a part of the product. A catalyst increases the rate of a reaction, after which, the catalyst may act on another reactant to form the product. A catalyst generally lowers the overall activation energy required for the reaction such that it proceeds more quickly or at a lower temperature. Thus, a reaction equilibrium may be more quickly attained. Examples of catalysts include enzymes, which are biological catalysts; heat, which is a non-biological catalyst; and metals used in fossil oil refining processes.

[0036] "Co-culture", and variants thereof such as "co-cultivate" and "co-ferment", refer to the presence of two or more types of cells in the same bioreactor. The two or more

types of cells may both be microorganisms, such as microalgae, or may be a microalgal cell cultured with a different cell type. The culture conditions may be those that foster growth and/or propagation of the two or more cell types or those that facilitate growth and/or proliferation of one, or a subset, of the two or more cells while maintaining cellular growth for the remainder.

[0037] “Coding Sequence” refers to that portion of a gene or expression cassette that encodes the RNA transcribed from that gene or expression cassette in a cell, specifically that portion of the mRNA that is translated into the protein encoded by that mRNA. Any non-translated portions of a gene between translated portions are referred to as “introns”.

[0038] “Cofactor” or “co-factor” is any molecule, other than the substrate, required for an enzyme to carry out its enzymatic activity.

[0039] “Complementary DNA” or “cDNA” is a DNA copy of mRNA, usually obtained by reverse transcription of messenger RNA (mRNA) or amplification (e.g., via polymerase chain reaction (“PCR”)).

[0040] “Control Sequence” refers to nucleic acid sequences in a gene or expression cassette that regulate transcription of a coding sequence and so include promoters, enhancers, transcription termination sequences, and translation initiation sequences.

[0041] “Cultivated”, and variants thereof such as “cultured” and “fermented”, refer to the intentional fostering of growth (increases in cell size, cellular contents, and/or cellular activity) and/or propagation (increases in cell numbers via mitosis) of one or more cells by use of selected and/or controlled conditions. The combination of both growth and propagation may be termed proliferation. Examples of selected and/or controlled conditions include the use of a defined medium (with known characteristics such as pH, ionic strength, and carbon source), specified temperature, oxygen tension, carbon dioxide levels, and growth in a bioreactor. Cultivate does not refer to the growth or propagation of microorganisms in nature or otherwise without human intervention; for example, natural growth of an organism that ultimately becomes fossilized to produce geological crude oil is not cultivation.

[0042] “Cytolysis” is the lysis of cells in a hypotonic environment. Cytolysis is caused by excessive osmosis, or movement of water, towards the inside of a cell (hyperhydration). The cell cannot withstand the osmotic pressure of the water inside, and so it explodes.

[0043] “Delipidated meal” and “delipidated microbial biomass” is microbial biomass after oil (including lipids) has been extracted or isolated from it, either through the use of mechanical (i.e., exerted by an expeller press) or solvent extraction or both. Delipidated meal has a reduced amount of oil/lipids as compared to before the extraction or isolation of oil/lipids from the microbial biomass but does contain some residual oil/lipid.

[0044] “Desaturase” are enzymes in the lipid synthesis pathway responsible for the introduction of double bonds (unsaturation) into the fatty acid chains of fatty acid or triacylglyceride molecules. Examples include but are not limited to stearyl-Acyl carrier protein desaturase (SAD) and fatty acid desaturase (FAD), also known as fatty acyl desaturase.

[0045] “Expression Cassette” refers to a coding sequence and a promoter, optionally in combination with one or more control sequences. Expression cassettes for enzymes include, for example and without limitation, a translation initiation control sequence.

[0046] “Expression vector” or “expression construct” or “plasmid” or “recombinant DNA construct” refer to a nucleic acid that has been generated via human intervention, including by recombinant means or direct chemical synthesis, with a series of specified nucleic acid elements that permit transcription and/or translation of a particular nucleic acid in a host cell. The expression vector may be part of a plasmid, virus, or nucleic acid fragment. Typically, the expression vector includes a nucleic acid to be transcribed operably linked to a promoter. Some expression cassettes are expression vectors, but expression vectors often contain more than one expression cassette, for example expression cassettes for selectable markers are sometimes included in expression vectors for introducing exogenous genes into host cells. One of skill in the art understands that a “recombinant nucleic acid” that encodes a particular gene, or portion thereof, is isolated from the specific context in which it naturally occurs.

[0047] “Exogenous gene” is a nucleic acid that codes for the expression of an RNA and/or protein that has been introduced (“transformed”) into a cell, and is also referred to as a “transgene”. A transformed cell may be referred to as a recombinant cell, into which additional exogenous gene(s) may be introduced. The exogenous gene may be from a different species (and so heterologous), or from the same species (and so homologous), relative to the cell being transformed. Thus, an exogenous gene can include a homologous gene that occupies a different location in the genome of the cell or is under different control, relative to the endogenous copy of the gene. An exogenous gene may be present

in more than one copy in the cell. An exogenous gene may be maintained in a cell as an insertion into the genome (nuclear or plastid) or as an episomal molecule.

[0048] “Exogenously provided” refers to a molecule provided to the culture media of a cell culture.

5 [0049] “Expeller pressing” is a mechanical method for extracting oil from raw materials such as soybeans and rapeseed. An expeller press is a screw type machine, which presses material through a caged barrel-like cavity. Raw materials enter one side of the press and spent cake exits the other side while oil seeps out between the bars in the cage and is collected. The machine uses friction and continuous pressure from the screw
10 drives to move and compress the raw material. The oil seeps through small openings that do not allow solids to pass through. As the raw material is pressed, friction typically causes it to heat up.

[0050] “Fatty acids” shall mean free fatty acids, fatty acid salts, or fatty acyl moieties in a glycerolipid.

15 [0051] “Fatty acid modification enzyme” or “fatty acid modifying enzyme” refers to an enzyme that alters the covalent structure of a fatty acid. Examples of fatty acid modification enzymes include lipase, fatty acyl-CoA/aldehyde reductase, fatty acyl-CoA reductase, fatty aldehyde reductase, fatty aldehyde decarbonylase.

[0052] “Fatty acid profile” refers to the distribution of fatty acids in a cell or oil
20 derived from a cell in terms of chain length and/or saturation pattern. In this context the saturation pattern can comprise a measure of saturated versus unsaturated acid or a more detailed analysis of the distribution of the positions of double bonds in the various fatty acids of a cell. The fatty acid profile in be readily determined, for example by using gas chromatography. In one method, the fatty acids of the triacylglycerol are converted into a
25 fatty acid methyl ester (FAME) using well known methods. The FAME molecules are then detected by gas chromatography. For example, a separate peak is observed for a fatty acid of 14 carbon atoms with no unsaturation (C14:0) compared to any other fatty acid such as C14:1. The peak area for each class of FAME determined using GC-FID is proportional to the weight percentages of the fatty acids. Unless specified otherwise, the
30 fatty acid profile is expressed as a weight percent of the total fatty acid content. When referring to fatty acid profiles, “at least 4% C8-C14” means that at least 4% by weight of the total fatty acids in a cell or in an extracted glycerolipid composition have a chain length that includes 8, 10, 12 or 14 carbon atoms.

[0053] “Fatty acid synthesis enzyme” refers to an enzyme that alters the chain length, saturation, or functional group modification of a fatty acid, or can otherwise lead to an altered fatty acid profile in a cell. Examples of fatty acid synthesis enzymes include fatty acyl-ACP thioesterase, desaturase, including stearoyl acyl carrier protein desaturase (SAD) and fatty acyl destaurase (FAD), fatty acyl hydroxylase, and β -keto-acyl-ACP synthase.

[0054] “Fatty acyl-ACP thioesterase” is an enzyme that catalyzes the cleavage of a fatty acid from an acyl carrier protein (ACP) during fatty acid synthesis.

[0055] “Fatty acyl-CoA/aldehyde reductase” is an enzyme that catalyzes the reduction of an acyl-CoA molecule to a primary alcohol.

[0056] “Fatty acyl-CoA reductase” is an enzyme that catalyzes the reduction of an acyl-CoA molecule to an aldehyde.

[0057] “Fatty aldehyde decarbonylase” is an enzyme that catalyzes the conversion of a fatty aldehyde to an alkane.

[0058] “Fatty aldehyde reductase” is an enzyme that catalyzes the reduction of an aldehyde to a primary alcohol.

[0059] “Fixed carbon source” is a molecule(s) containing carbon, typically an organic molecule, that is present at ambient temperature and pressure in solid or liquid form in a culture media that may be utilized by a microorganism cultured therein.

[0060] “Functional protein” refers to a protein whose activity has been retained even though it may be attenuated.

[0061] “Genetically engineered”, “genetically engineer”, and “genetic engineering” refers to alteration of the DNA and/or RNA of a living cell by human intervention. Typically, the alteration is mediated by the introduction of one or more expression vectors, but in some instances, functionally equivalent alterations may be achieved by mutagenesis alone.

[0062] “Glycerolipid” refers to a glycerol molecule esterified at the sn-1, sn-2 or sn-3 position of the glycerol with one or more phosphate, fatty acid, phosphoserine, phosphocholine, phosphoinositol, or phosphoethanolamine, or other moieties covalently attached to the glycerol backbone. Examples of glycerolipids include triacylglycerides (triglycerides), diacylglycerides (diglycerides), monoacylglycerides (monoglycerides),

glycerol-3-phosphate, lysophosphatidic acid, phosphatidic acid, phosphatidylcholine, phosphatidylserine, phosphatidylglycerol, and phosphatidylethanolamine.

[0063] “Glycerolipid synthesis enzyme” refers to an enzyme involved in the synthesis of glycerolipids. Glycerolipid synthesis enzymes function, for example, to covalently
5 attach acyl groups to a substituted glycerol. Examples of glycerolipid synthesis enzymes include glycerol-3-phosphate acyltransferase, lysophosphatidic acid acyltransferase, diacylglycerol acyltransferase, phospholipid diacylglycerol acyltransferase, and phosphatidic acid phosphatase.

[0064] “Glycerophospholipid” is a glycerolipid that at the sn-1, sn-2 or sn-3 positions
10 of the glycerol backbone has at least one or more covalently bound phosphate or a covalently bound phosphate containing moiety, for example, phosphocholine, phosphoserine, phosphoinositol, and phosphoethanolamine. Glycerophospholipids include phosphoglycerol, lysophosphatidic acid, phosphatidic acid, phosphatidylcholine, phosphatidylserine, phosphatidylglycerol, and phosphatidylethanolamine.

15 [0065] “Heterotrophic” as it pertains to culture conditions is culturing in the substantial absence of light while utilizing or metabolizing a fixed carbon source.

[0066] “Homogenate” is biomass that has been physically disrupted.

[0067] “Hydrogen:carbon ratio” is the ratio of hydrogen atoms to carbon atoms in a molecule on an atom-to-atom basis. The ratio may be used to refer to the number of
20 carbon and hydrogen atoms in a hydrocarbon molecule. For example, the hydrocarbon with the highest ratio is methane CH₄ (4:1).

[0068] “Hydrophobic fraction” is the portion, or fraction, of a material that is more soluble in a hydrophobic phase in comparison to an aqueous phase. A hydrophobic fraction is substantially insoluble in water and usually non-polar.

25 [0069] “Increase lipid yield” refers to an increase in the productivity of a microbial culture by, for example, increasing dry weight of cells per liter of culture, increasing the percentage of cells that constitute lipid, or increasing the overall amount of lipid per liter of culture volume per unit time.

[0070] “Inducible promoter” is a promoter that mediates transcription of an operably
30 linked gene in response to a particular stimulus. Examples of such promoters may be promoter sequences that are induced in conditions of changing pH or nitrogen levels.

[0071] “In operable linkage” is a functional linkage between two nucleic acid sequences, such a control sequence (typically a promoter) and the linked sequence (typically a sequence that encodes a protein, also called a coding sequence). A promoter is in operable linkage with an exogenous gene if it can mediate transcription of the gene.

5 [0072] “In situ” means “in place” or “in its original position”.

[0073] “Limiting concentration of a nutrient” is a concentration of a compound in a culture that limits the propagation of a cultured organism. A “non-limiting concentration of a nutrient” is a concentration that supports maximal propagation during a given culture period. Thus, the number of cells produced during a given culture period is lower in the
10 presence of a limiting concentration of a nutrient than when the nutrient is non-limiting. A nutrient is said to be “in excess” in a culture, when the nutrient is present at a concentration greater than that which supports maximal propagation.

[0074] “Lipase” is a water-soluble enzyme that catalyzes the hydrolysis of ester bonds in water-insoluble, lipid substrates. Lipases catalyze the hydrolysis of lipids into glycerols
15 and fatty acids.

[0075] “Lipids” are a class of molecules that are soluble in nonpolar solvents (such as ether and chloroform) and are relatively or completely insoluble in water. Lipid molecules have these properties, because they consist largely of long hydrocarbon tails which are hydrophobic in nature. Examples of lipids include fatty acids (saturated and unsaturated);
20 glycerides or glycerolipids (such as monoglycerides, diglycerides, triglycerides or neutral fats, and phosphoglycerides or glycerophospholipids); nonglycerides (sphingolipids, sterol lipids including cholesterol and steroid hormones, prenol lipids including terpenoids, fatty alcohols, waxes, and polyketides); and complex lipid derivatives (sugar-linked lipids, or glycolipids, and protein-linked lipids). As used herein, the term “triacylglycerides” and
25 “triglycerides” are interchangeable. “Fats” and “oils” are a subgroup of lipids called “triacylglycerides.” “Oil,” as distinguished from “fat” refers to triacylglycerides that are generally liquid at ordinary room temperature and pressure. Fatty acids are conventionally named by the notation that recites number of carbon atoms and the number of double
30 bonds separated by a colon. For example oleic acid can be referred to as C18:1 and capric acid can be referred to as C10:0.

[0076] “Lipid biosynthesis pathway” or “lipid biosynthetic pathway” or “lipid metabolic pathway” or “lipid pathway” refers to the synthesis or degradation of lipids. Thus enzymes of the lipid biosynthesis pathway (e.g. lipid pathway enzyme) include fatty

acid synthesis enzymes, fatty acid modification enzymes, and glycerolipid synthesis enzymes, as well as proteins (e.g. lipid pathway protein) that affect lipid metabolism, i.e., either lipid modification or degradation, and any proteins that chemically modify lipids, as well as carrier proteins. Lipid biosynthesis proteins also include transcription factors and
5 kinases that are involved in lipid metabolism.

[0077] “Lipid biosynthesis gene” is any gene that encodes a protein that is involved in lipid metabolism, either in lipid synthesis, modification, or degradation, and any protein that chemically modifies lipids including carrier proteins.

[0078] “Lipid pathway enzyme” is any enzyme that plays a role in lipid metabolism,
10 i.e., either lipid synthesis, modification, or degradation, and any proteins that chemically modify lipids, as well as carrier proteins.

[0079] “Lysate” is a solution containing the contents of lysed cells.

[0080] “Lysis” is the breakage of the plasma membrane and optionally the cell wall of a biological organism sufficient to release at least some intracellular content, often by
15 mechanical, viral or osmotic mechanisms that compromise its integrity.

[0081] “Lysing” is disrupting the cellular membrane and optionally the cell wall of a biological organism or cell sufficient to release at least some intracellular content.

[0082] “Microalgae” is a eukaryotic microbial organism that contains a chloroplast or plastid, and optionally that is capable of performing photosynthesis, or a prokaryotic
20 microbial organism capable of performing photosynthesis. Microalgae include obligate photoautotrophs, which cannot metabolize a fixed carbon source as energy, as well as heterotrophs, which can live solely off of a fixed carbon source. Microalgae include unicellular organisms that separate from sister cells shortly after cell division, such as *Chlamydomonas*, as well as microbes such as, for example, *Volvox*, which is a simple
25 multicellular photosynthetic microbe of two distinct cell types. Microalgae include cells such as *Chlorella*, *Dunaliella*, and *Prototheca*. Microalgae also include other microbial photosynthetic organisms that exhibit cell-cell adhesion, such as *Agmenellum*, *Anabaena*, and *Pyrobotrys*. Microalgae also include obligate heterotrophic microorganisms that have lost the ability to perform photosynthesis, such as certain dinoflagellate algae species and
30 species of the genus *Prototheca*.

[0083] “Microorganism” and “microbe” are microscopic unicellular organisms.

[0084] “Naturally co-expressed” with reference to two proteins or genes means that the proteins or their genes are co-expressed naturally in a tissue or organism from which they are derived, e.g., because the genes encoding the two proteins are under the control of a common regulatory sequence or because they are expressed in response to the same stimulus.

[0085] “Overexpression of a Gene” refers to (i) genetically engineering a gene so that it has, relative to a wild-type gene, different control sequences that result in increased amounts of a gene product (RNA and, if the RNA is an mRNA, the protein encoded by the mRNA) in a cell; (ii) genetically engineering a cell so that it has, relative to a wild-type cell, more copies of a gene and increased amounts of the corresponding gene product; and/or (iii) genetically engineering the coding sequence of a gene to either increase the stability and/or activity of the gene product (i.e., if the increase the stability of an RNA gene product, increase translation of an mRNA gene product, and/or increase the level of enzymatic activity of a protein encoded by the mRNA gene product, i.e., by making the protein more stable or more active (which may also be referred to as “Overexpression of an Enzymatic Product”). An “Overexpressed Gene” is the product of overexpression of a gene by any of the foregoing methods. Overexpression of a gene thus results in “Increased Expression of a Gene”.

[0086] “Promoter” is a nucleic acid control sequence that directs transcription of a nucleic acid. As used herein, a promoter includes necessary nucleic acid sequences near the start site of transcription, such as, in the case of a polymerase II type promoter, a TATA element. A promoter also optionally includes distal enhancer or repressor elements, which may be located as much as several thousand base pairs from the start site of transcription.

[0087] “*Prototheca* cell” refers to any cell, strain, and species of microalgae of the genus *Prototheca*. Illustrative *Prototheca* cells and strains include, without limitation, those of any of the following species: *Prototheca wickerhamii*, *Prototheca stagnora*, *Prototheca portoricensis*, *Prototheca moriformis*, and *Prototheca zopfii*. In one important embodiment, a *Prototheca* cell is a cell or strain of *Prototheca moriformis*. More generally, microalgal cells, strains, and species that share greater than 75% sequence identity with the 23s rRNA of *Prototheca moriformis* or that listed in SEQ ID NO: 5.

[0088] “Recombinant” is a cell, nucleic acid, protein or vector that has been modified due to the introduction of an exogenous nucleic acid or the alteration of a native nucleic

acid. Thus, e.g., recombinant cells express genes that are not found within the native (non-recombinant) form of the cell or express native genes differently than those genes are expressed by a non-recombinant cell. A “recombinant nucleic acid” is a nucleic acid originally formed in vitro, in general, by the manipulation of nucleic acid, e.g., using
5 polymerases and endonucleases, or otherwise is in a form not normally found in nature, including an isolated form, i.e., wherein the nucleic acid is separated from at least one other component with which the native form of the nucleic acid naturally occurs.

Recombinant nucleic acids may be produced, for example, to place two or more nucleic acids in operable linkage. Thus, an isolated nucleic acid or an expression vector formed in
10 vitro by ligating DNA molecules that are not normally joined in nature, are both considered recombinant for the purposes of this invention. Once a recombinant nucleic acid is made and introduced into a host cell or organism, it may replicate using the in vivo cellular machinery of the host cell; however, such nucleic acids, once produced recombinantly, although subsequently replicated intracellularly, are still considered
15 recombinant for purposes of this invention. Similarly, a “recombinant protein” is a protein made using recombinant techniques, i.e., through the expression of a recombinant nucleic acid.

[0089] The term “replacement” or “replace” or “replaced” when used in reference to modification of a gene sequence by another refers to the ablation or knockout of an
20 endogenous gene by homologous recombination with an exogenous gene sequence containing suitable flanking regions.

[0090] “Inhibitory RNA” refers to RNA that inhibits gene expression. Inhibitory RNA includes double-stranded interfering RNA. Inhibitory RNA includes long RNA hairpins, which, in some embodiments, are ~200 to 750 nucleotides in length, and comprise a
25 coding sequence of the target gene of 50 to 650 nucleotides and its complementary sequence separated by sequence long enough (typically 25 to 200 nucleotides) to allow the coding sequence and its complementary to form a double-stranded sequence. RNAi also includes microRNAs, which are shorter than long RNA hairpins comprising typically only 19-22 nucleotides of the coding sequence of the target gene and its complement together
30 with flanking sequences to engage the enzymes in the cell that mediate interference with gene expression by RNAi.

[0091] “Sonication” is a process of disrupting biological materials, such as a cell, by use of sound wave energy.

[0092] “Sucrose utilization gene” is a gene that, when expressed, aids the ability of a cell to utilize sucrose as an energy source. Proteins encoded by a sucrose utilization gene are referred to herein as “sucrose utilization enzymes” and include sucrose transporters, sucrose invertases, and hexokinases such as glucokinases and fructokinases.

5 [0093] “Up-regulation of an exogenous gene” refers to (i) genetically engineering a gene so that it has, relative to a wild-type gene, different control sequences that result in increased amounts of a gene product (RNA and, if the RNA is an mRNA, the protein encoded by the mRNA) in a cell; (ii) genetically engineering a cell so that it has, relative to a wild-type cell, more copies of a gene and increased amounts of the corresponding
10 gene product; and/or (iii) genetically engineering the coding sequence of a gene to either increase the stability and/or activity of the gene product (i.e., if the increase the stability of an RNA gene product, increase translation of an mRNA gene product, and/or increase the level of enzymatic activity of a protein encoded by the mRNA gene product, i.e., by making the protein more stable or more active (which may also be referred to as “Up-
15 regulation of an Enzymatic Product”). An “Up-regulated Gene” is the product of increased expression of a gene by any of the foregoing methods. Up-regulation of a gene thus results in “Increased Expression of a Gene”.

SECTION II *PROTOTHECA* LIPID BIOSYNTHESIS PATHWAY

20 [0094] In certain embodiments the present invention provides recombinant *Prototheca* cells that have been modified to alter the properties and/or proportions of lipids or fatty acids produced. The lipid biosynthesis pathway can further, or alternatively, be modified to alter the properties and/or proportions of various lipid molecules produced through enzymatic processing of lipids and intermediates in the lipid biosynthesis pathway. In
25 various embodiments, the recombinant *Prototheca* cells of the invention have, relative to their untransformed counterparts, optimized lipid yield per unit volume and/or per unit time, carbon chain length (e.g., for renewable diesel production or for industrial chemicals applications requiring lipid feedstock), reduced number of double or triple bonds, optionally to zero, and increasing the hydrogen:carbon ratio of a particular species of lipid
30 or of a population of distinct lipid. In other embodiments, the lipids have increased number of double bonds.

[0095] In particular embodiments, one or more key enzymes that control branch points of metabolism of fatty acids and glycerolipids have been up-regulated or down-regulated to improve lipid production. Up-regulation, or over-expression, of genes may be achieved,

for example, by transforming cells with expression constructs in which a gene encoding the enzyme of interest is expressed, e.g., using a strong promoter and/or enhancer elements that increase transcription. Such constructs can include a selectable marker such that the transformants may be subjected to selection, which can result in amplification of the construct and an increase in the expression level of the encoded enzyme. Down-regulation, or attenuation, of genes may be achieved, for example, by transforming cells with expression cassettes that ablate, through homologous recombination, all or a portion of the chromosomally-encoded corresponding gene. Expression levels of lipid pathway enzymes can also optionally be reduced through the use of inhibitory RNA constructs. Optionally, endogenous lipid pathway genes may be modified to alter individually or in combination their enzymatic specificity, level of expression, or cellular localization. The expression cassettes used in up- or down-regulation can replicate by integration into chromosomal DNA of the host cell or as a freely replicating vector.

[0096] Genes and gene products of the *Prototheca morifomis* (UTEX 1435) lipid biosynthesis pathway are listed in Table 1 and detailed in the subsections A-K below. Where noted, different alleles of genes are provided. Typically, modest amino acid changes are seen between the two proteins, with alleles typically being 0-2% polymorphic in exons, and 4-7% polymorphic in introns.

Table 1. *Prototheca* Lipid Biosynthesis Genes

NAD-dependent glycerol-3-phosphate dehydrogenase (SEQ ID NO: 204, nucleotide, SEQ ID NO: 205, protein),
Fumarate hydratase (SEQ ID NO: 170, nucleotide, SEQ ID NO: 171, protein),
Pyruvate dehydrogenase, Alpha subunit (SEQ ID NO: 238, nucleotide, SEQ ID NO: 239, protein)
Pyruvate dehydrogenase, Beta subunit (SEQ ID NO: 240, nucleotide, SEQ ID NO: 241, protein)
Pyruvate Dehydrogenase, DLAT E2 subunit (SEQ ID NO: 242, nucleotide, SEQ ID NO: 243, protein)
Acetate kinase 1 (ACK1) allele 1 (SEQ ID NO: 107, nucleotide, SEQ ID NO: 108, protein),
Acetate kinase 1 (ACK1) allele 2 (SEQ ID NO: 109, nucleotide, SEQ ID NO: 110, protein),

- Acetate kinase 2 (ACK2) (SEQ ID NO: 111, nucleotide, SEQ ID NO: 112, protein),
- Lactate dehydrogenase (SEQ ID NO: 117, nucleotide, SEQ ID NO: 118, protein),
- Phosphate acetyltransferase allele 1 (SEQ ID NO: 113, nucleotide, SEQ ID NO: 114, protein),
- 5 Phosphate acetyltransferase allele 2 (SEQ ID NO: 115, nucleotide, SEQ ID NO: 116, protein),
- Lactate dehydrogenase (SEQ ID NO: 117, nucleotide, SEQ ID NO: 118, protein)
- ACCase, Homomeric acetyl-CoA carboxylase, (SEQ ID NO: 262, nucleotide, SEQ ID NO: 263, protein)
- 10 Heteromeric acetyl-CoA carboxylase BC subunit allele 1 (SEQ ID NO: 104, nucleotide, SEQ ID NO: 93 protein),
- Heteromeric acetyl-CoA carboxylase BC subunit allele 2 (SEQ ID NO: 103, nucleotide, SEQ ID NO: 94 protein),
- Heteromeric acetyl-CoA carboxylase BCCP subunit allele 1 (SEQ ID NO: 101, nucleotide, SEQ ID NO: 95, protein),
- 15 Heteromeric acetyl-CoA carboxylase BCCP subunit allele 2 (SEQ ID NO: 102, nucleotide, SEQ ID NO: 96, protein),
- Acetyl-CoA carboxylase alpha-CT subunit allele 1 (SEQ ID NO: 92, nucleotide, SEQ ID NO: 97, protein),
- 20 Acetyl-CoA carboxylase alpha-CT subunit allele 2 (SEQ ID NO: 99, nucleotide, SEQ ID NO: 98, protein),
- Heteromeric acetyl-CoA carboxylase b-CT subunit, (SEQ ID NO: 222, nucleotide, SEQ ID NO: 223, protein)
- Plastidial ACP allele 2 (SEQ ID NO: 90, protein)
- 25 Plastidial Acyl-Carrier Protein (ACP) allele 1 (SEQ ID NO: 206, nucleotide, SEQ ID NO: 207, protein)
- Mitochondrial Acyl-Carrier Protein (ACP) allele 1 (SEQ ID NO: 64, nucleotide, SEQ ID NO: 65, protein)
- Mitochondrial Acyl-Carrier Protein (ACP) allele 2 (SEQ ID NO: 194, nucleotide, SEQ ID NO: 195, protein)
- 30

- Malonyl-CoA:ACP transacylase (MAT) (SEQ ID NO: 174, nucleotide, SEQ ID NO: 175 protein),
- Ketoacyl-ACP synthase I allele 1 (KASI allele 1, SEQ ID NO: 69, nucleotide, SEQ ID NO:70, protein),
- 5 Ketoacyl-ACP synthase I allele 2 (KASI allele 2, SEQ ID NO: 67, nucleotide, SEQ ID NO:68, protein),
- Ketoacyl-ACP synthase III (KASIII) (SEQ ID NO: 214, nucleotide, SEQ ID NO: 215, protein),
- Ketoacyl-ACP reductase (KAR), SEQ ID NO: 258, nucleotide, SEQ ID NO: 259, protein)
- 10 Ketoacyl-CoA reductase (KCR) (SEQ ID NO: 184, nucleotide, SEQ ID NO: 185, protein),
- 3-Hydroxyacyl-ACP dehydrase (HD) (SEQ ID NO: 208, nucleotide, SEQ ID NO: 209, protein),
- Enoyl-ACP reductase 1 version 1 (ENR1-1) (SEQ ID NO: 218 nucleotide, 220 protein)
- Enoyl-ACP reductase 1 version 2 (ENR1-1) (SEQ ID NO: 219 nucleotide, 221 protein)
- 15 Trans-2-enoyl-CoA reductase (SEQ ID NO: 196, nucleotide, SEQ ID NO: 197, protein),
- Stearoyl-ACP desaturase 1 allele 1 (SAD1 allele 1, SEQ ID NO: 87 nucleotide, SEQ ID NO:88, protein),
- Stearoyl-ACP desaturase 1 allele 2 (SAD1 allele 2, SEQ ID NO:86 nucleotide, SEQ ID NO:85, protein),
- 20 Fatty acyl-ACP thioesterase A (FATA) allele 1 (SEQ ID NO: 71, nucleotide, SEQ ID NO: 72, protein),
- Glycerol-3-phosphate acyltransferase (GPAT), (SEQ ID NO: 224, nucleotide, SEQ ID NO: 225, protein),
- Glycerol-3-phosphate acyltransferase (GPAT) (SEQ ID NO: 186, nucleotide, SEQ ID NO: 187, protein),
- 25 LPAAT-E, 1-Acyl-sn-glycerol-3-phosphate acyltransferase isoform E (SEQ ID NO:81 nucleotide, SEQ ID NO:82 protein),
- LPAAT-A, 1-Acyl-sn-glycerol-3-phosphate acyltransferase isoform A (SEQ ID NO:84 nucleotide, SEQ ID NO:83 protein),

Posphatidic acid phosphatase (PAP), (SEQ ID NO: 226, nucleotide, SEQ ID NO: 227, protein)

Long-chain acyl-CoA ligase (SEQ ID NO: 198, nucleotide, SEQ ID NO: 199, protein),

5 DGAT1-1, Acyl-CoA:Diacylglycerol Acyltransferase 1, Allele 1 (SEQ ID NO: 228, nucleotide, SEQ ID NO: 229, protein)

DGAT1-1, Acyl-CoA:Diacylglycerol Acyltransferase 1, Allele 2 (SEQ ID NO: 230, nucleotide, SEQ ID NO: 231, protein)

DGAT2, Acyl-CoA:Diacylglycerol Acyltransferase 2 (SEQ ID NO: 232, nucleotide, SEQ ID NO: 233, protein)

10 Diacylglycerol kinase (DGK) / Sphingosine Kinase (Spik) (SEQ ID NO: 256, nucleotide, SEQ ID NO: 257, protein)

Choline kinase (SEQ ID NO: 236, nucleotide, SEQ ID NO: 237, protein)

Leafy cotyledon2 (LEC2) (SEQ ID NO: 254, nucleotide, SEQ ID NO: 255, protein)

15 ACLA, ATP:Citrate Lyase, Subunit A (SEQ ID NO: 250, nucleotide, SEQ ID NO: 251, protein)

ACLB, ATP:Citrate Lyase, Subunit B (SEQ ID NO: 252, nucleotide, SEQ ID NO: 253 protein)

Malic Enzyme (SEQ ID NO: 248, nucleotide, SEQ ID NO: 249, protein)

20 ACBP1, Acyl-CoA Binding Protein 1 (SEQ ID NO: 244, nucleotide, SEQ ID NO: 245, protein)

ACBP2, Acyl-CoA Binding Protein 2 (SEQ ID NO: 246, nucleotide, SEQ ID NO: 247, protein)

Phosphatidate cytidyltransferase (SEQ ID NO: 176, nucleotide, SEQ ID NO: 177, protein),

25 Enoyl-CoA hydratase (SEQ ID NO: 192, nucleotide, SEQ ID NO: 193, protein),

Acyl-CoA oxidase (SEQ ID NO: 190, nucleotide, SEQ ID NO: 191, protein),

Lineolate FAD3 desaturase allele 1 (SEQ ID NO: 75, nucleotide, SEQ ID NO: 76, protein),

Lineolate FAD3 desaturase allele 2 (SEQ ID NO: 73, nucleotide, SEQ ID NO: 74, protein),

LS, lipote synthase, (SEQ ID NO: 260, nucleotide, SEQ ID NO: 261, protein)

5 Glyoxysomal fatty acid beta-oxidation multifunctional protein (SEQ ID NO: 166, nucleotide, SEQ ID NO: 167, protein),

Monoglyceride lipase (SEQ ID NO: 188, nucleotide, SEQ ID NO: 189, protein),

Triacylglycerol lipase (SEQ ID NO: 168, nucleotide, SEQ ID NO: 169, protein),

Glycerophosphodiester phosphodiesterase (SEQ ID NO: 180, nucleotide, SEQ ID NO: 181, protein),

10 Membrane bound O-acyl transferase domain-containing protein (SEQ ID NO: 202, nucleotide, SEQ ID NO: 203, protein),

Lipid droplet protein 1, LDP1 (SEQ ID NO: 119, nucleotide, SEQ ID NO: 120, protein),

Succinate semialdehyde dehydrogenase (SEQ ID NO: 172, nucleotide, SEQ ID NO: 173, protein),

15 Sterol 14 desaturase (SEQ ID NO: 212, nucleotide, SEQ ID NO: 213, protein)

Nitrogen Response Regulator (NRR1) (SEQ ID NO: 264, SEQ ID NO: 265)

Monoacylglycerol Acyltransferase (MGAT1) (LPLAT-MGAT-like acyltransferase, DAGAT-domain containing) (also known in the art as 2-acylglycerol O-acyltransferase) (SEQ ID NO: 266, nucleotide, SEQ ID NO: 267, protein)

20 Cellulase / Endoglucanase (EG1) (SEQ ID NO: 268, nucleotide, SEQ ID NO: 269, protein)

A. ACETYL-COA – MALONY-COA TO ACYL-ACP

[0097] The early stages of fatty acid synthesis involve the conversion of a fixed carbon
 25 (e.g., glucose, sucrose, etc.) or other carbon sources into pyruvate. Next, the pyruvate dehydrogenase complex (PDH), comprising pyruvate dehydrogenase, dihydrolipoyl transacetylase, and dihydrolipoyl dehydrogenase, converts the three carbon metabolite pyruvate into the two carbon metabolite acetyl-CoA. The acetyl-CoA carboxylase (ACC) complex, utilizing bicarbonate as a substrate, generates the 3-carbon compound malonyl-
 30 CoA. Malonyl-CoA:ACP acyltransferase (MAT) then catalyzes the transfer of a malonyl

group from malonyl-CoA to the acyl carrier protein (ACP), thereby generating malonyl-ACP. ACP is used as the acyl carrier for the various intermediate reactions in fatty acid biosynthesis. The metabolites acetyl-CoA and malonyl-CoA and the ACP protein are thus important starting points for fatty acid biosynthesis.

5 [0098] To genetically engineer a microbe for increased production of fatty acids and lipids, recombinant modifications may be made, either individually or in combination to obtain increased acetyl-CoA/malonyl-CoA/ACP production. For example, to increase malonyl-CoA production, an expression cassette may be generated and used to transform a microbe to overexpress polynucleotides encoding one or more components of the ACC
10 enzyme complex under the control of a constitutive or regulated promoter. Additional examples of enzymes suitable for up-regulation according to embodiments of the invention include enzymes of the pyruvate dehydrogenase complex (examples, some from microalgae, include GenBank Accession Numbers NP_415392; AAA53047; Q1XDM1; and CAF05587). Up-regulation of pyruvate dehydrogenase can increase production of
15 acetyl-CoA, and thereby increase fatty acid synthesis.

[0099] The acetyl-CoA carboxylase complex catalyzes the initial step in fatty acid synthesis. Accordingly, one or more enzymes comprising this complex may be up-regulated to increase production of fatty acids (examples, some from microalgae, include GenBank accession numbers BAA94752; AAA75528; AAA81471; YP_537052;
20 YP_536879; NP_045833; and BAA57908). Enzymes of the ACCase complex may include the heteromeric ACCase BC subunit 1, the heteromeric ACCase BCC subunit, the heteromeric ACCase a-CT subunit, and the heteromeric ACCase b-CT subunit 1.

[0100] The ACC exists in two forms, a cytosolic, homomeric (single subunit) form and a heteromeric (multi-subunit) form in the plastid. Both types of ACC are found in
25 *Prototheca* (UTEX 1435 strain). A plastid transit peptide was identified in the homomeric ACCase, indicating it may be localized to both the plastid and cytoplasm. In some embodiments, provided are sequences and related compositions and methods for expression of one or more a plastid-targeted homomeric ACC subunit(s). Plastid targeting can be achieved by attaching a leader plastid transit peptide selected from any of those
30 provided herein, using the endogenous plastid transit peptide on the gene of interest itself, or obtained from any endogenous plastid-targeted gene in *Prototheca* containing a plastid transit peptide, or by using a heterologous plastid transit peptide. In other embodiments, provided are sequences and related compositions and methods for overexpression of one

or more plastid-targeted, heteromeric ACC subunit(s). In some embodiments, the ACC overexpression decreases the ratio of C18:2 and C18:3 fatty acids by at least 5%, 10%, 15%, 20%, 30%, 40%, or 50%.

[0101] In some embodiments, ACC and one or more downstream lipid biosynthesis genes are simultaneously overexpressed. In one embodiment, provided are sequences, compositions, and methods for the simultaneous overexpression of ACC and DGAT. Starch levels, generally low in *Prototheca*, can be further reduced in favor of lipid biosynthesis by overexpressing ACC. In some embodiments, starch production is attenuated in concert with overexpression of one or more downstream lipid biosynthesis genes.

[0102] In other embodiments, provided are sequences, compositions, and methods for fatty acid production by up-regulation of polynucleotides encoding acyl carrier protein (ACP), which carries the growing acyl chains during fatty acid synthesis. Examples of ACPs, some from microalgae, include GenBank accession numbers A0T0F8; P51280; NP_849041; YP_874433 as well as SEQ ID NO: 207. Acyl carrier proteins can differ in subcellular localization (SEQ ID NOs: 63-65, SEQ ID NO: 195, and SEQ ID NO: 207). The sequences listed in SEQ ID NO: 194 and SEQ ID NO: 206 provide the coding sequences of the *Prototheca moriformis* amino acid sequence given in SEQ ID NO: 195 and SEQ ID NO: 207, respectively. In one embodiment of the present invention, it is advantageous to overexpress a first exogenous ACP and concomitantly down-regulate a different ACP.

[0103] Recombinant modifications may be made to increase the production of other intermediates in the lipid biosynthesis pathway. For example, to increase malonyl-ACP production, an expression cassette may be generated and used to transform a microbe to overexpress polynucleotides given in SEQ ID NO: 174 encoding *P. moriformis* malonyl-CoA:ACP transacylase (MAT) (SEQ ID NO: 175) active to transfer a malonyl group from malonyl-CoA to acyl carrier protein (ACP). This expression cassette may comprise a constitutive or inducible promoter active to drive expression of MAT.

[0104] In some embodiments, provide are sequences and methods for down regulation of Pyruvate Dehydrogenase Kinase (PDHK), a negative regulator of pyruvate dehydrogenase complex.

[0105] Enzymes that deplete pools of pyruvate or acetyl-CoA for the synthesis of metabolites other than fatty acids may compete with lipid biosynthesis pathway enzymes

for precursor metabolites. Attenuation of these competitor enzymes may increase the production of fatty acids or lipids. To genetically engineer a microbe for increased production of fatty acids and lipids, recombinant modifications may be made, either individually or in combination to attenuate enzymes that compete for metabolite

precursors. For example, to decrease the use of acetyl-CoA for acetate production, an expression cassette may be generated to ablate the gene or genes encoding *P. moriformis* (UTEX 1435) acetate kinase enzymes (SEQ ID NOs: 107-112). Attenuation of acetate kinase can also be achieved through the construction and use of expression cassettes comprising an antisense RNA under the control of a constitutive or regulated promoter.

Additional examples of enzymes suitable for down-regulation according to embodiments of the present invention include *P. moriformis* (UTEX 1435) lactate dehydrogenase (SEQ ID NO: 117, nucleotide, SEQ ID NO: 118, protein), which synthesizes lactate from pyruvate or *P. moriformis* (UTEX 1435) phosphate acetyltransferase (PTA, SEQ ID NOs: 113-116), which catalyzes the conversion of acetyl-CoA to acetylphosphate, a step in the metabolism of acetate.

B. ACYL-ACP TO FATTY ACID

[0106] A growing acyl-ACP chain is elongated in 2-carbon increments through a set of four enzymatic reactions involving condensation, a first reduction reaction, dehydration, and a second reduction reaction. These reactions are catalyzed by a condensing enzyme (β -ketoacyl-ACP synthase, KAS), a first reductase enzyme (β -ketoacyl-ACP reductase, KAR), a dehydrase (β -hydroxyacyl-ACP dehydrase, HR) and a second reductase (enoyl-ACP reductase, ENR).

[0107] The initial condensation reaction between malonyl-ACP and acetyl-CoA to produce a 4-carbon compound is catalyzed by β -ketoacyl-ACP synthase (KAS) III.

Successive 2-carbon additions to the elongating acyl-ACP chain, through C16:0, are catalyzed by KAS I. The enzyme KASII performs a 2-carbon extension of C16:0-ACP to C18:0-ACP. Depending on the desired properties of fatty acids to be produced, one or more KAS enzymes may be attenuated or over-expressed in the microbe. For example, to engineer a microbe for increased production of fatty acids with 18 or more carbon atoms, an expression cassette may be generated and used to transform a microbe to overexpress polynucleotides encoding a KASII enzyme under the control of a constitutive or inducible promoter. Examples of KASII enzymes suitable for use with embodiments of the present invention are the *Prototheca moriformis* KASII enzymes provided in SEQ ID NO: 106 (see Example 6) and SEQ ID NO: 211. The protein coding sequences for these enzymes

are provided in SEQ ID NO: 105 and SEQ ID NO: 210. To genetically engineer a microbe for the increased production of short or mid chain fatty acids, an expression cassette may be generated and used to transform microbes to decrease the expression of KASII, such as through targeted knockdown or the use of inhibitory RNA. Optionally, the knockout or
5 knockdown of KASII may be combined with use of polynucleotides to transform a microbe to overexpress polynucleotides encoding either a KASI or KASIII enzyme. An example of a KASIII enzyme suitable for use with embodiments of the present invention is the *P. moriformis* KASIII enzyme (provided in SEQ ID NO: 215), encoded by the polynucleotide sequence given in SEQ ID NO: 214. Examples of KASI enzymes suitable
10 for use with embodiments of the present invention are the *P. moriformis* KASI enzymes alleles 1 and 2 provided in SEQ ID NO: 70 and 68, respectively. The protein coding sequence for these enzymes are provided in SEQ ID NO: 69 and SEQ ID NO: 67 respectively.

[0108] In an additional embodiment, a recombinant cell is engineered to overexpress a
15 KASI and/or KASIII enzyme in a genetic background wherein KASII enzyme activity has not been attenuated. In still a further embodiment, recombinant polynucleotides are generated and used to transform a microbe to ablate or down-regulate a KASI enzyme to increase the production of fatty acids with greater than 16 carbons. Optionally, the knockout or knockdown of KASI may be combined with the use of polynucleotides to
20 transform a microbe to overexpress polynucleotides encoding KASII.

[0109] To genetically engineer a microbe for increased production of specific fatty acids and lipids, recombinant modifications may be made, either individually or in combination to attenuate or overexpress enzymes that participate in fatty acid elongation. Such enzymes include β -ketoacyl-ACP reductase (KAR), β -hydroxyacyl-ACP dehydrase
25 (HD), and enoyl-ACP reductase (ER). For example, an expression cassette may be generated and used to transform a microbe to overexpress polynucleotides encoding a KAR enzyme, a β -hydroxyacyl-ACP dehydrase, or an enoyl-ACP reductase under the control of a constitutive or inducible promoter. Alternatively, an expression cassette may be generated and used to transform a microbe to overexpress polynucleotides operable to
30 knockout or attenuate the expression of a KAR enzyme, an HD enzyme, or an ER enzyme. An example of a KAR enzyme suitable for use with embodiments of the present invention is the *P. moriformis* KAR enzyme (provided in SEQ ID NO: 185), encoded by the polynucleotide sequence given in SEQ ID NO: 184. An example of an HD enzyme suitable for use with embodiments of the present invention is the *P. moriformis* HD

enzyme (provided in SEQ ID NO: 209), encoded by the polynucleotide sequence given in SEQ ID NO: 208. Examples of ER enzymes suitable for use with embodiments of the present invention are *P. moriformis* ER (provided in SEQ ID NO: 201), encoded by the polynucleotide sequence given in SEQ ID NO: 200 and *P. moriformis* (UTEX 1435) trans-
5 2-enoyl-CoA reductase (SEQ ID NO: 197, encoded by the polynucleotide sequence given in SEQ ID NO: 196). According to embodiments of the present invention, it may be advantageous to up-regulate or down-regulate a KAR, HD, or ER enzyme under specific culture conditions, for example during lipid production and/or in a genetic background of a microbe that has been engineered to alter additional lipid biosynthesis genes or gene
10 products.

[0110] Fatty acyl-ACP thioesterase (TE) enzymes terminate elongation by hydrolyzing the acyl-ACP into free fatty acids and ACP. TEs may show specificity for acyl-ACPs of certain carbon lengths and degree of saturation or may be broad TEs, able to cleave acyl-ACP chains of varying length and level of saturation. The substrate specificity of TEs is
15 an important contributor to establishing the chain length and degree of saturation of fatty acids. Depending on the desired length or degree of saturation of the fatty acid to be produced, one or more genes encoding acyl-ACP thioesterases may be attenuated or over-expressed in the microbe. For example, an endogenous fatty acyl-ACP thioesterase gene showing preference for C18:1-ACP (may be knocked out or reduced in expression while
20 concomitantly a different TE, showing specificity for saturated C12 and C14-ACPs is overexpressed in the microbe, thereby altering the population of fatty acids in the microbe. An example of an acyl-ACP thioesterase suitable for use in embodiments of the present invention includes the *P. moriformis* acyl-ACP thioesterase FATA1 (provided in SEQ ID NO: 72), encoded by the polynucleotide sequence given in SEQ ID NO: 71. Example 5,
25 Example 34, and Example 36 describe the attenuation of a *P. moriformis* acyl-ACP thioesterase.

C. UNSATURATED FATTY ACIDS AND FATTY ACYL CHAINS

[0111] The introduction of carbon-carbon double bonds into a fatty acid, fatty acyl-CoA, or fatty acyl-ACP chains relies on the activity of desaturases. Desaturase enzymes
30 may show specificity for the carbon chain length and degree of saturation of their substrates. Specific desaturases can convert saturated fatty acids or saturated fatty acyl-ACPs to unsaturated fatty acids or unsaturated fatty acyl-ACPs. Other desaturase enzymes may increase the number of carbon-carbon double bonds of unsaturated fatty acids.

[0112] Stearoyl-ACP desaturase (see, e.g., GenBank Accession numbers AAF15308; ABM45911; AAY86086, and SEQ ID Nos: 59-62), for example, catalyzes the conversion of stearoyl-ACP to oleoyl-ACP. Up-regulation of this gene can increase the proportion of monounsaturated fatty acids produced by a cell; whereas down-regulation can reduce the proportion of monounsaturates. For illustrative purposes, SADs are responsible for the synthesis of C18:1 fatty acids from C18:0 precursors.

[0113] Additional desaturases are the fatty acyl desaturases (FADs), including the phosphatidylglycerol desaturase (FAD4), the plastidial oleate desaturase (FAD6), the plastidial linoleate desaturase (FAD7/FAD8), endoplasmic reticulum oleate desaturase (FAD2), the endoplasmic reticulum linoleate desaturase (FAD3), the delta 12 fatty acid desaturase ($\Delta 12$ FAD) and the delta 15 fatty acid desaturase ($\Delta 15$ FAD). These desaturases also provide modifications with respect to lipid saturation. For illustrative purposes, $\Delta 12$ fatty acid desaturases are responsible for the synthesis of C18:2 fatty acids from C18:1 precursors and $\Delta 15$ fatty acid desaturases are responsible for the synthesis of C18:3 fatty acids from C18:2 precursors.

[0114] Still additional desaturases, including the palmitate-specific monogalactosyldiacylglycerol desaturase (FAD5), the linoleoyl desaturase, ω -6 fatty acid desaturases, ω -3 fatty acid desaturases, and ω -6-oleate desaturases, provide modifications with respect to lipid saturation. The expression of one or more desaturases, such as ω -6 fatty acid desaturase, ω -3 fatty acid desaturase, or ω -6-oleate desaturase, may be controlled to alter the ratio of unsaturated to saturated fatty acids.

[0115] We have found *Prototheca* to have an extremely compact genome, with very compact promoter regions and extensive use of bi-directional promoters. For example, *Prototheca* has a single copy of FAD2 and FAD3 which are capable of multiple localization. FAD3 encodes two alternate start positions, one of which encodes a plastid transit peptide and the other of which does not. In addition FAD3 contains an ER retention signal in its 3' UTR which is shifted away from the COOH carboxy terminal in the plastid-targeted form, and shifted back in the ER-bound form. This allows the same gene to go to different sub-cellular locations depending on the needs of the organism.

[0116] Acyl-ACPs synthesized in the plastid are either used directly within that organelle to form lipids, including glycerolipids, or exported outside the plastid for synthesis of lipids including phospholipids, triacylglycerol, or waxes. Lipid biosynthesis genes may show specificity for activity in specific subcellular locations.

D. FATTY ACID TO FATTY ACYL-COA

[0117] Upon export from the plastid, fatty acids are re-esterified to CoA to form acyl-CoA via the catalytic action of acyl-CoA synthetase. Different acyl-CoA synthetase enzymes can differ in subcellular localization and show specificity for fatty acids of differing chain length. An example of an acyl-CoA synthetase is the enzyme long chain fatty acyl-CoA synthetase. Depending on the desired properties of the triacylglycerol to be produced, one or more genes encoding an acyl-CoA synthetase may be attenuated or over-expressed in the microbe.

E. FATTY ACYL-COA TO TRIACYLGLYCEROL

[0118] Triacylglycerides may be formed through three sequential acyl-CoA-dependent acylations of a *sn*-glycerol-3-phosphate molecule. The first acylation, the rate-limiting step of glycerolipid synthesis, is catalyzed by glycerol-3-phosphate acyltransferase (GPAT) to produce lyso-phosphatidic acid. The second acylation step is catalyzed by the enzyme acyl-CoA:lyso-phosphatidic acid acyltransferase (LPAAT).

Prior to the third acylation step, the enzyme phosphatidic acid phosphatase (PAP) (or lipins) carries out the removal of the phosphate group from phosphatidic acid to generate *sn*-1,2-diacylglycerol (DAG). The final acyl-CoA-dependent acylation is catalyzed by acyl-CoA:diacylglycerol acyltransferase (DGAT).

[0119] Microbes may be genetically engineered for increased production of lipids. For example, to increase the production of TAGs, an expression cassette may be generated and used to transform a microbe to polynucleotides operable to increase the expression of GPAT. This expression cassette may comprise a constitutive or inducible promoter active to drive expression of GPAT and may be utilized in the genetic background of a strain in which endogenous GPAT activity has been attenuated. An example of a GPAT suitable for use in an embodiment of the present invention is the *P. moriformis* GPAT, given here as SEQ ID NO: 187, encoded by the sequences given here in SEQ ID NO: 186.

[0120] Microbes may be genetically engineered for increased production of triacylglycerol molecules with desired properties. Certain acyltransferase enzymes, including GPATs, LPAATs, and DGATs may demonstrate specificity for a subcellular localization or substrate specificity for the length and degree of saturation of the acyl-CoA chain they transfer to the substituted glycerol backbone. Additionally, LPAAT and DGAT enzymes may show substrate specificity for the form of substituted glycerol to which they transfer an acyl-CoA. Depending on the desired properties of the triacylglyceride to be

produced, one or more genes encoding GPATs, LPAATs, DGATs, or other acyltransferases may be attenuated or over-expressed in the microbe. For example, to increase the production of TAGs with midchain fatty acids esterified at the sn-2 position, an expression cassette may be generated and used to transform a microbe to overexpress an LPAAT
5 having specificity for transferring midchains. This expression cassette may comprise a constitutive or inducible promoter active to drive expression of LPAAT and may be utilized in the genetic background of a strain in which endogenous LPAAT activity has been attenuated. Examples of LPAATs suitable for use in embodiments of the present invention include the *P. moriformis* LPAAT E and LPAAT A, given in SEQ ID NO: 82
10 and SEQ ID NO: 83, encoded by the sequences given in SEQ ID NO: 81 and SEQ ID NO: 84, respectively.

[0121] In a similar fashion, to increase production of TAGs, an expression cassette may be generated and used to transform a microbe to overexpress polynucleotides encoding a DGAT, active to transfer a acyl-CoA to a DAG molecule. This expression
15 cassette may comprise a constitutive or inducible promoter active to drive expression of DGAT2. An example of a DGAT suitable for use in the present invention is *P. moriformis* diacylglycerol acyltransferase type 2 (DGAT2) (SEQ ID NO: 183, encoded by the sequence given in SEQ ID NO: 182). According to the desired characteristics of the fatty acids or lipids to be produced by the recombinant microbe, it may be advantageous to
20 couple up-regulation of a TE characterized by substrate specificity with one or more GPAT, LPAAT, or DGAT enzymes showing the same substrate specificity.

[0122] Additional acyltransferases suitable for use in embodiments of the present invention include *P. moriformis* membrane bound O-acyl transferase domain-containing protein (SEQ ID NO: 203, encoded by the nucleotide sequence provided in SEQ ID NO:
25 202), *P. moriformis* putative 1-acyl-sn-glycerol-3-phosphate acyltransferase (SEQ ID NO: 179, encoded by the nucleotide sequence provided in SEQ ID NO: 178), and *P. moriformis* acyl transferase (SEQ ID NO: 217, encoded by the nucleotide sequence provided in SEQ ID NO: 216). The Monoacylglycerol Acyltransferase (MGAT) gene, catalyzed the synthesis of diacylglycerol, and can generally also catalyze the final step in
30 triacylglycerol biosynthesis. Hence, upregulation of the MGAT gene provided in this invention may be desirable.

[0123] Alternate lipid pathway enzymes can generate triacylglyceride molecules through a route separate from that above. Enzymes of the fatty acyl-CoA-independent

triacylglycerol pathway transfer fatty acyl groups between phosphatidylcholine (PC) moieties employing acyl-lysophosphatidylcholine acyl transferases that may exhibit selective substrate specificity, ultimately transferring them to diacylglycerol.

F. ADDITIONAL LIPID MOLECULES

5 [0124] In addition to their incorporation into DAGs and TAGs, fatty acids or fatty acyl molecules may be incorporated into a range of lipid molecules including but not limited to phospholipids, phosphatidylcholine (PC), phosphatidylserine (PS), phosphatidylinositol (PI), sphingolipids (SL), monogalactosyldiacylglycerol, digalactosyldiacylglycerol, liponucleotides, and wax esters. Enzymes that synthesize molecules of PC, PS, PI, SL, 10 wax esters, liponucleotides, or the galactolipids monogalactosyldiacylglycerol (MGDG) or digalactosyldiacylglycerol (DGDG) may compete with enzymes that lead to or ultimately synthesize DAGs and TAGs for substrates including fatty acids or fatty acyl molecules. Genes encoding proteins involved in the synthesis, utilization, or degradation of PC, PS, PI, SL, monogalactosyldiacylglycerol, digalactosyldiacylglycerol, or wax esters may 15 include diacylglycerol cholinephosphotransferase (DAG-CPT), cytidine diphosphate diacylglycerol synthase (CTP-DAG synthase), phosphatidylinositol synthase (PI synthase), choline kinase (CK), phosphatidylinositol-3-kinase (PI3-Kinase), phosphatidylinositol-4-kinase (PI4-Kinase), diacylglycerol kinase (DGK), phosphatidylglycerol-3-phosphate phosphatase (PGPP), cholinephosphate 20 cytidyltransferase (CPCT), phosphatidate cytidyltransferase, phosphatidylserine decarboxylase (PSD), phospholipase C (PliC), phospholipase D (PliD), sphingolipid desaturase (SD), monogalactosyldiacylglycerol synthase (MGDG synthase), digalactosyldiacylglycerol synthase (DGDG synthase), ketoacyl-CoA synthase (KCS), 3-ketoacyl reductase (KR), and wax synthase (WS). Depending on the desired properties of 25 the fatty acids or lipid molecules to be produced, one or more genes encoding enzymes that utilize fatty acids or fatty acyl molecules as substrates to produce lipid molecules may be attenuated or over-expressed in the microbe.

[0125] In one embodiment, provided are sequences, compositions, and methods for inhibition of DGK, which converts DAG to PA. For example, DGK can be inhibited 30 through use of RNAi, hairpin constructs, or double or single knockouts. In other embodiments, provided are sequences, compositions, and methods for overexpression of epsilon subtype (DGKe). In some aspects, overexpression of DGKe results in selective removal of DAGs with certain acyl groups such as C20:4.

[0126] To engineer a microbe for the increased production of triglycerides, it may be advantageous to attenuate enzymes that support phospholipid synthesis. For example, to decrease production of the phospholipid cytidine diphosphate (CDP)-diacylglycerol, an expression cassette may be generated and used to transform a microbe to attenuate
5 phosphatidate cytidylyltransferase, which catalyzes condensation of phosphatidic acid and cytidine triphosphate to produce CDP-diacylglycerol. An example of a phosphatidate cytidylyltransferase suitable for use in an embodiment of the present invention is *P. moriformis* phosphatidate cytidylyltransferase (SEQ ID NO: 177), encoded by the polynucleotide sequence given by SEQ ID NO: 176. This expression cassette may
10 comprise a constitutive or inducible promoter active to drive down-regulate expression of the phosphatidate cytidylyltransferase.

[0127] Further, additional lipid moieties other than triacylglycerides may utilize derivations of phosphorylated glycerol as a backbone. Enzymes such as phosphatidylglycerophosphate synthase (PGP Synthase), involved in the synthesis of
15 phospholipids may compete with enzymes that provide for triacylglycerols for substrates including phosphorylated forms of glycerol. Depending on the desired properties of the lipid molecule to be produced, one or more genes encoding phosphatidylglycerophosphate synthase may be attenuated or over-expressed in the microbe. Lipoate Synthase (LS), also called Lipoyl Synthase or Lipoic Acid Synthase, is generally localized to the mitochondria
20 and utilized in the synthesis of lipoic acid. Lipoic acid is an important co-factor and antioxidant.

G. FATTY ACID DEGRADATION

[0128] To genetically engineer a microbe for increased production of specific fatty acids and lipids, recombinant modifications may be made, either individually or in
25 combination, to decrease the degradation of fatty acids and lipids. As proteins such as acyl-CoA oxidase, 3-ketoacyl-CoA thiolase, acyl-CoA dehydrogenase, glyoxysomal fatty acid beta-oxidation multifunctional protein, and enoyl-CoA hydratase are involved in the degradation of fatty acids, these and other proteins may be attenuated in the microbe to slow or prevent fatty acid degradation. For example, to engineer a microbe to decrease
30 fatty acid degradation, an expression cassette may be generated and used to transform a microbe to down-regulate one or more of acyl-CoA oxidase, enoyl-CoA hydratase, and glyoxysomal fatty acid beta-oxidation multifunctional protein, either through a knockout or knockdown approach. An example of an acyl-CoA oxidase enzyme suitable for use with embodiments of the present invention is the *Prototheca moriformis* acyl-CoA oxidase

(provided in SEQ ID NO: 191), encoded by the polynucleotide sequence given in SEQ ID NO: 190). An example of an enoyl-CoA hydratase suitable for use with embodiments of the present invention is the *Prototheca moriformis* enoyl-CoA hydratase (provided in SEQ ID NO: 193), encoded by the polynucleotide sequence given in SEQ ID NO: 192. An example of a glyoxysomal fatty acid beta-oxidation multifunctional protein suitable for use in embodiments of the present invention is the *P. moriformis* (UTEX 1435) glyoxysomal fatty acid beta-oxidation multifunctional protein, presented in SEQ ID NO: 167, encoded by the polynucleotide sequence given in SEQ ID NO: 166. According to the desired chain length and degree of saturation of the fatty acids to be produced by the recombinant microbe, it may be advantageous to down-regulate fatty acid or lipid degradation enzymes in the genetic background of a microbe that has been engineered to alter additional lipid pathway genes or gene products.

[0129] Long-chain acyl-CoA synthetases (also known in the art as long-chain acyl-CoA ligases) convert free fatty acids into acyl-CoA thioesters. These acyl-CoA thioesters may then be degraded by enzymes involved in fatty β -oxidation. To engineer a microbe for decreased fatty acid degradation, an expression cassette may be generated and used to transform a microbe to down-regulate long-chain acyl-CoA synthetase, either through a knockout or knockdown approach. An example of long-chain acyl-CoA synthetase suitable for use in an embodiment of the present invention is the *P. moriformis* (UTEX 1435), long-chain acyl-CoA synthetase presented in SEQ ID NO: 199, encoded by the polynucleotide sequence given in SEQ ID NO: 198. According to the desired chain length and degree of saturation of the fatty acids to be produced by the recombinant microbe, it may be advantageous to down-regulate long-chain acyl-CoA synthetase in the genetic background of a microbe that has been engineered to alter additional lipid pathway genes or gene products.

H. MONOGLYCERIDE, TRIGLYCERIDE, AND LIPID DEGRADATION

[0130] A strategy to increase the recombinant microbial production of triglycerides is to prevent or reduce the enzymatic degradation of these molecules. Enzymes such as monoglyceride lipase and triacylglycerol lipase that hydrolyze triglycerides to fatty acids and glycerol are examples of proteins that may be attenuated in a microbe to slow or prevent degradation of triglycerides. For example, to engineer a microbe to decrease triglyceride degradation an expression cassette may be generated and used to transform a microbe to down-regulate monoglyceride lipase or triacylglycerol lipase, either through a knockout or knockdown approach. An example of a monoglyceride lipase suitable for use

in embodiments of the present invention is the *P. moriformis* (UTEX 1435) monoglyceride lipase presented in SEQ ID NO: 189, encoded by the polynucleotide sequence given in SEQ ID NO: 188. An example of a triacylglycerol lipase suitable for use in embodiments of the present invention is the *P. moriformis* (UTEX 1435) triacylglycerol lipase presented in SEQ ID NO: 169, encoded by the polynucleotide sequence given in SEQ ID NO: 168. According to embodiments of the present invention, it may be advantageous to attenuate one or more lipases under specific culture conditions, for example during lipid production.

I. GLOBAL REGULATORS

[0131] Furthermore, up- and/or down-regulation of genes may be applied to global regulators controlling the expression of the genes of the lipid biosynthetic pathway. Accordingly, one or more global regulators of lipid synthesis may be up- or down-regulated, as appropriate, to inhibit or enhance, respectively, the expression of a plurality of fatty acid synthetic genes and, ultimately, to increase lipid production. Examples include sterol regulatory element binding proteins (SREBPs), such as SREBP-1a and SREBP-1c (for examples see GenBank accession numbers NP_035610 and Q9WTN3). In one embodiment, a global regulator such as the endogenous LEC2 homolog, may be upregulated to increase lipid production. Decoupling or alteration of nitrogen sensing from the process of lipid biosynthesis may also be of value (Boyle et al, J. Biol. Chem, May 4, 2012). Also presented in this invention is a Nitrogen Response Regulator, NRR1, a Squamosa Binding protein. In some instance it may be desirable, for example, to increase the response to nitrogen starvation by enhancing expression of NRR1.

J. LIPID DROPLET PROTEINS

[0132] Eukaryotic cells store triacylglycerol molecules in distinct organelles, often called lipid droplets. Proteins associated with lipid droplet proteins, such as lipid droplet protein 1 (LDP1, SEQ ID NOs: 119-120), are crucial to lipid droplet structure, formation, size, and number. In some instances, attenuation of lipid droplet proteins results in increases in lipid droplet size. In other instances, overexpression of mutated sequences of lipid droplet proteins results in increased lipid droplet size and number. To genetically engineer a microbe for the production of fatty acids and lipids, recombinant modifications may be made, either individually or in combination to alter the expression of lipid droplet proteins. For example, an expression cassette may be generated to attenuate or ablate the gene or gene products encoding lipid droplet proteins (SEQ ID NOs: 119-120). Attenuation of gene products through the use of RNAi, RNA hairpin, or other antisense-mediated strategy may be coupled to an inducible or constitutive promoter. In an

additional embodiment, an expression cassette may be generated to overexpress one or more lipid droplet proteins. Overexpression of lipid droplet proteins may be driven by constitutive or inducible promoters. *P. moriformis* (UTEX 1435) does not contain oleosins, the plant equivalent of algal Lipid Droplet Proteins (LDPs). Instead, algal Lipid Droplet Proteins, such as the LDP1 presented here, are suitable for overexpression to alter lipid droplet morphology.

K. ALTERING CARBON METABOLISM

[0133] Numerous enzymatic pathways are involved in metabolizing sugars and metabolites into intermediates suitable for use in fatty acid or lipid synthesis or for other cellular pathways. In one embodiment of the present invention, it is advantageous to alter the regulation or activity of enzymes that contribute to production of metabolites involved in lipid synthesis or that utilize the intermediates or metabolites of lipid synthesis for pathways other than the fatty acid and lipid pathways. The Krebs's cycle is such a metabolic pathway that consumes acetyl-CoA to ultimately produce carbon dioxide.

Enzymatic participants of the Krebs's Cycle include fumarate hydratase (also known in the art as fumarase). To engineer a microbe for the increased production of specific fatty acids or lipids, an expression cassette may be generated and used to transform a microbe to attenuate fumarate hydratase, either through a knockout or knockdown approach. An example of a fumarate hydratase suitable for use in embodiments of the present invention is the *P. moriformis* (UTEX 1435) fumarate hydratase presented in SEQ ID NO: 171, encoded by the polynucleotide sequence given in SEQ ID NO: 170. According to embodiments of the present invention, it may be advantageous to attenuate fumarate hydratase under specific culture conditions, for example during lipid production.

[0134] An additional example of an enzyme involved in carbon metabolism is NAD-dependent glycerol-3-phosphate dehydrogenase that reversibly converts sn-glycerol 3-phosphate to dihydroxyacetone phosphate, (also known in the art as glycero phosphate). To increase the level of the triacylglycerol backbone precursor molecule, the sn-glycerol 3-phosphate metabolite, an expression cassette may be generated and used to transform a microbe to enhance expression of NAD-dependent glycerol-3-phosphate dehydrogenase. An example of an NAD-dependent glycerol-3-phosphate dehydrogenase suitable for use in embodiments of the present invention is the *P. moriformis* (UTEX 1435) NAD-dependent glycerol-3-phosphate dehydrogenase presented in SEQ ID NO: 205, encoded by the polynucleotide sequence given in SEQ ID NO: 204. According to embodiments of the present invention, it may be advantageous to attenuate NAD-

dependent glycerol-3-phosphate dehydrogenase under specific culture conditions, for example during lipid production. In some embodiments, it may be advantageous to combine the expression of several pathway enzymes for triacylglycerol production, for example the PAP, G3PDH, GPAT, LPPAT, and DGAT combination.

5 [0135] Other proteins, such as glycerophosphodiester phosphodiesterase, synthesize intermediates of the lipid pathway. Glycerophosphodiester phosphodiesterase hydrolyses a glycerophosphodiester to form sn-glycerol 3-phosphate, which may be used in lipid synthesis. To engineer a microbe for increased production of lipids, an expression cassette may be generated and used to transform a microbe to overexpress polynucleotides
10 encoding glycerophosphodiester phosphodiesterase. An example of a glycerophosphodiester phosphodiesterase suitable for use in embodiments of the present invention is *P. moriformis* (UTEX 1435) glycerophosphodiester phosphodiesterase (SEQ ID NO: 181), encoded by the the polynucleotide sequence given in SEQ ID NO: 180. Cellulases, such as endoglucanase, are useful for breaking down cellulosic compounds into
15 sugar utilizable by the cell. Provided is an endogenous endoglucanase that may potentially find uses in secretion to break down cellulosic material, or as a mechanism for loosening cell walls to allow for enhanced lipid droplet formation.

SECTION III. CULTIVATION

20 [0136] In certain embodiments, the present invention generally relates to cultivation of microbes, e.g., oleaginous microbes, such as microalgae, including *Chlorella* and *Prototheca* species and strains, and yeast, fungi, plant, and bacteria species and strains, for the production of microbial oil (lipids). In particular embodiments, the microbes are recombinant microbes. The following discussion focuses on *Prototheca* as an illustrative
25 species. For the convenience of the reader, this section is subdivided into subsections. Subsection 1 describes *Prototheca* species and strains and how to identify new *Prototheca* species and strains and related microalgae by genomic DNA comparison. Subsection 2 describes bioreactors useful for cultivation. Subsection 3 describes media for cultivation. Subsection 4 describes oil production in accordance with illustrative cultivation methods
30 of the invention.

1. PROTOTHECA SPECIES AND STRAINS

[0137] *Prototheca* is a remarkable microorganism for use in the production of lipid, because it can produce high levels of lipid, particularly lipid suitable for fuel production. The lipid produced by *Prototheca* has hydrocarbon chains of shorter chain length and a

higher degree of saturation than that produced by other microalgae. Moreover, *Prototheca* lipid is generally free of pigment (low to undetectable levels of chlorophyll and certain carotenoids) and in any event contains much less pigment than lipid from other microalgae. Moreover, recombinant *Prototheca* cells provided by the invention may be used to produce lipid in greater yield and efficiency, and with reduced cost, relative to the production of lipid from other microorganisms. Illustrative *Prototheca* strains for use in the methods of the invention include *Prototheca wickerhamii*, *Prototheca stagnora* (including UTEX 327), *Prototheca portoricensis*, *Prototheca moriformis* (including UTEX strains 1441, 1435), and *Prototheca zopfii*. In addition, this microalgae grows heterotrophically and may be genetically engineered. Species of the genus *Prototheca* are obligate heterotrophs.

[0138] Species of *Prototheca* for use in the invention may be identified by amplification of certain target regions of the genome. For example, identification of a specific *Prototheca* species or strain may be achieved through amplification and sequencing of nuclear and/or chloroplast DNA using primers and methodology using any region of the genome, for example using the methods described in Wu *et al.*, *Bot. Bull. Acad. Sin.* (2001) 42:115-121 Identification of *Chlorella spp.* isolates using ribosomal DNA sequences. Well established methods of phylogenetic analysis, such as amplification and sequencing of ribosomal internal transcribed spacer (ITS1 and ITS2 rDNA), 23S rRNA, 18S rRNA, and other conserved genomic regions may be used by those skilled in the art to identify species of not only *Prototheca*, but other hydrocarbon and lipid producing organisms with similar lipid profiles and production capability. For examples of methods of identification and classification of algae also see for example *Genetics*, 2005 Aug;170(4):1601-10 and *RNA*, 2005 Apr;11(4):361-4.

[0139] Thus, genomic DNA comparison may be used to identify suitable species of microalgae to be used in the present invention. Regions of conserved genomic DNA, such as but not limited to DNA encoding for 23S rRNA, may be amplified from microalgal species and compared to consensus sequences in order to screen for microalgal species that are taxonomically related to the preferred microalgae used in the present invention. Examples of such DNA sequence comparison for species within the *Prototheca* genus are shown below. Genomic DNA comparison can also be useful to identify microalgal species that have been misidentified in a strain collection. Often a strain collection will identify species of microalgae based on phenotypic and morphological characteristics. The use of these characteristics may lead to miscategorization of the species or the genus of a

microalgae. The use of genomic DNA comparison may be a better method of categorizing microalgae species based on their phylogenetic relationship.

[0140] Microalgae for use in the present invention typically have genomic DNA sequences encoding for 23S rRNA that have at least 99%, least 95%, at least 90%, or at
5 least 85% nucleotide identity to at least one of the sequences listed in SEQ ID NOs: 1-9.

[0141] For sequence comparison to determine percent nucleotide or amino acid identity, typically one sequence acts as a reference sequence, to which test sequences are compared. When using a sequence comparison algorithm, test and reference sequences are input into a computer, subsequence coordinates are designated, if necessary, and sequence
10 algorithm program parameters are designated. The sequence comparison algorithm then calculates the percent sequence identity for the test sequence(s) relative to the reference sequence, based on the designated program parameters.

[0142] Optimal alignment of sequences for comparison may be conducted, *e.g.*, by the local homology algorithm of Smith & Waterman, *Adv. Appl. Math.* 2:482 (1981), by the
15 homology alignment algorithm of Needleman & Wunsch, *J. Mol. Biol.* 48:443 (1970), by the search for similarity method of Pearson & Lipman, *Proc. Nat'l. Acad. Sci. USA* 85:2444 (1988), by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, WI), or by visual inspection (see generally Ausubel *et al.*, *supra*).
20

[0143] Another example algorithm that is suitable for determining percent sequence identity and sequence similarity is the BLAST algorithm, which is described in Altschul *et al.*, *J. Mol. Biol.* 215:403-410 (1990). Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information (at the web
25 address www.ncbi.nlm.nih.gov). This algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence, which either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as the neighborhood word score threshold (Altschul *et al.*, *supra*). These initial neighborhood word hits act as seeds
30 for initiating searches to find longer HSPs containing them. The word hits are then extended in both directions along each sequence for as far as the cumulative alignment score may be increased. Cumulative scores are calculated using, for nucleotide sequences, the parameters M (reward score for a pair of matching residues; always > 0) and N

(penalty score for mismatching residues; always < 0). For amino acid sequences, a scoring matrix is used to calculate the cumulative score. Extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. For identifying whether a nucleic acid or polypeptide is within the scope of the invention, the default parameters of the BLAST programs are suitable. The BLASTN program (for nucleotide sequences) uses as defaults a word length (W) of 11, an expectation (E) of 10, $M=5$, $N=-4$, 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. The TBLASTN program (using protein sequence for nucleotide sequence) uses as defaults a word length (W) of 3, an expectation (E) of 10, and a BLOSUM 62 scoring matrix. (see Henikoff & Henikoff, *Proc. Natl. Acad. Sci. USA* 89:10915 (1989)).

[0144] In addition to calculating percent sequence identity, the BLAST algorithm also performs a statistical analysis of the similarity between two sequences (see, e.g., 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, more preferably less than about 0.01, and most preferably less than about 0.001.

[0145] It is understood that the peptide and protein sequences provided herein can have conservative or non-essential amino acid substitutions that do not have a substantial effect on the function of the peptide. Conservative amino acid substitutions include replacement of an aliphatic amino acid, such as alanine, valine, leucine, and isoleucine, with another aliphatic amino acid; replacement of a serine with a threonine; replacement of a threonine with a serine; replacement of an acidic residue, such as aspartic acid and glutamic acid, with another acidic residue; replacement of a residue bearing an amide group, such as asparagine and glutamine, with another residue bearing an amide group; exchange of a basic residue, such as lysine and arginine, with another basic residue; and replacement of an aromatic residue, such as phenylalanine and tyrosine, with another aromatic residue. In some embodiments, the polypeptide has about 1, 2, 3, 4, 5, 6, 7, 8, 9,

10, 15, 20, 30, 40, 50, 60, 70, 80, 90, 100 or more amino acid substitutions, additions, insertions, or deletions.

[0146] Other considerations affecting the selection of microorganisms for use in the invention include, in addition to production of suitable lipids or hydrocarbons for
5 production of oils, fuels, and oleochemicals: (1) high lipid content as a percentage of cell weight; (2) ease of growth; (3) ease of genetic engineering; and (4) ease of biomass processing. In particular embodiments, the wild-type or genetically engineered microorganism yields cells that are at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, or at least 70% or more lipid. Preferred organisms grow
10 heterotrophically (on sugars in the absence of light).

[0147] Examples of algae that may be used to practice the present invention include, but are not limited to the following algae: *Prototheca wickerhamii*, *Prototheca stagnora*, *Prototheca portoricensis*, *Prototheca moriformis*, *Prototheca zopfii*.

15 2. BIOREACTOR

[0148] Microorganisms are cultured both for purposes of conducting genetic manipulations and for production of hydrocarbons (e.g., lipids, fatty acids, aldehydes, alcohols, and alkanes). The former type of culture is conducted on a small scale and initially, at least, under conditions in which the starting microorganism can grow. Culture
20 for purposes of hydrocarbon production is usually conducted on a large scale (e.g., 10,000 L, 40,000 L, 100,000 L or larger bioreactors) in a bioreactor. Microalgae, including *Prototheca* species are typically cultured in the methods of the invention in liquid media within a bioreactor. Typically, the bioreactor does not allow light to enter.

[0149] The bioreactor or fermentor is used to culture microalgal cells through the
25 various phases of their physiological cycle. Bioreactors offer many advantages for use in heterotrophic growth and propagation methods. To produce biomass for use in food, microalgae are preferably fermented in large quantities in liquid, such as in suspension cultures as an example. Bioreactors such as steel fermentors can accommodate very large culture volumes (40,000 liter and greater capacity bioreactors are used in various
30 embodiments of the invention). Bioreactors also typically allow for the control of culture conditions such as temperature, pH, oxygen tension, and carbon dioxide levels. For example, bioreactors are typically configurable, for example, using ports attached to tubing, to allow gaseous components, like oxygen or nitrogen, to be bubbled through a

liquid culture. Other culture parameters, such as the pH of the culture media, the identity and concentration of trace elements, and other media constituents can also be more readily manipulated using a bioreactor.

[0150] Bioreactors may be configured to flow culture media through the bioreactor

5 throughout the time period during which the microalgae reproduce and increase in number. In some embodiments, for example, media may be infused into the bioreactor after inoculation but before the cells reach a desired density. In other instances, a bioreactor is filled with culture media at the beginning of a culture, and no more culture media is infused after the culture is inoculated. In other words, the microalgal biomass is
10 cultured in an aqueous medium for a period of time during which the microalgae reproduce and increase in number; however, quantities of aqueous culture medium are not flowed through the bioreactor throughout the time period. Thus in some embodiments, aqueous culture medium is not flowed through the bioreactor after inoculation.

[0151] Bioreactors equipped with devices such as spinning blades and impellers,

15 rocking mechanisms, stir bars, or means for pressurized gas infusion may be used to subject microalgal cultures to mixing. Mixing may be continuous or intermittent. For example, in some embodiments, a turbulent flow regime of gas entry and media entry is not maintained for reproduction of microalgae until a desired increase in number of said microalgae has been achieved.

20 **[0152]** Bioreactor ports may be used to introduce, or extract, gases, solids, semisolids, and liquids, into the bioreactor chamber containing the microalgae. While many

bioreactors have more than one port (for example, one for media entry, and another for sampling), it is not necessary that only one substance enter or leave a port. For example, a port may be used to flow culture media into the bioreactor and later used for sampling, gas
25 entry, gas exit, or other purposes. Preferably, a sampling port may be used repeatedly without altering compromising the axenic nature of the culture. A sampling port may be configured with a valve or other device that allows the flow of sample to be stopped and started or to provide a means of continuous sampling. Bioreactors typically have at least one port that allows inoculation of a culture, and such a port can also be used for other
30 purposes such as media or gas entry.

[0153] Bioreactors ports allow the gas content of the culture of microalgae to be manipulated. To illustrate, part of the volume of a bioreactor may be gas rather than liquid, and the gas inlets of the bioreactor to allow pumping of gases into the bioreactor. Gases

that may be beneficially pumped into a bioreactor include air, air/CO₂ mixtures, noble gases, such as argon, and other gases. Bioreactors are typically equipped to enable the user to control the rate of entry of a gas into the bioreactor. As noted above, increasing gas flow into a bioreactor may be used to increase mixing of the culture.

5 [0154] Increased gas flow affects the turbidity of the culture as well. Turbulence may be achieved by placing a gas entry port below the level of the aqueous culture media so that gas entering the bioreactor bubbles to the surface of the culture. One or more gas exit ports allow gas to escape, thereby preventing pressure buildup in the bioreactor. Preferably a gas exit port leads to a “one-way” valve that prevents contaminating microorganisms
10 from entering the bioreactor.

3. MEDIA

[0155] Microalgal culture media typically contains components such as a fixed nitrogen source, a fixed carbon source, trace elements, optionally a buffer for pH maintenance, and phosphate (typically provided as a phosphate salt). Other components
15 can include salts such as sodium chloride, particularly for seawater microalgae. Nitrogen sources include organic and inorganic nitrogen sources, including, for example, without limitation, molecular nitrogen, nitrate, nitrate salts, ammonia (pure or in salt form, such as, (NH₄)₂SO₄ and NH₄OH), protein, soybean meal, cornsteep liquor, and yeast extract. Examples of trace elements include zinc, boron, cobalt, copper, manganese, and
20 molybdenum in, for example, the respective forms of ZnCl₂, H₃BO₃, CoCl₂·6H₂O, CuCl₂·2H₂O, MnCl₂·4H₂O and (NH₄)₆Mo₇O₂₄·4H₂O.

[0156] Microorganisms useful in accordance with the methods of the present invention are found in various locations and environments throughout the world. As a consequence of their isolation from other species and their resulting evolutionary divergence, the
25 particular growth medium for optimal growth and generation of lipid and/or hydrocarbon constituents may be difficult to predict. In some cases, certain strains of microorganisms may be unable to grow on a particular growth medium because of the presence of some inhibitory component or the absence of some essential nutritional requirement required by the particular strain of microorganism.

30 [0157] Solid and liquid growth media are generally available from a wide variety of sources, and instructions for the preparation of particular media that is suitable for a wide variety of strains of microorganisms may be found, for example, online at <http://www.utex.org/>, a site maintained by the University of Texas at Austin, 1 University

Station A6700, Austin, Texas, 78712-0183, for its culture collection of algae (UTEX). For example, various fresh water and salt water media include those described in PCT Pub. No. 2008/151149, incorporated herein by reference.

[0158] In a particular example, Proteose Medium is suitable for axenic cultures, and a

5 1L volume of the medium (pH ~6.8) may be prepared by addition of 1g of proteose peptone to 1 liter of Bristol Medium. Bristol medium comprises 2.94 mM NaNO_3 , 0.17 mM $\text{CaCl}_2 \cdot 2\text{H}_2\text{O}$, 0.3 mM $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$, 0.43 mM, 1.29 mM KH_2PO_4 , and 1.43 mM NaCl in an aqueous solution. For 1.5% agar medium, 15 g of agar may be added to 1 L of the solution. The solution is covered and autoclaved, and then stored at a refrigerated
10 temperature prior to use. Another example is the *Prototheca* isolation medium (PIM), which comprises 10g/L potassium hydrogen phthalate (KHP), 0.9g/L sodium hydroxide, 0.1g/L magnesium sulfate, 0.2g/L potassium hydrogen phosphate, 0.3g/L ammonium chloride, 10g/L glucose 0.001g/L thiamine hydrochloride, 20g/L agar, 0.25g/L 5-fluorocytosine, at a pH in the range of 5.0 to 5.2 (see Pore, 1973, App. Microbiology, 26:
15 648-649). Other suitable media for use with the methods of the invention may be readily identified by consulting the URL identified above, or by consulting other organizations that maintain cultures of microorganisms, such as SAG, CCAP, or CCALA. SAG refers to the Culture Collection of Algae at the University of Göttingen (Göttingen, Germany), CCAP refers to the culture collection of algae and protozoa managed by the Scottish
20 Association for Marine Science (Scotland, United Kingdom), and CCALA refers to the culture collection of algal laboratory at the Institute of Botany (Třeboň, Czech Republic). Additionally, US Patent No. 5,900,370 describes media formulations and conditions suitable for heterotrophic fermentation of *Prototheca* species.

[0159] For oil production, selection of a fixed carbon source is important, as the cost

25 of the fixed carbon source must be sufficiently low to make oil production economical. Thus, while suitable carbon sources include, for example, acetate, floridoside, fructose, galactose, glucuronic acid, glucose, glycerol, lactose, mannose, N-acetylglucosamine, rhamnose, sucrose, and/or xylose, selection of feedstocks containing those compounds is an important aspect of the methods of the invention. Suitable feedstocks useful in
30 accordance with the methods of the invention include, for example, black liquor, corn starch, depolymerized cellulosic material, milk whey, molasses, potato, sorghum, sucrose, sugar beet, sugar cane, rice, and wheat. Carbon sources can also be provided as a mixture, such as a mixture of sucrose and depolymerized sugar beet pulp. The one or more carbon source(s) may be supplied at a concentration of at least about 50 μM , at least about 100

μM, at least about 500 μM, at least about 5 mM, at least about 50 mM, and at least about 500 mM, of one or more exogenously provided fixed carbon source(s). Highly concentrated carbon sources as feedstock for fermentation are preferred. For example, in some embodiments glucose levels of at least 300g/L, at least 400g/L, at least 500g/L, or at least 600g/L or more of glucose level of the feedstock prior to the cultivation step, is added to a fed batch cultivation, in which the highly concentrated fixed carbon source is fed to the cells over time as the cells grow and accumulate lipid. In other embodiments, sucrose levels of at least 500g/L, at least 600g/L, at least 700g/L, at least 800g/L or more of sucrose prior to the cultivation is added to a fed batch cultivation, in which the highly concentrated fixed carbon source is fed to the cells over time as the cells grow and accumulate lipid. Non-limiting examples of highly concentrated fixed carbon source such as sucrose include thick cane juice, sugar cane juice, sugar beet juice and molasses. Carbon sources of particular interest for purposes of the present invention include cellulose (in a depolymerized form), glycerol, sucrose, and sorghum, each of which is discussed in more detail below.

[0160] In accordance with the present invention, microorganisms may be cultured using depolymerized cellulosic biomass as a feedstock. Cellulosic biomass (e.g., stover, such as corn stover) is inexpensive and readily available; however, attempts to use this material as a feedstock for yeast have failed. In particular, such feedstocks have been found to be inhibitory to yeast growth, and yeast cannot use the 5-carbon sugars produced from cellulosic materials (e.g., xylose from hemi-cellulose). By contrast, microalgae can grow on processed cellulosic material. Cellulosic materials generally include about 40-60% cellulose; about 20-40% hemicellulose; and 10-30% lignin.

[0161] Suitable cellulosic materials include residues from herbaceous and woody energy crops, as well as agricultural crops, i.e., the plant parts, primarily stalks and leaves, not removed from the fields with the primary food or fiber product. Examples include agricultural wastes such as sugarcane bagasse, rice hulls, corn fiber (including stalks, leaves, husks, and cobs), wheat straw, rice straw, sugar beet pulp, citrus pulp, citrus peels; forestry wastes such as hardwood and softwood thinnings, and hardwood and softwood residues from timber operations; wood wastes such as saw mill wastes (wood chips, sawdust) and pulp mill waste; urban wastes such as paper fractions of municipal solid waste, urban wood waste and urban green waste such as municipal grass clippings; and wood construction waste. Additional cellulose include dedicated cellulosic crops such as

switchgrass, hybrid poplar wood, and miscanthus, fiber cane, and fiber sorghum. Five-carbon sugars that are produced from such materials include xylose.

[0162] Cellulosic materials are treated to increase the efficiency with which the microbe can utilize the sugar(s) contained within the materials. The invention provides novel methods for the treatment of cellulosic materials after acid explosion so that the materials are suitable for use in a heterotrophic culture of microbes (e.g., microalgae and oleaginous yeast). As discussed above, lignocellulosic biomass is comprised of various fractions, including cellulose, a crystalline polymer of beta 1,4 linked glucose (a six-carbon sugar), hemicellulose, a more loosely associated polymer predominantly comprised of xylose (a five-carbon sugar) and to a lesser extent mannose, galactose, arabinose, lignin, a complex aromatic polymer comprised of sinapyl alcohol and its derivatives, and pectins, which are linear chains of an alpha 1,4 linked polygalacturonic acid. Because of the polymeric structure of cellulose and hemicellulose, the sugars (e.g., monomeric glucose and xylose) in them are not in a form that may be efficiently used (metabolized) by many microbes. For such microbes, further processing of the cellulosic biomass to generate the monomeric sugars that make up the polymers may be very helpful to ensuring that the cellulosic materials are efficiently utilized as a feedstock (carbon source).

[0163] In another embodiment of the methods of the invention, the carbon source is glycerol, including acidulated and non-acidulated glycerol byproduct from biodiesel transesterification. In one embodiment, the carbon source includes glycerol and at least one other carbon source. In some cases, all of the glycerol and the at least one other fixed carbon source are provided to the microorganism at the beginning of the fermentation. In some cases, the glycerol and the at least one other fixed carbon source are provided to the microorganism simultaneously at a predetermined ratio. In some cases, the glycerol and the at least one other fixed carbon source are fed to the microbes at a predetermined rate over the course of fermentation.

[0164] Some microalgae undergo cell division faster in the presence of glycerol than in the presence of glucose (see PCT Pub. No. 2008/151149). In these instances, two-stage growth processes in which cells are first fed glycerol to rapidly increase cell density, and are then fed glucose to accumulate lipids can improve the efficiency with which lipids are produced. The use of the glycerol byproduct of the transesterification process provides significant economic advantages when put back into the production process. Other feeding methods are provided as well, such as mixtures of glycerol and glucose. Feeding such

mixtures also captures the same economic benefits. In addition, the invention provides methods of feeding alternative sugars to microalgae such as sucrose in various combinations with glycerol.

[0165] In another embodiment of the methods of the invention, the carbon source is invert sugar. Invert sugar is produced by splitting the sucrose into its monosaccharide components, fructose and glucose. Production of invert sugar may be achieved through several methods that are known in the art. One such method is heating an aqueous solution of sucrose. Often, catalysts are employed in order to accelerate the conversion of sucrose into invert sugar. These catalysts may be biological, for example enzymes such as invertases and sucrases may be added to the sucrose to accelerate the hydrolysis reaction to produce invert sugar. Acid is an example of non-biological catalyst, when paired with heat, can accelerate the hydrolysis reaction. Once the invert sugar is made, it is less prone to crystallization compared to sucrose and thus, provides advantages for storage and in fed batch fermentation, which in the case of heterotrophic cultivation of microbes, including microalgae, there is a need for concentrated carbon source. In one embodiment, the carbon source is invert sugar, preferably in a concentrated form, preferably at least 800g/liter, at least 900 g/liter, at least 1000 g/liter or at least 1100 g/liter prior to the cultivation step, which is optionally a fed batch cultivation. The invert sugar, preferably in a concentrated form, is fed to the cells over time as the cells grow and accumulate lipid.

[0166] In another embodiment of the methods of the invention, the carbon source is sucrose, including a complex feedstock containing sucrose, such as thick cane juice from sugar cane processing. Because of the higher densities of the cultures for heterotrophic oil production, the fixed carbon source (e.g., sucrose, glucose, etc.) is preferably in a concentrated form, preferably at least 500 g/liter, at least 600 g/liter, at least 700 g/liter or at least 800 g/liter of the fixed carbon source prior to the cultivation step, which is optionally a fed batch cultivation in which the material is fed to the cells over time as the cells grow and accumulate lipid. In the some cases, the carbon source is sucrose in the form of thick cane juice, preferably in a concentrated form, preferably at least 60% solids or about 770 g/liter sugar, at least 70% solids or about 925 g/liter sugar, or at least 80% solids or about 1125 g/liter sugar prior to the cultivation step, which is optionally a fed batch cultivation. The concentrated thick cane juice is fed to the cells over time as the cells grow and accumulate lipid

[0167] In one embodiment, the culture medium further includes at least one sucrose utilization enzyme. In some cases, the culture medium includes a sucrose invertase. In one embodiment, the sucrose invertase enzyme is a secretable sucrose invertase enzyme encoded by an exogenous sucrose invertase gene expressed by the population of
5 microorganisms. Thus, in some cases, as described in more detail in Section IV, below, the microalgae has been genetically engineered to express a sucrose utilization enzyme, such as a sucrose transporter, a sucrose invertase, a hexokinase, a glucokinase, or a fructokinase.

[0168] Complex feedstocks containing sucrose include waste molasses from sugar
10 cane processing; the use of this low-value waste product of sugar cane processing can provide significant cost savings in the production of hydrocarbons and other oils. Another complex feedstock containing sucrose that is useful in the methods of the invention is sorghum, including sorghum syrup and pure sorghum. Sorghum syrup is produced from the juice of sweet sorghum cane. Its sugar profile consists of mainly glucose (dextrose),
15 fructose and sucrose.

SECTION IV-I. GENETIC ENGINEERING METHODS AND MATERIALS

[0169] The present invention provides methods and materials for genetically modifying *Prototheca* cells and recombinant host cells useful in the methods of the present
20 invention, including but not limited to recombinant *Prototheca moriformis*, *Prototheca zopfii*, *Prototheca krugani*, and *Prototheca stagnora* host cells. The description of these methods and materials is divided into subsections for the convenience of the reader. In subsection 1, transformation methods are described. In subsection 2, genetic engineering methods using homologous recombination are described. In subsection 3, expression
25 vectors and components are described. In subsection 4, selectable markers and components are described.

1. ENGINEERING METHODS - TRANSFORMATION

[0170] Cells may be transformed by any suitable technique including, e.g., biolistics, electroporation (see Maruyama et al. (2004), *Biotechnology Techniques* 8:821-826), glass
30 bead transformation and silicon carbide whisker transformation. Another method that may be used involves forming protoplasts and using CaCl₂ and polyethylene glycol (PEG) to introduce recombinant DNA into microalgal cells (see Kim et al. (2002), *Mar. Biotechnol.* 4:63-73, which reports the use of this method for the transformation of *Chorella*

ellipsoidea). Co-transformation of microalgae may be used to introduce two distinct vector molecules into a cell simultaneously (see for example Jakobiak et al. (2004) Protist;155(4):381-93).

[0171] Biolistic methods (see, for example, Sanford, Trends In Biotech. (1988) 6:299-302, U.S. Patent No. 4,945,050; electroporation (Fromm et al., Proc. Nat'l. Acad. Sci. (USA) (1985) 82:5824-5828); use of a laser beam, microinjection or any other method capable of introducing DNA into a microalgae can also be used for transformation of a *Prototheca* cell.

2. ENGINEERING METHODS - HOMOLOGOUS RECOMBINATION

[0172] Homologous recombination is the ability of complementary DNA sequences to align and exchange regions of homology. Transgenic DNA ("donor") containing sequences homologous to the genomic sequences being targeted ("template") is introduced into the organism and then undergoes recombination into the genome at the site of the corresponding genomic homologous sequences. The mechanistic steps of this process, in most cases, include: (1) pairing of homologous DNA segments; (2) introduction of double-stranded breaks into the donor DNA molecule; (3) invasion of the template DNA molecule by the free donor DNA ends followed by DNA synthesis; and (4) resolution of double-strand break repair events that result in final recombination products.

[0173] The ability to carry out homologous recombination in a host organism has many practical implications for what may be carried out at the molecular genetic level and is useful in the generation of an oleaginous microbe that can produce tailored oils. By its very nature homologous recombination is a precise gene targeting event, hence, most transgenic lines generated with the same targeting sequence will be essentially identical in terms of phenotype, necessitating the screening of far fewer transformation events.

Homologous recombination also targets gene insertion events into the host chromosome, resulting in excellent genetic stability, even in the absence of genetic selection. Because different chromosomal loci will likely impact gene expression, even from heterologous promoters/UTRs, homologous recombination may be a method of querying loci in an unfamiliar genome environment and to assess the impact of these environments on gene expression.

[0174] Particularly useful genetic engineering applications using homologous recombination is to co-opt specific host regulatory elements such as promoters/UTRs to drive heterologous gene expression in a highly specific fashion. For example, ablation or

knockout of desaturase genes/gene families with a heterologous gene encoding a selective marker might be expected to increase overall percentage of saturated fatty acids produced in the host cell. Example 3 describes the homologous recombination targeting constructs and a working example of such desaturase gene ablations or knockouts generated in

5 *Prototheca moriformis*. Another approach to decreasing expression of an endogenous gene is to use an RNA-induced down-regulation or silencing of gene expression including, but not limited to an RNAi or antisense approach, as well as a dsRNA approach. Antisense, RNAi, RNA hairpin, and dsRNA approaches are well known in the art and include the introduction of an expression construct that when expressed as mRNA would lead to the
10 formation of hairpin RNA or an expression construct containing a portion of the target gene that would be transcribed in the antisense orientation. All four approaches would result in the decreased expression of the target gene. Examples 3 and 4 describe expression constructs and working examples of the attenuation, or down-regulation of endogenous *Prototheca moriformis* lipid biosynthesis genes by an RNA hairpin approach.

15 [0175] Because homologous recombination is a precise gene targeting event, it may be used to precisely modify any nucleotide(s) within a gene or region of interest, so long as sufficient flanking regions have been identified. Therefore, homologous recombination may be used as a means to modify regulatory sequences impacting gene expression of RNA and/or proteins. It can also be used to modify protein coding regions in an effort to
20 modify enzyme activities such as substrate specificity, affinities and K_m , and thus affecting the desired change in metabolism of the host cell. Homologous recombination provides a powerful means to manipulate the host genome resulting in gene targeting, gene conversion, gene deletion, gene duplication, gene inversion and exchanging gene expression regulatory elements such as promoters, enhancers and 3'UTRs.

25 [0176] Homologous recombination may be achieved by using targeting constructs containing pieces of endogenous sequences to "target" the gene or region of interest within the endogenous host cell genome. Such targeting sequences can either be located 5' of the gene or region of interest, 3' of the gene/region of interest or even flank the gene/region of interest. Such targeting constructs may be transformed into the host cell either as a
30 supercoiled plasmid DNA with additional vector backbone, a PCR product with no vector backbone, or as a linearized molecule. In some cases, it may be advantageous to first expose the homologous sequences within the transgenic DNA (donor DNA) with a restriction enzyme. This step can increase the recombination efficiency and decrease the occurrence of undesired events. Other methods of increasing recombination efficiency

include using PCR to generate transforming transgenic DNA containing linear ends homologous to the genomic sequences being targeted.

[0177] For purposes of non-limiting illustration, regions of donor DNA sequences that are useful for homologous recombination include the KE858 region of DNA in *Prototheca moriformis*. KE858 is a 1.3 kb, genomic fragment that encompasses part of the coding region for a protein that shares homology with the transfer RNA (tRNA) family of proteins. Southern blots have shown that the KE858 sequence is present in a single copy in the *Prototheca moriformis* (UTEX 1435) genome. This region and Examples of using this region for homologous recombination targeting has been described in PCT Publication No. WO 2010/063032. Another region of donor DNA that is useful is portions of the 6S genomic sequence. The use of this sequence in homologous recombination in *Prototheca moriformis* is described below in the Examples.

3. VECTORS AND VECTOR COMPONENTS

[0178] Vectors for transformation of microorganisms in accordance with embodiments of the present invention may be prepared by known techniques familiar to those skilled in the art in view of the disclosure herein. A vector typically contains one or more genes, in which each gene codes for the expression of a desired product (the gene product) and is operably linked to one or more control sequences that regulate gene expression or target the gene product to a particular location in the recombinant cell. To aid the reader, this subsection is divided into subsections. Subsection A describes control sequences typically contained on vectors as well as novel control sequences provided by the present invention. Subsection B describes genes typically contained in vectors as well as novel codon optimization methods and genes prepared using them provided by the invention. Subsection C describes selectable markers contained on vectors and provided by the present invention. Subsection D describes methods and procedures used to identify genes.

A. Control Sequences

[0179] Control sequences are nucleic acids that regulate the expression of a coding sequence or direct a gene product to a particular location in or outside a cell. Control sequences that regulate expression include, for example, promoters that regulate transcription of a coding sequence and terminators that terminate transcription of a coding sequence. Another control sequence is a 3' untranslated sequence located at the end of a coding sequence that encodes a polyadenylation signal. Control sequences that direct gene

products to particular locations include those that encode signal peptides, which direct the protein to which they are attached to a particular location in or outside the cell.

[0180] Thus, an exemplary vector design for expression of an exogenous gene in a microalgae contains a coding sequence for a desired gene product (for example, a selectable marker, a lipid pathway enzyme, or a sucrose utilization enzyme) in operable linkage with a promoter active in microalgae. Alternatively, if the vector does not contain a promoter in operable linkage with the coding sequence of interest, the coding sequence may be transformed into the cells such that it becomes operably linked to an endogenous promoter at the point of vector integration. The promoterless method of transformation has been proven to work in microalgae (see for example Lumbreras, et. al. Plant Journal (1988), 14(4), pp.441-447.

[0181] Many promoters are active in microalgae, including promoters that are endogenous to the algae being transformed, as well as promoters that are not endogenous to the algae being transformed (i.e., promoters from other algae, promoters from higher plants, and promoters from plant viruses or algae viruses). Illustrative exogenous and/or endogenous promoters that are active in microalgae (as well as antibiotic resistance genes functional in microalgae) are described in PCT Pub. No. 2008/151149 and references cited therein

[0182] The promoter used to express an exogenous gene may be the promoter naturally linked to that gene or may be a heterologous gene. Some promoters are active in more than one species of microalgae. Other promoters are species-specific. Illustrative promoters include promoters such as β -tubulin from *Chlamydomonas reinhardtii*, used in the Examples below, and viral promoters, such as cauliflower mosaic virus (CMV) and chlorella virus, which have been shown to be active in multiple species of microalgae (see for example Plant Cell Rep. 2005 Mar;23(10-11):727-35; J Microbiol. 2005 Aug;43(4):361-5; Mar Biotechnol (NY). 2002 Jan;4(1):63-73). Another promoter that is suitable for use for expression of exogenous genes in *Prototheca* is the *Chlorella sorokiniana* glutamate dehydrogenase promoter/5'UTR. Optionally, at least 10, 20, 30, 40, 50, or 60 nucleotides or more of these sequences containing a promoter are used. Illustrative promoters useful for expression of exogenous genes in *Prototheca* are listed in the sequence listing of this application, such as the promoter of the *Chlorella* HUP1 gene (SEQ ID NO:10) and the *Chlorella ellipsoidea* nitrate reductase promoter (SEQ ID NO:11). Chlorella virus promoters can also be used to express genes in *Prototheca*, such

as sequence numbers 1 to 7 of U.S. Patent 6,395,965. Additional promoters active in *Prototheca* may be found, for example, in Biochem Biophys Res Commun. 1994 Oct 14;204(1):187-94; Plant Mol Biol. 1994 Oct;26(1):85-93; Virology. 2004 Aug 15;326(1):150-9; and Virology. 2004 Jan 5;318(1):214-23. Other useful promoters are
5 described in detail in the Examples below.

[0183] A promoter can generally be characterized as either constitutive or inducible. Constitutive promoters are generally active or function to drive expression at all times (or at certain times in the cell life cycle) at the same level. Inducible promoters, conversely, are active (or rendered inactive) or are significantly up- or down-regulated only in
10 response to a stimulus. Both types of promoters find application in embodiments of the present invention. Inducible promoters useful in embodiments of the present invention include those that mediate transcription of an operably linked gene in response to a stimulus, such as an exogenously provided small molecule (e.g, glucose, as in SEQ ID NO:10), temperature (heat or cold), lack of nitrogen in culture media, pH, etc. Suitable
15 promoters can activate transcription of an essentially silent gene or upregulate, preferably substantially, transcription of an operably linked gene that is transcribed at a low level. Examples below describe additional inducible promoters that are useful in *Prototheca* cells.

[0184] The termination region, also referred to as a 3' untranslated region, may be
20 native to the transcriptional initiation region (the promoter), may be native to the DNA sequence of interest, or may be obtainable from another source. See, for example, Chen and Orozco, Nucleic Acids Res. (1988), 16:8411.

[0185] In an embodiment of the present invention, control sequences that provide for the compartmentalized expression of an exogenous enzyme or protein are utilized to direct
25 the exogenous enzyme or protein to one or more intracellular organelles. Organelles for targeting are chloroplasts, plastids, mitochondria, and endoplasmic reticulum. An additional embodiment of the present invention provides recombinant polynucleotides that enable secretion of a protein outside the cell.

[0186] Proteins encoded in the nuclear genome of *Prototheca* may be targeted to the
30 plastid using plastid targeting signals. Plastid targeting sequences endogenous to *Chlorella* are known, such as genes in the *Chlorella* nuclear genome that encode proteins that are targeted to the plastid; see for example GenBank Accession numbers AY646197 and

AF499684, and in one embodiment, such control sequences are used in the vectors of the present invention to target expression of a protein to a *Prototheca* plastid.

[0187] The Examples below describe the use of algal plastid targeting sequences to target exogenous enzymes and proteins to the correct compartment in the host cell. Algal plastid targeting sequences were obtained from cDNA libraries made using *Prototheca moriformis* and *Chlorella protothecoides* cells and are described in PCT Publication No. WO 2010/063032.

[0188] In another embodiment, the expression of an exogenous enzyme or protein in *Prototheca* is targeted to the endoplasmic reticulum. The inclusion of an appropriate retention or sorting signal in an expression vector ensure that proteins are retained in the endoplasmic reticulum (ER) and do not go downstream into Golgi. For example, the IMPACTVECTOR1.3 vector, from Wageningen UR- Plant Research International, includes the well known KDEL retention or sorting signal. With this vector, ER retention has a practical advantage in that it has been reported to improve expression levels 5-fold or more. The main reason for this appears to be that the ER contains lower concentrations and/or different proteases responsible for post-translational degradation of expressed proteins than are present in the cytoplasm. ER retention signals functional in green microalgae are known. For example, see Proc. Natl. Acad. Sci. U S A. 2005 Apr 26;102(17):6225-30.

[0189] In another embodiment of the present invention, a polypeptide is targeted for secretion outside the cell into the culture media. See Hawkins et al., Current Microbiology Vol. 38 (1999), pp. 335–341 for examples of secretion signals active in *Chlorella* that may be used, in accordance with the methods of the invention, in *Prototheca*.

B. Genes and Codon Optimization

[0190] Typically, a gene includes a promoter, coding sequence, and termination control sequences. When assembled by recombinant DNA technology, a gene may be termed an expression cassette and may be flanked by restriction sites for convenient insertion into a vector that is used to introduce the recombinant gene into a host cell. The expression cassette may be flanked by DNA sequences from the genome or other nucleic acid target to facilitate stable integration of the expression cassette into the genome by homologous recombination. Alternatively, the vector and its expression cassette may remain unintegrated, in which case, the vector typically includes an origin of replication, which is capable of providing for replication of the heterologous vector DNA.

[0191] A common gene present on a vector is a gene that codes for a protein, the expression of which allows the recombinant cell containing the protein to be differentiated from cells that do not express the protein. Such a gene, and its corresponding gene product, is called a selectable marker. Any of a wide variety of selectable markers may be employed in a transgene construct useful for transforming *Prototheca*. Examples of suitable selectable markers include the G418 resistance gene, the nitrate reductase gene (see Dawson et al. (1997), Current Microbiology 35:356-362), the hygromycin phosphotransferase gene (HPT; see Kim et al. (2002), Mar. Biotechnol. 4:63-73), the neomycin phosphotransferase gene, and the ble gene, which confers resistance to phleomycin (Huang et al. (2007), Appl. Microbiol. Biotechnol. 72:197-205). Methods of determining sensitivity of microalgae to antibiotics are well known. For example, Mol Gen Genet. 1996 Oct 16;252(5):572-9.

[0192] Other selectable markers that are not antibiotic-based can also be employed in a transgene construct useful for transforming microalgae in general, including *Prototheca* species. Genes that confer the ability to utilize certain carbon sources that were previously unable to be utilized by the microalgae can also be used as a selectable marker. By way of illustration, *Prototheca moriformis* strains typically grow poorly, if at all, on sucrose. Using a construct containing a sucrose invertase gene can confer the ability of positive transformants to grow on sucrose as a carbon substrate. Additional details on using sucrose utilization as a selectable marker along with other selectable markers are discussed below.

[0193] For purposes of the present invention, the expression vector used to prepare a recombinant host cell of the invention will include at least two, and often three, genes, if one of the genes is a selectable marker. For example, a genetically engineered *Prototheca* of the invention may be made by transformation with vectors of the invention that comprise, in addition to a selectable marker, one or more exogenous genes, such as, for example, sucrose invertase gene or acyl ACP-thioesterase gene. One or both genes may be expressed using an inducible promoter, which allows the relative timing of expression of these genes to be controlled to enhance the lipid yield and conversion to fatty acid esters. Expression of the two or more exogenous genes may be under control of the same inducible promoter or under control of different inducible (or constitutive) promoters. In the latter situation, expression of a first exogenous gene may be induced for a first period of time (during which expression of a second exogenous gene may or may not be induced)

and expression of a second exogenous gene may be induced for a second period of time (during which expression of a first exogenous gene may or may not be induced).

[0194] In other embodiments, the two or more exogenous genes (in addition to any selectable marker) are: a fatty acyl-ACP thioesterase and a fatty acyl-CoA/aldehyde

5 reductase, the combined action of which yields an alcohol product.

[0195] Other illustrative vectors of embodiments of the invention that express two or more exogenous genes include those encoding both a sucrose transporter and a sucrose invertase enzyme and those encoding both a selectable marker and a secreted sucrose invertase. The recombinant *Prototheca* transformed with either type of vector produce
10 lipids at lower manufacturing cost due to the engineered ability to use sugar cane (and sugar cane-derived sugars) as a carbon source. Insertion of the two exogenous genes described above may be combined with the disruption of polysaccharide biosynthesis through directed and/or random mutagenesis, which steers ever greater carbon flux into lipid production. Individually and in combination, trophic conversion, engineering to alter
15 lipid production and treatment with exogenous enzymes alter the lipid composition produced by a microorganism. The alteration may be a change in the amount of lipids produced, the amount of one or more hydrocarbon species produced relative to other lipids, and/or the types of lipid species produced in the microorganism. For example, microalgae may be engineered to produce a higher amount and/or percentage of TAGs.

[0196] For optimal expression of a recombinant protein, it is beneficial to employ coding sequences that produce mRNA with codons preferentially used by the host cell to be transformed. Thus, proper expression of transgenes can require that the codon usage of the transgene matches the specific codon bias of the organism in which the transgene is being expressed. The precise mechanisms underlying this effect are many, but include the
25 proper balancing of available aminoacylated tRNA pools with proteins being synthesized in the cell, coupled with more efficient translation of the transgenic messenger RNA (mRNA) when this need is met. When codon usage in the transgene is not optimized, available tRNA pools are not sufficient to allow for efficient translation of the heterologous mRNA resulting in ribosomal stalling and termination and possible
30 instability of the transgenic mRNA.

[0197] The present invention provides codon-optimized nucleic acids useful for the successful expression of recombinant proteins in *Prototheca*. Codon usage in *Prototheca* species was analyzed by studying cDNA sequences isolated from *Prototheca moriformis*.

This analysis represents the interrogation over 24,000 codons and resulted in Table 2 below.

Table 2. Illustrative preferred codon usage in *Prototheca* strains.

5	Ala	GCG	345 (0.36)	Asn	AAT	8 (0.04)
		GCA	66 (0.07)		AAC	201 (0.96)
		GCT	101 (0.11)			
		GCC	442 (0.46)			
10	Cys	TGT	12 (0.10)	Pro	CCG	161 (0.29)
		TGC	105 (0.90)		CCA	49 (0.09)
					CCT	71 (0.13)
					CCC	267 (0.49)
15	Asp	GAT	43 (0.12)	Gln	CAG	226 (0.82)
		GAC	316 (0.88)		CAA	48 (0.18)
20	Glu	GAG	377 (0.96)	Arg	AGG	33 (0.06)
		GAA	14 (0.04)		AGA	14 (0.02)
					CGG	102 (0.18)
					CGA	49 (0.08)
25	Phe	TTT	89 (0.29)		CGT	51 (0.09)
		TTC	216 (0.71)		CGC	331 (0.57)
30	Gly	GGG	92 (0.12)	Ser	AGT	16 (0.03)
		GGA	56 (0.07)		AGC	123 (0.22)
		GGT	76 (0.10)		TCG	152 (0.28)
		GGC	559 (0.71)		TCA	31 (0.06)
35	His	CAT	42 (0.21)		TCT	55 (0.10)
		CAC	154 (0.79)		TCC	173 (0.31)
40	Ile	ATA	4 (0.01)	Thr	ACG	184 (0.38)
		ATT	30 (0.08)		ACA	24 (0.05)
		ATC	338 (0.91)		ACT	21 (0.05)
					ACC	249 (0.52)
45	Lys	AAG	284 (0.98)	Val	GTG	308 (0.50)
		AAA	7 (0.02)		GTA	9 (0.01)
					GTT	35 (0.06)
					GTC	262 (0.43)
	Leu	TTG	26 (0.04)	Trp	TGG	107 (1.00)
		TTA	3 (0.00)			
		CTG	447 (0.61)			
		CTA	20 (0.03)			
		CTT	45 (0.06)	Tyr	TAT	10 (0.05)
		CTC	190 (0.26)		TAC	180 (0.95)
	Met	ATG	191 (1.00)	Stop	TGA/TAG/TAA	

[0198] In other embodiments, the gene in the recombinant vector has been codon-optimized with reference to a microalgal strain other than a *Prototheca* strain. For example, methods of recoding genes for expression in microalgae are described in U.S.

Patent 7,135,290. Additional information for codon optimization is available, e.g., at the codon usage database of GenBank.

[0199] While the methods and materials of the invention allow for the introduction of any exogenous gene into *Prototheca*, genes relating to sucrose utilization and lipid pathway modification are of particular interest, as discussed in the following sections.

C. Selectable Markers

Sucrose Utilization

[0200] In one embodiment, the recombinant *Prototheca* cell of the invention further contains one or more exogenous sucrose utilization genes. In various embodiments, the one or more genes encode one or more proteins selected from the group consisting of a fructokinase, a glucokinase, a hexokinase, a sucrose invertase, a sucrose transporter. For example, expression of a sucrose transporter and a sucrose invertase allows *Prototheca* to transport sucrose into the cell from the culture media and hydrolyze sucrose to yield glucose and fructose. Optionally, a fructokinase may be expressed as well in instances where endogenous hexokinase activity is insufficient for maximum phosphorylation of fructose. Examples of suitable sucrose transporters are GenBank accession numbers CAD91334, CAB92307, and CAA53390. Examples of suitable fructokinases are GenBank accession numbers P26984, P26420 and CAA43322.

[0201] In one embodiment, the present invention provides a *Prototheca* host cell that secretes a sucrose invertase. Secretion of a sucrose invertase obviates the need for expression of a transporter that can transport sucrose into the cell. This is because a secreted invertase catalyzes the conversion of a molecule of sucrose into a molecule of glucose and a molecule of fructose, both of which may be transported and utilized by microbes provided by the invention. For example, expression of a sucrose invertase with a secretion signal (such as that of SEQ ID NO: 12 (from yeast), SEQ ID NO: 13 (from higher plants), SEQ ID NO: 14 (eukaryotic consensus secretion signal), and SEQ ID NO: 15 (combination of signal sequence from higher plants and eukaryotic consensus) generates invertase activity outside the cell. Expression of such a protein, as enabled by the genetic engineering methodology disclosed herein, allows cells already capable of utilizing extracellular glucose as an energy source to utilize sucrose as an extracellular energy source.

[0202] *Prototheca* species expressing an invertase in media containing sucrose are a preferred microalgal species for the production of oil. The expression and extracellular

targeting of this fully active protein allows the resulting host cells to grow on sucrose, whereas their non-transformed counterparts cannot. Thus, the present invention provides *Prototheca* recombinant cells with a codon-optimized invertase gene (SEQ ID NO: 16), including but not limited to the yeast invertase gene, integrated into their genome such that the invertase gene is expressed as assessed by invertase activity and sucrose hydrolysis.

[0203] Examples of suitable sucrose invertases include those identified by GenBank accession numbers CAB95010, NP_012104 (SEQ ID NO: 17), and CAA06839. Non-limiting examples of suitable invertases include those described in PCT Publication No. WO 2010/063032, incorporated herein by reference.

[0204] The secretion of an invertase to the culture medium by *Prototheca* enable the cells to grow as well on waste molasses from sugar cane processing as they do on pure reagent-grade glucose; the use of this low-value waste product of sugar cane processing can provide significant cost savings in the production of lipids and other oils. Thus, the present invention provides a microbial culture containing a population of *Prototheca* microorganisms, and a culture medium comprising (i) sucrose and (ii) a sucrose invertase enzyme. In various embodiments the sucrose in the culture comes from sorghum, sugar beet, sugar cane, molasses, or depolymerized cellulosic material (which may optionally contain lignin). In another aspect, the methods and reagents of the invention significantly increase the number and type of feedstocks that may be utilized by recombinant *Prototheca*. While the microbes exemplified here are altered such that they can utilize sucrose, the methods and reagents of the invention may be applied so that feedstocks such as cellulotics are utilizable by an engineered host microbe of the invention with the ability to secrete cellulases, pectinases, isomerases, or the like, such that the breakdown products of the enzymatic reactions are no longer just simply tolerated but rather utilized as a carbon source by the host.

D. Sequence Determination

[0205] A variety of methods may be employed for the identification of gene sequences and amino acid sequences of lipid biosynthetic pathway genes and enzymes. Sequences of polynucleotides (e.g., genomic DNA, cDNA, RNA, PCR-amplified nucleotides) may be determined through sequencing technologies including but not limited to Sanger sequencing, pyrosequencing, sequencing by synthesis, sequencing by oligonucleotide probe ligation, and real time sequencing. One skilled in the art may compare nucleotide sequences to published databases of genomic sequences or expressed sequences. Where a

DNA sequence is determined or disclosed, one skilled in the art may compare segments from published exon sequences, or may assemble exon sequences into a reconstructed sequence that does not contain intronic sequences. Sequences of polynucleotides may also be translated into amino acids, peptides, polypeptides or proteins through a variety of methods including but not limited to manual translation or computer-automated translation with bioinformatics software commonly known in the art. Comparison methods of sequenced DNA, RNA, amino acids, peptides, or proteins may include but are not limited to manual evaluation of the sequence or computer-automated sequence comparison and identification using algorithms such as BLAST (Basic Local Alignment Search Tool; Altschul, S. E, et al, (1993) /. *Mol. Biol.* 215:403-410; see also www.ncbi.nlm.nih.gov/BLAST/).

[0206] The instant specification teaches partial or complete amino acid and nucleotide sequences encoding one or more particular microbial genes and proteins. The skilled artisan, having the benefit of the sequences as reported herein, may now use all or a portion of the disclosed sequences for purposes known to those skilled in this art.

[0207] Genomic sequencing of *P. moriformis* (UTEX 1435) was performed using Illumina HiSeq and paired-end reads were obtained (100bp reads, ~450bp fragment size). Genomic DNA is prepared using standard protocols and fragmented by hydrodynamic shearing. Genomic data using Roche 454 technology was also obtained (400 bp fragment size) as were 8kb mate pair libraries. Transcriptome data consisted of Illumina HiSeq paired end data (100bp reads, ~450bp fragment size). Sequencing reads were quality trimmed and filtered using fastx tools. Genome data was assembled using Velvet (Zerbino et al, Velvet: algorithms for de novo short read assembly using de Bruijn graphs, *Genome Research*, May 2008) using an optimized kmer and other default parameters, and using the Pacific Biosciences ALLORA assembler. Annotation was performed using the MAKER pipeline and genes were identified by BLAST against the nr database (NCBI). In addition, a cDNA library was constructed, and ~1200 cDNAs from this library were sequenced using Sanger sequencing. These cDNAs were then annotated and used as initial entry points to find transcripts from Illumina transcriptomes and to help verify the accuracy of Illumina transcriptome and genome sequence data. Pacific Biosciences technology was also utilized to obtain long reads from genomic samples. A given base position is indicated with a code as shown in the table below. In some cases, two representative sequences were constructed, with the two possible bases respectively, as well as

corresponding protein translations for the two alternatives. Such sequences are referred to as “version 1” and “version 2”.

Base codes

A	Adenine
C	Cytosine
G	Guanine
T (or U)	Thymine (or Uracil)
R	A or G
Y	C or T
S	G or C
W	A or T
K	G or T
M	A or C
B	C or G or T
D	A or G or T
H	A or C or T
V	A or C or G
N	any base
. or -	gap
*	stop / nonsense codon
?	unknown amino acid

5 SECTION IV-II. GENETICALLY ENGINEERED *PROTOTHECA* CELLS

[0208] In a first aspect, the present invention provides a genetically engineered *Prototheca* cell in one or more lipid biosynthesis genes have been modified to increase or decrease expression of such one or more genes such that the fatty acid profile of the genetically engineered strain differs from that of the strain from which it was derived. In one embodiment, at least two genes have been modified. In various embodiments, the genetic modifications include one or more of the following modifications: (i) attenuation of a gene or its enzymatic product; and (ii) increased expression of a gene or its enzymatic product; (iii) altered activity of a gene or its enzymatic product.

[0209] In various embodiments, the genetically engineered cell has one or more attenuated genes, wherein the genes attenuated have been attenuated by a means selected from the group consisting of a homologous recombination event and introduction of an exogenous gene that codes for an interfering RNA. In various embodiments, one or more alleles of a gene are attenuated.

[0210] In various embodiments, the genetically engineered cell has one or more over-expressed genes, wherein the genes over-expressed have been up-regulated by a means selected from the group consisting of introduction of additional copies of said gene into said cell; introduction of new expression control elements for said gene; and alteration of

the protein-coding sequence of the gene. In various embodiments, one or more alleles of a gene are over-expressed.

[0211] In various embodiments, the modified genes of the genetically engineered cell are selected from the group consisting of *Prototheca* lipid biosynthesis genes presented in Table 1. In various embodiments, the genetically engineered cell comprises an exogenous gene selected from the group consisting of *Prototheca* lipid biosynthesis genes presented in Table 1. In various embodiments, the genetically engineered cell comprises one more over-expressed alleles of a gene, the gene selected from the group consisting of *Prototheca* lipid biosynthesis genes presented in Table 1. In various embodiments, the genetically engineered cell has an attenuated gene selected from the group consisting of *Prototheca* lipid biosynthesis genes presented in Table 1. In various embodiments, the genetically engineered cell has one more attenuated alleles of a gene, the gene selected from the group consisting of *Prototheca* lipid biosynthesis genes presented in Table 1.

[0212] In various embodiments, the genetically engineered cell has one or more overexpressed genes, wherein the expression of the genes have been increased by a means selected from the group consisting of introduction of additional copies of said gene into said cell; and introduction of new expression control elements for said gene. In various embodiments, the overexpressed gene is an exogenous gene.

[0213] In various embodiments, the modified genes of the genetically engineered cell are selected from the group consisting of *Prototheca* lipid biosynthesis genes presented in Table 1.

[0214] In various embodiments, the genetically engineered cell has an up-regulated gene selected from the group consisting of *Prototheca* lipid biosynthesis genes presented in Table 1. In various embodiments, the genetically engineered cell has an attenuated gene selected from the group consisting of *Prototheca* lipid biosynthesis genes presented in Table 1. In various embodiments, the genetically engineered cell has a fatty acid profile selected from the group consisting of: 3% to 60% C8:0, 3% to 60% C10:0, 3% to 70% C12:0, 3% to 95% C14:0, 3% to 95% C16:0, 3% to 95% C18:0, 3% to 95% C18:1, 3% to 60% C18:2, 1% to 60% C18:3 or combinations thereof. In various embodiments, the ratio of C10:0 to C12:0 is at least 3:1. In various embodiments, the ratio of C12:0 to C14:0 is at least 3:1. In some cases, the ratio of C10:0 to C14:0 is at least 10:1. In various embodiments, the genetically engineered cell has a fatty acid profile of at least 40% saturated fatty acids, of at least 60% saturated fatty acids, or at least 85% saturated fatty

acids. In various embodiments, the genetically engineered cell has a fatty acid profile of at least 85% unsaturated fatty acids, of at least 90% unsaturated fatty acids, of at least 95% unsaturated fatty acids, or at least 97% unsaturated fatty acids.

[0215] An embodiment of the present invention also provides recombinant *Prototheca* cells that have been modified to contain one or more exogenous genes encoding lipid biosynthesis enzymes such as, for example, a fatty acyl-ACP thioesterase (see Example 5) or a ketoacyl-ACP synthase II (see Example 6). In some embodiments, genes encoding a fatty acyl-ACP thioesterase and a naturally co-expressed acyl carrier protein are transformed into a *Prototheca* cell, optionally with one or more genes encoding other lipid biosynthesis genes. In other embodiments, the ACP and the fatty acyl-ACP thioesterase may have an affinity for one another that imparts an advantage when the two are used together in the microbes and methods of the present invention, irrespective of whether they are or are not naturally co-expressed in a particular tissue or organism. Thus, embodiments of the present invention contemplate both naturally co-expressed pairs of these enzymes as well as those that share an affinity for interacting with one another to facilitate cleavage of a length-specific carbon chain from the ACP.

[0216] In still other embodiments, an exogenous gene encoding a desaturase is transformed into the *Prototheca* cell in conjunction with one or more genes encoding other lipid biosynthesis genes to provide modifications with respect to lipid saturation. In other embodiments, an endogenous desaturase gene is overexpressed (e.g., through the introduction of additional copies of the gene) in a *Prototheca* cell. In some embodiments, the desaturase may be selected with reference to a desired carbon chain length, such that the desaturase is capable of making location specific modifications within a specified carbon-length substrate, or substrates having a carbon-length within a specified range. In another embodiment, if the desired fatty acid profile is an increase in monounsaturates (such as C16:1 and/or C18:1) overexpression of a SAD or expression of a heterologous SAD may be coupled with the silencing or inactivation (e.g., through mutation, RNAi, antisense, or knockout of an endogenous desaturase gene, etc.) of a fatty acyl desaturase (FAD) or another desaturase gene.

[0217] In other embodiments, the *Prototheca* cell has been modified to have an attenuated endogenous desaturase gene, wherein the attenuation renders the gene or desaturase enzyme inactive. In some cases, the mutated endogenous desaturase gene is a fatty acid desaturase (FAD). In other cases, the mutated endogenous desaturase gene is a

stearoyl acyl carrier protein desaturase (SAD). Example 4 describes the targeted ablation of stearoyl-ACP desaturases and delta 12 fatty acid desaturases. Example 4 also describes the use of RNA antisense constructs to decrease the expression of an endogenous desaturase gene. Example

5 [0218] In some cases, it may be advantageous to pair one or more of the genetic engineering techniques in order to achieve a transgenic cell that produces the desired fatty acid profile. In one embodiment, a *Prototheca* cell comprises an attenuated endogenous thioesterase gene and one or more exogenous gene. In non-limiting examples, a
10 *Prototheca* cell with an attenuated endogenous thioesterase gene can also express an exogenous fatty acyl-ACP thioesterase gene and/or a sucrose invertase gene. Example 5 below describes a transgenic *Prototheca* cell containing a targeted ablation or knockout of an endogenous thioesterase and also expresses a *Cuphea wrightii* FatB2 C10:0-C14:0 preferring thioesterase and a sucrose invertase.

[0219] In other embodiments, one allele of a *Prototheca* lipid biosynthesis gene has
15 been attenuated. In additional embodiments, two or more alleles of a *Prototheca* lipid biosynthesis gene have been attenuated. In additional embodiments, one or more alleles of different *Prototheca* lipid biosynthesis genes have been attenuated. Example 37 below describes the targeted knockout of multiple alleles of stearoyl-ACP desaturase. Example 34 below describes the targeted knockout of multiple alleles of acyl-ACP thioesterase.
20 Example 8 below describes the targeted knockout of acyl-ACP thioesterase and fatty acid desaturase. In some cases, the targeted knockout of different alleles of a gene may result in different effects on fatty acid profiles.

SECTION V. MICROBIAL OILS

25 [0220] For the production of oil in accordance with the methods of the invention, it is preferable to culture cells in the dark, as is the case, for example, when using extremely large (40,000 liter and higher) fermentors that do not allow light to strike the culture. *Prototheca* species are grown and propagated for the production of oil in a medium containing a fixed carbon source and in the absence of light; such growth is known as
30 heterotrophic growth.

[0221] As an example, an inoculum of lipid-producing microalgal cells are introduced into the medium; there is a lag period (lag phase) before the cells begin to propagate. Following the lag period, the propagation rate increases steadily and enters the log, or

exponential, phase. The exponential phase is in turn followed by a slowing of propagation due to decreases in nutrients such as nitrogen, increases in toxic substances, and quorum sensing mechanisms. After this slowing, propagation stops, and the cells enter a stationary phase or steady growth state, depending on the particular environment provided to the cells. For obtaining lipid rich biomass, the culture is typically harvested well after then end of the exponential phase, which may be terminated early by allowing nitrogen or another key nutrient (other than carbon) to become depleted, forcing the cells to convert the carbon sources, present in excess, to lipid. Culture condition parameters may be manipulated to optimize total oil production, the combination of lipid species produced, and/or production of a specific oil.

[0222] As discussed above, a bioreactor or fermentor is used to allow cells to undergo the various phases of their growth cycle. As an example, an inoculum of lipid-producing cells may be introduced into a medium followed by a lag period (lag phase) before the cells begin growth. Following the lag period, the growth rate increases steadily and enters the log, or exponential, phase. The exponential phase is in turn followed by a slowing of growth due to decreases in nutrients and/or increases in toxic substances. After this slowing, growth stops, and the cells enter a stationary phase or steady state, depending on the particular environment provided to the cells. Lipid production by cells disclosed herein can occur during the log phase or thereafter, including the stationary phase wherein nutrients are supplied, or still available, to allow the continuation of lipid production in the absence of cell division.

[0223] Preferably, microorganisms grown using conditions described herein and known in the art comprise at least about 20% by weight of lipid, preferably at least about 40% by weight, more preferably at least about 50% by weight, and most preferably at least about 60% by weight. Process conditions may be adjusted to increase the yield of lipids suitable for a particular use and/or to reduce production cost. For example, in certain embodiments, a microalgae is cultured in the presence of a limiting concentration of one or more nutrients, such as, for example, nitrogen, phosphorous, or sulfur, while providing an excess of fixed carbon energy such as glucose. Nitrogen limitation tends to increase microbial lipid yield over microbial lipid yield in a culture in which nitrogen is provided in excess. In particular embodiments, the increase in lipid yield is at least about: 10%, 50%, 100%, 200%, or 500%. The microbe may be cultured in the presence of a limiting amount of a nutrient for a portion of the total culture period or for the entire period. In particular embodiments, the nutrient concentration is cycled between a limiting concentration and a

non-limiting concentration at least twice during the total culture period. Lipid content of cells may be increased by continuing the culture for increased periods of time while providing an excess of carbon, but limiting or no nitrogen.

[0224] In another embodiment, lipid yield is increased by culturing a lipid-producing microbe (e.g., microalgae) in the presence of one or more cofactor(s) for a lipid pathway enzyme (e.g., a fatty acid synthetic enzyme). Generally, the concentration of the cofactor(s) is sufficient to increase microbial lipid (e.g., fatty acid) yield over microbial lipid yield in the absence of the cofactor(s). In a particular embodiment, the cofactor(s) are provided to the culture by including in the culture a microbe (e.g., microalgae) containing an exogenous gene encoding the cofactor(s). Alternatively, cofactor(s) may be provided to a culture by including a microbe (e.g., microalgae) containing an exogenous gene that encodes a protein that participates in the synthesis of the cofactor. In certain embodiments, suitable cofactors include any vitamin required by a lipid pathway enzyme, such as, for example: biotin, pantothenate. Genes encoding cofactors suitable for use in the invention or that participate in the synthesis of such cofactors are well known and may be introduced into microbes (e.g., microalgae), using constructs and techniques such as those described above.

[0225] The specific examples of bioreactors, culture conditions, and heterotrophic growth and propagation methods described herein may be combined in any suitable manner to improve efficiencies of microbial growth and lipid and/or protein production.

[0226] Microalgal biomass with a high percentage of oil/lipid accumulation by dry weight has been generated using different methods of culture, which are known in the art (see PCT Pub. No. 2008/151149). Microalgal biomass generated by the culture methods described herein and useful in accordance with the present invention comprises at least 2% microalgal oil by dry weight. In some embodiments, the microalgal biomass comprises at least 10%, at least 25%, at least 50%, at least 55%, or at least 60% microalgal oil by dry weight. In some embodiments, the microalgal biomass contains from 10-90% microalgal oil, from 25-75%.

[0227] The microalgal oil of the biomass described herein, or extracted from the biomass for use in the methods and compositions of the present invention can comprise glycerolipids with one or more distinct fatty acid ester side chains. Glycerolipids are comprised of a glycerol molecule esterified to one, two or three fatty acid molecules, which may be of varying lengths and have varying degrees of saturation. The length and

saturation characteristics of the fatty acid molecules (and the microalgal oils) may be manipulated to modify the properties or proportions of the fatty acid molecules in the microalgal oils of the present invention via culture conditions or via lipid pathway engineering, as described in more detail in Section IV, below. Thus, specific blends of algal oil may be prepared either within a single species of algae by mixing together the biomass or algal oil from two or more species of microalgae, or by blending algal oil of the invention with oils from other sources such as soy, rapeseed, canola, palm, palm kernel, coconut, corn, waste vegetable, Chinese tallow, olive, sunflower, cottonseed, chicken fat, beef tallow, porcine tallow, microalgae, macroalgae, microbes, Cuphea, flax, peanut, choice white grease, lard, Camelina sativa, mustard seed, cashew nut, oats, lupine, kenaf, calendula, hemp, coffee, linseed (flax), hazelnut, euphorbia, pumpkin seed, coriander, camellia, sesame, safflower, rice, tung tree, cocoa, copra, pium poppy, castor beans, pecan, jojoba, macadamia, Brazil nuts, avocado, petroleum, or a distillate fraction of any of the preceding oils.

[0228] The oil composition, *i.e.*, the properties and proportions of the fatty acid constituents of the glycerolipids, can also be manipulated by combining biomass or oil from at least two distinct species of microalgae. In some embodiments, at least two of the distinct species of microalgae have different glycerolipid profiles. The distinct species of microalgae may be cultured together or separately as described herein, preferably under heterotrophic conditions, to generate the respective oils. Different species of microalgae can contain different percentages of distinct fatty acid constituents in the cell's glycerolipids

[0229] Generally, *Prototheca* strains have very little or no fatty acids with the chain length C8-C14. For example, *Prototheca moriformis* (UTEX 1435), *Prototheca krugani* (UTEX 329), *Prototheca stagnora* (UTEX 1442) and *Prototheca zopfii* (UTEX 1438) contains no (or undetectable amounts) C8 fatty acids, between 0-0.01% C10 fatty acids, between 0.03-2.1% C12 fatty acids and between 1.0-1.7% C14 fatty acids.

[0230] Microalgal oil can also include other constituents produced by the microalgae, or incorporated into the microalgal oil from the culture medium. These other constituents may be present in varying amount depending on the culture conditions used to culture the microalgae, the species of microalgae, the extraction method used to recover microalgal oil from the biomass and other factors that may affect microalgal oil composition. Non-limiting examples of such constituents include carotenoids, present from 0.025-0.3 mcg/g,

preferably from 0.05 to 0.244 micrograms/gram, of oil; chlorophyll A present from 0.025-0.3 mcg/g, preferably from 0.045 to 0.268 micrograms/gram, of oil; total chlorophyll of less than 0.03 mcg/g, preferably less than 0.025 micrograms/gram, of oil; gamma tocopherol present from 35-175 mcg/g, preferably from 38.3-164 micrograms/gram, of oil; total tocopherols present from 50-300 mcg/g, preferably from 60.8 to 261.7 microgram/gram, of oil; less than 0.5%, preferably less than 0.25%, brassicasterol, campesterol, stigmasterol, or betasitosterol; total tocotrienols less than 300 micrograms/gram of oil; and total tocotrienols present from 225-350 mcg/g, preferably from 249.6 to 325.3 micrograms/gram, of oil.

[0231] Other constituents can include, without limitation, phospholipids, tocopherols, tocotrienols, carotenoids (e.g., alpha-carotene, beta-carotene, lycopene, etc.), xanthophylls (e.g., lutein, zeaxanthin, alpha-cryptoxanthin and beta-cryptoxanthin), and various organic or inorganic compounds. In some cases, the oil extracted from *Prototheca* species comprises between 0.001 to 0.05, preferably from 0.003 to 0.039, microgram lutein/gram of oil, less than 0.005, preferably less than 0.003, micrograms lycopene/gram of oil; and less than 0.005, preferably less than 0.003, microgram beta carotene/gram of oil.

[0232] The stable carbon isotope value $\delta^{13}\text{C}$ is an expression of the ratio of $^{13}\text{C}/^{12}\text{C}$ relative to a standard (e.g. PDB, carbonite of fossil skeleton of *Belemnite americana* from Peedee formation of South Carolina). The stable carbon isotope value $\delta^{13}\text{C}$ ($^{\circ}/_{\text{oo}}$) of the oils can be related to the $\delta^{13}\text{C}$ value of the feedstock used. In some embodiments the oils are derived from oleaginous organisms heterotrophically grown on sugar derived from a C_4 plant such as corn or sugarcane. In some embodiments the $\delta^{13}\text{C}$ ($^{\circ}/_{\text{oo}}$) of the oil is from -10 to -17 $^{\circ}/_{\text{oo}}$ or from -13 to -16 $^{\circ}/_{\text{oo}}$.

[0233] In a second aspect, the present invention provides methods for obtaining microbial oil comprising culturing a genetically engineered *Prototheca* cell of the invention under conditions such that oil is produced. In various embodiments, the microbial oil has a fatty acid profile selected from the group consisting of: 0.1% to 60% $\text{C}_8:0$, 0.1% to 60% $\text{C}_{10:0}$, 0.1% to 60% $\text{C}_{12:0}$, 0.1% to 95% $\text{C}_{14:0}$, 1% to 95% $\text{C}_{16:0}$, 0.1% to 95% $\text{C}_{18:0}$, 0.1% to 95% $\text{C}_{18:1}$, 0.1% to 60% $\text{C}_{18:2}$, 0.1% to 60% $\text{C}_{18:3}$ or combinations thereof. In various embodiments, the ratio of $\text{C}_{10:0}$ to $\text{C}_{12:0}$ is at least 3:1. In some cases, the ratio of $\text{C}_{10:0}$ to $\text{C}_{14:0}$ is at least 10:1. In some cases, the ratio of $\text{C}_{10:0}$ to $\text{C}_{14:0}$ is at least 2:1. In some cases, the ratio of $\text{C}_{12:0}$ to $\text{C}_{14:0}$ is at least 2:1. In some cases, the ratio of $\text{C}_{12:0}$ to $\text{C}_{14:0}$ is at least 3:1. In various embodiments, the genetically engineered cell has a fatty acid profile of at least 40% saturated fatty acids, of

at least 60% saturated fatty acids, or at least 85% saturated fatty acids. In various embodiments, the genetically engineered cell has a fatty acid profile of at least 85% unsaturated fatty acids, of at least 90% unsaturated fatty acids, of at least 95% unsaturated fatty acids, or at least 97% unsaturated fatty acids.

- 5 [0234] In a third aspect, the present invention provides microbial oils and foods, fuels, and chemicals containing said oil or a chemical derived therefrom.

SECTION VI. NUCLEIC ACIDS

- 10 [0235] In a fifth aspect, the present invention provides recombinant nucleic acids useful in methods for making genetically modified *Prototheca* and other cells. Embodiments of the present invention provide polynucleotides that encode some portion of a coding sequence of a *Prototheca* lipid biosynthesis gene.

- 15 [0236] In various embodiments, these nucleic acids include expression cassettes, which consist of a coding sequence and control sequences that regulate expression of the coding sequence, which may code for an mRNA that encodes a lipid biosynthesis enzyme or for an RNAi that acts to suppress expression of a fatty acid biosynthesis gene.

- 20 [0237] In other embodiments, these nucleic acids are expression vectors that include one or more expression cassettes and stably replicate in a *Prototheca* or other host cell, either by integration into chromosomal DNA of the host cell or as freely replicating plasmids.

- 25 [0238] In other embodiments, these nucleic acids comprise a portion of a *Prototheca* lipid biosynthesis gene, which portion may be a portion of a coding sequence, an exon, or a control element. Such nucleic acids are useful in the construction of expression cassettes for *Prototheca* and non-*Prototheca* host cells, for integration of exogenous DNA into *Prototheca* host cells, for regulating expression of exogenous genes expressed in *Prototheca* and non-*Prototheca* host cells, and for construction of nucleic acids useful for inactivating *Prototheca* lipid synthesis genes such as through homologous recombination or antisense RNA mediated knockdown.

30 EXAMPLES

EXAMPLE 1: METHODS FOR CULTURING *PROTOTHECA*

- [0239] *Prototheca* strains were cultivated to achieve a high percentage of oil by dry cell weight. Cryopreserved cells were thawed at room temperature and 500 ul of cells were

added to 4.5 ml of medium (4.2 g/L K_2HPO_4 , 3.1 g/L NaH_2PO_4 , 0.24 g/L $MgSO_4 \cdot 7H_2O$, 0.25 g/L Citric Acid monohydrate, 0.025 g/L $CaCl_2 \cdot 2H_2O$, 2g/L yeast extract) plus 2% glucose and grown for 7 days at 28°C with agitation (200 rpm) in a 6-well plate. Dry cell weights were determined by centrifuging 1 ml of culture at 14,000 rpm for 5 min in a pre-weighed Eppendorf tube. The culture supernatant was discarded and the resulting cell pellet washed with 1 ml of deionized water. The culture was again centrifuged, the supernatant discarded, and the cell pellets placed at -80°C until frozen. Samples were then lyophilized for 24 hrs and dry cell weights calculated. For determination of total lipid in cultures, 3 ml of culture was removed and subjected to analysis using an Ankom system (Ankom Inc., Macedon, NY) according to the manufacturer's protocol. Samples were subjected to solvent extraction with an Amkom XT10 extractor according to the manufacturer's protocol. Total lipid was determined as the difference in mass between acid hydrolyzed dried samples and solvent extracted, dried samples. Percent oil dry cell weight measurements are shown in Table 3.

Table 3. Percent oil by dry cell weight

Species	Strain	% Oil
<i>Prototheca stagnora</i>	UTEX 327	13.14
<i>Prototheca moriformis</i>	UTEX 1441	18.02
<i>Prototheca moriformis</i>	UTEX 1435	27.17

[0240] Microalgae samples from multiple strains from the genus *Prototheca* were genotyped. Genomic DNA was isolated from algal biomass as follows. Cells (approximately 200 mg) were centrifuged from liquid cultures 5 minutes at 14,000 x g.

Cells were then resuspended in sterile distilled water, centrifuged 5 minutes at 14,000 x g and the supernatant discarded. A single glass bead ~2mm in diameter was added to the biomass and tubes were placed at -80°C for at least 15 minutes. Samples were removed and 150 µl of grinding buffer (1% Sarkosyl, 0.25 M Sucrose, 50 mM NaCl, 20 mM EDTA, 100 mM Tris-HCl, pH 8.0, RNase A 0.5 ug/ul) was added. Pellets were resuspended by vortexing briefly, followed by the addition of 40 µl of 5M NaCl. Samples were vortexed briefly, followed by the addition of 66 µl of 5% CTAB (Cetyl trimethylammonium bromide) and a final brief vortex. Samples were next incubated at 65°C for 10 minutes after which they were centrifuged at 14,000 x g for 10 minutes. The supernatant was transferred to a fresh tube and extracted once with 300 µl of Phenol:Chloroform:Isoamyl alcohol 12:12:1, followed by centrifugation for 5 minutes at 14,000 x g. The resulting aqueous phase was transferred to a fresh tube containing 0.7 vol

of isopropanol (~190 µl), mixed by inversion and incubated at room temperature for 30 minutes or overnight at 4°C. DNA was recovered via centrifugation at 14,000 x g for 10 minutes. The resulting pellet was then washed twice with 70% ethanol, followed by a final wash with 100% ethanol. Pellets were air dried for 20-30 minutes at room temperature

5 followed by resuspension in 50 µl of 10mM TrisCl, 1mM EDTA (pH 8.0).

[0241] Five µl of total algal DNA, prepared as described above, was diluted 1:50 in 10 mM Tris, pH 8.0. PCR reactions, final volume 20 µl, were set up as follows. Ten µl of 2 x iProof HF master mix (BIO-RAD) was added to 0.4 µl primer SZ02613 (5'-

TGTTGAAGAATGAGCCGGCGAC-3') at 10mM stock concentration. This primer

10 sequence runs from position 567-588 in GenBank accession no. L43357 and is highly conserved in higher plants and algal plastid genomes. This was followed by the addition of 0.4 µl primer SZ02615 (5'-CAGTGAGCTATTACGCACTC-3') at 10 mM stock concentration. This primer sequence is complementary to position 1112-1093 in GenBank accession no. L43357 and is highly conserved in higher plants and algal plastid genomes.

15 Next, 5 µl of diluted total DNA and 3.2 µl dH₂O were added. PCR reactions were run as follows: 98°C, 45''; 98°C, 8''; 53°C, 12''; 72°C, 20'' for 35 cycles followed by 72°C for 1 min and holding at 25°C. For purification of PCR products, 20 µl of 10 mM Tris, pH 8.0, was added to each reaction, followed by extraction with 40 µl of

20 Phenol:Chloroform:isoamyl alcohol 12:12:1, vortexing and centrifuging at 14,000 x g for 5 minutes. PCR reactions were applied to S-400 columns (GE Healthcare) and centrifuged for 2 minutes at 3,000 x g. Purified PCR products were subsequently TOPO cloned into PCR8/GW/TOPO and positive clones selected for on LB/Spec plates. Purified plasmid DNA was sequenced in both directions using M13 forward and reverse primers. In total, twelve *Prototheca* strains were selected to have their 23S rRNA DNA sequenced and the

25 sequences are listed in the Sequence Listing. A summary of the strains and Sequence Listing Numbers is included below. The sequences were analyzed for overall divergence from the UTEX 1435 (SEQ ID NO: 5) sequence. Two pairs emerged (UTEX 329/UTEX 1533 and UTEX 329/UTEX 1440) as the most divergent. In both cases, pairwise alignment resulted in 75.0% pairwise sequence identity. The percent sequence identity to

30 UTEX 1435 is also included below:

<u>Species</u>	<u>Strain</u>	<u>% nt identity</u>	<u>SEQ ID NO.</u>
<i>Prototheca kruegani</i>	UTEX 329	75.2	SEQ ID NO: 1
<i>Prototheca wickerhamii</i>	UTEX 1440	99	SEQ ID NO: 2
35 <i>Prototheca stagnora</i>	UTEX 1442	75.7	SEQ ID NO: 3
<i>Prototheca moriformis</i>	UTEX 288	75.4	SEQ ID NO: 4

	<i>Prototheca moriformis</i>	UTEX 1439; 1441; 1435; 1437	100	SEQ ID NO: 5
	<i>Prototheca wikerhamii</i>	UTEX 1533	99.8	SEQ ID NO: 6
	<i>Prototheca moriformis</i>	UTEX 1434	75.9	SEQ ID NO: 7
5	<i>Prototheca zopfii</i>	UTEX 1438	75.7	SEQ ID NO: 8
	<i>Prototheca moriformis</i>	UTEX 1436	88.9	SEQ ID NO: 9

[0242] Lipid samples from a subset of the above-listed strains were analyzed for fatty acid profile using HPLC. Results are shown below in Table 4.

10 **Table 4. Diversity of fatty acid chains in *Prototheca* species**

Strain	C14:0	C16:0	C16:1	C18:0	C18:1	C18:2	C18:3	C20:0	C20:1
UTEX 327	0	12.01	0	0	50.33	17.14	0	0	0
UTEX 1441	1.41	29.44	0.70	3.05	57.72	12.37	0.97	0.33	0
UTEX 1435	1.09	25.77	0	2.75	54.01	11.90	2.44	0	0

[0243] Oil extracted from *Prototheca moriformis* UTEX 1435 (via solvent extraction or using an expeller press) was analyzed for carotenoids, chlorophyll, tocopherols, other sterols and tocotrienols. The results are summarized below in Table 5.

15 **Table 5. Carotenoid, chlorophyll, tocopherol/sterols and tocotrienol analysis in oil extracted from *Prototheca moriformis* (UTEX 1435).**

	Pressed oil (mcg/ml)	Solvent extracted oil (mcg/ml)
cis-Lutein	0.041	0.042
trans-Lutein	0.140	0.112
trans-Zeaxanthin	0.045	0.039
cis-Zeaxanthin	0.007	0.013
t-alpha-Cryptoxanthin	0.007	0.010
t-beta-Cryptoxanthin	0.009	0.010
t-alpha-Carotene	0.003	0.001
c-alpha-Carotene	none detected	none detected
t-beta-Carotene	0.010	0.009
9-cis-beta-Carotene	0.004	0.002
Lycopene	none detected	none detected
Total Carotenoids	0.267	0.238
Chlorophyll	<0.01 mg/kg	<0.01 mg/kg
Tocopherols and Sterols		
	Pressed oil (mg/100g)	Solvent extracted oil (mg/100g)
gamma Tocopherol	0.49	0.49
Campesterol	6.09	6.05
Stigmasterol	47.6	47.8
Beta-sitosterol	11.6	11.5
Other sterols	445	446
Tocotrienols		
	Pressed oil (mg/g)	Solvent extracted oil (mg/g)
alpha Tocotrienol	0.26	0.26
beta Tocotrienol	<0.01	<0.01
gamma Tocotrienol	0.10	0.10

detal Tocotrienol	<0.01	<0.01
Total Tocotrienols	0.36	0.36

[0244] Oil extracted from *Prototheca moriformis*, from four separate lots, was refined and bleached using standard vegetable oil processing methods. Briefly, crude oil extracted from *Prototheca moriformis* was clarified in a horizontal decanter, where the solids were separated from the oil. The clarified oil was then transferred to a tank with citric acid and water and left to settle for approximately 24 hours. After 24 hours, the mixture in the tank formed 2 separate layers. The bottom layer was composed of water and gums that were then removed by decantation prior to transferring the degummed oil into a bleaching tank. The oil was then heated along with another dose of citric acid. Bleaching clay was then added to the bleaching tank and the mixture was further heated under vacuum in order to evaporate off any water that was present. The mixture was then pumped through a leaf filter in order to remove the bleaching clay. The filtered oil was then passed through a final 5µm polishing filter and then collected for storage until use. The refined and bleached (RB) oil was then analyzed for carotenoids, chlorophyll, sterols, tocotrienols and tocopherols. The results of these analyses are summarized in Table 6 below. “Nd” denotes none detected and the sensitivity of detection is listed below:

Sensitivity of Detection

Carotenoids (mcg/g) nd = <0.003 mcg/g

Chlorophyll (mcg/g) nd = <0.03 mcg/g

Sterols (%) nd = 0.25%

Tocopherols (mcg/g); nd = 3 mcg/g

Table 6. Carotenoid, chlorophyll, sterols, tocotrienols and tocopherol analysis from refined and bleached *Prototheca moriformis* oil.

	Lot A	Lot B	Lot C	Lot D
Carotenoids (mcg/g)				
Lutein	0.025	0.003	nd	0.039
Zeaxanthin	nd	nd	nd	nd
cis-Lutein/Zeaxanthin	nd	nd	nd	nd
trans-alpha-Cryptoxanthin	nd	nd	nd	nd
trans-beta-Cryptoxanthin	nd	nd	nd	nd
trans-alpha-Carotene	nd	nd	nd	nd
cis-alpha-Carotene	nd	nd	nd	nd
trans-beta-Carotene	nd	nd	nd	nd
cis-beta-Carotene	nd	nd	nd	nd
Lycopene	nd	nd	nd	nd
Unidentified	0.219	0.066	0.050	0.026
Total Carotenoids	0.244	0.069	0.050	0.065
Chlorophyll (mcg/g)				
Chlorophyll A	0.268	0.136	0.045	0.166
Chlorophyll B	nd	nd	nd	nd
Total Chlorophyll	0.268	0.136	0.045	0.166

Sterols (%)				
Brassicasterol	nd	nd	nd	nd
Campesterol	nd	nd	nd	nd
Stigmasterol	nd	nd	nd	nd
beta-Sitosterol	nd	nd	nd	nd
Total Sterols	nd	nd	nd	nd
Tocopherols (mcg/g)				
alpha-Tocopherol	23.9	22.8	12.5	8.2
beta-Tocopherol	3.72	nd	nd	nd
gamma-Tocopherol	164	85.3	43.1	38.3
delta-Tocopherol	70.1	31.1	18.1	14.3
Total Tocopherols	262	139.2	73.7	60.8
Tocotrienols (mcg/g)				
alpha-Tocotrienol	190	225	253	239
beta-Tocotrienol	nd	nd	nd	nd
gamma-Tocotrienol	47.3	60.4	54.8	60.9
delta-Tocotrienol	12.3	16.1	17.5	15.2
Total Tocotrienols	250	302	325	315

[0245] The same four lots of *Prototheca moriformis* oil was also analyzed for trace elements and the results are summarized below in Table 7.

Table 7. Elemental analysis of refined and bleached *Prototheca moriformis* oil.

	Lot A	Lot B	Lot C	Lot D
Elemental Analysis (ppm)				
Calcium	0.08	0.07	< 0.04	0.07
Phosphorous	< 0.2	0.38	< 0.2	0.33
Sodium	< 0.5	0.55	< 0.5	< 0.5
Potassium	1.02	1.68	< 0.5	0.94
Magnesium	< 0.04	< 0.04	< 0.04	0.07
Manganese	< 0.05	< 0.05	< 0.05	< 0.05
Iron	< 0.02	< 0.02	< 0.02	< 0.02
Zinc	< 0.02	< 0.02	< 0.02	< 0.02
Copper	< 0.05	< 0.05	< 0.05	< 0.05
Sulfur	2.55	4.45	2.36	4.55
Lead	< 0.2	< 0.2	< 0.2	< 0.2
Silicon	0.37	0.41	0.26	0.26
Nickel	< 0.2	< 0.2	< 0.2	< 0.2
Organic chloride	< 1.0	< 1.0	< 1.0	2.2
Inorganic chloride	< 1.0	< 1.0	< 1.0	< 1.0
Nitrogen	4.4	7.8	4.2	6.9
Lithium	< 0.02	< 0.02	< 0.02	< 0.02
Boron	0.07	0.36	0.09	0.38
Aluminum	--	< 0.2	< 0.2	< 0.2
Vanadium	< 0.05	< 0.05	< 0.05	< 0.05
Lovibond Color (°L)				
Red	5.0	4.3	3.2	5.0
Yellow	70.0	70.0	50.0	70.0
Mono & Diglycerides by HPLC (%)				
Diglycerides	1.68	2.23	1.25	1.61
Monoglycerides	0.03	0.04	0.02	0.03
Free fatty acids (FFA)	1.02	1.72	0.86	0.83
Soaps	0	0	0	
Oxidized and Polymerized Triglycerides				

Oxidized Triglycerides (%)	3.41	2.41	4.11	1.00
Polymerized Triglycerides (%)	1.19	0.45	0.66	0.31
Peroxide Value (meg/kg)	0.75	0.80	0.60	1.20
p-Anisidine value (dimensionless)	5.03	9.03	5.44	20.1
Water and Other Impurities (%)				
Karl Fisher Moisture	0.8	0.12	0.07	0.18
Total polar compounds	5.02	6.28	4.54	5.23
Unsaponifiable matter	0.92	1.07	0.72	1.04
Insoluble impurities	<0.01	<0.01	0.01	< 0.01
Total oil (%)				
Neutral oil	98.8	98.2	99.0	98.9

EXAMPLE 2: GENERAL METHODS FOR BIOLISTIC TRANSFORMATION OF *PROTOTHECA*

[0246] Seashell Gold Microcarriers (550 nanometers) were prepared according to the protocol from manufacturer. Plasmid (20 µg) was mixed with 50 µl of binding buffer and 60 µl (30 mg) of S550d gold carriers and incubated in ice for 1 min. Precipitation buffer (100 µl) was added, and the mixture was incubated in ice for another 1 min. After vortexing, DNA-coated particles were pelleted by spinning at 10,000 rpm in an Eppendorf 5415C microfuge for 10 seconds. The gold pellet was washed once with 500 µl of cold 100% ethanol, pelleted by brief spinning in the microfuge, and resuspended with 50 µl of ice-cold ethanol. After a brief (1-2 sec) sonication, 10 µl of DNA-coated particles were immediately transferred to the carrier membrane.

[0247] *Prototheca* strains were grown in proteose medium (2g/L yeast extract, 2.94mM NaNO₃, 0.17mM CaCl₂•2H₂O, 0.3mM MgSO₄•7H₂O, 0.4mM K₂HPO₄, 1.28mM KH₂PO₄, 0.43mM NaCl) with 2% glucose on a gyratory shaker until it reaches a cell density of 2x10⁶ cells/ml. The cells were harvested, washed once with sterile distilled water, and resuspended in 50 µl of medium. 1 x 10⁷ cells were spread in the center third of a non-selective proteose media plate. The cells were bombarded with the PDS-1000/He Biolistic Particle Delivery system (Bio-Rad). Rupture disks (1350 psi) were used, and the plates are placed 6 cm below the screen/macrocarrier assembly. The cells were allowed to recover at 25°C for 12–24 h. Upon recovery, the cells were scraped from the plates with a rubber spatula, mixed with 100 µl of medium and spread on plates containing the appropriate antibiotic selection. After 7-10 days of incubation at 25°C, colonies representing transformed cells were visible on the plates. Colonies were picked and

spotted on selective (either antibiotic or carbon source) agar plates for a second round of selection.

EXAMPLE 3: FATTY ACID ANALYSIS BY FATTY ACID METHYL ESTER

5 DETECTION

[0248] Lipid samples were prepared from dried biomass. 20-40 mg of dried biomass was resuspended in 2 mL of 5% H₂SO₄ in MeOH, and 200 ul of toluene containing an appropriate amount of a suitable internal standard (C19:0) was added. The mixture was sonicated briefly to disperse the biomass, then heated at 70 -75°C for 3.5 hours. 2 mL of
10 heptane was added to extract the fatty acid methyl esters, followed by addition of 2 mL of 6% K₂CO₃ (aq) to neutralize the acid. The mixture was agitated vigorously, and a portion of the upper layer was transferred to a vial containing Na₂SO₄ (anhydrous) for gas chromatography analysis using standard FAME

15 EXAMPLE 4: ALTERING THE LEVELS OF SATURATED FATTY ACIDS IN THE MICROALGAE *PROTOTHECA MORIFORMIS*

A. DECREASING STEAROYL ACP DESATURASE AND DELTA 12 FATTY ACID DESATURASE EXPRESSION BY A GENE KNOCKOUT APPROACH

[0249] As part of a genomics screen using a bioinformatics based approach based on
20 cDNAs, Illumina transcriptome and Roche 454 sequencing of genomic DNA from *Prototheca moriformis* (UTEX 1435), two specific groups of genes involved in fatty acid desaturation were identified: stearoyl ACP desaturases (SAD) and delta 12 fatty acid desaturases (Δ 12 FAD). Stearoyl ACP desaturase enzymes are part of the lipid synthesis pathway and they function to introduce double bonds into the fatty acyl chains, for
25 example, the synthesis of C18:1 fatty acids from C18:0 fatty acids. Delta 12 fatty acid desaturases are also part of the lipid synthesis pathway and they function to introduce double bonds into already unsaturated fatty acids, for example, the synthesis of C18:2 fatty acids from C18:1 fatty acids. The genes encoding stearoyl ACP desaturases fell into two distinct families. Based on these results, three gene disruption constructs were designed to
30 potentially disrupt multiple gene family members by targeting more highly conserved coding regions within each family of desaturase enzymes.

[0250] Three homologous recombination targeting constructs were designed using: (1) highly conserved portions of the coding sequence of delta 12 fatty acid desaturase (d12FAD) or FAD2 family members and (2) two constructs targeting each of the two

distinct families of SAD (SAD1 and SAD2), each with conserved regions of the coding sequences from each family. This strategy would embed a selectable marker gene into these highly conserved coding regions (targeting multiple family members) rather than a classic gene replacement strategy where the homologous recombination would target

5 flanking regions of the targeted gene.

[0251] All constructs were introduced into the cells by biolistic transformation using the methods described above and constructs were linearized before being shot into the cells. Transformants were selected on sucrose containing plates/media and changes in lipid profile were assayed using the above-described method. Relevant sequences from
10 each of the three targeting constructs are listed below.

Description

SEQ ID NO:

5' sequence from coding region of d12FAD targeting construct

SEQ ID NO: 18

3' sequence from coding region of d12FAD targeting construct

SEQ ID NO: 19

d12FAD targeting construct cDNA sequence

SEQ ID NO: 20

15 5' sequence from coding region of SAD2A

SEQ ID NO: 21

3' sequence from coding region of SAD2A

SEQ ID NO: 22

SAD2A targeting construct cDNA sequence

SEQ ID NO: 23

5' sequence from coding region of SAD2B

SEQ ID NO: 24

3' sequence from coding region of SAD2B

SEQ ID NO: 25

20 SAD2B targeting construct cDNA sequence

SEQ ID NO: 26

[0252] Representative positive clones from transformations with each of the constructs were picked and the fatty acid profiles for these clones were determined.

Lipid samples were prepared from dried biomass from each transformant and fatty acid

25 profiles from these samples were analyzed using standard fatty acid methyl ester gas chromatography flame ionization (FAME GC/FID) detection methods as described in Example 3. The fatty acid profiles (expressed as Area % of total fatty acids) from the transgenic lines arising from transformation are shown in Table 8.

Table 8. Fatty acid profiles of desaturase knockouts.

Fatty Acid	d12FAD KO	SAD2A KO	SAD2B KO	wt UTEX 1435
C8:0	0	0	0	0
C10:0	0.01	0.01	0.01	0.01
C12:0	0.03	0.03	0.03	0.03
C14:0	1.08	0.985	0.795	1.46
C16:0	24.42	25.335	23.66	29.87
C18:0	6.85	12.89	19.555	3.345
C18:1	58.35	47.865	43.115	54.09

C18:2	7.33	10.27	9.83	9.1
C18:3 alpha	0.83	0.86	1	0.89
C20:0	0.48	0.86	1.175	0.325

[0253] Each construct had a measurable impact on the desired class of fatty acid and in all three cases C18:0 levels increased markedly, particularly with the two SAD knockouts. Further comparison of multiple clones from the SAD knockouts indicated that the SAD2B knockout lines had significantly greater reductions in C18:1 fatty acids than the C18:1 fatty acid levels observed with the SAD2A knockout lines.

[0254] Additional $\Delta 12$ FAD knockouts were generated in a *Prototheca moriformis* background using the methods described above. In order to identify potential homologous of $\Delta 12$ FADs, the following primers were used in order to amplify a genomic region

encoding a putative FAD:

Primer 1 5'-TCACTTCATGCCGGCGGTCC-3'

SEQ ID NO: 27

Primer 2 5'-GCGCTCCTGCTTGGCTCGAA-3'

SEQ ID NO: 28

The sequences resulting from the genomic amplification of *Prototheca moriformis*

genomic DNA using the above primers were highly similar, but indicated that multiple genes or alleles of $\Delta 12$ FADs exist in *Prototheca*.

[0255] Based on this result, two gene disruption constructs were designed that sought to inactivate one or more $\Delta 12$ FAD genes. The strategy would embed a sucrose invertase (*suc2* from *S. cerevisiae*) cassette (SEQ ID NO: 29), thus conferring the ability to

hydrolyze sucrose as a selectable marker, into highly conserved coding regions rather than use a classic gene replacement strategy. The first construct, termed pSZ1124, contained 5' and 3' genomic targeting sequences flanking a *C. reinhardtii* β -tubulin promoter (SEQ ID NO: 30) driving the expression of the *S. cerevisiae* *suc2* gene (SEQ ID NO: 31) and a *Chlorella vulgaris* nitrate reductase 3'UTR (SEQ ID NO: 32). The second construct, termed pSZ1125, contained 5' and 3' genomic targeting sequences flanking a *C.*

reinhardtii β -tubulin promoter driving the expression of the *S. cerevisiae* *suc2* gene and a *Chlorella vulgaris* nitrate reductase 3'UTR. The relevant sequences of the constructs are listed in the Sequence Listing:

S. cerevisiae *suc2* cassette

SEQ ID NO: 29

pSZ1124 (FAD2B) 5' genomic targeting sequence

SEQ ID NO: 33

pSZ1124 (FAD2B) 3' genomic targeting sequence

SEQ ID NO: 34

pSZ1125 (FAD2C) 5' genomic targeting sequence

SEQ ID NO: 35

pSZ1125 (FAD2C) 3' genomic targeting sequence

SEQ ID NO: 36

[0256] pSZ1124 and pSZ1125 were each introduced into a *Prototheca moriformis* background. Positive clones were selected based on the ability to hydrolyze sucrose.

- 5 Table 9 summarizes the fatty acid profiles (in Area %, generated using methods described above) obtained in two transgenic lines in which pSZ1124 and pSZ1125 targeting vectors were utilized.

Table 9. Fatty acid profiles of $\Delta 12$ FAD knockouts

	C10:0	C12:0	C14:0	C16:0	C16:1	C18:0	C18:1	C18:2	C18:3 α
parent	0.01	0.03	1.15	26.13	1.32	4.39	57.20	8.13	0.61
FAD2B	0.02	0.03	0.80	12.84	1.92	0.86	74.74	7.08	0.33
FAD2C	0.02	0.04	1.42	25.85	1.65	2.44	66.11	1.39	0.22

[0257] The transgenic containing the FAD2B (pSZ1124) construct gave a very

- 10 interesting and unexpected result in fatty acid profile, in that the C18:2 levels, which would be expected to decrease, only decreased by about one area %. However, the C18:1 fatty acid levels increased significantly, almost exclusively at the expense of the C16:0 levels, which decreased significantly. The transgenic containing the FAD2C (pSZ1125) construct also gave a change in lipid profile: the levels of C18:2 are reduced significantly
15 along with a corresponding increase in C18:1 levels.

B. RNA HAIRPIN APPROACH TO DOWN-REGULATION OF DELTA 12 DESATURASE (FADc) IN *PROTOTHECA* CELLS

[0258] Vectors constructed to down-regulate FADc (delta 12 desaturase gene) gene
20 expression by long hairpin RNAs were introduced into a *Prototheca moriformis* UTEX 1435 genetic background. The *Saccharomyces cerevisiae* *suc2* sucrose invertase gene was utilized as a selectable marker, conferring the ability to grow on sucrose as a sole-carbon source to positive clones, and two types of constructs were used. The first type of construct utilized a portion of the first exon of the FADc coding region linked in cis to its
25 first intron followed by a repeat unit of the first exon in reverse orientation. This type of construct was designed to form a hairpin when expressed as mRNA. Two constructs of this first type were created, one driven by the *Prototheca moriformis* Amt03 promoter (SEQ ID NO:37), termed pSZ1468, and a second driven by the *Chlamydomonas reinhardtii* β -tubulin promoter (SEQ ID NO:30), termed pSZ1469. The second type of
30 construct utilized the large FADc exon 2 in the antisense orientation driven by either the *Prototheca moriformis* Amt03 promoter (SEQ ID NO:37), termed pSZ1470, or driven by

the *Chlamydomonas reinhardtii* β -tubulin promoter (SEQ ID NO:30), termed pSZ1471. All four constructs had a *S. cerevisiae* *suc2* sucrose invertase cassette (SEQ ID NO:29) and a 5' (SEQ ID NO:38) and 3' (SEQ ID NO:39) homologous recombination targeting sequences (flanking the construct) to the 6S genomic region for integration into the

5 nuclear genome. Sequences of the FADc portions of each long hairpin RNA construct along with the relevant portions of each construct are listed in the Sequence Listing as:

	<u>Description</u>	<u>SEQ ID NO:</u>
	pSZ1468 FADc RNA hairpin cassette	SEQ ID NO:40
	Relevant portions of the pSZ1468 construct	SEQ ID NO:41
10	pSZ1469 FADc RNA hairpin cassette	SEQ ID NO:42
	Relevant portions of the pSZ1469 construct	SEQ ID NO:43
	pSZ1470 FADc exon 2 RNA hairpin cassette	SEQ ID NO:44
	Relevant portions of the pSZ1470 construct	SEQ ID NO:45
	pSZ1471 FADc exon 2 RNA hairpin cassette	SEQ ID NO:46
15	Relevant portions of the pSZ1471 construct	SEQ ID NO:47

[0259] Each of the four constructs was transformed into a *Prototheca moriformis* (UTEX 1435) background and positive clones were screened using plates with sucrose as the sole carbon source. Positive clones were picked from each transformation and a subset

20 were selected to determine the impact of the hairpin and antisense cassettes contained in pSZ1468, pSZ1469, pSZ1470 and pSZ1471 on fatty acid profiles. The selected clones from each transformation were grown under lipid producing conditions and the fatty acid profiles were determined using direct transesterification methods as described above. Representative fatty acid profiles from each of the transformations are summarized below

25 in Table 10. Wildtype 1 and 2 cells were untransformed *Prototheca moriformis* (UTEX 1435) cells that were run with each of the transformants as a negative control.

Table 10. Fatty acid profiles of *Prototheca moriformis* cells containing long hairpin RNA constructs to down-regulate the expression of delta 12 desaturase gene (FADc).

Strain	C10:0	C12:0	C14:0	C16:0	C18:0	C18:1	C18:2
wildtype 1	0.01	0.03	1.20	27.08	4.01	57.58	7.81
pSZ1468 clone A	0.01	0.04	1.33	25.95	3.68	65.60	1.25
pSZ1468 clone B	0.01	0.03	1.18	23.43	2.84	65.32	4.91
pSZ1468 clone C	0.01	0.04	1.34	23.18	4.27	63.65	5.17
pSZ1468 clone D	0.01	0.03	1.24	23.00	3.85	61.92	7.62

pSZ1470 clone A	0.01	0.03	1.23	24.79	4.33	58.43	8.92
pSZ1470 clone B	0.01	0.03	1.26	24.91	4.14	57.59	9.64
pSZ1470 clone C	0.01	0.03	1.21	23.35	4.75	58.52	9.70
wildtype 2	0.01	0.03	0.98	24.65	3.68	62.48	6.26
pSZ1469 clone A	0.01	0.03	1.05	21.74	2.71	71.33	1.22
pSZ1469 clone B	0.01	0.03	1.01	22.60	2.98	70.19	1.27
pSZ1469 clone C	0.01	0.03	1.03	19.82	2.38	72.95	1.82
pSZ1469 clone D	0.01	0.03	1.03	20.54	2.66	70.96	2.71
pSZ1471 clone A	0.01	0.03	1.03	18.42	2.63	66.94	8.55
pSZ1471 clone B	0.01	0.03	0.94	18.61	2.58	67.13	8.66
pSZ1471 clone C	0.01	0.03	1.00	18.31	2.46	67.41	8.71
pSZ1471 clone D	0.01	0.03	0.93	18.82	2.54	66.84	8.77

[0260] The above results show that the hairpin constructs pSZ1468 and pSZ1469 showed expected phenotypes: a reduction in C18:2 fatty acid levels and an increase in C18:1 fatty acid levels as compared to wildtype1 and wildtype 2, respectively. The antisense constructs, pSZ1470 and pSZ1471 did not result in a decrease in C18:2 fatty acid levels but instead showed a slight increase when compared to wildtype 1 and wildtype 2, respectively and a slight decrease in C16:0 fatty acid levels.

EXAMPLE 5: ENGINEERED MICROALGAE WITH ALTERED FATTY ACID PROFILES

[0261] As described above, integration of heterologous genes to attenuate specific endogenous lipid pathway genes, through knockout or knockdown, in *Prototheca* species can alter fatty acid profiles. Plasmid constructs (listed in Table 11) were created to assess whether the fatty acid profile of a host cell may be affected as a result of a knockout an endogenous fatty acyl-ACP thioesterase gene, FATA1.

[0262] A classically mutagenized derivative of *Protheca moriformis* UTEX 1435, Strain J, was transformed with one of the following plasmid constructs in Table 11 using

the methods of Example 2. Each construct contained a region for integration into the nuclear genome to interrupt the endogenous FATA1 gene and a *S. cerevisiae* *suc2* sucrose invertase coding region under the control of *C. reinhardtii* β -tubulin promoter/5'UTR (SEQ ID NO: 30) and *Chlorella vulgaris* nitrate reductase 3' UTR (SEQ ID NO: 32). This

5 *S. cerevisiae* *suc2* expression cassette is listed as SEQ ID NO: 29 and served as a selection marker. All protein coding regions were codon optimized to reflect the codon bias inherent in *Prototheca moriformis* UTEX 1435 (see Table 2) nuclear genes. Relevant sequences for the targeting regions for the FATA1 gene used for nuclear genome integration are shown below.

10

<u>Description</u>	<u>SEQ ID NO:</u>
5' sequence for integration into FATA1 locus	SEQ ID NO: 48
3' sequence for integration into FATA1 locus	SEQ ID NO: 49

15 **Table 11. Plasmid constructs used to transform *Prototheca moriformis* (UTEX 1435) STRAIN J.**

Plasmid Construct	Relevant Sequence Elements	SEQ ID NO:
pSZ1883	FATA1-CrbTub_yInv_nr-FATA1	SEQ ID NO: 50
pSZ1925	FATA1-CrbTub_yInv_nr::amt03_CwTE2_nr-FATA1	SEQ ID NO: 51

[0263] To introduce the *Cuphea wrightii* ACP-thioesterase 2 (CwFatB2) gene (Accession No: U56104) into STRAIN J at the FATA1-1 locus, a construct was generated

20 to express the protein coding region of the CwFatB2 gene under the control of the *Prototheca moriformis* Amt03 promoter/5'UTR (SEQ ID NO: 37) and *C. vulgaris* nitrate reductase 3'UTR (SEQ ID NO: 32). Relevant portions of this construct are provided in the Sequence Listing as SEQ ID NO: 51. The codon-optimized cDNA sequences and amino acid sequences of the *Cuphea wrightii* FatB2 thioesterase are listed in the Sequence

25 Listing as SEQ ID NO: 52 and SEQ ID NO: 53, respectively.

[0264] Upon transformation of FATA1-CrbTub_yInv_nr-FATA1 into STRAIN J, primary transformants were clonally purified and grown under standard lipid production conditions at pH 5.0 similar to the conditions as disclosed in Example 1. Lipid samples were prepared from dried biomass from each transformant and fatty acid profiles from

these samples were analyzed using standard fatty acid methyl ester gas chromatography flame ionization (FAME GC/FID) detection methods as described in Example 3. The fatty acid profiles (expressed as Area % of total fatty acids) from the transgenic lines arising from transformation with pSZ1883 into Strain J are shown in Table 12.

5 **Table 12. Fatty acid profiles of *Prototheca moriformis* cells containing a selectable marker to disrupt an endogenous FATA1 allele.**

Transformation	% C14:0	% C16:0	% C18:0	% C18:1	% C18:2
Wildtype	1.23	25.68	2.83	60.54	7.52
Transformant 1	0.86	16.95	1.75	68.44	9.78
Transformant 2	0.85	17.33	1.71	68.57	9.31
Transformant 3	0.82	17.40	1.78	68.55	9.22
Transformant 4	0.84	17.43	1.78	68.25	9.53
Transformant 5	0.75	17.64	2.02	69.02	8.61

[0265] These results show that ablation of the host's endogenous FATA1-1 allele alters the fatty acid profile of the engineered microalgae. The impact of targeting a
 10 selectable marker to the endogenous FATA1 allele is a clear diminution of C16:0 fatty acid production with an increase in C18:1 fatty acid production.

[0266] Upon transformation of FATA1-CrbTub_yInv_nr::amt03_CwTE2_nr-FATA1 into STRAIN J, primary transformants were clonally purified and grown under standard lipid production conditions at pH 7.0 with different carbon sources provided to a total
 15 concentration of 40 g/L. The sucrose concentration was 40 g/L. Where only glucose was used as the carbon source, glucose was provided at 40 g/L. Where glucose and fructose was used as the carbon source, glucoses was provided at 20 g/L and fructose was provided at 20 g/L. Lipid samples were prepared from dried biomass from each transformant and fatty acid profiles from these samples were analyzed using standard fatty acid methyl ester gas
 20 chromatography flame ionization (FAME GC/FID) detection methods as described in Example 3. The fatty acid profiles (expressed as Area % of total fatty acids) from the transgenic line arising from transformation with pSZ1925 into Strain J are shown in Table 13.

The resulting fatty acid profiles are listed in Table 13.

25 **Table 13. Fatty acid profiles of *Prototheca moriformis* cells containing a selectable marker and an exogenous thioesterase to disrupt an endogenous FATA1 allele.**

	Transformant	Carbon source	% C10:0	% C12:0	% C14:0	% C16:0	% C18:0	% C18:1	% C18:2
Strain J	Wildtype	Glucose	0.01	0.04	1.38	28.83	3.00	56.05	8.21
	Wildtype	Glucose	0.01	0.04	1.50	29.38	3.00	55.29	8.23
	Wildtype	Glucose/ Fructose	0.01	0.05	1.48	28.58	3.20	57.14	7.27
	Wildtype	Glucose/ Fructose	0.01	0.04	1.54	29.05	3.23	56.47	7.32
>2 copies	1	Glucose/ Fructose	4.29	19.98	9.17	20.68	3.47	34.38	6.37
	2	Glucose/ Fructose	3.11	16.17	9.91	15.97	1.57	45.72	5.81
	3	Sucrose	4.84	24.22	11.56	19.48	2.67	29.56	6.02
	4	Sucrose	3.24	16.67	10.39	16.34	1.43	44.41	6.00
1-2 copies	1	Glucose/ Fructose	0.18	1.64	1.85	14.43	2.12	70.30	7.63
	2	Glucose/ Fructose	0.18	1.56	1.74	13.56	2.25	71.04	7.72
	3	Sucrose	0.19	1.69	1.89	13.79	3.15	69.97	7.68
	4	Sucrose	0.15	1.26	1.49	13.44	2.73	71.46	7.77

[0267] Concordant with targeting a selectable marker alone to the host's FATA1-1 allele, integration of a selectable marker concomitant with an exogenous thioesterase alters the fatty acid profile of the engineered microalgae. As above, targeting an exogenous gene to the FATA1-1 allele results in a clear diminution of C16:0 fatty acid production. The additional expression of the CwTE2 thioesterase at the FATA1-1 locus also impacts mid chain fatty acids and C18:1 fatty acid production to an extent that is dependent upon the level of exogenous thioesterase activity present in the transformants analyzed. Genes bordered by repeat units such as the *C. vulgaris* nitrate reductase 3' UTR in constructs such as FATA1-CrbTub_yInv_nr::amt03_CwTE2_nr-FATA1, may be amplified upon integration in the host genome. There is good concordance between copy number of the amplified transgene at the target integration site and thioesterase levels as revealed either by impacts on fatty acid profiles or recombinant protein accumulation as assessed by western blotting.

[0268] Transgenic lines in which the CwTE2 gene has undergone amplification show a marked increase in mid chain (C10:0-C14:0) fatty acids and a concurrent decrease in C18:1 fatty acids. In contrast, those transformants in which CwTE2 has undergone little or no amplification (likely 1-2 copies) are consistent with lower expression of the exogenous thioesterase, resulting in a slight increase in mid chain fatty acids and a far greater impact on the increase of C18:1 fatty acids.

[0269] Collectively, these data show that ablation of the host's endogenous FATA1-1 allele alters the lipid profile of the engineered microalgae.

EXAMPLE 6: ALTERING FATTY ACID PROFILES OF MICROALGAE THROUGH OVEREXPRESSION OF A *PROTOTHECA* LIPID BIOSYNTHESIS GENE.

[0270] As described above, the β -ketoacyl-ACP synthase II (KASII) catalyzes the 2-carbon extension of C16:0-ACP to C18:0-ACP during fatty acid biosynthesis. It is an important lipid biosynthesis enzyme in establishing the fatty acid profile of the host organism and is critical for stearate and oleate production. Plasmid constructs were created to assess whether the fatty acid profile of a host cell may be affected as a result of expression of a KASII gene. Sources of KASII gene sequences were selected from *Protheca moriformis* UTEX 1435 or from higher plants (*Glycine max*, *Helianthus annuus*, or *Ricinus communis*).

[0271] A classically mutagenized derivative of *Protheca moriformis* UTEX 1435, STRAIN J, was transformed individually with one of the following plasmid constructs in Table 14 using the methods of Example 2. Each construct contained a region for integration into the nuclear genome at the 6S locus and a *S. cerevisiae* *suc2* sucrose invertase coding region under the control of *C. reinhardtii* β -tubulin promoter/5'UTR and *Chlorella vulgaris* nitrate reductase 3' UTR. This *S. cerevisiae* *suc2* expression cassette is listed as SEQ ID NO: 29 and served as a selection marker. For each construct, the KASII coding region was under the control of the *Prototheca moriformis* Amt03 promoter/5'UTR (SEQ ID NO: 37) and *C. vulgaris* nitrate reductase 3'UTR (SEQ ID NO: 32). The native transit peptide of each KASII enzyme was replaced with the *Chlorella protothecoides* stearyl-ACP desaturase transit peptide (SEQ ID NO: 54). All protein coding regions were codon optimized to reflect the codon bias inherent in *Prototheca moriformis* UTEX 1435 nuclear genes (see Table 2).

[0272] Relevant sequences for the targeting regions to the 6S locus for nuclear genome integration are shown below.

<u>Description</u>	<u>SEQ ID NO:</u>
5' sequence for integration into 6S locus	SEQ ID NO: 38
3' sequence for integration into 6S locus	SEQ ID NO: 39

Table 14. Plasmid constructs used to transform *Protheca moriformis* (UTEX 1435) STRAIN J.

Plasmid Construct	Source of KASII enzyme	Sequence Elements	SEQ ID. NO:
pSZ1747	<i>Glycine max</i>	6S:: β -tub:suc2:nr::Amt03:S106SAD:GlmKASII:nr::6S	SEQ ID NO: 55
pSZ1750	<i>Helianthus annuus</i>	6S:: β -tub:suc2:nr::Amt03:S106SAD:HaKASII:nr::6S	SEQ ID NO: 56
pSZ1754	<i>Ricinus communis</i>	6S:: β -tub:suc2:nr::Amt03:S106SAD:RcKASII:nr::6S	SEQ ID NO: 57
pSZ2041	<i>Protheca moriformis</i>	6S:: β -tub:suc2:nr::Amt03:S106SAD:PmKASII:nr::6S	SEQ ID NO: 58

[0273] The relevant nucleotide sequence of the construct 6S:: β -tub:suc2:nr::Amt03:S106SAD:PmKASII:nr::6S is provided in the sequence listings as SEQ ID. NO: 58. The codon-optimized sequence of PmKASII comprising a *Chlorella protothecoides* S106 stearyl-ACP desaturase transit peptide is provided the sequence listings as SEQ ID. NO: 105. SEQ ID NO: 106 provides the protein translation of SEQ ID NO. 105.

[0274] Upon individual transformation of each plasmid construct into Strain J, positive clones were screened on plates with sucrose as the sole carbon source. As in the previous examples, primary transformants were clonally purified and grown under standard lipid production conditions. Here, transformants were cultivated at pH 7 and lipid samples were prepared from dried biomass from each transformant as described above. Fatty acid profiles (expressed as Area %) of several positive transformants as compared to a wildtype negative control are summarized for each plasmid construct in Table 15 below.

Table 15. Fatty acid profiles of *Prototheca moriformis* cells engineered to overexpress KAS II genes.

Plasmid Construct	KASII Source	Transformant	%C14:0	%C16:0	%C18:0	%C18:1	%C18:2
None	no over-expression	1	1.36	28.69	2.92	56.36	8.16
		2	1.35	28.13	3.57	55.63	8.79
		3	1.22	25.74	2.82	60.6	7.31
		4	1.22	25.74	2.82	60.6	7.31
pSZ1747	Glm	1	2.23	25.34	2.69	57.35	9.53
		2	2.18	25.46	2.74	57.35	9.46

		3	2.18	25.33	2.89	57.34	9.5
		4	2.2	25.69	2.66	57.28	9.43
		5	2.17	25.38	3.03	56.99	9.72
pSZ1750	Ha	1	2.43	26.82	2.72	55.17	9.87
		2	2.44	27.14	2.62	54.89	9.81
		3	2.61	26.9	2.67	54.43	10.25
		4	1.96	30.32	2.87	53.87	8.26
		5	2.55	27.64	2.98	53.82	10.07
pSZ1754	Rc	1	1.84	24.41	2.89	59.26	9.08
		2	1.3	25.04	2.81	58.75	9.65
		3	1.27	25.98	2.76	58.33	9.22
		4	1.95	25.34	2.77	58.15	9.22
		5	1.3	26.53	2.75	57.87	9.09
pSZ2041	Pm	1	1.63	11.93	3.62	70.95	9.64
		2	1.85	11.63	3.34	69.88	10.93
		3	1.84	12.01	3.81	69.56	10.45
		4	1.63	14.22	3.72	68.86	9.6
		5	1.67	15.04	3.05	68.63	9.24

[0275] The data presented in Table 15 show that none of the higher plant KASII genes effected a change in the fatty acid profile of the transformed microalgal cells. Additional plasmid constructs expressing KASII genes from higher plants driven by promoters other than the *Prototheca* Amt03 promoter also failed to alter fatty acid profiles in transformed cells. In stark contrast, a clear diminution of C16:0 chain lengths with a concomitant increase in C18:1 length fatty acids was observed upon overexpression of the *Prototheca moriformis* KASII gene codon optimized using the codon frequency denoted in Table 2. Similar fatty acid profile changes were observed upon transformation of constructs expressing the PmKASII gene driven by a β -tublin promoter.

[0276] These results show that exogenous expression of a *Prototheca* lipid biosynthesis gene can alter the fatty acid profile of genetically engineered microalgae.

EXAMPLE 7: CHARACTERISTICS OF PROCESSED OIL PRODUCED FROM ENGINEERED MICROORGANISMS

[0277] Methods and effects of transforming *Prototheca moriformis* (UTEX 1435) with transformation vector pSZ1500 (SEQ ID NO: 137) have been previously described in PCT Application Nos. PCT/US2011/038463, PCT/US2011/038464, and PCT/US2012/023696.

[0278] A classically mutagenized (for higher oil production) derivative of *Protheca moriformis* (UTEX 1435), Strain A, was transformed with pSZ1500 according to biolistic transformation methods as described in herein and in PCT/US2009/066141,

PCT/US2009/066142, PCT/US2011/038463, PCT/US2011/038464, and PCT/US2012/023696. pSZ1500 comprised nucleotide sequence of the *Carthamus tinctorius* oleyl-thioesterase (CtOTE) gene, codon-optimized for expression in *P. moriformis* UTEX 1435. The pSZ1500 expression construct included 5' (SEQ ID NO: 138) and 3' (SEQ ID NO: 139) homologous recombination targeting sequences (flanking the construct) to the FADc genomic region for integration into the nuclear genome and a *S. cerevisiae* suc2 sucrose invertase coding region under the control of *C. reinhardtii* β -tubulin promoter/5'UTR (SEQ ID NO: 5) and *Chlorella vulgaris* nitrate reductase 3' UTR (SEQ ID NO: 126). This *S. cerevisiae* suc2 expression cassette is listed as SEQ ID NO: 127 and served as a selection marker. The CtOTE coding region was under the control of the *P. moriformis* Amt03 promoter/5'UTR (SEQ ID NO: 128) and *C. vulgaris* nitrate reductase 3'UTR, and the native transit peptide was replaced with the *C. protothecoides* stearyl-ACP desaturase transit peptide (SEQ ID NO: 129). The protein coding regions of CtOTE and suc2 were codon optimized to reflect the codon bias inherent in *P. moriformis* UTEX 1435 nuclear genes as described in PCT/US2009/066141, PCT/US2009/066142, PCT/US2011/038463, PCT/US2011/038464, and PCT/US2012/023696.

[0279] Primary pSZ1500 transformants of Strain A were selected on agar plates containing sucrose as a sole carbon source, clonally purified, and a single engineered line, Strain D was selected for analysis. Strain D was grown as described in

PCT/US2009/066141, PCT/US2009/066142, PCT/US2011/038463, PCT/US2011/038464, and PCT/US2012/023696. Hexane extraction of the oil from the generated biomass was then performed using standard methods, and the resulting triglyceride oil was determined to be free of residual hexane. Other methods of extraction of oil from microalgae using an expeller press are described in PCT Application No.

PCT/US2010/031108 and are hereby incorporated by reference.

[0280] Different lots of oil extracted from biomass of Strain D were refined, bleached, and deodorized using standard vegetable oil processing methods. These procedures generated oil samples RBD437, RBD469, RBD501, RBD 502, RBD503, and RBD529, which were subjected to analytical testing protocols according to methods defined through the American Oil Chemists' Society, the American Society for Testing and Materials, and the International Organization for Standardization. The results of these analyses are summarized below in Tables 16-21.

Table 16. Analytical results for oil sample RBD469.

Method Number	Test Description	Results	Units
---------------	------------------	---------	-------

AOCS Ca 3a-46	Insoluble impurities	<0.01	%
AOCS Ca 5a-40	Free Fatty Acids (Oleic)	0.02	%
AOCS Ca 5a-40	Acid Value	0.04	mg KOH/g
AOCS CA 9f-57	Neutral oil	98.9	%
D97	Cloud Point	-15	deg C
D97	Pour Point	-18	deg C
	Karl Fischer Moisture	0.01	%
AOCS Cc 13d-55 (modified)	Chlorophyll	<0.01	ppm
	Iodine Value	78.3	g I ₂ /100g
AOCS Cd 8b-90	Peroxide Value	0.31	meq/kg
ISO 6885	p-Anisidine Value	0.65	
AOCS Cc 18-80	Dropping Melting point (Mettler)	6.2	deg C
AOCS Cd 11d-96	Tricylglycerides	98.6	%
AOCS Cd 11d-96	Monoglyceride	<0.01	%
AOCS Cd 11d-96	Diglycerides	0.68	%
AOCS Cd 20-91	Total Polar Compounds	2.62	%
IUPAC, 2.507 and 2.508	Oxidized & Polymerized Tricylglycerides	17.62	%
AOCS Cc 9b-55	Flash Point	244	deg C
AOCS Cc 9a-48	Smoke Point	232	deg C
AOCS Cd 12b-92	Oxidative Stability Index Rancimat (110°C)	31.6	hours
AOCS Ca 6a-40	Unsaponified Matter	2.28	%

[0281] RBD469 oil was analyzed for trace element content, solid fat content, and Lovibond color according to AOCS methods. Results of these analyses are presented below in Table 17, Table 18, and Table 19.

Table 17. ICP Elemental Analysis of RBD469 oil.

Method Number	Test Description	Results in ppm
AOCS Ca 20-99 and AOCS Ca 17-01 (modified)	Phosphorus	1.09
	Calcium	0.1
	Magnesium	0.04
	Iron	<0.02
	Sulfur	28.8
	Copper	<0.05
	Potassium	<0.50
	Sodium	<0.50
	Silicon	0.51
	Boron	0.06
	Aluminum	<0.20
	Lead	<0.20
	Lithium	<0.02
	Nickel	<0.20
	Vanadium	<0.05
	Zinc	<0.02
	Arsenic	<0.20
	Mercury	<0.20
	Cadmium	<0.03
	Chromium	<0.02

	Manganese	<0.05
	Silver	<0.05
	Titanium	<0.05
	Selenium	<0.50
UOP779	Chloride organic	<1
UOP779	Chloride inorganic	7.24
AOCS Ba 4e-93	Nitrogen	6.7

Table 18. Solid Fat Content of RBD469 Oil

Method Number	Solid Fat Content	Result
AOCS Cd 12b-93	Solid Fat Content 10°C	0.13%
AOCS Cd 12b-93	Solid Fat Content 15°C	0.13%
AOCS Cd 12b-93	Solid Fat Content 20°C	0.28%
AOCS Cd 12b-93	Solid Fat Content 25°C	0.14%
AOCS Cd 12b-93	Solid Fat Content 30°C	0.08%
AOCS Cd 12b-93	Solid Fat Content 35°C	0.25%

Table 19. Lovibond Color of RBD469 Oil

Method Number	Color	Result	Unit
AOCS Cc 13j-97	red	2	Unit
AOCS Cc 13j-97	yellow	27	Unit

[0282] RBD469 oil was subjected to transesterification to produce fatty acid methyl esters (FAMES). The resulting FAME profile of RBD469 is shown in Table 20.

5 Table 20. FAME Profile of RBD469 Oil

Fatty Acid	Area %
C10	0.01

C12:0	0.04
C14:0	0.64
C15:0	0.08
C16:0	8.17
C16:1 iso	0.39
C16:1	0.77
C17:0	0.08
C18:0	1.93
C18:1	85.88
C18:1 iso	0.05
C18:2	0.05
C20:0	0.3
C20:1	0.06
C20:1	0.44
C22:0	0.11
C23:0	0.03
C24:0	0.1
Total FAMES Identified	99.13

[0283] The oil stability indexes (OSI) of 6 RBD oil samples without supplemented antioxidants and 3 RBD oil samples supplemented with antioxidants were analyzed according to the Oil Stability Index AOCS Method Cd 12b-92. Shown in Table 21 are the results of OSI AOCS Cd 12b-92 tests, conducted at 110°C, performed using a Metrohm 873 Biodiesel Rancimat. Results, except where indicated with an astericks (*), are the average of multiple OSI runs. Those samples not analyzed are indicated (NA).

Table 21. Oil Stability Index at 110°C of RBD oil samples with and without antioxidants

		OSI (hours) for each RBD Sample					
		RBD437	RBD469	RBD502	RBD501	RBD503	RBD529
Antioxidant added	Antioxidant Concentration						
None	0	65.41	38.33	72.10	50.32	63.04	26.68
Tocopherol & Ascorbyl Palmitate	35 ppm/16.7 ppm	77.72	48.60	82.67	NA	NA	NA
Tocopherol & Ascorbyl Palmitate	140 ppm/66.7 ppm	130.27	81.54*	211.49*	NA	NA	NA
Tocopherol & Ascorbyl Palmitate	1050 ppm/500 ppm	>157*	>144	242.5*	NA	NA	NA
Tocopherol	50 ppm	NA	46.97	NA	NA	NA	NA
TBHQ	20 ppm	63.37	37.4	NA	NA	NA	NA

[0284] The untransformed *P. moriformis* (UTEX 1435) acid profile comprises less than 60% C18:1 fatty acids and greater than 7% C18:2 fatty acids. In contrast, Strain D (comprising pSZ1500) exhibited fatty acid profiles with an increased composition of

C18:1 fatty acids (to above 85%) and a decrease in C18:2 fatty acids (to less than 0.06%). Upon refining, bleaching, and degumming, RBD oils samples prepared from the oil made from strain E exhibited OSI values > 26 hrs. With addition of antioxidants, the OSI of RBD oils prepared from oils of Strain D increased from 48.60 hours to greater than 242 hours.

EXAMPLE 8: IMPROVING THE LEVELS OF OLEIC ACID OF ENGINEERED MICROBES THROUGH ALLELIC DISRUPTION OF A FATTY ACID DESATURASE AND AN ACYL-ACP THIOESTERASE

[0285] This example describes the use of a transformation vector to disrupt a FATA locus of a *Prototheca moriformis* strain previously engineered for high oleic acid and low linoleic acid production. The transformation cassette used in this example comprised a selectable marker and nucleotide sequences encoding a *P. moriformis* KASII enzyme to engineer microorganisms in which the fatty acid profile of the transformed microorganism has been altered for further increased oleic acid and lowered palmitic acid levels.

[0286] Strain D, described in Example 7 and in PCT/US2012/023696, is a classically mutagenized (for higher oil production) derivative of *P. moriformis* (UTEX 1435) subsequently transformed with the transformation construct pSZ1500 (SEQ ID NO: 137) according to biolistic transformation methods in Example 2 and as described in PCT/US2009/066141, PCT/US2009/066142, PCT/US2011/038463, PCT/US2011/038464, and PCT/US2012/023696. This strain was used as the host for transformation with construct pSZ2276 to increase expression of a KASII enzyme while concomitantly ablating an endogenous acyl-ACP thioesterase genetic locus to generate Strain E. The pSZ2276 transformation construct included 5' (SEQ ID NO: 140) and 3' (SEQ ID NO: 141) homologous recombination targeting sequences (flanking the construct) to the FATA1 genomic region for integration into the *P. moriformis* nuclear genome, an *A. thaliana* THIC protein coding region under the control of the *C. protothecoides* actin promoter/5'UTR (SEQ ID NO: 142) and *C. vulgaris* nitrate reductase 3' UTR (SEQ ID NO: 126). This *At*THIC expression cassette is listed as SEQ ID NO: 143 and served as a selection marker. The *P. moriformis* KASII protein coding region was under the control of the *P. moriformis* Amt03 promoter/5'UTR (SEQ ID NO: 128) and *C. vulgaris* nitrate reductase 3'UTR, and the native transit peptide of the KASII enzyme was replaced with the *C. protothecoides* stearyl-ACP desaturase transit peptide (SEQ ID NO: 129). The codon-optimized sequence of PmKASII comprising a *C. protothecoides* S106 stearyl-ACP desaturase transit peptide is provided the sequence listings as SEQ ID NO:

144. SEQ ID NO: 145 provides the protein translation of SEQ ID NO: 144. The protein coding regions of PmKASII and *suc2* were codon optimized to reflect the codon bias inherent in *P. moriformis* UTEX 1435 nuclear genes as described in PCT/US2009/066141, PCT/US2009/066142, PCT/US2011/038463, PCT/US2011/038464, and

5 PCT/US2012/023696.

[0287] Primary pSZ2276 transformants of Strain D were selected on agar plates lacking thiamine, clonally purified, and a single engineered line, strain E was selected for analysis. Strain E was cultivated under heterotrophic lipid production conditions at pH5.0 and pH7.0 as described in PCT/US2009/066141, PCT/US2009/066142,

10 PCT/US2011/038463, PCT/US2011/038464, and PCT/US2012/023696. Lipid samples were prepared from dried biomass from each transformant and fatty acid profiles from these samples were analyzed using standard fatty acid methyl ester gas chromatography flame ionization (FAME GC/FID) detection methods as described in Example 3. The fatty acid profiles (expressed as Area % of total fatty acids) from the transgenic line arising
15 from transformation with pSZ2276 into Strain D are shown in Table 22.

Table 22. Fatty acid profiles of *Prototheca moriformis* (UTEX 1435) Strains A, D, and E engineered for increased oleic acid and lowered linoleic acid levels.

Strain	Transformation Construct(s)	pH	Area % Fatty Acid				
			C16:0	C18:0	C18:1	C18:2	C20:1
Strain A	None	pH 5	26.6	3.3	60.5	6.7	0.07
Strain A	None	pH 7	28.3	4.1	58	6.5	0.06
Strain D	pSZ1500	pH 5	17	3.6	77.1	0.01	0.14
Strain D	pSZ1500	pH 7	19.5	5.3	72.6	0.01	0.09
Strain E	pSZ1500 + pSZ2276	pH 5	4.1	2.36	88.5	0.04	3.1
Strain E	pSZ1500+ pSZ2276	pH 7	2.1	7.8	87.9	0.01	0.5

[0288] As shown in Table 22, targeted interruption of FADc alleles with a CtOTE expression cassette impacted the fatty acid profiles of transformed microorganisms. Fatty acid profiles of Strain D (comprising the pSZ1500 transformation vector) showed increased composition of C18:1 fatty acids with a concomitant decrease in C16:0 and C18:2 fatty acids relative to Strain A. Subsequent transformation of Strain D with

pSZ2276 to overexpress a *P. moriformis* (UTEX 1435) KASII protein while concomitantly ablating a FATA genetic locus (thereby generating Strain E) resulted in still further impact on the fatty acid profiles of the transformed microorganisms. Fatty acid profiles of Strain E showed increased composition of C18:1 fatty acids, with a further
5 decrease in C16:0 fatty acids relative to Strains A and D. Propagation of Strain E in culture conditions at pH 7, to induce expression from the Amt03 promoter, resulted in a fatty acid profile that was higher in C18:0 and C18:1 fatty acids and lower in C16:0 fatty acids, relative to the same strain cultured at pH 5.

[0289] These data demonstrate the utility of multiple genetic modifications to impact
10 the fatty acid profile of a host organism for increased levels of oleic acid with concomitant decreased levels of linoleic acid and palmitic acid. Further, this example illustrates the use of recombinant polynucleotides to target gene interruption of an endogenous FATA allele with a cassette comprising a pH-regulatable promoter to control expression of an exogenous KASII protein-coding region in order to alter the fatty acid profile of a host
15 microbe.

EXAMPLE 9: CONDITIONAL EXPRESSION OF A FATTY ACID DESATURASE

[0290] This example describes the use of a transformation vector to conditionally
express a delta 12 fatty acid desaturase (FAD) in a *Prototheca moriformis* strain
20 previously engineered for high oleic acid and very low linoleic acid production in both seed and lipid productivity stages of propagation. Very low linoleic acid levels in natural oils are sought for use in certain applications. However, absence of linoleic acid during cell division phase ("seed stage") of a host microbe is disadvantageous. Linoleic acid may be supplemented to the seed medium to hasten cell division and not added during lipid
25 production, but this addition imposes unwanted costs. To overcome this challenge, a transformation cassette was constructed for regulated expression of a FAD2 enzyme such that levels of linoleic acids sufficient for cell division could be achieved and oil with very low levels of linoleic acids could be produced during the oil production phase of culture of a microorganism. The transformation cassette used in this example comprised a selectable
30 marker, a pH-regulatable promoter, and nucleotide sequences encoding a *P. moriformis* FAD2 enzyme to engineer microorganisms in which the fatty acid profile of the transformed microorganism has been altered for increased oleic acid production and regulatable linoleic acid production.

[0291] Strain D, described in Examples 7 and 8 and in PCT/US2012/023696, is a classically mutagenized (for higher oil production) derivative of *P. moriformis* (UTEX 1435) subsequently transformed with the transformation construct pSZ1500 (SEQ ID NO: 137) according to biolistic transformation methods as described in PCT/US2009/066141, 5 PCT/US2009/066142, PCT/US2011/038463, PCT/US2011/038464, and PCT/US2012/023696. This strain was used as the host for transformation with construct pSZ2413 to introduce a pH-driven promoter for regulation of a *P. moriformis* FAD2 enzyme. The pSZ2413 transformation construct included 5' (SEQ ID NO: 121) and 3' (SEQ ID NO: 122) homologous recombination targeting sequences (flanking the 10 construct) to the 6S genomic region for integration into the *P. moriformis* nuclear genome, an *A. thaliana* THIC protein coding region under the control of the *C. protothecoides* actin promoter/5'UTR (SEQ ID NO: 142) and *C. vulgaris* nitrate reductase 3' UTR (SEQ ID NO: 126). This *At*THIC expression cassette is listed as SEQ ID NO: 143 and served as a selection marker. The *P. moriformis* FAD2 protein coding region was under the control of 15 the *P. moriformis* Amt03 promoter/5'UTR (SEQ ID NO: 128) and *C. vulgaris* nitrate reductase 3'UTR. The codon-optimized sequence of *Pm*FAD2 is provided the sequence listings as SEQ ID NO: 146. SEQ ID NO: 147 provides the protein translation of SEQ ID NO: 146. The protein coding regions of *Pm*FAD2 and *suc2* were codon optimized to reflect the codon bias inherent in *P. moriformis* UTEX 1435 nuclear genes as described in 20 PCT/US2009/066141, PCT/US2009/066142, PCT/US2011/038463, PCT/US2011/038464, and PCT/US2012/023696.

[0292] Primary pSZ2413 transformants of Strain D were selected on agar plates lacking thiamine, clonally purified, and isolates of the engineered line, Strain F were selected for analysis. These isolates were cultivated under heterotrophic lipid production 25 conditions at pH7.0 (to activate expression of FAD2 from the *PmAmt03* promoter) and at pH5.0, as described in PCT/US2009/066141, PCT/US2009/066142, PCT/US2011/038463, PCT/US2011/038464, and PCT/US2012/023696. Lipid samples were prepared from dried biomass from each transformant and fatty acid profiles from these samples were analyzed using standard fatty acid methyl ester gas chromatography 30 flame ionization (FAME GC/FID) detection methods as described in Example 3. The resulting profile of C18:2 fatty acids (expressed in Area %) from nine representative isolates of transgenic Strain F (F-1 through F-9) arising from transformation with pSZ2413 into Strain D are shown in Table 23.

Table 23. C18:2 fatty acid profiles of *Prototheca moriformis* (UTEX 1435) Strains A, D, and F.

Strain	Transformation Construct (s)	Area % C18:2	
		pH 5.0	pH 7.0
A	None	6.07	7.26
D	pSZ1500	0.01	0.01
F-1	pSZ1500 + pSZ2413	0.37	5.29
F-2	pSZ1500 + pSZ2413	0.45	6.87
F-3	pSZ1500 + pSZ2413	0.50	6.79
F-4	pSZ1500 + pSZ2413	0.57	5.06
F-5	pSZ1500 + pSZ2413	0.57	7.58
F-6	pSZ1500 + pSZ2413	0.60	6.88
F-7	pSZ1500 + pSZ2413	0.62	6.52
F-8	pSZ1500 + pSZ2413	0.63	5.79
F-9	pSZ1500 + pSZ2413	0.77	4.53

[0293] As shown in Table 23 the impact of regulated expression of the *PmFAD2* enzyme, effected though strain culture at different pH levels, is a clear increase in the composition of C18:2 fatty acids in the transformed microorganism. Linoleic acid comprises about 6% to about 7.3% of fatty acids of Strain A. In contrast, Strain D (comprising the pSZ1500 transformation vector to ablate both FAD2 alleles) is characterized by a fatty acid profile of 0.01% linoleic acid. Transformation of Strain D with pSZ2413 to generate Strain F results in a recombinant microbe in which the production of linoleic acid is regulated by the Amt03 promoter. Propagation of Strain F isolates in culture conditions at pH 7, to induce FAD2 expression from the Amt03 promoter, resulted in a fatty acid profile characterized by about 4.5% to about 7.5% linoleic acid. In contrast, propagation of Strain F isolates in culture conditions at pH 5 resulted in a fatty acid profile characterized by about 0.33 to about 0.77% linoleic acid.

[0294] These data demonstrate the utility of and effectiveness of recombinant polynucleotides permitting conditional expression of a FAD2 enzyme to alter the fatty acid profile of engineered microorganisms, and in particular in regulating the production of C18:2 fatty acids in microbial cells.

EXAMPLES 10-33: INTRODUCTION AND TABLES

[0295] Examples 10-33 below describe the engineering of various microorganisms in accordance with the present invention. To alter the fatty acid profile of a microorganism, microorganisms may be genetically modified wherein endogenous or exogenous lipid biosynthesis pathway enzymes are expressed, overexpressed, or attenuated. Steps to genetically engineer a microbe to alter its fatty acid profile as to the degree of fatty acid

unsaturation and to decrease or increase fatty acid chain length comprise the design and construction of a transformation vector (e.g., a plasmid), transformation of the microbe with one or more vectors, selection of transformed microbes (transformants), growth of the transformed microbe, and analysis of the fatty acid profile of the lipids produced by the engineered microbe.

[0296] Transgenes that alter the fatty acid profiles of host organisms may be expressed in numerous eukaryotic microbes. Examples of expression of transgenes in eukaryotic microbes including *Chlamydomonas reinhardtii*, *Chlorella ellipsoidea*, *Chlorella protothecoides*, *Chlorella saccharophila*, *Chlorella vulgaris*, *Chlorella kessleri*, *Chlorella sorokiniana*, *Haematococcus pluvialis*, *Gonium pectorale*, *Volvox carteri*, *Dunaliella tertiolecta*, *Dunaliella viridis*, *Dunaliella salina*, *Closterium peracerosum–strigosum–littorale* complex, *Nannochloropsis* sp., *Thalassiosira pseudonana*, *Phaeodactylum tricornutum*, *Navicula saprophila*, *Cylindrotheca fusiformis*, *Cyclotella cryptica*, *Symbiodinium microadriaticum*, *Amphidinium* sp., *Chaetoceros* sp., *Mortierella alpina*, and *Yarrowia lipolytica* may be found in the scientific literature. These expression techniques may be combined with the teachings of the present invention to produce engineered microorganisms with altered fatty acid profiles.

[0297] Transgenes that alter the fatty acid profiles of host organisms or alter the regiospecific distribution of glycerolipids produced by host organisms can also be expressed in numerous prokaryotic microbes. Examples of expression of transgenes in oleaginous microbes including *Rhodococcus opacus* may be found in the literature. These expression techniques may be combined with the teachings of the present invention to produce engineered microorganisms with altered fatty acid profiles.

Tables 24A-E. Codon preference listing

Table 24A

Amino Acid	Codon	<i>Chlorella sorokiniana</i>	<i>Chlorella vulgaris</i>	<i>Chlorella ellipsoidea</i>	<i>Chlorella kessleri</i>	<i>Dunaliella tertiolecta</i>	<i>Volvox carteri</i>	<i>Haematococcus pluvialis</i>
Ala	GCG	0.20	0.25	0.15	0.14	0.09	0.25	0.21
Ala	GCA	0.05	0.24	0.32	0.10	0.17	0.13	0.27
Ala	GCT	0.12	0.16	0.26	0.18	0.31	0.26	0.17
Ala	GCC	0.63	0.35	0.27	0.58	0.43	0.36	0.35
Arg	AGG	0.03	0.09	0.10	0.09	0.26	0.08	0.14
Arg	AGA	0.04	0.05	0.14	0.01	0.09	0.03	0.05
Arg	CGG	0.06	0.19	0.09	0.06	0.06	0.17	0.15
Arg	CGA	0.00	0.10	0.08	0.00	0.08	0.08	0.10

Arg	CGT	0.06	0.09	0.37	0.14	0.12	0.22	0.13
Arg	CGC	0.81	0.48	0.22	0.71	0.40	0.43	0.42
Asn	AAT	0.04	0.16	0.43	0.06	0.27	0.23	0.21
Asn	AAC	0.96	0.84	0.57	0.94	0.73	0.77	0.79
Asp	GAT	0.13	0.25	0.47	0.12	0.40	0.35	0.27
Asp	GAC	0.87	0.75	0.53	0.88	0.60	0.65	0.73
Cys	TGT	0.06	0.13	0.43	0.09	0.20	0.17	0.27
Cys	TGC	0.94	0.87	0.57	0.91	0.80	0.83	0.64
End	TGA	0.00	0.72	0.14	0.14	0.36	0.24	0.70
End	TAG	0.33	0.11	0.29	0.00	0.00	0.18	0.22
End	TAA	0.67	0.17	4.00	0.86	0.64	0.59	0.09
Gln	CAG	0.42	0.40	0.15	0.40	0.27	0.29	0.33
Gln	CAA	0.04	0.04	0.21	0.40	0.27	0.07	0.10
Glu	GAG	0.53	0.50	0.33	0.40	0.27	0.53	0.49
Glu	GAA	0.02	0.06	0.31	0.40	0.27	0.11	0.07
Gly	GGG	0.04	0.16	0.19	0.08	0.10	0.12	0.22
Gly	GGA	0.02	0.11	0.13	0.07	0.13	0.12	0.11
Gly	GGT	0.03	0.12	0.39	0.24	0.25	0.23	0.15
Gly	GGC	0.91	0.61	0.29	0.96	0.51	0.53	0.52
His	CAT	0.14	0.16	0.30	0.08	0.25	0.35	0.27
His	CAC	0.86	0.84	0.70	0.93	0.75	0.65	0.73
Ile	ATA	0.00	0.04	0.07	0.01	0.04	0.08	0.09
Ile	ATT	0.15	0.30	0.63	0.29	0.31	0.35	0.29
Ile	ATC	0.85	0.66	0.65	0.69	0.65	0.57	0.62
Leu	TTG	0.03	0.07	0.03	0.05	0.14	0.14	0.16
Leu	TTA	0.00	0.01	0.32	0.00	0.02	0.03	0.02
Leu	CTG	0.72	0.61	0.34	0.61	0.60	0.45	0.53
Leu	CTA	0.01	0.03	0.03	0.04	0.04	0.07	0.07
Leu	CTT	0.04	0.08	0.16	0.06	0.06	0.14	0.09
Leu	CTC	0.20	0.20	0.12	0.24	0.14	0.17	0.13
Lys	AAG	0.98	0.94	0.54	0.98	0.90	0.90	0.84
Lys	AAA	0.02	0.06	0.46	0.02	0.10	0.10	0.16
Met	ATG	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Phe	TTT	0.28	0.32	0.42	0.31	0.24	0.27	0.35
Phe	TTC	0.72	0.68	0.58	0.69	0.76	0.73	0.65
Pro	CCG	0.18	0.31	0.09	0.07	0.04	0.34	0.15
Pro	CCA	0.06	0.17	0.36	0.07	0.04	0.20	0.24
Pro	CCT	0.10	0.14	0.25	0.17	0.04	0.19	0.29
Pro	CCC	0.66	0.38	0.29	0.69	0.04	0.27	0.32
Ser	AGT	0.03	0.04	0.14	0.02	0.08	0.08	0.07
Ser	AGC	0.27	0.38	0.18	0.18	0.31	0.27	0.31
Ser	TCG	0.12	0.14	0.08	0.10	0.02	0.19	0.10
Ser	TCA	0.03	0.08	0.14	0.08	0.09	0.09	0.14
Ser	TCT	0.09	0.11	0.26	0.18	0.19	0.14	0.13
Ser	TCC	0.47	0.24	0.20	0.44	0.30	0.24	0.24
Thr	ACG	0.11	0.20	0.13	0.05	0.12	0.27	0.19

Thr	ACA	0.01	0.20	0.32	0.07	0.20	0.12	0.23
Thr	ACT	0.12	0.13	0.29	0.12	0.24	0.20	0.18
Thr	ACC	0.76	0.47	0.26	0.76	0.44	0.41	0.40
Trp	TGG	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Tyr	TAT	0.07	0.15	0.43	0.27	0.28	0.24	0.19
Tyr	TAC	0.93	0.85	0.57	0.73	0.72	0.76	0.81
Val	GTG	0.71	0.54	0.37	0.60	0.54	0.46	0.62
Val	GTA	0.00	0.05	0.25	0.03	0.09	0.07	0.09
Val	GTT	0.11	0.14	0.24	0.09	0.14	0.17	0.09
Val	GTC	0.18	0.27	0.14	0.28	0.23	0.30	0.21

Table 24B

Amino Acid	Codon	<i>Closterium peracerosum-strigosum-littorale</i> complex	<i>Dunaliella viridis</i>	<i>Dunaliella salina</i>	<i>Gonium pectorale</i>	<i>Phaeodactylum tricornutum</i>	<i>Chaetoceros compressum</i>
Ala	GCG	0.48	0.13	0.15	0.43	0.15	0.08
Ala	GCA	0.10	0.27	0.20	0.09	0.10	0.37
Ala	GCT	0.15	0.25	0.27	0.08	0.23	0.36
Ala	GCC	0.26	0.35	0.39	0.41	0.52	0.18
Arg	AGG	0.04	0.25	0.22	0.13	0.02	0.14
Arg	AGA	0.00	0.06	0.05	0.00	0.04	0.29
Arg	CGG	0.18	0.08	0.12	0.40	0.10	0.00
Arg	CGA	0.00	0.06	0.06	0.05	0.12	0.19
Arg	CGT	0.13	0.15	0.13	0.08	0.41	0.38
Arg	CGC	0.64	0.39	0.43	0.35	0.31	0.00
Asn	AAT	0.04	0.17	0.23	0.07	0.30	0.58
Asn	AAC	0.96	0.83	0.77	0.93	0.65	0.42
Asp	GAT	0.30	0.38	0.40	0.11	0.41	0.53
Asp	GAC	0.70	0.62	0.60	0.89	0.59	0.47
Cys	TGT	0.06	0.24	0.17	0.20	0.39	0.44
Cys	TGC	0.94	0.76	0.83	0.90	0.61	0.56
End	TGA	0.75	0.31	0.37	0.50	0.06	0.50
End	TAG	0.00	0.15	0.14	0.00	0.13	0.00
End	TAA	0.25	0.54	0.49	0.50	0.81	0.50
Gln	CAG	0.53	0.36	0.32	0.31	0.23	0.16
Gln	CAA	0.09	0.12	0.08	0.07	0.14	0.19
Glu	GAG	0.31	0.44	0.51	0.56	0.21	0.28
Glu	GAA	0.06	0.09	0.09	0.07	0.42	0.37
Gly	GGG	0.31	0.14	0.10	0.18	0.08	0.12
Gly	GGA	0.06	0.11	0.12	0.09	0.34	0.33
Gly	GGT	0.09	0.22	0.22	0.07	0.30	0.39
Gly	GGC	0.53	0.54	0.56	0.65	0.28	0.16
His	CAT	0.33	0.25	0.25	0.43	0.28	0.84

His	CAC	0.67	0.75	0.75	0.57	0.72	0.16
Ile	ATA	0.03	0.03	0.03	0.07	0.03	0.12
Ile	ATT	0.23	0.25	0.31	0.33	0.51	0.65
Ile	ATC	0.74	0.72	0.66	0.59	0.46	0.23
Leu	TTG	0.04	0.11	0.12	0.04	0.26	0.11
Leu	TTA	0.00	0.01	0.01	0.00	0.02	0.14
Leu	CTG	0.31	0.60	0.61	0.64	0.15	0.05
Leu	CTA	0.01	0.05	0.04	0.01	0.05	0.08
Leu	CTT	0.04	0.07	0.08	0.05	0.18	0.51
Leu	CTC	0.60	0.16	0.14	0.26	0.34	0.11
Lys	AAG	0.86	0.87	0.89	0.93	0.75	0.52
Lys	AAA	0.14	0.13	0.11	0.07	0.25	0.48
Met	ATG	1.00	1.00	1.00	1.00	1.00	1.00
Phe	TTT	0.09	0.25	0.29	0.10	0.44	0.65
Phe	TTC	0.91	0.75	0.71	0.90	0.56	0.35
Pro	CCG	0.28	0.10	0.08	0.53	0.29	0.05
Pro	CCA	0.15	0.10	0.17	0.09	0.12	0.45
Pro	CCT	0.12	0.10	0.30	0.04	0.20	0.33
Pro	CCC	0.44	0.10	0.45	0.34	0.40	0.17
Ser	AGT	0.04	0.09	0.06	0.02	0.12	0.14
Ser	AGC	0.05	0.31	0.32	0.20	0.12	0.07
Ser	TCG	0.22	0.04	0.06	0.42	0.19	0.08
Ser	TCA	0.16	0.08	0.10	0.09	0.06	0.31
Ser	TCT	0.05	0.17	0.15	0.07	0.15	0.23
Ser	TCC	0.47	0.31	0.30	0.20	0.35	0.18
Thr	ACG	0.30	0.16	0.13	0.42	0.23	0.10
Thr	ACA	0.06	0.21	0.18	0.03	0.13	0.38
Thr	ACT	0.22	0.18	0.23	0.08	0.19	0.27
Thr	ACC	0.42	0.46	0.46	0.47	0.45	0.25
Trp	TGG	1.00	1.00	1.00	1.00	1.00	1.00
Tyr	TAT	0.07	0.16	0.21	0.12	0.18	0.67
Tyr	TAC	0.93	0.84	0.79	0.88	0.82	0.33
Val	GTG	0.50	0.64	0.62	0.57	0.22	0.30
Val	GTA	0.02	0.03	0.05	0.04	0.09	0.27
Val	GTT	0.06	0.11	0.11	0.04	0.22	0.10
Val	GTC	0.42	0.22	0.23	0.35	0.47	0.33

Table 24C

Amino Acid	Codon	<i>Cylindro-theca fusiformis</i>	<i>Amphidinium carterae</i>	<i>Symbiodinium microadriaticum</i>	<i>Nannochloropsis</i> sp	<i>Cyclotella cryptica</i>	<i>Navi-cula pelliculosa</i>	<i>Thalassiosira pseudonana</i>	<i>C. reinhardtii</i>
Ala	GCG	0.07	0.17	0.22	0.24	0.11	0.00	0.11	0.35
Ala	GCA	0.14	0.33	0.26	0.10	0.16	0.13	0.25	0.08
Ala	GCT	0.35	0.29	0.20	0.17	0.45	0.44	0.33	0.13

Ala	GCC	0.43	0.20	0.32	0.48	0.27	0.44	0.30	0.43
Arg	AGG	0.09	0.15	0.27	0.00	0.09	0.05	0.18	0.05
Arg	AGA	0.14	0.03	0.27	0.00	0.05	0.10	0.17	0.01
Arg	CGG	0.06	0.08	0.09	0.00	0.04	0.05	0.06	0.20
Arg	CGA	0.16	0.18	0.09	0.29	0.08	0.35	0.11	0.04
Arg	CGT	0.34	0.18	0.09	0.14	0.47	0.20	0.34	0.09
Arg	CGC	0.22	0.40	0.18	0.57	0.28	0.25	0.15	0.62
Asn	AAT	0.42	0.37	0.21	0.00	0.25	0.47	0.43	0.09
Asn	AAC	0.58	0.63	0.79	1.00	0.75	0.53	0.57	0.91
Asp	GAT	0.54	0.54	0.50	0.20	0.52	0.20	0.56	0.14
Asp	GAC	0.46	0.46	0.50	0.80	0.48	0.80	0.44	0.86
Cys	TGT	0.44	0.75	0.50	0.00	0.29	0.10	0.54	0.10
Cys	TGC	0.56	0.25	0.50	1.00	0.71	0.90	0.46	0.90
End	TGA	0.13	0.50	1.00	0.00	0.10	0.00	0.31	0.27
End	TAG	0.10	0.00	0.00	0.00	0.00	0.00	0.38	0.22
End	TAA	0.77	0.50	0.00	1.00	0.90	1.00	0.31	0.52
Gln	CAG	0.12	0.33	0.28	0.41	0.19	0.21	0.16	0.38
Gln	CAA	0.25	0.15	0.17	0.00	0.17	0.28	0.19	0.04
Glu	GAG	0.23	0.41	0.50	0.59	0.38	0.17	0.40	0.55
Glu	GAA	0.39	0.10	0.06	0.00	0.26	0.34	0.26	0.03
Gly	GGG	0.06	0.19	0.32	0.10	0.10	0.03	0.12	0.11
Gly	GGA	0.47	0.10	0.12	0.05	0.45	0.28	0.51	0.06
Gly	GGT	0.35	0.34	0.16	0.25	0.22	0.13	0.23	0.11
Gly	GGC	0.12	0.37	0.40	0.60	0.24	0.56	0.14	0.72
His	CAT	0.39	0.12	0.40	0.00	0.42	1.00	0.50	0.11
His	CAC	0.61	0.88	0.60	1.00	0.58	0.00	0.50	0.89
Ile	ATA	0.06	0.05	0.00	0.00	0.04	0.00	0.08	0.03
Ile	ATT	0.42	0.53	0.38	0.14	0.53	0.73	0.38	0.22
Ile	ATC	0.52	0.42	0.63	0.86	0.42	0.27	0.54	0.75
Leu	TTG	0.26	0.35	0.39	0.22	0.20	0.16	0.29	0.04
Leu	TTA	0.09	0.01	0.00	0.00	0.03	0.00	0.05	0.01
Leu	CTG	0.09	0.22	0.39	0.09	0.06	0.12	0.08	0.73
Leu	CTA	0.05	0.00	0.04	0.00	0.03	0.04	0.06	0.03
Leu	CTT	0.37	0.31	0.13	0.04	0.39	0.36	0.20	0.05
Leu	CTC	0.13	0.12	0.04	0.65	0.29	0.32	0.32	0.15
Lys	AAG	0.60	0.93	0.85	1.00	0.70	0.83	0.76	0.95
Lys	AAA	0.40	0.07	0.15	0.00	0.30	0.17	0.24	0.05
Met	ATG	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Phe	TTT	0.37	0.21	0.25	0.20	0.31	0.78	0.38	0.16
Phe	TTC	0.63	0.79	0.75	0.80	0.69	0.22	0.62	0.84
Pro	CCG	0.11	0.14	0.18	0.08	0.10	0.21	0.16	0.33
Pro	CCA	0.33	0.42	0.09	0.08	0.16	0.29	0.31	0.08
Pro	CCT	0.32	0.22	0.41	0.25	0.35	0.21	0.31	0.13
Pro	CCC	0.24	0.22	0.32	0.58	0.39	0.29	0.23	0.47
Ser	AGT	0.12	0.13	0.09	0.00	0.09	0.13	0.18	0.04
Ser	AGC	0.09	0.24	0.14	0.13	0.08	0.28	0.11	0.35

Ser	TCG	0.13	0.03	0.05	0.00	0.15	0.25	0.17	0.25
Ser	TCA	0.12	0.25	0.05	0.00	0.12	0.08	0.12	0.05
Ser	TCT	0.30	0.16	0.23	0.13	0.39	0.25	0.23	0.07
Ser	TCC	0.24	0.19	0.45	0.75	0.18	0.03	0.19	0.25
Thr	ACG	0.09	0.14	0.10	0.28	0.10	0.18	0.21	0.30
Thr	ACA	0.15	0.28	0.10	0.00	0.15	0.09	0.19	0.08
Thr	ACT	0.39	0.12	0.10	0.17	0.33	0.41	0.28	0.10
Thr	ACC	0.37	0.47	0.70	0.56	0.43	0.32	0.32	0.52
Trp	TGG	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Tyr	TAT	0.38	0.32	0.20	0.00	0.38	0.20	0.39	0.10
Tyr	TAC	0.62	0.68	0.80	1.00	0.62	0.80	0.61	0.90
Val	GTG	0.11	0.65	0.67	0.31	0.16	0.18	0.29	0.67
Val	GTA	0.06	0.05	0.00	0.00	0.09	0.09	0.16	0.03
Val	GTT	0.38	0.08	0.11	0.15	0.42	0.09	0.28	0.07
Val	GTC	0.46	0.21	0.22	0.54	0.33	0.64	0.27	0.22

Table 24D

Amino Acid	Codon	<i>Yarrowia lipolytica</i>	<i>Mortierella alpina</i>	<i>Rhodococcus opacus</i>
Ala	GCG	0.08	0.14	0.35
Ala	GCA	0.11	0.12	0.14
Ala	GCT	0.35	0.29	0.09
Ala	GCC	0.46	0.45	0.43
Arg	AGG	0.05	0.05	0.05
Arg	AGA	0.13	0.06	0.02
Arg	CGG	0.12	0.06	0.26
Arg	CGA	0.52	0.09	0.12
Arg	CGT	0.11	0.32	0.11
Arg	CGC	0.07	0.42	0.44
Asn	AAT	0.17	0.15	0.21
Asn	AAC	0.83	0.85	0.79
Asp	GAT	0.35	0.42	0.24
Asp	GAC	0.65	0.58	0.76
Cys	TGT	0.46	0.13	0.26
Cys	TGC	0.54	0.87	0.74
End	TGA	0.16	0.05	0.72
End	TAG	0.38	0.25	0.17
End	TAA	0.46	0.70	0.11
Gln	CAG	0.33	0.36	0.28
Gln	CAA	0.08	0.06	0.06
Glu	GAG	0.44	0.49	0.45
Glu	GAA	0.14	0.09	0.22
Gly	GGG	0.05	0.03	0.18

Gly	GGA	0.28	0.29	0.15
Gly	GGT	0.32	0.32	0.20
Gly	GGC	0.34	0.36	0.48
His	CAT	0.34	0.27	0.20
His	CAC	0.66	0.73	0.80
Ile	ATA	0.03	0.01	0.05
Ile	ATT	0.44	0.33	0.14
Ile	ATC	0.53	0.66	0.81
Leu	TTG	0.09	0.27	0.09
Leu	TTA	0.02	0.00	0.01
Leu	CTG	0.37	0.26	0.41
Leu	CTA	0.05	0.02	0.03
Leu	CTT	0.18	0.12	0.06
Leu	CTC	0.29	0.32	0.40
Lys	AAG	0.84	0.91	0.80
Lys	AAA	0.16	0.09	0.20
Met	ATG	1.00	1.00	1.00
Phe	TTT	0.38	0.39	0.09
Phe	TTC	0.62	0.61	0.91
Pro	CCG	0.10	0.07	0.52
Pro	CCA	0.10	0.08	0.09
Pro	CCT	0.32	0.36	0.07
Pro	CCC	0.47	0.49	0.32
Ser	AGT	0.07	0.05	0.08
Ser	AGC	0.11	0.14	0.23
Ser	TCG	0.16	0.32	0.33
Ser	TCA	0.08	0.08	0.07
Ser	TCT	0.28	0.12	0.05
Ser	TCC	0.30	0.29	0.24
Thr	ACG	0.11	0.17	0.28
Thr	ACA	0.14	0.10	0.11
Thr	ACT	0.26	0.23	0.07
Thr	ACC	0.49	0.49	0.53
Trp	TGG	1.00	1.00	1.00
Tyr	TAT	0.18	0.20	0.18
Tyr	TAC	0.82	0.80	0.82
Val	GTG	0.33	0.22	0.37
Val	GTA	0.05	0.02	0.05
Val	GTT	0.26	0.27	0.10
Val	GTC	0.36	0.49	0.49

Table 24EPreferred codon usage in *Chlorella protothecoides*

TTC (Phe)	TAC (Tyr)	TGC (Cys)	TGA (Stop)
TGG (Trp)	CCC (Pro)	CAC (His)	CGC (Arg)
CTG (Leu)	CAG (Gln)	ATC (Ile)	ACC (Thr)
GAC (Asp)	TCC (Ser)	ATG (Met)	AAG (Lys)
GCC (Ala)	AAC (Asn)	GGC (Gly)	GTG (Val)
GAG (Glu)			

Table 25. Lipid biosynthesis pathway proteins.

5	<i>3-Ketoacyl ACP synthase</i>
10	<i>Cuphea hookeriana</i> 3-ketoacyl-ACP synthase (GenBank Acc. No. AAC68861.1), <i>Cuphea wrightii</i> beta-ketoacyl-ACP synthase II (GenBank Acc. No. AAB37271.1), <i>Cuphea lanceolata</i> beta-ketoacyl-ACP synthase IV (GenBank Acc. No. CAC59946.1), <i>Cuphea wrightii</i> beta-ketoacyl-ACP synthase II (GenBank Acc. No. AAB37270.1), <i>Ricinus communis</i> ketoacyl-ACP synthase (GenBank Acc. No. XP_002516228), <i>Gossypium hirsutum</i> ketoacyl-ACP synthase (GenBank Acc. No. ADK23940.1), <i>Glycine max</i> plastid 3-keto-acyl-ACP synthase II-A (GenBank Acc No. AAW88763.1), <i>Elaeis guineensis</i> beta-ketoacyl-ACP synthase II (GenBank Acc. No. AAF26738.2), <i>Helianthus annuus</i> plastid 3-keto-acyl-ACP synthase I (GenBank Acc. No. ABM53471.1), <i>Glycine max</i> 3-keto-acyl-ACP synthase I (GenBank Acc. No. NP_001238610.1), <i>Helianthus annuus</i> plastid 3-keto-acyl-ACP synthase II (GenBank Acc ABI18155.1), <i>Brassica napus</i> beta-ketoacyl-ACP synthetase 2 (GenBank Acc. No. AAF61739.1), <i>Perilla frutescens</i> beta-ketoacyl-ACP synthase II (GenBank Acc. No. AAC04692.1), <i>Helianthus annuus</i> beta-ketoacyl-ACP synthase II (GenBank Accession No. ABI18155), <i>Ricinus communis</i> beta-ketoacyl-ACP synthase II (GenBank Accession No. AAA33872), <i>Haematococcus pluvialis</i> beta-ketoacyl acyl carrier protein synthase (GenBank Accession No. HM560033.1), <i>Jatropha curcas</i> beta ketoacyl-ACP synthase I (GenBank Accession No. ABJ90468.1), <i>Populus trichocarpa</i> beta-ketoacyl-ACP synthase I (GenBank Accession No. XP_002303661.1),
15	<i>Coriandrum sativum</i> beta-ketoacyl-ACP synthetase I (GenBank Accession No. AAK58535.1), <i>Arabidopsis thaliana</i> 3-oxoacyl-[acyl-carrier-protein] synthase I (GenBank Accession No. NP_001190479.1), <i>Vitis vinifera</i> 3-oxoacyl-[acyl-carrier-protein] synthase I (GenBank Accession No. XP_002272874.2)
20	
25	
30	<i>Fatty acyl-ACP Thioesterases</i>
35	<i>Umbellularia californica</i> fatty acyl-ACP thioesterase (GenBank Acc. No. AAC49001), <i>Cinnamomum camphora</i> fatty acyl-ACP thioesterase (GenBank Acc. No. Q39473), <i>Umbellularia californica</i> fatty acyl-ACP thioesterase (GenBank Acc. No. Q41635), <i>Myristica fragrans</i> fatty acyl-ACP thioesterase (GenBank Acc. No. AAB71729), <i>Myristica fragrans</i> fatty acyl-ACP thioesterase (GenBank Acc. No. AAB71730), <i>Elaeis guineensis</i> fatty acyl-ACP thioesterase (GenBank Acc. No. ABD83939), <i>Elaeis guineensis</i> fatty acyl-ACP thioesterase (GenBank Acc. No. AAD42220), <i>Populus tomentosa</i> fatty acyl-ACP thioesterase (GenBank Acc. No. ABC47311), <i>Arabidopsis thaliana</i> fatty acyl-ACP thioesterase (GenBank Acc. No. NP_172327), <i>Arabidopsis thaliana</i> fatty acyl-ACP thioesterase (GenBank Acc. No. CAA85387), <i>Arabidopsis thaliana</i> fatty acyl-ACP thioesterase (GenBank Acc. No. CAA85388), <i>Gossypium hirsutum</i> fatty acyl-ACP thioesterase (GenBank Acc. No. Q9SQI3), <i>Cuphea lanceolata</i> fatty acyl-ACP thioesterase
40	

(GenBank Acc. No. CAA54060), *Cuphea hookeriana* fatty acyl-ACP thioesterase (GenBank Acc. No. AAC72882), *Cuphea calophylla subsp. mesostemon* fatty acyl-ACP thioesterase (GenBank Acc. No. ABB71581), *Cuphea lanceolata* fatty acyl-ACP thioesterase (GenBank Acc. No. CAC19933), *Elaeis guineensis* fatty acyl-ACP thioesterase (GenBank Acc. No. AAL15645), *Cuphea hookeriana* fatty acyl-ACP thioesterase (GenBank Acc. No. Q39513), *Gossypium hirsutum* fatty acyl-ACP thioesterase (GenBank Acc. No. AAD01982), *Vitis vinifera* fatty acyl-ACP thioesterase (GenBank Acc. No. CAN81819), *Garcinia mangostana* fatty acyl-ACP thioesterase (GenBank Acc. No. AAB51525), *Brassica juncea* fatty acyl-ACP thioesterase (GenBank Acc. No. ABI18986), *Madhuca longifolia* fatty acyl-ACP thioesterase (GenBank Acc. No. AAX51637), *Brassica napus* fatty acyl-ACP thioesterase (GenBank Acc. No. ABH11710), *Brassica napus* fatty acyl-ACP thioesterase (GenBank Acc. No. CAA52070.1), *Oryza sativa* (indica cultivar-group) fatty acyl-ACP thioesterase (GenBank Acc. No. EAY86877), *Oryza sativa* (japonica cultivar-group) fatty acyl-ACP thioesterase (GenBank Acc. No. NP_001068400), *Oryza sativa* (indica cultivar-group) fatty acyl-ACP thioesterase (GenBank Acc. No. EAY99617), *Cuphea hookeriana* fatty acyl-ACP thioesterase (GenBank Acc. No. AAC49269), *Ulmus Americana* fatty acyl-ACP thioesterase (GenBank Acc. No. AAB71731), *Cuphea lanceolata* fatty acyl-ACP thioesterase (GenBank Acc. No. CAB60830), *Cuphea palustris* fatty acyl-ACP thioesterase (GenBank Acc. No. AAC49180), *Iris germanica* fatty acyl-ACP thioesterase (GenBank Acc. No. AAG43858), *Iris germanica* fatty acyl-ACP thioesterase (GenBank Acc. No. AAG43858.1), *Cuphea palustris* fatty acyl-ACP thioesterase (GenBank Acc. No. AAC49179), *Myristica fragrans* fatty acyl-ACP thioesterase (GenBank Acc. No. AAB71729), *Myristica fragrans* fatty acyl-ACP thioesterase (GenBank Acc. No. AAB71729.1), *Cuphea hookeriana* fatty acyl-ACP thioesterase (GenBank Acc. No. U39834), *Umbellularia californica* fatty acyl-ACP thioesterase (GenBank Acc. No. M94159), *Cinnamomum camphora* fatty acyl-ACP thioesterase (GenBank Acc. No. U31813), *Ricinus communis* fatty acyl-ACP thioesterase (GenBank Acc. No. ABS30422.1), *Helianthus annuus* acyl-ACP thioesterase (GenBank Accession No. AAL79361.1), *Jatropha curcas* acyl-ACP thioesterase (GenBank Accession No. ABX82799.3), *Zea mays* oleoyl-acyl carrier protein thioesterase, (GenBank Accession No. ACG40089.1), *Haematococcus pluvialis* fatty acyl-ACP thioesterase (GenBank Accession No. HM560034.1)

Desaturase Enzymes

Linum usitatissimum fatty acid desaturase 3C, (GenBank Acc. No. ADV92272.1), *Ricinus communis* omega-3 fatty acid desaturase, endoplasmic reticulum, putative, (GenBank Acc. No. EEF36775.1), *Vernicia fordii* omega-3 fatty acid desaturase, (GenBank Acc. No. AAF12821), *Glycine max* chloroplast omega 3 fatty acid desaturase isoform 2, (GenBank Acc. No. ACF19424.1), *Prototheca moriformis* FAD-D omega 3 desaturase (SEQ ID NO: 155), *Prototheca moriformis* linoleate desaturase (SEQ ID NO: 156), *Carthamus tinctorius* delta 12 desaturase, (GenBank Accession No. ADM48790.1), *Gossypium hirsutum* omega-6 desaturase, (GenBank Accession No. CAA71199.1), *Glycine max* microsomal desaturase (GenBank Accession No. BAD89862.1), *Zea mays* fatty acid desaturase (GenBank Accession No. ABF50053.1), *Brassica napa* linoleic acid desaturase (GenBank Accession No. AAA32994.1), *Camelina sativa* omega-3 desaturase (SEQ ID NO: 157), *Prototheca moriformis* delta 12 desaturase allele 2 (SEQ ID NO: 158), *Camelina sativa* omega-3 FAD7-1 (SEQ ID NO: 159), *Helianthus annuus* stearyl-ACP desaturase, (GenBank Accession No. AAB65145.1), *Ricinus communis* stearyl-ACP desaturase, (GenBank Accession No. AACG59946.1), *Brassica juncea* plastidic delta-9-stearyl-ACP

desaturase (GenBank Accession No. AAD40245.1), *Glycine max* stearoyl-ACP desaturase (GenBank Accession No. ACJ39209.1), *Olea europaea* stearoyl-ACP desaturase (GenBank Accession No. AAB67840.1), *Vernicia fordii* stearoyl-acyl-carrier protein desaturase, (GenBank Accession No. ADC32803.1), *Descurainia sophia* delta-12 fatty acid desaturase (GenBank Accession No. ABS86964.2), *Euphorbia lagascae* delta12-oleic acid desaturase (GenBank Acc. No. AAS57577.1), *Chlorella vulgaris* delta 12 fatty acid desaturase (GenBank Accession No. ACF98528), *Chlorella vulgaris* omega-3 fatty acid desaturase (GenBank Accession No. BAB78717), *Haematococcus pluvialis* omega-3 fatty acid desaturase (GenBank Accession No. HM560035.1), *Haematococcus pluvialis* stearoyl-ACP-desaturase GenBank Accession No. EF586860.1, *Haematococcus pluvialis* stearoyl-ACP-desaturase GenBank Accession No. EF523479.1

Oleate 12-hydroxylase Enzymes

Ricinus communis oleate 12-hydroxylase (GenBank Acc. No. AAC49010.1), *Physaria lindheimeri* oleate 12-hydroxylase (GenBank Acc. No. ABQ01458.1), *Physaria lindheimeri* mutant bifunctional oleate 12-hydroxylase:desaturase (GenBank Acc. No. ACF17571.1), *Physaria lindheimeri* bifunctional oleate 12-hydroxylase:desaturase (GenBank Accession No. ACQ42234.1), *Physaria lindheimeri* bifunctional oleate 12-hydroxylase:desaturase (GenBank Acc. No. AAC32755.1), *Arabidopsis lyrata subsp. Lyrata* (GenBank Acc. No. XP_002884883.1)

Glycerol-3-phosphate Enzymes

Arabidopsis thaliana glycerol-3-phosphate acyltransferase BAA00575, *Chlamydomonas reinhardtii* glycerol-3-phosphate acyltransferase (GenBank Acc. No. EDP02129), *Chlamydomonas reinhardtii* glycerol-3-phosphate acyltransferase (GenBank Acc. No. Q886Q7), *Cucurbita moschata* acyl-(acyl-carrier-protein):glycerol-3-phosphate acyltransferase (GenBank Acc. No. BAB39688), *Elaeis guineensis* glycerol-3-phosphate acyltransferase, ((GenBank Acc. No. AAF64066), *Garcinia mangostana* glycerol-3-phosphate acyltransferase (GenBank Acc. No. ABS86942), *Gossypium hirsutum* glycerol-3-phosphate acyltransferase (GenBank Acc. No. ADK23938), *Jatropha curcas* glycerol-3-phosphate acyltransferase (GenBank Acc. No. ADV77219), *Jatropha curcas* plastid glycerol-3-phosphate acyltransferase (GenBank Acc. No. ACR61638), *Ricinus communis* plastidial glycerol-phosphate acyltransferase (GenBank Acc. No. EEF43526), *Vicia faba* glycerol-3-phosphate acyltransferase (GenBank Accession No. AAD05164), *Zea mays* glycerol-3-phosphate acyltransferase (GenBank Acc. No. ACG45812)

Lysophosphatidic acid acyltransferase Enzymes

Arabidopsis thaliana 1-acyl-sn-glycerol-3-phosphate acyltransferase (GenBank Accession No. AEE85783), *Brassica juncea* 1-acyl-sn-glycerol-3-phosphate acyltransferase (GenBank Accession No. ABQ42862), *Brassica juncea* 1-acyl-sn-glycerol-3-phosphate acyltransferase (GenBank Accession No. ABM92334), *Brassica napus* 1-acyl-sn-glycerol-3-phosphate acyltransferase (GenBank Accession No. CAB09138), *Chlamydomonas reinhardtii* lysophosphatidic acid acyltransferase (GenBank Accession No. EDP02300), *Cocos nucifera* lysophosphatidic acid acyltransferase (GenBank Acc. No. AAC49119), *Limnanthes alba* lysophosphatidic acid acyltransferase (GenBank Accession No. EDP02300), *Limnanthes douglasii* 1-acyl-sn-glycerol-3-phosphate acyltransferase (putative) (GenBank Accession No. CAA88620), *Limnanthes douglasii* acyl-CoA:sn-1-acylglycerol-3-phosphate acyltransferase (GenBank Accession No. ABD62751),

Limnanthes douglasii 1-acylglycerol-3-phosphate O-acyltransferase (GenBank Accession No. CAA58239), *Ricinus communis* 1-acyl-sn-glycerol-3-phosphate acyltransferase (GenBank Accession No. EEF39377)

5 Diacylglycerol acyltransferase Enzymes

Arabidopsis thaliana diacylglycerol acyltransferase (GenBank Acc. No. CAB45373), *Brassica juncea* diacylglycerol acyltransferase (GenBank Acc. No. AAY40784), *Elaeis guineensis* putative diacylglycerol acyltransferase (GenBank Acc. No. AEQ94187), *Elaeis guineensis* putative diacylglycerol acyltransferase (GenBank Acc. No. AEQ94186), *Glycine max* acyl-CoA:diacylglycerol acyltransferase (GenBank Acc. No. AAT73629), *Helianthus annuus* diacylglycerol acyltransferase (GenBank Acc. No. ABX61081), *Olea europaea* acyl-CoA:diacylglycerol acyltransferase 1 (GenBank Acc. No. AAS01606), *Ricinus communis* diacylglycerol acyltransferase (GenBank Acc. No. AAR11479)

Phospholipid diacylglycerol acyltransferase Enzymes

Arabidopsis thaliana phospholipid:diacylglycerol acyltransferase (GenBank Acc. No. AED91921), *Elaeis guineensis* putative phospholipid:diacylglycerol acyltransferase (GenBank Acc. No. AEQ94116), *Glycine max* phospholipid:diacylglycerol acyltransferase 1-like (GenBank Acc. No. XP_003541296), *Jatropha curcas* phospholipid:diacylglycerol acyltransferase (GenBank Acc. No. AEZ56255), *Ricinus communis* phospholipid:diacylglycerol acyltransferase (GenBank Acc. No. ADK92410), *Ricinus communis* phospholipid:diacylglycerol acyltransferase (GenBank Acc. No. AEW99982)

EXAMPLE 10: ENGINEERING *CHLORELLA SOROKINIANA*

[0298] Expression of recombinant genes in accordance with the present invention in *Chlorella sorokiniana* may be accomplished by modifying the methods and vectors taught by Dawson *et al.* as discussed herein. Briefly, Dawson *et al.*, *Current Microbiology* Vol. 35 (1997) pp. 356–362, reported the stable nuclear transformation of *Chlorella sorokiniana* with plasmid DNA. Using the transformation method of microprojectile bombardment, Dawson introduced the plasmid pSV72-NRg, encoding the full *Chlorella vulgaris* nitrate reductase gene (NR, GenBank Accession No. U39931), into mutant *Chlorella sorokiniana* (NR-mutants). The NR-mutants are incapable of growth without the use of nitrate as a source of nitrogen. Nitrate reductase catalyzes the conversion of nitrate to nitrite. Prior to transformation, *Chlorella sorokiniana* NR-mutants were unable to grow beyond the microcolony stage on culture medium comprising nitrate (NO_3^-) as the sole nitrogen source. The expression of the *Chlorella vulgaris* NR gene product in NR-mutant *Chlorella sorokiniana* was used as a selectable marker to rescue the nitrate metabolism deficiency. Upon transformation with the pSV72-NRg plasmid, NR-mutant

Chlorella sorokiniana stably expressing the *Chlorella vulgaris* NR gene product were obtained that were able to grow beyond the microcolony stage on agar plates comprising nitrate as the sole carbon source. Evaluation of the DNA of the stable transformants was performed by Southern analysis and evaluation of the RNA of the stable transformants

5 was performed by RNase protection. Selection and maintenance of the transformed *Chlorella sorokiniana* (NR mutant) was performed on agar plates (pH 7.4) comprising 0.2 g/L MgSO₄, 0.67 g/L KH₂PO₄, 3.5 g/L K₂HPO₄, 1.0 g/L Na₃C₆H₅O₇·H₂O and 16.0 g/L agar, an appropriate nitrogen source (e.g., NO₃⁻), micronutrients, and a carbon source.

Dawson also reported the propagation of *Chlorella sorokiniana* and *Chlorella sorokiniana* NR mutants in liquid culture medium. Dawson reported that the plasmid pSV72-NRg and the promoter and 3' UTR/terminator of the *Chlorella vulgaris* nitrate reductase gene were suitable to enable heterologous gene expression in *Chlorella sorokiniana* NR-mutants.

Dawson also reported that expression of the *Chlorella vulgaris* nitrate reductase gene product was suitable for use as a selectable marker in *Chlorella sorokiniana* NR-mutants.

15 **[0299]** In an embodiment of the present invention, vector pSV72-NRg, comprising nucleotide sequence encoding the *Chlorella vulgaris* nitrate reductase (CvNR) gene product for use as a selectable marker, is constructed and modified to further comprise a lipid biosynthesis pathway expression cassette sequence, thereby creating a transformation vector. The lipid biosynthesis pathway expression cassette encodes one or more lipid

20 biosynthesis pathway proteins selected from Table 25, each protein-coding sequence codon-optimized for expression in *Chlorella sorokiniana* to reflect the codon bias inherent in nuclear genes of *Chlorella sorokiniana* in accordance with Tables 24A-D. For each lipid biosynthesis pathway protein of Table 25, the codon-optimized gene sequence can individually be operably linked to the CvNR promoter upstream of the protein-coding

25 sequence and operably linked to the CvNR 3'UTR/terminator at the 3' region, or downstream, of the protein-coding sequence. The transformation construct may additionally comprise homology regions to the *Chlorella sorokiniana* genome for targeted genomic integration of the transformation vector. Homology regions may be selected to disrupt one or more genomic sites of endogenous lipid biosynthesis pathway genes. Stable

30 transformation of *Chlorella sorokiniana* with the transformation vector is achieved through well-known transformation techniques including microprojectile bombardment or other known methods. Activity of the CvNR gene product may be used as a selectable marker to rescue the nitrogen assimilation deficiency of *Chlorella sorokiniana* NR mutant strains and to select for *Chlorella sorokiniana* NR-mutants stably expressing the

transformation vector. Growth media suitable for *Chlorella sorokiniana* lipid production include, but are not limited to 0.5 g/L KH₂PO₄, 0.5g/L K₂HPO₄, 0.25 g/L MgSO₄-7H₂O, with supplemental micronutrients and the appropriate nitrogen and carbon sources (Patterson, *Lipids* Vol.5:7 (1970), pp.597-600). Evaluation of fatty acid profiles of

5 *Chlorella sorokiniana* lipids may be assessed through standard lipid extraction and analytical methods described herein.

EXAMPLE 11: ENGINEERING *CHLORELLA VULGARIS*

[0300] Expression of recombinant genes in accordance with the present invention in

10 *Chlorella vulgaris* may be accomplished by modifying the methods and vectors taught by Chow and Tung *et al.* as discussed herein. Briefly, Chow and Tung *et al.*, *Plant Cell Reports*, Volume 18 (1999), pp. 778-780, reported the stable nuclear transformation of *Chlorella vulgaris* with plasmid DNA. Using the transformation method of electroporation, Chow and Tung introduced the plasmid pIG121-Hm (GenBank Accession

15 No. AB489142) into *Chlorella vulgaris*. The nucleotide sequence of pIG121-Hm comprised sequence encoding a beta-glucuronidase (GUS) reporter gene product operably-linked to a CaMV 35S promoter upstream of the GUS protein-coding sequence and further operably linked to the 3' UTR/terminator of the nopaline synthase (*nos*) gene downstream of the GUS protein-coding sequence. The sequence of plasmid pIG121-Hm further

20 comprised a hygromycin B antibiotic resistance cassette. This hygromycin B antibiotic resistance cassette comprised a CaMV 35S promoter operably linked to sequence encoding the hygromycin phosphotransferase (*hpt*, GenBank Accession No. BAH24259) gene product. Prior to transformation, *Chlorella vulgaris* was unable to be propagated in culture medium comprising 50 ug/ml hygromycin B. Upon transformation with the

25 pIG121-Hm plasmid, transformants of *Chlorella vulgaris* were obtained that were propagated in culture medium comprising 50 ug/ml hygromycin B. The expression of the *hpt* gene product in *Chlorella vulgaris* enabled propagation of transformed *Chlorella vulgaris* in the presence of 50 ug/mL hygromycin B, thereby establishing the utility of the a hygromycin B resistance cassette as a selectable marker for use in *Chlorella vulgaris*.

30 Detectable activity of the GUS reporter gene indicated that CaMV 35S promoter and *nos* 3'UTR are suitable for enabling heterologous gene expression in *Chlorella vulgaris*. Evaluation of the genomic DNA of the stable transformants was performed by Southern analysis. Selection and maintenance of transformed *Chlorella vulgaris* was performed on agar plates comprising YA medium (agar and 4 g/L yeast extract). The propagation of

Chlorella vulgaris in liquid culture medium was conducted as discussed by Chow and Tung. Propagation of *Chlorella vulgaris* in media other than YA medium has been described (for examples, see Chader et al., *Revue des Energies Renouvelables*, Volume 14 (2011), pp. 21-26 and Illman et al., *Enzyme and Microbial Technology*, Vol. 27 (2000), pp. 631-635). Chow and Tung reported that the plasmid pIG121-Hm, the CaMV 35S promoter, and the *Agrobacterium tumefaciens* nopaline synthase gene 3'UTR/terminator are suitable to enable heterologous gene expression in *Chlorella vulgaris*. In addition, Chow and Tung reported the hygromycin B resistance cassette was suitable for use as a selectable marker in *Chlorella vulgaris*. Additional plasmids, promoters, 3'UTR/terminators, and selectable markers suitable for enabling heterologous gene expression in *Chlorella vulgaris* have been discussed in Chader et al., *Revue des Energies Renouvelables*, Volume 14 (2011), pp. 21-26.

[0301] In an embodiment of the present invention, pIG121-Hm, comprising the nucleotide sequence encoding the hygromycin B gene product for use as a selectable marker, is constructed and modified to further comprise a lipid biosynthesis pathway expression cassette sequence, thereby creating a transformation vector. The lipid biosynthesis pathway expression cassette encodes one or more lipid biosynthesis pathway proteins selected from Table 25, each protein-coding sequence codon-optimized for expression in *Chlorella vulgaris* to reflect the codon bias inherent in nuclear genes of *Chlorella vulgaris* in accordance with Tables 24A-D. For each lipid biosynthesis pathway protein of Table 25, the codon-optimized gene sequence can individually be operably linked to the CaMV 35S promoter upstream of the protein-coding sequence and operably linked to the *Agrobacterium tumefaciens* nopaline synthase gene 3'UTR/terminator at the 3' region, or downstream, of the protein-coding sequence. The transformation construct may additionally comprise homology regions to the *Chlorella vulgaris* genome for targeted genomic integration of the transformation vector. Homology regions may be selected to disrupt one or more genomic sites of endogenous lipid biosynthesis pathway genes. Stable transformation of *Chlorella vulgaris* with the transformation vector is achieved through well-known transformation techniques including electroporation or other known methods. Activity of the hygromycin B resistance gene product may be used as a marker to select for *Chlorella vulgaris* transformed with the transformation vector on, but not limited to, agar medium comprising hygromycin. Growth media suitable for *Chlorella vulgaris* lipid production include, but are not limited to BG11 medium (0.04 g/L KH₂PO₄, 0.075 g/L CaCl₂, 0.036 g/L citric acid, 0.006 g/L Ammonium Ferric Citrate, 1mg/L

EDTA, and 0.02 g/L Na₂CO₃) supplemented with trace metals, and optionally 1.5 g/L NaNO₃. Additional media suitable for culturing *Chlorella vulgaris* for lipid production include, for example, Watanabe medium (comprising 1.5 g/L KNO₃, 1.25 g/L KH₂PO₄, 1.25 g l⁻¹ MgSO₄·7H₂O, 20 mg l⁻¹ FeSO₄·7H₂O with micronutrients and low-nitrogen
 5 medium (comprising 203 mg/l (NH₄)₂HPO₄, 2.236 g/l KCl, 2.465 g/l MgSO₄, 1.361 g/l KH₂PO₄ and 10 mg/l FeSO₄) as reported by Illman et al., *Enzyme and Microbial Technology*, Vol. 27 (2000), pp. 631–635. Evaluation of fatty acid profiles of *Chlorella vulgaris* lipids may be assessed through standard lipid extraction and analytical methods described herein.

10

EXAMPLE 12: ENGINEERING *CHLORELLA ELLIPSOIDEA*

[0302] Expression of recombinant genes in accordance with the present invention in *Chlorella ellipsoidea* may be accomplished by modifying the methods and vectors taught by Chen *et al.* as discussed herein. Briefly, Chen *et al.*, *Current Genetics*, Vol. 39:5
 15 (2001), pp. 365-370, reported the stable transformation of *Chlorella ellipsoidea* with plasmid DNA. Using the transformation method of electroporation, Chen introduced the plasmid pBinUΩNP-1 into *Chlorella ellipsoidea*. The nucleotide sequence of pBinUΩNP-1 comprised sequence encoding the neutrophil peptide-1 (NP-1) rabbit gene product operably linked to a *Zea mays* Ubiquitin (ubi1) gene promoter upstream of the
 20 NP-1 protein-coding region and operably linked to the 3' UTR/terminator of the nopaline synthase (*nos*) gene downstream of the NP-1 protein-coding region. The sequence of plasmid pBinUΩNP-1 further comprised a G418 antibiotic resistance cassette. This G418 antibiotic resistance cassette comprised sequence encoding the aminoglycoside 3'-phosphotransferase (*aph 3'*) gene product. The *aph 3'* gene product confers resistance to
 25 the antibiotic G418. Prior to transformation, *Chlorella ellipsoidea* was unable to be propagated in culture medium comprising 30 ug/mL G418. Upon transformation with the pBinUΩNP-1 plasmid, transformants of *Chlorella ellipsoidea* were obtained that were propagated in selective culture medium comprising 30 ug/mL G418. The expression of the *aph 3'* gene product in *Chlorella ellipsoidea* enabled propagation of transformed
 30 *Chlorella ellipsoidea* in the presence of 30 ug/mL G418, thereby establishing the utility of the G418 antibiotic resistance cassette as selectable marker for use in *Chlorella ellipsoidea*. Detectable activity of the NP-1 gene product indicated that the ubi1 promoter and *nos* 3' UTR are suitable for enabling heterologous gene expression in *Chlorella ellipsoidea*. Evaluation of the genomic DNA of the stable transformants was performed

by Southern analysis. Selection and maintenance of the transformed *Chlorella ellipsoidea* was performed on Knop medium (comprising 0.2 g/L K₂HPO₄, 0.2 g/L MgSO₄·7H₂O, 0.12 g/L KCl, and 10 mg/L FeCl₃, pH 6.0-8.0 supplemented with 0.1% yeast extract and 0.2% glucose) with 15 ug/mL G418 (for liquid cultures) or with 30 ug/mL G418 (for solid
5 cultures comprising 1.8% agar). Propagation of *Chlorella ellipsoidea* in media other than Knop medium has been reported (see Cho *et al.*, Fisheries Science, Vol. 73:5 (2007), pp. 1050-1056, Jarvis and Brown, *Current Genetics*, Vol. 19 (1991), pp.317-321 and Kim *et al.*, *Marine Biotechnology*, Vol. 4 (2002), pp.63-73). Additional plasmids, promoters, 3'UTR/terminators, and selectable markers suitable for enabling heterologous gene
10 expression in *Chlorella ellipsoidea* have been reported (see Jarvis and Brown and Kim *et al.*, *Marine Biotechnology*, Vol. 4 (2002), pp.63-73). Chen reported that the plasmid pBinUΩNP-1, the ubi1 promoter, and the *Agrobacterium tumefaciens* nopaline synthase gene 3'UTR/terminator are suitable to enable exogenous gene expression in *Chlorella ellipsoidea*. In addition, Chen reported that the G418 resistance cassette encoded on
15 pBinUΩNP-1 was suitable for use as a selectable marker in *Chlorella ellipsoidea*.

[0303] In an embodiment of the present invention, vector pBinUΩNP-1, comprising the nucleotide sequence encoding the *aph 3'* gene product, conferring resistance to G418, for use as a selectable marker, is constructed and modified to further comprise a lipid biosynthesis pathway expression cassette sequence, thereby creating a transformation
20 vector. The lipid biosynthesis pathway expression cassette encodes one or more lipid biosynthesis pathway proteins selected from Table 25, each protein-coding sequence codon-optimized for expression in *Chlorella ellipsoidea* to reflect the codon bias inherent in nuclear genes of *Chlorella ellipsoidea* in accordance with Tables 24A-D. For each lipid biosynthesis pathway protein of Table 25, the codon-optimized gene sequence can
25 individually be operably linked to the *Zea mays* ubi1 promoter upstream of the protein-coding sequence and operably linked to the *Agrobacterium tumefaciens* nopaline synthase gene 3'UTR/terminator at the 3' region, or downstream, of the protein-coding sequence. The transformation construct may additionally comprise homology regions to the *Chlorella ellipsoidea* genome for targeted genomic integration of the transformation
30 vector. Homology regions may be selected to disrupt one or more genomic sites of endogenous lipid biosynthesis pathway genes. Stable transformation of *Chlorella ellipsoidea* with the transformation vector is achieved through well-known transformation techniques including electroporation or other known methods. Activity of the *aph 3'* gene product may be used as a marker to select for *Chlorella ellipsoidea* transformed with the

transformation vector on, but not limited to, Knop agar medium comprising G418.

Growth media suitable for *Chlorella ellipsoidea* lipid production include, but are not limited to, Knop medium and those culture medium reported by Jarvis and Brown and Kim *et al.* Evaluation of fatty acid profiles of *Chlorella ellipsoidea* lipids may be assessed through standard lipid extraction and analytical methods described herein.

EXAMPLE 13: ENGINEERING *CHLORELLA KESSLERI*

[0304] Expression of recombinant genes in accordance with the present invention in *Chlorella kessleri* may be accomplished by modifying the methods and vectors taught by El-Sheekh *et al.* as discussed herein. Briefly, El-Sheekh *et al.*, *Biologia Plantarum*, Vol. 42:2 (1999), pp. 209-216, reported the stable transformation of *Chlorella kessleri* with plasmid DNA. Using the transformation method of microprojectile bombardment, El-Sheekh introduced the plasmid pBI121 (GenBank Accession No. AF485783) into *Chlorella kessleri*. Plasmid pBI121 comprised a kanamycin/neomycin antibiotic resistance cassette. This kanamycin/neomycin antibiotic resistance cassette comprised the *Agrobacterium tumefaciens* nopaline synthase (*nos*) gene promoter, sequence encoding the neomycin phosphotransferase II (*nptII*) gene product (GenBank Accession No. AAL92039) for resistance to kanamycin and G418, and the 3' UTR/terminator of the *Agrobacterium tumefaciens* nopaline synthase (*nos*) gene. pBI121 further comprised sequence encoding a beta-glucuronidase (GUS) reporter gene product operably linked to a CaMV 35S promoter and operably linked to a 3' UTR/terminator of the *nos* gene. Prior to transformation, *Chlorella kessleri* was unable to be propagated in culture medium comprising 15 ug/L kanamycin. Upon transformation with the pBI121 plasmid, transformants of *Chlorella kessleri* were obtained that were propagated in selective culture medium comprising 15 mg/L kanamycin. The expression of the *nptII* gene product in *Chlorella kessleri* enabled propagation in the presence of 15 mg/L kanamycin, thereby establishing the utility of the kanamycin/neomycin antibiotic resistance cassette as selectable marker for use in *Chlorella kessleri*. Detectable activity of the GUS gene product indicated that the CaMV 35S promoter and *nos* 3' UTR are suitable for enabling heterologous gene expression in *Chlorella kessleri*. Evaluation of the genomic DNA of the stable transformants was performed by Southern analysis. As reported by El-Sheekh, selection and maintenance of transformed *Chlorella kessleri* was conducted on semisolid agar plates comprising YEG medium (1% yeast extract, 1% glucose) and 15 mg/L kanamycin. El-Sheekh also reported the propagation of *Chlorella kessleri* in YEG liquid

culture media. Additional media suitable for culturing *Chlorella kessleri* for lipid production are disclosed in Sato *et al.*, *BBA Molecular and Cell Biology of Lipids*, Vol. 1633 (2003), pp. 27-34). El-Sheekh reported that the plasmid pBI121, the CaMV promoter, and the nopaline synthase gene 3'UTR/terminator are suitable to enable
5 heterologous gene expression in *Chlorella kessleri*. In addition, El-Sheekh reported that the kanamycin/neomycin resistance cassette encoded on pBI121 was suitable for use as a selectable marker in *Chlorella kessleri*.

[0305] In an embodiment of the present invention, vector pBI121, comprising the nucleotide sequence encoding the kanamycin/neomycin resistance gene product for use as
10 a selectable marker, is constructed and modified to further comprise a lipid biosynthesis pathway expression cassette sequence, thereby creating a transformation vector. The lipid biosynthesis pathway expression cassette encodes one or more lipid biosynthesis pathway proteins selected from Table 25, each protein-coding sequence codon-optimized for expression in *Chlorella kessleri* to reflect the codon bias inherent in nuclear genes of
15 *Chlorella kessleri* in accordance with Tables 24A-D. For each lipid biosynthesis pathway protein of Table 25, the codon-optimized gene sequence can individually be operably linked to the CaMV 35S promoter upstream of the protein-coding sequence and operably linked to the *Agrobacterium tumefaciens* nopaline synthase gene 3'UTR/terminator at the 3' region, or downstream, of the protein-coding sequence. The transformation construct
20 may additionally comprise homology regions to the *Chlorella kessleri* genome for targeted genomic integration of the transformation vector. Homology regions may be selected to disrupt one or more genomic sites of endogenous lipid biosynthesis pathway genes. Stable transformation of *Chlorella kessleri* with the transformation vector is achieved through well-known transformation techniques including microprojectile bombardment or other
25 known methods. Activity of the *nptII* gene product may be used as a marker to select for *Chlorella kessleri* transformed with the transformation vector on, but not limited to, YEG agar medium comprising kanamycin or neomycin. Growth media suitable for *Chlorella kessleri* lipid production include, but are not limited to, YEG medium, and those culture media reported by Sato *et al.* Evaluation of fatty acid profiles of *Chlorella kessleri* lipids
30 may be assessed through standard lipid extraction and analytical methods described herein.

EXAMPLE 14: ENGINEERING *DUNALIELLA TERTIOLECTA*

[0306] Expression of recombinant genes in accordance with the present invention in *Dunaliella tertiolecta* may be accomplished by modifying the methods and vectors taught
5 by Walker *et al.* as discussed herein. Briefly, Walker *et al.*, *Journal of Applied Phycology*, Vol. 17 (2005), pp. 363-368, reported stable nuclear transformation of *Dunaliella tertiolecta* with plasmid DNA. Using the transformation method of electroporation, Walker introduced the plasmid pDbleFLAG1.2 into *Dunaliella tertiolecta*.

pDbleFLAG1.2 comprised sequence encoding a bleomycin antibiotic resistance cassette,
10 comprising sequence encoding the *Streptoalloteichus hindustanus* Bleomycin binding protein (*ble*), for resistance to the antibiotic phleomycin, operably linked to the promoter and 3' UTR of the *Dunaliella tertiolecta* ribulose-1,5-bisphosphate carboxylase/oxygenase small subunit gene (*rbcS1*, GenBank Accession No. AY530155). Prior to transformation, *Dunaliella tertiolecta* was unable to be propagated in culture medium comprising 1 mg/L
15 phleomycin. Upon transformation with the pDbleFLAG1.2 plasmid, transformants of *Dunaliella tertiolecta* were obtained that were propagated in selective culture medium comprising 1 mg/L phleomycin. The expression of the *ble* gene product in *Dunaliella tertiolecta* enabled propagation in the presence of 1 mg/L phleomycin, thereby establishing the utility of the bleomycin antibiotic resistance cassette as selectable marker
20 for use in *Dunaliella tertiolecta*. Evaluation of the genomic DNA of the stable transformants was performed by Southern analysis. As reported by Walker, selection and maintenance of transformed *Dunaliella tertiolecta* was conducted in Dunaliella medium (DM, as described by Provasoli *et al.*, *Archiv fur Mikrobiologie*, Vol. 25 (1957), pp. 392-428) further comprising 4.5 g/L NaCl and 1 mg/L pheomycin. Additional media suitable
25 for culturing *Dunaliella tertiolecta* for lipid production are discussed in Takagi *et al.*, *Journal of Bioscience and Bioengineering*, Vol. 101:3 (2006), pp. 223-226 and in Massart and Hanston, Proceedings Venice 2010, *Third International Symposium on Energy from Biomass and Waste*. Walker reported that the plasmid pDbleFLAG1.2 and the promoter and 3' UTR of the *Dunaliella tertiolecta* ribulose-1,5-bisphosphate carboxylase/oxygenase
30 small subunit gene are suitable to enable heterologous expression in *Dunaliella tertiolecta*. In addition, Walker reported that the bleomycin resistance cassette encoded on pDbleFLAG1.2 was suitable for use as a selectable marker in *Dunaliella tertiolecta*.

[0307] In an embodiment of the present invention, vector pDbleFLAG1.2, comprising the nucleotide sequence encoding the *ble* gene product for use as a selectable marker, is

constructed and modified to further comprise a lipid biosynthesis pathway expression cassette sequence, thereby creating a transformation vector. The lipid biosynthesis pathway expression cassette encodes one or more lipid biosynthesis pathway proteins selected from Table 25, each protein-coding sequence codon-optimized for expression in

5 *Dunaliella tertiolecta* to reflect the codon bias inherent in nuclear genes of *Dunaliella tertiolecta* in accordance with Tables 24A-D. For each lipid biosynthesis pathway protein of Table 25, the codon-optimized gene sequence can individually be operably linked to the *rbcS1* promoter upstream of the protein-coding sequence and operably linked to the *rbcS1* 3'UTR/terminator at the 3' region, or downstream, of the protein-coding sequence. The

10 transformation construct may additionally comprise homology regions to the *Dunaliella tertiolecta* genome for targeted genomic integration of the transformation vector. Homology regions may be selected to disrupt one or more genomic sites of endogenous lipid biosynthesis pathway genes. Stable transformation of *Dunaliella tertiolecta* with the transformation vector is achieved through well-known transformation techniques

15 including electroporation or other known methods. Activity of the *ble* gene product may be used as a marker to select for *Dunaliella tertiolecta* transformed with the transformation vector on, but not limited to, DM medium comprising pheomycin. Growth medium suitable for *Dunaliella tertiolecta* lipid production include, but are not limited to DM medium and those culture media described by Takagi *et al.* and Massart and Hanston.

20 Evaluation of fatty acid profiles of *Dunaliella tertiolecta* lipids may be assessed through standard lipid extraction and analytical methods described herein.

EXAMPLE 15: ENGINEERING *VOLVOX CARTERI*

[0308] Expression of recombinant genes in accordance with the present invention in

25 *Volvox carteri* may be accomplished by modifying the methods and vectors taught by Hallman and Rappel *et al.* as discussed herein. Briefly, Hallman and Rappel *et al.*, *The Plant Journal*, Volume 17 (1999), pp. 99-109, reported the stable nuclear transformation of *Volvox carteri* with plasmid DNA. Using the transformation method of microprojectile bombardment, Hallman and Rappel introduced the pzeoE plasmid into *Volvox carteri*.

30 The pzeoE plasmid comprised sequence encoding a bleomycin antibiotic resistance cassette, comprising sequence encoding the *Streptoalloteichus hindustanus* Bleomycin binding protein (*ble*), for resistance to the antibiotic zeocin, operably linked to and the promoter and 3' UTR of the *Volvox carteri* beta-tubulin gene (GenBank Accession No. L24547). Prior to transformation, *Volvox carteri* was unable to be propagated in culture

medium comprising 1.5 ug/ml zeocin. Upon transformation with the pzeoE plasmid, transformants of *Volvox carteri* were obtained that were propagated in selective culture medium comprising greater than 20 ug/ml zeocin. The expression of the *ble* gene product in *Volvox carteri* enabled propagation in the presence of 20 ug/ml zeocin, thereby

5 establishing the utility of the bleomycin antibiotic resistance cassette as selectable marker for use in *Volvox carteri*. Evaluation of the genomic DNA of the stable transformants was performed by Southern analysis. As reported by Hallman and Rappel, selection and maintenance of transformed *Volvox carteri* was conducted in Volvox medium (VM, as described by Provasoli and Pintner, The Ecology of Algae, Special Publication No. 2

10 (1959), Tyron, C.A. and Hartman, R.T., eds., Pittsburgh: Univeristy of Pittsburgh, pp. 88-96) with 1 mg/L pheomycin. Media suitable for culturing *Volvox carteri* for lipid production are also discussed by Starr in Starr R,C., *Dev Biol Suppl.*, Vol. 4 (1970), pp.59-100). Hallman and Rappel reported that the plasmid pzeoE and the promoter and 3' UTR of the *Volvox carteri* beta-tubulin gene are suitable to enable heterologous expression in

15 *Volvox carteri*. In addition, Hallman and Rappel reported that the bleomycin resistance cassette encoded on pzeoE was suitable for use as a selectable marker in *Volvox carteri*. Additional plasmids, promoters, 3'UTR/terminators, and selectable markers suitable for enabling heterologous gene expression in *Volvox carteri* and suitable for use as selective markers *Volvox carteri* in have been reported (for instance see Hallamann and Sumper,

20 *Proceedings of the National Academy of Sciences*, Vol. 91 (1994), pp 11562-11566 and Hallman and Wodniok, *Plant Cell Reports*, Volume 25 (2006), pp. 582-581).

[0309] In an embodiment of the present invention, vector pzeoE, comprising the nucleotide sequence encoding the *ble* gene product for use as a selectable marker, is constructed and modified to further comprise a lipid biosynthesis pathway expression

25 cassette sequence, thereby creating a transformation vector. The lipid biosynthesis pathway expression cassette encodes one or more lipid biosynthesis pathway proteins selected from Table 25, each protein-coding sequence codon-optimized for expression in *Volvox carteri* to reflect the codon bias inherent in nuclear genes of *Volvox carteri* in accordance with Tables 24A-D. For each lipid biosynthesis pathway protein of Table 25,

30 the codon-optimized gene sequence can individually be operably linked to the *Volvox carteri* beta-tubulin promoter upstream of the protein-coding sequence and operably linked to the *Volvox carteri* beta-tubulin 3'UTR/terminator at the 3' region, or downstream, of the protein-coding sequence. The transformation construct may additionally comprise homology regions to the *Volvox carteri* genome for targeted

genomic integration of the transformation vector. Homology regions may be selected to disrupt one or more genomic sites of endogenous lipid biosynthesis pathway genes. One skilled in the art can identify such homology regions within the sequence of the *Volvox carteri* genome (referenced in the publication by Prochnik et al., *Science*, Vol. 329:5988 (2010), pp223-226). Stable transformation of *Volvox carteri* with the transformation vector is achieved through well-known transformation techniques including microprojectile bombardment or other known methods. Activity of the *ble* gene product may be used as a marker to select for *Volvox carteri* transformed with the transformation vector on, but not limited to, VM medium comprising zeocin. Growth medium suitable for *Volvox carteri* lipid production include, but are not limited to VM medium and those culture media discussed by Starr. Evaluation of fatty acid profiles of *Volvox carteri* lipids may be assessed through standard lipid extraction and analytical methods described herein.

EXAMPLE 16: ENGINEERING *HAEMATOCOCCUS PLUVIALIS*

[0310] Expression of recombinant genes in accordance with the present invention in *Haematococcus pluvialis* may be accomplished by modifying the methods and vectors taught by Steinbrenner and Sandmann *et al.* as discussed herein. Briefly, Steinbrenner and Sandmann *et al.*, *Applied and Environmental Microbiology*, Vol. 72:12 (2006), pp.7477-7484, reported the stable nuclear transformation of *Haematococcus pluvialis* with plasmid DNA. Using the transformation method of microprojectile bombardment, Steinbrenner introduced the plasmid pPlat-pds-L504R into *Haematococcus pluvialis*. The plasmid pPlat-pds-L504R comprised a norflurazon resistance cassette, which comprised the promoter, protein-coding sequence, and 3'UTR of the *Haematococcus pluvialis* phytoene desaturase gene (Pds, GenBank Accession No. AY781170), wherein the protein-coding sequence of Pds was modified at position 504 (thereby changing a leucine to an arginine) to encode a gene product (Pds-L504R) that confers resistance to the herbicide norflurazon. Prior to transformation with pPlat-pds-L504R, *Haematococcus pluvialis* was unable to propagate on medium comprising 5 uM norflurazon. Upon transformation with the pPlat-pds-L504R plasmid, transformants of *Haematococcus pluvialis* were obtained that were propagated in selective culture medium comprising 5 uM norflurazon. The expression of the Pds-L504R gene product in *Haematococcus pluvialis* enabled propagation in the presence of 5 uM norflurazon, thereby establishing the utility of the norflurazon herbicide resistance cassette as selectable marker for use in *Haematococcus pluvialis*. Evaluation of

the genomic DNA of the stable transformants was performed by Southern analysis. As reported by Steinbrenner, selection and maintenance of transformed *Haematococcus pluvialis* was conducted on agar plates comprising OHA medium (OHM (0.41 g/L KNO₃, 0.03 g/L Na₂HPO₄, 0.246 g/L MgSO₄·7H₂O, 0.11 g/L CaCl₂·2H₂O, 2.62 mg/L Fe_(III)citrate x H₂O, 0.011 mg/L CoCl₂·6H₂O, 0.012 mg/L CuSO₄·5H₂O, 0.075 mg/L Cr₂O₃, 0.98 mg/L MnCl₂·4H₂O, 0.12 mg/L Na₂MoO₄ x 2H₂O, 0.005 mg/L SeO₂ and 25 mg/L biotin, 17.5 mg/L thiamine, and 15 mg/L vitamin B12), supplemented with 2.42 g/L Tris-acetate, and 5mM norflurazon. Propagation of *Haematococcus pluvialis* in liquid culture was performed by Steinbrenner and Sandmann using basal medium (basal medium as described by Kobayashi *et al.*, *Applied and Environmental Microbiology*, Vol. 59 (1993), pp.867-873). Steinbrenner and Sandmann reported that the pPlat-pds-L504R plasmid and promoter and 3' UTR of the *Haematococcus pluvialis* phytoene desaturase gene are suitable to enable heterologous expression in *Haematococcus pluvialis*. In addition, Steinbrenner and Sandmann reported that the norflurazon resistance cassette encoded on pPlat-pds-L504R was suitable for use as a selectable marker in *Haematococcus pluvialis*. Additional plasmids, promoters, 3'UTR/terminators, and selectable markers suitable for enabling heterologous gene expression in *Haematococcus pluvialis* have been reported (see Kathiresan *et al.*, *Journal of Phycology*, Vol. 45 (2009), pp 642-649).

[0311] In an embodiment of the present invention, vector pPlat-pds-L504R, comprising the nucleotide sequence encoding the Pds-L504R gene product for use as a selectable marker, is constructed and modified to further comprise a lipid biosynthesis pathway expression cassette sequence, thereby creating a transformation vector. The lipid biosynthesis pathway expression cassette encodes one or more lipid biosynthesis pathway proteins selected from Table 25, each protein-coding sequence codon-optimized for expression in *Haematococcus pluvialis* to reflect the codon bias inherent in nuclear genes of *Haematococcus pluvialis* in accordance with Tables 24 A-D. For each lipid biosynthesis pathway protein of Table 25, the codon-optimized gene sequence can individually be operably linked to the *Haematococcus pluvialis* pds gene promoter upstream of the protein-coding sequence and operably linked to the *Haematococcus pluvialis* pds gene 3'UTR/terminator at the 3' region, or downstream, of the protein-coding sequence. The transformation construct may additionally comprise homology regions to the *Haematococcus pluvialis* genome for targeted genomic integration of the transformation vector. Homology regions may be selected to disrupt one or more genomic sites of endogenous lipid biosynthesis pathway genes. Stable transformation of

Haematococcus pluvialis with the transformation vector is achieved through well-known transformation techniques including microprojectile bombardment or other known methods. Activity of the Pds-L504R gene product may be used as a marker to select for *Haematococcus pluvialis* transformed with the transformation vector on, but not limited to, OHA medium comprising norflurazon. Growth media suitable for *Haematococcus pluvialis* lipid production include, but are not limited to basal medium and those culture media described by Kobayashi *et al.*, Kathiresan *et al.*, and Gong and Chen, *Journal of Applied Phycology*, Vol. 9:5 (1997), pp. 437-444). Evaluation of fatty acid profiles of *Haematococcus pluvialis* lipids may be assessed through standard lipid extraction and analytical methods described herein.

EXAMPLE 17: ENGINEERING *CLOSTERIUM PERACEROSUM-STRIGOSUM-LITTORALE* COMPLEX

[0312] Expression of recombinant genes in accordance with the present invention in *Closterium peracerosum-strigosum-littorale* complex may be accomplished by modifying the methods and vectors taught by Abe *et al.* as discussed herein. Briefly, Abe *et al.*, *Plant Cell Physiology*, Vol. 52:9 (2011), pp. 1676-1685, reported the stable nuclear transformation of *Closterium peracerosum-strigosum-littorale* complex with plasmid DNA. Using the transformation methods of microprojectile bombardment, Abe introduced the plasmid pSA106 into *Closterium peracerosum-strigosum-littorale* complex. Plasmid pSA106 comprised a bleomycin resistance cassette, comprising sequence encoding the *Streptoalloteichus hindustanus* Bleomycin binding protein gene (ble, GenBank Accession No. CAA37050) operably linked to the promoter and 3' UTR of the *Closterium peracerosum-strigosum-littorale* complex *Chlorophyll a/b-binding protein* gene (CAB, GenBank Accession No. AB363403). Prior to transformation with pSA106, *Closterium peracerosum-strigosum-littorale* complex was unable to propagate on medium comprising 3 ug/ml phleomycin. Upon transformation with pSA106, transformants of *Closterium peracerosum-strigosum-littorale* complex were obtained that were propagated in selective culture medium comprising 3 ug/ml phleomycin. The expression of the ble gene product in *Closterium peracerosum-strigosum-littorale* complex enabled propagation in the presence of 3 ug/ml phleomycin, thereby establishing the utility of the bleomycin antibiotic resistance cassette as selectable marker for use in *Closterium peracerosum-strigosum-littorale* complex. Evaluation of the genomic DNA of the stable transformants was performed by Southern analysis. As reported by Abe, selection and

maintenance of transformed *Closterium peracerosum-strigosum-littorale* complex was conducted first in top agar with C medium (0.1 g/L KNO₃, 0.015 g/L Ca(NO₃)₂·4H₂O, 0.05 g/L glycerophosphate-Na₂, 0.04 g/L MgSO₄·7H₂O, 0.5 g/L Tris (hydroxymethyl) aminomethane, trace minerals, biotin, vitamins B₁ and B₁₂) and then subsequently isolated to agar plates comprising C medium supplemented with phleomycin. As reported by Abe, propagation of *Closterium peracerosum-strigosum-littorale* complex in liquid culture was performed in C medium. Additional liquid culture medium suitable for propagation of *Closterium peracerosum-strigosum-littorale* complex are discussed by Sekimoto *et al.*, *DNA Research*, 10:4 (2003), pp. 147-153. Abe reported that the pSA106 plasmid and promoter and 3' UTR of the *Closterium peracerosum-strigosum-littorale* complex CAB gene are suitable to enable heterologous gene expression in *Closterium peracerosum-strigosum-littorale* complex. In addition, Abe reported that the bleomycin resistance cassette encoded on pSA106 was suitable for use as a selectable marker in *Closterium peracerosum-strigosum-littorale* complex. Additional plasmids, promoters, 3'UTR/terminators, and selectable markers suitable for enabling heterologous gene expression in *Closterium peracerosum-strigosum-littorale* complex have been reported (see Abe *et al.*, *Plant Cell Physiology*, Vol. 49 (2008), pp. 625-632).

[0313] In an embodiment of the present invention, vector pSA106, comprising the nucleotide sequence encoding the *ble* gene product for use as a selectable marker, is constructed and modified to further comprise a lipid biosynthesis pathway expression cassette sequence, thereby creating a transformation vector. The lipid biosynthesis pathway expression cassette encodes one or more lipid biosynthesis pathway proteins selected from Table 25, each protein-coding sequence codon-optimized for expression in *Closterium peracerosum-strigosum-littorale* complex to reflect the codon bias inherent in nuclear genes of *Closterium peracerosum-strigosum-littorale* complex in accordance with Tables 24A-D. For each lipid biosynthesis pathway protein of Table 25, the codon-optimized gene sequence can individually be operably linked to the *Closterium peracerosum-strigosum-littorale* complex CAB gene promoter upstream of the protein-coding sequence and operably linked to the *Closterium peracerosum-strigosum-littorale* complex CAB gene 3'UTR/terminator at the 3' region, or downstream, of the protein-coding sequence. The transformation construct may additionally comprise homology regions to the *Closterium peracerosum-strigosum-littorale* complex genome for targeted genomic integration of the transformation vector. Homology regions may be selected to disrupt one or more genomic sites of endogenous lipid biosynthesis pathway genes. Stable

transformation of *Closterium peracerosum-strigosum-littorale* complex with the transformation vector is achieved through well-known transformation techniques including microprojectile bombardment or other known methods. Activity of the *ble* gene product may be used as a marker to select for *Closterium peracerosum-strigosum-littorale* complex transformed with the transformation vector on, but not limited to, C medium comprising phleomycin. Growth media suitable for *Closterium peracerosum-strigosum-littorale* complex lipid production include, but are not limited to C medium and those culture media reported by Abe *et al.* and Sekimoto *et al.* Evaluation of fatty acid profiles of *Closterium peracerosum-strigosum-littorale* complex lipids may be assessed through standard lipid extraction and analytical methods described herein.

EXAMPLE 18: ENGINEERING *DUNALIELLA VIRIDIS*

[0314] Expression of recombinant genes in accordance with the present invention in *Dunaliella viridis* may be accomplished by modifying the methods and vectors taught by Sun *et al.* as discussed herein. Briefly, Sun *et al.*, *Gene*, Vol. 377 (2006), pp.140-149, reported the stable transformation of *Dunaliella viridis* with plasmid DNA. Using the transformation method of electroporation, Sun introduced the plasmid pDVNR, encoding the full *Dunaliella viridis* nitrate reductase gene into mutant *Dunaliella viridis* (*Dunaliella viridis* NR-mutants.) The NR-mutants are incapable of growth without the use of nitrate as a source of nitrogen. Nitrate reductase catalyzes the conversion of nitrate to nitrite. Prior to transformation, *Dunaliella viridis* NR-mutants were unable to propagate in culture medium comprising nitrate (NO_3^-) as the sole nitrogen source. The expression of the *Dunaliella viridis* NR gene product in NR-mutant *Dunaliella viridis* was used as a selectable marker to rescue the nitrate metabolism deficiency. Upon transformation with the pDVNR plasmid, NR-mutant *Dunaliella viridis* stably expressing the *Dunaliella viridis* NR gene product were obtained that were able to grow on agar plates comprising nitrate as the sole carbon source. Evaluation of the DNA of the stable transformants was performed by Southern analysis. Selection and maintenance of the transformed *Dunaliella viridis* (NR mutant) was performed on agar plates comprising 5 mM KNO_3 . Sun also reported the propagation of *Dunaliella viridis* and *Dunaliella viridis* NR mutants in liquid culture medium. Additional media suitable for propagation of *Dunaliella viridis* are reported by Gordillo *et al.*, *Journal of Applied Phycology*, Vol. 10:2 (1998), pp. 135-144 and by Moulton and Burford, *Hydrobiologia*, Vols. 204-205:1 (1990), pp. 401-408. Sun reported that the plasmid pDVNR and the promoter and 3' UTR/terminator of the

Dunaliella viridis nitrate reductase gene were suitable to enable heterologous expression in *Dunaliella viridis* NR-mutants. Sun also reported that expression of the *Dunaliella viridis* nitrate reductase gene product was suitable for use as a selectable marker in *Dunaliella viridis* NR-mutants.

5 [0315] In an embodiment of the present invention, vector pDVNR, comprising the nucleotide sequence encoding the *Dunaliella viridis* nitrate reductase (*DvNR*) gene product for use as a selectable marker, is constructed and modified to further comprise a lipid biosynthesis pathway expression cassette sequence, thereby creating a transformation vector. The lipid biosynthesis pathway expression cassette encodes one or more lipid
10 biosynthesis pathway proteins selected Table 25, each protein-coding sequence codon-optimized for expression in *Dunaliella viridis* to reflect the codon bias inherent in nuclear genes of *Dunaliella viridis* in accordance with Tables 24A-D. For each lipid biosynthesis pathway protein of Table 25, the codon-optimized gene sequence can individually be operably linked to the *DvNR* promoter upstream of the protein-coding sequence and
15 operably linked to the *DvNR* 3'UTR/terminator at the 3' region, or downstream, of the protein-coding sequence. The transformation construct may additionally comprise homology regions to the *Dunaliella viridis* genome for targeted genomic integration of the transformation vector. Homology regions may be selected to disrupt one or more genomic sites of endogenous lipid biosynthesis pathway genes. Stable transformation of *Dunaliella*
20 *viridis* NR mutants with the transformation vector is achieved through well-known transformation techniques including electroporation or other known methods. Activity of the *DvNR* gene product may be used as a selectable marker to rescue the nitrogen assimilation deficiency of *Dunaliella viridis* NR mutant strains and to select for *Dunaliella viridis* NR-mutants stably expressing the transformation vector. Growth media
25 suitable for *Dunaliella viridis* lipid production include, but are not limited to those discussed by Sun *et al.*, Moulton and Burford, and Gordillo *et al.* Evaluation of fatty acid profiles of *Dunaliella viridis* lipids may be assessed through standard lipid extraction and analytical methods described herein.

30 EXAMPLE 19: ENGINEERING *DUNALIELLA SALINA*

[0316] Expression of recombinant genes in accordance with the present invention in *Dunaliella salina* may be accomplished by modifying the methods and vectors taught by Geng *et al.* as discussed herein. Briefly, Geng *et al.*, *Journal of Applied Phycology*, Vol. 15 (2003), pp. 451-456, reported the stable transformation of *Dunaliella salina* with

plasmid DNA. Using the transformation method of electroporation, Geng introduced the pUΩHBsAg-CAT plasmid into *Dunaliella salina*. pUΩHBsAg-CAT comprises a hepatitis B surface antigen (HBsAg) expression cassette comprising sequence encoding the hepatitis B surface antigen operably linked to a *Zea mays* ubi1 promoter upstream of the HBsAg protein-coding region and operably linked to the 3'UTR/terminator of the *Agrobacterium tumefaciens* nopaline synthase gene (*nos*) downstream of the HBsAg protein-coding region. pUΩHBsAg-CAT further comprised a chloramphenicol resistance cassette, comprising sequence encoding the chloramphenicol *acetyltransferase* (CAT) gene product, conferring resistance to the antibiotic chloramphenicol, operably linked to the simian virus 40 promoter and enhancer. Prior to transformation with pUΩHBsAg-CAT, *Dunaliella salina* was unable to propagate on medium comprising 60 mg/L chloramphenicol. Upon transformation with the pUΩHBsAg-CAT plasmid, transformants of *Dunaliella salina* were obtained that were propagated in selective culture medium comprising 60 mg/L chloramphenicol. The expression of the CAT gene product in *Dunaliella salina* enabled propagation in the presence of 60 mg/L chloramphenicol, thereby establishing the utility of the chloramphenicol resistance cassette as selectable marker for use in *Dunaliella salina*. Detectable activity of the HBsAg gene product indicated that ubi1 promoter and nos 3'UTR/terminator are suitable for enabling gene expression in *Dunaliella salina*. Evaluation of the genomic DNA of the stable transformants was performed by Southern analysis. Geng reported that selection and maintenance of the transformed *Dunaliella salina* was performed on agar plates comprising Johnson's medium (J1, described by Borowitzka and Borowitzka (eds), Microalgal Biotechnology. Cambridge University Press, Cambridge, pp. 460-461) with 60 mg/L chloramphenicol. Liquid propagation of *Dunaliella salina* was performed by Geng in J1 medium with 60 mg/L chloramphenicol. Propagation of *Dunaliella salina* in media other than J1 medium has been discussed (see Feng *et al.*, *Mol. Bio. Reports*, Vol. 36 (2009), pp.1433-1439 and Borowitzka *et al.*, *Hydrobiologia*, Vols. 116-117:1 (1984), pp. 115-121). Additional plasmids, promoters, 3'UTR/terminators, and selectable markers suitable for enabling heterologous gene expression in *Dunaliella salina* have been reported by Feng *et al.* Geng reported that the plasmid pUΩHBsAg-CAT, the ubi1 promoter, and the *Agrobacterium tumefaciens* nopaline synthase gene 3'UTR/terminator are suitable to enable exogenous gene expression in *Dunaliella salina*. In addition, Geng reported that the CAT resistance cassette encoded on pUΩHBsAg-CAT was suitable for use as a selectable marker in *Dunaliella salina*.

[0317] In an embodiment of the present invention, vector pUΩHBsAg-CAT, comprising the nucleotide sequence encoding the CAT gene product for use as a selectable marker, is constructed and modified to further comprise a lipid biosynthesis pathway expression cassette sequence, thereby creating a transformation vector. The lipid biosynthesis pathway expression cassette encodes one or more lipid biosynthesis pathway proteins selected Table 25, each protein-coding sequence codon-optimized for expression in *Dunaliella salina* to reflect the codon bias inherent in nuclear genes of *Dunaliella salina* in accordance with Tables 24A-D. For each lipid biosynthesis pathway protein of Table 25, the codon-optimized gene sequence can individually be operably linked to the *ubi1* promoter upstream of the protein-coding sequence and operably linked to the *Agrobacterium tumefaciens* nopaline synthase gene 3'UTR/terminator at the 3' region, or downstream, of the protein-coding sequence. The transformation construct may additionally comprise homology regions to the *Dunaliella salina* genome for targeted genomic integration of the transformation vector. Homology regions may be selected to disrupt one or more genomic sites of endogenous lipid biosynthesis pathway genes. Stable transformation of *Dunaliella salina* with the transformation vector is achieved through well-known transformation techniques including electroporation or other known methods. Activity of the CAT gene product may be used as a selectable marker to select for *Dunaliella salina* transformed with the transformation vector in, but not limited to, J1 medium comprising chloramphenicol. Growth medium suitable for *Dunaliella salina* lipid production include, but are not limited to J1 medium and those culture media described by Feng *et al.* and Borowitzka *et al.* Evaluation of fatty acid profiles of *Dunaliella salina* lipids may be assessed through standard lipid extraction and analytical methods described herein.

EXAMPLE 20: ENGINEERING *GONIUM PECTORAL*

[0318] Expression of recombinant genes in accordance with the present invention in *Gonium pectoral* may be accomplished by modifying the methods and vectors taught by Lerche and Hallman *et al.* as discussed herein. Briefly, Lerche and Hallman *et al.*, *BMC Biotechnology*, Volume 9:64, 2009, reported the stable nuclear transformation of *Gonium pectorale* with plasmid DNA. Using the transformation method of microprojectile bombardment, Lerche introduced the plasmid pPmr3 into *Gonium pectorale*. Plasmid pPmr3 comprised a paromomycin resistance cassette, comprising a sequence encoding the aminoglycoside 3'-phosphotransferase (*aphVIII*) gene product (GenBank Accession No.

AAB03856) of *Streptomyces rimosus* for resistance to the antibiotic paromomycin, operably linked to the *Volvox carteri* hsp70A-rbcS3 hybrid promoter upstream of the *aphVIII* protein-coding region and operably linked to the 3' UTR/terminator of the *Volvox carteri* rbcS3 gene downstream of the *aphVIII* protein-coding region. Prior to

5 transformation with pPmr3, *Gonium pectorale* was unable to propagate on medium comprising 0.06 ug/ml paromomycin. Upon transformation with pPmr3, transformants of *Gonium pectorale* were obtained that were propagated in selective culture medium comprising 0.75 and greater ug/ml paromomycin. The expression of the *aphVIII* gene product in *Gonium pectorale* enabled propagation in the presence of 0.75 and greater

10 ug/ml paromomycin, thereby establishing the utility of the paromomycin antibiotic resistance cassette as selectable marker for use in *Gonium pectorale*. Evaluation of the genomic DNA of the stable transformants was performed by Southern analysis. Lerche and Hallman reported that selection and maintenance of the transformed *Gonium pectorale* was performed in liquid Jaworski's medium (20 mg/L $\text{Ca}(\text{NO}_3)_2 \cdot 4\text{H}_2\text{O}$, 12.4 mg/L KH_2PO_4 , 50 mg/L $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$, 15.9 mg/L NaHCO_3 , 2.25 mg/L EDTA-FeNa, 2.25

15 mg/L EDTA Na_2 , 2.48 g/L H_3BO_3 , 1.39 g/L $\text{MnCl}_2 \cdot 4\text{H}_2\text{O}$, 1 mg/L $(\text{NH}_4)_6\text{MO}_7\text{O}_{24} \cdot 4\text{H}_2\text{O}$, 0.04 mg/L vitamin B12, 0.04 mg/L Thiamine-HCl, 0.04 mg/L biotin, 80 mg/L NaNO_3 , 36 mg/L $\text{Na}_4\text{HPO}_4 \cdot 12\text{H}_2\text{O}$) with 1.0 ug/ml paromomycin. Additional plasmids, promoters, 3'UTR/terminators, and selectable markers suitable for enabling heterologous gene

20 expression in *Gonium pectorale* are further discussed by Lerche and Hallman. Lerche and Hallman reported that the plasmid pPmr3, *Volvox carteri* hsp70A-rbcS3 hybrid promoter, and the 3' UTR/terminator of the *Volvox carteri* rbcS3 gene are suitable to enable exogenous gene expression in *Gonium pectorale*. In addition, Lerche and Hallman reported that the paromomycin resistance cassette encoded pPmr3 was suitable for use as a

25 selectable marker in *Gonium pectorale*.

[0319] In an embodiment of the present invention, vector pPmr3, comprising the nucleotide sequence encoding the *aphVIII* gene product for use as a selectable marker, is constructed and modified to further comprise a lipid biosynthesis pathway expression cassette sequence, thereby creating a transformation vector. The lipid biosynthesis

30 pathway expression cassette encodes one or more lipid biosynthesis pathway proteins selected Table 25, each protein-coding sequence codon-optimized for expression in *Gonium pectorale* to reflect the codon bias inherent in nuclear genes of *Gonium pectorale* in accordance with Tables 24A-D. For each lipid biosynthesis pathway protein of Table 25, the codon-optimized gene sequence can individually be operably linked to the *Volvox*

carteri hsp70A-rbcS3 hybrid promoter upstream of the protein-coding sequence and operably linked to the *Volvox carteri* rbcS3 gene 3'UTR/terminator at the 3' region, or downstream, of the protein-coding sequence. The transformation construct may additionally comprise homology regions to the *Gonium pectorale* genome for targeted genomic integration of the transformation vector. Homology regions may be selected to disrupt one or more genomic sites of endogenous lipid biosynthesis pathway genes. Stable transformation of *Gonium pectorale* with the transformation vector may be achieved through well-known transformation techniques including microprojectile bombardment or other known methods. Activity of the *aphVIII* gene product may be used as a selectable marker to select for *Gonium pectorale* transformed with the transformation vector in, but not limited to, Jaworski's medium comprising paromomycin. Growth media suitable for *Gonium pectorale* lipid production include Jawaorski's medium and media reported by Stein, American Journal of Botany, Vol. 45:9 (1958), pp. 664-672. Evaluation of fatty acid profiles of *Gonium pectorale* lipids may be assessed through standard lipid extraction and analytical methods described herein.

EXAMPLE 21: ENGINEERING *PHAEODACTYLUM TRICORNUTUM*

[0320] Expression of recombinant genes in accordance with the present invention in *Phaeodactylum tricornutum* may be accomplished by modifying the methods and vectors taught by Apt *et al.* as discussed herein. Briefly, Apt *et al.*, *Molecular and General Genetics*, Vol. 252 (1996), pp. 572-579, reported the stable nuclear transformation of *Phaeodactylum tricornutum* with vector DNA. Using the transformation technique of microprojectile bombardment, Apt introduced the plasmid pfcA into *Phaeodactylum tricornutum*. Plasmid pfcA comprised a bleomycin resistance cassette, comprising sequence encoding the *Streptoalloteichus hindustanus* Bleomycin binding protein (*ble*), for resistance to the antibiotics phleomycin and zeocin, operably linked to the promoter of the *Phaeodactylum tricornutum* fucoxanthin chlorophyll a binding protein gene (*fcpA*) upstream of the *ble* protein-coding region and operably linked to the 3' UTR/terminator of the *Phaeodactylum tricornutum* *fcpA* gene at the 3' region, or downstream of the *ble* protein-coding region. Prior to transformation with pfcA, *Phaeodactylum tricornutum* was unable to propagate on medium comprising 50 ug/ml zeocin. Upon transformation with pfcA, transformants of *Phaeodactylum tricornutum* were obtained that were propagated in selective culture medium comprising 50 ug/ml zeocin. The expression of the *ble* gene product in *Phaeodactylum tricornutum* enabled propagation in the presence of

50 ug/ml zeocin, thereby establishing the utility of the bleomycin antibiotic resistance cassette as selectable marker for use in *Phaeodactylum tricornutum*. Evaluation of the genomic DNA of the stable transformants was performed by Southern analysis. Apt reported that selection and maintenance of the transformed *Phaeodactylum tricornutum* was performed on agar plates comprising LDM medium (as reported by Starr and Zeikus, *Journal of Phycology*, Vol. 29, Supplement, (1993)) with 50 mg/L zeocin. Apt reported liquid propagation of *Phaeodactylum tricornutum* transformants in LDM medium with 50 mg/L zeocin. Propagation of *Phaeodactylum tricornutum* in medium other than LDM medium has been discussed (by Zaslavskaia *et al.*, *Science*, Vol. 292 (2001), pp. 2073-2075, and by Radokovits *et al.*, *Metabolic Engineering*, Vol. 13 (2011), pp. 89-95). Additional plasmids, promoters, 3'UTR/terminators, and selectable markers suitable for enabling heterologous gene expression in *Phaeodactylum tricornutum* have been reported in the same report by Apt *et al.*, by Zaslavskaia *et al.*, and by Radokovits *et al.*). Apt reported that the plasmid pfcfA, and the *Phaeodactylum tricornutum* fcpA promoter and 3' UTR/terminator are suitable to enable exogenous gene expression in *Phaeodactylum tricornutum*. In addition, Apt reported that the bleomycin resistance cassette encoded on pfcfA was suitable for use as a selectable marker in *Phaeodactylum tricornutum*.

[0321] In an embodiment of the present invention, vector pfcfA, comprising the nucleotide sequence encoding the *ble* gene product for use as a selectable marker, is constructed and modified to further comprise a lipid biosynthesis pathway expression cassette sequence, thereby creating a transformation vector. The lipid biosynthesis pathway expression cassette encodes one or more lipid biosynthesis pathway proteins selected Table 25, each protein-coding sequence codon-optimized for expression in *Phaeodactylum tricornutum* to reflect the codon bias inherent in nuclear genes of *Phaeodactylum tricornutum* in accordance with Tables 24A-D. For each lipid biosynthesis pathway protein of Table 25, the codon-optimized gene sequence can individually be operably linked to the *Phaeodactylum tricornutum* fcpA gene promoter upstream of the protein-coding sequence and operably linked to the *Phaeodactylum tricornutum* fcpA gene 3'UTR/terminator at the 3' region, or downstream, of the protein-coding sequence. The transformation construct may additionally comprise homology regions to the *Phaeodactylum tricornutum* genome for targeted genomic integration of the transformation vector. Homology regions may be selected to disrupt one or more genomic sites of endogenous lipid biosynthesis pathway genes. One skilled in the art can identify such homology regions within the sequence of the *Phaeodactylum tricornutum* genome

(referenced in the publication by Bowler et al., *Nature*, Vol. 456 (2008), pp. 239-244). Stable transformation of *Phaeodactylum tricornutum* with the transformation vector is achieved through well-known transformation techniques including microprojectile bombardment or other known methods. Activity of the *ble* gene product may be used as a marker to select for *Phaeodactylum tricornutum* transformed with the transformation vector in, but not limited to, LDM medium comprising paromomycin. Growth medium suitable for *Phaeodactylum tricornutum* lipid production include, but are not limited to f/2 medium as reported by Radokovits *et al.* Evaluation of fatty acid profiles of *Phaeodactylum tricornutum* lipids may be assessed through standard lipid extraction and analytical methods described herein.

EXAMPLE 22: ENGINEERING *CHAETOCEROS SP.*

[0322] Expression of recombinant genes in accordance with the present invention in *Chaetoceros sp.* may be accomplished by modifying the methods and vectors taught by Yamaguchi *et al.* as discussed herein. Briefly, Yamaguchi *et al.*, Phycological Research, Vol. 59:2 (2011), pp.113-119, reported the stable nuclear transformation of *Chaetoceros sp.* with plasmid DNA. Using the transformation method of microprojectile bombardment, Yamaguchi introduced the plasmid pTpfcf/nat into *Chaetoceros sp.* pTpfcf/nat comprised a nourseothricin resistance cassette, comprising sequence encoding the nourseothricin acetyltransferase (*nat*) gene product (GenBank Accession No. AAC60439) operably linked to the *Thalassiosira pseudonana* fucoxanthin chlorophyll a/c binding protein gene (*fcf*) promoter upstream of the *nat* protein-coding region and operably linked to the *Thalassiosira pseudonana* *fcf* gene 3' UTR/ terminator at the 3' region (downstream of the *nat* protein coding-sequence). The *nat* gene product confers resistance to the antibiotic nourseothricin. Prior to transformation with pTpfcf/nat, *Chaetoceros sp.* was unable to propagate on medium comprising 500 ug/ml nourseothricin. Upon transformation with pTpfcf/nat, transformants of *Chaetoceros sp.* were obtained that were propagated in selective culture medium comprising 500 ug/ml nourseothricin. The expression of the *nat* gene product in *Chaetoceros sp.* enabled propagation in the presence of 500 ug/ml nourseothricin, thereby establishing the utility of the nourseothricin antibiotic resistance cassette as selectable marker for use in *Chaetoceros sp.* Evaluation of the genomic DNA of the stable transformants was performed by Southern analysis. Yamaguchi reported that selection and maintenance of the transformed *Chaetoceros sp.* was performed on agar plates comprising f/2 medium (as

reported by Guilard, R.R., Culture of Phytoplankton for feeding marine invertebrates, *In* Culture of Marine Invertebrate Animals, Smith and Chanley (eds) 1975, Plenum Press, New York, pp. 26-60) with 500 ug/ml nourseothricin. Liquid propagation of *Chaetoceros* sp. transformants, as performed by Yamaguchi, was carried out in f/2 medium with 500 mg/L nourseothricin. Propagation of *Chaetoceros* sp. in additional culture medium has been reported (for example in Napolitano *et al.*, *Journal of the World Aquaculture Society*, Vol. 21:2 (1990), pp. 122-130, and by Volkman *et al.*, *Journal of Experimental Marine Biology and Ecology*, Vol. 128:3 (1989), pp. 219-240). Additional plasmids, promoters, 3'UTR/terminators, and selectable markers suitable for enabling heterologous gene expression in *Chaetoceros* sp. have been reported in the same report by Yamaguchi *et al.* Yamaguchi reported that the plasmid pTpfcf/nat, and the *Thalassiosira pseudonana* fcp promoter and 3' UTR/terminator are suitable to enable exogenous gene expression in *Chaetoceros* sp. In addition, Yamaguchi reported that the nourseothricin resistance cassette encoded on pTpfcf/nat was suitable for use as a selectable marker in *Chaetoceros* sp.

[0323] In an embodiment of the present invention, vector pTpfcf/nat, comprising the nucleotide sequence encoding the *nat* gene product for use as a selectable marker, is constructed and modified to further comprise a lipid biosynthesis pathway expression cassette sequence, thereby creating a transformation vector. The lipid biosynthesis pathway expression cassette encodes one or more lipid biosynthesis pathway proteins selected from Table 25, each protein-coding sequence codon-optimized for expression in the closely-related *Chaetoceros compressum* to reflect the codon bias inherent in nuclear genes of *Chaetoceros compressum* in accordance with Tables 24A-D. For each lipid biosynthesis pathway protein of Table 25, the codon-optimized gene sequence can individually be operably linked to the *Thalassiosira pseudonana* fcp gene promoter upstream of the protein-coding sequence and operably linked to the *Thalassiosira pseudonana* fcp gene 3'UTR/terminator at the 3' region, or downstream, of the protein-coding sequence. The transformation construct may additionally comprise homology regions to the *Chaetoceros* sp. genome for targeted genomic integration of the transformation vector. Homology regions may be selected to disrupt one or more genomic sites of endogenous lipid biosynthesis pathway genes. Stable transformation of *Chaetoceros* sp. with the transformation vector is achieved through well-known transformation including microprojectile bombardment or other known methods. Activity of the *nat* gene product may be used as a selectable marker to select for *Chaetoceros* sp.

transformed with the transformation vector in, but not limited to, f/2 agar medium comprising nourseothricin. Growth medium suitable for *Chaetoceros sp.* lipid production include, but are not limited to, f/2 medium, and those culture media discussed by Napolitano *et al.* and Volkman *et al.* Evaluation of fatty acid profiles of *Chaetoceros sp* lipids may be assessed through standard lipid extraction and analytical methods described herein.

EXAMPLE 23: ENGINEERING *CYLINDROTHECA FUSIFORMIS*

[0324] Expression of recombinant genes in accordance with the present invention in *Cylindrotheca fusiformis* may be accomplished by modifying the methods and vectors taught by Poulsen and Kroger *et al.* as discussed herein. Briefly, Poulsen and Kroger *et al.*, *FEBS Journal*, Vol. 272 (2005), pp.3413-3423, reported the transformation of *Cylindrotheca fusiformis* with plasmid DNA. Using the transformation method of microprojectile bombardment, Poulsen and Kroger introduced the pCF-ble plasmid into *Cylindrotheca fusiformis*. Plasmid pCF-ble comprised a bleomycin resistance cassette, comprising sequence encoding the *Streptoalloteichus hindustanus* Bleomycin binding protein (*ble*), for resistance to the antibiotics zeocin and phleomycin, operably linked to the *Cylindrotheca fusiformis* fucozanthin chlorophyll a/c binding protein gene (*fcpA*, GenBank Accession No. AY125580) promoter upstream of the *ble* protein-coding region and operably linked to the *Cylindrotheca fusiformis* *fcpA* gene 3'UTR/terminator at the 3' region (down-stream of the *ble* protein-coding region). Prior to transformation with pCF-ble, *Cylindrotheca fusiformis* was unable to propagate on medium comprising 1 mg/ml zeocin. Upon transformation with pCF-ble, transformants of *Cylindrotheca fusiformis* were obtained that were propagated in selective culture medium comprising 1 mg/ml zeocin. The expression of the *ble* gene product in *Cylindrotheca fusiformis* enabled propagation in the presence of 1 mg/ml zeocin, thereby establishing the utility of the bleomycin antibiotic resistance cassette as selectable marker for use in *Cylindrotheca fusiformis*. Poulsen and Kroger reported that selection and maintenance of the transformed *Cylindrotheca fusiformis* was performed on agar plates comprising artificial seawater medium with 1 mg/ml zeocin. Poulsen and Kroger reported liquid propagation of *Cylindrotheca fusiformis* transformants in artificial seawater medium with 1 mg/ml zeocin. Propagation of *Cylindrotheca fusiformis* in additional culture medium has been discussed (for example in Liang *et al.*, *Journal of Applied Phycology*, Vol. 17:1 (2005), pp. 61-65, and by Orcutt and Patterson, *Lipids*, Vol. 9:12 (1974), pp. 1000-1003).

Additional plasmids, promoters, and 3'UTR/terminators for enabling heterologous gene expression in *Chaetoceros sp.* have been reported in the same report by Poulsen and Kroger. Poulsen and Kroger reported that the plasmid pCF-ble and the *Cylindrotheca fusiformis* fcp promoter and 3' UTR/terminator are suitable to enable exogenous gene expression in *Cylindrotheca fusiformis*. In addition, Poulsen and Kroger reported that the bleomycin resistance cassette encoded on pCF-ble was suitable for use as a selectable marker in *Cylindrotheca fusiformis*.

[0325] In an embodiment of the present invention, vector pCF-ble, comprising the nucleotide sequence encoding the *ble* gene product for use as a selectable marker, is constructed and modified to further comprise a lipid biosynthesis pathway expression cassette sequence, thereby creating a transformation vector. The lipid biosynthesis pathway expression cassette encodes one or more lipid biosynthesis pathway proteins selected Table 25, each protein-coding sequence codon-optimized for expression in *Cylindrotheca fusiformis* to reflect the codon bias inherent in nuclear genes of *Cylindrotheca fusiformis* in accordance with Tables 24A-D. For each lipid biosynthesis pathway protein of Table 25, the codon-optimized gene sequence can individually be operably linked to the *Cylindrotheca fusiformis* fcp gene promoter upstream of the protein-coding sequence and operably linked to the *Cylindrotheca fusiformis* fcp gene 3'UTR/terminator at the 3' region, or downstream, of the protein-coding sequence. The transformation construct may additionally comprise homology regions to the *Cylindrotheca fusiformis* genome for targeted genomic integration of the transformation vector. Homology regions may be selected to disrupt one or more genomic sites of endogenous lipid biosynthesis pathway genes. Stable transformation of *Cylindrotheca fusiformis* with the transformation vector is achieved through well-known transformation techniques including microprojectile bombardment or other known methods. Activity of the *ble* gene product may be used as a selectable marker to select for *Cylindrotheca fusiformis* transformed with the transformation vector in, but not limited to, artificial seawater agar medium comprising zeocin. Growth media suitable for *Cylindrotheca fusiformis* lipid production include, but are not limited to, artificial seawater and those media reported by Liang et al. and Orcutt and Patterson. Evaluation of fatty acid profiles of *Cylindrotheca fusiformis* lipids may be assessed through standard lipid extraction and analytical methods described herein.

EXAMPLE 24: ENGINEERING *AMPHIDINIUM SP.*

[0326] Expression of recombinant genes in accordance with the present invention in *Amphidinium sp.* may be accomplished by modifying the methods and vectors taught by ten Lohuis and Miller *et al.* as discussed herein. Briefly, ten Lohuis and Miller *et al.*, *The Plant Journal*, Vol. 13:3 (1998), pp. 427-435, reported the stable transformation of *Amphidinium sp.* with plasmid DNA. Using the transformation technique of agitation in the presence of silicon carbide whiskers, ten Lohuis introduced the plasmid pMT NPT/GUS into *Amphidinium sp.* pMT NPT/GUS comprised a neomycin resistance cassette, comprising sequence encoding the neomycin phosphotransferase II (*nptII*) gene product (GenBank Accession No. AAL92039) operably linked to the *Agrobacterium tumefaciens* nopaline synthase (*nos*) gene promoter upstream, or 5' of the *nptII* protein-coding region and operably linked to the 3' UTR/terminator of the *nos* gene at the 3' region (down-stream of the *nptII* protein-coding region). The *nptII* gene product confers resistance to the antibiotic G418. The pMT NPT/GUS plasmid further comprised sequence encoding a beta-glucuronidase (GUS) reporter gene product operably-linked to a CaMV 35S promoter and further operably linked to the CaMV 35S 3' UTR/terminator. Prior to transformation with pMT NPT/GUS, *Amphidinium sp.* was unable to be propagated on medium comprising 3 mg/ml G418. Upon transformation with pMT NPT/GUS, transformants of *Amphidinium sp.* were obtained that were propagated in selective culture medium comprising 3 mg/ml G418. The expression of the *nptII* gene product in *Amphidinium sp.* enabled propagation in the presence of 3 mg/ml G418, thereby establishing the utility of the neomycin antibiotic resistance cassette as selectable marker for use in *Amphidinium sp.* Detectable activity of the GUS reporter gene indicated that CaMV 35S promoter and 3'UTR are suitable for enabling gene expression in *Amphidinium sp.* Evaluation of the genomic DNA of the stable transformants was performed by Southern analysis. ten Lohuis and Miller reported liquid propagation of *Amphidinium sp.* transformants in medium comprising seawater supplemented with F/2 enrichment solution (provided by the supplier Sigma) and 3 mg/ml G418 as well as selection and maintenance of *Amphidinium sp.* transformants on agar medium comprising seawater supplemented with F/2 enrichment solution and 3 mg/ml G418. Propagation of *Amphidinium sp.* in additional culture medium has been reported (for example in Mansour *et al.*, *Journal of Applied Phycology*, Vol. 17:4 (2005) pp. 287-v300). An additional plasmid, comprising additional promoters, 3'UTR/terminators, and a selectable marker for

enabling heterologous gene expression in *Amphidinium sp.* have been reported in the same report by ten Lohuis and Miller. ten Lohuis and Miller reported that the plasmid pMT NPT/GUS and the promoter and 3' UTR/terminator of the *nos* and CaMV 35S genes are suitable to enable exogenous gene expression in *Amphidinium sp.* In addition, ten Lohuis and Miller reported that the neomycin resistance cassette encoded on pMT NPT/GUS was suitable for use as a selectable marker in *Amphidinium sp.*

[0327] In an embodiment of the present invention, vector pMT NPT/GUS, comprising the nucleotide sequence encoding the *nptII* gene product for use as a selectable marker, is constructed and modified to further comprise a lipid biosynthesis pathway expression cassette sequence, thereby creating a transformation vector. The lipid biosynthesis pathway expression cassette encodes one or more lipid biosynthesis pathway proteins selected from Table 25, each protein-coding sequence codon-optimized for expression in *Amphidinium sp.* to reflect the codon bias inherent in nuclear genes of the closely-related species, *Amphidinium carterae* in accordance with Tables 24A-D. For each lipid biosynthesis pathway protein of Table 25, the codon-optimized gene sequence can individually be operably linked to the *Agrobacterium tumefaciens* nopaline synthase (*nos*) gene promoter upstream of the protein-coding sequence and operably linked to the *nos* 3'UTR/terminator at the 3' region, or downstream, of the protein-coding sequence. The transformation construct may additionally comprise homology regions to the *Amphidinium sp.* genome for targeted genomic integration of the transformation vector. Homology regions may be selected to disrupt one or more genomic sites of endogenous lipid biosynthesis pathway genes. Stable transformation of *Amphidinium sp.* with the transformation vector is achieved through well-known transformation techniques including silicon fibre-mediated microinjection or other known methods. Activity of the *nptII* gene product may be used as a selectable marker to select for *Amphidinium sp.* transformed with the transformation vector in, but not limited to, seawater agar medium comprising G418. Growth media suitable for *Amphidinium sp.* lipid production include, but are not limited to, artificial seawater and those media reported by Mansour *et al.* and ten Lohuis and Miller. Evaluation of fatty acid profiles of *Amphidinium sp.* lipids may be assessed through standard lipid extraction and analytical methods described herein.

EXAMPLE 25: ENGINEERING *SYMBIODINIUM MICROADRIACTICUM*

[0328] Expression of recombinant genes in accordance with the present invention in *Symbiodinium microadriacticum* may be accomplished by modifying the methods and

vectors taught by ten Lohuis and Miller *et al.* as discussed herein. Briefly, ten Lohuis and Miller *et al.*, *The Plant Journal*, Vol. 13:3 (1998), pp. 427-435, reported the stable transformation of *Symbiodinium microadriaticum* with plasmid DNA. Using the transformation technique of silicon fibre-mediated microinjection, ten Lohuis introduced

5 the plasmid pMT NPT/GUS into *Symbiodinium microadriaticum*. pMT NPT/GUS comprised a neomycin resistance cassette, comprising sequence encoding the neomycin phosphotransferase II (*nptII*) gene product (GenBank Accession No. AAL92039) operably linked to the *Agrobacterium tumefaciens* nopaline synthase (*nos*) gene promoter upstream, or 5' of the *nptII* protein-coding region and operably linked to the 3' UTR/terminator of

10 the *nos* gene at the 3' region (down-stream of the *nptII* protein-coding region). The *nptII* gene product confers resistance to the antibiotic G418. The pMT NPT/GUS plasmid further comprised sequence encoding a beta-glucuronidase (GUS) reporter gene product operably-linked to a CaMV 35S promoter and further operably linked to the CaMV 35S 3' UTR/terminator. Prior to transformation with pMT NPT/GUS, *Symbiodinium*

15 *microadriaticum* was unable to be propagated on medium comprising 3 mg/ml G418. Upon transformation with pMT NPT/GUS, transformants of *Symbiodinium microadriaticum* were obtained that were propagated in selective culture medium comprising 3 mg/ml G418. The expression of the *nptII* gene product in *Symbiodinium microadriaticum* enabled propagation in the presence of 3 mg/ml G418, thereby

20 establishing the utility of the neomycin antibiotic resistance cassette as selectable marker for use in *Symbiodinium microadriaticum*. Detectable activity of the GUS reporter gene indicated that CaMV 35S promoter and 3'UTR are suitable for enabling gene expression in *Symbiodinium microadriaticum*. Evaluation of the genomic DNA of the stable transformants was performed by Southern analysis. ten Lohuis and Miller reported liquid

25 propagation of *Symbiodinium microadriaticum* transformants in medium comprising seawater supplemented with F/2 enrichment solution (provided by the supplier Sigma) and 3 mg/ml G418 as well as selection and maintenance of *Symbiodinium microadriaticum* transformants on agar medium comprising seawater supplemented with F/2 enrichment solution and 3 mg/ml G418. Propagation of *Symbiodinium microadriaticum* in additional

30 culture medium has been discussed (for example in Iglesias-Prieto *et al.*, *Proceedings of the National Academy of Sciences*, Vol. 89:21 (1992) pp. 10302-10305). An additional plasmid, comprising additional promoters, 3'UTR/terminators, and a selectable marker for enabling heterologous gene expression in *Symbiodinium microadriaticum* have been discussed in the same report by ten Lohuis and Miller. ten Lohuis and Miller reported that

the plasmid pMT NPT/GUS and the promoter and 3' UTR/terminator of the *nos* and CaMV 35S genes are suitable to enable exogenous gene expression in *Symbiodinium microadriaticum*. In addition, ten Lohuis and Miller reported that the neomycin resistance cassette encoded on pMT NPT/GUS was suitable for use as a selectable marker in *Symbiodinium microadriaticum*.

[0329] In an embodiment of the present invention, vector pMT NPT/GUS, comprising the nucleotide sequence encoding the *nptII* gene product for use as a selectable marker, is constructed and modified to further comprise a lipid biosynthesis pathway expression cassette sequence, thereby creating a transformation vector. The lipid biosynthesis pathway expression cassette encodes one or more lipid biosynthesis pathway proteins selected Table 25, each protein-coding sequence codon-optimized for expression in *Symbiodinium microadriaticum* to reflect the codon bias inherent in nuclear genes of *Symbiodinium microadriaticum* in accordance with Tables 24A-D. For each lipid biosynthesis pathway protein of Table 25, the codon-optimized gene sequence can individually be operably linked to the *Agrobacterium tumefaciens* nopaline synthase (*nos*) gene promoter upstream of the protein-coding sequence and operably linked to the *nos* 3'UTR/terminator at the 3' region, or downstream, of the protein-coding sequence. The transformation construct may additionally comprise homology regions to the *Symbiodinium microadriaticum* genome for targeted genomic integration of the transformation vector. Homology regions may be selected to disrupt one or more genomic sites of endogenous lipid biosynthesis pathway genes. Stable transformation of *Symbiodinium microadriaticum* with the transformation vector is achieved through well-known transformation techniques including silicon fibre-mediated microinjection or other known methods. Activity of the *nptII* gene product may be used as a selectable marker to select for *Symbiodinium microadriaticum* transformed with the transformation vector in, but not limited to, seawater agar medium comprising G418. Growth media suitable for *Symbiodinium microadriaticum* lipid production include, but are not limited to, artificial seawater and those media reported by Iglesias-Prieto *et al.* and ten Lohuis and Miller. Evaluation of fatty acid profiles of *Symbiodinium microadriaticum* lipids may be assessed through standard lipid extraction and analytical methods described herein.

EXAMPLE 26: ENGINEERING *NANNOCHLOROPSIS SP.*

[0330] Expression of recombinant genes in accordance with the present invention in *Nannochloropsis sp.* W2J3B may be accomplished by modifying the methods and vectors

taught by Kilian *et al.* as discussed herein. Briefly, Kilian *et al.*, *Proceedings of the National Academy of Sciences*, Vol. 108:52 (2011) pp.21265-21269, reported the stable nuclear transformation of *Nannochloropsis* with a transformation construct. Using the transformation method of electroporation, Kilian introduced the transformation construct

5 C2 into *Nannochloropsis sp.* W2J3B. The C2 transformation construct comprised a bleomycin resistance cassette, comprising the coding sequence for the *Streptoalloteichus hindustanus* Bleomycin binding protein (*ble*), for resistance to the antibiotics phleomycin and zeocin, operably linked to and the promoter of the *Nannochloropsis sp.* W2J3B violaxanthin/chlorophyll a-binding protein gene VCP2 upstream of the *ble* protein-coding

10 region and operably linked to the 3'UTR/terminator of the *Nannochloropsis sp.* W2J3B violaxanthin/chlorophyll a-binding gene VCP1 downstream of the *ble* protein-coding region. Prior to transformation with C2, *Nannochloropsis sp.* W2J3B was unable to propagate on medium comprising 2 ug/ml zeocin. Upon transformation with C2, transformants of *Nannochloropsis sp.* W2J3B were obtained that were propagated in

15 selective culture medium comprising 2 ug/ml zeocin. The expression of the *ble* gene product in *Nannochloropsis sp.* W2J3B enabled propagation in the presence of 2 ug/ml zeocin, thereby establishing the utility of the bleomycin antibiotic resistance cassette as selectable marker for use in *Nannochloropsis*. Evaluation of the genomic DNA of the stable transformants was performed by PCR. Kilian reported liquid propagation of

20 *Nannochloropsis sp.* W2J3B transformants in F/2 medium (reported by Guilard and Ryther, *Canadian Journal of Microbiology*, Vol. 8 (1962), pp. 229-239) comprising fivefold levels of trace metals, vitamins, and phosphate solution, and further comprising 2 ug/ml zeocin. Kilian also reported selection and maintenance of *Nannochloropsis sp.* W2J3B transformants on agar F/2 medium comprising artificial seawater 2 mg/ml zeocin.

25 Propagation of *Nannochloropsis* in additional culture medium has been discussed (for example in Chiu *et al.*, *Bioresour Technol.*, Vol. 100:2 (2009), pp. 833-838 and Pal *et al.*, *Applied Microbiology and Biotechnology*, Vol. 90:4 (2011), pp. 1429-1441.). Additional transformation constructs, comprising additional promoters and 3'UTR/terminators for enabling heterologous gene expression in *Nannochloropsis sp.* W2J3B and selectable

30 markers for selection of transformants have been described in the same report by Kilian. Kilian reported that the transformation construct C2 and the promoter of the *Nannochloropsis sp.* W2J3B violaxanthin/chlorophyll a-binding protein gene VCP2 and 3' UTR/terminator of the *Nannochloropsis sp.* W2J3B violaxanthin/chlorophyll a-binding protein gene VCP1 are suitable to enable exogenous gene expression in *Nannochloropsis*

sp. W2J3B. In addition, Kilian reported that the bleomycin resistance cassette encoded on C2 was suitable for use as a selectable marker in *Nannochloropsis sp.* W2J3B.

[0331] In an embodiment of the present invention, transformation construct C2, comprising the nucleotide sequence encoding the *ble* gene product for use as a selectable
5 marker, is constructed and modified to further comprise a lipid biosynthesis pathway expression cassette sequence, thereby creating a transformation vector. The lipid biosynthesis pathway expression cassette encodes one or more lipid biosynthesis pathway proteins selected from Table 25, each protein-coding sequence codon-optimized for expression in *Nannochloropsis sp.* W2J3B to reflect the codon bias inherent in nuclear
10 genes of *Nannochloropsis sp.* in accordance with Tables 24A-D. For each lipid biosynthesis pathway protein of Table 25, the codon-optimized gene sequence can individually be operably linked to the *Nannochloropsis sp.* W2J3B VCP2 gene promoter upstream of the protein-coding sequence and operably linked to the *Nannochloropsis sp.* W2J3B VCP1 gene 3'UTR/terminator at the 3' region, or downstream, of the protein-
15 coding sequence. The transformation construct may additionally comprise homology regions to the *Nannochloropsis sp.* W2J3B genome for targeted genomic integration of the transformation vector. Homology regions may be selected to disrupt one or more genomic sites of endogenous lipid biosynthesis pathway genes. Stable transformation of *Nannochloropsis sp.* W2J3B with the transformation vector is achieved through well-
20 known transformation techniques including electroporation or other known methods. Activity of the *ble* gene product may be used as a selectable marker to select for *Nannochloropsis sp.* W2J3B transformed with the transformation vector in, but not limited to, F/2 medium comprising zeocin. Growth media suitable for *Nannochloropsis sp.* W2J3B lipid production include, but are not limited to, F/2 medium and those media
25 reported by Chiu et al. and Pal et al. Evaluation of fatty acid profiles of *Nannochloropsis sp.* W2J3B lipids may be assessed through standard lipid extraction and analytical methods described herein.

EXAMPLE 27: ENGINEERING *CYCLOTELLA CRYPTICA*

30 [0332] Expression of recombinant genes in accordance with the present invention in *Cyclotella cryptica* may be accomplished by modifying the methods and vectors taught by Dunahay et al. as discussed herein. Briefly, Dunahay et al., *Journal of Phycology*, Vol. 31 (1995), pp. 1004-1012, reported the stable transformation of *Cyclotella cryptica* with plasmid DNA. Using the transformation method of microprojectile bombardment,

Dunahay introduced the plasmid pACCNPT5.1 into *Cyclotella cryptica*. Plasmid pACCNPT5.1 comprised a neomycin resistance cassette, comprising the coding sequence of the neomycin phosphotransferase II (*nptII*) gene product operably linked to the promoter of the *Cyclotella cryptica* acetyl-CoA carboxylase (ACCase) gene (GenBank
5 Accession No. L20784) upstream of the *nptII* coding-region and operably linked to the 3'UTR/terminator of the *Cyclotella cryptica* ACCase gene at the 3' region (downstream of the *nptII* coding-region). The *nptII* gene product confers resistance to the antibiotic G418. Prior to transformation with pACCNPT5.1, *Cyclotella cryptica* was unable to propagate on 50% artificial seawater medium comprising 100 ug/ml G418. Upon transformation
10 with pACCNPT5.1, transformants of *Cyclotella cryptica* were obtained that were propagated in selective 50% artificial seawater medium comprising 100 ug/ml G418. The expression of the *nptII* gene product in *Cyclotella cryptica* enabled propagation in the presence of 100 ug/ml G418, thereby establishing the utility of the neomycin antibiotic resistance cassette as selectable marker for use in *Cyclotella cryptica*. Evaluation of the
15 genomic DNA of the stable transformants was performed by Southern analysis. Dunahay reported liquid propagation of *Cyclotella cryptica* in artificial seawater medium (ASW, as discussed by Brown, L., *Phycologia*, Vol. 21 (1982), pp. 408-410) supplemented with 1.07 mM sodium silicate and with 100 ug/ml G418. Dunahay also reported selection and maintenance of *Cyclotella cryptica* transformants on agar plates comprising ASW medium
20 with 100 ug/ml G418. Propagation of *Cyclotella cryptica* in additional culture medium has been discussed (for example in Sriharan *et al.*, *Applied Biochemistry and Biotechnology*, Vol. 28-29:1 (1991), pp. 317-326 and Pahl *et al.*, *Journal of Bioscience and Bioengineering*, Vol. 109:3 (2010), pp. 235-239). Dunahay reported that the plasmid pACCNPT5.1 and the promoter of the *Cyclotella cryptica* acetyl-CoA carboxylase
25 (ACCase) gene are suitable to enable exogenous gene expression in *Cyclotella cryptica*. In addition, Dunahay reported that the neomycin resistance cassette encoded on pACCNPT5.1 was suitable for use as a selectable marker in *Cyclotella cryptica*.

[0333] In an embodiment of the present invention, vector pACCNPT5.1, comprising the nucleotide sequence encoding the *nptII* gene product for use as a selectable marker, is
30 constructed and modified to further comprise a lipid biosynthesis pathway expression cassette sequence, thereby creating a transformation vector. The lipid biosynthesis pathway expression cassette encodes one or more lipid biosynthesis pathway proteins selected from Table 25, each protein-coding sequence codon-optimized for expression in *Cyclotella cryptica* to reflect the codon bias inherent in nuclear genes of *Cyclotella*

cryptica in accordance with Tables 24A-D. For each lipid biosynthesis pathway protein of Table 25, the codon-optimized gene sequence can individually be operably linked to the *Cyclotella cryptica* ACCase promoter upstream of the protein-coding sequence and operably linked to the *Cyclotella cryptica* ACCase 3'UTR/terminator at the 3' region, or

5 downstream, of the protein-coding sequence. The transformation construct may additionally comprise homology regions to the *Cyclotella cryptica* genome for targeted genomic integration of the transformation vector. Homology regions may be selected to disrupt one or more genomic sites of endogenous lipid biosynthesis pathway genes. Stable transformation of *Cyclotella cryptica* with the transformation vector is achieved through

10 well-known transformation techniques including microprojectile bombardment or other known methods. Activity of the *nptII* gene product may be used as a marker to select for for *Cyclotella cryptica* transformed with the transformation vector in, but not limited to, agar ASW medium comprising G418. Growth media suitable for *Cyclotella cryptica* lipid production include, but are not limited to, ASW medium and those media reported by

15 Sriharan *et al.*, 1991 and Pahl *et al.* Evaluation of fatty acid profiles of *Cyclotella cryptica* lipids may be assessed through standard lipid extraction and analytical methods described herein.

EXAMPLE 28: ENGINEERING *NAVICULA SAPROPHILA*

20 [0334] Expression of recombinant genes in accordance with the present invention in *Navicula saprophila* may be accomplished by modifying the methods and vectors taught by Dunahay *et al.* as discussed herein. Briefly, Dunahay *et al.*, *Journal of Phycology*, Vol. 31 (1995), pp. 1004-1012, reported the stable transformation of *Navicula saprophila* with plasmid DNA. Using the transformation method of microprojectile bombardment,

25 Dunahay introduced the plasmid pACCNPT5.1 into *Navicula saprophila*. Plasmid pACCNPT5.1 comprised a neomycin resistance cassette, comprising the coding sequence of the neomycin phosphotransferase II (*nptII*) gene product operably linked to the promoter of the *Cyclotella cryptica* acetyl-CoA carboxylase (ACCase) gene (GenBank Accession No. L20784) upstream of the *nptII* coding-region and operably linked to the

30 3'UTR/terminator of the *Cyclotella cryptica* ACCase gene at the 3' region (downstream of the *nptII* coding-region). The *nptII* gene product confers resistance to the antibiotic G418. Prior to transformation with pACCNPT5.1, *Navicula saprophila* was unable to propagate on artificial seawater medium comprising 100 ug/ml G418. Upon transformation with pACCNPT5.1, transformants of *Navicula saprophila* were obtained that were propagated

in selective artificial seawater medium comprising 100 ug/ml G418. The expression of the *nptII* gene product in *Navicula saprophila* enabled propagation in the presence of G418, thereby establishing the utility of the neomycin antibiotic resistance cassette as selectable marker for use in *Navicula saprophila*. Evaluation of the genomic DNA of the stable

5 transformants was performed by Southern analysis. Dunahay reported liquid propagation of *Navicula saprophila* in artificial seawater medium (ASW, as discussed by Brown, L., *Phycologia*, Vol. 21 (1982), pp. 408-410) supplemented with 1.07 mM sodium silicate and with 100 ug/ml G418. Dunahay also reported selection and maintenance of *Navicula saprophila* transformants on agar plates comprising ASW medium with 100 ug/ml G418.

10 Propagation of *Navicula saprophila* in additional culture medium has been discussed (for example in Tadros and Johansen, *Journal of Phycology*, Vol. 24:4 (1988), pp. 445-452 and Sriharan *et al.*, *Applied Biochemistry and Biotechnology*, Vol. 20-21:1 (1989), pp. 281-291). Dunahay reported that the plasmid pACCNPT5.1 and the promoter of the *Cyclotella cryptica* acetyl-CoA carboxylase (ACCase) gene are suitable to enable exogenous gene

15 expression in *Navicula saprophila*. In addition, Dunahay reported that the neomycin resistance cassette encoded on pACCNPT5.1 was suitable for use as a selectable marker in *Navicula saprophila*.

[0335] In an embodiment of the present invention, vector pACCNPT5.1, comprising the nucleotide sequence encoding the *nptII* gene product for use as a selectable marker, is

20 constructed and modified to further comprise a lipid biosynthesis pathway expression cassette sequence, thereby creating a transformation vector. The lipid biosynthesis pathway expression cassette encodes one or more lipid biosynthesis pathway proteins selected from Table 25, each protein-coding sequence codon-optimized for expression in *Navicula saprophila* to reflect the codon bias inherent in nuclear genes of the closely-

25 related *Navicula pelliculosa* in accordance with Tables 24A-D. For each lipid biosynthesis pathway protein of Table 25, the codon-optimized gene sequence can individually be operably linked to the *Cyclotella cryptica* ACCase gene promoter upstream of the protein-coding sequence and operably linked to the *Cyclotella cryptica* ACCase gene 3'UTR/terminator at the 3' region, or downstream, of the protein-coding

30 sequence. The transformation construct may additionally comprise homology regions to the *Navicula saprophila* genome for targeted genomic integration of the transformation vector. Homology regions may be selected to disrupt one or more genomic sites of endogenous lipid biosynthesis pathway genes. Stable transformation of *Navicula saprophila* with the transformation vector is achieved through well-known transformation

techniques including microprojectile bombardment or other known methods. Activity of the *nptII* gene product may be used as a selectable marker to select for *Navicula saprophila* transformed with the transformation vector in, but not limited to, agar ASW medium comprising G418. Growth media suitable for *Navicula saprophila* lipid production include, but are not limited to, ASW medium and those media reported by Sriharan *et al.* 1989 and Tadros and Johansen. Evaluation of fatty acid profiles of *Navicula saprophila* lipids may be assessed through standard lipid extraction and analytical methods described herein.

10 **EXAMPLE 29: ENGINEERING *THALASSIOSIRA PSEUDONANA***

[0336] Expression of recombinant genes in accordance with the present invention in *Thalassiosira pseudonana* may be accomplished by modifying the methods and vectors taught by Poulsen *et al.* as discussed herein. Briefly, Poulsen *et al.*, *Journal of Phycology*, Vol. 42 (2006), pp. 1059-1065, reported the stable transformation of *Thalassiosira pseudonana* with plasmid DNA. Using the transformation method of microprojectile bombardment, Poulsen introduced the plasmid pTpfcf/nat in to *Thalassiosira pseudonana*. pTpfcf/nat comprised a nourseothricin resistance cassette, comprising sequence encoding the nourseothricin acetyltransferase (*nat*) gene product (GenBank Accession No. AAC60439) operably linked to the *Thalassiosira pseudonana* fucoxanthin chlorophyll *a/c* binding protein gene (*fcf*) promoter upstream of the *nat* protein-coding region and operably linked to the *Thalassiosira pseudonana* *fcf* gene 3' UTR/ terminator at the 3' region (downstream of the *nat* protein coding-sequence). The *nat* gene product confers resistance to the antibiotic nourseothricin. Prior to transformation with pTpfcf/nat, *Thalassiosira pseudonana* was unable to propagate on medium comprising 10 ug/ml nourseothricin. Upon transformation with pTpfcf/nat, transformants of *Thalassiosira pseudonana* were obtained that were propagated in selective culture medium comprising 100 ug/ml nourseothricin. The expression of the *nat* gene product in *Thalassiosira pseudonana* enabled propagation in the presence of 100 ug/ml nourseothricin, thereby establishing the utility of the nourseothricin antibiotic resistance cassette as selectable marker for use in *Thalassiosira pseudonana*. Evaluation of the genomic DNA of the stable transformants was performed by Southern analysis. Poulsen reported that selection and maintenance of the transformed *Thalassiosira pseudonana* was performed in liquid culture comprising modified ESAW medium (as discussed by Harrison *et al.*, *Journal of Phycology*, Vol. 16 (1980), pp. 28-35) with 100 ug/ml nourseothricin. Propagation of

Thalassiosira pseudonana in additional culture medium has been discussed (for example in Volkman *et al.*, *Journal of Experimental Marine Biology and Ecology*, Vol. 128:3 (1989), pp. 219-240). An additional plasmid, comprising additional selectable markers suitable for use in *Thalassiosira pseudonana* has been discussed in the same report by Poulsen. Poulsen reported that the plasmid pTpfcf/nat, and the *Thalassiosira pseudonana* fcf promoter and 3' UTR/terminator are suitable to enable exogenous gene expression in *Thalassiosira pseudonana*. In addition, Poulsen reported that the nourseothricin resistance cassette encoded on pTpfcf/nat was suitable for use as a selectable marker in *Thalassiosira pseudonana*.

[0337] In an embodiment of the present invention, vector pTpfcf/nat, comprising the nucleotide sequence encoding the *nat* gene product for use as a selectable marker, is constructed and modified to further comprise a lipid biosynthesis pathway expression cassette sequence, thereby creating a transformation vector. The lipid biosynthesis pathway expression cassette encodes one or more lipid biosynthesis pathway proteins selected from Table 25, each protein-coding sequence codon-optimized for expression in *Thalassiosira pseudonana* to reflect the codon bias inherent in nuclear genes of *Thalassiosira pseudonana* in accordance with Tables 24A-D. For each lipid biosynthesis pathway protein of Table 25, the codon-optimized gene sequence can individually be operably linked to the *Thalassiosira pseudonana* fcf gene promoter upstream of the protein-coding sequence and operably linked to the *Thalassiosira pseudonana* fcf gene 3'UTR/terminator at the 3' region, or downstream, of the protein-coding sequence. The transformation construct may additionally comprise homology regions to the *Thalassiosira pseudonana* genome for targeted genomic integration of the transformation vector. Homology regions may be selected to disrupt one or more genomic sites of endogenous lipid biosynthesis pathway genes. One skilled in the art can identify such homology regions within the sequence of the *Thalassiosira pseudonana* genome (referenced in the publication by Armbrust *et al.*, *Science*, Vol. 306: 5693 (2004): pp. 79-86). Stable transformation of *Thalassiosira pseudonana* with the transformation vector is achieved through well-known transformation techniques including microprojectile bombardment or other known methods. Activity of the *nat* gene product may be used as a marker to select for *Thalassiosira pseudonana* transformed with the transformation vector in but not limited to, ESAW agar medium comprising nourseothricin. Growth media suitable for *Thalassiosira pseudonana* lipid production include, but are not limited to, ESAW medium, and those culture media discussed by Volkman *et al.* and Harrison *et al.* Evaluation of

fatty acid profiles of *Thalassiosira pseudonana* lipids may be assessed through standard lipid extraction and analytical methods described herein.

EXAMPLE 30: ENGINEERING *CHLAMYDOMONAS REINHARDTII*

- 5 [0338] Expression of recombinant genes in accordance with the present invention in *Chlamydomonas reinhardtii* may be accomplished by modifying the methods and vectors taught by Cerutti *et al.* as discussed herein. Briefly, Cerutti *et al.*, *Genetics*, Vol. 145:1 (1997), pp. 97–110, reported the stable nuclear transformation of *Chlamydomonas reinhardtii* with a transformation vector. Using the transformation method of
- 10 microprojectile bombardment, Cerutti introduced transformation construct P[1030] into *Chlamydomonas reinhardtii*. Construct P[1030] comprised a spectinomycin resistance cassette, comprising sequence encoding the aminoglucoside 3''-adenyltransferase (*aadA*) gene product operably linked to the *Chlamydomonas reinhardtii* ribulose-1,5-bisphosphate carboxylase/oxygenase small subunit gene (*RbcS2*, GenBank Accession No. X04472)
- 15 promoter upstream of the *aadA* protein-coding region and operably linked to the *Chlamydomonas reinhardtii* *RbcS2* gene 3' UTR/ terminator at the 3' region (downstream of the *aadA* protein coding-sequence). The *aadA* gene product confers resistance to the antibiotic spectinomycin. Prior to transformation with P[1030], *Chlamydomonas reinhardtii* was unable to propagate on medium comprising 90 ug/ml spectinomycin.
- 20 Upon transformation with P[1030], transformants of *Chlamydomonas reinhardtii* were obtained that were propagated in selective culture medium comprising 90 ug/ml spectinomycin, thereby establishing the utility of the spectinomycin antibiotic resistance cassette as a selectable marker for use in *Chlamydomonas reinhardtii*. Evaluation of the genomic DNA of the stable transformants was performed by Southern analysis. Cerutti
- 25 reported that selection and maintenance of the transformed *Chlamydomonas reinhardtii* was performed on agar plates comprising Tris-acetate-phosphate medium (TAP, as described by Harris, *The Chlamydomonas Sourcebook*, Academic Press, San Diego, 1989) with 90 ug/ml spectinomycin. Cerutti additionally reported propagation of *Chlamydomonas reinhardtii* in TAP liquid culture with 90 ug/ml spectinomycin.
- 30 Propagation of *Chlamydomonas reinhardtii* in alternative culture medium has been discussed (for example in Dent *et al.*, *African Journal of Microbiology Research*, Vol. 5:3 (2011), pp. 260-270 and Yantao *et al.*, *Biotechnology and Bioengineering*, Vol. 107:2 (2010), pp. 258-268). Additional constructs, comprising additional selectable markers suitable for use in *Chlamydomonas reinhardtii* as well as numerous regulatory sequences,

including protomers and 3' UTRs suitable for promoting heterologous gene expression in *Chlamydomonas reinhardtii* are known in the art and have been discussed (for a review, see Radakovits *et al.*, *Eukaryotic Cell*, Vol. 9:4 (2010), pp. 486-501). Cerutti reported that the transformation vector P[1030] and the *Chlamydomonas reinhardtii* promoter and 3' UTR/terminator are suitable to enable exogenous gene expression in *Chlamydomonas reinhardtii*. In addition, Cerutti reported that the spectinomycin resistance cassette encoded on P[1030] was suitable for use as a selectable marker in *Chlamydomonas reinhardtii*.

[0339] In an embodiment of the present invention, vector P[1030], comprising the nucleotide sequence encoding the *aadA* gene product for use as a selectable marker, is constructed and modified to further comprise a lipid biosynthesis pathway expression cassette sequence, thereby creating a transformation vector. The lipid biosynthesis pathway expression cassette encodes one or more lipid biosynthesis pathway proteins selected from Table 25, each protein-coding sequence codon-optimized for expression in *Chlamydomonas reinhardtii* to reflect the codon bias inherent in nuclear genes of *Chlamydomonas reinhardtii* in accordance with Tables 24A-D. For each lipid biosynthesis pathway protein of Table 25, the codon-optimized gene sequence can individually be operably linked to the *Chlamydomonas reinhardtii* *RbcS2* promoter upstream of the protein-coding sequence and operably linked to the *Chlamydomonas reinhardtii* *RbcS2* 3'UTR/terminator at the 3' region, or downstream, of the protein-coding sequence. The transformation construct may additionally comprise homology regions to the *Chlamydomonas reinhardtii* genome for targeted genomic integration of the transformation vector. Homology regions may be selected to disrupt one or more genomic site of an endogenous lipid biosynthesis pathway gene. One skilled in the art can identify such homology regions within the sequence of the *Chlamydomonas reinhardtii* genome (referenced in the publication by Merchant *et al.*, *Science*, Vol. 318:5848 (2007), pp. 245-250). Stable transformation of *Chlamydomonas reinhardtii* with the transformation vector is achieved through well-known transformation techniques including microprojectile bombardment or other known methods. Activity of the *aadA* gene product may be used as a marker to select for *Chlamydomonas reinhardtii* transformed with the transformation vector on, but not limited to, TAP agar medium comprising spectinomycin. Growth media suitable for *Chlamydomonas reinhardtii* lipid production include, but are not limited to, ESAW medium, and those culture media discussed by Yantao *et al.* and Dent *et al.* Evaluation of fatty acid profiles of *Chlamydomonas reinhardtii* lipids may be assessed

through standard lipid extraction and analytical methods described herein.

EXAMPLE 31: ENGINEERING *YARROWIA LIPOLYTICA*

[0340] Expression of recombinant genes in accordance with the present invention in
5 *Yarrowia lipolytica* may be accomplished by modifying the methods and vectors taught by
Fickers et al. as discussed herein. Briefly, Fickers *et al.*, *Journal of Microbiological
Methods*, Vol. 55 (2003), pp. 727-737, reported the stable nuclear transformation of
Yarrowia lipolytica with plasmid DNA. Using a lithium acetate transformation method,
Fickers introduced the plasmid JMP123 into *Yarrowia lipolytica*. Plasmid JMP123
10 comprised a hygromycin B resistance cassette, comprising sequence encoding the
hygromycin B phosphotransferase gene product (*hph*), operably-linked to the *Yarrowia
lipolytica* LIP2 gene promoter (GenBank Accession No. AJ012632) upstream of the *hph*
protein-coding region and operably linked to the *Yarrowia lipolytica* LIP2 gene
3'UTR/terminator downstream of the *hph* protein-coding region. Prior to transformation
15 with JMP123, *Yarrowia lipolytica* were unable to propagate on medium comprising 100
ug/ml hygromycin. Upon transformation with JMP123, transformed *Yarrowia lipolytica*
were obtained that were able to propagate on medium comprising 100 ug/ml hygromycin,
thereby establishing the hygromycin B antibiotic resistance cassette as a selectable marker
for use in *Yarrowia lipolytica*. The nucleotide sequence provided on JMP123 of the
20 promoter and 3'UTR/terminator of the *Yarrowia lipolytica* LIP2 gene served as donor
sequences for homologous recombination of the *hph* coding sequence into the LIP2 locus.
Evaluation of the genomic DNA of the stable transformants was performed by Southern.
Fickers reported that selection and maintenance of the transformed *Yarrowia lipolytica*
was performed on agar plates comprising standard YPD medium (Yeast Extract Peptone
25 Dextrose) with 100 ug/ml hygromycin. Liquid culturing of transformed *Yarrowia
lipolytica* was performed in YPD medium with hygromycin. Other media and techniques
used for culturing *Yarrowia lipolytica* have been reported and numerous other plasmids,
promoters, 3' UTRs, and selectable markers for use in *Yarrowia lipolytica* have been
reported (for example see Pignede *et al.*, *Applied and Environmental Biology*, Vol. 66:8
30 (2000), pp. 3283-3289, Chuang *et al.*, *New Biotechnology*, Vol. 27:4 (2010), pp. 277-282,
and Barth and Gaillardin, (1996), In: K.W. (Ed.), *Nonconventional Yeasts in
Biotechnology*. Springer-Verlag, Berlin-Heidelberg, pp. 313-388). Fickers reported that the
transformation vector JMP123 and the *Yarrowia lipolytica* LIP2 gene promoter and 3'
UTR/terminator are suitable to enable heterologous gene expression in *Yarrowia*

lipolytica. In addition, Fickers reported that the hygromycin resistance cassette encoded on JMP123 was suitable for use as a selectable marker in *Yarrowia lipolytica*.

[0341] In an embodiment of the present invention, vector JMP123, comprising the nucleotide sequence encoding the *hph* gene product for use as a selectable marker, is constructed and modified to further comprise a lipid biosynthesis pathway expression cassette sequence, thereby creating a transformation vector. The lipid biosynthesis pathway expression cassette encodes one or more lipid biosynthesis pathway proteins selected from Table 25, each protein-coding sequence codon-optimized for expression in *Yarrowia lipolytica* to reflect the codon bias inherent in nuclear genes of *Yarrowia lipolytica* in accordance with Tables 24A-D. For each lipid biosynthesis pathway protein of Table 25, the codon-optimized gene sequence can individually be operably linked to the *Yarrowia lipolytica* *LIP2* gene promoter upstream of the protein-coding sequence and operably linked to the *Yarrowia lipolytica* *LIP2* gene 3'UTR/terminator at the 3' region, or downstream, of the protein-coding sequence. The transformation construct may additionally comprise homology regions to the *Yarrowia lipolytica* genome for targeted genomic integration of the transformation vector. Homology regions may be selected to disrupt one or more genomic sites of endogenous lipid biosynthesis pathway genes. One skilled in the art can identify such homology regions within the sequence of the *Yarrowia lipolytica* genome (referenced in the publication by Dujun *et al.*, *Nature*, Vol. 430 (2004), pp. 35-44). Stable transformation of *Yarrowia lipolytica* with the transformation vector is achieved through well-known transformation techniques including lithium acetate transformation or other known methods. Activity of the *hph* gene product may be used as a marker to select for *Yarrowia lipolytica* transformed with the transformation vector on, but not limited to, *YPD* medium comprising hygromycin. Growth media suitable for *Yarrowia lipolytica* lipid production include, but are not limited to, *YPD* medium, and those culture media described by Chuang *et al.* Evaluation of fatty acid profiles of *Yarrowia lipolytica* lipids may be assessed through standard lipid extraction and analytical methods described herein.

EXAMPLE 32: ENGINEERING *MORTIERELLA ALPINE*

[0342] Expression of recombinant genes in accordance with the present invention in *Mortierella alpine* may be accomplished by modifying the methods and vectors taught by Mackenzie *et al.* as discussed herein. Briefly, Mackenzie *et al.*, *Applied and Environmental Microbiology*, Vol. 66 (2000), pp. 4655-4661, reported the stable nuclear

transformation of *Mortierella alpina* with plasmid DNA. Using a protoplast transformation method, MacKenzie introduced the plasmid pD4 into *Mortierella alpina*. Plasmid pD4 comprised a hygromycin B resistance cassette, comprising sequence encoding the hygromycin B phosphotransferase gene product (*hpt*), operably-linked to the

5 *Mortierella alpina* histone H4.1 gene promoter (GenBank Accession No. AJ249812) upstream of the *hpt* protein-coding region and operably linked to the *Aspergillus nidulans* N-(5'-phosphoribosyl)anthranilate isomerase (*trpC*) gene 3'UTR/terminator downstream of the *hpt* protein-coding region. Prior to transformation with pD4, *Mortierella alpina* were unable to propagate on medium comprising 300 ug/ml hygromycin. Upon transformation

10 with pD4, transformed *Mortierella alpina* were obtained that were propagated on medium comprising 300 ug/ml hygromycin, thereby establishing the hygromycin B antibiotic resistance cassette as a selectable marker for use in *Mortierella alpina*. Evaluation of the genomic DNA of the stable transformants was performed by Southern. Mackenzie reported that selection and maintenance of the transformed *Mortierella alpina* was

15 performed on PDA (potato dextrose agar) medium comprising hygromycin. Liquid culturing of transformed *Mortierella alpina* by Mackenzie was performed in PDA medium or in S2GYE medium (comprising 5% glucose, 0.5% yeast extract, 0.18% NH₄SO₄, 0.02% MgSO₄·7H₂O, 0.0001% FeCl₃·6H₂O, 0.1%, trace elements, 10 mM K₂HPO₄–NaH₂PO₄), with hygromycin. Other media and techniques used for culturing *Mortierella alpina* have

20 been reported and other plasmids, promoters, 3' UTRs, and selectable markers for use in *Mortierella alpina* have been reported (for example see Ando *et al.*, *Applied and Environmental Biology*, Vol. 75:17 (2009) pp. 5529-35 and Lu *et al.*, *Applied Biochemistry and Biotechnology*, Vol. 164:7 (2001), pp. 979-90). Mackenzie reported that the transformation vector pD4 and the *Mortierella alpina* histone H4.1 promoter and *A. nidulans* *trpC* gene 3' UTR/terminator are suitable to enable heterologous gene expression

25 in *Mortierella alpina*. In addition, Mackenzie reported that the hygromycin resistance cassette encoded on pD4 was suitable for use as a selectable marker in *Mortierella alpina*.

[0343] In an embodiment of the present invention, vector pD4, comprising the nucleotide sequence encoding the *hpt* gene product for use as a selectable marker, is

30 constructed and modified to further comprise a lipid biosynthesis pathway expression cassette sequence, thereby creating a transformation vector. The lipid biosynthesis pathway expression cassette encodes one or more lipid biosynthesis pathway proteins selected from Table 25, each protein-coding sequence codon-optimized for expression in *Mortierella alpina* to reflect the codon bias inherent in nuclear genes of *Mortierella alpina*

in accordance with Tables 24A-D. For each lipid biosynthesis pathway protein of Table 25, the codon-optimized gene sequence can individually be operably linked to the *Mortierella alpina* histone H4.1 gene promoter upstream of the protein-coding sequence and operably linked to the *A. nidulans trpC* 3'UTR/terminator at the 3' region, or
5 downstream, of the protein-coding sequence. The transformation construct may additionally comprise homology regions to the *Mortierella alpina* genome for targeted genomic integration of the transformation vector. Homology regions may be selected to disrupt one or more genomic sites of endogenous lipid biosynthesis pathway genes. One skilled in the art can identify such homology regions within the sequence of the
10 *Mortierella alpina* genome (referenced in the publication by Wang et al., *PLOS One*, Vol. 6:12 (2011)). Stable transformation of *Mortierella alpina* with the transformation vector is achieved through well-known transformation techniques including protoplast transformation or other known methods. Activity of the *hpt* gene product may be used as a marker to select for *Mortierella alpina* transformed with the transformation vector on,
15 but not limited to, PDA medium comprising hygromycin. Growth media suitable for *Mortierella alpina* lipid production include, but are not limited to, *S2GYE* medium, and those culture media described by Lu *et al.* and Ando *et al.* Evaluation of fatty acid profiles of *Mortierella alpina* lipids may be assessed through standard lipid extraction and analytical methods described herein.

20

EXAMPLE 33: ENGINEERING *RHODOCOCCUS OPACUS* PD630

[0344] Expression of recombinant genes in accordance with the present invention in *Rhodococcus opacus* PD630 may be accomplished by modifying the methods and vectors taught by Kalscheuer *et al.* as discussed herein. Briefly, Kalscheuer *et al.*, *Applied and*
25 *Environmental Microbiology*, Vol. 52 (1999), pp. 508-515, reported the stable transformation of *Rhodococcus opacus* with plasmid DNA. Using the transformation method of electroporation, Kalscheuer introduced the plasmid pNC9501 into *Rhodococcus opacus* PD630. Plasmid pNC9501 comprised a thiostrepton resistance (thio^r) cassette, comprising the full nucleotide sequence of the *Streptomyces azureus* 23S rRNA A1067
30 methyltransferase gene, including the gene's promoter and 3' terminator sequence. Prior to transformation with pNC9501, *Rhodococcus opacus* was unable to propagate on medium comprising 1 mg/ml thiostrepton. Upon transformation of *Rhodococcus opacus* PD630 with pNC9501, transformants were obtained that propagated on culture medium comprising 1 mg/ml thiostrepton, thereby establishing the use of the thiostrepton

resistance cassette as a selectable marker in *Rhodococcus opacus* PD630. A second plasmid described by Kalscheuer, pAK68, comprised the resistance thio^r cassette as well as the gene sequences of the *Ralstonia eutropha* beta-ketothiolase (phaB), acetoacetyl-CoA reductase (phaA), and poly3-hydroxyalkanoic acid synthase (phaC) genes for polyhydroxyalkanoate biosynthesis, driven by the lacZ promoter. Upon pAK68 transformation of a *Rhodococcus opacus* PD630 strain deficient in polyhydroxyalkanoate biosynthesis, transformed *Rhodococcus opacus* PD630 were obtained that produced higher amounts of polyhydroxyalkanoates than the untransformed strain. Detectable activity of the introduced *Ralstonia eutropha* phaB, phaA, and phaC enzymes indicated that the regulatory elements encoded on the pAK68 plasmid were suitable for heterologous gene expression in *Rhodococcus opacus* PD630. Kalscheuer reported that selection and maintenance of the transformed *Rhodococcus opacus* PD630 was performed on standard Luria Broth (LB) medium, nutrient broth (NB), or mineral salts medium (MSM) comprising thioestrepton. Other media and techniques used for culturing *Rhodococcus opacus* PD630 have been described (for example see Kurosawa *et al.*, *Journal of Biotechnology*, Vol. 147:3-4 (2010), pp. 212-218 and Alvarez *et al.*, *Applied Microbial and Biotechnology*, Vol. 54:2 (2000), pp.218-223). Kalscheuer reported that the transformation vectors pNC9501 and pAK68, the promoters of the *Streptomyces azureus* 23S rRNA A1067 methyltransferase gene and lacZ gene are suitable to enable heterologous gene expression in *Rhodococcus opacus* PD630. In addition, Kalscheuer reported that the thio^r cassette encoded on pNC9501 and pAK68 was suitable for use as a selectable marker in *Rhodococcus opacus* PD630.

[0345] In an embodiment of the present invention, vector pNC9501, comprising the nucleotide sequence encoding the thio^r gene product for use as a selectable marker, is constructed and modified to further comprise a lipid biosynthesis pathway expression cassette sequence, thereby creating a transformation vector. The lipid biosynthesis pathway expression cassette encodes one or more lipid biosynthesis pathway proteins selected from Table 25, each protein-coding sequence codon-optimized for expression in *Rhodococcus opacus* PD630 to reflect the codon bias inherent in nuclear genes of *Rhodococcus opacus* in accordance with Tables 24A-D. For each lipid biosynthesis pathway protein of Table 25, the codon-optimized gene sequence can individually be operably linked to the lacZ gene promoter upstream of the protein-coding sequence. The transformation construct may additionally comprise homology regions to the *Rhodococcus opacus* PD630 genome for targeted genomic integration of the transformation vector.

Homology regions may be selected to disrupt one or more genomic sites of endogenous lipid biosynthesis pathway genes. One skilled in the art can identify such homology regions within the sequence of the *Rhodococcus opacus PD630* genome (referenced in the publication by Holder *et al.*, *PLOS Genetics*, Vol. 7:9 (2011). Transformation of

5 *Rhodococcus opacus PD630* with the transformation vector is achieved through well-known transformation techniques including electroporation or other known methods. Activity of the *Streptomyces azureus* 23S rRNA A1067 methyltransferase gene product may be used as a marker to select for *Rhodococcus opacus PD630* transformed with the transformation vector on, but not limited to, LB medium comprising thiostrepton. Growth

10 media suitable *Rhodococcus opacus PD630* lipid production include, but are not limited to those culture media discussed by Kurosawa *et al.* and Alvarez *et al.* Evaluation of fatty acid profiles of *Rhodococcus opacus PD630* lipids may be assessed through standard lipid extraction and analytical methods described herein.

15 **EXAMPLE 34: ENGINEERING MICROALGAE FOR FATTY ACID AUXOTROPHY**

[0346] *Prototheca moriformis* (UTEX 1435) engineered to express a *Cuphea wrightii* thioesterase (CwTE2), was used as the host organism for further genetic modification to knockout both endogenous thioesterase alleles, FATA1-1 and FATA1-2. Here, a first

20 transformation construct was generated to integrate a neomycin expression cassette at the FATA1-1 locus. This construct, pSZ2226, included 5' (SEQ ID NO: 150) and 3' (SEQ ID NO: 151) homologous recombination targeting sequences (flanking the construct) to the FATA1-1 locus of the nuclear genome and a neomycin resistance protein-coding sequence under the control of the *C. reinhardtii* β -tubulin promoter/5'UTR (SEQ ID NO: 5) and the

25 *Chlorella vulgaris* nitrate reductase 3' UTR (SEQ ID NO: 126). This NeoR expression cassette is listed as SEQ ID NO: 135 and served as a selectable marker.

[0347] Upon transformation of pSZ2226, individual transformants were selected on agar plates comprising sucrose and G418. A single isolate, Strain H, was selected for further genetic modification. A second transformation construct, pSZ2236, was generated

30 to integrate polynucleotides enabling expression of a thiamine selectable marker into Strain H at the FATA1-2 locus. pSZ2236 included 5' (SEQ ID NO: 152) and 3' (SEQ ID NO: 153) homologous recombination targeting sequences (flanking the construct) to the FATA1-2 genomic region for integration into the *P. moriformis* nuclear genome and an *A. thaliana* THIC protein coding region under the control of the *C. protothecoides* actin

promoter/5'UTR (SEQ ID NO: 142) and *C. vulgaris* nitrate reductase 3' UTR (SEQ ID NO: 126). This *At*THIC expression cassette is listed as SEQ ID NO: 143 and served as a selectable marker. Upon transformation of Strain H with pSZ2236 to generate Strain I, individual transformants, were selected on agar plates comprising free fatty acids. Strain I
 5 was able to propagate on agar plates and in medium lacking thiamine and supplemented with free fatty acids.

EXAMPLE 35: ENGINEERING MICROORGANISMS FOR INCREASED PRODUCTION OF STEARIC ACID

10 [0348] A classically mutagenized strain of *Prototheca moriformis* (UTEX 1435), Strain J, was transformed with the plasmid construct pSZ2281 according to biolistic transformation methods as described herein and in PCT/US2009/066141, PCT/US2009/066142, PCT/US2011/038463, PCT/US2011/038464, and PCT/US2012/023696. pSZ2281 included polynucleotides encoding RNA hairpins
 15 (SAD2hpC, SEQ ID NO: 154) to down-regulate the expression of stearyl-ACP desaturase, 5' (SEQ ID NO: 121) and 3' (SEQ ID NO: 122) homologous recombination targeting sequences (flanking the construct) to the 6S genomic region for integration into the nuclear genome, and a *S. cerevisiae* *suc2* sucrose invertase coding region (SEQ ID NO: 124), to express the protein sequence given in SEQ ID NO: 123, under the control of
 20 *C. reinhardtii* β -tubulin promoter/5'UTR (SEQ ID NO: 125) and *Chlorella vulgaris* nitrate reductase 3' UTR (SEQ ID NO: 126). This *S. cerevisiae* *suc2* expression cassette is listed as SEQ ID NO: 127 and served as a selectable marker. The polynucleotide sequence encoding the SAD2hpC RNA hairpin was under the control of the *C. protothecoides* actin promoter/5'UTR (SEQ ID NO: 142) and *C. vulgaris* nitrate reductase 3' UTR (SEQ ID
 25 NO: 126).

[0349] Upon transformation of Strain J with construct pSZ2281 thereby generating Strain K, positive clones were selected on agar plates containing sucrose as a sole carbon source. Individual transformants were clonally purified and propagated under heterotrophic conditions suitable for lipid production as those detailed herein and in
 30 PCT/US2009/066141, PCT/US2009/066142, PCT/US2011/038463, PCT/US2011/038464, and PCT/US2012/023696. Lipid samples were prepared from dried biomass and analyzed using standard fatty acid methyl ester gas chromatography flame ionization detection methods as described in Example 3 (also see PCT/US2012/023696). The fatty acid profiles (expressed as Area % of total fatty acids) of

P. moriformis UTEX Strain J propagated on glucose as a sole carbon source and three representative isolates of Strain K, propagated on sucrose as a sole carbon source, are presented in Table 26.

Table 26. Fatty acid profiles of *Prototheca moriformis* cells engineered to express a hairpin RNA construct targeting stearoyl ACP desaturase gene/gene products.

Area % Fatty acid	Strain J	Strain K-1	Strain K-2	Strain K-3
C10:0	0.01	0.00	0.02	0.02
C12:0	0.03	0.05	0.05	0.05
C14:0	1.22	0.89	0.87	0.77
C16:0	26.75	29.23	28.96	27.55
C18:0	3.06	37.39	36.76	36.41
C18:1	59.62	23.90	24.76	26.92
C18:2	7.33	5.44	5.54	5.54

[0350] The data presented in Table 26 show a clear impact of the expression of a SAD2 hairpin RNA construct on the C18:0 and C18:1 fatty acid profiles of the transformed organism. The fatty acid profiles of Strain K transformants comprising a SAD2 hairpin RNA construct demonstrated an increase in the percentage of saturated C18:0 fatty acids with a concomitant diminution of unsaturated C18:1 fatty acids. Fatty acid profiles of the untransformed strain comprise about 3% C18:0. Fatty acid profiles of the transformed strains comprise about 37% C18:0. These data illustrate the successful expression and use of polynucleotides enabling expression of a SAD RNA hairpin construct in *Prototheca moriformis* to alter the percentage of saturated fatty acids in the engineered host microbes, and in particular in increasing the concentration of C18:0 fatty acids and decreasing C18:1 fatty acids in microbial cells.

EXAMPLE 36: ENGINEERING MICROORGANISMS FOR INCREASED PRODUCTION OF OLEIC ACID THROUGH KNOCKDOWN OF AN ENDOGENOUS ACYL-ACP THIOESTERASE

[0351] A classically mutagenized strain of *Prototheca moriformis* (UTEX 1435), Strain J, was transformed independently with each of the constructs pSZ2402-pSZ2407 according to biolistic transformation methods as described herein and in PCT/US2009/066141, PCT/US2009/066142, PCT/US2011/038463, PCT/US2011/038464, and PCT/US2012/023696. Each of the constructs pSZ2402-pSZ2407 included different polynucleotides encoding a hairpin RNA targeted against *Prototheca moriformis* FATA1

mRNA transcripts to down-regulate the expression of fatty acyl-ACP thioesterase, 5' (SEQ ID NO: 121) and 3' (SEQ ID NO: 122) homologous recombination targeting sequences (flanking the construct) to the 6S genomic region for integration into the nuclear genome, and a *S. cerevisiae* *suc2* sucrose invertase coding region (SEQ ID NO: 124) to express the protein sequence given in SEQ ID NO: 123 under the control of *C. reinhardtii* β -tubulin promoter/5'UTR (SEQ ID NO: 125) and *Chlorella vulgaris* nitrate reductase 3' UTR (SEQ ID NO: 126). This *S. cerevisiae* *suc2* expression cassette is listed as SEQ ID NO: 127 and served as a selectable marker. Sequence listing identities for the polynucleotides corresponding to each hairpin are listed in Table 27. The polynucleotide sequence encoding each RNA hairpin was under the control of the *C. reinhardtii* β -tubulin promoter/5'UTR (SEQ ID NO: 125) and *C. vulgaris* nitrate reductase 3' UTR (SEQ ID NO: 126).

Table 27. Plasmid constructs used to transform *Prototheca moriformis* (UTEX 1435) Strain J.

1. Plasmid construct	2. Hairpin designation	3. SEQ ID NO:
4. pSZ2402	5. PmFATA-hpB	6. SEQ ID NO: 160
7. pSZ2403	8. PmFATA-hpC	9. SEQ ID NO: 161
10. pSZ2404	11. PmFATA-hpD	12. SEQ ID NO: 162
13. pSZ2405	14. PmFATA-hpE	15. SEQ ID NO: 163
16. pSZ2406	17. PmFATA-hpF	18. SEQ ID NO: 164
19. pSZ2407	20. PmFATA-hpG	21. SEQ ID NO: 165

[0352] Upon independent transformation of Strain J with each of the constructs listed in Table 27, positive clones were selected on agar plates containing sucrose as a sole carbon source. Individual transformants were clonally purified and propagated under heterotrophic conditions suitable for lipid production as those detailed in PCT/US2009/066141, PCT/US2009/066142, PCT/US2011/038463, PCT/US2011/038464, and PCT/US2012/023696. Lipid samples were prepared from dried biomass and analyzed using standard fatty acid methyl ester gas chromatography flame ionization detection methods as described in Example 3. The fatty acid profiles (expressed as Area % of total fatty acids) of *P. moriformis* UTEX Strain J propagated on glucose as a sole carbon source and representative isolates of each transformation of Strain J, propagated on sucrose as a sole carbon source, are presented in Table 28.

Table 28. Fatty acid profiles of *Prototheca moriformis* cells engineered to express hairpin RNA constructs targeting fatty acyl-ACP thioesterase gene/gene products.

Construct	Area % Fatty Acid						
	C10:0	C12:0	C14:0	C16:0	C18:0	C18:1	C18:2
Strain J untransformed	0	0.05	1.32	26.66	3.1	59.07	7.39
PmFATA-hpB	0.04	0.07	1.36	24.88	2.24	61.92	6.84
	0	0.08	1.33	25.34	2.39	61.72	6.5
	0	0.07	1.29	25.44	2.26	61.7	6.69
	0	0.06	1.33	25.1	2.37	61.56	6.87
PmFATA-hpC	0	0.08	1.18	22.03	1.71	63.8	8.63
	0	0.07	1.21	24.5	2.23	62.32	7.19
	0	0.08	1.29	24.93	2.24	62.02	7.01
	0.05	0.06	1.29	25.45	2.26	61.81	6.76
PmFATA-hpD	0	0.02	0.68	15.8	1.88	72.64	6.96
	0	0.03	0.78	17.56	1.7	71.8	6.03
	0	0.03	0.92	19.04	2.03	68.82	7.05
	0	0.04	1.27	23.14	2.25	65.27	6.07
PmFATA-hpE	0	0.03	0.79	18.55	2.13	69.66	6.77
	0	0.04	1.11	21.01	1.74	65.18	8.55
	0	0.03	1.08	21.11	1.54	64.76	8.87
	0	0.03	1.17	21.93	1.71	63.89	8.77
PmFATA-hpF	0.03	0.04	0.34	8.6	1.69	78.08	8.87
	0	0.03	0.49	10.2	1.52	76.97	8.78
	0	0.03	1	20.47	2.22	66.34	7.45
	0	0.03	1.03	21.61	1.88	65.39	7.76
PmFATA-hpG	0	0.03	1.03	20.57	2.36	64.73	8.75
	0	0.03	1.2	24.39	2.47	61.9	7.49
	0	0.04	1.29	24.14	2.29	61.41	8.22

[0353] The data presented in Table 28 show a clear impact of the expression of FATA hairpin RNA constructs on the C18:0 and C18:1 fatty acid profiles of the transformed organism. The fatty acid profiles of Strain J transformants comprising a FATA hairpin RNA construct demonstrated an increase in the percentage of C18:1 fatty acids with a concomitant diminution of C16:0 and C18:0 fatty acids. Fatty acid profiles of the untransformed Strain J are about 26.66% C16:0, 3% C18:0, and about 59% C18:1 fatty acids. In contrast, the fatty acid profiles of the transformed strains comprise as low as 8.6% C16:0 and 1.54% C18:0 and greater than 78% C18:1 fatty acids. These data indicate that FATA1 enzyme of *Prototheca moriformis* (UTEX 1435) displays a preferential specificity for hydrolysis of fatty acids of length C18.

[0354] These data illustrate the utility and successful use of polynucleotide FATA RNA hairpin constructs in *Prototheca moriformis* to alter the fatty acids profile of

engineered microbes, and in particular in increasing the concentration of C18:1 fatty acids and decreasing C18:0 and C16:0 fatty acids in microbial cells.

EXAMPLE 37: ALTERING THE LEVELS OF FATTY ACIDS OF ENGINEERED MICROBES THROUGH MULTIPLE ALLELIC DISRUPTION OF A FATTY ACID DESATURASE

[0355] This example describes the use of a transformation vector to disrupt the FADc loci of *Protheca moriformis* with a transformation cassette comprising a selectable marker and sequence encoding an exogenous SAD enzyme to engineer microorganisms in which the fatty acid profile of the transformed microorganism has been altered.

[0356] A classically mutagenized (for higher oil production) derivative of *Protheca moriformis* (UTEX 1435), strain A, was transformed with the transformation construct pSZ1499 (SEQ ID NO: 246) according to biolistic transformation methods detailed in Example 2. pSZ1499 comprised nucleotide sequence of the *Olea europaea* stearyl-ACP desaturase gene, codon-optimized for expression in *Protheca moriformis* UTEX 1435. The pSZ1499 expression construct contained 5' (SEQ ID NO: 247) and 3' (SEQ ID NO: 248) homologous recombination targeting sequences (flanking the construct) to the FADc genomic region for integration into the nuclear genome and a *S. cerevisiae* suc2 sucrose invertase coding region under the control of *C. reinhardtii* β -tubulin promoter/5'UTR and *Chlorella vulgaris* nitrate reductase 3' UTR. This *S. cerevisiae* suc2 expression cassette is listed as SEQ ID NO: 159 and served as a selection marker. The *Olea europaea* stearyl-ACP desaturase coding region was under the control of the *Protheca moriformis* Amt03 promoter/5'UTR (SEQ ID NO: 89) and *C. vulgaris* nitrate reductase 3'UTR, and the native transit peptide was replaced with the *Chlorella protothecoides* stearyl-ACP desaturase transit peptide (SEQ ID NO: 49). The entire *O. europaea* SAD expression cassette was termed pSZ1499 and may be written as FADc5'_btub-Suc2-nr_amt03-S106SAD-OeSAD-nr-FADc3'.

[0357] Primary transformants were selected on plates containing sucrose as a sole carbon source. Individual transformants were clonally purified and grown under standard lipid production conditions at pH 7.0, similar to the conditions as disclosed in Example 1. Fatty acid profiles were analyzed using standard fatty acid methyl ester gas chromatography flame ionization (FAME GC/FID) detection methods as described in Example 3. The resulting fatty acid profiles from a set of representative clones arising from the transformations of the transformation vector are shown in Table 29. Fatty acid

profiles of lipids obtained from the untransformed strain C strain grown under lipid production conditions comprising glucose as a sole carbon source (pH 5.0) are additionally presented in Table 29.

Table 29. Fatty acid profiles of *Prototheca moriformis* (UTEX 1435) multiply engineered to knockout endogenous FADc alleles and to express an *O. europaea* stearoyl-ACP desaturase.

Strain	Transformant	% C16:0	% C18:0	% C18:1	% C18:2
strain A	untransformed	28.50	3.72	57.70	7.04
strain A	untransformed	28.57	3.69	57.61	7.07
strain A pSZ1499	Transformant 1	20.37	1.13	74.38	0.01
	Transformant 2	19.98	1.16	74.60	0.00
	Transformant 3	20.10	1.16	74.70	0.00
	Transformant 4	21.13	1.21	73.86	0.00
	Transformant 5	19.95	1.11	74.58	0.00
	Transformant 6	20.20	1.14	74.61	0.00
	Transformant 7	20.72	1.15	74.15	0.00
	Transformant 8	20.06	1.11	74.44	0.00
	Transformant 9	19.86	1.18	74.88	0.00

[0358] As shown in Table 29, transformation of strain C with pSZ1499 impacts the fatty acid profiles of the transformed microbes. The untransformed *Prototheca moriformis* (UTEX 1435) strain C strain exhibits a fatty acid profile comprising less than 60% C18:1 fatty acids and greater than 7% C18:2 fatty acids. In contrast, strain C strains transformed with pSZ1499 exhibited fatty acid profiles with an increased composition of C18:1 fatty acids and a concomitant decrease in C18:0 and C18:2 fatty acids. C18:2 fatty acids were undetected in the fatty acid profiles of strain C transformed with pSZ1499. The absence of detectable C18:2 fatty acids in pSZ1499 transformants indicated that the transformation with pSZ1499, bearing homologous recombination targeting sequences for integration into multiple FADc genomic loci, had abolished FAD activity.

[0359] Southern blot analysis was conducted to verify that multiple FADc alleles were interrupted by the pSZ1499 transformation vector. Genomic DNA was extracted from strain C and pSZ1499 transformants using standard molecular biology methods. DNA from each sample was run on 0.8% agarose gels after digestion with the restriction enzyme PstI. DNA from this gel was transferred onto a Nylon+ membrane (Amersham), which was then hybridized with a P32-labeled polynucleotide probe corresponding to FADc 3' region. Figure 1 shows maps of the pSZ1499 transformation cassette, the two sequenced

FADc alleles of *Prototheca moriformis* (UTEX 1435), and the predicted sizes of the alleles disrupted by the pSZ1499 transformation vector. FADc allele 1 comprises a PstI restriction site, whereas FADc allele 2 does not. Integration of the SAD cassette would introduce a PstI restriction site into the disrupted FADc allele, resulting in a ~6kb
5 fragment resolved on the Southern, regardless of which allele was disrupted. Figure 2 shows the results of Southern blot analysis. A hybridization band at ~6kb is detected in both transformants. No smaller hybridization bands, that would be indicative of uninterrupted alleles, were detected. These results indicate that both FADc alleles were disrupted by pSZ1499.

10 [0360] The ablation of both alleles of the FADc fatty acid desaturase with a SAD expression cassette results in fatty acid profiles comprising about 74% C18:1.

Collectively, these data demonstrate the utility and effectiveness of polynucleotides permitting knockout of FAD alleles and concomitant exogenous expression of stearyl-ACP desaturase enzymes to alter the fatty acid profile of engineered microorganisms.

15 [0361] All references cited herein, including patents, patent applications, and publications, including Genbank Accession numbers, are hereby incorporated by reference in their entireties, whether previously specifically incorporated or not. The publications mentioned herein are cited for the purpose of describing and disclosing reagents, methodologies and concepts that may be used in connection with the present invention.

20 Nothing herein is to be construed as an admission that these references are prior art in relation to the inventions described herein. In particular, the following patent applications are hereby incorporated by reference in their entireties for all purposes: PCT Application No. PCT/US2008/065563, filed June 2, 2008, entitled "Production of Oil in Microorganisms", PCT Application No. PCT/US2010/31108, filed April 14, 2010, entitled

25 "Methods of Microbial Oil Extraction and Separation", PCT Publication No. WO 2010/063032, filed November 30, 2009, entitled "Production of Tailored Oils in Heterotrophic Microorganisms", PCT Application No. PCT/US2011/038463, filed May 27, 2011, entitled "Tailored Oils Produced From Recombinant Heterotrophic Microorganisms", and PCT Application No. PCT/US/2012/023696, filed February 2,

30 2012, entitled "Tailored Oils Produced from Recombinant Heterotrophic Microorganisms."

SEQUENCE LISTING

SEQ ID NO: 1

UTEX 329 *Prototheca kruegeri*

5

TGTTGAAGAATGAGCCGGCGAGTTAAAAAGAGTGGCATGGTTAAAGAAAATACTCTGGAGCCATAGCGAAAAG
CAAGTTTGTAGTAAGCTTAGGTCATTCTTTTTAGACCCGAAACCGAGTGATCTACCCATGATCAGGGTGAAGTG
TTAGTAAAATAACATGGAGGCCCGAACCGACTAATGTTGAAAAATTAGCGGATGAATTGTGGGTAGGGGCGA
AAAACCAATCGAACTCGGAGTTAGCTGGTTCTCCCCGAAATGCGTTTAGGCGCAGCAGTAGCAGTACAAATA
10 GAGGGGTAAAGCACTGTTTCTTTTGTGGGCTTCGAAAGTTGTACCTCAAAGTGGCAAACCTCTGAATACTCTA
TTTAGATATCTACTAGTGAGACCTTGGGGGATAAGCTCCTTGGTCAAAGGGGAAACAGCCCAGATCACCAGT
TAAGGCCCCAAAATGAAAATGATAGTGAAGGATGTGGGTATGTCAAACCTCCAGCAGGTTAGCTTAGA
AGCAGCAATCCTTTCAAGAGTGCGTAATAGCTCACTG

SEQ ID NO: 2

UTEX 1440 *Prototheca wickerhamii*

15

TGTTGAAGAATGAGCCGGCGACTTAAAAATAAATGGCAGGCTAAGAGATTTAATAAATCGAAACCTAAGCGAA
AGCAAGTCTTAATAGGGCGTCAATTTAACAAAACCTTTAAATAAATTATAAAGTCATTTATTTTAGACCCGAA
20 CCTGAGTGATCTAACCATGGTCAGGATGAAACTTGGGTGACACCAAGTGGAAAGTCCGAACCGACCGATGTTG
AAAAATCGGCGGATGAACCTGTGGT'TAGTGGTGAATACCAGTCGAACTCAGAGCTAGCTGGTTCTCCCCGAA
ATGCGTTGAGGCGCAGCAATATATCTCGTCTATCTAGGGGTAAAGCACTGTTTCGGTGCGGGCTATGAAAAT
GGTACCAAATCGTGGCAAACCTCTGAATACTAGAAATGACGATATATTAGTGAGACTATGGGGGATAAGCTCC
ATAGTCGAGAGGGGAAACAGCCCAGACCACAGTTAAGGCCCCAAAATGATAATGAAGTGGTAAAGGAGGTGA
25 AAATGCAAATACAACCAGGAGGTTGGCTTAGAAGCAGCCATCCTTTAAAGAGTGCGTAATAGCTCACTG

SEQ ID NO: 3

UTEX 1442 *Prototheca stagnora*

30

TGTTGAAGAATGAGCCGGCGAGTTAAAAAATGGCATGGTTAAAGATATTTCTCTGAAGCCATAGCGAAAAG
CAAGTTTTACAAGCTATAGTCATTTTTTTTTAGACCCGAAACCGAGTGATCTACCCATGATCAGGGTGAAGTG
TTGGTCAAATAACATGGAGGCCCGAACCGACTAATGGTGAATAATTAGCGGATGAATTGTGGGTAGGGGCGA
AAAACCAATCGAACTCGGAGTTAGCTGGTTCTCCCCGAAATGCGTTTAGGCGCAGCAGTAGCAACACAAAATA
35 GAGGGGTAAAGCACTGTTTCTTTTGTGGGCTTCGAAAGTTGTACCTCAAAGTGGCAAACCTCTGAATACTCTA
TTTAGATATCTACTAGTGAGACCTTGGGGGATAAGCTCCTTGGTCAAAGGGGAAACAGCCCAGATCACCAGT
TAAGGCCCCAAAATGAAAATGATAGTGAAGGACGTGAGTATGTCAAACCTCCAGCAGGTTAGCTTAGA
AGCAGCAATCCTTTCAAGAGTGCGTAATAGCTCACTG

SEQ ID NO: 4

UTEX 288 *Prototheca moriformis*

40

TGTTGAAGAATGAGCCGGCGAGTTAAAAAGAGTGGCATGGTTAAAGATAATTCTCTGGAGCCATAGCGAAAAG
CAAGTTTAAACAAGCTAAAGTCACCCTTTTTTAGACCCGAAACCGAGTGATCTACCCATGATCAGGGTGAAGTG
45 TTGGTAAAATAACATGGAGGCCCGAACCGACTAATGGTGAATAATTAGCGGATGAATTGTGGGTAGGGGCGA
AAAACCAATCGAACTCGGAGTTAGCTGGTTCTCCCCGAAATGCGTTTAGGCGCAGCAGTAGCAACACAAAATA
GAGGGGTAAAGCACTGTTTCTTTTGTGGGCTTCGAAAGTTGTACCTCAAAGTGGCAAACCTCTGAATACTCTA
TTTAGATATCTACTAGTGAGACCTTGGGGGATAAGCTCCTTGGTCAAAGGGGAAACAGCCCAGATCACCAGT
TAAGGCCCCAAAATGAAAATGATAGTGAAGGACGTGAGTATGTCAAACCTCCAGCAGGTTAGCTTAGA
50 AGCAGCAATCCTTTCAAGAGTGCGTAATAGCTCACTG

SEQ ID NO: 5

UTEX 1439, UTEX 1441, UTEX 1435, UTEX 1437 *Prototheca moriformis*

50

TGTTGAAGAATGAGCCGGCGACTTAAAAATAAATGGCAGGCTAAGAGAATTAATAAATCGAAACCTAAGCGAA
AGCAAGTCTTAATAGGGCGCTAATTTAACAAAACATTAATAAATCTAAAGTCATTTATTTTAGACCCGAA
55 CCTGAGTGATCTAACCATGGTCAGGATGAAACTTGGGTGACACCAAGTGGAAAGTCCGAACCGACCGATGTTG
AAAAATCGGCGGATGAACCTGTGGT'TAGTGGTGAATACCAGTCGAACTCAGAGCTAGCTGGTTCTCCCCGAA
ATGCGTTGAGGCGCAGCAATATATCTCGTCTATCTAGGGGTAAAGCACTGTTTCGGTGCGGGCTATGAAAAT
GGTACCAAATCGTGGCAAACCTCTGAATACTAGAAATGACGATATATTAGTGAGACTATGGGGGATAAGCTCC
60 ATAGTCGAGAGGGGAAACAGCCCAGACCACAGTTAAGGCCCCAAAATGATAATGAAGTGGTAAAGGAGGTGA
AAATGCAAATACAACCAGGAGGTTGGCTTAGAAGCAGCCATCCTTTAAAGAGTGCGTAATAGCTCACTG

SEQ ID NO: 6

UTEX 1533 *Prototheca wickerhamii*

65

TGTTGAAGAATGAGCCGTGCACTTAAAAATAAATGGCAGGCTAAGAGAATTAATAAATCGAAACCTAAGCGAA
AGCAAGTCTTAATAGGGCGCTAATTTAACAAAACATTAATAAATCTAAAGTCATTTATTTTAGACCCGAA
CCTGAGTGATCTAACCATGGTCAGGATGAAACTTGGGTGACACCAAGTGGAAAGTCCGAACCGACCGATGTTG
AAAAATCGGCGGATGAACCTGTGGT'TAGTGGTGAATACCAGTCGAACTCAGAGCTAGCTGGTTCTCCCCGAA

ATGCGTTGAGGCGCAGCAATATATCTCGTCTATCTAGGGGTAAAGCACTGTTTCGGTGCGGGCTATGAAAAT
GGTACCAAATCGTGGCAAACCTCTGAATACTAGAAATGACGATATATTAGTGAGACTATGGGGGATAAGCTCC
ATAGTCGAGAGGGAAACAGCCCAGACCACAGTTAAGGCCCAAATGATAATGAAGTGGTAAAGGAGGTGA
AAATGCAAATACAACCAGGAGGTTGGCTTAGAAGCAGCCATCCTTTAAAGAGTGCGTAATAGCTCACTG

5

SEQ ID NO: 7
UTEX 1434 *Prototheca moriformis*

TGTTGAAGAATGAGCCGCGCAGTTAAAAAGAGTGGCGTGGTTAAAGAAAATTCTCTGGAACCATAGCGAAAG
CAAGTTTAAACAAGCTTAAGTCACTTTTTTTTAGACCCGAAACCGAGTGATCTACCCATGATCAGGGTGAAGTG
TTGGTAAAATAACATGGAGGCCCGAACCAGCTAATGGTGAAGAAATTAGCGGATGAATTGTGGGTAGGGGCGA
AAAACCAATCGAACTCGGAGTTAGCTGGTTCTCCCCGAAATGCGTTTAGGCGCAGCAGTAGCAACACAAAATA
GAGGGGTAAAGCACTGTTTCTTTTGTGGGCTCCGAAAGTTGTACCTCAAAGTGGCAAACCTCTGAATACTCTA
TTTAGATATCTACTAGTGAGACCTTGGGGGATAAGCTCCTTGGTCAAAAGGGAAACAGCCCAGATCACCAGT
TAAGGCCCCCAAATGAAAATGATAGTGACTAAGGATGTGAGTATGTCAAACCTCCAGCAGGTTAGCTTAGA
AGCAGCAATCCTTTCAAGAGTGCGTAATAGCTCACTG

10

15

SEQ ID NO: 8
UTEX 1438 *Prototheca zopfii*

TGTTGAAGAATGAGCCGCGCAGTTAAAAAGAGTGGCATGGTTAAAGAAAATTCTCTGGAGCCATAGCGAAAG
CAAGTTTAAACAAGCTTAAGTCACTTTTTTTTAGACCCGAAACCGAGTGATCTACCCATGATCAGGGTGAAGTG
TTGGTAAAATAACATGGAGGCCCGAACCAGCTAATGGTGAAGAAATTAGCGGATGAATTGTGGGTAGGGGCGA
AAAACCAATCGAACTCGGAGTTAGCTGGTTCTCCCCGAAATGCGTTTAGGCGCAGCAGTAGCAACACAAAATA
GAGGGGTAAAGCACTGTTTCTTTTGTGGGCTCCGAAAGTTGTACCTCAAAGTGGCAAACCTCTGAATACTCTA
TTTAGATATCTACTAGTGAGACCTTGGGGGATAAGCTCCTTGGTCAAAAGGGAAACAGCCCAGATCACCAGT
TAAGGCCCCCAAATGAAAATGATAGTGACTAAGGATGTGAGTATGTCAAACCTCCAGCAGGTTAGCTTAGA
AGCAGCAATCCTTTCAAGAGTGCGTAATAGCTCACTG

20

25

SEQ ID NO: 9
UTEX 1436 *Prototheca moriformis*

TGTTGAAGAATGAGCCGCGCAGTTAGAAAAGGTGGCATGGTTAAAGAAAATTCTCCGAAGCCGTAGCAAAAGC
GAGTCTGAATAGGGCGATAAAAATATATTAATATTTAGAATCTAGTCATTTTTTTCTAGACCCGAACCCGGGTG
ATCTAACCATGACCAGGATGAAGCTTGGGTGATACCAAGTGAAGGTCCGAACCGACCGATGTTGAAAAATCG
GCGGATGAGTTGTGGTTAGCGGTGAAATACCAAGTCAACCCGAGCTAGCTGGTTCTCCCCGAAATGCGTTG
AGGCGCAGCAGTACATCTAGTCTATCTAGGGGTAAAGCACTGTTTCGGTGCGGGCTGTGAGAACGGTACCAA
ATCGTGGCAAACCTCTGAATACTAGAAATGACGATGTAGTAGTGAGACTGTGGGGGATAAGCTCCATTGTCAA
GAGGGAAACAGCCCAGACCACAGCTAAGGCCCAAATGGTAATGTAGTGACAAAGGAGGTGAAAATGCAA
ATACAACCAGGAGGTTGGCTTAGAAGCAGCCATCCTTTAAAGAGTGCGTAATAGCTCACTG

30

35

40

SEQ ID NO: 10
HUP promoter from *Chlorella* (subsequence of GenBank accession number
X55349)

GATCAGACGGGCGCTGACCTGCGAGATAATCAAGTGCTCGTAGGCAACCAACTCAGCAGCTGCTTGGTGTGG
GTCTGCAGGATAGTGTGTCAGGGCCCCAAGGACAGCAGGGGAACCTTACACCTTGTCCCCGACCCAGTTTTAT
GGAGTGCAATTGCCTCAAGAGCCTAGCCGGAGCGCTAGGCTACATACTTGCCGCACCCGGTATGAGGGGATATA
GTACTCGCACTGCGCTGTCTAGTGAGATGGGCAGTGCTGCCATAAACAACCTGGCTGCTCAGCCATTTGTTG
GCGGACCATTTCTGGGGGGCCAGCAATGCCCTGACTTTCGGGTAGGGTGAAAACCTGAACAAAGACTACCAAAA
CAGAAATTTCTCTCTCTCTGGAGGTAAGCGCAGGCCGCGCCCGCTGCGCCCATGCGCGCTCCGAACACCTCC
ATAGCTGTAAGGGCGCAAACATGGCCGAGCTGTTGTGACGACTCTTTTCATGGCCATACAAGGTCATGTCGAG
ATTAGTGCTGAGTAAGACACTATCACCCCATGTTTCGATTGAAGCCGTGACTTCATGCCAACCTGCCCTGGG
CGTAGCAGACGTATGCCATCATGACCACTAGCCGACATGCGCTGTCTTTTGCCACCAAAACAACCTGGTACAC
CGCTCGAAGTCGTGCCGCACACCTCCGGGAGTGAGTCCGGCGACTCCTCCCCGGCGGGGCCGCGGCCCTACCT
GGGTAGGGTCGCCATACGCCACGACCAAACGACGAGGAGGGGATTGGGGTAGGGAATCCCAACCGACCTA
ACCAAGACGCGCACTATAATAATAGGTGGGGGACTAACAGCCCTATATCGCAAGCTTTGGGTGCCTATCTT
GAGAAGCACGAGTTGGAGTGGCTGTGTACGGTCGACCCTAAGGTGGGTGTGCCGCAGCCTGAAACAAAGCGT
CTAGCAGCTGCTTCTATAATGTGTGACCCGTTGTGTTTCAGTTATATTGTATGCTATTGTTTGTTCGTGCTA
GGGTGGCGCAGGCCACCTACTGTGGCGGGCCATTGGTTGGTGCTTGAATTGCCTCACCATCTAAGGTCTGA
ACGCTCACTCAAACGCCCTTTGTACAACCTGCAGAACTTTCCTTGGCGCTGCAACTACAGTGTGCAAACAGCA
CATAGCACTCCCTTACATCACCCAGCAGTACAACA

45

50

55

60

SEQ ID NO: 11
Chlorella ellipsoidea nitrate reductase promoter from AY307383

CGCTGCGCACCAAGGGCCGCCAGCTCGCTGATGTCGCTCCAAATGCGGTCCCCCGATTGTTTGTCTTTCATCT
TCTCCACCTTGGTGGCTTCTTGGCCAGGGCCTTCAGCTGCATGCGCACAGACCGTTGAGCTCCTGATCAGC
ATCCTCAGGAGGCCCTTTGACAAGCAAGCCCCTGTGCAAGCCCATTACGGGGTACCAGTGGTGCTGAGGTA
GATGGGTTTGAAAAGGATTGCTCGGTGATGCTGCTCATGGAATTGGCATGTGCATGCATGTTTACAATAT

65

70

GCCACCAGGCTTTGGAGCAAGAGAGCATGAATGCCTTCAGGCAGGTTGAAAGTTCCTGGGGGTGAAGAGGCA
GGGCCGAGGATTGGAGGAGGAAAGCATCAAGTCGTCGCTCATGCTCATGTTTTTCAGTCAGAGTTTGCCAAGC
TCACAGGAGCAGAGACAAGACTGGCTGCTCAGGTGTTGCATCGTGTGTGGTGGGGGGGGGGGGTTAATA
CGGTACGAAATGCACTTGGAATCCCACCTCATGCCAGCGGACCCACATGCTTGAATTCGAGGCCTGTGGGG
5 TGAGAAATGCTCACTCTGCCCTCGTTGCTGAGGTACTTCAGGCCGCTGAGCTCAAAGTCGATGCCCTGCTCG
TCTATCAGGGCCTGCACCTCTGGGCTGACCGGCTCAGCCTCCTTCGCGGGCATGGAGTAGGCGCCGGCAGCG
TTCATGTCCGGGGCCAGGGCAGCGGTGGTGCCATAAATGTCGGTGATGGTGGGGAGGGGGGCCGTGCCACA
CCATTGCCGTTGCTGGCTGACGCATGCACATGTGGCCTGGCTGGCACCGGCAGCACTGGTCTCCAGCCAGCC
10 AGCAAGTGGCTGTTTCAGGAAAGCGGCCATGTTGTTGGTCCCTGCGCATGTAATTCCTCCAGATCAAAGGAGGG
AACAGCTTGGATTTGATGTAGTGCCCAACCGGACTGAATGTGCGATGGCAGGTCCCTTTGAGTCTCCCGAAT
TACTAGCAGGGCACTGTGACCTAACGCAGCATGCCAACCGCAAAAAAATGATTGACAGAAAATGAAGCGGTG
TGTCAAATATTTGCTGTATTTATTCGTTTAAATCAGCAACCAAGTTCGAAACGCAACTATCGTGGTGATCAAG
TGAACCTCATCAGACTTACCTCGTTCCGGCAAGGAAACGGAGGCACCAATTCCAATTTGATATTATCGCTTG
15 CCAAGCTAGAGCTGATCTTTGGGAAACCAACTGCCAGACAGTGGACTGTGATGGAGTGCCCCGAGTGGTGG
GCCTCTTCGATTTCGTTAGTCATTACTAACGTGAACCCTCAGTGAAGGGACCATCAGACCAGAAAGACCAGA
TCTCCTCCTCGACACCGAGAGAGTGTTCGGGCAGTAGGACGACAAG

SEQ ID NO: 12

Yeast secretion signal

MLLQAFLLLAGFAAKISAS

SEQ ID NO: 13

Higher plants secretion signal

MANKSLLLLLLLLGLASG

SEQ ID NO: 14

Consensus eukaryotic secretion signal

MARLPLAALG

SEQ ID NO: 15

Combination higher plant/eukaryotic secretion signal

MANKLLLLLLLLLLLLPLAASG

SEQ ID NO: 16

Codon-optimized yeast sucrose invertase

ATGCTGCTGCAGGCCTTCTGTCTCTGCTGGCCGGCTTCGCCCAAGATCAGCGCTCCATGACGAACGAGACGTCCGACCGCCCCCTG
GTGCACTTACCCCCAACAAGGCTGGATGAACGACCCCAACGGCCTGTGGTACGACGAGAAGGACGCCAAGTGGCACCTGTACTTCCAGT
ACAACCCGAACGACACCGCTCTGGGGGACGCCCTTGTCTGGGGCCACGCCACGTCCGACGACCTGACCAACTGGGAGGACAGCCCATCGC
40 CATCGCCCCGAAGCGCAACGACTCCGGCGCCTTCTCCGGCTCCATGGTGGGACTACAACAACCTCCGGCTTCTTCAACGACACCATC
GACCCGCGCCAGCGCTGCGTGGCCATCTGGAGCTTACAACACCCCGAGTCCGAGGAGCAGTACATCTCTACAGCCTGGACGGCGGCTACA
CCTTCACCGAGTACCAGAAGAACCCCGTGTGGCCGCCAACTCCACCCAGTTCGCGGACCCGAAGGTCTTCTGGTACGAGCCCTCCAGAA
GTGGATCATGACCGCGGCCAAGTCCCAGGACTACAAGATCGAGATCTACTCTCCGACGACCTGAAGTCTTGAAGCTGGAGTCCGCGTTC
GCCAACGAGGGCTTCTTCGGCTACCAAGTACGAGTGCCTCCGGCTGATCGAGGTCCCCACCGAGCAGGACCCAGCAAGTCTTACTGGGTGA
45 TGTTTCATCTCCATCAACCCCGCGCCCCGGCGCGGCTCCTTCAACCAAGTACTTCGTCGGCAGCTTCAACGGCACCCACTTCGAGGCCTT
CGACAACCAAGTCCCGGTGGGACTTCGGCAAGCACTACGCCCTCGAGACCTTCTTCAACACCGACCCCTACGGGAGCGCCCTG
GGCATCGCGTGGGCTCCAACCTGGGAGTACTCGGCCTTCGTGCCACCAACCCCTGGCGCTCCTCCATGTCCCTCGTGCAGAACTTCTCCC
TCAACACCGAGTACCAGGCCAACCCGGAGACGGAGCTGATCAACCTGAAGGCCGAGCCGATCCTGAACATCAGCAACGCGGCCCTGGAG
CCGGTTCGCCACCAACACCGTTGACGAAGGCCAACAGCTACAACGTCGACCTGTCCAACAGCACCGGCACCTGGAGTTCGAGCTGGTG
50 TACGCCGTCAACACCCAGACGATCTCCAAGTCCGTGTCGCGGACCTCTCCCTCTGGTTCAAGGCCTGGAGGACCCCGAGGATACC
TCCGCATGGGCTTCGAGGTGTCGCGCTCCTCTTCTTCTGGACCGCGGGAACAGCAAGGTGAAGTTCGTGAAGGAGAACCCCTACTTCAC
CAACCGCATGAGCGTGAACAACAGCCCTTCAAGAGCGAGAACGACCTGCTCTACTACAAGGTGTACGCTTGTGGACGAGAACATCTCTG
GAGCTGTACTTCAACGACGGCGACGTCGTGTCCACCAACCTACTTTCATGACCACCGGAACGCCCTGGGCTCCGTGAACATGACGACGG
GGGTGGACAACCTGTCTACATCGACAAGTTCAGGTGCGCGAGGTCAAGTGA

SEQ ID NO: 17

Yeast sucrose invertase, GenBank accession no. NP_012104.1

MTNETSDRPLVHFTPNKGWMNDPNGLWYDEKDAKWHLYFQYNPNDTVWGTPLFWGHATSDDLNTWEDQPIAI
APKRNDSGAFSGSMVVDYNNTSGFFNDTIDPRQRCVAIWYNTPESEEQYISYSLDGGYTFTEYQKNPVLAA
60 NSTQFRDPKVFWEPSQKWIMTAAKSQDYKIEIYSSDDLKSWKLESAFANEGFLGYQYECPLIEVPTEQDP
SKSYWVMFISINPGAPAGGSFNQYFVGSFNGTHFEAFDNQSRVVDGKDYALQTFNTDPTYGSALGIAWA
SNWEYSAFVPTNPWRSSMSLVKFSNLNEYQANPETELINLKAEPILNISNAGPWSRFATNTTLTKANSYNV
DLNSTGTLEFELVYAVNTTQTISKSVFADLSLWFKGLEDPPEEYLRMGFEVSASSFFLDRGNSKVKFVKENP
YFTNRMSVNNQPFKSENDLSYYKVYGLLDQNILELYFNDGDVVSNTYFMTTGNALGSVNMTTGVDNLFYID
65 KFQVREVK

SEQ ID NO: 18

5' donor DNA sequence of *Prototheca moriformis* delta 12 FAD knockout homologous recombination targeting construct

GCTCTTCGGGTTTGCTCACCCGCGAGGTGCGACGCCCAGCATGGCTATCAAGACGAACAGGCAGCCTGTGGAG
 AAGCCTCCGTTACGATCGGGACGCTGCGCAAGGCCATCCCCGCGCACTGTTTCGAGCGCTCGGCGCTTCGT
 AGCAGCATGTACCTGGCCTTTGACATCGCGGTGATGTCCCTGCTCTACGTGCGGTGACGTACATCGACCT
 GCGCCGGTGCCCTACGTGGGTCAAGTATGGCGTCATGTGGCCGCTCTACTGGTTCTTCCAGGTGTGTGTGAGG
 GTTGTGGTTGCCGTATCGAGGTCTTGGTGGCGCGCATGGGGGAGAAGGCGCCTGTCCCGCTGACCCCCCG
 GCTACCCCTCCCGGCACCTTCCAGGGCGCCTTCGGCACGGGTGTCTGGGTGTGCGCGCACGAGTGCGGCCACC
 AGGCCTTTTCTCCAGCCAGGCCATCAACGACGGCGTGGGCCTGGTGTTCACAGCCTGCTGCTGGTGCCCT
 ACTACTCCTGGAAGCACTCGCACCGCCGCCACCACTCCAACACGGGGTGCCTGGACAAGGACGAGGTGTTTG
 TGCCGCGCACCGCGCAGTGGCGCACGAGGGCCTGGAGTGGGAGGAGTGGCTGCCCATCCGCATGGGCAAGG
 TGCTGGTCAACCTGACCTGGGCTGGCCGCTGTACCTCATGTTCAACGTCGCCTCGCGGCCGTACCCGCGCT
 TCGCCAACCACTTTGACCCGTGGTGCCTCATCTTACGCAAGCGCGAGGTACCCCTTTCTTGCCTATGACACT
 TCCAGCAAAAGGTAGGGCGGGCTGCGAGACGGCTTCCCGGCGCTGCATGCAACACCGATGATGCTTCGACCC
 CCCGAAGCTCCTTCGGGGCTGCATGGGCGCTCCGATGCCGCTCCAGGGCGAGCGCTGTTTAAATAGCCAGGC
 CCCCATTGCAAAGACATTATAGCGAGCTACCAAAGCCATATTCAAACACCTAGATCACTACCACTTCTACA
 CAGGCCACTCGAGCTTGTGATCGCACTCCGCTAAGGGGCGCCTCTTCTCTTCGTTTCAGTCACAACCCGC
 AAACGGCGCGCC

SEQ ID NO: 19

3' donor DNA sequence of *Prototheca moriformis* delta 12 FAD knockout homologous recombination targeting construct

CAATTGGCAGCAGCAGCTCGGATAGTATCGACACACTCTGGACGCTGGTTCGTGTGATGGACTGTTGCCGCCA
 CACTTGCTGCCTTGACCTGTGAATATCCCTGCCGCTTTTATCAAACAGCCTCAGTGTGTTTGATCTTGTGTG
 TACGCGCTTTTTCGAGATTGCTAGCTGCTTGTGCTATTTGCGAATACCACCCCGAGCATCCCCTTCCCTCGTT
 TCATATCGCTTGCATCCCAACCGCAACTTATCTACGCTGTCTGCTATCCCTCAGCGCTGCTCCTGCTCCTG
 CTCACTGCCCCCTCGCACAGCCTTGGTTTGGGCTCCGCCTGTATTCTCCTGGTACTGCAACCTGTAAACAGC
 ACTGCAATGCTGATGCACGGGAAGTAGTGGGATGGGAACACAATGGAGCATCGAGGTGGTCATCTCCGACC
 TCGCGTTGGTGGCGGTGCTCAGCGGGCTCAGCGTGTGGGCCGACCATGGGCTGGGCTGGTCAAGA
 CCTACGTGGTGGCCTACATGATCGTGAACATGTGGCTGGTGTCTCATCAGCTGCTCCAGCACACGCACCCGG
 CCCTGCCGCACTACTTCGAGAAGGACTGGGACTGGCTACGCGGCGCCATGGCCACCGTCGACCGCTCCATGG
 GCCCGCCCTTCATGGACAGCATCCTGCACCACATCTCCGACACCCACGTGCTGCACCACCTCTTCAGCACCA
 TCCCGCACTACCACGCCGAGGAGGCTCCGCCGCCATCCGGCCCATCCTGGGCAAGTACTACCAATCCGACA
 GCCGCTGGGTGCGCCGCGCCCTGTGGGAGGACTGGCGCACTGCCGCTACGTCGTCCCCGACGCGCCCGAGG
 ACGACTCCGCGCTCTGGTTCCACAAGTGAGCGCGCCTGCGCGAGGACGCAACAACGCTGCCGCCGTGTCT
 TTTGCACGCGGACTCCGGCGCTTCGCTGGTGGCACCCCCATAAAGAAACCTCAATTCTGTTTGTGGAAGA
 CACGCTGTACCCCCACCCACCTGCACCTCTATTATTGGTATTATTGACGCGGGAGTGGGCGTTGTACC
 CTACAACGTAGCTTCTCTAGTTTTCAGCTGGCTCCACCATTGTAAAGAGCCTCTAGAGTGCACCTGCAGGC
 ATGCAAGCTTGGCGTAATCATGGTTCATAGCTGTTTCTGTGTGAAATTGTTATCCGCTCACAATTCCACACA
 ACATACGAGCCGGAAGCATAAAGTGTAAAGCTGGGCTCCGCTAATGAGTGAGCTAACTCACATTAATTGCGT
 TGCGCTCACTGCCCGCTTTCAGTTCGGGAAACCTGTGCTGCCAGCTGCATTAATGAATCGGCCAACGCGCGG
 GGAGAGGCGGTTTTCGCTATTGGGCGCTCTTCC

SEQ ID NO: 20

Prototheca moriformis delta 12 FAD knockout homologous recombination targeting construct

GCTCTTCGGGTTTGCTCACCCGCGAGGTGCGACGCCCAGCATGGCTATCAAGACGAACAGGCAGCCTGTGGAG
 AAGCCTCCGTTACGATCGGGACGCTGCGCAAGGCCATCCCCGCGCACTGTTTCGAGCGCTCGGCGCTTCGT
 AGCAGCATGTACCTGGCCTTTGACATCGCGGTGATGTCCCTGCTCTACGTGCGGTGACGTACATCGACCT
 GCGCCGGTGCCCTACGTGGGTCAAGTATGGCGTCATGTGGCCGCTACTACTGGTTCTTCCAGGTGTGTGTGAGG
 GTTGTGGTTGCCGTATCGAGGTCTTGGTGGCGCATGCGGGGAGAAGGCGCTGTCCCGCTGACCCCGCG
 GCTACCCCTCCCGGCACCTTCCAGGGCGCCTTCGGCACGGGTGTCTGGGTGTGCGCGCACGAGTGCGGCCACC
 AGGCCTTTTCTCCAGCCAGGCCATCAACGACGGCGTGGGCCTGGTGTTCACAGCCTGCTGCTGGTGCCCT
 ACTACTCCTGGAAGCACTCGCACCGCCGCCACCACTCCAACACGGGGTGCCTGGACAAGGACGAGGTGTTTG
 TGCCGCGCACCGCGCAGTGGCGCACGAGGGCCTGGAGTGGGAGGAGTGGCTGCCCATCCGCATGGGCAAGG
 TGCTGGTCAACCTGACCTGGGCTGGCGCTGTACCTCATGTTCAACGTGCGCTCGCGGCGTACCCGCGCT
 TCGCCAACCACTTTGACCCGTGGTGCCTCATCTTCAGCAAGCGCGAGGTACCCTTTCTTCCGCTATGACACT
 TCCAGCAAAAGGTAGGGCGGGCTGCGAGACGGCTTCCCGGCGCTGCATGCAACACCGATGATGCTTCGACCC
 CCCGAAGCTCCTTCGGGGCTGCATGGGCGCTCCGATGCCGCTCCAGGGCGAGCGCTGTTTAAATAGCCAGGC
 CCCCATTGCAAAGACATTATAGCGAGCTACCAAAGCCATATTCAAACACCTAGATCACTACCACTTCTACA
 CAGGCCACTCGAGCTTGTGATCGCACTCCGCTAAGGGGCGCCTCTTCTCTTCGTTTCAGTCACAACCCGC
 AAACGGCGCGCCATGCTGCTGCAGGCCCTTCTGTTCTGCTGGCCGGCTTCGCCGCGCAAGATCAGCGCCTCC
 ATGACGAACGAGACGTCCGACCGCCCTGGTGCATCTCACCCCAACAAGGGCTGGATGAACGACCCCAAC
 GGCTGTGGTACGACGAGAAGGACGCCAAGTGGCACCTGTACTTCCAGTACAACCCGAACGACACCGTCTGG
 GGGACGCCCTTGTCTGGGGCCACGCCACGTCCGACGACCTGACCAACTGGGAGGACAGCCCATCGCCATC
 GCCCCGAAGCGCAACGACTCCGGCGCCTTCTCCGGCTCCATGGTGGTGGACTACAACAACACCTCCGGCTTC
 TTCAACGACACCATCGACCCGCGCCAGCGCTGCGTGGCCATCTGGACCTACAACACCCCGGAGTCCGAGGAG
 CAGTACATCTCTACAGCCTGGACGGCGGCTACACCTTACCGAGTACCAGAAGAACCCCGTGTGGCCGCC

AACTCCACCCAGTTCCGCGACCCGAAGGTCTTCTGGTACGAGCCCTCCCAGAAGTGGATCATGACCGCGGCC
 AAGTCCCAGGACTACAAGATCGAGATCTACTCTCCGACGACCTGAAGTCTTGAAGCTGGAGTCCGCGTTT
 GCCAACGAGGGCTTCCCTCGGCTACCAGTACGAGTGGCCCGGCTGATCGAGGTCCCACCGAGCAGGACCCC
 AGCAAGTCCCTACTGGGTGATGTTTCATCTCCATCAACCCCGGCGCCCCGGCCGGCGGCTCCTTCAACACAGTAC
 5 TTCGTCCGCGAGCTTCAACGGCACCCACTTCGAGGCCTTCGACAACCAGTCCCGCGTGGTGGACTTCGGCAAG
 GACTACTACGCCCTGCAGACCTTCTTCAACACCGACCCGACCTACGGGAGCGCCCTGGGCGATCGCGTGGGCC
 TCCAACTGGGAGTACTCCGCTTTCGTGCCCCACCAACCCCTGGCGCTCCTCCATGTCCCTCGTGCAGTTC
 TCCCTCAACACCGAGTACCAGGCCAACCCGGAGACGGAGCTGATCAACCTGAAGGCCGAGCCGATCCTGAAC
 10 ATCAGCAACGCCGGCCCCCTGGAGCCGGTTCGCCACCAACACCACGTTGACGAAGGCCAACAGCTACAACGTC
 GACCTGTCCAACAGCACCGGCACCCTGGAGTTCGAGCTGGTGTACGCCGTCAACACCACCCAGACGATCTCC
 AAGTCCGTGTTTCGCGGACCTCTCCCTCTGGTTCAAGGGCCTGGAGGACCCCGAGGAGTACCTCCGCATGGGC
 TTCGAGGTGTCCGCTCCCTCTTCTTCCCTGGACCGCGGGAACAGCAAGGTGAAGTTTCGTGAAGGAGAACCCC
 TACTTCACCAACCGCATGAGCGTGAACAACACGACCTTCAAGAGCGAGAACGACCTGTCTACTACAAGGTG
 TACGGCTTGCTGGACCAGAACATCCTGGAGCTGTACTTCAACGACGGCGACGTCTGTCTCCACCAACACCTAC
 15 TTCATGACCACCGGAACGCCCTGGGCTCCGTGAACATGACGACGGGGGTGGACAACCTGTTTCTACATCGAC
 AAGTTCACAGGTGCGCGAGGTCAAGTGACAATTGGCAGCAGCAGCTCGGATAGTATCGACACACTCTGGACGC
 TGGTCTGTGATGGACTGTTGCCGCCACACTTGTCTGCCTTGACCTGTGAATATCCCTGCCGCTTTTATCAAA
 CAGCCTCAGTGTGTTTGTGTTGTGTGTACGCGCTTTTTCGAGTTGCTAGCTGCTTGTGCTATTTGCGAATA
 CCACCCCGAGCATCCCCCTTCCCTCGTTTTCATATCGCTTGATCCCAACCGCAACTTATCTACGCTGTCTGCTG
 20 TATCCCTCAGCGCTGCTCCTGCTCCTGCTCACTGCCCTCGCACAGCCTTGGTTTGGGCTCCGCTGTATTC
 TCCTGGTACTGCAACCTGTAAACAGCACTGCAATGCTGATGCACGGGAAGTAGTGGGATGGGAACACAAAT
 GGAGCATCGAGGTGGTTCATCTCCGACCTCGCGTTGGTGGCGGTGCTCAGCGGGCTCAGCGTGTGGGCCGCA
 CCATGGGCTGGGCTGGCTGGTCAAGACCTACGTGGTCCCTACATGATCGTGAACATGTGGCTGGTGTCTCA
 TCACGCTCAGTCCAGCACACGACCCCGCCCTGCCGCACTACTTCGAGAAGGACTGGGACTACCGCGCG
 25 CCATGGCCACCGTTCGACCGCTCCATGGGCCCGCCCTTTCATGGACAGCATCCTGCACCACATCTCCGACACCC
 ACGTGTCTGCACCACCTCTTCAGCACCATCCCGCACTACCACGCCGAGGAGGCCTCCGCCGCCATCCGGCCCA
 TCCTGGGCAAGTACTACCAATCCGACAGCCGCTGGGTTCGGCCGCGCCCTGTGGGAGGACTGGCGCGACTGCC
 GCTACGTCTGTCCTCCCGACGCGCCCGAGGACGACTCCGCGCTCTGGTTCCACAAGTGAGCGCGCCTGCGCGAGG
 ACGCAGAACACGCTGCCGCCGTGTCTTTTGCACGCGGACTCCGGCGCTTCGCTGGTGGCACCCCCATAAA
 30 GAAACCTCAATTCTGTTTGTGGAAGACACGGTGTACCCCAACCCACCCACCTGCACCTTATTATTGGTAT
 TATTGACGCGGGAGTGGGCGTTGTACCTACAACGTAGCTTCTCTAGTTTTTCAGCTGGCTCCACCATTTGTA
 AAGAGCCTCTAGAGTCGACCTGCAGGCATGCAAGCTTGGCGTAATCATGGTTCATAGCTGTTTCTGTGTGAA
 ATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAGCCTGGGGTGCCTAAT
 GAGTGAGCTAATCTACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCTGTGCCAGC
 35 TGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTTCGTATTGGGCGCTCTTCC

SEQ ID NO: 21

5' donor DNA sequence of *Prototheca moriformis* SAD2A knockout homologous recombination targeting construct

GCTCTTCCGCTTGGAGCTGGTGCAGAGCATGGGTTCAGTTTTCGCGAGGAGAGGGTGCTCCCCGTGCTGCACCC
 CGTGGACAAGCTGTGGCAGCCGACGACTTCTGCCCCGACCCGAGTCCCGGACTTCGAGGACCAGGTGGC
 GGAGCTGCGCGCGCGGCCAAGGACCTGCCGACGAGTACTTTGTGGTGTGCTGGTGGGCGACATGATCACGGA
 45 GGAGGCGCTGCCGACCTACATGGCCATGCTCAACACCTTGGACGGTGTGCGCGACGACACGGGCGCGGCTGA
 CCACCCGTGGGCGCGCTGGACGCGGCAGTGGGTGGCCGAGGAGAACCAGGACGCGGACCTGCTGAACAAGTA
 CTGTTGGCTGACGGGGCGCGTCAACATGCGGGCCGTGGAGGTGACCATCAACAACCTGATCAAGAGCGGCAT
 GAACCCGACGACGACAAACAACCTTACTTGGGCTTCGTCTACACCTCCTTCCAGGAGCGCGCCACCAAGTA
 GTACC

SEQ ID NO: 22

3' donor DNA sequence of *Prototheca moriformis* SAD2A knockout homologous recombination targeting construct

CAATTGGCAGCAGCAGCTCGGATAGTATCGACACACTCTGGACGCTGGTCTGTGATGGACTGTTGCCGCCA
 55 CACTTGTCTGCCCTTGACCTGTGAATATCCCTGCCGCTTTTATCAAACAGCCTCAGTGTGTTTGTGTTG
 TACGCGCTTTTTCGAGTTGCTAGCTGCTTGTGCTATTTGCGAATAACACCCCGAGCATCCCCCTTCCCTCGTT
 TCAATATCGCTTGCATCCCAACCGCAACTTATCTACGCTGTCTGCTATCCCTCAGCGCTGCTCCTGCTCCTG
 CTCACTGCCCCCTCGCACAGCCTTGGTTTGGGCTCCGCCTGTATTCTCCTGGTACTGCAACCTGTAAACAGC
 ACTGCAATGCTGATGCACGGGAAGTAGTGGGATGGGAACACAAATGGAAGGATCGTAGAGCTCCAGCCACGG
 60 CAACCCGCGCGCCTGGCGGCGAGCACGGCGACAAGGGCCTGAGCAAGATCTGCGGGCTGATCGCCAGCGA
 CGAGGGCCCGGCACGAGATCGCCTACACGCGCATCGTGGACGAGTTCTTCCGCTCGACCCCGAGGGCGCCGT
 CGCCGCCCTACGCCAACATGATGCGCAAGCAGATCACCATGCCGCGCACCTCATGGACGACATGGGCCACGG
 CGAGGCCAACCCGGGCGCAACCTCTTCGCCGACTTCTCCGCCGTGCGCGAGAAGATCGACGTCTACGACGC
 CGAGGACTACTGCCGATCCTGGAGCACCTCAACGCGCGCTGGAAGGTGGACGAGCGCCAGGTGAGCGGCCA
 65 GGCCCGCGGACAGGAGTACGTTCTGGGCTGCCCCAGCGCTTCCGGAACCTCGCCGAGAAGACCGCCGC
 CAAGCGCAAGCGCTGCGCGCAGGCGCTGCGCTTCTCCTGGAGAGAAGAGCCTCTAGAGTCGACCTGACG
 GCATGCAAGCTTGGCGTAATCATGGTTCATAGCTGTTTCTGTGTGAAATTGTTATCCGCTCACAATTCCACA
 CAACATACGAGCCGGAAGCATAAAGTGTAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGC
 70 GTTGCCTCACTGCCCGCTTTCAGTCGGGAAACCTGTCTGTGCCAGCTGCATTAATGAATCGGCCAACGCGC
 GGGGAGAGGCGGTTTTCGTATTGGGCGCTCTTCC

SEQ ID NO: 23

Prototheca moriformis SAD2A knockout homologous recombination targeting construct

5 GCTCTTCCGCGCTGGAGCTGGTGCAGAGCATGGGTGAGTTTGCAGGAGGAGAGGGTGCTCCCCGTGCTGCACCC
CGTGGACAAGCTGTGGCAGCCGAGGACTTCCTGCCCCGACCCGAGTCGCCCCGACTTCGAGGACCAGGTGGC
GGAGCTGCGCGCGCGCGCCAAGGACCTGCCCCGACGAGTACTTTGTGGTGCTGGTGGGCGACATGATCACGGA
GGAGGCGCTGCCGACCTACATGGCCATGCTCAACACCTTGGACGGTGTGCGCGACGACACGGGCGCGGCTGA
10 CCACCCGTGGGCGCGCTGGACGCGGCAGTGGGTGGCCGAGGAGAACCGGCACGGCGACCTGCTGAACAAGTA
CTGTTGGCTGACGGGGCGCGTCAACATGCGGGGCGTGGAGGTGACCATCAACAACCTGATCAAGAGCGGCAT
GAACCCGACAGCGACAACAACCCCTTACTTGGGCTTCGTCTACACCTCCTTCCAGGAGCGCGCCACCAAGTA
GGTACCCTTTCTTGGCTATGACACTTCCAGCAAAAGGTAGGGCGGGCTGCGAGACGGCTTCCCGGCGCTGC
ATGCAACACCGATGATGCTTCGACCCCCCGAAGCTCCTTCGGGGCTGCATGGGCGCTCCGATGCCGCTCCAG
15 GGCGAGCGCTGTTTAAATAGCCAGGCCCCGATTGCAAAGACATTATAGCGAGCTACCAAAGCCATATTCAA
ACACCTAGATCACTACCCTTCTACACAGGCCACTCGAGCTTGTGATCGCACTCCGCTAAGGGGGCGCCTCT
TCCTCTTCGTTTCAGTCACAACCCGCAAACGGCGCGCCATGCTGCTGCAGGCCTTCCTGTTCTGCTGGCCG
GCTTCGCGGCCAAGATCAGCGCCTCCATGACGAACGAGACGTCCGACCGCCCCCTGGTGCACCTTACCCCCA
ACAAAGGCTGGATGAACGACCCCAACGGCCTGTGGTACGACGAGAAGGACGCCAAGTGGGACCTGTACTTCC
20 AGTACAACCCGAACGACACCGTCTGGGGGACGCCCTTGTCTGGGGCCACGCCACGTCCGACGACCTGACCA
ACTGGGAGGACCAGCCCATCGCCATCGCCCCGAAGCGCAACGACTCCGGCGCCTTCTCCGGCTCCATGGTGG
TGGACTACAACAACACCTCCGGCTTCTTCAACGACACCATCGACCCGCGCCAGCGCTGCGTGGCCATCTGGA
CCTACAACACCCCGGAGTCCGAGGAGCAGTACATCTCCTACAGCCTGGACGGCGGCTACACCTTACCGAGT
ACCAGAAGAACCCCGTGTGGCCGCCAATCCACCCAGTTCGCGACCCGAAGGTCTTCTGGTACGAGCCCT
CCGAGAAGTGGATCATGACCGCGGCCAAGTCCAGGACTACAAGATCGAGATCTACTCTCCGACGACCTGA
25 AGTCCTGGAAGCTGGAGTCCGCGTTCGCCAACGAGGGCTTCCTCGGCTACCAGTACGAGTGCCCCGGCCTGA
TCGAGGTCCCCACCGAGCAGGACCCAGCAAGTCTACTGGGTGATGTTTATCTCCATCAACCCCGGCGCCC
CGGCCGGCGGCTCCTTCAACAGTACTTTCGTGCGCAGCTTCAACGGCACCCACTTCGAGGCCTTCGACAACC
AGTCCCGCGTGGTGGACTTCGGCAAGGACTACTACGCCCTGCAGACCTTCTTCAACACCGACCCGACCTACG
GGAGCGCCCTGGGCATCGCGTGGGCCTCCAACCTGGGAGTACTCCGCCTTCGTGCCCCACCAACCCCTGGCGCT
30 CCTCCATGTCCTTCGTGCGCAAGTTCTCCCTCAACACCGAGTACCAGGCCAACCCGAGACGGATGATCA
ACCTGAAGGCCGAGCCGATCCTGAACATCAGCAACGCCGGCCCCCTGGAGCCGGTTCGCCACCAACACCAAGT
TGACGAAGGCCAACAGCTACAACGTGACCTGTCCAACAGCACCGGCACCCCTGGAGTTCGAGCTGGTGTACG
CCGTCAACACCACCCAGACGATCTCCAAGTCCGTGTTTCGCGGACCTCTCCCTCTGTTTCAAGGGCCTGGAGG
ACCCCGAGGAGTACCTCCGCATGGGCTTCGAGGTGTCCGCGTCTCCTTCTTCTTCTGGACCGCGGGAACAGCA
35 AGGTGAAGTTTCGTGAAGGAGAACCCCTACTTCAACACCGCATGAGCGTGAACAACAGCCCTTCAAGAGCG
AGAACGACCTGTCTTACTACAAGGTGTACGGCTTGTGAGGACGAGAATCCTGGAGCTGTACTTCAACGACG
GCGACGTGCTGTCCACCAACACCTACTTTCATGACACCGGGAACGCCCTGGGCTCCGTGAACATGACGACGG
GGGTGGACAACCTGTTTCTACATCGACAAGTTCAGGTGCGCGAGGTCAAGTGACAATTGGCAGCAGCAGCTC
GGATAGTATCGACACACTCTGGACGCTGGTGTGATGAGTGTGCGGCCACACTTGCTGCCTTGACCTG
40 TGAATATCCCTGCCGCTTTTATCAAAACAGCCTCAGTGTGTTTATGATCTTGTGTGTACGCGCTTTTTCGAGTTG
CTAGCTGCTTGTGCTATTTTGGCAATACCAACCCCGACATCCCTTCCCTCGTTCATATCGCTTGCATCCCA
ACCGCAACTTATCTACGCTGTCTGCTATCCCTCAGCGCTGCTCCTGCTCCTGCTCACTGCCCCCTCGCACAG
CCTTGGTTTGGGCTCCGCTGTATTTCTCCTGGTACTGCAACCTGTAAACAGCACTGCAATGCTGATGCACG
45 GGAAGTAGTGGGATGGGAACACAAATGGAAGGATCGTAGAGCTCCAGCCACGGCAACACCGCGCGCCTGGCG
GCCGAGCACGGCGACAAGGGCCTGAGCAAGATCTGCGGGCTGATCGCCAGCGACGAGGGCCGGCACGAGATC
GCCTACACGCGCATCGTGGACGAGTTCCTCCGCCTCGACCCCGAGGGCGCGCTCGCCGCCTACGCCAACATG
ATGCGCAAGCAGATCACCATGCCCGCGCACCTCATGGACGACATGGGCCACGGCGAGGCCAACCCGGGCGCG
AACCTCTTCGCGGACTTCTCCGCGCTCGCCGAGAAGATCGACGTCTACGACGCGGAGGACTACTGCCGATC
CTGGAGCACCTCAACGCGCGCTGGAAGGTGGACGAGCGCCAGGTGAGCGGCCAGGCCCGCGGACAGGAG
50 TACGTTCTGGGCTGCCCCAGCGCTTCCGGAACTCGCCGAGAAGACCGCCGCAAGCGCAAGCGCTGCGG
CGCAGGCCCGTTCGCTTCTCCTGGAGAGAAGAGCCTCTAGAGTCGACCTGCAGGCATGCAAGCTTGGCGTAA
TCATGGTCATAGCTGTTTCTGTGTGAAATTTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGC
ATAAAGTGTAAGCCTGGGGTGCCTAATGAGTGAGCTAATCACATTAATTGCGTTGCGCTCACTGCCCGCT
TTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAAATCGGCCAACGCGCGGGGAGAGGCGGTTTTCGT
55 ATTGGGCGCTCTTCC

SEQ ID NO: 24

5' donor DNA sequence of *Prototheca moriformis* SAD2B knockout homologous recombination targeting construct

60 GCTCTTCCCGCCTGGAGCTGGTGCAGAGCATGGGGCAGTTTGCAGGAGGAGAGGGTGCTCCCCGTGCTGCACC
CCGTGGACAAGCTGTGGCAGCCGAGGACTTCCTGCCCCGACCCGAGTCGCCCCGACTTCGAGGACCAGGTGG
CGGAGCTGCGCGCGCGCGCCAAGGACCTGCCCCGACGAGTACTTTGTGGTGCTGGTGGGCGACATGATCACGG
AGGAGGCGCTGCCGACCTACATGGCCATGCTCAACACCTTGGACGGTGTGCGCGACGACACGGGCGCGGCTG
65 ACCACCCGTGGGCGCGCTGGACGCGGCAGTGGGTGGCCGAGGAGAACCGGCACGGCGACCTGCTGAACAAGT
ACTGTTGGCTACGGGGCGCGTCAACATGCGGGCGTGGAGGTGACCATCAACAACCTGATCAAGAGCGGCA
TGAACCCGACAGCGACAACAACCCCTTACTTGGGCTTCGTCTACACCTCCTTCCAGGAGCGCGCCACCAAGT
AGGTACC

70 SEQ ID NO: 25

3' donor DNA sequence of *Prototheca moriformis* SAD2B knockout homologous recombination targeting construct

CAGCCACGGCAACACCGCGCGCCTTGCGGGCCGAGCACGGCGACAAGAACCTGAGCAAGATCTGCGGGCTGAT
CGCCAGCGACGAGGGCCGGCACGAGATCGCCTACACGCGCATCGTGGACGAGTTCTTCCGCCTCGACCCCGA
GGGCGCGTTCGCCCTACGCCAACATGATGCGCAAGCAGATCACCATGCCCGCGCACCTCATGGACGACAT
GGGCCACGGCGAGGCCAACCCGGGGCCGCAACCTCTTCGCCGACTTCTCCGCGGTGCGCGAGAAGATCGACGT
CTACGACGCCGAGGACTACTGCCGCATCCTGGAGCACCTCAACGCGCGCTGGAAGGTGGACGAGCGCCAGGT
CAGCGGCCAGGCCGCCGCGGACCAGGAGTACGTCTGGGGCTGCCCCAGCGCTTCCGGAAACTCGCCGAGAA
GACCGCCGCCAAGCGCAAGCGCGTCGCGCGCAGGCCCGTTCGCTTCTCCTGGAGAAGAGCCTCTAGAGTCGA
CCTGCAGGCATGCAAGCTTGGCGTAATCATGGTCATAGCTGTTTCTGTGTGAAATTGTTATCCGCTCACAA
TTCACACAAACATACGAGCCGGAAGCATAAAGTTGAAAGCCTGGGGTGCCCTAATGAGTGAGCTAACTCACAT
TAATTGCGTTGCGTCTACTGCCCGCTTTCCAGTCGGGAAACCTGTCTGTGCCAGCTGCATTAATGAATCGGCC
AACGCGCGGGGAGAGGCGGTTTGCCTATTGGGCGCTCTTCC

SEQ ID NO: 26

Prototheca moriformis SAD2B knockout homologous recombination targeting construct

GCTCTTCCCGCCTGGAGCTGGTGCAGAGCATGGGGCAGTTTGGCGAGGAGAGGGTGCTCCCCGTGCTGCACC
CCGTGGACAAGCTGTGGCAGCCGAGGACTTCTGCCCCGACCCGAGTCGCCCCGACTTTCGAGGACCAGGTGG
CGGAGCTGCGCGCGCGCGCCAAAGGACCTGCCCGACGAGTACTTTGTGGTGCTGGTGGGGCGACATGATCACGG
AGGAGGCGCTGCCGACCTACATGGCCATGCTCAACACCTTGGACGGTGTGCGCGACGACACGGGGCGCGGTG
ACCACCCGTGGGCGCGCTGGACGCGGCAGTGGGTGGCCGAGGAGAACCAGGACGCGGACCTGTGAACAAGT
ACTGTTGGCTGACGGGGCGCGTCAACATGCGGGCCGTGGAGGTGACCATCAACAACCTGATCAAGAGCGGCA
TGAACCCGCGAGACGGACAACAACCCCTTACTTGGGCTTCGTCTACACCTCCTTCCAGGAGCGCGCCACCAAGT
AGGTACCCTTTCTTGCCTATGACACTTCCAGCAAAAGGTAGGGCGGGCTGCGAGACGGCTTCCCGGCGCTG
CATGCAACACCGATGATGCTTCGACCCCCCGAAGCTCCTTCGGGGCTGCATGGGCGCTCCGATGCCGCTCCA
GGGCGAGCGCTGTTTAAATAGCCAGGCCCCGATTGCAAAGACATTATAGCGAGCTACCAAAGCCATATTCA
AACACCTAGATCACTACCCTTCTACACAGGCCACTCGAGCTTGTGATCGCACTCCGCTAAGGGGGCGCCTC
TTCCTCTTCTGTTTCAAGTCAACACCCGCAACCGCGCGCCATGCTGCTGCAGGCCTTCTGTTCTGCTGGCC
GGCTTCGCCGCCAAGATCAGCGCCTCCATGACGAACGAGACGTCCGACCGCCCCCTGGTGCACCTTCAACCCC
AACAAGGGCTGGATGAACGACCCCAACGGCCTGTGGTACGACGAGAAGGACGCCAAGTGGCACCTGTACTTC
CAGTACAACCCGAACGACACCGTCTGGGGGACGCCCTTGTTCCTGGGGCCACGCCACGTCCGACGACCTGACC
AATGGGAGGACCAGCCCATCGCCATCGCCCCGAAGCGCAACGACTCCGGCGCCTTCTCCGGCTCCATGGTG
GTGGACTACAACAACACCTCCGGCTTCTTCAACGACACCATCGACCCGCGCCAGCGCTGCGTGGCCATCTGG
ACCTACAACACCCCGAGTCCGAGGAGCAGTACATCTCCTACAGCCTGGACGGCGGCTACACCTTCAACGAG
TACCAGAAGAACCCCGTGTGTCGGCCGCAACTCCACCCAGTTCGCGACCCGAAGGTCTTCTGGTACGAGCCC
TCCCAGAAGTGGATCATGACCGCGGCCAAGTCCCAGGACTACAAGATCGAGATCTACTCCTCCGACGACCTG
AAGTCTTGAAGCTGGAGTCCGCGTTCGCCAACGAGGGCTTCTCCTGGCTACCAGTACGAGTGCCCCGGCCTG
ATCGAGGTCCCCACCGAGCAGGACCCAGCAAGTCTTACTGGGTGATGTTTATCTCCATCAACCCCGGCGCC
CCGGCCGGCGGCTCCTTCAACAGTACTTCTCGTCGGCAGCTTCAACGGCACCCACTTCGAGGCTTCGACAAC
CAGTCCCGCGTGGTGGACTTCGGCAAGGACTACTACGCCCTGCAGACCTTCTTCAACACCGACCCGACCTAC
GGGAGCGCCCTGGGCATCGCGTGGGCCTCCAACCTGGGAGTACTCCGCCTTCGTGCCCACCAACCCCTGGCGC
TCCTCCATGTCCCTCGTGCAGCAAGTTCTCCCTCAACACCGAGTACCAGGCCAACCCGGAGACGGAGCTGATC
AACCTGAAGGCCGAGCCGATCCTGAACATCAGCAACGCCGGCCCCCTGGAGCCGGTTTCGCCACCAACACACG
TTGACGAAGGCCAACAGCTACAACGTGACCTGTCCAACAGCACCGGCACCCCTGGAGTTCGAGCTGGTGTAC
GCCGTCAACACCAACCCAGACGATCTCCAAGTCCGTGTTTCGCGGACCTCTCCCTCTGTTTCAAGGGCCTGGAG
GACCCCGAGGAGTACCTCCGATGGGCTTCGAGGTGTCCGCGTCTCCTTCTTCTGACCGCGGGAAACAGC
AAGGTGAAGTTCTGTAAGGAGAACCCCTACTTCACCAACCGCATGAGCGTGAACAACAGCCCTTCAAGAGC
GAGAACGACCTGTCTACTACAAGGTGTACGGCTTGTGAGACGAGAACATCCTGGAGCTGTACTTCAACGAC
GGCGACGTGCTGTCCACCAACACCTACTTTCATGACCACCGGGAACGCCCTGGGCTCCGTGAACATGACGACG
GGGGTGACAACCTGTTCTACATCGACAAGTTCAGGTGCGCGAGGTCAAGTGACAATTGGCAGCAGCAGCT
CGGATAGTATCGACACACTCTGGACGCTGATCTGTGATGAGTGTGCGGACACTTGTGCGCCACACTTGCTGCCTGACCT
GTGAATATCCCTGCCGCTTTTATCAACAGCCTCAGTGTGTTTGATCTTGTGTGTACGCGCTTTTGCAGATT
GCTAGCTGCTTGTGCTATTTGCGAATACCACCCCGAGCATCCCCTTCCCTCGTTTTCATATCGCTTGCATCCC
AACCGCAACTTATCTACGCTGTCTGCTATCCCTCAGCGCTGCTCCTGCTCCTGCTCACTGCCCCCTCGCACA
GCCTTGGTTTGGGCTCCGCCTGTATTCTCCTGGTACTGCAACCTGTAAACAGCACTGCAATGCTGATGCAC
GGGAAGTAGTGGGATGGGAACACAAATGGACAGCCACGGCAACACCGCGCGCCTTGGCGGCCGAGACCGGCA
CAAGAACCTGAGCAAGATCTGCGGGCTGATCGCCAGCAGCAGGAGGCGGACAGAGATCGCCTACACGCGCAT
CGTGGACGAGTTCTTCCGCCTCGACCCCGAGGGCGCGCTCGCCGCCTACGCCAACATGATGCGCAAGCAGAT
CACCATGCCCGCGCACCTCATGGACGACATGGGCCACGGCGAGGCCAACCCGGGCGCAACCTCTTCGCCGA
CTTCTCCGCGGTGCGCGAGAAGATCGACGTCTACGACGCCGAGGACTACTGCCGCATCCTGGAGCACCTCAA
CGCGCGCTGGAAGGTGGACGAGCGCCAGGTGAGCGGCCAGGCCGCCGCGGACCAGGAGTACGTCTGGGGCT
GCCCCAGCGCTTCCGGAAACTCGCCGAGAAGACCGCGCCCAAGCGCAAGCGCGTCCGCGCAGGCCCCGTGCG
CTTCTCCTGGAGAAGAGCCTCTAGAGTCGACCTGAGGCTGCAAGCTTGGCGTAATCATGCTGATGATGCTGT
TTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCT
GGGGTGCCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACC
GTCTGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCCTATTGGGCGCTCTTCC

SEQ ID NO: 27

Primer 1 5'-TCACTTCATGCCGGCGGTCC-3'

SEQ ID NO: 28

5 Primer 2 5'- GCGCTCCTGCTTGGCTCGAA-3'

SEQ ID NO: 29

S. cerevisiae suc2 cassette

CTTTCTTGCGCTATGACACTTCCAGCAAAAGGTAGGGCGGGCTGCGAGACGGCTTCCCGGCGCTGCATGCAA
 CACCGATGATGCTTCGACCCCCCGAAGCTCCTTCGGGGCTGCATGGGCGCTCCGATGCCGCTCCAGGGCGAG
 CGCTGTTTAAATAGCCAGGCCCCCGATTGCAAAGACATTATAGCGAGCTACCAAAGCCATATTCAAACACCT
 AGATCACTACCACTTCTACACAGGCCACTCGAGCTTGTGATCGCACTCCGCTAAGGGGGCGCCTCTTCTCTCT
 TCGTTTCAGTCACAACCCGCAAACGGCGCGCCATGCTGCTGCAGGCCTTCTGTTCTCTGCTGGCCGGCTTCG
 CCGCCAAGATCAGCGCCTCCATGACGAACGAGACGTCCGACCGCCCCCTGGTGCACCTTACCCCCAACAAAGG
 GCTGGATGAACGACCCCCAACGGCCTGTGGTACGACGAGAAGGACGCCAAGTGGCACCTGTACTTCCAGTACA
 ACCCGAACGACACCGTCTGGGGGACGCCCTTGTTCGGGGCCACGCCACGTCCGACGACCTGACCAACTGGG
 AGGACCGAGCCCATCGCCATCGCCCCGAAGCGCAACGACTCCGGCGCCTTCTCCGGCTCCATGGTGGTGGACT
 ACAACAACACCTCCGGCTTCTTCAACGACACCATCGACCCGCGCCAGCGCTGCGTGGCCATCTGGACCTACA
 ACACCCCGGAGTCCGAGGAGCAGTACATCTCCTACAGCCTGGACGGCGGCTACACCTTACCGAGTACCAGA
 AGAACCCCGTGTGGCCGCGCAACTCCACCCAGTTCGCGACCCGAAGGTCTTCTGGTACGAGCCCTCCCAGA
 AGTGGATCATGACCGCGGCCAAGTCCCAGGACTACAAGATCGAGATCTACTCCTCCGACGACCTGAAGTCTCT
 GGAAGTCTGGAGTCCGCGTTCGCCAACGAGGGCTTCTCGGTACCAGTACGAGTACGAGTCCCGGGCCTGATCGAGG
 TCCCCACCGAGCAGGACCCAGCAAGTCTACTGGGTGATGTTTCATCTCCATCAACCCCGGCGCCCCGGCCG
 GCGGCTCCTTCAACCAGTACTTCGTGCGCAGCTTCAACGGCACCCACTTCGAGGCCTTCGACAACCAGTCCC
 GCGTGGTGGACTTCGGCAAGGACTACTACGCCCTGCAGACCTTCTTCAACACCGACCCGACCTACGGGAGCG
 CCTGGGCATCGCGTGGGCCTCCAACCTGGGAGTACTCCGCCTTCGTGCCACCAACCCCTGGCGCTCCTCCA
 TGTCCCTCGTGCGCAAGTTCCTCCCTCAACACCGAGTACCAGGCCAACCCGGAGACGGAGCTGATCAACCTGA
 AGCCCGAGCCGATCCTGAACATCAGCAACGCCGGCCCCCTGGAGCCGGTTCGCCACCAACACACCTGTGACGA
 AGGCCAACAGCTACAACGTGACCTGTCCAACAGCACCGGCACCCCTGGAGTTCGAGCTGGTGTACGCCGTCA
 ACACACCCAGACGATCTCCAAGTCCGTGTTTCGCGGACCTCTCCCTCTGGTTCAAGGGCCTGGAGGACCCCG
 AGGAGTACCTCCGCATGGGCTTCGAGGTGTCCGCGTCTCCTTCTTCTTCTGGACCGCGGGAACAGCAAGGTGA
 AGTTCGTGAAGGAGAACCCCTACTTCAACCAACCGCATGAGCGTGAACAACAGCCCTTCAAGAGCGAGAACG
 ACCTGTCTTACTACAAGGTGTACGGCTTGTGGACCAGAACATCCTGGAGCTGTACTTCAACGACGGCGACG
 TCGTGTCCACCAACACCTACTTTCATGACCAACCGGGAACGCCCTGGGCTCCGTGAACATGACGACGGGGGTGG
 ACAACCTGTTCTACATCGACAAGTTCAGGTGCGCGAGGTCAAGTGACAATTGGCAGCAGCAGCTCGGATAG
 TATCGACACACTCTGGACGCTGGTCTGTGTGATGGACTGTTGCCGCCACACTTGTCTGCCTTGACCTGTGAATA
 TCCCTGCCGCTTTTATCAAACAGCCTCAGTGTGTTTGATCTTGTGTGTACGCGCTTTTGCAGTGTGTAGCT
 GCTTGTGCTATTTGCGAATACCACCCCCAGCATCCCTTCCCTCGTTTCATATCGCTTGCATCCCCAACCGCA
 ACTTATCTACGCTGTCTGTCTATCCCTCAGCGCTGCTCCTGCTCCTGCTCACTGCCCCCTCGCACAGCCTTGG
 TTTGGGCTCCGCGCTGATTCTCTGGTACTGCAACCTGTAAACCAGCACTGCAATGCTGATGCACGGGAAGT
 AGTGGGATGGGAACACAAATGGA

SEQ ID NO: 30

Chlamydomonas reinhardtii TUB2 promoter/5' UTR

CTTTCTTGCGCTATGACACTTCCAGCAAAAGGTAGGGCGGGCTGCGAGACGGCTTCCCGGCGCTGCATGCAA
 CACCGATGATGCTTCGACCCCCCGAAGCTCCTTCGGGGCTGCATGGGCGCTCCGATGCCGCTCCAGGGCGAG
 CGCTGTTTAAATAGCCAGGCCCCCGATTGCAAAGACATTATAGCGAGCTACCAAAGCCATATTCAAACACCT
 AGATCACTACCACTTCTACACAGGCCACTCGAGCTTGTGATCGCACTCCGCTAAGGGGGCGCCTCTTCTCTCT
 TCGTTTCAGTCACAACCCGCAAAC

SEQ ID NO: 31

Codon-optimized *suc2* gene

ATGCTGCTGCAGGCCTTCTGTTCCTGCTGGCCGGCTTCGCCGCCAAGATCAGCGCCTCCATGACGAACGAG
 ACGTCCGACCGCCCCCTGGTGCACCTTACCCCCAACAAAGGGCTGGATGAACGACCCCCAACGGCCTGTGGTAC
 GACGAGAAGGACGCCAAGTGGCACCTGTACTTCCAGTACAACCCGAACGACACCGTCTGGGGGACGCCCTTG
 TTCTGGGGCCACGCCACGTCCGACGACCTGACCAACTGGGAGGACCAGCCCATCGCCATCGCCCCGAAGCGC
 AACGACTCCGGCGCCTTCTCCGGCTCCATGGTGGTGGACTACAACAACACCTCCGGCTTCTTCAACGACACC
 ATCGACCCGCGCCAGCGCTGCGTGGCCATCTGGACCTACAACACCCCGGAGTCCGAGGAGCAGTACATCTCC
 TACAGCCTGGACGGCGGCTACACCTTACCGAGTACCAGAAGAACCCGTGCTGGCCGCCAACTCCACCCAG
 TTCCGCGACCCGAAGGTCTTCTGGTACGAGCCCTCCAGAAGTGGATCATGACCGCGGCCAAGTCCCAGGAC
 TACAAGATCGAGATCTACTCCTCCGACGACCTGAAGTCTTGAAGCTGGAGTCCGCGTTCGCCAACGAGGGC
 TTCTCTGGCTACCACTACGAGTGGCCCGGCTGATCGAGTCCCCACCGAGCAGGACCCAGCAAGTCTTAC
 TGGGTGATGTTTCATCTCCATCAACCCCGGCGCCCCGGCGGCTCCTTCAACCACTACTTCTGTCGCGAGC
 TTCAACGGCACCCACTTCGAGGCCTTCGACAACCAGTCCCGCGTGGTGGACTTCGGCAAGGACTACTACGCC
 CTGCAGACCTTCTTCAACACCGACCCGACCTACGGGAGCGCCCTGGGCATCGCGTGGGCCTCCAACCTGGGAG
 TACTCCGCGCTTCGTGCCACCAACCCCTGGCGCTCCTCCATGTCCCTCGTGCAGCAAGTCTCCCTCAACACC
 GAGTACCAGGCCAACCCGGAGACGGAGCTGATCAACCTGAAGGCCGAGCCGATCCTGAACATCAGCAACGCC
 GGCCCCTGAGCCGGTTCGCCACCAACACCACGTTGACGAAGGCCAACAGCTACAACGTGACCTGTCCAAC

AGCACCGGCACCCTGGAGTTTCGAGCTGGTGTACGCCGTCAACACCACCCAGACGATCTCCAAGTCCGTGTTT
GCGGACCTCTCCCTCTGGTTCAAGGGCCTGGAGGACCCCGAGGAGTACCTCCGCATGGGCTTCGAGGTGTCC
GCGTCTCTCTTCTTCTTGACCGCGGGAACAGCAAGGTGAAGTTTCGTGAAGGAGAACCCCTACTTCACCAAC
CGCATGAGCGTGAACAACCAGCCCTTCAAGAGCGAGAACGACCTGTCTACTACAAGGTGTACGGCTTGCTG
5 GACCAGAACATCCTGGAGCTGTACTTCAACGACGGCGACGTCTGTCCACCAACACCTACTTCATGACCACC
GGGAACGCCCTGGGCTCCGTGAACATGACGACGGGGGTGGACAACCTGTTCTACATCGACAAGTTCAGGTG
CGCGAGGTCAAGTGA

SEQ ID NO: 32

Chlorella vulgaris nitrate reductase 3'UTR

GCGAGCAGCAGCTCGGATAGTATCGACACACTCTGGACGCTGGTCTGTGATGGACTGTTGCCGCCACACTTG
CTGCCCTTGACCTGTGAATATCCCTGCCGCTTTTATCAAACAGCCTCAGTGTGTTTGATCTTGTGTGTACGCG
CTTTTGCGAGTTGCTAGCTGCTTGTGCTATTTGCGAATACCACCCCGAGCATCCCCTTCCCTCGTTTCATAT
15 CGCTTGCATCCCAACCGCAACTTATCTACGCTGTCTGTATCCCTCAGCGCTGCTCCTGCTCCTGCTCACT
GCCCCCTCGCACAGCCTTGGTTTGGGCTCCGCCTGTATTCTCCTGGTACTGCAACCTGTAAACCAGCACTGCA
ATGCTGATGCACGGGAAGTAGTGGGATGGGAACACAAATGGAGGATCC

SEQ ID NO: 33

pSZ1124 (FAD2B) 5' genomic targeting sequence

GCTCTTCGAGACGTGGTCTGAATCCTCCAGGCGGGTTTCCCCGAGAAAGAAAGGGTGCCGATTTCAAAGCAG
AGCCATGTGCCGGGCCCTGTGGCCTGTGTTGGCGCCTATGTAGTCACCCCCCTCACCCAATTGTGCGCCAGT
TTGCGCAATCCATAAACTCAAACCTGCAGCTTCTGAGCTGCGCTGTTCAAGAACACCTCTGGGGTTTGCTCA
25 CCGCGAGGTTCGACGCCCAGCATGGCTATCAAGACGAACAGGCAGCCTGTGGAGAAGCCTCCGTTACGATC
GGGACGCTGCGCAAGGCCATCCCCGCGCACTGTTTCGAGCGCTCGGCGCTTCGTAGCAGCATGTACCTGGCC
TTTGACATCGCGGTTCATGTCCCTGCTCTACGTGCGCTCGACGTACATCGACCCTGCGCCGGTGCCCTACGTGG
GTCAAGTATGGCGTCATGTGGCCGCTCTACTGGTTCCTCCAGGTGTGTGTGAGGGTTGTGGTTGCCCGTATC
GAGTCCCTGGTGGCGCGCATGGGGGAGAAGGCGCCTGTCCCGCTGACCCCCCGGGTACCCTCCCGGCACCT
30 TCCAGGGCGCCTTCGGCACGGGTGTCTGGGTGTGCGCGCACAGTGCAGGCCACAGGCCTTTTCTCCAGCC
AGGCCATCAACGACGGCGTGGGCCTGGTGTTCACAGCCTGCTGCTGGTGCCCTACTACTCCTGGAAGCACT
CGCACCGGGTACC

SEQ ID NO: 34

pSZ1124 (FAD2B) 3' genomic targeting sequence

CCGCCACCACTCCAACACGGGGTGCCTGGACAAGGACGAGGTGTTTGTGCCGCCGCACCGCGCAGTGGCGCA
CGAGGGCCTGGAGTGGGAGGAGTGGCTGCCCATCCGCATGGGCAAGGTGCTGGTCAACCTGACCCTGGGCTG
40 GCCGCTGTACCTCATGTTCAACGTGCGCTCGCGGCCGTACCCGCGCTTCGCCAACCACCTTTGACCCGTGGTC
GCCCATCTTCAGCAAGCGCGAGCGCATCGAGGTGGTTCATCTCCGACCTGGCGCTGGTGGCGGTGCTCAGCGG
GCTCAGCGTGTGGGCCCGCACCATGGGCTGGGCTGGTGGTCAAGACCTACGTGGTGCCCTACCTGATCGT
GAACATGCTGCTGCTCATCAGCTGCTCCAGCACAGCACC CGCGCTGCCGCACTACTTCGAGAAGGA
CTGGGACTGGCTGCGCGCGCCATGGCCACCGTGGACCGCTCCATGGGCCCGCCCTTCATGGACAACATCCT
45 GCACCACATCTCCGACACCCACGTGCTGCACCACCTCTTCAGCACCATCCCGCACTACCACGCGGAGGAGGC
CTCCGCGGCCATCAGGCCATCCTGGGCAAGTACTACCAGTCCGACAGCCGCTGGGTGCGCCGCGCCCTGTG
GGAGGACTGGCGGACTGCCGCTACGTCTCCGCGACGCGCCCGAGGACGACTCCGCGCTCTGGTTCCACAA
GTGAGTGAGTGAGAAGAGC

SEQ ID NO: 35

pSZ1125 (FAD2C) 5' genomic targeting sequence

GCTCTTCGAGGGGCTGGTCTGAATCCTTCAGGCGGGTGTACCCGAGAAAGAAAGGGTGCCGATTTCAAAGC
AGACCCATGTGCCGGGCCCTGTGGCCTGTGTTGGCGCCTATGTAGTCACCCCCCTCACCCAATTGTGCGCA
GTTTGCGCACTCCATAAACTCAAACAGCAGCTTCTGAGCTGCGCTGTTCAAGAACACCTCTGGGGTTTGCT
55 CACCCGCGAGGTTCGACGCCCAGCATGGCTATCAAGACGAACAGGCAGCCTGTGGAGAAGCCTCCGTTACGA
TCGGGACGCTGCGCAAGGCCATCCCCGCGCACTGTTTCGAGCGCTCGGCGCTTCGTAGCAGCATGTACCTGG
CCTTTGACATCGCGGTTCATGTCCCTGCTCTACGTGCGCTCGACGTACATCGACCCTGCACCGGTGCCTACGT
GGGTCAAGTACGGCATCATGTGGCCGCTCTACTGGTTCCTCCAGGTGTGTTTGAGGGTTTTGGTTGCCCGTA
TTGAGTCCCTGGTGGCGCGCATGGAGGAGAAGGCGCCTGTCCCGCTGACCCCCCGGGTACCCTCCCGGCAC
60 CTTCCAGGGCGCCTTCGGCACGGGTGTCTGGGTGTGCGCGCACAGTGCAGGCCACAGGCCTTTTCTCCAG
CCAGGCCATCAACGACGGCGTGGGCCTGGTGTTCACAGCCTGCTGCTGGTGCCCTACTACTCCTGGAAGCA
CTCGCACCGGGTACC

SEQ ID NO: 36

pSZ1125 (FAD2C) 3' genomic targeting sequence

CCGCCACCACTCCAACACGGGGTGCCTGGACAAGGACGAGGTGTTTGTGCCGCCGCACCGCGCAGTGGCGCA
CGAGGGCCTGGAGTGGGAGGAGTGGCTGCCCATCCGCATGGGCAAGGTGCTGGTCAACCTGACCCTGGGCTG
70 GCCGCTGTACCTCATGTTCAACGTGCGCTCGCGGCCGTACCCGCGCTTCGCCAACCACCTTTGACCCGTGGTC
GCCCATCTTCAGCAAGCGCGAGCGCATCGAGGTGGTTCATCTCCGACCTGGCGCTGGTGGCGGTGCTCAGCGG
GCTCAGCGTGTGGGCCCGCACCATGGGCTGGGCTGGTGGTCAAGACCTACGTGGTGCCCTACCTGATCGT

GAACATGTGGCTCGTGCTCATCACGCTGCTCCAGCACACGCACCCGGCGCTGCCGCACTACTTCGAGAAGGA
CTGGGACTGGCTGCGCGCGGCCATGGCCACCGTGGACCGCTCCATGGGCCCGCCCTTCATGGACAACATCCT
GCACCACATCTCCGACACCCACGTGCTGCACCACCTCTTCAGCACCATCCCGCACTACCACGCCGAGGAGGC
CTCCGCCGCCATCAGGCCCATCCTGGGCAAGTACTACCAGTCCGACAGCCGCTGGGTTCGGCCGCGCCCTGTG
GGAGGACTGGCGCGACTGCCGCTACGTCGTCCCGGACGCGCCCGAGGACGACTCCGCGCTCTGGTTCCACAA
GTGAGTGAGTGAGAAGAGC

SEQ ID NO: 37
amt03 promoter/UTR sequence

GGCCGACAGGACGCGCGTCAAAGGTGCTGGTCTGTATGCCCTGGCCGGCAGGTTCGTTGCTGCTGGTTA
GTGATTCCGCAACCTGATTTTGGCGTCTTATTTTGGCGTGGCAAACGCTGGCGCCCGCGAGCCGGGCGGC
GGCGATGCGGTGCCCCACGGCTGCCGGAATCCAAGGGAGGCAAGAGCGCCCGGGTCAGTTGAAGGGCTTTAC
GCGCAAGGTACAGCCGCTCCTGCAAGGCTGCGTGGTGAATTGGACGTGCAGGTCCTGCTGAAGTTCCTCCA
CCGCTCACCAGCGGACAAAGCACCGGTGTATCAGGTCCGTGTATCCACTCTAAAGAGCTCGACTACGACC
TACTGATGGCCCTAGATTCTTCATCAAAAACGCCTGAGACACTTGCCAGGATTGAAACTCCCTGAAGGGAC
CACCAGGGGCCCCGAGTTGTTCTTCCCCCGTGGCGAGCTGCCAGCCAGGCTGTACCTGTGATCGAGGCTG
GCGGGAATAAGCTTCGTGTGCTCAGGTATGGGAGGTGCAGGACAGCTCATGAAACGCCAACAATCGCAC
AATTCATGTCAAGCTAATCAGCTATTTCTCTTACAGAGCTGTAATTGTCCAAAATTCTGGTCTACCGGGG
GTGATCCTTCGTGTACGGGCCCTTCCCTCAACCCTAGGTATGCGCGCATGCGGTGCGCCGCGCAACTCGCGCG
AGGGCCGAGGGTTTGGGACGGGCCGTCCCGAAATGCAGTTGCACCCGGATGCGTGGCACCTTTTTTTCGATA
ATTTATGCAATGGAAGTCTGCTGCAAAATTCTGGCTCTGTGCGCAACCCTAGGATCAGCGGCGTAGGATTTTCG
TAATCATTCGTCTGATGGGAGCTACCGACTACCTAATATCAGCCGACTGCCTGACGCCAGCGTCCACT
TTTGTGCACACATTCATTCGTGCCAAGACATTTATTGTGGTGCGAAGCGTCCCCAGTTACGCTCACCTG
TTTCCCGACCTCCTTACTGTTCTGTGACAGAGCGGGCCACAGGCCGGTTCGACGCC

SEQ ID NO: 38
5' 6S genomic donor sequence of *Prototheca moriformis*

GCTCTTCGCGCGCCCACTCCTGCTCGAGCGCGCCGCGCGTGCGCCGCCAGCGCCTTGGCCTTTTCGCCGC
GCTCGTGCGCGTGCCTGATGTCCATCACCAGGTCCATGAGGTCTGCCTTGCGCCGGCTGAGCCACTGCTTCG
TCCGGGCGGCCAAGAGGAGCATGAGGGAGGACTCCTGGTCCAGGTCCTGACGTGGTTCGCGGCTCTGGGAGC
GGGCCAGCATCATCTGGCTCTGCCGCACCGAGGCCGCCTCCAAGTGGTCTCCAGCAGCCGAGTCGCCGCC
GACCTTGGCAGAGGAAGACAGGTGAGGGGGGTATGAATTGTACAGAACAACCACGAGCCTTGTCTAGGCAGA
ATCCCTACCAGTCATGGCTTTACCTGGATGACGGCCTGCAACAGCTGTCCAGCGACCTCGCTGCCGCCGC
TTCTCCCGCACGTTCTTCCAGCACCGTGTATGGCGCGAGCCAGCGCCGACGCTGGCGCTGCGCTTCGCCG
ATCTGAGGACAGTCGGGGAACCTCTGATCAGTCTAAACCCCTTGC CGGTTAGTGTTGCCATCCTTTGCAGAC
CGGTGAGAGCCGACTTGTGTGCGCCACCCCCACACCACCTCCTCCAGACCAATTCTGTACCTTTTTTGG
CGAAGGCATCGCCCTCGCCCTGCAGAGAGGACAGCAGTGCACGCCGCTGGGGGTTGGCGGATGCACGCTCA
GGTACC

SEQ ID NO: 39
3' 6S genomic donor sequence of *Prototheca moriformis*

GAGCTCCTTGTTTTCCAGAAGGAGTTGCTCCTTGAGCCTTTTATTCTCAGCCTCGATAACCTCCAAAGCCGC
TCTAATTGTGGAGGGGTTTCGAATTTAAAAGCTTGAATGTTGGTTTCGTGCGTCTGGAACAAGCCAGACTT
GTTGCTCACTGGGAAAAGGACCATCAGCTCCAAAAAAGTTCGCGCTCAAACCGCGTACCTCTGCTTTTCGCGC
AATCTGCCCTGTTGAAATCGCCACCACATTATGTTGACGCTTGAGCAGTCTGTAATTGCCTCAGAATGT
GGAATCATCTGCCCCCTGTGCGAGCCCATGCCAGGCATGTGCGGGCGAGGACACCCGCCACTCGTACAGCA
GACCATTATGCTACCTCACAATAGTTTATAACAGTGACCATATTTCTCGAAGCTCCCCAACGAGCACCTCCA
TGCTCTGAGTGGCCACCCCCCGGCCCTGGTGTGTCGGAGGGCAGGTCAACCGGCATGGGGCTACCGAAATC
CCCGACCGGATCCCACCACCCCGCGATGGGAAGAATCTCTCCCGGGATGTGGGGCCACCACCAGCACAAAC
CTGCTGGCCCGAGCGAGCGTCAAACCATACACAAATATCCTTGGCATCGGCCCTGAATTCCTTCTGCCG
CTCTGCTACCCGTTGCTTCTGTCCGAAGCGGGGTTGCTAGGGATCGCTCCGAGTCCGCAAAACCTTGTGCG
GTGGCGGGGCTTGTTCGAGCTTGAAGAGC

SEQ ID NO: 40
FADc portion of the long hairpin RNA expression cassette from pSZ1468

ACTAGTATGGCTATCAAGACGAACAGGCAGCCTGTGGAGAAGCCTCCGTTACGATCGGGACGCTGCGCAAG
GCCATCCCCGCGCACTGTTTCGAGCGCTCGGCGCTTCGTAGCAGCATGTACCTGGCCTTTGACATCGCGGTC
ATGTCCCTGCTCTACGTCGCGTCGACGTACATCGACCCTGCACCGGTGCCTACGTGGGTCAAGTACGGCATC
ATGTGGCCGCTCTACTGGTTCTTCCAGGTGTGTTTGGGGTTTTGGTTGCCCGTATTGAGGTCCTGGTGGCG
CGCATGGAGGAGAAGGCGCCTGTCCCGCTGACCCCCCGGCTACCTCCCGGCACCTTCAGGGCGCGTACG
GGAAGAACCAGTAGAGCGGCCACATGATGCCGTACTTGACCCACGTAGGCACCGGTGCAGGGTCGATGTACG
TCGACGCGACGTAGACGAGGACATGACCGCATGTCAAAGGCCAGGTACATGCTGTACGAAAGCGCCGAGC
GCTCGAAACAGTGC GCGGGGATGGCCTTGC GCGAGCGTCCCGATCGTGAACGGAGGCTTCTCCACAGGCTGCC
TGTTCTGCTTTGATAGCCAT

SEQ ID NO: 41

Relevant portion of the FADc long hairpin RNA expression cassette from pSZ1468

5 GCTCTTCGCGCCGCCACTCCTGCTCGAGCGCGCCGCGCGTGCGCCGCCAGCGCCTTGGCCTTTTCGCCCGC
 GCTCGTGC CGCTCGCTGATGTCCATCACCAGGTCCATGAGGTCTGCCTTGCGCCGGCTGAGCCACTGTCTCG
 TCCGGGCGGCCAAGAGGAGCATGAGGAGGACTCTGGTCCAGGGTCTGACGTGGTGC CGGCTCGGGAGC
 GGGCCAGCATCATCTGGCTCTGCCGCACCGAGGCCGCTCCAACTGGTCCTCCAGCAGCCGCGAGTCGCCGCC
 10 GACCCCTGGCAGAGGAAGACAGGTGAGGGGGGTATGAATTGTACAGAACAACCACGAGCCTTGTCTAGGCAGA
 ATCCCTACCAGTCATGGCTTTACCTGGATGACGGCCTGCGAACAGCTGTCCAGCGACCCCTCGCTGCCGCCGC
 TTCTCCCGCACGCTTCTTTCCAGCACCGTGATGGCGCGAGCCAGCGCCGCACGCTGGCGCTGCGCTTCGCCCG
 ATCTGAGGACAGTCGGGGAACCTCTGATCAGTCTAAACCCCTTGC CGCTTAGTGTTGCCATCCTTTGCAGAC
 CGGTGAGAGCCGACTTGTGTGCGCCACCCCCACACCCACTCCTCCAGACCAATTCTGTACCTTTTGG
 CGAAGGCATCGGCCTCGGCCTGCAGAGAGGACAGCAGTGCCAGCCGCTGGGGGTTGGCGGATGCACGCTCA
 15 GGTACCCTTTCTTGCCTATGACACTTCCAGCAAAGGTAGGGCGGGCTGCGAGACGGCTTCCCGGCGCTGC
 ATGCAACACCGATGATGCTTCGACCCCCCGAAGCTCCTTCGGGGCTGCATGGGCGCTCCGATGCCGCTCCAG
 GGCGAGCGCTGTTTAAATAGCCAGGCCCCCGATTGCAAAGACATTATAGCGAGCTACCAAAGCCATATTCAA
 ACACCTAGATCACTACCACTTCTACACAGGCCACTCGAGCTTGTGATCGCACTCCGCTAAGGGGGCGCCTCT
 TCCTCTTCGTTTTCAGTCACAACCCGCAACCGCGCCATCTGCTGTCAGGCCTTCTGTTCTCTGCTGGCCG
 20 GCTTCGCCCGCAAGATCAGCGCCTCCATGACGAACGAGACGTCCGACCGCCCCCTGGTGCACTTCACCCCCA
 ACAAGGGCTGGATGAACGACCCCAACGGCCTGTGGTACGACGAGAAGGACGCCAAGTGGCACCTGTACTTCC
 AGTACAACCCGAACGACACCGTCTGGGGGACGCCCTTGTTCGGGGCCACGCCACGTCCGACGACCTGACCA
 ACTGGGAGGACCAGCCCATCGCCATCGCCCCGAAGCGCAACGACTCCGGCGCCTTCTCCGGCTCCATGGTGG
 TGGACTACAACAACACCTCCGGCTTCTTCAACGACACCATCGACCCGCGCCAGCGCTGCGTGGCCATCTGGA
 CCTACAACACCCCGAGTCCGAGGAGCAGTACATCTCTACGCTTGACGCGGCTACACCTTCCCGAGT
 25 ACCAAGAAGACCCCGTGCTGGCGGCCAACTCCACCCAGTTCGCGACCCGAAGGTCTTCTGGTACGAGCCCT
 CCCAGAAGTGGATCATGACCGCGGCCAAGTCCCAGGACTACAAGATCGAGATCTACTCCTCCGACGACCTGA
 AGTCTTGGAAGCTGGAGTCCGCGTTCGCCAACGAGGGCTTCTCGGCTACCAGTACGAGTGCCCCGGCCTGA
 TCGAGGTCCCCACCGAGCAGGACCCAGCAAGTCTACTGGGTGATGTTTCATCTCCATCAACCCCCGGCGCCC
 CGGCCGCGCGCTCCTTCAACAGTACTTCTGTCGGCAGCTTCAACGGCACCCACTTCGAGGCTTTCGACAACC
 30 AGTCCCGCTGGTGGACTTCGGCAAGGACTACTACGCCCTGCAGACCTTCTTCAACACCCGACCCGACTACG
 GGAGCGCCCTGGGCATCGCGTGGGCCTCCAACTGGGAGTACTCCGCCTTCTGTCGCCACCAACCCCTGGCGCT
 CCTCCATGTCCCTCGTGC GCAAGTCTCCCTCAACACCGAGTACCAGGCCAACCCGGAGACGGAGCTGATCA
 ACCTGAAGGCCGAGCCGATCCTGAACATCAGCAACGCCGGCCCCCTGGAGCCGGTTCGCCACCAACACCACGT
 TGACGAAGGCCAACAGCTACAACGTGACCTGTCCAACAGCACCGGCACCCCTGGAGTTCGAGCTGGTGTACG
 35 CCGTCAACACCACCAGACGATCTCAAAGTCCGTGTTTCGCGGACCTCTCCCTCTGGTTCAAGGGCCTGGAGG
 ACCCCGAGGAGTACCTCCGCATGGGCTTCGAGGTGTCCGCGTCTCCTTCTTCTGGACCCGGGAACAGCA
 AGGTGAAGTTCGTGAAGGAGAACCCCTACTTCAACCAACCGCATGAGCGTGAACAACAGCCCTTCAAGAGCG
 AGAACGACCTGTCTACTACAAGGTGTACGGCTTGTGGACCAGAACATCCTGGAGCTGTACTTCAACGACG
 GCGACGTCTGTCCACCAACACCTACTTCATGACCACCGGAACGCCCTGGGCTCCGTGAACATGACGACGG
 40 GGGTGGACAACCTGTTCTACATCGACAAGTTCAGGTGCGCGAGGTCAAGTGACAATTGGCAGCAGCAGCTC
 GGATAGTATCGACACACTCTGGACGCTGGTCTGTGATGGACTGTTGCCGCCACACTTGTCTGCCTTGACCTG
 TGAATATCCCTGCCCTTTTATCAAACAGCCTCAGTGTCTGTTGATCTTGTGTGATACGCGCTTTTGCAGTTG
 CTAGCTGCTTGTGCTATTTGCGAATACCACCCCCAGCATCCCCCTTCCCTCGTTTTCATATCGCTTGCATCCCA
 45 ACCGCAACTTATCTACGCTGTCTGCTATCCCTCAGCGCTGCTCCTGCTCCTGCTCACTGCCCCCTCGCACAG
 CCTTGGTTTGGGCTCCGCCTGTATTCTCCTGGTACTGCAACCTGTAAACCAGCACTGCAATGCTGATGCACG
 GGAAGTAGTGGGATGGGAACACAAATGGAGGATCCCGCGTCTCGAACAGAGCGCGCAGAGGAACGCTGAAGG
 TCTCGCCTCTGTGCGACCTCAGCGCGGCATACACCACAATAACCACCTGACGAATGCGCTTGGTTCTTCTGTC
 CATTAGCGAAGCGTCCGGTTCACACACGTGCCACGTTGGCGAGGTGGCAGGTGACAATGATCGGTGGAGCTG
 50 ATGGTTCGAAACGTTTACAGCCTAGGGATATCGAATTCCGCCGACAGGACGCGCGTCAAAGGTGCTGGTCTG
 TATGCCCTGGCCGGCAGGTCTGTGCTGCTGCTGTTAGTGATTCCGCAACCCTGATTTTGGCGTCTTATTTT
 GGCGTGGCAAACGCTGGCGCCCGCAGCCGGGCCGGCGCGATGCGGTGCCCCACGGCTGCCGGAATCCAAG
 GGAGGCAAGAGCGCCCGGTCACTTGAAGGGCTTTACGCGCAAGGTACAGCCGCTCCTGCAAGGTGCGTGG
 TGGAAATTGGACGTGACAGTCTGCTGAAGTCTCTCCACCCGCTCACCAGCGGACAAAGCACCGGTGTATCAG
 55 GTCCGTGTCATCTCACTCTAAAGAGCTCAGCTACGACTACTGATGGCCCTAGATTCTTCATCAAAAACGCT
 GAGACACTTGCCCAGGATTGAAACTCCCTGAAGGGACCAACAGGGGCCCTGAGTTGTTTCTTCCCCCGTGG
 CGAGCTGCCAGCCAGGCTGTACCTGTGATCGAGGCTGGCGGGAAAATAGGCTTCGTGTGCTCAGGTGATGGG
 AGGTGCAGGACAGCTCATGAAACGCCAACAATCGCACAAATCATGTCAAGCTAATCAGCTATTTCTCTTCA
 CGAGCTGTAATTGTCCAAAATTTCTGGTCTACCGGGGGTGATCCTTCGTGTACGGGCCCTTCCCTCAACCT
 60 AGGTATGCGCGCATGCGGTGCGCCGCAACTCGCGCGAGGGCCGAGGGTTTGGGACGGGCCGCTCCCGAAATG
 CAGTTGCAACCCGATGCGTGGCACCTTTTTCGATAATTTATGCAATGGACTGCTCTGCAAAAATTTCTGGCT
 CTGTGCGCAACCTTAGGATCAGCGGCGTAGGATTTCTGTAATCATTCGTCTGATGGGGAGCTACCGACTACC
 CTAATATCAGCCCCGACTGCCTGACGCCAGCGTCCACTTTTGTGCACACATTCCATTCTGTGCCCAAGACATTT
 CATTGTGGTGC GAAGCGTCCCCAGTTACGCTCACCTGTTTCCCGACCTCCTTACTGTTCTGTGACAGAGCG
 65 GGCCACAGGCGCGTTCGACGCCACTAGTATGGCTATCAAGACGAACAGGCAGCCTGTGGAGAAGCCTCCGTT
 CACGATCGGGACGCTGCGCAAGGCCATCCCCGCTCAGTGTTCGAGCGCTCGGCGCTTCGTAGCAGCATGTA
 CCTGGCCCTTGTACATCGCGGTCTGCTCCTGCTACTGCTGCGTTCGACGTACATCGACCTGACCCGCTGCC
 TACGTGGGTCAAGTACGGCATCATGTGGCCGCTCTACTGGTTCTTCCAGGTGTGTTTGGAGGTTTTGGTTGC
 CCGTATTGAGGTCTGGTGGCGCGCATGGAGGAGAAGGCGCCTGTCCCGCTGACCCCCCGGCTACCCCTCCC
 70 GGCACCTTCCAGGGCGCGTACGGGAAGAACCAGTAGAGCGGCCACATGATGCCGTACTTGACCCACGTAGGC
 ACCGGTGCAGGGTTCGATGTACGTGACGCGACGTAGAGCAGGGACATGACCGCGATGTCAAAGGCCAGGTAC
 ATGCTGCTACGAAGCGCCGAGCGCTCGAAACAGTGCGCGGGGATGGCCTTGC CGCAGCGTCCCGATCGTGAAC

GGAGGCTTCTCCACAGGCTGCCTGTTTCGTCTTGATAGCCATCTCGAGGCAGCAGCAGCTCGGATAGTATCGA
 CACACTCTGGACGCTGGTTCGTGTGATGGACTGTTGCCGCCACACTTGCTGCCTTGACCTGTGAATATCCCTG
 CCGCTTTTATCAAACAGCCTCAGTGTGTTTGATCTTGTGTGTACGCGCTTTTGCAGAGTTGCTAGCTGCTTGT
 5 GCTATTTGCGAATACCACCCCCAGCATCCCCCTCCCTCGTTTCATATCGCTTGCATCCCCAACCGCAACTTAT
 CTACGCTGTCCTGCTATCCCTCAGCGCTGCTCCTGCTCCTGCTCACTGCCCCCTCGCACAGCCTTGGTTTGGG
 CTCCGCTGTATTCTCCTGGTACTGCAACCTGTAACACGACTGCAATGCTGATGCACGGGAAGTAGTGGG
 ATGGGAACACAAATGGAAAGCTGTAGAGCTCTTGTTCAGAAAGGAGTTGCTCCTTGAGCCTTTTCATTCTC
 AGCCTCGATAACCTCCAAAGCCGCTCTAATTGTGGAGGGGTTTGAATTTAAAAGCTTGAATGTTGGTTTCG
 10 TGCCTCTGGAACAAGCCAGACTTGTGCTCACTGGGAAAAGGACCATCAGCTCCAAAAAAGCTTGGCGCTCA
 AACCGCGTACCTCTGCTTTCGCGCAATCTGCCCTGTTGAAATCGCCACCACATTTCATATTGTGACGCTTGAG
 CAGTCTGTAATTGCCTCAGAATGTGGAATCATCTGCCCCCTGTGCGAGCCCATGCCAGGCATGTGCGGGGCG
 AGGACACCCGCCACTCGTACAGCAGACCATTATGCTACCAATAGTTTATAACAGTGCATATTTCTC
 GAAGCTCCCCAACGAGCACCTCCATGCTCTGAGTGGCCACCCCCCGGCCCTGGTGCTTGCAGAGGGCAGGTC
 15 AACCGGCATGGGGCTACCGAAATCCCCGACCGGATCCCACCACCCCCGCGATGGGAAGAATCTCTCCCCGGG
 ATGTGGGGCCACCACCAGCACAACTGCTGGGCCAGGCGAGCGTCAAACCATAACACAAATATCCTTGGC
 ATCGGCCCTGAATTCCTTCTGCCGCTCTGCTACCCGGTGCTTCTGTCCGAAGCAGGGGTTGCTAGGGATCGC
 TCCGAGTCCGCAAACCTTGTGCGTGGCGGGGCTTGTTCGAGCTTGAAGAGC

SEQ ID NO: 42

FADc portion of the long hairpin RNA expression cassette from pSZ1469

ACTAGTATGGCTATCAAGACGAACAGGCAGCCTGTGGAGAAGCCTCCGTTACGATCGGGACGCTGCGCAAG
 GCCATCCCCCGCGCACTGTTTCGAGCGCTCGGCGCTTCGTAGCAGCATGTACCTGGCCTTTGACATCGCGGT
 25 ATGTCCCTGCTCTACGTGCGCTCGAGTACCTGACCCGTGCACCGGTGCCTACGTGGGTCAAGTACGGC
 ATGTGGCCGCTCTACTGGTTCTTCCAGGTGTGTTTGAGGGTTTTGGTTGCCCGTATTGAGGTCTGGTGGCG
 CGCATGGAGGAGAAGGCGCCTGTCCCGCTGACCCCCCGGCTACCCTCCCGGCACCTTCCAGGGCGCGTACG
 GGAAGAACCAGTAGAGCGGCCACATGATGCCGTACTTGACCCACGTAGGCACCGGTGCAGGGTTCGATGTACG
 TCGACGCGACGTAGAGCAGGGACATGACCGCGATGTCAAAGGCCAGGTACATGCTGCTACGAAGCGCCGAGC
 30 GCTCGAAACAGTGCAGCGGGATGGCCTTGCGCAGCGTCCCGATCGTGAACGGAGGCTTCTCCACAGGCTGCC
 GTTCGTCTTGATAGCCAT

SEQ ID NO: 43

Relevant portion of the FADc long hairpin RNA expression cassette from pSZ1469

GCTCTTCGCCGCCGCCACTCCTGCTCGAGCGCGCCGCGCGTGCGCCGCCAGCGCCTTGGCCTTTTCGCCGC
 GCTCGTGCAGCTGCTGATGTCCATCACCAGGTCCATGAGGTCTGCCTTGCGCCGGCTGAGCCACTGCTTCG
 TCCGGGCGGCCAAGAGGAGCATGAGGGAGGACTCCTGGTCCAGGGTCTGACGTGGTTCGCGGCTCTGGGAGC
 40 GGGCCAGCATCATCTGGCTCTGCCGCACCGAGGCGCCTCCAAGTGGTCTCCAGCAGCCGAGTCCGCCGC
 GACCTTGGCAGAGGAAGACAGGTGAGGGGGGTATGAATTGTACAGAACAACCACGAGCCTTGTCTAGGCAGA
 ATCCCTACCAGTCTATGGCTTTACCTGGATGACGGCCTGCGAACAGCTGTCCAGCGACCCCTCGCTGCCGCCG
 TTCTCCCGCAGCTTCTTTCCAGCACCGTGTATGGCGCGAGCCAGCGCCGACGCTGGCGCTTCGCTTCCCG
 ATCTGAGGACAGTCCGGGAACCTTGATCAGTCTAAACCCCTTGCAGCTTAGTGTGCCATCCTTTGCAGAC
 45 CGGTGAGAGCCGACTTGTGTGCGCCACCCCCACACCACCTCCTCCAGACCAATTCTGTACCTTTTTTGG
 CGAAGGCATCGGCCTCGGCCTGCAGAGAGGACAGCAGTGCCAGCCGCTGGGGGTGGCGGATGCACGCTCA
 GGTACCTTTCTTGCGCTATGACACTTCCAGCAAAAGGTAGGGCGGGCTGCGAGACGGCTTCCCGGCGCTGC
 ATGCAACACCGATGATGCTTCGACCCCCGAAGCTCCTTCGGGGCTGCATGGGCGCTCCGATGCCGCTCCAG
 50 GGCGAGCGCTGTTTAAATAGCCAGGCCCCGATTGCAAGACATTATAGCGAGTACCAAAGCCATATTCAA
 ACACCTAGATCACTACCCTTCTACACAGGCCACTCGAGCTTGTGATCGCACTCCGCTAAGGGGGCGCCTCT
 TCCTCTTCGTTTCAGTCACAACCCGCAAACGGCGCGCCATGCTGCTGCAGGCCTTCTGTTCTGCTGGCCG
 GCTTCGCCGCCAAGATCAGCGCCTCCATGACGAACGAGACGTCCGACCGCCCCCTGGTGCACCTTACCCCCA
 ACAAGGGCTGGATGAACGACCCCAACGGCCTGTGGTACGACGAGAAGGACGCCAAGTGGCACCTGTACTTCC
 55 AGTACAACCCGAACGACACCGTCTGGGGGACGCCCTGTTCTGGGGCCACGCCACGTCCGACGACCTGACCA
 ACTGGGAGGACCAAGCCATCGCCATCGCCCCGAAGCTGCTCAACGACACCATCGACCCGCGCCAGCGCTGCGTGGCCATCTGGA
 TGGACTACAACAACACCTCCGGCTTCTTCAACGACACCATCGACCCGCGCCAGCGCTGCGTGGCCATCTGGA
 CCTACAACACCCCGGAGTCCGAGGAGCAGTACATCTCCTACAGCCTGGACGGCGGCTACACCTTACCGAGT
 60 ACCAGAAGAACCCCGTGTGGCCGCCAAGTCCACCCAGTTCGCGACCCGAAGGTCTTCTGGTACGAGCCCT
 CCCAGAAGTGGATCATGACCGCGGCCAAGTCCCAGGACTACAAGATCGAGATCTACTCCTCCGACACCTGA
 AGTCTTGAAGCTGGAGTCCGCGTTTCGCCAACAGGGCTTCTCGGCTACCACTACAGTACAGTCCGCGCTGA
 TCGAGTCCCCACCGAGCAGGACCCGAGTCTTCAAGAGCATTATAGCGAGTACCAAAGCCATATTCAA
 65 CGGCCGGCGGCTCCTTCAACAGTACTTTCGTGCGCAGCTTCAACGGCACCCACTTCGAGGCCTTCGACAACC
 AGTCCCGCGTGGTGGACTTCGGCAAGGACTACTACGCCCTGCAGACCTTCTTCAACACCGACCCGACCTACG
 GGAGCGCCCTGGGCATCGCGTGGGCCTCCAAGTGGGAGTACTCCGCTTCTGTGCCACCAACCCCTGGCGCT
 CCTCCATGTCCCTCGTGCAGAACTTCTCCCTCAACACCGAGTACCAGGCCAACCCGGAGACGGAGCTGATCA
 70 ACCTGAAGGCCGAGCCGATCCTGAACATCAGCAACCGCGCCCTGGAGCCGGTTTCGCCACCAACACCACTG
 TGACGAAGGCCAACAGCTACAACGTGACCTTCAACAGCAGCACCGGACCCCTGGAGTTGAGTTTCAGAGTGTGATACG
 CCGTCAACACCAACCCAGACGATCTCCAAGTCCGTGTTTCGCGGACCTTCTCCCTCTGGTTCAAGGGCCTGGAGG
 ACCCCGAGGAGTACCTCCGATGGGCTTCGAGGTGTCCGCGTCTCCTTCTTCTGGACCGCGGGAACAGCA
 AGGTGAAGTTTCGTGAAGGAGAACCCCTACTTCACCAACCGCATGAGCGTGAACAACAGCCCTTCAAGAGCG
 AGAACGACCTGTCTACTACAAGGTGTACGGCTTGCTGGACCAGAACATCCTGGAGCTGTACTTCAACGACG
 GCGACGTCTGTCCACCAACACCTACTTCATGACCACCGGAACGCCCTGGGCTCCGTGAACATGACGACGG

GGGTGGACAACCTGTTCTACATCGACAAGTTCAGGTGCGCGAGGTCAAGTGACAATTGGCAGCAGCAGCTC
 GGATAGTATCGACACACTCTGGACGCTGGTCTGTGATGGACTGTTGCCGCCACACTTGCTGCCTTGACCTG
 TGAATATCCCTGCCGCTTTTATCAAACAGCCTCAGTGTGTTTGATCTTGTGTGTACGCGCTTTTGCGAGTTG
 CTAGCTGCTTGTGCTATTTGCGAATACCACCCCCAGCATCCCCCTTCCCTCGTTTCATATCGCTTGCAATCCCA
 5 ACCGCAACTTATCTACGCTGTCTGCTATCCCTCAGCGCTGCTCCTGCTCCTGCTCACTGCCCTCGCACAG
 CCTTGGTTTGGGCTCCGCTGTATTCTCCTGGTACTGCAACCTGTAAACCAGCACTGCAATGCTGATGACAG
 GGAAGTAGTGGGATGGGAACACAAATGGAGGATCCGCGTCTCGAACAGAGCGCGCAGAGGAACGCTGAAGG
 TCTCGCTCTGTGCGACCTCAGCGCGGCATACACCACAATAACCACCTGACGAATGCGCTTGGTTCTTCGTC
 CATTAGCGAAGCGTCCGTTTACACACAGTGCCACGTTGGCGAGGTGGCAGGTGACAATGATCGGTGGAGCTG
 10 ATGGTCGAAAACGTTACACAGCCTAGGGATATCGAATTCCTTTCTTGCCTATGACACTTCCAGCAAAAGGTAG
 GCGGGCTGCGAGACGGCTTCCCGCGCTGCTGCAACACCGATGATGCTTCGACCCCCCGAAGCTCCTTCG
 GGGTCGATGGGCGTCCGATGCCGCTCCAGGGCGAGCGCTGTTTAAATAGCCAGGCCCCCGATTGCAAAGA
 CATTATAGCGAGCTACCAAAGCCATATTCAAACACCTAGATCACTACCACTTCTACACAGGCCACTCGAGCT
 TGTGATCGCACTCCGCTAAGGGGGCGCCTCTTCTCTTCTGTTTCAGTCACAACCCGCAAACTAGTATGGC
 15 TATCAAGACGAACAGGCAGCCTGTGGAGAAGCCTCCGTTACGATCGGGACGCTGCGCAAGGCCATCCCCGC
 GCACTGTTTCGAGCGCTCGGCGCTTCGTAGCAGCATGTACCTGGCCTTTGACATCGCGGTGATGTCCCTGCT
 CTACGTCGCGTCGACGTACATCGACCCTGCACCGGTGCCTACGTGGGTCAAGTACGGCATCATGTGGCCGCT
 CTACTGTTTCTTCCAGTGTGTTTGGGGTTCCTGGTTCGCGCTATTGAGGTCTGGTGGCGCGCATGGAGGA
 GAAGGCGCTTGTCCCGCTGACCCCCCGGCTACCCTCCCGGCACCTTCCAGGGCGCGTACGGGAAGAACCAG
 20 TAGAGCGGCCACATGATGCCGTACTTGACCCACGTAGGCACCGGTGCAGGGTGCATGTACGTGACGCGGACG
 TAGAGCAGGGACATGACCGCGATGTCAAAGGCCAGGTACATGCTGCTACGAAGCGCCGAGCGCTCGAAACAG
 TGCGCGGGGATGGCCTTGCGCAGCGTCCCGATCGTGAACGGAGGCTTCTCCACAGGCTGCCTGTTCTGCTTG
 ATAGCCATCTCGAGGCAGCAGCTCGGATAGTATCGACACACTCTGGACGCTGGTCTGTGATGGACTGT
 TGCCGCGCTTGTGCTGCCCTTGACCTGTGAATATCCCTGCCGCTTTTATCAAACAGCCTCAGTGTGTTGAT
 25 CTTGTGTGTACGCGCTTTTTCGAGTTGCTAGCTGCTTGTGCTATTTGCGAATACCACCCCCAGCATCCCCCT
 CCCTCGTTTCATATCGCTTGCATCCCAACCGCACTTATCTACGCTGTCTGCTATCCCTCAGCGCTGCTCC
 TGCTCCTGCTCACTGCCCCCTCGCACAGCCTTGGTTTGGGCTCCGCTGTATTCTCCTGGTACTGCAACCTGT
 AAACCAGCACTGCAATGCTGATGCACGGGAAGTAGTGGGATGGGAACACAAATGGAAAGCTGTAGAGCTCTT
 GTTTTCCAGAAGGAGTTGCTCCTTGAGCCTTTCATTCTCAGCCTCGATAACCTCCAAAGCCGCTCTAATTGT
 30 GGAGGGGTTTCGAATTTAAAGCTTGGAAATGTTGGTTGCTGCGCTCTGGAACAAGCCAGCTTGTGTGCTCAC
 TGGGAAAAGGACCATTAGCTCCAAAAAATTTGCCGCTCAAACCGCGTACCTCTGCTTTTCGCGCAATCTGCCC
 TGTGAAATCGCCACCACATTATATTGTGACGCTTGAGCAGTCTGTAATTGCCTCAGAATGTGGAATCATC
 TGCCCCCTGTGCGAGCCCATGCCAGGCATGTGCGGGGCGAGGACACCCGCCACTCGTACAGCAGACCATTAT
 GCTACCTCACAATAGTTTCATAACAGTGACCATATTTCTCGAAGCTCCCCAACGAGCACCTCCATGCTCTGAG
 35 TGGCCACCCCCCGCCCTGGTGTCTTGGCGAGGCGAGTCAACCGGCATGGGGCTACCGGAAATCCCCGACCGG
 ATCCCACCCCGCGATGGGAAGAATCTCTCCCGGGATGTGGGCCACACCAGCAGGCTGTGGGCCCTGTGGCC
 CAGGCGAGCGTCAAACCATAACCACACAAATATCCTTGGCATCGGCCCTGAATTCCTTCTGCCGCTCTGCTAC
 CCGGTGCTTCTGTCCGAAGCAGGGGTTGCTAGGGATCGCTCCGAGTCCGCAAACCTTGTGCGGTGGCGGGG
 40 CTTGTTTCGAGCTTGAAGAGC

SEQ ID NO: 44

FADc portion of the long hairpin RNA expression cassette from pSZ1470

ACTAGTTCACCTTGTGGAACCAGAGCGCGGAGTCTGCTCCTCGGGCGCGTCCGGGACGACGTAGCGGCAGTCCGG
 45 CCAGTCTCCACAGGGCGCGGCCGACCCAGCGGCTGTGCGACTGGTAGTACTTGCCAGGATGGGCCTGAT
 GCGGCGGAGGCCCTCCTCGGCGTGGTAGTGCGGGATGGTGTGTAAGAGGTGGTGCAGCACGTGGGTGTGCGGA
 GATGTGGTGCAGGATGTTGTCCATGAAGGGCGGGCCCATGGAGCGGTCCACGGTGGCCATGGCGCCGCGCAG
 CCAGTCCCAGTCTTCTCGAAGTAGTGCGGACGCGCGGCTGCGTGTGCTGGAGCAGCGTGATGAGCACGAG
 50 CCACATGTTTCAGATCAGGTAGGGCACACGTAGGTCTTGACCAGCCAGGCCAGCCCATGGTGGCGCCAG
 CACGCTGAGCCCGCTGAGCACCGCCACCAGCGCCAGGTGCGAGATGACCACCTCGATGCGCTCGCGCTTGCT
 GAAGATGGGCGACCACGGGTCAAAGTGGTTGGCGAAGCGCGGTACGGCCGCGAGGCGACGTTGAACATGAG
 GTACAGCGGCCAGCCAGGGTCAGGGTGACCAGCACCTTGCCCATGCGGATGGGCAGCCACTCCTCCCACTC
 CAGGCCCTCGTGCGCCACTGCGCGGTGCGGCGGACAAACACCTCGTCTTGTCCAGGACCCCGTGTGGA
 GTGGTGGCGGCGGTGCGAGTGCTTCCAGGAGTAGAGGACACCAGCAGCAGGCTGTGGAACACCAGGCCAC
 55 GCGCTCGTTGATGGCCTGGCTGGAGGAAAAGGCTTGGTGGCCGCACTCGTGCGCGCACACCCAGACACCCGT
 GCCGAAGGCGCCCTGGAAGGTGCCGGGAGGGTAGCCGGGGGGGTGAGCGGGACAGGCGCCTTCTCCTCCATG
 CGCGCCACCAGGACCTCAATACGGGCAACCAAACCTCAAACACACCTGGAAGAACCAGTAGAGCGGCCAC
 ATGATGCCGTACTTGACCCACGTAGGCACCGGTGCAGGGTGCATGTACGTGACGCGGACGTAGAGCAGGGAC
 60 ATGACCGCGATGTCAAAGGCCAGGTACATGCTGCTACGAAGCGCCGAGCGCTCGAAACAGTGCGCGGGGATG
 GCCTTGCGCAGCGTCCGATCGTGAACGGAGGCTTCTCCACAGGCTGCCTGTTCTGCTTTGATAGCCAT

SEQ ID NO: 45

Relevant portion of the FADc long hairpin RNA expression cassette from pSZ1470

GCTCTTCGCCGCCGCCACTCCTGCTCGAGCGCGCCGCGCGTGCGCCGCCAGCGCCTTGGCCTTTTCGCCGC
 GCTCGTGCGCTGCTGATGTCCATACCAGGTCCATGAGGTCTGCCTTGCGCCGGCTGAGCCACTGCTTCG
 TCCGGGCGGCCAAGAGGAGCATGAGGGAGGACTCCTGGTCCAGGGTCTGACGTGGTGCAGGCTCTGGGAGC
 70 GGGCCAGCATCATCTGGCTCTGCCGCACCGAGGCCGCTCCAACCTGGTCTCCTCAGCAGCCGAGTCGCCGCC
 GACCTTGGCAGAGGAAGACAGGTGAGGGGGGTATGAATTGTACAGAACAACACGAGCCTTGTCTAGGCAGA
 ATCCCTACCAGTCATGGCTTTACCTGGATGACGGCTGCGAACAGCTGTCCAGCGACCTCGCTGCCGCCGC

TTCTCCCGCACGCTTCTTTCCAGCACCGTGATGGCGCGAGCCAGCGCCGCACGCTGGCGCTGCGCTTTCGCCG
 ATCTGAGGACAGTCGGGGAACCTCTGATCAGTCTAAACCCCTTGCCTGTTAGTGTGCCATCCTTTGCAGAC
 CGGTGAGAGCCGACTTGTGTGCGCCACCCCCACACCACCTCCTCCAGACCAATTCTGTACCTTTTTTGG
 CGAAGGCATCGGCCTCGGCCTGCAGAGAGGACAGCAGTGCCAGCCGCTGGGGGTTGGCGGATGCACGCTCA
 5 GGTACCCTTTCTTGCCTATGACACTTCCAGCAAAGAGTGGGCGGGCTGCGAGACGGCTTCCCGGCGCTGC
 ATGCAACACCGATGATGCTTCGACCCCCCAAGCTCCTCGGGGCTGCATGGGCGCTCCGATGCCGCTCCAG
 GGCGAGCGCTGTTTAAATAGCCAGGCCCCCGATTGCAAAGACATTATAGCGAGCTACCAAAGCCATATTCAA
 ACACCTAGATCACTACCCTTCTACACAGGCCACTCGAGCTTGTGATCGCACTCCGCTAAGGGGGCGCTCT
 10 TCCTCTTCGTTTTCAGTCACAACCCGCAAACGGCGCGCCATGCTGCTGCAGGCCTTCTGTTCCTGCTGGCCG
 GCTTCGCCCGCAAGATCAGCGCCTCCATGACGAACGAGACGTCCGACCGCCCCCTGGTGCACTTACCCCCA
 ACAAGGCTGGATGAACGACCCCAACGGCCTGTGGTACGACGAGAAGGACGCCAAGTGGCACCTGTACTTCC
 AGTACAACCCGAACGACACCGTCTGGGGGACGCCCTTGTCTGGGGCCACGCCACGTCCGACGACCTGACCA
 ACTGGGAGGACGAGCCCATCGCCATCGCCCCGAAGCGCAACGACTCCGGCGCCTTCTCCGGCTCCATGGTGG
 15 TGGACTACAACAACACCTCCGGCTTCTTCAACGACACCATCGACCCGCGCCAGCGCTGCGTGGCCATCTGGA
 CCTACAACACCCCGAGTCCGAGGAGCAGTACATCTCCTACAGCCTGGACGGCGGCTACACCTTACCGAGT
 ACCAGAAGAACCCCGTGCTGGCCGCCAACTCCACCCAGTTCGCGACCCGAAGGTCTTCTGGTACGAGCCCT
 CCCAGAAGTGGATCATGACCGCGGCCAAGTCCCAGGACTACAAGATCGAGATCTACTCCTCCGACGACCTGA
 AGTCTTGAAGCTGGATCCGCTTCCGCCAACGAGGCTTCTCGGCTACCAGTACGATGCCCGCGCTGA
 20 TCGAGGTCCCCACCGAGCAGGACCCAGCAAGTCTACTGGGTGATGTTTCATCTCCATCAACCCCGCGCGCC
 CGGCCGGCGGCTCCTTCAACAGTACTTCTCGCGCAGCTTCAACGGCACCCACTTCGAGGCCTTCGACAACC
 AGTCCCGCGTGGTGGACTTCGGCAAGGACTACTACGCCCTGCAGACCTTCTTCAACACCGACCCGACCTACG
 GGAGCGCCCTGGGCATCGCGTGGGCCTCCAACCTGGGAGTACTCCGCCTTCGTGCCCCACCAACCCCTGGCGCT
 CCTCCATGTCCCTCGTGCAGTCTCTCCCTCAACACCGAGTACCAGGCCAACCCGAGACGGAGCTGATCA
 25 ACCTGAAGCCGAGCCGATCCTGAACATCAGCAACCGGCCCTTGAGGCCGTTTCGCCACCAACACACGCT
 TGACGAAGGCCAACAGCTACAACGTGACCTGTCCAACAGCACCGGCACCCTGGAGTTCGAGCTGGTGTACG
 CCGTCAACACCAACCCAGACGATCTCCAAGTCCGTGTTCCGCGACCTCTCCCTCTGGTTCAAGGGCCTGGAGG
 ACCCCGAGGAGTACCTCCGCATGGGCTTCGAGGTGTCCGCGTCTCCTTCTTCTTGACCGCGGGAACAGCA
 AGGTGAAGTTCGTGAAGGAGAACCCCTACTTCAACCAACCGCATGAGCGTGAACAACAGCCCTTCAAGAGCG
 30 AGAACGACCTGTCTACTACAAGGTGTACGGCTTGTGGACCAGAACATCCTGGAGCTGTACTTCAACGACG
 GCGACGTCTGTCCACCAACACCTACTTTCATGACACCGGGAACGCCCTGGGCTCCGTGAACATGACGACGG
 GGGTGGACAACCTGTTCTACATCGACAAGTTCAGGTGCGCGAGGTCAAGTGACAATTGGCAGCAGCAGCTC
 GGATAGTATCGACACACTCTGGACGCTGGTCTGTGATGGACTGTTGCCGCCACACTTGCTGCCTTGACCTG
 TGAATATCCCTGCCGCTTTTATCAAACAGCCTCAGTGTGTTTGATCTTGTGTGTACGCGCTTTTGCAGTGTG
 35 CTAGCTGCTTGTGCTATTTGCGAATACCACCCCCAGCATCCCTTCCCTCGTTTTCATATCGCTTGCAATCCCA
 ACCGCAACTTATCTACGCTGTCTGCTATCCCTCAGCGCTGCTCCTGCTCCTGCTCACTGCCCTCGCACAG
 CCTTGGTTTGGGCTCCGCTGTATTCTCTGGTACTGCAACCTGTAAACAGCACTGCAATGCTGATGCAAG
 GGAAGTAGTGGGATGGGAACACAAATGGAGGATCCCGCTCTCGAACAGAGCGCGCAGAGGAACGCTGAAGG
 TCTCGCCTCTGTGCGACCTCAGCGCGGCATACACCACAATAACCACTGACGAATGCGCTTGGTTCTTCTGTC
 40 CATTAGCGAAGCGTCCGTTTACACACGTGCCACGTTGGCGAGGTGGCAGGTGACAATGATCGGTGGAGCTG
 ATGGTCAAAACGTTTACAGCCTAGGGATATCGAATTCCGCCGACAGGACGCGCGTCAAAGGTGCTGGTCTGT
 TATGCCCTGGCCGCGAGGCTGTTGCTGCTGCTGGTGTAGTGATTCCGCAACCTGATTTTGGCGCTTATTTT
 GGCGTGGCAACCGTGGCGCCCGGAGCCGGCGCGGATGCGGTGCCCGACGCTCCGGAATCCCAAG
 GGAGGCAAGAGCGCCCGGTGAGTTGAAGGGCTTACGCGCAAGGTACAGCCGCTCCTGCAAGGCTGCGTGG
 45 TGGAATTGGACGTGCAGGTCTGCTGAAGTTCCTCCACCGCCTCACCAGCGGACAAAGCACCGGTGTATCAG
 GTCCGTGTATCCACTCTAAAGAGCTCGACTACGACCTACTGATGGCCCTAGATTCTTCATCAAAAACGCT
 GAGACACTTGCCCAGGATTGAAACTCCCTGAAGGGACCAACAGGGGCCCTGAGTTGTTTCTTCCCCCGTGG
 CGAGCTGCCAGCCAGGCTGTACCTGTGATCGAGGCTGGCGGGAAAATAGGCTTCTGTGTGCTCAGGTATGGG
 50 AGGTGACGACAGCTCATGAAACGCCAACAATCGCACAAATTCATGTCAAGCTAATCAGCTATTCTCTCTTCA
 CGAGCTGTAATTGTCCAAAATTTCTGGTCTACCGGGGGTGATCCTTCTGTGTACGGGCCCTTCCCTCAACCT
 AGGTATGCGCGCATGCGGTGCGCGCGCAACTCGCGCGAGGGCCGAGGGTTTGGGACGGGCGCTCCCGAAATG
 CAGTTGCACCCGATGCGTGGCACCTTTTTTGCATAATTTATGCAATGGACTGCTCTGCAAAATTTGCGT
 CTGTGCGCAACCTTAGGATCAGCGGCGTAGGATTTCTGTAATCATTCGTCTGATGGGGAGCTACCGACTACC
 55 CTAATATCAGCCGAGCTGCCAGCGGACGCTCCACTTTTGTGACACATTCCATTCTGTGCCCAAGACATTT
 CATTTGTCGGAAGCGTCCCGAGTTACGCTCAGCTGTTTCCCGACCTCCTTACTGTTCTGTGCGACAGCG
 GGCCACAGGCGGTCGCGAGCCACTAGTTTCACTTGTGGAACAGAGCGCGGAGTCTGCTCCTCGGGCGCGTCCG
 GGACGACGTAGCGGACGTGCGGCCAGTCTCCACAGGGCGCGGCGACCCAGCGGCTGTGCGACTGGTAGT
 60 ACTTGCCCGAGGATGGGCCTGATGGCGGCGGAGGCCTCCTCGGCGTGGTAGTGCGGGATGGTGTGAAGAGGT
 GGTGCAGCACGTGGGTGTGCGAGATGTGGTGCAGGATGTTGTCCATGAAGGGCGGGGCCATGGAGCGGTCCA
 CGGTGGCCATGGCGCCGCGCAGCCAGTCCAGTCTTCTCGAAGTAGTGCGGACGCGCGGCTGCGTGTGCT
 GGAGCAGCGTGTAGACACGAGCCACATGTTTACGATCAGTAGGGACACCAGTACGTTCTTGACGACGAGG
 65 CCCAGCCCATGGTGGCGCCAGCACGCTGAGCCCGTGGACACCGCCACCAGCGCCAGGTGCGAGATGACCA
 CCTCGATGCGCTGCGCTTGTGTAAGATGGGCGACACGGGTCAAAGTGGTTGGCGAAGCGCGGGTACGGCC
 GCGAGGCGACGTTGAACATGAGGTACAGCGGCCAGCCAGGGTCAGGGTGACAGCACCTTGCCCATGCGGA
 TGGGCGAGCCACTCCTCCCACTCCAGGCCCTCGTGCGCCACTGCGCGGTGCGGCGGCACAAACACCTCGTCT
 70 TGTCCAGGCACCCCGTGTGGAGTGGTGGCGGCGGTGCGAGTGCTTCCAGGAGTAGTAGGGCACCGACGCA
 GGCTGTGGAACACACCGGACCGCTGTTGATGGCTGGCTGGAGGAAAAGGCTGGTGGCGGCGACTCGT
 GCGCGCACACCCAGACACCCGTGCCGAAGGCGCCCTGGAAGGTGCCGGGAGGGTAGCCGGGGGGGTGACGCG
 GACAGGCGCCTTCTCCTCCATGCGCGCCACCAGGACCTCAATACGGGCAACCAAAACCTCAAACACACCTG
 GAAGAACCAGTAGAGCGGCCACATGATGCCGTACTTGACCCACGTAGGCACCGGTGCAGGGTGCATGTACGT
 CGACGCGACGTAGAGCAGGGACATGACCGCGATGTCAAAGGCCAGGTACATGCTGTACGAAGCGCCGAGCG
 CTCGAAACAGTGCAGGGGATGGCCTTGGCGAGCGTCCCGATCGTGAACGGAGGCTTCTCCACAGGCTGCTT

GTTCGTCTTGATAGCCATCTCGAGGCAGCAGCAGCTCGGATAGTATCGACACACTCTGGACGCTGGTTCGTGT
 GATGGACTGTTGCCGCCACACTTGCTGCCTTGACCTGTGAATATCCCTGCCGCTTTTATCAAACAGCCTCAG
 TGTGTTTGATCTTGTGTGTACGCGCTTTTGCAGATTGCTAGCTGCTTGTGCTATTTGCGAATACCACCCCCA
 GCATCCCCCTTCCCTCGTTTTCATATCGCTTGCATCCCAACCGCAACTTATCTACGCTGTCTGTATCCCTCA
 5 GCGCTGCTCCTGCTCCTGCTCACTGCCCCCTCGCACAGCCTTGGTTTGGGCTCCGCTGTATTCTCCTGGTAC
 TGCAACCTGTAAACACGACTGCAATGCTGATGCACGGGAAGTAGTGGGATGGGAACACAAATGGAAAGCTG
 TAGAGCTCTTGTTTTTCCAGAAGGAGTTGCTCCTTGAGCCTTTTATTCTCAGCCTCGATAACCTCCAAAGCCG
 CTCTAATTGTGGAGGGGGTTTGAATTTAAAGCTTGAATGTTGGTTCTGTCGTCTGGAACAAGCCCAGACT
 10 TGTGCTCACTGGGAAAAGGACCATCAGCTCCAAAAAAGCTTGGCGCTCAAACCGCGTACCTCTGCTTTTCGCG
 CAATCTGCCCTGTTGAAATCGCCACCACATTATATTGTGACGCTTGAGCAGTCTGTAATTGCCTCAGAATG
 TGAATCATCTGCCCCCTGTGCGAGCCCATGCCAGGCATGTGCGGGGCGAGGACACCCGCCACTCGTACAGC
 AGACCTTATGCTACCTACAATAGTTTATAACAGGTACCATATTTCTCGAAGCTCCCCAACGAGCACCTCC
 ATGCTCTGAGTGGCCACCCCCCGGCCCTGGTGCTTGGGAGGGCAGGTCAACCGGCATGGGGCTACCGAAAT
 15 CCCCAGCCGATCCCACCACCCCCGCGATGGGAAGAATCTCTCCCCGGGATGTGGGCCCCACCACCAGCACAA
 CCTGCTGGCCCAGGCGAGCGTCAAACCATACCACACAAATATCCTTGGCATCGGCCCTGAATTCTTCTGCC
 GCTCTGCTACCCGGTGCTTCTGTCCGAAGCAGGGGTTGCTAGGGATCGCTCCGAGTCCGCAAACCTTGTGCG
 CGTGGCGGGGCTTGTTCGAGCTTGAAGAGC

SEQ ID NO: 46

FADc portion of the long hairpin RNA expression cassette from pSZ1471

ACTAGTTCACCTTGTGGAACCAGAGCGCGGAGTTCGTCTCGGGCGCGTCCGGGACGACGTAGCGGCAGTCCGCG
 CCAGTCCCTCCACAGGGCGCGGCCGACCCAGCGGCTGTGCGACTGGTAGTACTTGGCCAGGATGGGCCTGAT
 25 GGCGCGGAGGCCCTCCTCGGCGTGGTAGTGCAGGATGGTGCTGAAGAGGTGGTGACACAGTGGGTGTTCGGA
 GATGTGGTGCAGGATGTTGTCCATGAAGGGCGGGCCCATGGAGCGGTCCACGGTGGCCATGGCGCCGCGCAG
 CCAGTCCCAGTCTTCTCGAAGTAGTGCAGCGCGCGGGTGCCTGTGCTGGAGCAGCGTGTGAGCAGCAG
 CCACATGTTACAGATCAGGTAGGGCACCACGTAGGTCTTGACCAGCCAGGCCAGCCCATGGTGCAGGCCAG
 CACGCTGAGCCCGCTGAGCACCGCCACCAGCGCCAGGTCCGAGATGACCACCTCGATGCGCTCGCGCTTGCT
 GAAGATGGGCGACACGGGTCAAAGTGGTTGGCGAAGCGCGGTACGGCCGCGAGGGCAGCTTGAACATGAG
 30 GTACAGCGGCCAGCCCAGGGTCAGGGTGACCAGCACCTTGCCCATGCGGATGGGCAGCCACTCCTCCACTC
 CAGGCCCTCGTGCGCCACTGCGCGGTGCGGCGGCACAAACACCTCGTCTTGTCCAGGCACCCCGTGTGGA
 GTGGTGGCGGCGGTGCGAGTGCTTCCAGGAGTAGTAGGGCACCAGCAGCAGGCTGTGGAACACCAGGGCCAC
 GCCGTGCTTGATGGCCTGGCTGGAGGAAAAGGCCTGGTGGCCGCACTCGTGCGCGCACACCCAGACACCCGT
 35 GCCGAAGGCGCCCTGGAAGGTGCCGGGAGGGTAGCCGGGGGGGTGAGCGGGACAGGCGCCTTCTCCTCCATG
 CGCGCCACCAGGACCTCAATACGGGCAACCAAAACCTCAAACACACCTGGAAGAACCAGTAGAGCGGCCAC
 ATGATGCCGTACTTGACCCACGTAGGCACCGGTGACGGGTGATGTACGTGACGTGACGCGACGTAGAGCAGGGAC
 ATGACCGCGATGTCAAAGGCCAGGTACATGCTGCTACGAAGCGCCGAGCGCTCGAAACAGTGCAGCGGGGATG
 GCCTTGCGCAGCGTCCCGATCGTGAACGGAGGCTTCTCCACAGGCTGCCTGTTCTGCTTGTATAGCCAT

SEQ ID NO: 47

Relevant portion of the FADc long hairpin RNA expression cassette from pSZ1471

GCTCTTCGCCGCCGCCACTCCTGCTCGAGCGCGCCGCGCGTGCGCCGCCAGCGCCTTGGCCTTTTCGCCGC
 45 GCTCGTGCGCTGCTGATGTCCATCACCAGGTCCATGAGGTCTGCCTTGCGCCGGCTGAGCCACTGCTTCG
 TCCGGGCGGCCAAGAGGAGCATGAGGGAGGACTCCTGGTCCAGGGTCTGACGTGGTTCGCGGCTCTGGGAGC
 GGGCCAGCATCATCTGGCTCTGCCGCACCGAGGCCGCTCCAACTGGTCTCTCAGCAGCCGAGTTCGCCGCC
 GACCTGGCAGAGGAAGACAGGTGAGGGGGGTATGAATTGTACAGAACAACACAGCAGCCTTGTCTAGGCAGA
 50 ATCCCTACCAGTCATGGCTTTACCTGGATGACGGCCTGCGAACAGCTGTCCAGCGACCCCTCGCTGCCGCCGC
 TTCTCCCGCACGCTTCTTTCCAGCACCGTGTGCGCGAGCCAGCGCCGCACGCTGGCGCTGCGCTTTCGCCG
 ATCTGAGGACAGTCGGGGAACCTCTGATCAGTCTAAACCCCTTGCGCGTTAGTGTGTCATCCTTTGCAGAC
 CGGTGAGAGCCGACTTGTGTGCGCCACCCCCCACACCACCTCCTCCAGACCAATTCTGTACCTTTTTTGG
 CGAAGGCATCGGCCCTCGGCTGCAGAGAGGACAGCAGTGCCAGCCGCTGGGGGTGGCGGATGCACGCTCA
 55 GGTACCCCTTCTTTCGCTATGACACTTCAGCAAAAGGTAGGGCGGGCTGCGAGACGGCTTCCCGCGCTGC
 ATGCAACACCGATGATGCTTCGACCCCCCGAAGCTCCTTCGGGGCTGCATGGGCGCTCCGATGCCGCTCCAG
 GGCGAGCGCTGTTTAAATAGCCAGGCCCCCGATTGCAAAGACATTATAGCGAGCTACCAAAGCCATATTCAA
 ACACCTAGATCACTACCACTTCTACACAGGCCACTCGAGCTTGTGATCGCACTCCGCTAAGGGGGCGCCTCT
 TCCTCTTCGTTTTCAGTCACAACCCGCAACCGCGCGCCATGCTGCTGCAGGCCTTCTGTTCTGCTGGCCG
 60 GCTTCGCCCGCCAAAGTACAGCCCTCCATGACGAACAGAGACGTCCGACCGCCCCCTGGTGCACTTACCCCCA
 ACAAGGCTGGATGAACGACCCCCAACCGGCTGTGGTACGACGAGAAGGACGCGCAAGTGGCAGCTGTACTTCC
 AGTACAACCCGAACGACACCGTCTGGGGGACGCCCTTGTCTGGGGCCACGCCACGTCCGACGACCTGACCA
 ACTGGGAGGACCAGCCCATCGCCATCGCCCCGAAGCGCAACGACTCCGGCGCCTTCTCCGGCTCCATGGTGG
 TGGACTACAACAACACCTCCGGCTTCTTCAACGACACCATCGACCCGCGCCAGCGCTGCGTGGCCATCTGGA
 65 CCTACAACACCCCGGAGTCCGAGGAGCAGTACATCTCCTACAGCCTGGACGGCGGCTACACCTTACCGAGT
 ACCAGAAGAACCCGCTGCTGGCCGCCAACTCCACCCAGTTCCGCGACCCGAAGGTCTTCTGTTGACGAGCCCT
 CCCAGAAGTGGATCATGACCGCGGCCAAGTCCGAGACTACAAGATCGAGATCTACTCTCCGACGACCTGA
 AGTCTTGAAGCTGGAGTCCGCGTTCGCCAACGAGGGCTTCTCGGCTACCACTACGAGTGCCCCGGCCTGA
 TCGAGGTCCCCACCGAGCAGGACCCAGCAAGTCTACTGGGTGATGTTTCATCTCCATCAACCCCGGCGCCC
 70 CGGCCGCGGCTCCTTCAACAGTACTTCTGTCGGCAGCTTCAACGGCACCCACTTCGAGGCCTTCGACAACC
 AGTCCCGCGTGGTGGACTTCGGCAAGGACTACTACGCCCTGCAGACCTTCTTCAACACCGACCCGACCTACG
 GGAGCGCCCTGGGCATCGCGTGGGCCTCCAACTGGGAGTACTCCGCTTCTGTGCCACCAACCCCTGGCGCT

CCTCCATGTCCCTCGTGCGCAAGTTCTCCCTCAACACCGAGTACCAGGCCAACCCGGAGACGGAGCTGATCA
ACCTGAAGGCCGAGCCGATCCTGAACATCAGCAACGCCGCCCTGGAGCCGGTTCGCCACCAACACACGT
TGACGAAGGCCAACAGCTACAACGTCGACCTGTCCAACAGCACCGGCACCCTGGAGTTTCGAGCTGGTGTACG
CCGTCAACACCACCCAGACGATCTCCAAGTCCGTGTTTCGCGGACCTCTCCCTCTGGTTCAAGGGCCTGGAGG
5 ACCCCGAGGAGTACCTCCGCATGGGCTTCGAGGTGTCGCGTCTCCTTCTTCTGGACCGCGGGAACAGCA
AGGTGAAGTTCTGTAAGGAGAACCCCTACTTCAACAAACCGCATGAGCGTGAACAACCCCTTCAAGGACG
AGAACGACCTGTCTACTACAAGGTGTACGGCTTGCTGGACCAGAACATCCTGGAGCTGTACTTCAACGACG
GCGACGTCGTGTCCACCAACACCTACTTTCATGACCACCGGGAACGCCCTGGGCTCCGTGAACATGACGACGG
GGGTGGACAACCTGTTCTACATCGACAAGTTCCAGGTGCGCGAGGTCAAGTGACAATTGGCAGCAGCAGCTC
10 GGATAGTATCGACACACTCTGGACGCTGGTGTGTGATGGACTGTTGCCGCCACACTTGCTGCCTTGACCTG
TGAATATCCCTGCCGCTTTTATCAAACAGCCTCAGTGTGTTTGATCTTGTGTGTACGCGCTTTTGCGAGTTG
CTAGCTGCTTGTGCTATTGTGCGAATACCACCCCGAGCATCCCTTCCCTCGTTTCATATCGCTTGGCATCCCA
ACCGCAACTTATCTACGCTGTCTGTCTATCCCTCAGCGCTGCTCCTGCTCCTGCTCACTGCCCCCTCGCACAG
CCTTGTTTTGGGCTCCGCTGTATTCTCCTGGTACTGCAACCTGTAAACCAGCACTGCAATGCTGATGCACG
15 GGAAGTAGTGGGATGGGAACACAAATGGAGGATCCCGCGTCTCGAACAGAGCGCGCAGAGGAACGCTGAAGG
TCTCGCCTCTGTGCGACCTCAGCGCGGCATACACCACAATAACCACCTGACGAATGCGCTTGTTCTTCGTC
CATTAGCGAAGCGTCCGGTTCACACACGTGCCACGTTGGCGAGGTGGCAGGTGACAATGATCGGTGGAGCTG
ATGGTCGAAACGTTACAGCCTAGGGATATCGAATTCCTTTCTTGCGCTATGACACTTCCAGCAAAAGGTAG
GGCGGGCTGCGAGCAGGCTTCCCGGCGCTGCATGCAACACCGATGATGCTTCGACCCCCGAAGCTCCTTCG
20 GGGCTGCATGGGCGCTCCGATGCCGCTCCAGGGCGAGCGCTGTTTAAATAGCCAGGCCCCCGATTGCAAAGA
CATTATAGCGAGCTACCAAAGCCATATTCAAACACCTAGATCACTACCACTTCTACACAGGGCCACTCGAGCT
TGTGATCGCACTCCGCTAAGGGGGGCGCCTCTTCTCTCTCGTTTTAGTCACAACCCGCAAAACACTAGTTCACT
TGTGGAACAGAGCGCGGAGTCGTCTCGGGCGGTCCGGGACGACGTAGCGGCAGTCGCGCCAGTCTCCCTCC
25 ACAGGGCGCGGACCCAGCGGCTGTCCGACACCTGCAAGTTCCTTGGCCAGGATGGGCTGATGGCTGGCAGG
CCTCCTCGGCGTGGTAGTGCGGGATGGTGTGTAAGAGGTGGTGCAGCACGTGGGTGTGCGAGATGTGGTGCA
GGATGTTGTCCATGAAGGGCGGGCCCATGGAGCGGTCCACGGTGGCCATGGCGCCGCGCAGCCAGTCCCAGT
CCTTCTCGAAGTAGTGCGGCAGCGCCGGGTGCGTGTGCTGGAGCAGCGTGATGAGCACGAGCCACATGTTCA
CGATCAGGTAGGGCACCACGTAGGTCTTGACCAGCCAGGCCCAGCCCATGGTGCGGCCAGCACGCTGAGCC
CGCTGAGCACCGCCACCAGCGCCAGGTCCGAGATGACCACCTCGATGCGCTCGCGCTTGTGTAAGATGGGCG
30 ACCACGGGTCAAAGTGGTTGGCGAAGCGCGGGTACGGCCGCGAGGCGACGTTGAACATGAGGTACAGCGGCC
AGCCCAGGGTCAGGGTGACCAGCACCTTGCCCATGCGGATGGGCAGCCACTCCTCCCACTCCAGGCCCTCGT
GCGCCACTGCGCGGTGCGGCGGCACAAACACCTCGTCTTGTCCAGGCACCCCGTGTGGAGTGGTGGCGGC
GGTGCGAGTGCTTCCAGGAGTAGTAGGGCACCAGCAGCAGGCTGTGGAACACCAGGCCACGCCGTGTTGA
TGGCCTGGCTGGAGGAAAAGGCCCTGGTGGCCGCACTCGTGCGCGCACACCCAGACACCCGTGCCGAAGGCGC
35 CCTGGAAGGTGCCGGGAGGGTAGCCGGGGGGGTGAGCGGGACAGGCGCCTTCTCCTCCATGCGCGCCACCAG
GACCTCAATACGGGCAACCAAAACCCCTCAAACACACCTGGAAGAACAGTAGAGCGGCCACATGATGCCGTA
CTTGACCCACGTAGGCACCGGTGCAGGGTTCGATGTACGTGACGCGACGTAGAGCAGGGACATGACCGCGAT
GTCAAAGGCCAGGTACATGCTGCTACGAAGCGCCGAGCGCTCGAAACAGTGCGCGGGGATGGCCTTGCGCAG
CGTCCCGATCGTGAACGGAGGCTTCTCCACAGGCTGCCTGTTCTGCTTGTATAGCCATCTCGAGGCAGCAGCA
40 GCTCGGATAGTATCGACACACTCTGGACGCTGGTGTGTGATGGACTGTTGCCGCCACACTTGCTGCCTTGA
CCTGTGAATATCCCTGCCGCTTTTATCAAACAGCCTCAGTGTGTTTGATCTTGTGTGTACGCGCTTTTGCGA
GTTGCTAGCTGCTTGTGCTATTGTGCGAATACCACCCCGAGCATCCCTTCCCTCGTTTCATATCGCTTGCAT
CCCAACCGCAACTTATCTACGCTGTCTGTCTATCCCTCAGCGCTGCTCCTGCTCCTGCTCACTGCCCCCTCGC
45 ACAGCCTTGGTTTTGGGCTCCGCTGTATTCTCCTGGTACTGCAACCTGTAAACCAGCACTGCAATGCTGATG
CACGGGAAGTAGTGGGATGGGAACACAAATGGAAAGCTGTAGAGCTCTTGTTTTCCAGAAGGAGTTGCTCCT
TGAGCCTTTTCATTTCTCAGCCTCGATAACCTCCAAAGCCGCTCTAATTGTGGAGGGGGTTGCAATTTAAAAGC
TTGGAATGTTGGTTCGTGCGTCTGGAACAAGCCAGACTTGTGCTCACTGGGAAAAGGACCATCAGCTCCA
50 AAAAACTTGCCGCTCAAACCGCGTACCTCTGCTTTCGCGCAATCTGCCCTGTTGAAATCGCCACCACATTCA
TATTGTGACGCTTGAGCAGTCTGTAATTGCCTCAGAATGTGGAATCATCTGCCCCCTGTGCGAGCCCATGCC
AGGCATGTCGCGGGCGAGGACACCCGCCACTCGTACAGCAGACCATTATGCTACCTCACAATAGTTTCATAAC
AGTGACCATATTTCTCGAAGCTCCCCAACGAGCACCTCCATGCTCTGAGTGGCCACCCCCCGGCCCTGGTGC
TTGCGGAGGGCAGGTCAACCGGCATGGGGCTACCGAAATCCCCGACCGGATCCCACCACCCCCGCGATGGGA
AGAATCTCTCCCGGGATGTGGGCCACACCAAGCACAACCTGCTGGCCAGGCGAGCGTCAAACCATACCA
55 CACAAATATCCTTGGCATCGGCCCTGAATCTTCTGAGGCTCTGCTACCCGGTGCTTCTGTCGCGCAGG
GGTTGCTAGGGATCGCTCCGAGTCCGCAAAACCCCTTGTCGCGTGGCGGGGCTTGTTCGAGCTTGAAGAGC

SEQ ID NO: 48

5' donor DNA sequence of *Prototheca moriformis* FATA1 knockout homologous recombination targeting construct

GCTCTTCGGAGTCACTGTGCCACTGAGTTCGACTGGTAGCTGAATGGAGTCGCTGCTCCACTAAACGAATTG
TCAGCACCGCCAGCCGGCCGAGGACCCGAGTCATAGCGAGGGTAGTAGCGCGCCATGGCACCGACAGCCTG
CTTGCCAGTACTGGCGTCTCTTCCGCTTCTCTGTGGTCTCTGCGCGCTCCAGCGCGTGCCTTTTCCGGTG
60 GATCATGCGGTCCGTGGCGCACCGCAGCGGCCGCTGCCATGCAGCGCCGCTGCTTCCGAACAGTGGCGGT
AGGGCCGACCCGCGTAGCCGTCCGTCCGGAACCCGCCCAAGAGTTTTGGGAGCAGCTTGAGCCCTGCAAG
ATGGCGGAGGACAAGCGCATCTTCTGGAGGAGCACCGGTGCGTGGAGGTCCGGGGCTGACCGCCGTGCGA
TTCAACGTAATCAATCGCATGATGATCAGAGGACACGAAGTCTTGGTGGCGGTGGCCAGAAACACTGTCCAT
TGCAAGGGCATAGGGATGCGTTCTTACCTCTCATTTCTCATTTCTGAATCCCTCCCTGCTCACTCTTTCT
70 CCTCCTCCTTCCCGTTCACGCAGCATTCGGGGTACC

SEQ ID NO: 49

3' donor DNA sequence of *Prototheca moriformis* FATA1 knockout homologous recombination targeting construct

GACAGGGTGGTTGGCTGGATGGGGAAACGCTGGTCGCGGGATTGATCCTGCTGCTTATATCCTCCCTGGAA
 GCACACCCACGACTCTGAAGAAGAAAACGTGCACACACAACCCAAACCGGCCGAATATTTGCTTCCTTATC
 CCGGGTCCAAGAGAGACTGCGATGCCCGTCAATCAGCATCCTCCTCCCTGCCGCTTCAATCTTCCCTGCT
 TGCCTGCGCCCGCGGTGCGCCGTCTGCCCGCCAGTCAGTCACTCCTGCACAGGCCCTTGTGCGCAGTGCT
 CCTGTACCCCTTTACCGCTCCTTCCATTCTGCGAGGCCCTTATTGAATGTATTGCTTGCCTGTGTGGCCAAG
 CGGGCTGCTGGGCGCGCCGCGTCTGGGCAGTGCTCGGCGACTTTGGCGGAAGCCGATTGTTCTTCTGTAAGC
 CACGCGCTTGCTGCTTTGGGAAGAGAAGGGGGGGGGTACTGAATGGATGAGGAGGAGAAGGAGGGGTATTGG
 TATTATCTGAGTTGGGTGAAGAGC

SEQ ID NO: 50

GCTCTTCGGAGTCACTGTGCCACTGAGTTCGACTGGTAGCTGAATGGAGTCGCTGCTCCACTAAACGAATTG
 TCAGCACCGCCAGCCGGCCGAGGACCCGAGTCATAGCGAGGGTAGTAGCGCGCCATGGCACCGACCGAGCCTG
 CTTGCCAGTACTGGCGTCTCTTCCGCTTCTCTGTGGTCTCTGCGCGCTCCAGCGCGTGCCTTTTCCGGTG
 GATCATGCGGTCCGTGGCGCACCGCAGCGGCCGTGACCATGACGCGCCGTGCTTCCGAACAGTGGCGGTC
 AGGGCCGCACCCGCGGTAGCCGTCCGTCCGGAACCCGCCAAGAGTTTGGGAGCAGCTTGAGCCCTGCAAG
 ATGGCGGAGGACAAGCGCATCTTCTGGAGGAGCACCGGTGCGTGGAGGTCCGGGGCTGACCGGCCGTGCGA
 TTCAACGTAATCAATCGCATGATGATCAGAGGACACGAAGTCTTGGTGGCGGTGGCCAGAAACACTGTCCAT
 TGCAAGGGCATAGGGATGCGTTCCTTACCTCTCATTTCTCATTTCTGAATCCCTCCCTGCTCACTCTTTCT
 CCTCCTCCTTCCCGTTACGCGAGCATTCGGGGTACCCTTTCTTGCCTATGACACTTCCAGCAAAAGGTAGG
 GCGGGCTGCGAGACGGCTTCCCGGCGCTGCATGCAACACCGATGATGCTTGCACCCCGAAGCTCCTTCGG
 GGCTGCATGGGCGCTCCGATGCCGCTCCAGGGCGAGCGCTGTTAAATAGCCAGGCCCGGATTGCAAAGAC
 ATTATAGCGAGCTACCAAAGCCATATTCAAACACCTAGATCACTACCACTTCTACACAGGCCACTCGAGCTT
 GTGATCGCACTCCGCTAAGGGGGCGCCTCTTCTCTTCTGTTTCACTCACAACCCGCAACCGGCGCGCCATGC
 TGCTGCAGGCCTTCTCTGTTCTGCTGGCCGGCTTCGCCGCCAAGATCAGCGCCTCCATGACGAACGAGACGT
 CCGACCGCCCCCTGGTGCATTCACCCCAACAAGGGCTGGATGAACGACCCCAACGGCCTGTGGTACGACG
 AGAAGGACGCCAAGTGGCACCTGTACTTCAAGTACAAGCCGACGACCCGTGTTGGGGGACGCCCTTGTCT
 GGGGCCACGCCACGTCCGACGACCTGACCAACTGGGAGGACCAGCCCATCGCCATCGCCCCGAAGCGCAACG
 ACTCCGGCGCCTTCTCCGGCTCCATGGTGGTGGACTACAACAACACCTCCGGCTTCTTCAACGACACCATCG
 ACCCGCGCCAGCGCTGCGTGGCCATCTGGACCTACAACACCCCGAGTCCGAGGAGCAGTACATCTCCTACA
 GCCTGGACGGCGGCTACACCTTACCGAGTACCAGAAGAACCCCGTGTGGCCGCCAACTCCACCCAGTTCC
 GCGACCCGAAGTCTTCTGGTACGAGCCCTCCAGAAGTGGATCATGACCGCGGCCAAGTCCCAGGACTACA
 AGATCGAGATCTACTCCTCCGACGACCTGAAGTCTTGAAGCTGGAGTCCGCGTTCCGCCAACGAGGGCTTCC
 TCGGCTACAGTACGAGTGCCCCGGCCTGATCGAGGTCCCCACCGAGCAGGACCCAGCAAGTCTACTGGG
 TGATGTTTCATCTCCATCAACCCCGGCGCCCCGGCGCGGCTCCTTCAACAGTACTTCTGTCGGCAGCTTCA
 ACGGCACCCACTTCGAGGCCTTCGACAACAGTCCCGCGTGGTGGACTTCGGCAAGGACTACTACGCCCTGC
 AGACCTTCTTCAACACCGACCCGACCTACGGGAGCGCCCTGGGCATCGCGTGGGCCTCCAACCTGGGAGTACT
 CCGCCTTCGTGCCCACCAACCCCTGGCGCTCCTCCATGTCCCTCGTGCAGAGTTCTCCCTCAACACCGAGT
 ACCAGGCCAACC CGGAGACGGAGCTGATCAACCTGAGCTCCGAGCCGATCCTGAACATCGACCCCGGCC
 CCTGGAGCCGGTTCCGCCACCAACACCACGTGACGAAGGCCAACAGCTACAACGTGACCTGTCCAACAGCA
 CCGGCACCCCTGGAGTTCGAGCTGGTGTACGCCGTCAACACCACCCAGACGATCTCCAAGTCCGTGTTCCGCG
 ACCTCTCCCTCTGTTCAAGGGCCTGGAGGACCCCGAGGAGTACCTCCGATGGGCTTCGAGGTGTCCGCGT
 CCTCCTTCTTCTGGACCGCGGAACAGCAAGGTGAAGTTCTGTAAGGAGAACCCCTACTTCAACCAACCGCA
 TGAGCGTGAACAACAGCCCTTCAAGAGCGAGAACGACCTGTCTACTACAAGGTGTACGGCTTGTGGACC
 AGAACATCCTGGAGCTGTACTTCAACGACGGCGACGTGCTGTCCACCAACACCTACTTCATGACACCGGGA
 ACGCCCTGGGCTCCGTGAACATGACGACGGGGGTGGACAACCTGTTCTACATCGACAAGTTCAGGTGCGCG
 AGGTCAAGTGACAATTGGCAGCAGCAGCTCGGATAGTATCGACACACTCTGGACGCTGGTGTGTGATGGAC
 TGTGCGGCCACACTTGCTGCCTTGACCTGTGAATATCCCTGCCGCTTTTATCAAACAGCCTCAGTGTGTTT
 GATCTTGTGTGTACGCGCTTTTTCGAGTTGCTAGCTGCTTGTGCTATTTGCGAATACCACCCCCAGCATCCC
 CTTCCCTCGTTTCATATCGCTTGCATCCCAACCGCACTTATCTACGCTGTCTGTCTATCCCTCAGCGCTGC
 TCCGTCTCCTTCACTGCCCTCGCACAGCCTTGGTTTGGCTCCGCTCCGCTGTATTCTCCTGGTACTGCAACC
 TGTAACACGCACTGCAATGCTGATGCACGGGAAGTAGTGGGATGGGAACACAAATGGAGGATCGTAGAGCT
 CACTAGTATCGATTTTCAAGACAGGGTGGTGGCTGGATGGGGAAACGCTGGTTCGCGGGATTGATCCTGCT
 GCTTATATCCTCCCTGGAAGCACACCCACGACTCTGAAGAAGAAAACGTGCACACACACAACCCAAACCGGCC
 GAATATTTGCTTCTTATCCCGGGTCCAAGAGAGACTGCGATGCCCCCTCAATCAGCATCCTCCTCCCTGC
 CGCTTCAATCTTCCCTGCTTGCCTGCGCCCGGTGCGCCGTGCCCCGCCAGTCAGTCACTCCTGACAG
 GCCCCTGTGCGCAGTGCTCCTGTACCTTTACCGTCTTCCATTCTGCGAGGGCCCCCTATTGAATGTATT
 CGTTGCGCTGTGTGGCCAAGCGGGCTGCTGGGCGCGCCCGCTCGGGCAGTGCTCGGCGACTTTGGCGGAAGC
 CGATTGTTCTTCTGTAAGCCACGCGCTTGTGCTTTGGGAAGAGAAGGGGGGGGGTACTGAATGGATGAGGA
 GGAGAAGGAGGGGTATTGGTATTATCTGAGTTGGGTGAAGAGC

SEQ ID NO: 51

GCTCTTCGGAGTCACTGTGCCACTGAGTTCGACTGGTAGCTGAATGGAGTCGCTGCTCCACTAAACGAATTG
 TCAGCACCGCCAGCCGGCCGAGGACCCGAGTCATAGCGAGGGTAGTAGCGCGCCATGGCACCGACCGAGCCTG
 CTTGCCAGTACTGGCGTCTCTTCCGCTTCTCTGTGGTCTCTGCGCGCTCCAGCGCGTGCCTTTTCCGGTG
 GATCATGCGGTCCGTGGCGCACCGCAGCGGCCGTGCCCATGACGCGCCGTGCTTCCGAACAGTGGCGGTC
 AGGGCCGCACCCGCGGTAGCCGTCCGTCCGGAACCCGCCAAGAGTTTGGGAGCAGCTTGAGCCCTGCAAG

ATGGCGGAGGACAAGCGCATCTTCTGGAGGAGCACCGGTGCGTGGAGGTCCGGGGCTGACCGGCCGTGCGA
 TTCAACGTAATCAATCGCATGATGATCAGAGGACACGAAGTCTTGGTGGCGGTGGCCAGAAACACTGTCCAT
 TGCAAGGGCATAGGGATGCGTTCCCTTACCTCTCATTTCTCATTTCTGAATCCCTCCCTGCTCACTCTTTCT
 5 CCTCCTCCTTCCCCTTACGCGAGCATTCGGGGTACCCTTTCTTGCCTATGACACTTCCAGCAAAAGGTAGG
 GCGGGCTGCGAGACGGCTTCCCGGCGCTGCATGCAACACCGATGATGCTTCGACCCCCCGAAGCTCCTTCGG
 GGCTGCATGGGCGCTCCGATGCCGCTCCAGGGCGAGCGCTGTTTAAATAGCCAGGCCCGGATTGCAAGAAC
 ATTATAGCGAGCTACCAAAGCCATATTCAAACACCTAGATCACTACCACTTCTACACAGGCCACTCGAGCTT
 GTGATCGCACTCCGCTAAGGGGGCGCTTCTCCTCTTCGTTTCAGTCACAACCCGCAACCGCGCGCCATGC
 10 TGCTGCAGGCCCTTCTGTTCTGCTGGCCGGCTTCGCCGCCAAGATCAGCGCCTCCATGACGAACGAGACGT
 CCGACCGCCCCCTGGTGCATTCACCCCCAACAAAGGGCTGGATGAACGACCCCCAACGGCCTGTGGTACGACG
 AGAAGGACGCCAAGTGGCACCTGTACTTCCAGTACAACCCGAACGACACCGTCTGGGGGACGCCCTTGTCT
 GGGGCGACGCCACTGCCGACGACCTGACCAACTGGGAGGACCAGCCCATCGCCATCGCCCCGAAGCGCAACG
 ACTCCGGCGCCTTCTCCGGCTCCATGGTGGTGGACTACAACAACACCTCCGGCTTCTTCAACGACACCATCG
 15 ACCCGCGCCAGCGCTGCGTGGCCATCTGGACCTACAACACCCCGAGTCCGAGGAGCAGTACATCTCCTACA
 GCCTGGACGGCGGCTACACCTTACCGAGTACCAGAAGAACCCCGTGTGGCCGCCAACTCCACCCAGTTCC
 GCGACCCGAAGTCTTCTGGTACGAGCCCTCCAGAAAGTGGATCATGACCGCGGCCAAGTCCCAGGACTACA
 AGATCGAGATCTACTCCTCCGACGACCTGAAGTCTTGAAGCTGGAGTCCGCGTTTCGCCAACGAGGGCTTCC
 TCGGTACAGTACGATGCCCGCGCTGATCGAGGTCCACCCAGCAGGACCCCAAGTCTTACTTCTAGTGGG
 20 TGATGTTTCATCTCCATCAACCCCCGCGCCCCGGCGCGGCTCCTTCAACAGTACTTTCGTGCGCAGCTTCA
 ACGGCACCCACTTCGAGGCCCTTCGACAACCAAGTCCCGCTGGTGGACTTCGGAAGGACTACTACGCCCTGC
 AGACCTTCTTCAACACCGACCCGACCTACGGGAGCGCCCTGGGCATCGCGTGGGCCTCCAAGTGGGAGTACT
 CCGCCTTCTGTGCCACCAACCCCTGGCGCTCCTCCATGTCCCTCGTGCAGCAAGTCTCCCTCAACACCGAGT
 ACCAGGCCAACCCGAGACGGAGCTGATCAACCTGAAGGCCGAGCCGATCCTGAACATCAGCAACCGCGGCC
 25 CCGGACCGGCTCCGTGAACATGACGACGGGGGTGGACCCACCTGTTCTACATCGACAAGTTCAGGTGCGCG
 CCGGCACCCCTGGAGTTCGAGCTGGTGTACGCCGTCAACACCAACCCAGACGATCTCCAAGTCCGTGTTTCGCG
 ACCTCTCCCTCTGGTTCAAGGGCCTGGAGGACCCCGAGGAGTACCTCCGATGGGCTTCGAGGTGTCCGCGT
 CCTCCTTCTTCTGGACCGCGGAACAGCAAGGTGAAGTTCGTGAAGGAGAACCCCTACTTCACCAACCGCA
 TGAGCGTGAACAACAGCCCTTCAAGAGCGAGAACGACCTGTCTACTACAAGGTGTACGGCTTGCTGGACC
 30 AGAATCATCTGGAGCTGTACTTCAACGACGGCGAGCTCGTGTCCACCAACACCTACTTCATGACCAACCGGA
 ACGCCTCGGCTCCGTGAACATGACGACGGGGGTGGACCCACCTGTTCTACATCGACAAGTTCAGGTGCGCG
 AGGTCAAGTGACAATTGGCAGCAGCAGCTCGGATAGTATCGACACACTCTGGACGCTGGTGTGTGATGGAC
 TGTTCGCCGCACACTTGCTGCCTTGACCTGTGAATATCCCTGCCGCTTTTATCAAACAGCCTCAGTGTGTTT
 GATCTTGTGTGTACGCGCTTTTTCGAGTGTGTAGCTGCTTGTGCTATTTGCGAATACCACCCCCAGCATCCC
 35 CTTCCTCTCGTTTCATATCGCTTGCATCCCAACCGCAACTTATCTACGCTGTCTCTGCTATCCCTCAGCGCTGC
 TCCTGCTCCTGCTCACTGCCCTTCGCACAGCCTTGGTTTGGGCTCCGCTGTATTCTCCTGGTATCGCAACC
 TGTAAACAGCACTGCAATGCTGATGCACGGGAAGTAGTGGATGGGAACACAATGGAGGATCCCGCTCT
 CGAACAGAGCGCGCAGAGGAACGCTGAAGGTCTCGCCTCTGTGCACCTCAGCGCGGCATACACCACAATAA
 CCACCTGACGAATGCGCTTGGTCTTTCGTCCATTAGCGAAGCGTCCGTTTACACACGTGCCACGTTGGCGA
 40 GGTGGCAGGTGACAATGATCGGTGGAGCTGATGGTCAAACGTTTACAGCCTAGGGATATCGAATTTCGGCCG
 ACAGGACGCGCGTCAAAGGTGCTGGTGTGTATGCCCTGGCCGGCAGGTGCTTGTGCTGCTGGTTAGTGAT
 TCCGCAACCCGTATTTTGGCGCTTATTTTGGCGTGGCAACGCTGGCGCCCGGAGCGGGGCGCGGCGGA
 TCGGCTGCCACCGCTGCCGGAATCCAAGGGAGGCAAGAGCGCCCGGTGAGTTGAAGGTTTACGCGCA
 AGGTACAGCCGCTCCTGCAAGGCTGCGTGGTGGAAATGGACGTGCAGGTCTGCTGAAGTTCCTCCACCGCC
 45 TCACCAGCGGACAAAGCACCGGTGTATCAGGTCCGTGTATCCACTCTAAAGAGCTCGACTACGACCTACTG
 ATGGCCCTAGATTCTTTCATCAAAAACGCTGAGACACTTGGCCAGGATTGAACTCCCTGAAGGGACCACCA
 GGGGCCCTGAGTTGTTTCTTCCCCCGTGGCGAGCTGCCAGCCAGGCTGTACCTGTGATCGAGGCTGGCGGG
 AAAATAGGCTTCGTGTGCTCAGGTCTAGGTGAGGTGCAGGACAGCTCATGAAACGCCAACATCGCACAAATTC
 50 ATGTCGAAGCTAATCAGCTATTTCTCTTACAGAGCTGTAATTGTCCAAAATTCTGGTCTACCGGGGTGAT
 CCTTCGTGTACGGGCCCTTCCCTCAACCTAGGTATGCGCGCATGCGGTGCGCGCGCAACTCGCGCGAGGGC
 CGAGGGTTTGGGACGGGCCGTCCCGAAATGCAGTTGCACCCGGATGCGTGGCACCTTTTTTTCGATAATTTA
 TGCAATGGACTGCTCTGCAAAATTCTGGCTCTGTGCGCAACCTAGGATCAGCGGCGTAGGATTTTCGTAATC
 ATTCGTCTGATGGGGAGCTACCGACTACCTAATATCAGCCCCGACTGCCTGACGCCAGCGTCCACTTTTGT
 GCACACATTCATTCGTGCCAAGACATTTTCATTGTGGTGCGAAGCGTCCCCAGTTACGCTCACCTGTTTCC
 55 CGACCTTCTTCTGTCGACAGAGCGGGCCACAGGCGCGTCCGAGCTAGTATGTTGTTGTCGCGC
 CGCCGCGCAGCAGCGCTTCTTCCCCGTGCCCGCCCCCGCCCCAACCCCAAGCCCGGCAAGTTTCGGCAACTG
 GCCCAGCAGCCTGAGCCAGCCCTTCAAGCCCCAAGAGCAACCCCAACGGCCGCTTCCAGGTGAAGGCCAACGT
 GAGCCCCACGGGCGCGCCCCAAGGCCAACGGCAGCGCCGTGAGCCTGAAGTCCGGCAGCCTGAACACCCCT
 GGAGGACCCCCCAGCAGCCCCCCCCCCCCGACCTTCTTGAACAGCTGCCCCGACTGGAGCCGCTGCGCAC
 60 CGCCATCACCACCGTGTTCGTGGCCCGCAGAACGAGTTCACCCGCTGGACCGCAAGAGCAAGCGCCCCGA
 CATGCTGGTGGACTGGTTTCGGCAGCAGACCATCGTGCAGGACGGCCTGGTGTTCGCGAGCGCTTCAGCAT
 CCGCAGCTACGAGATCGGCGCCGACCGCACCGCCAGCATCGAGACCTGATGAACCACCTGCAGGACACCAG
 CCTGAACCACTGCAAGAGCGTGGGCCTGCTGAACGACGGCTTCGGCCGCACCCCCGAGATGTGCACCCGCGA
 CCTGATCTGGGTGCTGACCAAGATGCAGATCGTGGTGAACCGCTACCCACCTGGGGCGACACCGTGGAGAT
 65 CAACAGCTGGTTACGCCAGAGCGGCAAGATCGGCATGGGCCGCGAGTGGCTGATCAGCGACTGCAACACCGG
 CGAGATCCTGGTGGCGGCCACAGCGCTGGGCCATGATGAACAGAACAGACCCGCGCTTCAGCAAGCTGCC
 CTGCGAGGTGCGCCAGGAGATCGCCCCCACTTCTGTGACGACGCCCCCGGTGATCGAGGACAACGACCGCAA
 GCTGCACAAGTTTCGACGTGAAGACCGGCGACAGCATCTGCAAGGGCCTGACCCCGGCTGGAACGACTTCGA
 CGTGAACCAAGCAGTGCAGCAACGTGAAGTACATCGGCTGGATTCTGGAGAGCATGCCACCGAGGTGCTGGA
 70 GACCCAGGAGCTGTGCAGCCTGACCTGGAGTACCGCCGCGAGTGCAGCCGCGAGAGCGTGGTGGAGAGCGT
 GACCAGCATGAACCCAGCAAGGTGGGCGACCGCAGCCAGTACCAGCACCTGCTGCGCCTGGAGGACGGCGC
 CGACATCATGAAGGGCCGACCGAGTGGCGCCCCAAGAACGCGGCGACCAACCGCGCCATCAGCACCTGATT

AATTAACCTCGAGGCAGCAGCAGCTCGGATAGTATCGACACACTCTGGACGCTGGTCTGTGATGGACTGTTG
 CCGCCACACTTGTGTCCTTGACCTGTGAATATCCCTGCCGCTTTTATCAAACAGCCTCAGTGTGTTTGATCT
 TGTGTGTACGCGCTTTTTCGAGTTGCTAGCTGCTTGTGCTATTTGCGAATACCACCCCCAGCATCCCCCTCC
 5 CTGCTTTTCATATCGCTTGCATCCCAACCGCAACTTATCTACGCTGTCTGCTATCCCTCAGCGCTGCTCCTG
 CTCTGCTCACTGCCCCCTCGCACAGCCTTGGTTTGGGCTCCGCCTGTATTCTCCTGGTACTGCAACCTGTAA
 ACCAGCACTGCAATGCTGATGCACGGGAAGTAGTGGGATGGGAACACAAATGGAAAGCTTGAGCTCTTGTTT
 TCCAGAAGGAGTTGCTCCTTGAGCCTTTTCATTCTCAGCCTCGATAACCTCCAAAGCCGCTCTAATTGTGGAG
 GGGGTTTCAAGACAGGGTGGTTGGCTGGATGGGGAAACGCTGGTCGCGGGATTTCGATCCTGCTGCTTATATC
 10 CTCCCTGGAAGCACACCCACGACTCTGAAGAAGAAAACGTGCACACACACAACCCAACCGGCCGAATATTTG
 CTTCCTTATCCCGGGTCCAAGAGAGACTGCGATGCCCCCTCAATCAGCATCCTCCTCCCTGCCGCTTCAAT
 CTTCCTTGCTTGCTGCGCCCGCGGTGCGCCGTCTGCCCGCCAGTCAGTCACTCCTGCACAGGCCCTTGT
 GCGCAGTGTCTACCTTTACCGCTCCTTCCATTCTGCGAGGCCCTTATTGAATGTATTCTGTTGCCGTG
 TGTGGCCAAGCGGGCTGCTGGGCGCGCCGCGCTCGGGCAGTGCTCGGCGACTTTGGCGGAAGCCGATTGTTT
 15 TTCTGTAAGCCACGCGCTTGTGCTTTGGGAAGAGAAGGGGGGGGTACTGAATGGATGAGGAGGAGAAGGA
 GGGGTATTGGTATTATCTGAGTTGGGTGAAGAGC

SEQ ID NO: 52

Codon optimized sequence for *Cuphea wrightii* FatB2 (CwTE2) thioesterase

20 ATGGTGGTGGCCGCCGCCAGCAGCGCCTTCTTCCCCGTGCCCGCCCCCGCCCCACCCCCAAGCCCGGC
 AAGTTCGGCAACTGGCCCGAGCAGCCTGAGCCAGCCCTTCAAGCCCAAGAGCAACCCCCAACGGCCGCTTCCAG
 GTGAAGGCCAACGTGAGCCCCACGGGCGCGCCCCAAGGCCAACGGCAGCGCCGTGAGCCTGAAGTCCGGC
 AGCCTGAACACCTGGAGGACCCCCCAGCAGTCCCCCCCCCGCCACCTTCTGAACAGTGGCCGATG
 25 AGCCGCTGCGCACCGCCATCACCACCGTGTTCGTGGCGCCGAGAAGCAGTTACCCGCTGGACCGCAAG
 AGCAAGCGCCCCGACATGCTGGTGGACTGGTTTCGGCAGCGAGACCATCGTGCAGGACGGCCTGGTGTTCGC
 GAGCGCTTACGATCCGAGCTACGAGATCGGCGCCGACCGCACCGCCAGCATCGAGACCCTGATGAACCAC
 CTGCAGGACACCAGCCTGAACCACTGCAAGAGCGTGGGCCTGCTGAACGACGGCTTCGGCCGACCCCCGAG
 ATGTGCACCCGCGACCTGATCTGGGTGCTGACCAAGATGCAGATCGTGGTGAACCGCTACCCCCACCTGGGGC
 GACACCGTGGAGATCAACAGCTGGTTTCAGCCAGAGCGGCAAGATCGGCATGGGCCCGGATGGCTGATCAGC
 30 GACTGCAACACCGCGAGATCCTGGTGCAGCGCCACCAGCGCCTGGGCCATGATGAACCAGAAGACCCCGCGC
 TTCAGCAAGCTGCCCTGCGAGGTGCGCCAGGAGATCGCCCCCACTTCTGTGGACGCCCCCCCCCGTGTATCGAG
 GACAACGACCGCAAGCTGCACAAGTTCGACGTGAAGACCGGCGACAGCATCTGAAGGGCCTGACCCCCGGC
 TGGAACGACTTCGACGTGAACCAGCAGCTGAGCAACGTGAAGTACATCGGCTGGATTCTGGAGAGCATGCCC
 ACCGAGGTGCTGGAGACCCAGGAGCTGTGCAGCCTGACCCTGGAGTACCGCCGCGAGTGCGGCCGCGAGAGC
 35 GTGGTGGAGAGCGTGACAGCATGAACCCAGCAAGGTGGGCGACCGCAGCCAGTACCAGCACCTGCTGCGC
 CTGGAGGACGGCGCCGACATCATGAAGGGCCGACCGAGTGGCGCCCCAAGAACGCGGCACCAACCGCGCC
 ATCAGCACCTGA

SEQ ID NO: 53

40 Protein sequence for *Cuphea wrightii* FatB2 (CwTE2) thioesterase; GenBank
 Accession No. U56104

MVAAAAASSAFFVPAPRPTPKPGKFGNWPSSLSQPFKPKSNPNGRFQVKANVSPHPKANGSAVSLKSGSLN
 45 TLEDPPSSPPPRFTFLNQLPDWSRLRTAITTVFAAEKQFTRLDRKSKRPDMLVDWFGSETIVQDGLVFRERF
 SIRS YEIGADRTAS IETLMNHLQD TSLNHCKSVGLLNDGFRTPMEMCTRD LIWVLTKMQIVVNRYPTWGD TV
 EINSWFSQSGKIGMGREWLISDCNTGEILVRATS AWAMNQKTRRFSKLPCEVRQE IAPHFVDAPPVIEDND
 RKLHKFDVKTGDSICKGLTPGWND FDNQHVSNVKYIGWILESMPT EVLETQELCSLTLEYRRECGRRESVVE
 SVTSMNPSKVGDRSQYQHLLRLEDGADIMKGRTEWRPKNAGTNRAIST

50 SEQ ID NO:54

Codon optimized *Chlorella protothecoides* (UTEX 250) stearyl ACP
 desaturase transit peptide cDNA sequence.

55 ACTAGTATGGCCACCGCATCCACTTTCTCGGCGTTCAATGCCCGCTGCGGCGACCTGCGTGCCTCGGCGGGC
 TCCGGGCCCCGGCGCCAGCGAGGCCCTCCCCGTGCGCGGGCGCGCC

SEQ ID. NO: 55

Codon-optimized *Glycine max* KASII

60 ATGGCCACCGCATCCACTTTCTCGGCGTTCAATGCCCGCTGCGGCGACCTGCGTGCCTCGGCGGGCTCCGGG
 CCCCCGGCGCCAGCGAGGCCCTCCCCGTGCGCGGGCGCGCCTCCACCCACTCCGGCAAGACCATGGCCGTG
 GCCCTGCAGCCACCCAGGAGATCACCACCATCAAGAAGCCCCCACCAAGCAGCGCCGCGTGGTGGTGACC
 GGCTGGGCGTGGTGACCCCCCTGGGCCACGAGCCCGACATCTTCTACAACAACCTGCTGGACGGCGCCTCC
 65 GGCATCTCCGAGATCGAGACCTTCGACTGCGCCGAGTACCCCAACCGCATCGCCGGCGAGATCAAGTCTTTC
 TCCACCGACGGTGGGTGGCCCCCAAGCTGTCCAAGCGCATGGACAAGTTTATGCTGTACATGCTGACCGCC
 GGCAAGAAGGCCCTGGTGGACGGCGGCATCACCGACGACGTGATGGACGAGCTGAACAAGGAGAAGTGCGGC
 GTGCTGATCGGCTCCGCCATGGGCGGCATGAAGGTGTTCAACGACGCCATCGAGGCCCTGCGCATCTCTAC
 AAGAAGATGAACCCCTTCTGCGTGCCCTTCGCCACCACCAACATGGGCTCCGCCATGCTGGCCATGGACCTG
 70 GGCTGGATGGGCCCCAACTACTCCATCTCCACCGCCTGCGCCACCTCCAACCTTCTGCATCCTGAACGCCGCC
 AACCACATCATCCGCGCGAGGCCGACGTGATGCTGTGCGGCGGCTCCGACGCCGCCATCATCCCCATCGGC

CTGGGCGGCTTCGTGGCCTGCCGCGCCCTGTCCAGCGCAACACCGACCCACCAAGGCCTCCCGCCCTGG
 GACATCAACCGCGACGGCTTCGTGATGGGCGAGGGCGCCGGCGTGTCTGCTGGAGGAGCTGGAGCACGCC
 AAGGAGCGCGGCGCCACCATCTACGCCGAGTTCCTGGGCGGCTCCTTCACCTGCGACGCCTACCACGTGACC
 5 GAGCCCCGCCCCGACGGCGCCGGCGTGATCCTGTGCATCGAGAAGGCCCTGGCCAGTCCGGCGTGTCCAAG
 GAGGACGTGAACTACATCAACGCCACGCCACCTCCACCCCCGCGGCGACCTGAAGGAGTACCAGGCCCTG
 ATGCACTGCTTCGGCCAGAACCCCGAGCTGGCGGTGAATCCACCAAGTCCATGATCGGCCACCTGCTGGGG
 GCGCGCGCGCGTGGAGGGCGTGCCACCATCCAGGCCATCCGCACCGGCTGGGTGCACCCCCAACATCAAC
 10 CTGGAGAACCCCGACAACGGCGTGGACGCCAAGGTGCTGGTGGGCTCCAAGAAGGAGCGCCTGGACGTGAAG
 GCCGCCCTGTCCAACCTCCTTCGGCTTCGGCGGCCACAACCTCCTCCATCATCTTCGCCCCCTACATGGACTAC
 AAGGACCACGACGGCGACTACAAGGACCACGACATCGACTACAAGGACGACGACGACAAGTGA

SEQ ID. NO: 56

Codon-optimized *Helianthus annuus* KASII

15 ATGGCCACCGCATCCACTTTCTCGGCGTTCAATGCCCGCTGCGGCGACCTGCGTGCCTCGGCGGGCTCCGGG
 CCGCGCGCCAGCGAGGCCCTCCCGCTGCGCGGGCGCGCCTGCCGCCGCGGGCGGCCGCTGGCCATGGCC
 ATCGCCATCCAGCCCTCCTCCATCGAGATGGAGGAGGAGACCACCTGACCAAGCGCAAGCAGCCCCCACC
 AAGCAGCGCCGCTGGTGGTGACCGCATGGCGGTGGAGACCCCATCGGCAACAACCCCGACAGTTCTAC
 20 AACAACTGCTGCGGGCGTGTCCGGCATCACCCAGATCGAGGCCCTTCGACTGCTCCTCCTACCCACCCGC
 ATCGCCGGCGAGATCAAGAACTTCTCCACCGACGGCTGGGTGGCCCCCAAGCTGTCCAAGCGCATGGACCGC
 TTCATGCTGTACATGCTGACCGCCGGCAAGAAGGCCCTGGCCGACGCCGGCATCTCCCCCTCCGACTCCGAC
 GAGATCGACAAGTCCCGCTGCGGCGTGTGATCGGCTCCGCCATGGGCGGCATGAAGGTGTTCAACGACGCC
 ATCGAGGCCCTGCGCGTGTCTACCGCAAGATGAACCCCTTCTGCGTGCCCTTCGCCACCACCAACATGGGC
 25 TCCGCCATGCTGGCCATGGACCTGGGCTGGATGGGCCCCAACTACTCCATCTCCACCGCCTGCGCCACCTCC
 AACTTCTGCATCCTGAACGCGGCCAACCACATCATCCGCGGCGAGGCCGACATGATGCTGTGCGGCGGCTCC
 GACGCCGTGATCATCCCCATCGGCCCTGGGCGGCTTCGTGGCCTGCCGCGCCCTGTCCGAGCGCAACACCGAC
 CCGCCAAGGCCCTCCCGCCCTGGGACTCCGGCCGCGACGGCTTCGTGATGGGCGAGGGCGCCGGCGTGTG
 CTGCTGGAGGAGCTGGAGCACGCCAAGAAGCGCGGCGCCAAGATCTACGCCGAGTTCTTGGGCGGCTCCTTC
 30 ACCTGCGACGCCCTACCACATGACCGAGCCCCACCCGAGGGCGCCGGCGTGTGCTGTGCATCGAGAAGGCC
 CTGTCCCGAGCGCGGTGCGCCGCGAGGACGTGAATACATCAACGCCACGCCACCTCCACCCCGCGCGC
 GACCTGAAGGAGTACCACGCCCTGCTGCACTGCTTCGGCAACAACAGGAGCTGCGCGTGAATCCACCAAG
 TCCATGATCGGCCACCTGCTGGGCGCCGCCGGCGCGTGGAGGCCGTGGCCACCGTGCAGGCCATCCGCACC
 GGCTGGATCCACCCCAACATCAACCTGGAGAACCCGACAGGGCGTGGACACCAAGGTGCTGGTGGGCTCC
 35 AAGAAGGAGCGCCTGAACGTGAAGGTGGGCCTGTCCAACCTCCTTCGGCTTCGGCGGCCACAACCTCCTCCATC
 CTGTTCCGCCCCCTTCCAGATGGACTACAAGGACCACGACGGCGACTACAAGGACCACGACATCGACTACAAG
 GACGACGACGACAAGTGA

SEQ ID. NO: 57

Codon-optimized *Ricinus communis* KASII

40 ATGGCCACCGCATCCACTTTCTCGGCGTTCAATGCCCGCTGCGGCGACCTGCGTGCCTCGGCGGGCTCCGGG
 CCGCGCGCCAGCGAGGCCCTCCCGCTGCGCGGGCGCGCCTCCCACTACTACTCCTCCAACGGCCTGTTT
 CCCAACACCCCCCTGCTGCCAAGCGCCACCCCGCCTGCACACCGCCTGCCCGCTCCGGCGAGGCCATG
 45 GCGGTGGCCGTGCAGCCCGAGAAGGAGGTGGCCACCAACAAGAAGCCCTGATGAAGCAGCGCCGCGTGGTG
 GTGACCGGCATGGGCGTGGTGTCCCCCTGGGCCACGACATCGACGTGTACTACAACAACCTGCTGGACGGC
 TCCTCCGGCATCTCCAGATCGACTCCTTCGACTGCGCCAGTTCCCCACCCGCATCGCCGGCGAGATCAAG
 TCCTTCTCCACCGACGGCTGGGTGGCCCCAAGCTGTCCAAGCGCATGGACAAGTTCATGCTGTACATGCTG
 50 ACCGCCGCAAGAAGGCCCTGGCCGACGGCGGCATCACCGAGGACATGATGGACGAGCTGGACAAGGCCGCG
 TGCGGCGTGTGATCGGCTCCGCCATGGGCGGCATGAAGGTGTTCAACGACGCCATCGAGGCCCTGCGCATC
 TCCTACCGCAAGATGAACCCCTTCTGCGTGCCCTTCGCCACCACCAACATGGGCTCCGCCATGCTGGCCATG
 GACCTGGGCTGGATGGGCCCAACTACTCCATCTCCACCGCCTGCGCCACCTCCAACCTTCTGCATCCTGAAC
 55 GCGCCAACACATCATCCGCGGCGAGGCCGACATCATGCTGTGCGGCGGCTCCGACGCCGCCATCATCCCC
 ATCGGCTGGGCGGCTTCGTGGCTTCCGCGCCCTGTCCAGCGCAACGACGACCCACCAAGGCCTCCCGC
 CCTGGGACATGAACCGGACGGCTTCGTGATGGGCGAGGCCGCGCGTGTGCTGCTGGAGGAGCTGGAG
 60 CACGCCAAGAAGCGCGGCGCCAACATCTACGCCGAGTTCTTGGGCGGCTCCTTCACCTGCGACGCCTACCAC
 ATGACCGAGCCCCGCCCCGACGGCGTGGGCGTGATCCTGTGCATCGAGAAGGCCCTGGCCCGCTCCGGCGTG
 TCCAAGGAGGAGGTGAATACATCAACGCCACGCCACCTCCACCCCGCGGCGACCTGAAGGAGTACGAG
 GCGCTGATGCGTGTCTTCTCCAGAACCCGACCTGCGCGTGAATCCACCAAGTCCATGATCGGCCACCTG
 CTGGGCGCCCGCGCGCGTGGAGGCCATCGCCACCATCCAGGCCATCCGCACCGGCTGGGTGACACCCCAAC
 65 ATCAACCTGGAGAACCCCGAGGAGGCGTGACCAACCAAGGTGCTGGTGGGCCCCAAGAAGGAGCGCCTGGAC
 ATCAAGGTGGCCCTGTCCAACCTCCTTCGGCTTCGGCGGCCACAACCTCCTCCATCATCTTCGCCCCCTACAAG
 ATGGACTACAAGGACCACGACGGCGACTACAAGGACCACGACATCGACTACAAGGACGACGACGACAAGTGA

SEQ ID NO: 58

Relevant portions of pSZ2041 including a codon-optimized *Prototheca morifomis* KASII

65 GCTCTTCGCGCGGCCACTCCTGCTCGAGCGCGCCGCGCGTGGCGCGCCAGCGCCTTGGCCTTTTCGCGCG
 GCTCGTGGCGTGCCTGATGTCCATCACAGGTCCATGAGGTCTGCCTTGGCGCGGCTGAGCCACTGCTTCG
 TCCGGGCGGCCAAGAGGAGCATGAGGGAGGACTCCTGGTCCAGGGTCTGACGTGGTGGCGGCTTGGGAGC
 70 GGGCCAGCATCATCTGGCTTGGCGACCGAGGCCGCTCCAACCTGGTCTCCTCAGCAGCCGAGTCGCGGCC
 GACCTGGCAGAGGAAGACAGGTGAGGGGGTATGAATTGTACAGAACAACACGAGCCTTGTCTAGGCAGA

ATCCCTACCAGTTCATGGCTTTACCTGGATGACGGCCTGCGAACAGCTGTCCAGCGACCTCGCTGCCGCCG
TTCTCCCGCACGCTTCTTTCCAGCACCGTGATGGCGCGAGCCAGCGCCGCACGCTGGCGCTGCGCTTCGCCG
ATCTGAGGACAGTCGGGGAACCTCTGATCAGTCTAAACCCCTTGCGCGTTAGTGTTGCCATCCTTTGCAGAC
CGGTGAGAGCCGACTTGTGTGCGCCACCCCCACACCACCTCCTCCAGACCAATTCTGTACCTTTTTTGG
5 CGAAGGCATCGGCCCTCGGCCGTCAGAGAGGACAGCAGTGCCCGAGCCGCTGGGGGTTGGCGGATGCACGCTCA
GGTACCTTTCTTGCGCTATGACACTTCAGCAAAAAGGTAGGGCGGGCTGCGAGACGGCTTCCCGGCTGCG
ATGCAACACCGATGATGCTTCGACCCCCCGAAGCTCCTTCGGGGCTGCATGGGCGCTCCGATGCCGCTCCAG
GGCGAGCGCTGTTTAAATAGCCAGGCCCGGATTGCAAAGACATTATAGCGAGCTACCAAAGCCATATTCAA
10 ACACCTAGATCACTACCCTTCTACACAGGCCACTCGAGCTTGTGATCGCACTCCGCTAAGGGGGCGCCTCT
TCCTCTTCGTTTCAGTCACAACCCGCAAACTCTAGAATATCAATGCTGCTGCAGGCCTTCTCTGTTCTGCTG
GCCGGCTTCGCCGCCAAGATCAGCGCCTCCATGACGAACGAGACGTCGACCGCCCCCTGGTGCACCTTACC
CCCAACAAGGGCTGGATGAACGACCCCAACGGCCTGTGGTACGAGTACGAGAAGGACACCGAGTGGACCTGTAC
TTCCAGTACAACCCGAACGACACCGTCTGGGGGACGCCCTTGTCTGGGGCCACGCCACGTCGACGACCTG
15 ACCAAGTGGGAGGACCAGCCCATCGCCATCGCCCCGAAGCGCAACGACTCCGGCGCCTTCTCCGGCTCCATG
GTGGTGGACTACAACAACACCTCCGGCTTCTTCAACGACACCATCGACCCGCGCCAGCGCTGCGTGGCCATC
TGGACCTACAACACCCCGGAGTCCGAGGAGCAGTACATCTCCTACAGCCTGGACGGCGGCTACACCTTACC
GAGTACCAGAAGAACCCCGTGCTGGCCGCCAACTCCACCCAGTTCGCGACCCGAAGGTCTTCTGGTACGAG
CCCTCCGCAAGTGGATCATGACCGCGGCAAGTCCAGGACTACAAGATCGAGATCTACTCTCCGACGAC
20 CTGAAGTCTCTGGAAGCTGGAGTCCGCGTTTCGCCAACGAGGGCTTCTCGGCTACAGTACGAGTGCCTCCG
CTGATCGAGGTCCCCACCGAGCAGGACCCAGCAAGTCTTACTGGGTGATGTTTCATCTCCATCAACCCCGGC
GCCCCGGCGCGGCTCCTTCAACACAGTACTTCTGTCGGCAGCTTCAACGGCACCCACTTCGAGGCCTTCGAC
AACCAGTCCCGCGTGGTGGACTTCGGCAAGGACTACTACGCCCTGCAGACCTTCTTCAACACCGACCCGACC
TACGGGAGCGCCCTGGGCATCGCGTGGGCTCCAACTGGGAGTACTCCGCTTCTGTCGCCACCAACCCCTGG
25 CGCTCCTCATGTCCTTACTACAAGGTGACGGCTTCTCCCTCAACAGGAGTACCAGGCCAACCCGAGCAGGATG
ATCAACCTGAAGGCCGAGCCGATCCTGAACATCAGCAACGCCGCCCCCTGGAGCCGGTTTCGCCACCAACACC
ACGTTGACGAAGGCCAACAGCTACAACGTCGACCTGTCCAACAGCACCGGCACCCCTGGAGTTTCGAGCTGGT
TACGCCGTCAACACCACCCAGACGATCTCCAAGTCCGTGTTTCGCGACCTCTCCCTCTGGTTCAAGGGCCTG
GAGGACCCCGAGGAGTACCTCCGCATGGGCTTCGAGGTGTCCGCGTCTCCTTCTTCTGACCCGCGGAAC
30 AGCAAGGTGAAGTTCTGTAAGGAGAACCCCTACTTCAACACCGCATGAGCGTGAACAACAGCCCTTCAAG
AGCGAGAACGACCTGTCTTACTACAAGGTGACGGCTTCTGCTGGACGAGAACATCTGGAGTGTCTTCAAC
GACGGCGACGTCGTGTCCACCAACACCTACTTTCATGACACCGGGAACGCCCTGGGCTCCGTGAACATGACG
ACGGGGGTGGACAACCTGTTCTACATCGACAAGTTCAGGTGCGCGAGGTCAAGTGACAATTGGCAGCAGCA
GCTCGGATAGTATCGACACACTCTGGACGCTGGTGTGTGATGAGTGTTCGCCCCACACTTGCTGCCTTGA
35 CCGTGTGAATATCCCTGCCGCTTTTATCAAAACAGCCTCAGTGTGTTTGATCTTGTGTGTACGCGCTTTTGCGA
GTTGCTAGCTGCTTGTGCTATTTGCGAATACCAACCCAGCATCCCTTCCCTCGTTTCATATCGCTTGCAT
CCCAACCGCAACTTATCTACGCTGTCTGCTATCCCTCAGCGCTGCTCCTGCTCCTGCTCACTGCCCCCTGC
ACAGCCTTGGTTTGGGCTCCGCTGTATTCTCCTGGTACTGCAACCTGTAAACAGCACTGCAATGCTGATG
CACGGGAAGTAGTGGGATGGGAACACAAATGGAGGATCCCGCTCTCGAACAGAGCGCGCAGAGGAACGCTG
40 AAGGTCTCGCCTCTGTGCGACCTCAGCGCGGCATACACCACAATAACCACCTGACGAATGCGCTTGGTTCTT
CGTCCATTAGCGAAGCGTCCGTTTACACACGCTGCCACGTTGGCGAGGTGGCAGGTGACAATGATCGGTGGA
GCTGATGGTTCGAAACGTTTACAGCCTAGGGATATCGAATTTCGGCCGACAGGACGCGCTCAAAGGTGCTGGT
CGTGTATGCCCTGGCCGCGAGGTGCTGCTGCTGCTGTTAGTGATTCCGCAACCCCTGATTCTTGGCGCTTA
TTTGGCGTGGCAAACGCTGGCGCCCGCAGCCGGGCGGCGGCGATGCGGTGCCCCACGGCTGCCGGAATC
45 CAAGGGAGGCAAGAGCGCCCGGTGAGTTGAAGGGCTTTACGCGCAAGGTACAGCCGCTCCTGCAAGGCTGC
GTGGTGGAAATTGGACGTGCAGGTCTGTGTAAGTTCTCCACCGCCTCACCAGCGGACAAAGCACCGGTGTA
TCAGGTCCGTGTATCCACTCTAAAGAACTCGACTACGACCTACTGATGGCCCTAGATTCTTCATCAAAAAC
GCCTGAGACACTTGCCCAGGATTGAAACTCCCTGAAGGGACCAACAGGGGCCCTGAGTTGTTTCTTCCCCC
GTGGCAGCTGCCAGCCAGGCTGTACCTGTGATCGAGGTGGCGGGAATAAGGCTTCTGTGTCTCAGGTCA
50 TGGGAGGTGCAGGACAGCTCATGAAACGCCAACAAATCGCACAAATTCATGTCAAGCTAATCAGCTATTTCTC
TTCACGAGCTGTAATTGTCCAAAATTCTGGTCTACCGGGGGTGATCCTTCTGTGTACGGGCCCTTCCCTCAA
CCCTAGGTATGCGCGCATGCGGTGCGCGCGCAACTCGCGCGAGGGCCGAGGGTTTGGGACGGGCGCTCCCGA
AATGCAGTTGCACCCGGATGCGTGGCACCTTTTTTTCGATAATTTATGCAATGGACTGCTCTGCAAAATTCT
GGCTCTGTGCGCAACCCCTAGGATCAGCGCGTAGGATTTCTGTAATCATTCTGCTCTGATGGGGAGCTACCGAC
55 TACCCTAATATCAGCCCGACTGCCTGACCGCAGCTCACTTTTGTGACACATTCTGATCGTCCCAAGAC
ATTTTCATTGTGGTGCAGCGTCCCCAGTTACGCTCACCTGTTTCCCGACCTCCTTACTGTTCTGTGACAG
AGCGGGCCACAGGCCGTCGCGAGCCACTAGTATGGCCACCGCATCCACTTTCTCGGCGTTCAATGCCCGCT
GCGCGACCTGCGTCTGCGCGGGCTCCGGGCCCCGCGCGCCAGCGAGGCCCCCTCCCCGTGCGCGGGCGCG
CCGCGCGCGCGCGGACGCCAACCCCGCCCCCGAGCGCGCGCTGGTGATCACCGGCCAGGGCGTGGTGA
60 CCTCCCTGGGCCAGACCATCGAGCAGTTCTACTCTCTCTGCTGGAGGGCGTGTCCGGCATCTCCGACATCC
AGAACTTCGACACACCGGCTACACCACCATCGCGCGGAGATCAAGTCCCTGCGAGCTGGACCCCTACG
TGCCCAAGCGCTGGGCCAAGCGCGTGGACGACGTGATCAAGTACGTGTACATCGCCGGCAAGCAGGCCCTGG
AGTCCGCGCGCTGCCCATCGAGGCCGCGGCTGGCCGGCGCGGCGCTGGACCCCGCCCTGTGCGGCGTGC
TGATCGGCACCGCCATGGCCGGCATGACCTCCTTCGCCGCGGCGTGGAGGCCCTGACCCGCGGCGGCGTGC
65 GCAAGATGAACCCCTTCTGCATCCCCCTTCTCCATCTCCAACATGGGCGGCGCCATGCTGGCCATGGACATCG
GCTTCATGGGCCCCAACTACTCCATCTCCACGCTGCGCCACCGGCAACTACTGCATCTTGGGCGCGCGG
ACCCATCCGCGCGGCGGACGCGATGATGCTGGCGGCGGCGGCGGACGCGCCATCTCCCTCCGCGCA
TCGGCGGCTTTCATCGCTGCAAGGCCCTGTCCAAGCGCAACGACGAGCCCGAGCGCGCTCCCGCCCCCTGGG
ACGCGGACCGCGACGGCTTCTGTGATGGGCGAGGGCGCGGCGTGTGGTGTGAGGAGCTGGAGCACGCCA
70 AGCGCGCGGCGGCCACCATCTGGCCGAGCTGGTGGGCGGCGCGGCCACCTCCGACGCGCCACCATGACCG
AGCCCGACCCCGAGGGCGCGGCGTGCCTGTGCTGGAGCGCGCCCTGGAGCGCGCGCCCTGGGCCCGCG
AGCGGTGGGCTACGTGAACGCCACCGCACCTCCACCCCGCGGCGACGTGGCCGAGTACCGCGCCATCC

GCGCCGTGATCCCCAGGACTCCCTGCGCATCAACTCCACCAAGTCCATGATCGGCCACCTGCTGGGCGGCG
CCGCGCCGTGGAGGCCGTGGCCGCCATCCAGGCCCTGCGCACCGGCTGGCTGCACCCCAACCTGAACCTGG
AGAACCCCGCCCCGGCGTGGACCCCGTGGTGTGGTGGGCCCCCGCAAGGAGCGCGCCGAGGACCTGGACG
TGGTGTCTTCCAACCTCCTTCGGCTTCGGCGGCCACAACCTCCTGCGTGATCTTCCGCAAGTACGACGAGATGG
5 ACTACAAGGACCACGACGGCGACTACAAGGACCACGACATCGACTACAAGGACGACGACGACAAGTGAATCG
ATAGATCTCTTAAGGCAGCAGCAGCTCGGATAGTATCGACACACTCTGGACGCTGGTGTGTGATGGACTGT
TGCCGCCACACTTGTGCTGCCCTTGACCTGTGAATATCCCTGCCGCTTTTATCAAACAGCCTCAGTGTGTTTGAT
CTTGTGTGTACGCGCTTTTGGGAGTTGCTAGCTGCTTGTGCTATTTGCGAATACCACCCCGAGCATCCCCTT
10 CCCTCGTTTTCATATCGCTTGCATCCCAACCGCAACTTATCTACGCTGTCCTGCTATCCCTCAGCGCTGCTCC
TGCTCCTGCTCACTGCCCCCTCGCACAGCCTTGGTTTTGGGCTCCGCCTGTATTCTCCTGGTACTGCAACCTGT
AAACCAGCACTGCAATGCTGATGCACGGGAAGTAGTGGGATGGGAACACAAATGGAAGCTTAATTAAGAGC
TCTTGTTCCTCAGAAGGAGTTGCTCCTTGAGCCTTTTATTCTCAGCCTCGATAACCTCCAAAGCCGCTCTAA
TTGTGGAGGGGGTTTGAATTTAAAGCTTGAATGTTGGTTCTGTCGTCTGGAACAAGCCCAGACTTGTTCG
15 TCACTGGGAAAAGGACCATCAGCTCCAAAAAAGCTTGGCGCTCAAACCGCGTACCTCTGCTTTTCGCGCAATCT
GCCCTGTTGAAATCGCCACCACATTCATATTGTGACGCTTGAGCAGTCTGTAATTGCCTCAGAATGTGGAAT
CATCTGCCCCCTGTGCGAGCCCATGCCAGGCATGTGCGGGGCGAGGACACCCGCCACTCGTACAGCAGACCA
TTATGCTACCTCACAATAGTTTCATAACAGTGACCATATTTCTCGAAGCTCCCCAACGAGCACCTCCATGCTC
TGAGTGGCCACCCCGGCCCTGGTGTCTTGGGAGGGCAGGTCAACCGGCATGGGGCTACCGAAATCCCCGA
20 CCGGATCCCACCACCCCGCGATGGGAAGAATCTCTCCCGGGATGTGGGCCACACCAGCACAACCTGCT
GGCCCAGGCGAGCGTCAAACCATAACACACAAATATCCTTGGCATCGGCCCTGAATTCCTTCTGCGGCTCTG
CTACCCGGTGTCTTCTGTCCGAAGCAGGGGTGCTAGGGATCGCTCCGAGTCCGCAAACCCCTTGTGCGGTGGC
GGGGCTTGTTCGAGCTTGAAGAGC

SEQ ID NO: 59

Prototheca moriformis (UTEX 1435) SAD2 allele 2

ATGGCGTCTGCCGTACCTTTGCGTGCGCCCCCTCCCCGCGGCGCGGTGCGCGCGCCGGGTGCGCGCGCTGCC
TCGCGTCCCCTGGTGGTGCACGCCGTGCCAGCGAGGCCCGCTGGGCGTGCCGCCCTCGGTGCAGCGCCCC
TCCCCCGTGGTCTACTCCAAGCTGGACAAGCAACACCGCCTGACGCCCCGAGCGCCTGGAGCTGGTGCAGAGC
30 ATGGGTCACTTGCAGGAGAGAGGGTGTCCCCGTGCTGCACCCCGTGGACAAGCTGTGGCAGCCGAGGACT
TCCTGGCCCCGACCCCGAGTCGCCCCGACTTCGAGGACCAGGTGGCGGAGCTGCGCGCGCGCGCCAAGGACCTG
CCCGACGAGTACTTTGTGGTGTGGTGGGCGACATGATCACGGAGGAGGCGCTGCCGACCTACATGGCCATG
CTCAACACCTTGGACGGTGTGCGCGACGACACGGGCGCGGTGACACCCGTGGGCGCGCTGGACGCGGCAG
35 TGGGTGGCCGAGGAGAACCAGGACGGCGACCTGCTGAACAAGTACTGTTGGCTGACGGGGCGCGTCAACATG
CGGGCCGTGGAGGTGACCATCAACAACCTGATCAAGAGCGGCATGAACCCGAGACGGACAACAACCCCTTAC
TTGGGCTTCTGCTACACCTCCTTCCAGGAGCGCGCGACCAAGTACAGCCACGGCAACACCGCGCGCCTGGCG
GCCGAGCACGGCGACAAGGGCCTGAGCAAGATCTGCGGGCTGATCGCCAGCGACGAGGGCCGGCACGAGATC
GCCTACACGCGCATCGTGGACGAGTTCTTCGCTCGACCCCGAGGGCGCCGTGCGCGCCTACGCCAACATG
40 ATGCGCAAGCAGATCACCATGCCCGCGCACCTCATGGACGACATGGGCCACGGCGAGGCCAACCCGGGCCGC
AACCTCTTCGCCGACTTCTCCGCCGTGCGCGAGAAGATCGACGTCTACGACGCGGAGGACTACTGCCGCATC
CTGGAGCACCTCAACGCGCGCTGGAAGGTGGACGAGCGCCAGGTGAGCGGCCAGGCCGCGCGGACCAGGAG
TACGTTCTGGGCTGCCCGAGCGCTCCGGAAGTTCGCCGAGAAGACCGCCGCAAGCGCAAGCGCGTTCGCG
CGCAGGCCGCTGCGCTTCTCCTGGATCTCCGGACGCGAGATTATGGTCTAG

SEQ ID NO: 60

Prototheca moriformis (UTEX 1435) SAD2 allele 2

MASAVTFACAPPRGAVAAPGRRAASRPLVVHAVASEAPLGVPPSVQRPSPVVS KLDKQHRLTPERLELVQS
MGQLRRRGCSPPCTPWTSCGSRRTSWDPESPDEFDQVAELRARAKDLPDEYFVVLVGDIMITEALPTYMAM
50 LNTLDGVRDDTGAADHPWARWTRQWVAEENRHGDLNLCYIASDEGRHEIAYTRIVDEFRLDPEGAVAAANM
LGFVYTSFQERATKYSHGNTARLAAEHGDKLSKICGLIASDEGRHEIAYTRIVDEFRLDPEGAVAAANM
MRKQITMPAHLMDDMGHGEANPGRNLFADFSAVA EKIDVYDAEDYCRILEHLNARWKVDERQVSGQAADQE
YVLGLPQRFRLAEKTA AKRKRVARRPVAFSWISGREIMV

SEQ ID NO: 61

Prototheca moriformis (UTEX 1435) SAD2 allele 1

ATGGCGTCTGCCGTACCTTTGCGTGCGCCCCCTCCCCGCGGCGCGGTGCGCGCGCCGGGTGCGCGCGCTGCC
TCGCGTCCCCTGGTGGTGCAGCGCGGTGCCAGCGAGGCCCGCTGGGCGTTCCGCCCTCGGTGCAGCGCCCC
60 TCCCCCGTGGTCTACTCCAAGCTGGACAAGCAGCACCGCCTGACGCCCCGAGCGCCTGGAGCTGGTGCAGAGC
ATGGGGCAGTTTGGCGAGGAGAGGGTGTGCCCGTGTGACCCCGTGGACAAGCTGTGGCAGCCGCGAGGAC
TTTTTGCCCGACCCGAGTCGCCCCGACTTCGAGGATCAGGTGGCGGAGCTGCGCGCGCGCCCAAGGACCTG
CCCGACGAGTACTTTGTGGTGTGGTGGGGAATGATCACGGAGGAGGCGCTGCCGACCTACATGGCCATG
65 CTCAACACGCTGGACGGCGTGCAGCGACGACACGGGCGCGGCCGACCACCCGTGGGCGCGCTGGACGCGGCAG
TGGGTGGCCGAGGAGAACCAGGACGGCGACCTGCTGAACAAGTACTGCTGGCTGACGGGGCGCGTCAACATG
CGGGCCGTGGAGGTGACCATCAACAACCTGATCAAGAGCGGCATGAACCCGAGACGGACAACAACCCCTTAT
TTGGGGTTCTGCTACACCTCCTTCCAGGAGCGGCCCAAGTACAGCCACGGCAACACCGCGCGCTTGGC
GCCGACACGGCGACAAGAACCCTGAGCAAGATCTGCGGCTGATCGCCAGCGACGAGGGCCGCGCAGATC
70 GCCTAGCACGCGCATCGTGGACGAGTTCTTCGCTCGACCCCGAGGGCGCGCTGCGCGCCTACGCCAACATG
ATGCGCAAGCAGATCACCATGCCCGCGCACCTCATGGACGACATGGGCCACGGCGAGGCCAACCCGGGCCGCG

AACCTCTTCGCCGACTTCTCCGCGGTGCGCGAGAAGATCGACGTCTACGACGCCGAGGACTACTGCCGCATC
CTGGAGCACCTCAACGCGCGCTGGAAGGTGGACGAGCGCCAGGTGACGCGCCAGGCCGCCGCGGACCAGGAG
TACGTCTTGGGCTGCCCCAGCGCTTCCGGAACTCGCCGAGAAGACCGCCGCCAAGCGCAAGCGGTGCGG
CGCAGGCCCGTGCCTTCTCCTGGATCTCCGGGCGCGAGATCATGGTCTAG

5

SEQ ID NO: 62

Prototheca moriformis (UTEX 1435) SAD2 allele 1

MASAVTFACAPPRGAVAAPGRRAASRPLVVRAVASEAPLGVPPSVQRPSPVVYSKLDKQHRLTPERLELVQS
MGQFAEERVLPLVHPVDKLWQPQDFLPDPESPDFEDQVAELRARAKDLPDEYFVVLVGDMITTEALPTYMAM
LNTLDGVRDDTGAADHPWARWTRQWVAEENRHGDLNLYCWLTGRVNMRAVEVTINNLIKSGMNPQTDNNPY
LGFVYTSFQERATKYSHGNTARLAAEHGDKNLSKICGLIASDEGRHEIAYTRIVDEFFRLDPEGAVAAYANM
MRKQITMPAHLMDDMGHGEANPGRNLFADFSVAEKIDVYDAEDYCRILEHLNARWKVDERQVSGQAAADQE
YVLGLPQRFRLAEKTAARKRVARRPVAFSWISGREIMV

10

SEQ ID NO: 63

Prototheca moriformis (UTEX 1435) mitochondrial ACP allele 1

ATGGCGTTTCTTGCAGCGGACGAGCGCGCTGGTGCGCCAGGGTGTCTGTCTCGCCTCGCCGTGCAGGCCAGC
CCCCCCCTGAACAGCGTCCGGGCCTTCGCCTCGGCGTCTTACCTGGACAAGAATGAGGTGACGCACCGCGTG
CTGTGCGATTGTCAAGAACTTTGACCGCGTCGACGCCGGAAGGTACGCGACTCTGCCAACTTCCAGTCCGAC
CTGGGTCTGGACTCACTAGACACCGTGGAGCTGGTTCATGGCCCTGGAGGAGGAGTTTTCGCGATCGAGATCCCG
GATGCGGAGGCNGACAAGATCCTCTCCGTCCCGGAGGCGATTTCCTACATTGCCGCGAACCCCATGGCCAAG
TAG

20

SEQ ID NO: 64

Prototheca moriformis (UTEX 1435) Mitochondrial ACP allele 1

ATGGCGTTTCTTGCAGCGGACGAGCGCGCTGGTGCGCCAGGGTGTCTGTCTCGCCTCGCCGTGCAGGCCAGC
CCCCCCCTGAACAGCGTCCGGGCCTTCGCCTCGGCGTCTTACCTGGACAAGAATGAGGTGACGCACCGCGTG
CTGTGCGATTGTCAAGAACTTTGACCGCGTCGACGCCGGAAGGTACGCGACTCTGCCAACTTCCAGTCCGAC
CTGGGTCTGGACTCACTAGACACCGTGGAGCTGGTTCATGGCCCTGGAGGAGGAGTTTTCGCGATCGAGATCCCG
GATGCGGAGGCNGACAAGATCCTCTCCGTCCCGGAGGCGATTTCCTACATTGCCGCGAACCCCATGGCCAAG
TAG

25

SEQ ID NO: 65

Prototheca moriformis (UTEX 1435) Mitochondrial ACP allele 1

MAFLQRTSALVRQGVLSRLAVQASPALNSVRAFASASYLDKKEVTDRVLSIVKNFDRVDAGKVTDSANFQSD
LGLDSLDTVELVMALEEEFAIEIPDAEADKILSVPEAISYIAANPMAK

40

SEQ ID NO: 66

Prototheca moriformis (UTEX 1435) LPAAT-E

ATGGTTGCGGCTCAGGAAGACGGCTCCATGGAGAGGAGGACCTCGGCCAACCCCTCCGTGGGTCTTAGCAGC
TTAAAGCACGTTCCGCTGAGTGCGGTGGACCTCCCGAATTTGAGTCCGCTCCCAAGGGTGCAGGCCAGCCCC
GATGTTCCATCGTCAGATAGAGGAACCTGGTGGATAAGGCACGCGCAGATCGGTACCCGGCCTCGTTGTGCTC
CTAGCCGATGTGTGGACATCAGCGGGCCGCTGCAGGACGCGGCCTCGGCGATGGTGGACGACTCCTTCCTG
CGCTGCTTACCTCCACCATGGACGAGCCCTGGAACCTGGAACCTTTTACCTGTTCCCGCTCTGGGCGCTCGGC
GTCGTGCTCCGCAACCTGATCCTGTTCCCCCTGCGCCTCCTCACCATCGTGCTGGGGACGCTGCTCCTCGTG
CTGGCCTTTGCGGTGACGGGCTTTCTCCCCAAGGAGAAGCGGCTCGCGGCCGAGCAGCGGTGCGTCCAGTTC
ATGGCGCAGGCGTTCTGTTGGCTCCTGGACAGGCGTGATCCGGTACCACGGCCCCCGCCCCGTGGCGGCGCCC
AACC CGTGTGGGTGCGCAACACACGCTCCATGATCGACTACGCGGTGCTGTGCGCCTACTGCCCGTTTGCG
GCCATCATGCAGTGCACCCCGGCTGGGTGGCGCTCTTCCAGACGCGCTACCTGGCCTCGTGGGCTGCCTC
TGTTTCAACCGCACGCAGGCCAAGGACCGCACGCTCGTGGCCAAGCGCATGCGCGAGCACGTGGCCAGCGCG
ACCAGCACGCCGCTGCTCATCTTCCCCGAGGGCACCTGCGTCAACAACGAGTACTGCGTCATGTTCCGCAAG
GGCGCCTTTGACCTGGGCGCCACCGTCTGCCCCGTGGCCATCAAGTACAACAAGATCTTCGTGACGCGCTTT
TGGAACAGCAAGCGCCAGTCTTCTCCGCGCACCTCATGAAGATCATGCGCTCGTGGGCCCTCGTCTGCGAC
GTCTACTTCTGGAGCCCCAGACGCGGAGGGAGGGCGAGTGGTTCGAGGCGTTTGCGAACCGC

45

50

55

60

SEQ ID NO: 67

Prototheca moriformis KASI allele 2, nucleotide sequence

ATGGTCGCGACCCCTGTCCCTCGCCGGCCCCCGCCTGCAACACGCAGTGCGTATCCAGCAAGCGGGTTGTGCGC
TTCAACCGCCCCCATGTTGGCGTCCGGGCTCGATCAGCCGTGGTTCGCCCCGAGCTGCGAAGCGGGACCCCCACC
CAGCGCATTTGTATCACCAGGAATGGGCGTGGCCTCCGTGTTTGGCAACGATGTCGAGACCTTTTACGACAAG
CTTCTGGAAGGAACGAGCGGCGTGGACCTGATTTCCAGGTTTGACATCTCCGAGTTCCCGACCAAGTTTTCG
GCGCAGATCACC GGCTTCTCCGTGGAGGACTGCGTGGACAAGAAGAACGCGCGGCGGTACGACGACGCGCTG
TCGTACGCGATGGTGGCCTCCAAGAAGGCCCTGCGCCAGGCAGGCCTGGAGAAGGACAAGTGCCCCGAGGGC
TACGGGGCGCTGGACAAGACGCGCACGGGCGTGGTTCGGCTCGGGCATGGGCGGGCTGACGGTCTTCCAG
GACGGCGTCAAGGCGTGGTGGAGAAGGGCTACAAGAAGATGAGCCCCCTTCTTCATCCCCCTACGCCATCACC

70

AACATGGGCTCCGCGCTGGTGGGCATCGACCAGGGCTTCATGGGCCCCAACTACTCCGTCTCCACAGCCTGC
GCGACGTCCAACCTACGCATTTGTGAACGCGGCCAACACATCCGCAAGGGCGACGCGGACGTTCATGGTCGTC
GGCGGCACCGAGGCCCTCCATCGTGCCCGTGGGCCTGGGCGGCTTTGTGGCCTGCCGCGCGCTGTCCACGCGC
AACGACGAGCCCAAGCGCGCGAGCCGGCCGTGGGACGAGGGCCGCGACGGCTTTGTGATGGGCGAGGGCGCG
5 GCGGTGCTGGTCATGGAGTCGCTGGAGCACGCGCAGAAGCGTGGCGCGACCATCCTGGGCGAGTACCTGGGC
GGCGCCATGACCTGCGACGCGCACCATGACGCGACCCGACCCCGAGGGCCTGGGCGTGAGCACCTGCATC
CGCCTGGCGCTCGAGGACGCGCGCGTCTCGCCCGACGAGGTCAACTACGTCAACGCGCACGCCACCTCCACC
CTGGTGGGCGACAAGGCCGAGGTGCGCGCGGTCAAGTCGGTCTTTGGCGACATGAAGGGTATCAAGATGAAC
GCCACCAAGAGTATGATCGGGCACTGCCTGGGCGCCGCCGGCGGCATGGAGGCCGTGCGGACGCTCATGGCC
10 ATCCGCACCGGCTGGGTGCACCCACCATCAACCACGACAACCCCATCGCCGAGGTGATGGCCTGGACGTC
GTCGCCAACGCCAAGGCCAGCACGACATCAACGTGCGCATCTCCAACCTCCTTCGGCTTTGGCGGGCACAAAC
TCCGTGCTCGCCTTTGCGCCCTTCCGCGAGTAG

SEQ ID NO: 68

Prototheca moriformis KASI allele 2, protein sequence

MVATLSLAGPACNTQCVSSKRIVAFNRPHVGVRRASAVVARAAKRDPTQRIVITGMGVASVFGNDVETFYDK
LLEGTSGVDLISRFDISEFPTKFAAQITGFSVEDCVDKKNARRYDDALSYAMVASKALRQAGLEKDKCPEG
YGALDKTRTGVLVGSGMGLTVFQDGVKALVEKGYKKMSPFFIPYAITNMGSAVGVIDQGFMPNYSVSTAC
20 ATSNYAFVNAANHIRKGDADVMVVGTEASIVPVLGGFVACRALSTRNDEPKRASRPWDEGRDGFVMGEA
AVLVMESLEHAQKRGATILGEYLGGAMTCDAAHMTDPHPEGLGVSTCIRLALEDAGVSPDEVNYVNAHATST
LVGDKAEVRAVKSVMGDMKGIKMNATKSMIGHCLGAAGGMEAVATLMAIRTGWVHPTINHNDNPIAEVDGLDV
VANAKAQHDINVAISNSFGFGHNSVVAFAFPRE*

SEQ ID NO: 69

Prototheca moriformis (UTEX 1435) KASI allele 1, nucleotide sequence

ATGGTCGCGACCCTCTCCCTCGCCGGCCCCGCTGCAACACGCAGTGCGTATCCGGCAAGCGGGCTGTGCGC
TTCAACCGCCCCCATGTTGGCGTCCGGGCTCGATCAGCCGTGGTCGCCCGGGCTGCGAAGCGGGACCCACCC
30 CAGCGCATTTGTATCACCGGAATGGGCGTGGCCTCCGTGTTTGGCAACGATGTGAGACCTTTTACAACAAG
CTTCTGGAAGGAACGAGCGGCGTGGACCTGATTTCCAGGTTTGACATCTCCGAGTTCCCGACCAAGTTTTCG
GCGCAGATCACCGGCTTCTCCGTGGAGGACTGCGTGGACAAGAAGAAGCGCGCGGCGGTACGACGACGCGCTG
TCGTACGCGATGGTGGCCTCCAAGAAGGCCCTGCGCCAGGCGGGACTGGAGAAGGACAAGTGCCCCGAGGGC
TACGGAGCGCTGGATAAGACGCGCGCGGGCGTGTGGTGGGCTCGGGCATGGGCGGGCTGACGGTCTTCCAG
35 GACGGCGTCAAGCGCTGGTGGAGAAGGGCTACAAGAAGATGAGCCCTTCTTTCATCCCTACGCCATCACCC
AACATGGGCTCCGCGCTGGTGGGCATCGACAGGGCTTCATGGGGCCCACTACTCCGTTCACCGGCTGCG
GCGACCTCCAACCTACGCCTTTGTGAACGCGGCCAACACATCCGCAAGGGCGACGCGGACGTTCATGGTCGTG
GGCGGCACCGAGGCCCTCCATCGTGCCCGTGGGCCTGGGCGGCTTTGTGGCCTGCCGCGCGCTGTCCACGCGC
AACGACGAGCCCAAGCGCGCGAGCCGGCCGTGGGACGAGGGCCGCGACGGCTTCGTGATGGGCGAGGGCGCG
40 GCGGTGCTGGTCATGGAGTCGCTGGAGCACGCGCAGAAGCGCGGCGCGACCATCCTGGGCGAGTACCTGGGG
GGCGCCATGACCTGCGACGCGCACCATATGACGCGGACCCGACCCCGAGGGCCTGGGCGTGAGCACCTGCATC
CGCCTGGCGCTCGAGGACGCGCGCGTCTCGCCGACGAGGTCAACTACGTCAACGCGCACGCCACCTCCACC
CTGGTGGGCGACAAGGCCGAGGTGCGCGCGGTCAAGTCGGTCTTTGGCGACATGAAGGGCATCAAGATGAAC
GCCACCAAGTCCATGATCGGGCACTGCCTGGGCGCGCGCGGCGGCATGGAGGCCGTGCGCACGCTCATGGCC
45 ATCCGCACCGGCTGGGTGCACCCACCATCAACCACGACAACCCCATCGCCGAGGTGACGCGCCTGGACGTC
GTCGCCAACGCCAAGGCCAGCACAAAATCAACGTGCGCATCTCCAACCTCCTTCGGCTTCGGCGGGCACAAAC
TCCGTGCTCGCCTTTGCGCCCTTCCGCGAGTAG

SEQ ID NO: 70

Prototheca moriformis (UTEX 1435) KASI allele 1, protein sequence

MVATLSLAGPACNTQCVSGKRAVAFNRPHVGVRRASAVVARAAKRDPTQRIVITGMGVASVFGNDVETFYNK
LLEGTSGVDLISRFDISEFPTKFAAQITGFSVEDCVDKKNARRYDDALSYAMVASKALRQAGLEKDKCPEG
YGALDKTRAGVLVSGMGLTVFQDGVKALVEKGYKKMSPFFIPYAITNMGSAVGVIDQGFMPNYSVSTAC
55 ATSNYAFVNAANHIRKGDADVMVVGTEASIVPVLGGFVACRALSTRNDEPKRASRPWDEGRDGFVMGEA
AVLVMESLEHAQKRGATILGEYLGGAMTCDAAHMTDPHPEGLGVSTCIRLALEDAGVSPDEVNYVNAHATST
LVGDKAEVRAVKSVMGDMKGIKMNATKSMIGHCLGAAGGMEAVATLMAIRTGWVHPTINHNDNPIAEVDGLDV
VANAKAQHKINVAISNSFGFGHNSVVAFAFPRE

SEQ ID NO: 71

Prototheca moriformis (UTEX 1435) FATA allele 1

ATGGCACCGACCAGCCTGCTTGCCAGTACTGGCGTCTCTTCCGCTTCTCTGTGGTCTCTGCGCGCTCCAGC
GCGTGCGCTTTTCCGGTGGATCATGCGGTCCGTGGCGCACCGCAGCGGCCGTGCCCATGCAGCGCCGCTGC
65 TTCCGAACAGTGGCGGTGAGGGCCGACCCGCGGTAGCCGTCCGTCCGGAACCCGCCCAAGAGTTTGGGAG
CAGCTTGAGCCCTGCAAGATGGCGGAGGACAAAGCGCATCTTCTGGAGGAGCACCAGTTCGGGCAACGAG
GTGGGCCCCCTCGCAGCGGTGACGATCACGGCGGTGGCCAACATCCTGCAGGAGGCGGCGGGCAACACGCG
GTGGCCATGTGGGGCCGAGCTCGGAGGGTTTCGCGACGCGACCCGAGCTGCAGGAGGCGGGTCTCATCTTT
GTGATGACGCGCATGCAGATCCAGATGTACCGCTACCCGCGCTGGGCGACCTGATGCAGGTGGAGACCTGG
70 TTCCAGACGGCGGGCAAGCTGGGCGCGCAGCGCGAGTGGGTGCTGCGCGACAAGCTGACCGGCGAGGCGCTG
GGCGGGCCACCTCGAGCTGGGTGATGATCAACATCCGACGCGCGCGGCGGCTGCCGATGCCGGAGCTCGTC

CGCGTCAAGTCGGCCTTCTTCGCGCGCGAGCCGCCGCGCCTGGCGCTGCCGCCCCGCGGTACGCGTGCCAAG
 CTGCCCCAACATCGCGACGCCGCGCCGCTGCGCGGGCACCAGGTCGCGCGCCGACCGACATGGACATG
 AACGGGCACGTGAACAACGTGGCCTACCTGGCCTGGTGCCTGGAGGCCGTGCCCGAGCACGTCTTCAGCGAC
 TACCACCTCTACCAGATGGAGATCGACTTCAAGGCCGAGTGCCACGCGGGGCGACGTATCTCTCTCCAGGCC
 5 GAGCAGATCCCCCCCCAGGAGGCGCTCACGCACAACGGCGCCGCGCAACCCCTCTGCTTCGTCCATAGC
 ATTCTGCGCGCCGAGACCGAGCTCGTCCGCGCGCAACCACATGGTCGGCCCCCATCGACGCGCCCGCCGCC
 AAGCCGCCCAAGGCGAGCCACTGA

SEQ ID NO: 72

Prototheca moriformis (UTEX 1435) FATA allele 1

MAPTSLLASTGVSSASLWSSARSSACAFVDHAVRGAPQRPLPMQRRCFRTVAVRAAPAVAVRPEPAQEFWE
 QLEPCKMAEDKRI FLEEHRIRGNEVGPSQRLTITAVANILQEAAGNHAVAMWGRSSEGFATDPQLQEAGLIF
 VMTRMQIQMYRPRWGLMQVETWFQTAGKLGAQREWVLRDKLTGEALGAATSSWVMINIRTRPCRMPELV
 15 RVKSAFFAREPPRLALPPAVTRAKLPNIATPAPLRGHRQVARRTDMDMNGHVNNVAYLAWCLEAVPEHVFS
 YHLYQMEIDFKAECHAGDVISSQAEQIPQEAALTHNGAGRNPSCFVHSILRAETELVRARTTWSAPIDAPAA
 KPPKASH

SEQ ID NO: 73

Prototheca moriformis (UTEX 1435) FAD linoleate Allele 2

ATGGTCGCTGCGGGCGACGTGCTGGAGCCACCGCCTCTGAGGGAGATGGCCGGGGCCCTGGCCAACCCGCCG
 CCCTTCACGCTGGCCGACATCCGCAACGCCATCCCCAAGGACTGCTTCCGCAAGAGCGCGCCCAAGTCGTTG
 GCGTACCTGGGTCTGGACCTTTCATCGTCACGGGCATGGCTGTGCTGGCCTACAAAATCAACTCGCCGTGG
 25 CTCTGGCCCCCTGTACTGGTTTGCCAGGGAACGATGTTTTGGGCCCTGTTTGTGCTCGGCCACGACTGTGGC
 CACCAGAGCTTTAGCACGAGCAAGCGGCTGAACGACGCGGTGCGTCTTTTGTGTAACCTCCATCATCGGCGTG
 CCGTACCACGGCTGGCGCATCTCCACCGCACCCACCACAACAACACGCGCCACGTGGAGAACGACGAGTCC
 TGGTACCCGCCCCACCGAGAGCGGGCTCAAGGCGATGACCGACATGGGCCCGGCAGGGGCGCTTCCACTTTCCC
 ACCATGCTCTTTGTCTACCCCTTTTACCTCTTCTGGCGGTCCCCGGGCAAGACCGGGTCCCCTTTAGCCCC
 30 TCCACCGACCTCTTCGCCTCGTGGGAGGCCCCCTGATCCGGACGTCAAACGCGTGCCAGCTGGCGTGGCTG
 GCGCGCTGGCGCGGGGACCTGGGCGCTGGGCGTGCTGCCCATGCTCAACCTGTACCTCGCGCCCTACGTG
 ATATCGGTGGCGTGGCTGGATCTGGTGACGTACCTGCACCACCACGGCCCCCTCGGACCCGCGCGAGGAGATG
 CCCTGGTACCAGCGCCGCGAGTGGAGCTACATGCGCGGCGGGCTGACCACCATCGACCGCGACTACGGCCTC
 TTCAACAAGGTCCACCACGACATTTGGGACGCACGTGCTGCACCACCTGTTCCCGCAGATCCCGCACTACAAC
 35 CTCTGCCGCGCCACGGAGGCCGCCAAGAAGGTGCTGGGGCCCTACTACCGCGAGCCCGAGCGCTGCCCGCTG
 GGCTGCTGCCGTCACCTGCTCGCGCCCCCTGCTCCGCTCCCTCGGCCAGGACCACTTTGTGACGACGCG
 GGCTCCGTCTCTTCTACCGCCGCGCCGAGGGCATCAACCCCTGGATCCAGAAGCTCCTGCCCT

SEQ ID NO: 74

Prototheca moriformis (UTEX 1435) FAD linoleate Allele 2

MVAAGDVAGATASEGDGRALANPPPFITLADIRNAIPKDCFRKSAKSLAYLGLDLSIVTGMVAVLAYKINSPW
 LWPLYWFAQGTMFALFVVGHDGCHQSFSSTSKRLNDAVGLFVHSIIGVPYHGWRIHRTHNNHGHVENDS
 WYPPTESGLKAMTDMGRQGRFHFPTMLFVYPFLFWRSPGKTGSHFSPSTDLFASWEAPLIRTSNACQLAWL
 45 GALAAGTWALGVLPMNLVLYAPYVISVAWLDLVTYLHHHGPSDPREEMPWYRGREWSYMRGGLTTIDRDYGL
 FNKVHHDIGTHVHHLFPQIPHYNLCRATEAAKKVLGPYYREPERCPLGLLPVHLLAPLLRLSLGQDHFVDDA
 GSVLFYRRAEINPWIQKLLP

SEQ ID NO: 75

Prototheca moriformis (UTEX 1435) FAD linoleate Allele 1

ATGGTCGCTGCGGGCGACGTGCTAGAGCCACCGCCTCTGAGGGAGATGGCCGGGGCCCTGGCCAACCCGCCG
 CCCTTCACGCTGGCCGACATCCGCAACGCCATCCCCAAGGACTGCTTCCGCAAGAGCGCGCCCAAGTCGTTG
 GCGTACCTGGGTCTGGACCTTTCATCATCACGGGCATGGCTGTGCTGGCCTACAAAATCAACTCGCCGTGG
 55 CTCTGGCCCCCTGTACTGGTTTGCCAGGGAACGATGTTTTGGGCCCTGTTTGTGCTCGGACACGACTGTGGC
 CACCAGAGCTTTAGCACGAGCAAGCGGCTGAACGACGCGGTGCGTCTTTTGTGTAACCTCCATCATCGGCGTG
 CCGTACCACGGCTGGCGCATCTCCACCGCACCCACCACAACAACACGCGCCACGTGGAGAACGACGAGTCC
 TGGCAAAGCGGGCTCAAGGCGATGACCGATATGGGCCCGGCAGGGCCGATTCCACTTTCCCTCCATGCTCTTT
 GTCTACCCCTTTTACCTCTTCTGGCGGTCCCCGGGCAAGACCGGGTCCCCTTTAGCCCCGCCACAGACCTC
 60 TTGCCTTGTGGGAGGCCCCCTGATCCGGACGTCAAACGCGTGCCAGCTGGCGTGGCTGGGCGCGCTGGCG
 GCGGGGACCTGGGCGCTGGGCGTGCTGCCCATGCTGAACCTGTACCTCGCGCCCTACGTGATCTCGGTGGCC
 TGGCTGGACCTGGTGACGTACCTGCACCACCACGGCCCCCTCGGACCCGCGCGAGGAGATGCCCTGGTACCGC
 GGCCGGGAGTGGAGCTACATGCGCGGCGGGCTGACCACAATCGACCGCGACTACGGTCTTTTCAACAAGGTC
 CACCACGACATCGGGACGCACGTGCTGCACCACCTGTTCCCGCAGATCCCGCACTACAACCTCTGCCGCGCC
 65 ACGAAGGCCGCCAAGAAGGTGCTGGGGCCCTACTACCGCGAGCCCGAGCGCTGCCCGCTGGGCCTGCTGCCC
 GTGCACCTGCTCGCGCCCTGCTCCCTCGGCCAGGACCACTTTGTGACGACGCGGGCTCCGTCTCTC
 TTCTACCGGCGCGCCGAGGGCATCAACCCCTGGATCCAGAAGCTCCTGCCCT

SEQ ID NO: 76

Prototheca moriformis (UTEX 1435) FAD linoleate Allele 1

MVAAGDVARATASEGDRALANPPFFTADIRNAIPKDCFRKSAKSLAYLGLDLSIITGMAVLAYKINSPW
 LWPLYWFAQGTMTFWALFVVGHDCGHQSFSTSKRLNDAVGLFVHSIIGVPYHGWRIHRTHNNHGHVENDES
 CESGLKAMTDMGRQGRFHFPSMLFVYPFYLFWRSPGKTGSHFSPATDLFALWEAPLIRTSNACQLAWLGALA
 AGTWALGVLPMNLNLYLAPYVISVAWLDLVLYLHHHGPSDPREEMPWYRGREWSYMRGGLTTIDRDYGLFNKV
 5 HHDIGTHVHHLPQIPHYNLCRATKAACKVLGPYYREPERCPLGLLPVHLLAPLLRSLGQDHFVDDAGSVL
 FYRRAEGINPWIKLLP

SEQ ID NO: 77

Prototheca moriformis (UTEX 1435) FAD-C delta-12 desaturase (FADc)
 Allele 2

MAIKTNRPVEKPPFTIGTLRKAIPAHCFERSALRSSMYLAFDIAVMSLLYVASTYIDPAPVPTWVKYIGIMW
 PLYWFFQAGFTGVWVCAHECGHQAFSSSQAINDGVLVFSLLLVPYYSWKHSHRRHHSNTGCLDKDEVFV
 PPHRAVAHEGLEWEELPIRMGKVLVLTLLGWPLYLMFNVASRPPYPRFANHFDWPSPIFSKRERIEVVISDL
 15 ALVAVLSGLSVLGRMTGWAWLVKTYVVPYMIIVNMWLVLTLLQHTHPALPHYFEKDWDWLRGAMATVDRSMG
 PPFMDNILHHISDTHVLHHLFSTIPHYHAEESAAIRPILGKYYQSDSRVWGRALWEDWRDCRYVVPDAPED
 DSALWFFHK

SEQ ID NO: 78

Prototheca moriformis (UTEX 1435) FADc allele 2

ATGGCTATCAAGACGAACAGGCAGCCTGTGGAGAAGCCTCCGTTACAGATCGGGACGCTGCGCAAGGCCATC
 CCCGCGCACTGTTTCGAGCGCTCGGCGCTTCGTAGCAGCATGTACCTGGCCTTTGACATCGCGGTATGTCC
 CTGCTCTACGTGCGCTCGACGTACATCGACCCCTGCACCGGTGCCTACGTGGGTCAAGTACGGCATCATGTGG
 25 CCGCTCTACTGTTCTTCCAGGGCGCCTTCGGCACGGGTGTCTGGGTGTGCGCGCACGAGTGCGGTACCCAG
 GCCTTTTCTCCAGCCAGGCCATCAACGACGGCGTGGGCCTGGTGTTCACAGCCTGCTGCTGGTGCCCTAC
 TACTCCTGGAAGCACTCGCACCGCGCCCACTCCAACACGGGGTGCCTGGACAAGGACGAGGTGTTTGTG
 CCGCCGCAACCGTGCGGTGGCGCACGAGGGCCTGGAGTGGGAGGAGTGGCTGCCCATCCGCATGGGCAAGGTG
 CTGGTCACCTTGACCTGGGCTGGCCGCTGTACCTCATGTTCAACGTGCGCTCCCGCCCTTACCCGCGCTTC
 30 GCGAACCCTTTGACCCGTGGTCGCCCATCTTCAGCAAGCGCGAGCGCATCGAGGTGGTCATCTCCGACCTC
 GCGTTGGTGGCGGTGCTCAGCGGGCTCAGCGTGTGGGCCGACCATGGGCTGGGCTGGTCAAGACC
 TACGTGGTGCCCTACATGATCGTGAACATGTGGCTGGTGTCTCATCAGCTGCTCCAGCACACGACCCGGCC
 CTGCCGCACTACTTCGAGAAGGACTGGGACTGGCTACGCGGCGCCATGGCCACCGTCGACCGCTCCATGGGC
 CCGCCCTTCATGGACAGCATCCTGCACCACATCTCCGACACCCACGTGCTGCACCACCTCTTCAGCACCATC
 35 CCGCACTACCACGCGGAGGAGGCTCCGCCGCGCATCCGCCCATCTGGGCAAGTACTACCAATCCGACAGC
 CGTGGGTGCGCCGCGCCCTGTGGGAGGACTGGCGCGACTGCCGCTACGTGTCCTCCCGACGCGCCCGAGGAC
 GACTCCGCGCTCTGGTTCACAAAGTGA

SEQ ID NO: 79

Prototheca moriformis (UTEX 1435) Oleate Desturase

ATGGCTATCAAGACGAACAGGCAGCCTGTGGAGAAGCCTCCGTTACAGATCGGGACGCTGCGCAAGGCCATC
 CCCGCGCACTGTTTCGAGCGCTCGGCGCTTCGTAGCAGCATGTACCTGGCCTTTGACATCGCGGTATGTCC
 CTGCTCTACGTGCGCTCGACGTACATCGACCCCTGCGCCGGTGCCTACGTGGGTCAAGTATGGCGTCATGTGG
 45 CCGCTCTACTGTTCTTCCAGGTGTGTGTGAGGGTGTGGTGGCCGTCATCGAGGTCTGGTGGCGCGCATG
 GGGGAGAAGGCGCTGTCCCGCTGACCCCCCGGCTACCTCCCGGCACCTTCAGGGCGCCTTCGGCACGG
 GTGTCTGGGTGTGCGCGCACGAGTGCGGCCACAGGCCTTTTCTCCAGCCAGGCCATCAACGACGGCGTGG
 GCCTGGTGTTCACAGCCTGCTGCTGGTGGCCCTACTACTCCTGGAAGCACTCGCACCGCGCCCACTCCA
 ACACGGGGTGCTGGACAAGGACGAGGTGTTTGTGCCGCGCACCGCGCAGTGGCGCACGAGGGCCTGGAGT
 50 GGGAGGAGTGGTGGCATCGCCATCGGATGGGCAAGGTGCTGGTCAACCCTGACCCCTGGCCGCTGTACCTCA
 TGTTCAACGTGCGCTCGCGGCCGTACCCGCGCTTCGCCAACCCTTTGACCCGTGGTGGCCCATCTTCAGCA
 AGCGCGAGCGCATCGAGGTGGTCATCTCCGACCTGGCGCTGGTGGCGGTGCTCAGCGGGCTCAGCGTGTGG
 GCCGACCATGGGCTGGGCTGGCTGGTCAAGACCTACGTGGTGGCTACCTGATCGTGAACATGTGGCTCG
 TGCTCATCAGCTGCTCCAGCACACGACCCGCGCTGCCGCACTACTTCGAGAAGGACTGGGACTGGCTGC
 55 GCGGCGCCATGGCCACCGTGGACCGCTCCATGGGCCCGCCCTTCATGGACAACATCCTGCACCACATCTCCG
 ACACCCAGTGTGCACCACCTCTTCAGCACCATCCCGCACTACCACGCGGAGGAGGCTCCGCCGCCATCA
 GGCCCATCTGGGCAAGTACTACCAGTCCGACAGCCGCTGGGTGCGCCGCGCCCTGTGGGAGGACTGGCGCG
 ACTGCCGCTACGTGTCCTCCGACGCGCCCGAGGACGACTCCGCGCTCTGGTTCACAAAGTGA

SEQ ID. NO:80

Prototheca moriformis (UTEX 1435) FAD-B delta-12 desaturase (FADc Allele
 1), protein

MAIKTNRPVEKPPFTIGTLRKAIPAHCFERSALRSSMYLAFDIAVMSLLYVASTYIDPAPVPTWVKYIGVMW
 PLYWFFQAGFTGVWVCAHECGHQAFSSSQAINDGVLVFSLLLVPYYSWKHSHRRHHSNTGCLDKDEVFV
 PPHRAVAHEGLEWEELPIRMGKVLVLTLLGWPLYLMFNVASRPPYPRFANHFDWPSPIFSKRERIEVVISDL
 65 AXVAVLSGLSVLGRMTGWAWLVKTYVVPYLIIVNMWLVLTLLQHTHPALPHYFEKDWDWLRGAMATVDRSMG
 PPFMDNILHHISDTHVLHHLFSTIPHYHAEESAAIRPILGKYYQSDSRVWGRALWEDWRDCRYVVPDAPED
 DSALWFFHK

SEQ ID. NO: 81

Prototheca moriformis (UTEX 1435) LPAAT-E

5 ATGGTTGCGGCTCAGGAAGACGGCTCCATGGAGAGGAGGACCTCGGCCAACCCCTCCGTGGGTCTTAGCAGC
TTAAAGCACGTTCCGCTGAGTGCGGTGGACCTCCCGAATTTTCGAGTCCGCTCCCAAGGGTGCGGCCAGCCCC
GATGTCATCGTCAGATAGAGGAACCTGGTGGATAAGGCACGCGCAGATCGGTACCGGCCCTCGTTGTTCGCTC
CTAGCCGATGTGCTGGACATCAGCGGGCCGCTGCAGGACGCGGCCTCGGCGATGGTGGACGACTCCTTCCTG
CGCTGCTTCACCTCCACCATGGACGAGCCCTGGAACCTGGAACCTTTTACCTGTTCCCGCTCTGGGCGCTCGGC
10 GTCGTGTCGTCGCAACCTGATCCTGTTCCCCCTGCGCCTCCTCACCATCGTGCTGGGGACGCTGCTCTTCGTG
CTGGCCTTTGCCGTGACGGGCTTTCTCCCCAAGGAGAAGCGGCTCGCGGCCGAGCAGCGGTGCGTCCAGTTC
ATGGCGCAGGCGTTTCGTGGCCTCCTGGACAGGCGTGATCCGGTACCACGGCCCCCGCCCCGTGGCGGCGCCCC
AACCGCGTGTGGGTGCGCAACCACACGTCCATGATCGACTACGCGGTGCTGTGCGCCTACTGCCCGTTTGCG
GCCATCATGCAGCTGCACCCCGGCTGGGTGGGCGTCTTCCAGACGCGCTACCTGGCCTCGTGCGGCTGCCTC
15 TGGTTCAACCGCACGCAGGCCAAGGACCGCACGCTCGTGGCCAAGCGCATGCGCGAGCACGTGGCCAGCGCG
ACCAGCACGCCGCTGCTCATCTTCCCCGAGGGCACCTGCGTCAACAACGAGTACTGCGTCATGTTCCGCAAG
GGCGCCTTTGACCTGGGCGCCACCGTCTGCCCCGTGGCCATCAAGTACAACAAGATCTTCGTGACGCGCTTT
TGGAACAGCAAGCGCCAGTCTTCTCCGCGCACCTCATGAAGATCATGCGCTCGTGGGCCCTCGTCTGCGAC
GTCTACTTCTTGAGCCCCAGACGCGGAGGGAGGGCGAGTCGGTCGAGGCGTTTGCGAACCGC

SEQ ID. NO: 82

Prototheca moriformis (UTEX 1435) LPAAT-E, protein

25 MVAAQEDGSMERRTSANPSVGLSSLKHVPLSAVDLPNFESAPKGAASPDVHRQIEELVDKARADRSPASLSL
LADVLDISGPLQDAASAMVDDSLRCLFTSTMDEPNWNFYLFLPLWALGVVVRNLILFPLRLLTIVLGTLLFV
LAFAVTGFLPKERLAAEQRCVQFMAQAFVASWTGVIRYHGPRPVAAPNRVWVANHTSMIDYAVLCAYCPFA
AIMQLHPGWGVFQTRYLASLGCLWFNRTQAKDRTLAKRMREHVASATSTPLLIIFEGTCVNNEYCVMFRRK
GAFDLGATVCPVAIKYNKIFVDAFWNSKRQSFSAHLMKIMRSWALVCDVYFLEPQTRREGESVEAFANR

SEQ ID. NO:83

Prototheca moriformis (UTEX 1435) LPAAT-A, protein

30 MPPPEAPTNGDLAAPFVRKDRFGEYGMAPQPPMVKVVLALALVLLPLRAVSFVILVVLYWLICNASVALPP
KYCAAVTVTAGRLVCRWALFCFGFYIKWVNLAGAEEGPRPGGIVSNHCSYLDILLHMSDSFPAFVARQSTA
KLFIGIISQIMSCLYVNRDRSGPNHVGADLVKQRMQDEAEGKTPPEYRPLLLFPEGTTSNGDYLLPFTKG
AFLAGVPVQPVVLHYHRLPVYVPNEEEKADPKLYAQNVKAMMEVAGTKDTTAVFEDKMRYLNSLKRKYGKP
VPKKIE

SEQ ID. NO:84

Prototheca moriformis (UTEX 1435) LPAAT-A

40 AAATTATTGGCCTCATCCGCGAGGGTCGGAGCCTTTCAAACGTCACAAGCTGTATATTTGAGTCCCTCTCAA
TTCGTTTGAATGAAGGAGCAAGTGGTTAATGCCCCCCCCCTGAGGCCCCACGAATGGGGACCTTGCGGCCC
CCTTTGTGCGGAAGGACCGGTTGCGGGAGTATGGCATGGCCCCGAGCCCCGATGGTGAAGGTGGTGGTGG
CACTCGAGGCCTTGGTGTCTTGGCCGCTGCGCGCGGTGAGCGTGTTTCATCCTCGTCTGTCTATACTGGCTCA
TTTGCAATGCTTCGGTTCGACTCCCGCCCCAAGTACTGCGCGGCCGTACGGTCACCGCTGGGCGCTTGGTCT
GCCGTGGGCGCTCTTCTGCTTCGGATTCCACTACATCAAGTGGGTGAACCTGGCGGGCGCGAGGAGGGCC
CCCGCCCCGGGCGGCATTGTTAGCAACCACTGCAGCTACCTGGACATCCTGCTGCACATGTCCGATTCTTCC
CCGCTTTGTGGCGCGCCAGTCGACGGCCAAGCTGCCCTTTATCGGCATCATCAGCCAAATTATGAGCTGCC
TCTACGTGAACCGCGACCGCTCGGGGCCCAACCACGTGGGTGTGGCCGACCTGGTGAAGCAGCGCATGCAGG
45 ACGAGGCCGAGGGGAAGACCCCGCCGAGTACCGGCCGCTGCTCCTCTTCCCCGAGGGCACCACTCCAACG
GCGACTACCTGCTTCCCTTCAAGACCGCGCCTTCTTGGCCGGGTGCCCCGTCCAGCCCGCTCGTCCCTCCACT
ACCACAGGTTGCCCGTGTACGTCCCAATGAGGAGGAAAGGCCGACCCCAAGCTGTACGCCCAAAACGTCC
GCAAAGCCATGATGGAGGTCGCCGGGACCAAGGACACGACGCGCGGTGTTTGAGGACAAGATGCGCTACCTGA
50 ACTCCCTGAAGAGAAAGTACGGCAAGCCTGTGCCTAAGAAAATTGAGTGAACCCCCGTGCTCGACCAAAGAG

SEQ ID. NO:85

Prototheca moriformis (UTEX 1435) SAD1 allele 2, protein

55 MASAAFTMSACPAMTGRAPEARRRSGRPVATRLRAVAAPPRSQGLTPTIVRQDVLSATDLQIEVIKDLTPVF
ESKILPLLPGVDDLDWQPTDYLPA NSERFFDEIGELRERSASLSDELLVCLVGD MITEEALPTYMAMINTLD
GMRDETGRDAHPYARWTRSWVAEENRHGDLNLYLWLTGRVDM LAVERTIQRLISTGMDPGTENHPYHGFVF
TSFQERATKLSHGATARIAHAAGDEALAKICGTIARDES RHEMAYTRTMEAI FERDP SGAVVAFAHMMLRKI
TMPAHLMD DGRHERATGRSLFDDYAAVAERIGVYAKDYISILKHLIKRWKVEGLTGLTPPEARQAQDLLCSL
PTRFERLARHKS KKVTDTPPAQVQFSWVFQRPVAM

SEQ ID. NO:86

Prototheca moriformis (UTEX 1435) SAD1 allele 2

ATGGCTTCCGCGGCATTACCATGTGCGGCGTGCCCCGCGATGACTGGCAGGGCCCCCTGAGGCACGTGCGTCC
 GGACGGCCAGTCGCCACCCGCCCTGAGGGCTGTGGCGGCCCCGCCGCGTTCACAGGGCCTCACCCCCACCATC
 GTCCGCCAAGATGTGTGCTGCACTCTGCCACGGACCTGCAGATCGAGGTGATCAAGGACCTGACCCCCGTGT
 TTGAGTCCAAGATCCTGCCCTGCTGCCCGGCGTGGACGACCTGTGGCAGCCGACGGACTACCTTCCCGCCT
 5 CCAACTCGGAGCGTTTCTTCGACGAGATCGGGGAGCTGCGCGAGCGCTCGGCGTGTGCGACGAGCTTC
 TGGTGTGCTTGGTGGGCGACATGATCACGGAGGAGGCGCTGCCGACCTACATGGCCATGATCAACACGCTGG
 ACGGGATGCGGGACGAGACGGGGCGCGACGCGCACCCGTACGCGCGCTGGACGCGCAGCTGGGTGGCGGAGG
 AGAACCGCCACGGCGACCTGCTGAACAAGTACCTCTGGCTGACGGGGCGCGTGGACATGCTGGCCGTGGAGC
 GCACGATCCAGCGCCTCATCTCCACGGGCATGGACCCGGGCACGGAGAACCACCCCTACCACGGCTTCGTCT
 10 TCACCTCCTTCCAGGAGCGCGCGACCAAGCTGAGCCACGGCGCCACCGCGCGCATCGCGCACGCCGCCGGCG
 ACGAGGCGCTGGCCAAGATCTGCGGCACCATCGCGCGCGACGAGTTCGCGCCACGAGATGGCCTACACGCGCA
 CGATGGAGGCCATCTTCGAGCGCGACCCCTCGGGCGCGGTGCTCGCCTTTGCGCACATGATGCTGCGCAAGA
 TCACCATGCGCGCGCACCTCATGGACGACGGCGGCCACGAGCGCGCCACGGGACGCTCGCTCTTCGACGACT
 ACGCGGCCGTGCGCGAGCGCATCGGCGTCTACACCGCCAAGGACTACATCTCCATCCTCAAGCACCTCATCA
 15 AGCGCTGGAAGGTGGAGGGCCTGACCGGGCTGACCCCGGAGGCGCGCCAGGCGCAGGACCTGCTCTGCTCCC
 TCCCCACCCGCTTCGAGCGCCTGGCCAGGCACAAGTCCAAGAAGGTGACCGACACGCCCCCGGCCAGGTCC
 AGTTCTCCTGGGTCTTCCAGAGGCCCCGTGCGCATGTAAAGAGATTGGCGCCTGATTGCGTTTGGATCCGAGG
 ATTTCCAATCGGTGAGAGGACTGGGTGCCCAACTACCGACCTTGCCACCATCGCCCTTGCACTATTATTATCC
 CACTTTCTGCTGCGCCTGCGGGCGATTGCGGGCGTTTCTGCCCTTGACGTATTAATTTGCCCCCTTGCTGGC
 20 GCGAGGGTTCTTCAAGTTGATAAGAACGCACTCCCGCCAGCTCTGTACTTTTTCTGCGGGGCCCCCTGCATGG
 CTTGTTCCCTATGCTTGCTCGATCGACGGCGCCATTGCCACGGCGTTCGCCGATCCATGTGAAGAAACAC
 GG

SEQ ID. NO: 87

Prototheca moriformis (UTEX 1435) SAD1 allele 1

ATGGCTTCCGCGGCATTACCATGTGCGGCGTGCCCCGCGATGACTGGCAGGGCCCCCTGGGGCACGTGCGTCC
 GGACGGCCAGTCGCCACCCGCCCTGAGGGCTGTGGCGGCCCCGCCGCGTTCACAGGGCCTCACCCCCACCATC
 GTCCGCCAAGATGTGTGCTGCACTCTGCCACGGACCTGCAGATCGAGGTGATCAAGGACCTGACCCCCGTGT
 30 TTGAGTCCAAGATCCTGCCCTGCTGCCCGGCGTGGACGACCTGTGGCAGCCGACGGACTACCTGCCCGCCT
 CCAACTCGGAGCGCTTCTTCGACGAGATCGGGGAGCTGCGCGAGCGCTCGGCGGCGTGTGCGACGAGCTGC
 TGGTGTGCTTGGTGGGCGACATGATCACGGAGGAGGCGCTGCCGACCTACATGGCCATGATCAACACGCTGG
 ACGGGATGCGGGACGAGACGGGGCGCGACGCGCACCCGTACGCGCGCTGGACACGCAGCTGGGTGGCGGAGG
 AGAACCGCCACGGCGACCTGCTGAACAAGTACCTCTGGCTGACGGGGCGCGTGGACATGCTGGCCGTGGAGC
 35 GCACCATCCAGCGCCTCATCTCCACGGGCATGGACCCGGGCACGGAGAACCACCCCTACCACGGCTTCGTCT
 TCACCTCCTTCCAGGAGCGCGCGACCAAGCTGAGCCACGGCGCCACCGCGCGCATCGCGCACGCCCGCGCG
 ACGAGGCCCTGGCCAAGATCTGCGGCACCATCGCGCGCGACGAGTACGCCACGAGATGGCCTACACGCGCA
 CGATGGAGGCCATCTTCGAGCGCGACCCCTCGGGCGCGGTGCTCGCCTTTGCGCACATGATGCTGCGCAAGA
 TCACCATGCGCGCGCACCTCATGGACGACGGGCGGCACGAGCGCGCCACGGGGCGCTCGCTCTTCGACGACT
 40 ACGCGGCCGTGGCCGAGCGCATCGGCGTCTACACCGCCAAGGACTACATCTCCATCCTCAAGCACCTCATCA
 AGCGCTGGAAGGTGGAGGGCCTGACCGGGCTGACCCCGGAGGCGCGCCAGGCCCAGGACCTGCTCTGCTCCC
 TCCCCGCCGCTTTGAGCGCCTGGCCAAGCACAAAGTCCAAGAAGGTGACCGACACGCCCCCGGCCAGGTCC
 AGTTCTCCTGGGTCTTCCAGAGGCCCCGTGCGCATGTAAAGTGGCAGAGATTGGCGCCTGATTGCAATTTGGAT
 CCAAGGATCTCCAATCGGTGATGGGGACTGAGTGCCCAACTACCACCCTTGCACTATCGTCCTCGCACTATT
 45 TATTCCCACCTTCTGCTGCGCCTGCCGGGCGATTGCGGGCGTTTCTGCCCTTGACGTATCAATTTGCCCCCT
 GCTGGCGCGAGGATCTTTCATTTAATAAGAACTCACTCCCGCCAGCTCTGTACTTTTTCTGCGGGGCCCCCT
 GCATGGCTTGTTCCTCAATGCTTGCTCGATCGACGGCGCCATTGCCACGGCGTTCGCCGATCCATGTGAAG
 AAACACGG

SEQ ID. NO:88

Prototheca moriformis (UTEX 1435) SAD1 allele 1

MASAAFTMSACPAMTGRAPGARSGRPVATRLRAVAAPPRSQGLTPTIVRQDVLHSATDLQIEVIKDLTPVF
 ESKILPLLPGVDDLWQPTDYLPA NSERFFDEIGELRERSAALSDELLVCLVGD MITEEALPTYMAMINTLD
 55 GMRDETGRDAHPYARWTRSWVAEENRHGDLNLYLWLTGRVDM LAVERTIQRLISTGMDPGTENHPYHGFVF
 TSFQERATKLSHGATARIAHAAGDEALAKICGTIARDES RHEMAYTRTMEAI FERDP SGAVVAFAHMMLRKI
 TMPAHLMD DGRHERATGRSLFDDYAAVAERIGVYAKDYISILKHLIKRWKVEGLTGLTPEARQAQDLLCSL
 PARFERLAKHKS KKVTDTPPAQVQFSWVFQRPVAM

SEQ ID. NO:89

Prototheca moriformis (UTEX 1435) Plastidial ACP allele 1

MAMSMTSCRAVCAPRATLRVQAPRVAVRPFRAQRMICRAVDKASVLS DVRVIIAEQLGTDVEKVNADAKFAD
 LGADSLDTVEIMMALEEKFDLQLDEEGA EKITTVQEAA DLISAQIGA

SEQ ID. NO:90

Prototheca moriformis (UTEX 1435) Plastidial ACP allele 2

MAMSMTSCRAVCAPRATLRVQAPRVAVRPFRAQRMICRAVDKASVLS DVRVIIAEQLGTDVEKVNADAKFAD
 LGADSLDTVEXMMALEEKFDLQLDEEGA EKITTVQEAA DLISAQXGA

SEQ ID. NO:91

Prototheca moriformis (UTEX 1435) Acetyl-CoA carboxylase alpha-CT subunit allele 2

ATGCTGCTCTCGCGCTTCAACCCCATCAGCGAGAAGGCCACCAACGCCACGGTGCTCGACTTTGAGAAACCT
 TTGGTGGAACTTGATCATAAAATTCGCGAGGTCGCGAAAGTCGCGGAGGAGAACGGCGTGACGTCACGGAG
 CAGATCCGCGAGCTGGAGGACCGCGCGCAGCAGGTAGTCGAGTCTGTGTGGAGAGGGTTGCGAAAGGAGACG
 TACTCCAAGCTGACGCCGATCCAGCGGCTGCAGGTGGCGCGCCACCCCAACCGCCCCACTTTCTTGGACATT
 GCGCTCAACATCAGCGACAAGTTTGTGGAGCTGCACGGCGATCGCGCGGGGTACGACGACCCGGCGCTCGTC
 TGCGGCATCGCGTCGATCGACAACGTGAGCTTCATGTTTCATGGGCCACCAGAAGGGCCGCAACACCAAGGAG
 AACATCTACCGCAATTTTCGGCATGCCGCAGCCCAACGGCTACCGCAAGGCGCTGCGCTTCATGCGCCTGGCG
 GACAAAGTTTGGCTTCCCCATCGTCACGTTTGTGGACACGCCCCGGCGCGTACGCGGGGCTCAAGGCCGAGGAG
 CTGGGCCAGGGCGAGGCCATCGCGGTCAACCTGCGCGAGATGTTTGGCTTCCGCGTGCCCATCATCTCCATC
 GTCATCGGCGAGGGCGGCTCCGGCGGCGCGCTGGCCATCGGCTGCGCCAACCGCATGCTCATCATGGAGAAC
 GCGGTGTACTACGTGGCCTCTCCCGAGGCCTGCGCGGCCATCCTGTGGAAGAGCCGCGACAAGGCCGGCCTG
 GCCACCGAGGCGCTCAAGATCACGCCGCCGACCTGGTCAAGCTCAAGGTCATGGACGAGATCGTGCCGGAG
 CCGCTGGGCGCCGCGCACACCGACCCCATGGGCGCCTTCCCGGCCGTGCGCGAGGCCATCCTGCGCCAC

SEQ ID. NO: 92

Prototheca moriformis Acetyl-CoA carboxylase alpha-CT subunit allele 1

ATGTTCAATAGATACTATCTTGCAACAAGATTGCACTTGGGTTTGGTCGTCTGCTGTTGTGCACTACCATGT
 GACGTTGTGCTCGTCCATTCTGTTCCCGTCCAGGTCCGCAAAGTCGCGGAGGAGAACGGCGTGACGTCACG
 GAGCAGATCCGCGAGCTGGAGGACCGCGCGCAGCAGGTAGTCGAGTCTGTGTGGAGAGGGGTGGGATACTTT
 GGAAGAATGGGTGCTGTGGAGTCCAGTCATTTGAAGGAAAGGGTAGTCGACTTCCCTCTGTGGTCTGATCTA
 TCTTCTCCAAGCATATCTCACTTTGCCCTCTTCTCCGCCCCCTCCATTACCAACCCGTCCACAGTTGCGA
 AAGGAGACGTACTCCAAGCTGACGCCGATCCAGCGGCTGCAGGTGGCGCGCCACCCCAACCGCCCCACTTTC
 TTGGACATTGCGCTCAACATCACGGACAAGTTTGTGGAGCTGCACGGCGATCGCGCGGGGTACGACGACCCG
 NNNNNNNNAGAACGGCGTGGACGTCACGGAGCAGATCCGCGAGCTGGAGGACCGCGCGCAGCAGTGTAGT
 CAACTTTTCTGTGGTTCTATCCATCTTCCCCAAGTATATCTCACCCCTCTTCCCCCTCTCCCCCTTTAAC
 AACACCCCGTCCACAGTTGCGAAAGGAGACATACTCCAAGCTGACGCCGATCCAGCGGCTGCAGGTGGCGCG
 CCACCCCAACCGCCCTACTTTCTTGGACATTGCGCTCAACATCACGGACAAGTTTGTGGAGCTGCACGGCGA
 CCGCGCGGGGTACGACGACCCCGCGCTCGTCTGCGGCATCGCGTCGATCGACAACGTGAGCTTCATGTTTCAT
 GGGCCACCAGAAGGGCCGCAACACCAAGGAGAACATCTACCGCAATTTGGCATGCCGCAGCCCAACGGCTA
 CCGCAAGGCGCTGCGCTTCATGCGCCTGGCGNNNNNNNNNNCGACAACGTGAGCTTCATGTTTCATGGGCCAC
 CAGAAGGGCCGCAACACCAAGGAGAACATCTACGAAACTTTGGCATGCCGCAGCCCAACGGCTACCGCAAG
 GCGCTGCGCTTCATGCGCCTGGCCGACAAGTTTGGCTTCCCCATCGTCACGTTTGTGGACACGCCCCGGCGCG
 TACGCGGGGCTCAAGGCCGAGGAGCTGGGCCAGGGCGAGGCCATCGCGGTCAACCTGCGCGAGATGTTTGGC
 TTCCGCGTGCCCATCATCTCCATCGTCATCGGCGAGGGCGGCTCCGGCGGCGCGCTGGCCATCGGCTGCGCC
 AACCGCATGCTCATCATGGAGAACGCGGTGTACTACGTGGCCTCTCCCGAGGCCTGCGCGGCCATCCTGTGG
 AAGAGCCGCGCAAGGCGGGCTGGCCACCGAGGCGCTCAAGATCACGCCCGCCGACCTGGTCAAGCTCAAG
 GTCATGGACGAGATCGTGCCGAGCCGCTGGGCGCCGCGCACACCGACCCCATGGGCGCCTTCCCGGCCGAG
 GAGGCCGACTACATCGAGCAGATGGCCGAGCGCTGGCGGTGGTGGGACAGCCTGCTGGAGGACAAGAAGGAC
 ATCGTCAACCCGCCAAAGACCGAGCACCCGGGCTGGCTCATCCCGGGCTTTGCCGAGTTCGCCATCGGCGTC
 GCCAAGGCCGCCGACGCGCGCCAGGCGCGCATGGGGCTCGAGGGCGAGAGCGCCGCGCACGCCAACGACGCC
 GGCATCGGCACCGAGAGGACGGCGGCGCGAGGGCAACGGCGCGAACCGCGAGAACGGGGCGCACGACGAG
 TCGTCTATCGCGGAGACCGAGTGA

SEQ ID. NO:93

Prototheca moriformis (UTEX 1435) 1b-ACCcase-BC-A

MAPASVNLTQVCGGANVEFSPVQGDMDHGLIQHVDISIAFRATPCPAAVAGRRQFPRGAGPSRRSQRSTGS
 ARLTPSASTPSRLVARAETGTNGSAALTPIHKVLIANRGEIAVRVIRACKELGIKTAVYSTADRECLHAQL
 ADEAVCIGEAPSSSESYLNIPTILSAAMSRGADAIHPGYGFLSENAEFVDICKDHGINFIGPGSEHIRVMGDK
 ATARETMKQAGVPTVPGSDGLVETPEEAELEVADQVGFPVMIKATAGGGGRGMRMLAMTRDEFLPLLRQAQGEA
 QAAFNGAVYLERVYKDPRIEFQVLADKHGNIHLGERDCSIQRRNQKLLEAPSPALTPEVRAAMGEAAT
 NAARSIGYVGVGTIEFLWEPAGFYFMEFNTRIXXXXCARSGHSGVCPGGAGETQAWRAIAHSPGTKEVIGWV
 CKDSKERPFKRPRLAGRALESSNSSTGAASRSAANSVGDGINIQRGGARANRLSKENNTQETRAPFGQTK
 RERRLEGQVGTQAVEGAPRHM

SEQ ID. NO: 94

Prototheca moriformis Heteromeric acetyl-CoA carboxylase BC subunit allele 2

MAPASVNLTQVCCSAGHGLRPARATPCPAAVAGRRQFPRGAGPSRRSQRSTGSARLTPSASTPSRLVARAET
 GTNGSAALTPIHKVLIANRGEIAVRVIRACKELGIKTAVYSTADRECLHAQLADEAVCIGEAPSSSESYLNI
 PTILSAAMSRGADAIHPGYGFLSENAEFVDICKDHGINFIGPGSEHIRVMGDKATARETMKQAGVPTVPGSD
 GLVETPEEAELEVADQVGFPVMIKATAGGGGRGMRMLAMTRDEFLPLLRQAQGEAQAAGNGAVYLERVYKDPRI
 HIEFQVLADKHGNIHLGERDCSIQRRNQKLLEAPSPALTPEVRAAMGEAATNAARSIGYVGVGTIEFLWE
 PAGFYFMEFNTRIQQVEHPVTEMITGVDLIQEQIRAAAMGEPLRLTQEDIVIKGHSIECRINAEDPFKGFRRPGP

GRVTAYLPPGGPNVRMDSHVYPDYLVPSNYDSLLGKLIWGEDRQAAIERMKRALDEMVVAGVPTTAPYHLL
ILDHEAFRAGDVTGFIPKHGDDLILPPEAARKKPNVVVEAAKRAHSRSKASV*

SEQ ID. NO:95

- 5 *Prototheca moriformis* (UTEX 1435) Heteromeric acetyl-CoA carboxylase
BCCP-A subunit allele 1

MSGTFYSSPAPGEPPFVKVGDSVKKGQTVCIIEAMKLMNEIEAESSGTIVKVVAEHGKPVTPGAPLFIKIP

SEQ ID. NO:96

- 10 *Prototheca moriformis* (UTEX 1435) Heteromeric acetyl-CoA carboxylase
BCCP-A subunit allele 2

MPSFTKALPRLGGGLVPKPAVSVTDSGKSDLLMRFFESDWFDAWIALTYVALGSEGRSCVEMAGLRAVRRFG
TDWGHPPGEVWVLERVGLLREVWVWPGRSRQSDALGPCIGVGGRARAGFKIARVQLTPLGDDYVTHYKPTRFP
SRPPPLTAGCGDLPVAAHPAVHAAAQQLARARPDRLALAAHCGQDVLAAPRHQPGQPHRHARHRAPRSMR
15 AGRPGRHLGTRGEETEGGRTXXXXPAGGPCRPPAAPPPAPAVEGVEISSPMSGTFYSSPAPGEPPFVKVG
SVKKGQTVCIIEAMKLMNEIEAEVGD SGVVAVKEG

SEQ ID. NO:97

- 20 *Prototheca moriformis* (UTEX 1435) Acetyl-CoA carboxylase alpha-CT
subunit allele 1

MSDPVASGCDMSMTSSQCRVGAWASPMAVARTRAVARPIPRSRNPRLRVAGIDPPDLPKPREEEKKHGDWFO
MLLSRFNPISEKATNATVLD FEKPLVELDHKIREVRKVAEENGVDVTEQIRELEDRAQQLRKETYSKLTPIQ
RLQVARHPNRPTFLDIALNITDKFVELHGDRAGYDDPALVCGIASIDNV SFMFMGHQKGRNTKENIYRNFGM
PQPNGYRKALRFMR LADKFGFPIVTFVDTPGAYAGLKAEE LGQGEAIAVNLRMFGRVPIIISIVIGEGGSG
25 GALAIGCANRMLIMENAVYYVASPEACAA ILWKS RDKAGLATEALKITPADLVKLKVMDEIVPEPLGAAHTD
PMGAFPAVREAILRHWR ELQPLSPVQMRADRYMKFRHIGVYAEH MVPGGQIEAVEAFRESLP GAKTKAGTYA
PTQEEADYIEQMAERWRWDSLLEDKKDIVNPPKTEHPGWLIPGFAEFAIGVAKAADARQARQGLEGESAAH
ANDAGIGTERDGS AEVNGANAKNGAHDES LAEAE*

SEQ ID. NO:98

- 30 *Prototheca moriformis* (UTEX 1435) Acetyl-CoA carboxylase alpha-CT
subunit allele 2

proteinMLLSRFNPISEKATNATVLD FEKPLVELDHKIREVRKVAEENGVDVTEQIRELEDRAQQVVESVW
RGLRKETYSKLTPIQRLQVARHPNRPTFLDIALNITDKFVELHGDRAGYDDPALVCGIASIDNV SFMFMGHQ
35 KGRNTKENIYRNFGMPQPNGYRKALRFMR LADKFGFPIVTFVDTPGAYAGLKAEE LGQGEAIAVNLRMFGR
RVPIIISIVIGEGGSGGAL AIGCANRMLIMENAVYYVASPEACAA ILWKS RDKAGLATEALKITPADLVKLK
MDEIVPEPLGAAHTDPMGAFPAVREAILRH

SEQ ID. NO:99

- 40 *Prototheca moriformis* (UTEX 1435) Acetyl-CoA carboxylase alpha-CT
subunit allele 2

ATGCTGCTCTCGCGCTTCAACCCCATCAGCGAGAAGGCCACCAACGCCACGGTGCTCGACTTTGAGAAACCT
TTGGTGGAACCTTGATCATAAAATTCGCGAGGTCCGCAAAGTCGCGGAGGAGAACGGCGTGACGTCACGGAG
CAGATCCGCGAGCTGGAGGACCGCGCGCAGCAGGTAGTCGAGTCTGTGTGGAGAGGGTTGCGAAAGGAGACG
45 TACTCCAAGCTGACGCCGATCCAGCGGCTGCAGGTGGCGCGCCACCCCAACCGCCCCACTTTCTTGACATT
GCGCTCAACATCACGGACAAGTTTGTGGAGCTGCACGGCGATCGCGCGGGGTACGACGACCCGGCGCTCGTC
TGCGGCATCGCGTCGATCGACAACGTGAGCTTCATGTTTCATGGGCCACCAGAAGGGCCGCAACACCAAGGAG
AACATCTACCGCAATTTCCGGCATGCCGAGCCCAACGGCTACCGCAAGGCGCTGCGCTTCATGCGCCTGGCG
GACAAGTTTGGCTTCCCCATCGTCACGTTTGTGGACACGCGCGCGGTACGCGGGGCTCAAGGCCGAGGAG
50 CTGGGCCAGGGCGAGGCCATCGCGGTCAACCTGCGCGAGATGTTTGGCTTCCGCGTGCCCATCATCTCCATC
GTCATCGGCGAGGGCGGCTCCGGCGGCGCGCTGGCCATCGGCTGCGCCAACCGCATGCTCATATGGAGAAC
GCGGTGTACTACGTGGCTCTCCCGAGGCCTGCGCGGCCATCCTGTGGAAGAGCCGCGACAAGGCGGGCCTG
GCCACCGAGGCGCTCAAGATCACGCGCGCCGACCTGGTCAAGCTCAAGGT CATGGACGAGATCGTGCCGGAG
CCGCTGGGCGCGCGCACACCGACCCCATGGGCGCCTTCCCGGCCGTGCGCGAGGCCATCCTGCGCCAC

SEQ ID. NO:100

- 55 *Prototheca moriformis* (UTEX 1435) 1d-ACCcase-CT-alpha-A

ATGGATCTCTTCCGCGAGGAGGAGGTGCACCTGATGCAGCTCATGATCCCCGCGGAGGTGCGCCACGACACG
ATGGTGGCCCTGGGCGAGGTGGGGATGCTGCAGTTCAAGGACCTCAACCCCGACAAGCCGGCCTTCCAGCGC
ACGTTCCGCCAACAGATCAAGCGGTGCGACGAGATGGCGCGGACGCTGCGCTTCTTCCAGGCGGAGCTGGAC
AAGACGAGGTGCCCGCTTCCGCGCGCGCGGGCGACCCGCGGACCTGGACGAGCTGGAGTGCAG
CTTGCGGCGCTGGAGTCGGAGCTGGTGCAGCTCAACGGCAACGCGGAGCGCCTGGGACTAGGCGGTGCGCC
GCGCCTGTGCTCCCTTTGGAGGAGGAGAAGAAGCACGGCGACTGGTTCCAAATGCTGCTCTCGCGCTTCAAC
CCCATCAGCGAGAAGGCCACCAACGCCACGGTGCTCGACTTTGAGAAACCTTTGGTGGAACTTGATCATAAA
ATTGCGGAGTTGCGAAAGGAGACATACTCCAAGCTGACGCCGATCCAGCGGCTGCAGGTGGCGCGCCACCC

AACCGCCCTACTTTCTTGGACATTGCGCTCAACATCACGGACAAGTTTGTGGAGCTGCACGGCGACCGCGCG
GGGTACGACGACCCCCGCGCTCGTCTGCGGCATCGCGTCGATCGACAACGTGAGCTTCATGTTTCATGGGCCAC
CAGAAGGGCCGCAACACCAAGGAGAACATCTACCGAACTTTGGCATGCCGAGCCCAACGGCTACCGCAAG
5 GCGCTGCGCTTCATGCGCCTGGCCGACAAGTTTGGCTTCCCCATCGTCACGTTTGTGGACACGCCCCGGCGCG
TACGCGGGGCTCAAGGCCGAGGAGCTGGGCCAGGGCGAGGCCATCGCGGTCAACCTGCGCGAGATGTTTGGC
TTCCGCGTGCCCATCATCTCCATCGTCATCGGCAGGGGCGGCTCCGGCGGGCGCGCTGGCCATCGGCTGCGCC
AACCGCATGCTCATCATGGAGAACGCGGTGTACTACGTGGCGTCCCCGGAAGCCTGTGCGGCCATCCTGTGG
AAGAGCCGCGACAAGGCCGGCCTGGCCACCGAGGCGCTCAAGATCACGCCCGCCGACCTGGTCAAGCTCAAG
10 GTCATGGACGAGATCGTACCCGAGCCGCTGGGCGCCGCGCACACCGACCCCATGGGCGCCTTCCCGGCCGCTG
CGCGAGGCCATTGCGCCAGAAAAACGTGCTGGAACCTTTGAGGCGGCGCTGGAGTACAACCCCCGAGGCGTTT
GGCACCGTCAACCATGCTGTACGTGCCCATGACGGTCAACGGCACGCCCCCTCAAGGCCTTTATCGATTGCGGC
GCACAGATGACCATGCTGTCGCGAGCTGCGCGGAGCGGTGCGGGCTGCTCCGTCTCATGGACACGCGGTGG
CAGGGGACAGCCGTGCGGCTGCGGCTCGGCCAAGATCCTGGGGAGGATCCACATGGCTCCGCTGGTTCGCGGCC
15 GGCCACCACCTGCCCATCTCCATCACCGTGTGGATCAGGAGGGCATGGACTTTCTCTTTGGGCTGGACAAC
CTCAAGCGGCACCAATGCTGCATCGACCTGAAGGCAAACGCGCTTCGGTTTGGGTGACAGAGGCCGAGATC
CCGTTTCTGCCGGAGCACGAGGTCCCCAGGCATCAGCACCAGCTCGCCGGGGAGGACGCCGCGAGCGGTAGA
ACCGCAGCCACGGCCCCCTCCCAGCACCGCTCCCGCACCGACCAACCAGCCCACAGTGAAGCAGCTCCGCCA
GCACCAAATGCTCAGCTGGAGGAAAAGGTGTCGAAACTCATGGGCCTGGGCGGCGTGATCCAGCTATGGGAT
20 TACCGCATGGGCACGCTCATCGATCGATTGAGGAGCATGAGGGACCCGTGCGTGGCGTGCACTTTTCATCCC
ACCCAGCCGCTCTTCGTGTCTGGCGGGACGACTACAAGATCAAAGTGTGGAACATAAGACGCGC

SEQ ID. NO:101

Prototheca moriformis (UTEX 1435) Heteromeric acetyl-CoA carboxylase
BCCP-A subunit allele 1

25 ATGTCCGGCACTTTCTACTCTCCCCCGCCCCGGCGAGCCACCGTTTGTCAAGGTGCGCGACAGCGTCAAG
AAGGGCCAGACGGTCTGCATCATCGAGGCCATGAAGCTGATGAACGAGATCGAGGCCGAGTCTTCTGGAACA
ATCGTCAAGTTGTGGCCGAGCACGGCAAGCCTGTACCCCCGGCGCCCCCTGTTTCATCATCAAGCCCTGA

SEQ ID. NO:102

30 *Prototheca moriformis* (UTEX 1435) Heteromeric acetyl-CoA carboxylase
BCCP-A subunit allele 2

35 ATGCCGAGCTTTACCAAGGCGCTCCCCAGGCTGGGAGGGGGACTGGTCCCCAAGCCCGCAGTGAGCGTGACG
GACAGCGGCAAGTCCGACTTGCTCATGCGTTTCTTTGAGAGCGATTGGTTTCGACGCGCTGGATTGCTCTTACG
TACGTTGCGCTCGGTTTCGGAGGGGCGTTTCTGTGTGAGATGGCGGGGTTGCGAGCTGTGAGACGTTTTTGA
ACGGATTGGGGGCACCTGGGGAGTGGGTTTGGCTGGAGCGGCGGGTGGGCTTGTGAGGGAGGTGGTCTGG
CCTGGTCTGTTCAAGGCAGTCTGACGCATTGGGGCCGTGCATTGGCGTTCGGGGGGCGTGCCCCGCGCTGGTTTC
AAAATCGCTCGCGTCCAATTGACGCCATTAGGTGATGATTATGTGACGCATTACAAACCCACCCGGCCCCCG
40 TCGCGCCCCCGTTGACAGCGGGATGTGGAGACCTACCTGTGCGAGCTCACCAGCTGTGCATGCTGCA
GCCCCAACAGCTCGCTCGAGCGCGTCTGATCGACCTCTGCTCGCGCTCGCTGCGCATTGCGGTCAAGACGTA
CTGGCTGCTCCTCGCCATCAGCCAGGACAACCCACAGACACGCACGTATCGCGCTCCGCGATCGATGCGA
GCAGGCCGCCCTGGAAGGCACCTGGGTACGAGGGGAGAGGAGACAGAGGGGGGAGGACNNNNNNNNNNCA
CCGGCCGGCGGGCCTTGTGCGCCTCCCGCCGCGCCGCGCCGCCCCCGCCGCTCGAGGGCGTGGAGATC
AGCTCGCCCATGTCCGGCACCTTCTACTCTCCCCCGCCCCGGCGAGCCGCGCTTGTCAAGGTGCGCGAC
45 AGCGTCAAGAAGGGCCAGACGGTCTGCATCATCGAGGCCATGAAGCTCATGAACGAGATCGAGGCCGAGGTG
GGCGACAGCGGAGTGGTGGCGGTGAAAGAGGGGTAA

SEQ ID. NO:103

Prototheca moriformis (UTEX 1435) Heteromeric acetyl-CoA carboxylase BC
subunit allele 2

50 ATGATCACNGGCGTGACCTGATCCAGGAGCAGATCCGCGCCGCCATGGGCGAGCCGCTGCGCCTGAGACAG
GAGGACATCGTGATCAAGGGCCACAGCATCGAGTGCCGCATCAACGCGGAGGACCCNTTCAAGGGCTTCCGC
CCGGGGCCGGGCGCGTGACCGCTACCTGCCGCCCGGGCGGGCCCAACGTGCGCATGGACAGCCACGTCTAC
CCCGACTACCTGGTGCCCTCCAACCTACGACTCCCTGTGGGCAAGCTCATCGTCTGGGGCGAGGACCGCCAG
GCGGCCATCGAGCGCATGAAGCGCGCGCTGGACGAGATG

SEQ ID. NO:104

55 *Prototheca moriformis* ACC-BC-1 nucleotide sequence

60 ATGGCGCCAGCCTCTGTCAACCTGACCCAGGTTTGTGCTCCGCGGGCATGGTCTGCGTCCCGCGAGGGCG
ACGCCCCGCTGCGCGGTGGCCGGCGGAGGAGTTCCCCCGCGGCGGGCCCCCTCCCGTCCGCTCCAG
CGCTCGACGGGCTCCGCGAGGTTGACGCCCTCTGCCTCCACGCCCTCCGCGCTGGTCCGCCGCGCGAGACG
GGCACCAATGGCTCCGCCGCCCTGACCCCCATCCACAAGGTGCTGATCGCCAACCGTGGCGAGATCGCGGTT
CGCGTGATCCGGGCCCTGCAAGGAGCTGGGCATCAAGACGGTGGCCGTCTACTCCACGGCCGACCGCGAGTGC
CTGCACGCGCAGCTGGCGGACGAGGCGGTGTGCATCGGCGAGGCGCCAGCTCCGAGTCTTACCTCAACATC
CCGACCATCTCTCGGCGGCCATGTGCGCGGCGCGGACGCGATCCACCCCGGCTACGGCTTTCTGTCCGAG
AACGCCGAGTTTGTGGACATCTGCAAGGACCACGGCATCAACTTCATCGGCCCGGGTGGAGCACATCCGC

GTGATGGGCGACAAGGCCACCGCGCGCGAGACGATGAAGCAGGCGGGCGTGCCGACGGTGCCCGGCTCGGAC
 GGGCTGGTGGAGACGCCCCGAGGAGGCGCTGGAGGTGGCGGACCAGGTGGGCTTCCCCGTGATGATCAAGGCC
 ACGGCCGGCGGCGGCGGCGGCGGCGCATGCGCCTGGCCATGACGCGCGACGAGTTTTTGGCGCTGCTGCGCGCG
 5 GCGCAGGGCGAGGCGCAGGCGGCTTTGGCAACGGCGCCGTCTACCTGGAGCGCTACGTGAAGGACCCGCGG
 CACATCGAGTTCAGGTGCTGGCGGACAAGCACGGCAACGTGATCCACCTGGGCGAGCGCGACTGCTCCATC
 CAGCGCCGCAACCAGAAGCTGCTGGAGGAGGCGCCAGCCCCGCGCTACGCCCCGAGGTGCGCGCGGCCATG
 GCGGAGGCGGCCACCAACGCGCGCGCAGCATCGGCTACGTGGGCGTGCGGACCATCGAGTTCCTGTGGGAG
 CCGGCCGGCTTCTACTTTCATGGAGATGAACACGCGCATCCAGGTGGAGCACCCGGTGACGGAGATGATCACG
 10 GGCGTGGACCTGATCCAGGAGCAGATCCGCGCCGCCATGGGCGAGCCGCTGCGCCTGACGCAGGAGGACATC
 GTGATCAAGGGCCACAGCATCGAGTGCCGCATCAACGCGGAGGACCCGTTCAAGGGCTTCCGCCCCGGGGCCG
 GGGCGGTGACCGCTTACCTGCCGCCCGCGGGCCCCAACGTGCGCATGGACAGCCACGTCTACCCCGACTAC
 CTGGTGCCCTCGAAGTACGACTCCCTGCTGGGCAAGTTCATCGTCTGGGGCGAGGACCCGAGGCGGCCATC
 GAGCGCATGAAGCGCGCGCTGGACGAGATGGTCTGTCGCGGCGTGCCACCACCGCGCCCTACCACCTGCTC
 15 ATCCTCGACCACGAGGCCCTTCCGCGCGGGCGACGTGACACGGGCTTCATCCCCAAGCACGGCGACGACCTC
 ATCCTGCCCCCGAGGCGCGCGCAAGAAGCCAAACGTGCTGTCGAGGCCGCCAAGCGCGCCACAGCCGC
 AGCAAGGCCTCCGTCTGA

SEQ ID NO: 105

Codon-optimized *Prototheca moriformis* (UTEX 1435) KASII

ATGGCCACCGCATCCACTTTCTCGGCGTTCAATGCCCGCTGCGGCGACCTGCGTTCGCTCGGCGGGCTCCGGG
 CCCCCGCGCCCAGCGAGGCCCCCTCCCCGTGCGCGGGGCGCGCCGCCGCCGCGCCGACGCCAACCCCGCCCCG
 CCGAGCGCCGCGTGGTATCACCAGGCCAGGCGGTGGTGACCTCCCTGGGCCAGACCATCGAGCAGTTCTAC
 20 TCCTCCCTGCTGGAGGCGGTGTCCGGCATCTCCAGATCCAGAAGTTCGACACCACCGGCTACACCACCACC
 ATCGCCGCGCAGATCAAGTCCCTGCAGCTGGACCCCTACGTGCCAAGCGCTGGGCCAAGCGCGTGGACGAC
 GTGATCAAGTACGTGTACATCGCCGGAAGCAGGCCCTGGAGTCCGCGGCGCTGCCCATCGAGGCCGCGCGG
 CTGGCCGGCGCCGCGCTGGACCCCGCCCTGTGCGGCGTGCTGATCGGCACCGCCATGGCCGGCATGACCTCC
 TTCGCGCGCGCGTGGAGGCCCTGACCCGCGGCGGCGTGCGCAAGATGAACCCCTTCTGCATCCCCCTTCTCC
 30 ATCTCCAACATGGGCGGCGCCATGCTGGCCATGGACATCGGCTTCATGGGCCCAACTACTCCATCTCCACC
 GCCTGCGCCACCGCAACTACTGCATCCTGGGCGCCGCCGACCATCCGCGCGGCGAGCCCAACGTGATG
 CTGGCCGGCGCGCGGCGGCGCCATCATCCCCCTCCGGCATCGGCGGCTTCATCGCCTGCAAGGCCCTGTCC
 AAGCGCAACGACGAGCCCGAGCGCGCTCCCGCCCCCTGGGACGCGGACCGCGACGGCTTCGTGATGGGCGAG
 GGCGCCGGCGTGCTGGTGTGGAGGAGCTGGAGCACGCCAAGCGCCGCGGCGCCACCATCCTGGCCGAGCTG
 35 GTGGGCGGCGCCGCGCCACCTCCGACGCCCCACCATGACCGAGCCCGACCCCCAGGGCCGCGGCGTGCGCCTG
 TGCCTGGAGCGCGCCCTGGAGCGCGCCCGCTGGCCCCCGAGCGCGTGGGCTACGTGAACGCCCCACGGCACC
 TCCACCCCGCGCGGCGAGTGGCCGAGTACCGCGCCATCCGCGCCGTGATCCCCCAGGACTCCCTGCGCATC
 AACTCCACCAAGTCCATGATCGGCCACCTGCTGGGCGGCGCGCGCGCGTGGAGGCCGTGGCCGCCATCCAG
 GCCCTGCGCACCGGCTGGCTGCACCCCAACCTGAACCTGGAGAACCCCGCCCCCGGCGTGGACCCCGTGGT
 40 CTGGTGGGCCCCCGCAAGGAGCGCGCGGAGGACCTGGACGTGGTGTGTCCAACCTCTTCGGCTTCGGCGGC
 CACAACCTCCTGCGTGATCTTCCGCAAGTACGACGAGATGGACTACAAGGACCACGACGGCGACTACAAGGAC
 CACGACATCGACTACAAGGACGACGACGACAAGTGA

SEQ ID NO: 106

Prototheca moriformis (UTEX 1435) KASII

MATASTFSAFNARCGDLRRSAGSGPRRPARPLPVRGRAAAAADANPARPERRVVITGQGVVTSLGQTIEQFY
 SSLLEGVSGISQIQKFDTTGYTTTIAGEIKSLQLDPYVPKRWAKRVDDVIKYVYIAGKQALESAGLP IEAAG
 LAGAGLDPALCGVLIGTAMAGMTSFAAGVEALTRGGVRKMNPFCIPFSISNMGGAMLAMDIGFMPNYSIST
 ACATGNYCILGAADHIRRGDANVMLAGGADAIIIPSGIGGFIACKALSKRNDEPERASRPWDADRDGFVMGE
 45 GAGVLVLEELEHAKRRGATILAEVLVGAATSDAHMTPEPDQGRGVRLCLERALERARLAPERVGYVNAHGT
 STPAGDVAEYRAIRAVIPQDSLRLNSTKSMIGHLLGGAGAVEAVAAIQALRTGWLHPNLNLENPAPGVDPVV
 50 LVGPRKERAEDLDVLSNSFGFGGHNSCVIFRKYDEMDYKDHDG

SEQ ID NO: 107

Prototheca moriformis (UTEX 1435) Acetate kinase 1 (ACK1) allele 1

ATGGTTTTGAACGCGGGGAGTTCGTCCCTCAAGTTCAGGTGTTTGACAAGATCGGGGGCAAGCTGCAGCCC
 CTGGCGTCCGGTCTGTGCGAGCGCATCGGCGACACCGCCAACCTCCCGCATGAAGGCCTTTACGGAAAAGGAC
 GGCAACGTGCTGGTGGAGGAGGTTTCCAGGACCACCAGGACGCGCTCGGGGTCTGTGCGGCTTTCTTGGT
 60 GATGCCTTTTTCGTCCGGCTTACGGAGCGCATGCACAGCGTGGGACACCGCGTGGTGCACGGCCTGACGATC
 AGCGAGCCCGTGCTGATCGACGAGGGCGTGCTCGCGACCATCCGCGAGGCGATCGACCTCGCGCCGCTGCAC
 AACC CGCCGGGCTGCAGGGCATCGAGGCGGCCATGCGCACCTTTGCCTCGGCGCGCACGTGGCCGTGTTT
 GACACGGCCTTCCACCGCTCTGCGCGCGCAACCTGGGGCTTACCTGGCGGTGCGCGCGGACGGCCACGGG
 CGGCCGCGAGTACCGCGTGTACTGCGCGGTCCACTCGGCGCGCGAGCAGGAGCGCGACCGCAAGGCCTGGCAG
 65 GCGCGGCGAGGAGGCGCGCGCCTCGGCCGACCTGGAGGCGCGCGCGGCGGCGGCGGCTCGCGCGCCTGGCC
 GAGCTGGGCTGGGTCTGCTCCTCCAGGTGCACGACGAGGTTCATCCTGGAAGGGCCGCGGAGTCCGCCGAC
 GAGGCGCGCGAGCGGTCTGCTGCTGCTGATGCGCAGCCCTTTACGGGCGCCACCGCGCAGCCCCTGCTCGTC
 GACCTGGTCTGTCAGCCCAAGCACGCGGACACCTGGTACGACGCCAAGTAG

SEQ ID NO: 108

Prototheca moriformis (UTEX 1435) Acetate kinase 1 (ACK1) allele 1, protein

5 MVLNAGSSSLKFKVFDKIGGKLQPLASGLCERIGDTANSRMKAFTEKDGNVLVEEGFQDHDQDALGVVSRFLG
DAFSSGFTERMHVSHRVVHGLTISEPVLIDEGVLATIREAIDLAPLHNPAGLQGIEAAMRTFASAPHVAVF
DTAFHPLCARNLGLYLAVRADGHGRPQYRVYCAVHSAREQERDRKAWQARQEAARASADLEARRGQALARLA
ELGWVLLLQVHDEVILEGPRESADEARERVVACMRSPFTGATAQPLLVDLVDAKHADTWYDAK

SEQ ID NO: 109

10 *Prototheca moriformis* (UTEX 1435) Acetate kinase 1 (ACK1) allele 2

ATGAAGGTATTTCGAGATTTTGGGGGCGGACTTGGGGCTGGCGTGGGTTCGCCTTGCTCCGCCCATGGGGTT
TGGGTGCAGATCACGATGAATGAGAGAGAGAGGGAGGAGAGTGGGGCACTGCGTGGTGGGTGGGGTTTGTG
15 GTCTGCTCCTCTCCAGGTCTCCAGAAGACAAGGAGAAATGGAGGGCCTTTACGGAAGGACGGCAACGTGCT
GGTGGAGGAGGGTTTCCAGGACCACCAGGACGCGCTCGGGGTCTGTGTCGCGCTTTCTTGGCGATGCCTTTTC
GTCCGGCTTCACGGAGCGCATGCACAGCGTGGGGCACCGCGTGGTGCACGGGCTGACGATCAGCGAGCCCGT
GCTGATCGACGAGGGCGTGATCGCGACCATCCGCGAGGCGATCGACCTGGCGCCGCTGCACAACCCGGCCGG
GCTGCAGGGCATCGAGGCGGCCATGCGCACCTTTGCCTCGGCGCCGCACGTSGCCCGTGGCGTGCGCCGCTA
20 CGGCTTCCACGGCTCCAGCTACGCGTACCTCGTCCGCGCGCCGCGGCCGCTGCTGGGCAAGCCCGCCGCGA
GCTGMMSCGCCAGGCGCTCAGGTCCGCCATCCAGCCCTTCCACACGGGGCCGTCGAAGGCTGGACGGTGGCG
GGGCGCTGCTTGTGCGGCGCTGGTGCCGCTCCTCTTCGTGCTGCTGGTGTGCTGACCTCATCGGGCTGTGCGAG
CAGGCTGGCGTTGCGAGCCGTGGGTGCGAGCTCCACGAGGACCCTGCCGTGGTGTGCGGGTGGGGGTGGGTG
AGCGGGACAGCCGTGGTGAGCGGGCTGCGAATGCAGTTGCGCCAATCAGGCGCCTGA

SEQ ID NO: 110

25 *Prototheca moriformis* (UTEX 1435) Acetate kinase 1 (ACK1) allele 2, protein

30 MKVFADFGGGLGAGVGSPPSAHGVWVQITMNEREREESGALRGGWGLLVAPLQVSRQEGEMEGLYGKGRQRA
GGGGFPGPGRARGRVALSWRCLFVRLHGAHAQRGAPRGARADDQARADRRGRDRDHPRGDRPGAAAQPGR
AAGHRGGHAHLCLGAAR?PWRAPLRLPRLQLRVPRPARRGRAGQARRA??PGAQVRHPALPHGAVEGWTV
GPLLVGAGAVLFVAAGADLIGLSQQAGVAGRGLQLHEDPAVVCGWVSGTAVVSGLRMQRLRQSGA

SEQ ID NO: 111

35 *Prototheca moriformis* (UTEX 1435) Acetate kinase 1 (ACK2)

ATGGTTTTGAACGCGGGGAGTTTCGTCCCTGAAGTTCAAGGTGTTTGACAAGATCGGAGGGAAGCTGCAGCCC
CTGGCGTCCGGTCTGTGCGAGCGCATCGGCGACACCGCCAACCTCCCGCATGAAGGCCTTTACGGAAGGAC
40 GGCAACGTGCTGGTGGAGGAGGGTTTCCAGGACCACAGGACGCGCTCGGGGTCTGTGTCGCGCTTTCTTGGT
GATGCTCTTTTCGTCCGGCTTCACGGAGCGCATGCACAGCGTGGGGCACCGCGTGGTGCACGGGCTGACGATC
AGCGAGCCCGTGTGATCGACGAGGGCGTGATCGCGACCATCCGCGAGGCGATCGACCTGGCGCCGCTGCAC
AACCCGGCCGGGCTGCAGGGCATCGAGGCGGCCATGCGCACCTTTGCCTCGGCGCCGCACGTGCGCGTGTTC
GACACGGCTTTCCACCAGACCATGCCGCCCAGCGCCTTCATGTACGCGCTGCCCTACGAGCTGTACGAGGCG
45 CAGGGCGTGCGCCGCTACGGCTTCCACGGCTCCAGCTACGCCTACCTCGTCCCGCGCGCGGCGGCCGCGCTG
GGCAAGCCCGCCGCGGAGCTGAACGCGGTCTGTGCCACCTGGGCGCCGCGCCAGCATGGCCGCCATCCGC
GGCGGGCGCTCGGTGGACACCACCATGGGCCTCACCCGCTGGAGGGCCTKGTCTATGGGCACGCGCTGCGGC
GACATCGACCCCGCCATCGTGACCTACCTGCTGGGCAAGGGCCACTCGGCCAAGGACGTGGACGCGCTCATG
AACAAGAAGAGCGGCTTYGCCGGSCTGGCMGGGCACGCCGACCTGCGCCTGGTTCATCGAGGGCGCCGACAAG
50 GGCGACGAGCGCTGCCGCGTGCAGCTGGAGGTCTGGCGCCACCGCATCCGCAAGTACATCGSGCCTACGCG
ATGCAGCTGGGCGCGCCSCTGGACGCCATCGTCTTCTCCGCCGGCATCGGCGGAGAACAGCTCCATCCTGCGC
AAGTACCTGCTGGAGGACCTGGAGTGGTGGGGCATCGAGCTGGACGAGGGCAAGAACAGTCTCTGCGTGGAC
GGCGTCGCGGGGACATCTCCCGCGAYGGCTCYAAGACCCGCTGCTCGTCATCCCCACCGACGAGGAGCTS
AGCATCGCCGAACAGAGCATCGACGTCGTCAACGGCCTGGCCGCTGA

SEQ ID NO: 112

55 *Prototheca moriformis* (UTEX 1435) Acetate kinase 1 (ACK2), protein

60 MVLNAGSSSLKFKVFDKIGGKLQPLASGLCERIGDTANSRMKAFTEKDGNVLVEEGFQDHDQDALGVVSRFLG
DAFSSGFTERMHVSHRVVHGLTISEPVLIDEGVIAITIREAIDLAPLHNPAGLQGIEAAMRTFASAPHVAVF
DTAFHQTMPPSAFMYALPYELYEAQGVRRYGFHGSSYAYLVPRAAAALGKPAAELNAVVCVLGAGASMAAIR
GGRSVDTTMGLTPLEGLVMGTRCGDIDPAIVTYLLGKHSKADVDALMNKSGFAGLAGHADLRLVIEGADK
GDERCRVALEVRHRIRKYIGAYAMQLGAPLDAIVFSAGIGENSSILRKYLLLEDLEWWGIELDEGKNSLCVD
GVAGDISRDGSKTRLLVIPTDEELSIAEQSIDVNVGLAA

SEQ ID NO: 113

70 *Prototheca moriformis* Phosphate acetyltransferase allele 1

ATGAGTCTGTGCTCTTCTCGGAGTTTGCGGCGGACGCTGAGCCTGCTGAGGCCGGTGGCGGGCTCCACGCCC
CTGTTTGCAGCGCCACCTGTGAGCCGCGGCATCCCGGACGCGGACACGGTCTACTGCACGGACGTCTCCGCG
CAGCGCGCAACTCGCCCATCTGCTGGGCCTGTTTGAGTACTTCCAGCGCCAGTCGCCCCAGTGGGCTTC

TTCCAGCCCATCGGCGCGGACCCGCTGAAGAGCTACAAGCCGGCCCCCGGCGCGCCCAAGCACATCGCGCTC
 ATCCACCAGGCCATGGGCCCTGAAGGACTCGCCCCGAGTCGATGTACGGCGTGACCGAGGACGAGGCCAAGCGC
 CTGCTGACGGCGGGGCGCTACGACGAGCTGGTGGACCGCGTGTTTACCAAGTTCCAGGCGGTCAAGCGCAGC
 AAGGAGATCGTGGTGATCGAGGGCGCCACCGTGCAGGGCATCGGCAACCACAACGAGCTCAACGCGCGCCTG
 5 G CCTCCGAGCTGGACAGCGGCGTGCTGATGATCATGGACCTCAACAGCGACGCGCGGGGAGCGGCGAGGAC
 ATCGCCGGGCGCGCGCTGCGGTACAAGCGCGAGTTTCGAGGCGGAGCGCGCGCGCTGAGCGGGCTCATCCTG
 AACAAAGGTGCCCTTTGGCGCGGAGGAGGCGTGATCGAGGCGGTCCGGCGCGAGCTGGAGGGCCGGGACCTG
 CCTTTGCGGGCGGCTGCCCTACAGCCGCGTCATCGGCACCGCACGCGTGAACGAGATCGTCGAGTCGCTG
 10 GGCGCCAAGCTGCTGTTTCGGGCGGCAGGAGCTGATCGACAGCGACGTCACGGGGTACATGATCGCCGCGCAG
 GCCATCGACGCCTTTCTGACGAACTAGAGACGCTGCGCTCGCAGCGCACGAGCGAGGGCCAGGACTTTTTTC
 CGGCCGCTGGTGGTGACCAACAAGGACCGGACGCTGATCCTGGGCCTGGCGGCCCGCTCCCTGGCGGGC
 ACGGCCCGCGAGCTGGCGGGCTCATCCTCGCGGACGCGGACCTGACCGACCTGACCCCGAGACGCGCGC
 ATCGTTAGCAAGCTCAACCCGGACCTGCCCATGCTGGAGGTGCCCTACGGCACCTTTGAGACGGCGCGCCGC
 15 CTGACGCGCGTGACGCCGGGCATCCTGCCCAGCTCGGTGCGCAAGGTTTCGCGAGGCCAAGGCGCTCTTCCAC
 CGGTACGTGGACGACGACGTCATCGCGGCCGGGCTGGTGC GGCCGAGCGCCCCAGCCTGACCCCCAAGCGC
 TTCTGTACAGATCGAGGAGATCTGCAAGGGCGAGCTGCAGAACGTGGTGCTCCCCGAGTCGCACGACCAC
 CGTGTGCTCACGGCCGCGGCCGAGGTGCGGCCAAGGGCCTGGCGCGCGTGACGCTGCTGGGCAGCCCCGAG
 GAGGTGGATCGGCCGCGCGCCGCTTCAACGTCGACATCAGCAAGTGCGACGTGATCGACTACAAGCATCTCG
 20 CCGAGCTGGACCGGTACGCGGACTGGCTGGTGGAGAAGCGCAAGGCCAAGGGCCTGACCAAGGAGGGCGCG
 CTGGACCAGCTGCAGGACCTCAACATGTACGCCACCGTCATGGTGGCCGTGCGCGACGCGGACGGCATGGTG
 TCGGGCGCGACCTGCACCACGGCCAACACCATCCGGCCTGCGCTGCAGGTGCTCAAGACGCCCCGACCGCAAG
 CTCGTCTCCTCCATCTTCTTCATGTGCCTGCCGACAAGGTCCTCGTCTACGGCGACTGCGCGGTCAACGTG
 ACGCCCCGCGCGGGCGAGCTGGCGCAGATCGCGACCGTGTGCGCCGACACGGCCGCGCCTTTGGCATCGAC
 CCGCGCTGGCCATGCTCTCTACTCGACCTGGGCTTGGCTCGGGGCCGAGGTGGAGTGGATCGGTCACCGAG
 25 GCCACGCGCCTCGCGCGCGAGGCGCGCCCCGACCTCAACATTGAGGGGCCATCCAGTACGACGCGCGCTGTC
 GACCCCGCCGTGGCCGCCACCAAGATCAAGGCGCACTCCGAGGTGCGGGGGCGCGCCTCCGTCTGCATCTTC
 CCGGATCTCAACACGGGGAACAACACCTACAAGGCCGTGCAGCAGTCCACGGGCGCCATCGCCATCGGACCG
 CTGATGCAGGGCCTCGCGCGCCCCGTCAACGATCTCTCCCGCGGTGCACCGTGC CGCGACATCGTCAACACC
 30 GTCGCTGCACGGCCGTCCAGGCCATCGGGCTCAAAGCGCGCGACGCCGAGGCGGCGCGCGGGGGCCGCG
 GCTGCTGCTGCTTGA

SEQ ID NO: 114

Prototheca moriformis Phosphate acetyltransferase, protein

35 MSLSSSRSLRRTL SLLRPVAASTPLFARHLSSRGIPDADTVYCTDVSAQRGNPILLGLFEYFQRQSPDVG
 FQPIGADPLKSYKPAPGAPKHIALIHQAMGLKDSPEMYGVTEDEAKRLLTAGRYDELVDVRVFTKFQAVKRS
 KEIVVIEGATVQIGIHNHNLNARLASELDSGLVLMIMDLNSDAPSGEDIAGRALRYKREFEAERARVSGLIL
 NKVPFGAEEALIEAVRRELEGRDLPFAGGLPYSRVIGTARVNEIVESLGAKLLFGRQELIDSDVTGYMIAAQ
 40 AIDAFLLKLETLRSQRTSEGQDFRPLVVTNKDRQDLILGLAAASLAGTGPQLGGLILADADLTDLTPETRA
 IVSKLNPDLPMLEVPYGTTFETARRLTRVTPGILPSSVRKVREAKALFHRYVDADVIAAGLVRPQRPSLTPKR
 FLYQIEEICKGELQNVLPESHDRVLTAEEVAAGKLARVTLGSPREEVESAAARRFNVDISKCDVIDYKHS
 PELDRYADWLVEKRKAKGLTKEGALDQLDLNMYATVMVAVGDADGMVSGATCTTANTIRPALQVLKTPDRK
 LVSSIIFMCLPDKVLVYGDCAVNVTPAPGELAQIATVSADTAAAFGIDPRVAMLSYSTLGS GSGPQVDMVTE
 45 ATRLAREARPD LNI EGPIQYDAAVDPAVAATKIKAHSEVAGRASVCIFPD LNTGNNTYKAVQQSTGAIAIGP
 LMQGLARPVN DLSRGCTVADIVNTVACTAVQAIGLKARDAEAAARAGAAAAA*

SEQ ID NO: 115

Prototheca moriformis (UTEX 1435) Phosphate acetyltransferase allele 2

50 ATGGTCAACGACTTGGTGGGAGCCATGGCGATGGTCGGCGACGGCAAGCACGGGCGCAGGGCAGCACGGGGC
 GGTCAGCGAGGGGCTTGGGTGCAACTGTGTGGGTGCTCAGCAGGCCCTAATCAATCCAGGGACCTTCTTCA
 GCGCAATTGGCCTCGTTCATCGCGATCGCGATGACATATACTCTCGTCACCTCAAGGGGTTGAGCGAGTTC
 GCATTCAAGTCAAGTGGCATCGGTCTTGCCGGGTTTGGGCATTGTGCCTGGGTCTGATCTGGAACACCGAC
 55 TCGCCGAAGTGGCCATCGGCCGCTCATCGCGGAGTCCCGCCACCCCGCCGCGCTCGTTGATTGGCAGCG
 GAGTCTGTCTTCTCGGAGTTTGCGGCGGACGCTGAGCCTGCTGAGGTGCGATGCAAGGGCGTCCCTGGG
 CGATTTGACGGGCGGTTACCTAAGGCGGAGCGACGTGCGGCTCCGGCGGGCAGCCGACACGTGGGCTTGTG
 CCATGGACAGGCCAGATCCCGATCCGATGGGCCCTGTAACGTTTGCTTATTTGCGACGCCCTTTGCCACG
 60 GCCTGCCCCGCCAGAAATGCCGGGCGAGCGCCCCGTGCGCCGCTTCCCCATGCTTCTCGCCTCGAATCACATT
 TCGACACTCTCCGACGCTCTCTCCACCCCTCCCCGAATCACGAGGCCAGTGGCGGGCTCCACGCACTGTT
 TGCGCGCCACCTGTGAGCGCGCGCATCCCGAGCGGACGCGTACTGACGAGCGTCTCCGCGAGGACGCG
 CGGCAACTCGCCCATCTGCTGGGCTGTTTCAGTACTTCCAGCGCCAGTCCCGGACGCTGGGCTTCTTCCA
 65 GCCCATCGGCGCGGACCCGCTGAGGAGCTACAAGCCTGCCCCGGCGCGCCCAAGCACATCGCGCTCATCCA
 CCAGGCCATGGGCCTGAAGGACTCGCCGAGTGCATGTACGGCGTGACCGAGGACGAGGCCAAGCGCCTGCT
 GACGGCGGGGCGCTACGACGAGCTGGTGGACCGCGTGTTCACCAAGTTCAGGCGGTCAAACGCGAGCAAGGA
 GATTGTGGTGATCGAGGGCGCCACCGTGCAGGGCATCGGCAACCACAACGAGCTCAACGCGCGCCTGGCCTC
 70 CGAGCTGGACAGCGCGCTGCTGATGATCATGGACCTSAACMGSGACGCSGSGSAGCGGCGAGGACATCGC
 CGGGCGCGCGCTGCGGTACAAGCGCGAGTTTCAGGCGGAGCGCGCGCGCTGAGCGGGCTCATCTGAACAA
 GGTGCCCTTTGGCGCGGAGGAGGCGCTGATCGAGGCGGTCCGGCGGAGCTGGAGGGCCGGGACCTGCCCTT
 TGCGGGCGGCGCTGCCCTACAGCCGCGTCATCGGCACCGCGCGCGTGAACGAGATCGTCGAGTCGCTGGGCGC
 CAAGCTGCTGTTTCGGGCGGCASGAGCTGATCGACAGCGACGTCACGGGGTACATGATCGCCGCGCAGGCCAT
 CGACGCSSTTYCTGACSAARCTRGAGACGCTGCGCTCGCAGCGCACGAGCGAGGGCCAGGACTTTTTCCGGCC

GCTGGTGGTGACCAACAAGGACCGGCAGGACCTGATCCTGGGCCTGGCGGCCGCGTCSCTGGCGGGCACGGG
 CCCGAGCTGGGCGGGCTCATCCTSGCSGACGCGGACCTGACCGACCTGACCCCCGAGACGCGCGCGATCGT
 TAG

5 SEQ ID NO: 116
Prototheca moriformis (UTEX 1435) Phosphate acetyltransferase allele 2,
 protein

MVNDLVGAMAMVGDGKHGRRRAARGGQRGAWVQLCGSSAGPNQSRDPSSAQLASFYRDRDDIYSRHLKGLSEF
 10 AFISQWHRSCRWALCLGLIWKPDSPKVPIGRSSRDPTPAGLVDWQHESVVFSEFAADAEPAEVRCKGVPG
 RFDGRLPKAERRAAPAGSRHVGLLPWTGPDPPMGVPVTFAYLPTPFATACPPPEMPGSAPLRFPMLLASNHI
 STLLRRLSTPPRITQASGGVHAPVCAPPVEPRHPGRGHGRLHGRLLRAARQLAHPAGPVRVLPAPVARRGLLP
 AHRRGPAEELQACPRRAQAHRAHPPGHGPEGLARVDVRRDRGRQAPADGGALRRAGGPRVHVQVPGGQTQQG
 DCGDRGRHRAGHRQPQRAQRAPGLRAGQRRADDHGP

15 SEQ ID NO: 117
Prototheca moriformis Lactate dehydrogenase

ATGGCTGAGGGGAGCATCGACTACAAGGTGGCTGTCTTCAGCTCCAGCGAGTGGGTGACCGACTCGTTTCGCC
 20 GAGCCCCCTCAAGATCTTCAAGGAGGTCTCCTACCTGCCGGCCCGCCTGGAGCCGACCAGCGCTGCCCTGGCG
 CACGGCTTTGACGCGGTGTGCATCTTTGTGAACGACGACGCTGGCGCGGAGACGCTCAAGGTTCTGGCGGAG
 GGCGGCGTGAAAGCTGATCGCCATGCGCTGCGCAGGGTACGACCGCGTGGACCTGGCCGCGGCCAAGGAGCTG
 GGCATCCGCGTGGTGCCTGCGGCTGCCGGCTACAGTCCGCGCTCGGTGGCGGAGCACGCGCTGGCGCTGATGATG
 25 GCGTACGAGGGCGAAGCTCAAGGCTCGAGATCAAGTACGCGCGCCGGTCTACAGCTCAACGGCCTGGTG
 GGCGTGGAGCTGACGGGCAAGACCTTTGGCGTGGTGGGACGCGGCCCATCGGGCGCGAGTTCTGTCAGGCTG
 CTCAAGGGCTTTGAGGGCAAGGTGCTGGGCTACGACCCCTACCCGACCGACGCGGCGCGGAGCTGGGCGTG
 GAGTACGTCGACCTGGACACGCTGCTGGCGCAGTCGGACGTCATCTCGCTGCACGTGCCGCTGCTGCCCAGC
 ACGCAGAAGCTGCTGGGCACCGAGTCGCTGAGCAAGCTCAAGCCCACCGCGCTGGTCATCAACGTGAGCCGC
 30 GGCGGGCTCATCGACACCGAGGCTGCATCCAGGCGCTGCAGGAGGACAAGTTTGGCGGCTGGCCGTGCAG
 GTCTACGAGGGCGAGGGCGCGCTCTTCTTCCAGGACTGGGCCAACATGAACGCGCGCCGCTGAAGCAG
 TGGGACCAGAAGTTCATCGCGCTCAAGTCCATGCCCCAGGTGTCGTACGCGCACATCGCCTTCTCAGC
 CACAGCGCGCTCGACGCCATCGCCGCCACCACGCGGACAACCTGCTCGCCGCCGCGCAGGGCAAGGACCTC
 GTCAACGAGGTCAA

35 SEQ ID NO: 118
Prototheca moriformis lactate dehydrogenase protein

MAEGSIDYKVAVFSSSEWVTDSFAEPLKIFKEVSYLPARLEPTSAALAHGFDAVCIFVNDDAGAETLKVLAE
 40 GGVKLIAMRCAGYDRVDLAAAKELGIRVVRVPAYSPRSVAEHALALMMALARNIKASQIKMAAGSYTLNGLV
 GVELTGKTFGVGTGAIGREFVKLLKFEGKVLGYDPYPTDAARELGVEYVDLDTLLAQSDVISLHVPLLP
 TQKLLGTESLSKLKPTALVINVSRLIDTEACIQALQEDKFAGLAVDVYEGEGALFFQDWANMNAARRMKQ
 WDQKFIALKSMPQVIVTPHIAFLTHSALDAIAATTRDNLAAQGKDLVNEV

45 SEQ ID NO: 119
Prototheca moriformis (UTEX 1435) LDP1

ATGGCNGCTGTCGCTGAGAACCCCGCTCCCCCGCCCGAACAGCCTTAAGAAGCTGGGCTTCGTTCCCGAG
 GCCGGCGGAAGAGCTACGAGATTGCCACCAATGTCTACACTACGGCAAAGAGCTACGTTCCCGCTAGCCTG
 CAGCCCAAGCTGGAGAAGGTGGAGGAGTCTGTGACCAGCGTGAGCGCCCGTACGTGGCCAAAGGCCAGGAC
 50 AAGACCGCGACCTGCTCAAGGTGGCCGACGAGAAGGTTGATGAGGCCTTCGCCAAGGCCACCCAGGTCTAC
 CTTAACAACAGCAAGTACGTCGCCGACACCCTGGACAAGCAGCGCGCCTACCACGCGCAGAACCTCGAGTCC
 TACAAGGCCGCGCGGAGCACTACCTCAAGGTGCTCGAAAGCGGCGTTCGAGTACGTGAAGAAGAACGGCATC
 TCCGGCACCGCCAAGGCCGCGCCGACGAGGTGGCCGCGCGCGTTCGAGCGAGGCGCGCGCTCCCGGCGCC
 CTGCTGCACCGCGTGCAGGAGGCGCTGGACAAGCTGCTGGCCAGCACGCCCCTGCACGGCACGGTTCGAGCGC
 55 CTGAAGCCGGCCCTGGACAGCGCCTACCAGCGCTACAGCTCCCTGCACGACAACGTGCTCTCCAGCGACCA
 TACCGCAAGGTGCTGAAGACCGGCTCGGACGTGCTCGCCCGCGTCGAGGCCTCGCCCATCTTCGCCAAGTCC
 AAGGCGACCTGTACCCCTACGTGCGCGCCTACGCGGAGCCCGCGGACGCGCGCTGCAGCCCTACTACCTC
 CGGGTGGTCGAGCACCTCGCGCCCAAGGACGCTGAATTTGTGGACGCGGCGGTTCGACTGCTGCCGCGTTG
 CGCGGTCTCGCTTGATCGCGGGGCTGCCCCCTTGGAGCGCTTGTTCGCGCCGGCGCGACGCGCGCTCCAGT
 60 TGATTGTCGGGGCGGGGCGCGCTCACCGCGCGTGGGCCATCGGAATTGGATCTGCTGAGAATGCAAAGGGCG
 ACCGGTGA

SEQ ID NO: 120
Prototheca moriformis (UTEX 1435) LDP1 protein

MAAVAENPAPPRPNSLKKLGFVPEAGAKSYEIAITNVYTTAKSYVPASLQPKLEKVEESVTSVSAPYVAKAQD
 65 KTATLLKVADEKVDFAKATQVYLNNKYVADTLQRAYHAQNLESYKAAREHYLKVVESGVEYVKKNGI
 SGTAKAAAEVAARVSEARAVPGALLHRVQEAVDKLLASTPVHGTVERLKPALDSAYQRYSSLHDNVVSSDQ
 YRKVVKTGSDVLARVEASPIFAKSKATLYPYVAPYAEPAARLQPYLRVVEHLAPKDA

SEQ ID NO: 121

6S 5' genomic donor sequence

GCTCTTCGCGCGCCGCCACTCCTGCTCGAGCGCGCCCGCGCGTGCGCCGCCAGCGCCTTGGCCTTTTCGCCGC
 GCTCGTGC CGCTCGCTGATGTCCATCACCAGGTCCATGAGGTCTGCCTTGCGCCGGCTGAGCCACTGCTTCG
 TCCGGGCGGCCAAGAGGAGCATGAGGGAGGACTCCTGGTCCAGGGTCTGACGTGGTTCGCCGCTCTGGGAGC
 GGGCCAGCATCATCTGGCTCTGCCGCACCGAGGCGCCTCCAACCTGGTCCTCCAGCAGCCGCAGTCGCCGCC
 GACCCCTGGCAGAGGAAGACAGGTGAGGGGGGTATGAATTGTACAGAACAACCACGAGCCTTGTCTAGGCAGA
 ATCCCTACCAGTCATGGCTTTACCTGGATGACGGCCTGCGAACAGCTGTCCAGCGACCCCTCGCTGCCGCCGC
 TTCTCCCGCACGCTTCTTTCCAGCACCGTGATGGCGCGAGCCAGCGCCGCACGCTGGCGCTGCGCTTCGCCGC
 ATCTGAGGACAGTCGGGGAACCTCTGATCAGTCTAAACCCCTTGC GCGTTAGTGTTGCCATCCTTTGCAGAC
 CGGTGAGAGCCGACTTGTGTGCGCCACCCCCACACCACCTCCTCCAGACCAATTCTGTACACCTTTTGG
 CGAAGGCATCGGCCTCGGCCTGCAGAGAGGACAGCAGTGCCAGCCGCTGGGGGTTGGCGGATGCACGCTCA
 GGTACC

SEQ ID NO: 122

6S 3' genomic donor sequence

GAGCTCCTTGTTTTCCAGAAGGAGTTGCTCCTTGAGCCTTTTCAATTCTCAGCCTCGATAACCTCCAAAGCCGC
 TCTAATTGTGGAGGGGGTTTCAATTTAAAAGCTTGAATGTTGGTTCGTGCGTCTGGAACAAGCCAGACTT
 GTTGCTCACTGGGAAAAGGACCATCAGCTCCAAAAAAGTTCGCCGCTCAAACCGCGTACCTCTGCTTTTCGCCGC
 AATCTGCCCTGTTGAAATCGCCACCACATTATATTGTGACGCTTGAGCAGTCTGTAATTGCCTCAGAATGT
 GGAATCATCTGCCCCCTGTGCGAGCCCATGCCAGGCATGTGCGGGCGAGGACACCCGCCACTCGTACAGCA
 GACCATATGCTACCTCACAATAGTTTATAACAGTGACCATATTTCTCGAAGCTCCCAACGAGCACCTCCA
 TGCTCTGAGTGCGCCACCCCGCGCCCTGGTGCTTGC GAGGGCAGGTCAACCGGCATGGGGCTACCGAAATC
 CCCGACCGGATCCCAACACCCCGCGATGGGAAGAATCTCTCCCGGGATGTGGGGCCACCACCAGCACAAAC
 CTGCTGGCCAGGCGAGCGTCAAACCATAACACAAATATCCTTGGCATCGGCCCTGAATTCCTTCTGCCGC
 CTCTGCTACCCGGTGCTTCTGTCCGAAGCAGGGGTTGCTAGGGATCGCTCCGAGTCCGCAAACCTTGTGCGC
 GTGGCGGGGCTTGTTCGAGCTTGAAGAGC

SEQ ID NO: 123

S. cerevisiae invertase protein sequence

MLLQAFLLFLLAGFAAKISASMTNETSDRPLVHFTPNKGWMNDPNGLWYDEKDAKWHLYFQYNPNDTVWGTP
 FWGHATSDDL TNWEDQPIAIAPKRND SGAFSGSMVVDYNNNTSGFFNDTIDPRQRCVAIWYNTPESEEQYIS
 YSLDGGYTFTEYQKNPVLAA NSTQFRDPKVFWEPSQKWIMTA AKSQDYKIEIYSSDDLKSWKLESAFANE
 FLGYQYECPLIEVPTEQDPSKSYWVMFISINPGAPAGGSFNQYFVGSFNGTHFEAFDNQSRVDFGKDYYA
 LQTFNTDPTYGSALGIAWASNWEYSAFVPTNPWRSSMSLVRKFSLNTEYQANPETELINLKAEPILNISNA
 GPWSRFATNTTLTKANSYNVDLSNSTGTLEFELVYAVNTTQTISKSVFADLSLWFKGLEDPEEYLRMGFEVS
 ASSFFLDRGNSKVFKVKNPYFTNRMSVNNQPFKSENDLSYKYVGLLDQNI LELYFNDGDV VSTNTYFMTT
 GNALGSVNMTTGVDNLFYIDKFQVREVK

SEQ ID NO: 124

S. cerevisiae invertase protein coding sequence codon optimized for expression in *P. moriformis* (UTEX 1435)

ATGctgctgcaggccttctgcttctgctggcggccttcgcccgaagatcagcgcctccatgacgaacgag
 acgtccgaccgccccctggtgcacttcaccccccaacaagggctggatgaacgacccccaacggcctgtggtac
 gacgagaaggacgccaagtggcacctgtacttccagtacaacccgaacgacaccgtctgggggacgcccctg
 ttctggggccacgcccacgtccgacgacctgaccaactgggaggaccagcccacgcccacgcccgaagcgc
 aacgactccggcgccttctccggctccatggtggtgactacaacaacacctccggcttcttcaacgacacc
 atcgaccccgccagcgctgctggccatctggacctacaacacccccggagtcaggaggacgtacatctcc
 tacagcctggacggcggtacaccttcacggactaccagaagaacccccgtgctggcgcccaactccacccag
 ttcccgacccgaaggtcttctggtacgacctccagctggatcatgaccgcccgaagtccaggac
 tacaagatcgagatctactcctccgacgacctgaagtcctggaagctggagtcgcggttcgccaacgagggc
 ttctctcggtaccagtacgagtgccccggcctgatcgaggtccccaccgagcaggacccagcaagtcctac
 tgggtgatgttcatctccatcaacccccggcgcggcgccggcggtccttcaaccagtaactcgtcggcagc
 ttcaacggcaccacacttcgaggccttcgacaaccagtcggcggtggtggacttcgggaaggactactacgcc
 ctgcagaccttcttcaacaccgacccgacctacgggagcgccctgggcatcgcggtgggcctccaactgggag
 tactccgcttcgtgcccaccaacccctggcgctcctccatgtccctcgtgcgcaagttctccctcaacacc
 gagtaccaggcccaacccggagacggagctgatcaacctgaaggccgagccgatcctgaacatcagcaacgcc
 ggccccctggagccggttcgccaccaacaccacgttgacgaaggccaacagctacaacgtcgacctgtccaac
 agcaccggcaccctggagttcgagctggtgtacggcgtcaacaccaccagacgatctccaagtccgtgttc
 gcggacctctccctctggttcaagggcctggaggacccccgaggagtacctccgcatgggcttcgaggtgtcc
 gcgtcctccttcttctggtgaccgcgggaacagcaaggtgaagttcgtgaaggagaacccctacttccaacac
 cgcatagagctgaacaacagcccttcaagagcgaagacgacctgtcctactacaaggtgtacggctgtctg
 gaccagaacatcctggagctgtacttcaacgacggcgacgtcgtgtccaccaacacactacttcatgacaccc
 gggaacgcccctgggctccgtgaacatgacgacgggggtggacaacctgttctacatcgacaagttccaggtg
 cgcgaggtcaagTGA

SEQ ID NO: 125

Chlamydomonas reinhardtii TUB2 (B-tub) promoter/5' UTR

CTTTCTTGCGCTATGACACTTCCAGCAAAAGGTAGGGCGGGCTGCGAGACGGCTTCCCGGCGCTGCATGCAA
 CACCGATGATGCTTCGACCCCCCGAAGCTCCTTCGGGGCTGCATGGGCGCTCCGATGCCGCTCCAGGGCGAG
 CGCTGTTTAAATAGCCAGGCCCCCGATTGCAAAGACATTATAGCGAGCTACCAAAGCCATATTCAAACACCT
 AGATCACTACCACTTCTACACAGGCCACTCGAGCTTGTGATCGCACTCCGCTAAGGGGGCGCCTCTTCTCTCT
 TCGTTTCAGTCACAACCCGCAAAAC

SEQ ID NO: 126

Chlorella vulgaris nitrate reductase 3'UTR

GCGAGCAGCAGCTCGGATAGTATCGACACACTCTGGACGCTGGTCTGTGATGGACTGTTGCCGCCACACTTG
 CTGCCCTTGACCTGTGAATATCCCTGCCGCTTTTATCAAACAGCCTCAGTGTGTTTGATCTTGTGTGTACGCG
 CTTTTGCGAGTTGCTAGCTGCTTGTGCTATTTGCGAATACCACCCCGAGCATCCCCTTCCCTCGTTTCATAT
 CGCTTGTCATCCCAACCGCAACTTATCTACGCTGTCTGCTATCCCTCAGCGCTGCTCCTGCTCCTGCTCACT
 GCCCCCTCGCACAGCCTTGGTTTGGGCTCCGCCTGTATTCTCCTGGTACTGCAACCTGTAAACCAGCACTGCA
 ATGCTGATGCACGGGAAGTAGTGGGATGGGAACACAAATGGAAAGCTT

SEQ ID NO: 127

Nucleotide sequence of the codon-optimized expression cassette of *S. cerevisiae* *suc2* gene with *C. reinhardtii* β -tubulin promoter/5'UTR and *C. vulgaris* nitrate reductase 3' UTR

CTTTCTTGCGCTATGACACTTCCAGCAAAAGGTAGGGCGGGCTGCGAGACGGCTTCCCGGCGCTGCATGCAA
 CACCGATGATGCTTCGACCCCCCGAAGCTCCTTCGGGGCTGCATGGGCGCTCCGATGCCGCTCCAGGGCGAG
 CGCTGTTTAAATAGCCAGGCCCCCGATTGCAAAGACATTATAGCGAGCTACCAAAGCCATATTCAAACACCT
 AGATCACTACCACTTCTACACAGGCCACTCGAGCTTGTGATCGCACTCCGCTAAGGGGGCGCCTCTTCTCTCT
 TCGTTTCAGTCACAACCCGCAAAACGGCGCGCCATGCTGCTGCAGGCCTTCTGTTCTCTGCTGGCCGGCTTCG
 CCGCCAAGATCAGCGCCTCCATGACGAACGAGACGTCCGACCGCCCCCTGGTGCACCTTACCCCCAACAAAGG
 GCTGGATGAACGACCCCCAACGGCCTGTGGTACGACGAGAAGGACGCCAAGTGGCACCTGTACTTCCAGTACA
 ACCCGAACGACACCGCTCTGGGGGACGCCCTTGTCTGGGGCCACGCCACGTCCGACGACCTGACCAACTGGG
 AGGACCAGCCCATCGCCATCGCCCCGAAGCGCAACGACTCCGGCGCCTTCTCCGGCTCCATGGTGGTGGACT
 ACAACAACACCTCCGGCTTCTTCAACGACACCATCGACCCGCGCCAGCGCTGCGTGGCCATCTGGACCTACA
 ACACCCCGGAGTCCGAGGAGCAGTACATCTCCTACAGCCTGGACGGCGGCTACACCTTACCGAGTACCAGA
 AGAACCCCGTGTGGCCGCCAACTCCACCCAGTTCGCGGACCCGAAGGTCTTCTGGTACGAGCCCTCCCAGA
 AGTGGATCATGACCGCGGCCAAGTCCCAGGACTACAAGATCGAGATCTACTCCTCCGACGACCTGAAGTCTCT
 GGAAGCTGGAGTCCGCGTTCGCCAACGAGGGCTTCTCCTCGGCTACCAGTACGAGTGCCCCGGCCTGATCGAGG
 TCCCCACCGAGCAGGACCCAGCAAGTCTACTGGGTGATGTTTATCTCCATCAACCCCGGCGCCCCGGCCG
 GCGGCTCCTTCAACCAGTACTTTCGTCGGCAGCTTCAACGGCACCCACTTCGAGGCCTTCGACAACCAGTCCC
 GCGTGGTGGACTTCGGCAAGGACTACTACGCCCTGCAGACCTTCTTCAACACCGACCCGACCTACGGGAGCG
 CCTGGGCATCGCGTGGGCCCTCCAAC'TGGGAGTACTCCGCCTTCGTGCCCACCAACCCCTGGCGCTCCTCCA
 TGTCCCTCGTGCAGCAAGTCTCCTCAACACCGAGTACCAGGCCAACCCGGAGACGGAGCTGATCAACCTGA
 AGGCCGAGCCGATCCTGAACATCAGCAACCGCGGCCCTGGAGCCGGTTTCGCCACCAACACACAGTTGACGA
 AGGCCAACAGCTACAACGTGACCTGTCCAACAGCACCGGCACCCCTGGAGTTCGAGCTGGTGTACGCCGTCA
 ACACCACCCAGACGATCTCCAAGTCCGTGTTTCGCGGACCTCTCCCTCTGGTTCAAGGGCCTGGAGGACCCCG
 AGGAGTACCTCCGCATGGGCTTCGAGGTGTCCGCGTCTCCTTCTTCTCTGACCGCGGGAACAGCAAGGTGA
 AGTTCGTGAAGGAGAACCCCTACTTCAACCAACCGCATGAGCGTGAACAACAGCCCTTCAAGAGCGGAGAAG
 ACCTGTCTTACTACAAGGTGTACGGCTTGTGAGCAGAACATCTGGAGCTGTACTTCAACGACGGCGAGC
 TCGTGTCCACCAACACCTACTTTCATGACACCGGGAACGCCCTGGGCTCCGTGAACATGACGACGGGGGTGG
 ACAACCTGTTCTACATCGACAAGTTCAGGTGCGCGAGGTCAAGTGACAATTGGCAGCAGCAGCTCGGATAG
 TATCGACACACTCTGGACGCTGGTCTGTGTGATGGACTGTTGCCGCCACACTTGCTGCCTTGACCTGTGAATA
 TCCCTGCCGCTTTTATCAAACAGCCTCAGTGTGTTTGATCTTGTGTGTACGCGCTTTTGCAGGTTGCTAGCT
 GCTTGTGCTATTTGCGAATACCACCCCGAGCATCCCCTTCCCTCGTTTCATATCGCTTGCATCCCAACCGCA
 ACTTATCTACGCTGTCTGCTATCCCTCAGCGCTGCTCCTGCTCCTGCTCACTGCCCTCGCACAGCCTTGG
 TTTGGGCTCCGCTGTATTCTCCTGGTACTGCAACCTGTAAACCAGCACTGCAATGCTGATGCACGGGAAGT
 AGTGGGATGGGAACACAAATGGAGGATCC

SEQ ID NO: 128

Prototheca moriformis (UTEX 1435) Amt03 promoter

GGCCGACAGGACGCGCGTCAAAGGTGCTGGTCTGTATGCCCTGGCCGGCAGGTGCTTGCTGCTGCTGGTTA
 GTGATTCCGCAACCCGTGATTTTGGCGTCTTATTTTGGCGTGGCAAACGCTGGCGCCCGCGAGCCGGGCGCGC
 GGCGATGCGGTGCCCCACGGCTGCCGGAATCCAAGGGAGGCAAGAGCGCCCGGGTCAGTTGAAGGGCTTTAC
 GCGCAAGGTACAGCCGCTCCTGCAAGGCTGCGTGGTGGAAATTGGACGTGCAGGTCCTGCTGAAGTTCTTCCA
 CCGCCTCACCAGCGGACAAAGCACCGGTGTATCAGGTCCGTGTATCCACTCTAAAGAGCTCGACTACGACC
 TACTGATGGCCCTAGATTCTTTCATCAAAAACGCCCTGAGACACTTGCCAGGATTGAAACTCCCTGAAGGGAC
 CACGAGGGCCCTGAGTTGTTTCTTCCCGCGTGGCGAGCTGCCAGCCAGGCTGTACCTGTGATCGAGGCTG
 GCGGGAAAATAGGCTTCGTGTGCTCAGGTGATGGGAGGTGCAGGACAGCTCATGAAACGCCAACAAATCGCAC
 AATTTCATGTCAAGCTAATCAGCTATTTCTCTTACAGAGCTGTAATTGTCCAAAATTCTGGTCTACCGGGG
 GTGATCCTTCGTGTACGGGCCCTTCCCTCAACCCTAGGTATGCGCGCATGCGGTGCGCCGCGCAACTCGCGCG
 AGGGCCGAGGGTTTGGGACGGGCCGTCCCGAAATGCAGTTGCACCCGGATGCGTGGCACCTTTTTTTCGATA

ATTTATGCAATGGACTGCTCTGCAAAATTCTGGCTCTGTGCGCAACCCTAGGATCAGCGGCGTAGGATTTTCG
TAATCATTCGTCCTGATGGGGAGCTACCGACTACCCTAATATCAGCCCGACTGCCTGACGCCAGCGTCCACT
TTTGTGCACACATTCCATTCTGTGCCAAGACATTTATTGTGGTGCGAAGCGTCCCCAGTTACGCTCACCTG
TTTCCCGACCTCCTTACTGTTCTGTGACAGAGCGGGCCACAGGCCGGTTCGCAGCC

5

SEQ ID NO: 129

Chlorella protothecoides (UTEX 250) stearoyl ACP desaturase transit
peptide cDNA sequence codon optimized for expression in *P. moriformis*.

10 ACTAGTATGGCCACCGCATCCACTTTCTCGGCGTTCAATGCCCCGTGCGGCGACCTGCGTCGCTCGGCGGGC
TCCGGGCCCCGGCGCCCAGCGAGGCCCTCCCCGTGCGCGGGCGCGCC

SEQ ID NO: 130

15 ATGGTGGTGGCCGCCGCCAGCAGCGCCTTCTTCCCCGTGCCCCGCCCCCGCCCCACCCCCAAGCCCGGC
AAGTTCGGCAACTGGCCCAGCAGCCTGAGCCAGCCCTTCAAGCCCCAAGAGCAACCCCCAACGGCCGCTTCCAG
GTGAAGGCCAACGTGAGCCCCACGGGCGCGCCCCAAGGCCAACGGCAGCGCCGTGAGCCTGAAGTCCGGC
AGCCTGAACACCCCTGGAGGACCCCCCAGCAGCCCCCCCCCGCACCTTCTGAACAGCTGCCCGACTGG
AGCCGCTGCGCACCGCCATCACCACCGTGTTCGTGGCCGCCGAGAAGCAGTTACCCGCTGGACCGCAAG
20 AGCAAGCGCCCCGACATGCTGGTGGACTGGTTCGGCAGCGAGACCATCGTGCAGGACGGCCTGGTGTTCGCG
GAGCGCTTACAGCATCCGCAGCTACGAGATCGGCGCCGACCGCACCGCCAGCATCGAGACCCTGATGAACCAC
CTGCAGGACACCAGCCTGAACCACTGCAAGAGCGTGGGCCTGCTGAACGACGGCTTCGGCCGACCCCCGAG
ATGTGCACCCGCGACCTGATCTGGGTGCTGACCAAGATGCAGATCGTGGTGAACCGCTACCCCCACCTGGGGC
GACACCTGGAGATCAACAGCTGGTTCAGCCAGAGCGGCAAGATCGGCATGGGCCGCGAGTGGCTGATCAGC
25 GACTGCAACACCGCGAGATCCTGGTGCAGCGCCACCAGCGCCTGGGCCATGATGAACCAGAAGACCCGCGC
TTCAGCAAGCTGCCCTGCGAGGTGCGCCAGGAGATCGCCCCCACTTCTGTGGACGCCCCCCCCCGTGTATCGAG
GACAACGACCGCAAGCTGCACAAGTTCGACGTGAAGACCGGCGACAGCATCTGCAAGGGCCTGACCCCCGGC
TGGAACGACTTCGACGTGAACCAGCAGCTGAGCAACGTGAAGTACATCGGCTGGATTCTGGAGAGCATGCCC
ACCGAGGTGCTGGAGACCCAGGAGCTGTGCAGCCTGACCCTGGAGTACCGCCGCGAGTGCAGGCCGCGAGAGC
30 GTGGTGGAGAGCGTGACAGCATGAACCCAGCAAGGTGGGCCGACCGCAGCCAGTACCAGCACCTGTGCGC
CTGGAGGACGGCGCCGACATCATGAAGGGCCGACCGAGTGGCGCCCCAAGAACGCGGCGACCAACCGCGCC
ATCAGCACCTGA

SEQ ID NO: 131

35 *Cuphea wrightii* FatB2 thioesterase; Gen Bank Accession No. U56104

MVAAAAASSAFFPVPAPRPTPKPGKFGNWPSSLSQPFKPKSNPNGRFQVKANVSPHPKANGSAVSLKSGSLN
TLEDPPSSPPRPTFLNQLPDWSRLRTAITTVFAAEKQFTRLDRKSKRPDMLVDWFGSETIVQDGLVFRERF
40 SIRS YEIGADRTAS IETLMNHLQDTSLNHCKSVGLLNDGFGRTPEMCTRDLIWVLTKMQIVVNRYPTWGDTV
EINSWFSQSGKIGMGREWLISDCNTGEILVRATSAWAMMNQKTRRFSLPCEVRQEIAHPFVDAPPVIEDND
RKLHKFDVKTGDSICKGLTPGWNDFDVNQHVSNVKYIGWILESMPTVLETQELCSLTLEYRRECGRESVVE
SVTSMNP SKVGDRSQYQHLLRLEDGADIMKGRTEWRPKNAGTNRAIST

SEQ ID NO: 137

pSZ1500

GGGCTGGTCTGAATCCTTCAGGCGGGTGTACCCGAGAAAGAAAGGGTGCCGATTTCAAAGCAGACCCATGT
GCCGGGGCCCTGTGGCCTGTGTTGGCGCCTATGTAGTCACCCCCCTCACCCAATTGTGCGCCAGTTTGCGCAC
50 TCCATAAACTCAAAACAGCAGCTTCTGAGCTGCGCTGTTCAAGAACACCTCTGGGGTTTGTCTACCCGCGAG
GTGACGCGCCAGCATGGCTATCAAGACGAACAGGCAGCCTGTGGAGAAGCCTCCGTTACAGATCGGGACGCT
GCGCAAGGCCATCCCCGCGCACTGTTTCGAGCGCTCGGCGCTTCGTAGCAGCATGTACCTGGCCTTTGACAT
CGCGGTATGTCCCTGCTCTACGTCGCGTCGACGTACATCGACCCTGCACCGGTGCCTACGTGGGTCAAGTA
CGGCATCATGTGGCCGCTCTACTGGTCTTCCAGGTGTGTTTGAGGGTTTTGGTTGCCCGTATTGAGGTCTT
55 GGTGGCGCGCATGGAGGAGAAGGCGCCTGTCCCGCTGACCCCCCGGCTACCCTCCCGGCACCTTCCAGGGC
GCCTTCGGCACGGGTGTCTGGGTGTGCGCGCACGAGTGCGGCCACCAGGCCTTTTCTCCAGCCAGGCCATC
AACGACGCGCTGGGCCTGGTGTTCACAGCCTGCTGCTGGTGCCCTACTACTCCTGGAAGCACTCGCACCGG
GTACCCCTTCTTGCGCTATGACACTTCCAGCAAAAGGTAGGGCGGGCTGCGAGACGGCTTCCCGGCGCTGCA
TGCAACACCGATGATGCTTCGACCCCCGAAAGCTCTTGGGGCTGCATGGGCGCTCCGATGCCGCTCCAGG
60 GCGAGCGCTGTTTAAATAGCCAGGCCCCGATTGCAAGACATTATAGCGAGCTACCAAAGCCATTATCAAA
CACCTAGATCACTACCCTTCTACACAGGCCACTCGAGCTTGTGATCGCACTCCGCTAAGGGGGCGCCTCTT
CCTCTTCTGTTTTCAGTCACAACCCGCAAACTCTAGAATATCAATGCTGCTGCAGGCCTTCTGTTCTGCTGG
CCGGCTTCGCGGCCAAGATCAGCGCCTCCATGACGAACGAGACGTCCGACCGCCCCCTGGTGCATTCACCC
CCAACAAGGGCTGGATGAACGACCCCAACGGCCTGTGGTACGACGAGAAGGACGCCAAGTGGCACCTGTACT
65 TCCAGTACAACCCGAACGACACCGCTTGGGGGACGCCCTTGTCTGGGGCCACGCCACGTCCGACGACCTGA
CCAACTGGGAGGACCAAGCCATCGCCCTCGCCCCGAAGCGCAACGACTCCGGCGCCTTCTCCGGCTCCATGG
TGGTGGACTACAACAACACCTCCGGCTTCTTCAACGACACCATCGACCCGCGCCAGCGCTGCGTGGCCATCT
GGACCTACAACACCCCGGAGTCCGAGGAGCAGTACATCTCTACAGCCTGGACGGCGGCTACACCTTACCCG
AGTACCAGAAGAACCCCGTGTGCGCGCCAACTCCACCCAGTTCCGCGACCCGAAGGTCTTCTGGTACGAGC
70 CCTCCAGAAAGTGGATCATGACCGCGGCCAAGTCCAGGACTACAAGATCGAGATCTACTCCTCCGACGACC
TGAAGTCTTGGAAGCTGGAGTCCGCGTTTCGCCAAGGAGGGCTTCTCGGCTACCAGTACGAGTGCCCCGGCC

TGATCGAGGTCCCCACCGAGCAGGACCCCCAGCAAGTCTACTGGGTGATGTTTCATCTCCATCAACCCCGGCG
CCCCGGCCGGCGGCTCCTTCAACCAGTACTTCGTGCGGAGCTTCAACGGCACCCACTTCGAGGCCTTCGACA
ACAGTCCCAGCTGGTGGACTTCGGCAAGGACTACTACGCCCTGCAGACCTTCTTCAACACCGACCCGACCT
ACGGGAGCGCCCTGGGCATCGCGTGGGCCTCCAACCTGGGAGTACTCCGCCTTCGTGCCCACCAACCCCTGGC
5 GCTCCTCCATGTCCCTCGTGCGCAAGTTCCTCCCTCAACACCGAGTACCAGGCCAACCCGGAGACGGAGCTGA
TCAACCTGAAGCCGAGCCGATCCTGAACATCAGCAACGCCGGGCCCTGGAGCCGGTTCGCCACCAACACCA
CGTTGACGAAGGCCAACAGCTACAACGTCGACCTGTCCAACAGCACCGGCACCCTGGAGTTCGAGCTGGTGT
ACGCCGTCAACACCCACCCAGACGATCTCCAAGTCCGTGTTGCGGACCTCTCCCTCTGGTTCAAGGGCCTGG
10 AGGACCCCGAGGAGTACCTCCGCATGGGCTTCGAGGTGTCCGCGTCTCCTTCTTCTGGACCCGCGGGAACA
GCAAGGTGAAGTTCGTGAAGGAGAACCCCTACTTCACCAACCGCATGAGCGTGAACAACAGCCCTTCAAGA
GCGAGAACGACCTGTCTTACTACAAGGTGTACGGCTTGCTGGACCAGAACATCCTGGAGCTGTACTTCAACG
ACGGCGACGTCGTGTCCACCAACACCTACTCATGACCACCGGGAACGCCCTGGGCTCCGTGAACATGACGA
CGGGGGTGGACAACCTGTTCTACATCGACAAGTTCAGGTGCGCGAGGTCAAGTGACAATTGGCAGCAGCAG
15 CTGGGATAGTATCGACACACTCTGGACGCTGGTGTGTGATGGACTGTTGCCGCCACACTTGCTGCCTTGAC
CTGTGAATATCCCTGCCGCTTTTATCAAACAGCCTCAGTGTGTTTGATCTTGTGTGTACGCGCTTTTGCGAG
TTGCTAGCTGCTTGTGCTATTTGCGAATACCACCCCGAGCATCCCCTTCCCTCGTTTTCATATCGCTTGCATC
CCAACCGCAACTTATCTACGCTGTCTGCTATCCCTCAGCGCTGCTCCTGCTCCTGCTCACTGCCCCCTCGCA
CAGCCTTGCTTGGCTCCGCTGTATCTCTGGTACTGCAACCTGTAAACCCAGCAGTATGCTGATGTCGATG
20 ACGGGAAGTAGTGGGATGGGAACACAAATGGAGGATCCGCGCTCTCGAACAGAGCGCGCAGAGGAACGCTGA
AGGTCTCGCCTCTGTGCGACCTCAGCGCGGCATACACCACAATAACACCTGACGAATGCGCTTGCTTCTTC
GTCCATTAGCGAAGCGTCCGGTTCACACACGTGCCACGTTGGCGAGGTGGCAGGTGACAATGATCGGTGGAG
CTGATGGTCAAACGTTACAGCCTAGGGATATCGAATTCGGCCGACAGGACGCGCGTCAAAGGTGCTGGTC
GTGTATGCCCTGGCCGGCAGGTGCTGCTGCTGGTGTAGTGATTCCGCAACCTGATTTTGGCGCTTAT
25 TTTGGCTGGCAAACGCTGGCGCCGCGAGCCGGGCGGCGATGCGGTGCCCGGACGTTTCCGGAATCC
AAGGGAGGCAAGAGCGCCCGGGTCAGTTGAAGGGCTTTACGCGCAAGGTACAGCCGCTCCTGCAAGGCTGCG
TGGTGGAAATTGGACGTGCAGGTCTGCTGAAGTTCCTCCACCGCTCACCAGCGGACAAAGCACCGGTGTAT
CAGGTCCGTGTCTACCTCTAAAGAACTCGACTACGACCTACTGATGGCCCTAGATTCTTCATCAAAAACG
30 CTTGAGACACTTGCCCGAGGATTGAACTCCCTGAAGGGACACCAGGGGCCCTGAGTTGTTCTTCCCCCG
TGGCGAGCTGCCAGCCAGGCTGTACCTGTGATCGAGGTGGCGGGAAAATAGGCTTCGTGTGCTCAGGTTCAT
GGAGGTGCGAGCAGCTCATGAAACGCCAACAAATCGACAATTTCATGTCAAGCTAATCAGCTATTTCTCT
TCACGAGCTGTAATTGTCCAAAATTTCTGGTCTACCGGGGTGATCCTTCGTGTACGGGCCCTTCCCTCAAC
CCTAGGTATGCGCGCATGCGGTGCGCGCGCAACTCGCGCGAGGGCCGAGGGTTTGGGACGGGCCGTCCCGAA
ATGCAAGTTGCACCCGATGCGTGGCACCTTTTTTTCGATAATTTATGCAATGGACTGCTCTGCAAAATTCTG
35 GCTCTGTGCGCAACCCCTAGGATCAGCGCGTAGGATTTCTGAATCATTCTGCTCCTGATGGGGAGCTACCGACT
ACCTAATATCAGCCCGACTGCCCTGAGCCAGCGTCCACTTTTGTGCACACATTCCATTCTGTGCCCAAGACA
TTTCATTGTGGTGGCAAGCGTCCCGAGTTACGCTACCTGTTTCCCGACCTCCTTACTGTTCTGTGCGACAGA
GCGGGCCACAGGCCGGTTCGCGCCACTAGTATGGCCACCGCATCCACTTTCTCGGCGTTCAATGCCCGCTG
CGGCGACCTGCGTTCGCTCGGCGGGCTCCGGGCCCGCGCGCCAGCGAGGGCCCTCCCGTGGCGGGCGCGC
40 CGCCACCGCGGAGCAGCCCTCCGGCGTGGCCTCCCTGCGCGAGGCGGACAAGGAGAAGTCCCTGGGCAACCG
CCTGCGCCTGGGCTCCCTGACCGAGGACGGCCTGTCTACAAGGAGAAGTTCGTGATCCGCTGCTACGAGGT
GGGCATCAACAAGACCGCCACCATCGAGACCATCGCCAACCTGCTGCAGGAGGTGGGCGGCAACACGCCCCA
GGGCGTGGGCTTCTCCACCGACGGCTTCGCCACCACCAACCATGCGCAAGCTCCTTACTGTTCTGTGCGACAG
45 CGCCCGCATGCACATCGAGATCTACCGCTACCCCGCTGGTCCGACGTGATCGAGATCGAGACCTGGGTGCA
GGGCGAGGGCAAGGTGGGCACCCGCGCGACTGGATCCTGAAGGACTACGCCAACGGCGAGGTGATCGGCCG
CGCCACCTCCAAGTGGGTGATGATGAACGAGGACACCCGCGCCTGCAGAAGGTGTCCGACGACGTGCGCGA
GGAGTACCTGGTGTCTGCCCCCGCACCCCTGCGCCTGGCCTTCCCCGAGGAGAAACAACAACTCCATGAAGAA
50 GATCCCCAAGCTGGAGACCCGCGGAGTACTCCCGCTGGGCTGGTGGCCCCGCTCCGACCTGGACAT
GAACAAGCAGCTGAACAACGTGACCTACATCGGTGGGCTGGAGTCCATCCCCCGAGATCATCGACAC
CCACGAGCTGCAGGCCATCACCCTGGACTACCGCGCGAGTGCCAGCGCGACGACATCGTGGACTCCCTGAC
CTCCCGGAGCCCCCTGGGCAACGCCCGCGGCTGAAGTTCAAGGAGATCAACGGCTCCGTGTCCCCAAGAA
GGACGAGCAGGACCTGTCCCGCTTCATGCACCTGCTGCGCTCCGCGGCTCCGGCCTGGAGATCAACCGCTG
CCGCACCGAGTGGCGCAAGAAGCCCGCCAAGCGCATGGACTACAAGGACCACGACGGCGACTACAAGGACCA
CGACATCGACTACAAGGACGACGACGACAAGTGAATCGATGATCTCTTAAGGCAGCAGCAGCTCGGATAGT
55 ATCGACACACTCTGGACGCTGGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT
CCCTGCCGCTTTTATCAAACAGCCTCAGTGTGTTTGATCTTGTGTGTACGCGCTTTTGCGAGTTGCTAGCTG
CTTGTGCTATTTGCGAATACCACCCCGAGCATCCCCTTCCCTCGTTTTCATATCGCTTGCATCCCAACCGCAA
CTTATCTACGCTGTCTGCTATCCCTCAGCGCTGCTCCTGCTCCTGCTCACTGCCCCCTGCGACAGCCTTGGT
TTGGGCTCCGCTGTATTTCTCTGGTACTGCAACCTGTAAACAGCACTGCAATGCTGATGCACGGGAAGTA
60 GTGGGATGGGAACACAAATGGAAGCTTAATTAAGAGCTCCCGCCACCACTCCAACACGGGGTGCCTGGACA
AGGACGAGGTGTTTGTGCCCGCGCAGTCAGTGGCGCACGAGGGCCTGGAGTGGGAGGAGTGGCTGGCCCA
TCCGCATGGGCAAGGTGCTGGTACCCCTGACCCTGGGCTGGCGCTGTACCTCATGTTCAACGTGCGCTCGC
GGCCGTACCCGCGCTTCGCCAACCCTTTGACCCGTGGTGGCCATCTTCAGCAAGCGCGAGCGCATCGAGG
TGGTCATCTCCGACCTGGCGCTGGTGGCGGTGCTCAGCGGGCTCAGCGTGTGGGCCGACCATGGGCTGGG
CCTGGCTGGTCAAGACCTACGTGGTGGCTTACCTGATCGTGAACATGTGGCTCGTGTCTCATACGCTGCTCC
65 AGCACACGACCCGGCGCTGCCGCACTACTTCGAGAAGGACTGGGACTGGCTGCGCGGGCCCATGGCCACCG
TGGACCGTCCATGGGCGCCCTTTCATGGACAACATCTGCACCAATCTCCGACCAACCCAGTGTGACAC
ACCTCTTCAGCACCATCCCGCACTACCACGCGGAGGAGGCTCCGCGCCCATCAGGCCCATCCTGGGCAAGT
ACTACCAGTCCGACAGCCGCTGGGTGGCGCGGCCCTGTGGGAGGACTGGCGCGACTGCCGCTACGTGCTCC
70 CGGACGCGCCCGAGGACGACTCCGCGCTCTGGTTCCACAAGTGAGTGAGTGA

SEQ ID NO: 138

5' FADc genomic region donor DNA

GGGCTGGTCTGAATCCTTCAGGCGGGTGTACCCGAGAAAGAAAGGGTGCCGATTTCAAAGCAGACCCATGT
GCCGGGGCCCTGTGGCCTGTGTTGGCGCCTATGTAGTCACCCCCCTCACCCAATTGTGCGCCAGTTTGCGCAC
TCCATAAACTCAAAACAGCAGCTTCTGAGCTGCGCTGTTCAAGAACACCTCTGGGGTTTGTCTACCCGCGAG
GTCGACGCCCCAGCATGGCTATCAAGACGAACAGGCAGCCTGTGGAGAAGCCTCCGTTACAGATCGGGACGCT
GCCGAAGGCCATCCCCGCGCACTGTTTCGAGCGCTCGGCGCTTCGTAGCAGCATGTACCTGGCCTTTGACAT
CGCGGTTCATGTCCCTGCTCTACGTCGCGTCGACGTACATCGACCCTGCACCGGTGCCTACGTGGGTCAAGTA
CGGCATCATGTGGCCGCTCTACTGGTTCCTCCAGGTGTGTTTGAGGGTTTTGGTTGCCCCGATTGAGGTCCCT
GGTGGCGCGCATGGAGGAGAAGGCGCCTGTCCCGCTGACCCCCCGGCTACCTCCCGGCACCTTCCAGGGC
GCCTTCGGCACGGGTGTCTGGGTGTGCGCGCACGAGTGCGGCCACCAGGCCTTTTCTCCAGCCAGGCCATC
AACGACGGCGTGGGCTGGTGTTCACAGCCTGCTGCTGGTGCCCTACTACTCCTGGAAGCACTCGCACCG

SEQ ID NO: 139

3' FADc genomic region donor DNA

CCGCCACCACTCCAACACGGGGTGCCTGGACAAGGACGAGGTGTTTGTGCCGCCGCACCGCGCAGTGGCGCA
CGAGGGCCTGGAGTGGGAGGAGTGGCTGCCCATCCGCATGGGCAAGGTGCTGGTCACCTGACCTGGGCTG
GCCGCTGTACCTCATGTTCAACGTCGCCTCGCGGCCGTACCCGCGCTTCGCCAACCACCTTTGACCCGTGGTC
GCCCCATCTTCAGCAAGCGCGAGCGCATCGAGGTGGTTCATCTCCGACCTGGCGCTGGTGGCGGTGCTCAGCGG
GCTCAGCGTGCTGGGCCGCGACCATGGGCTGGGCTGGTCAAGACCTACGTGGTGCCCTACCTGATCGT
GAACATGTGGCTCGTGCTCATCACGCTGCTCCAGCACACGCACCCGCGCTGCCGCACTACTTCGAGAAGGA
CTGGGACTGGCTGCGCGCGGCCATGGCCACCGTGGACCGCTCCATGGGCCCGCCCTTCATGGACAACATCCT
GCACCACATCTCCGACACCCACGTGCTGCACCACCTCTTCAGCACCATCCCGCACTACCACGCCGAGGAGGC
CTCCGCCGCCATCAGGCCCATCCTGGGCAAGTACTACCAGTCCGACAGCCGCTGGGTGCGCCGCGCCCTGTG
GGAGGACTGGCGCGACTGCCGCTACGTGCTCCCGACGCGCCCGAGGACGACTCCGCGCTCTGGTTCCACAA
GTGAGTGAGTGA

SEQ ID NO: 140

5' donor DNA sequence of *Prototheca moriformis* FATA1 knockout homologous recombination targeting construct

GCTCTTCGGAGTCACTGTGCCACTGAGTTCGACTGGTAGCTGAATGGAGTCGCTGCTCCACTAAACGAATTG
TCAGCACCGCCAGCCGGCCGAGGACCCGAGTCATAGCGAGGGTAGTAGCGCGCCATGGCACCGACCGAGCCTG
CTTGCCAGTACTGGCGTCTCTTCCGCTTCTGTGGTCTCTGCGCGCTCCAGCGCGTGGCGTTTTCCGGTG
GATCATGCGGTCCGTGGCGCACCGCAGCGGGCCGCTGCCATGCAGCGCCGCTGCTTCCGAACAGTGGCGGT
AGGGCCGCACCCGCGGTAGCCGTCCGTCCGGAACCCGCCAAGAGTTTTGGGAGCAGCTTGAGCCCTGCAAG
ATGGCGGAGGACAAGCGCATCTTCTGGAGGAGCACCGGTGCGTGGAGGTCCGGGGCTGACCGGCCGTGCGA
TTCAACGTAATCAATCGCATGATGATCAGAGGACACGAAGTCTTGGTGGCGGTGGCCAGAAACACTGTCCAT
TGCAAGGGCATAGGGATGCGTTCTTCCCTCTCATTTCTCATTTCTGAATCCCTCCCTGCTCACTCTTTCT
CCTCCTCCTTCCCGTTACGCAGCATTCGGGGTACC

SEQ ID NO: 141

3' donor DNA sequence of *Prototheca moriformis* FATA1 knockout homologous recombination targeting construct

GACAGGGTGGTTGGCTGGATGGGGAACGCTGGTTCGCGGGATTTCGATCCTGCTGCTTATATCCTCCCTGGAA
GCACACCCACGACTCTGAAGAAGAAAACGTGCACACACACAACCCAACCGGCCGAATATTTGCTTCCTTATC
CCGGGTCCAAGAGAGACTGCGATGCCCCCTCAATCAGCATCCTCCTCCCTGCGCTTCAATCTTCCCTGCT
TGCTGCGCCCGCGGTGCGCCGTCTGCCCGCCAGTCAGTCACTCCTGCACAGGCCCCCTGTGCGCAGTGCT
CCTGTACCCCTTACCGCTCCTTCCATTCTGCGAGGCCCCCTATTGAATGTATTGTTGCTGTGTGGCCAAG
CGGGTGTGGGCGCGCCGCGTCCGGGAGTGCTCGGCGACTTTGGCGGAAGCCGATTGTTCTTCTGTAAGC
CACGCGCTTGTGCTTTGGGAAGAGAAGGGGGGGGTACTGAATGGATGAGGAGGAGAAGGAGGGGTATTGG
TATTATCTGAGTTGGGTGAAGAGC

SEQ ID NO: 142

Chlorella protothecoides actin promoter/5'UTR

ACTAGAGAGTTTAGGTCCAGCGTCCGTGGGGGGGACGGGCTGGGAGCTTGGGCCGGGAAGGGCAAGACGAT
CGAGTCCCTCTGGGAGTCACAGCCGACTGTGTGTGTTGCACTGTGCGGCCCGCAGCACTCACACGCAAAAT
GCCTGGCCGACAGGCAGGCCCTGTCCAGTGCAACATCCACGGTCCCTCTCATCAGGCTCACCTTGCTCATTTG
ACATAACGGAATGCGTACCGCTCTTTCAGATCTGTCCATCCAGAGAGGGGAGCAGGCTCCCCACCGACGCTG
TCAAACCTTGCTTCCCTGCCCAACCGAAAACATTATTGTTTGAGGGGGGGGGGGGGGGGGGAGATTGCATGGCG
GGATATCTCGTGAGGAACATCACTGGGACACTGTGGAACACAGTGAGTGAGTATGCAGAGCATGTATGCTA
GGGGTCAGCGCAGGAAGGGGGCCTTTCCAGTCTCCCATGCCACTGCACCGTATCCACGACTCACCAGGACC
AGCTTCTTGATCGGCTTCCGCTCCCGTGGACACCACTGTGTAGCCTCTGGACTCCAGGTATGCGTGCACCGC
AAAGGCCAGCCGATCGTGCCGATTCTTGGGGTGGAGGATATGAGTCAGCCAACCTTGGGGCTCAGAGTGACACA
CTGGGGCACGATACGAAACAACATCTACACCGTGTCTCCATGCTGACACACCACAGCTTCGCTCCACCTGA
ATGTGGGCGCATGGGCCCCGAATCACAGCCAATGTGCTGCTGCCATAATGTGATCCAGACCCTCTCCGCCCA
GATGCCGAGCGGATCGTGGGCGTGAATAGATTCTGTTTCGATCACTGTTTGGGTCTTTCTTTTCTGCTCT
CGGATGCGCGTCTCGAAACAGGCTGCGTGGGCTTTCCGATCCCTTTTGCTCCCTCCGTACCATCCTGCGC

CGCGGGCAAGTTGCTTGACCCCTGGGCTGGTACCAGGGTTGGAGGGTATTACCGCGTCAGGCCATTCCCAGCCC
GGATTCAATTCAAAGTCTGGGCCACCACCCTCCGCCGCTCTGTCTGATCACTCCACATTCGTGCATACACTA
CGTTCAAGTCTCTGATCCAGGCGTGTCTCGGGACAAGGTGTGCTTGAGTTTGAATCTCAAGGACCCACTCCAG
CACAGCTGCTGGTTGACCCCGCCCTCGCAA

5

AACGCCTTTGTACAACTGCA

SEQ ID NO: 143

10 AtTHIC expression cassette comprising *Chlorella protothecoides* actin promoter/5'UTR, *Arabidopsis thaliana* THIC protein coding sequence codon-optimized for expression in *Prototheca moriformis*, and *Chlorella vulgaris* nitrate reductase 3' UTR

15 agtttaggtccagcgtccgtggggggggacgggctgggagcttggggccgggaagggcaagacgatgcagtc
ctctggggagtcacagccgactgtgtgtgttgactgtgcggcccgagcactcacacgcaaaatgcctggc
cgacaggcagggccctgtccagtgcaacatccacggctccctctcatcaggctcaccttgctcattgacataac
ggaatgctaccgctctttcagatctgtccatccagagaggggagcaggctccccaccgacgctgtcaaact
tgcttctgcccacgaaaacattattgtttgagggggggggggggggggggcagattgcatggcgggatatc
20 tctgagggaacatcactgggacactgtggaacacagtgagtgagtgatgcagagcatgtatgctaaggggtca
gcgaggaagggggcctttccagctctccatgccactgcacgctatccagcactcaccaggaccagcttct
tgatcggcttccgctcccggtggacaccagtggtagcctctggactccaggtatgcgtgcaccgcaaagggc
agccgatcgtgccgattcctggggtggaggatagagtcagccaacttggggctcagagtgacactggggc
acgatacgaacaacatctacacgctgtcctccatgctgacacaccacagcttcgctccacctgaatgtggg
25 cgcagtgggccgaatcacagccaatgtcgctgctgccataatgtgatccagacctctcgccagatgcgg
agcggatcgtgggcgctgaatagattcctggttcgatcactggttgggtcctttcctttctcgtctcggatgc
gctctcgaaacaggctgcgtcgggctttcggatcccttttgcctccctccgtcaccatcctgcgcggggca
agttgcttgacctgggctgtaccagggttgagggtattaccgcgtcaggccattcccagcccggtattcaa
ttcaaagtctgggccaccaccctccgcccgtctgtctgatcactccacattcgtgcatacactacgttcaag
tctgatccaggcgtgtctcgggacaagggtgtgcttgagttgaatctcaaggacccactccagcacagctg
30 ctggttgaccgcctcgcattctagaATGgccgcgtccgtccactgcacctgatgtccgtggtctgcaa
caacaagaaccactccgcccgcggcccaagctgccaaactcctccctgctgccgggttcgacgtgggtggtcca
ggccgcggccaccgcgttcaagaaggagacgacgaccaccgcggccacgctgacgttcgacccccccacgac
caactccgagcgcgccaagcagcgaagcacaccatcgacccctcctccccgacttccagcccatccctc
cttcgaggagtgttccccaaagtccacgaaggagcacaaggaggtggtgcacgaggagtccggccacgtcct
35 gaaggtgcccttccgcccgtgcacctgtccggcggcgagcccgcccttcgacaactacgacacgtccggccc
ccagaacgtcaacgcccacatcggcctggcgaagctgcgaaggagtggtcgacccgcgcgagaagctggg
cacgccccgtacacgcagatgtactacgcgaagcagggcatcatcacggaggagatgctgtaactgcgcgac
gcgcgagaagctggacccccgagttcgtccgctccgaggtcgcgccggggccgcgccatcatccctccaaca
gaagcacctggagctggagcccatgatcgtgggcccgaagttcctggtgaaggtgaacgcgaacatcggcaa
40 ctccgccgtggcctcctccatcgaggaggaggtctacaaggtgcagtgggccaccatgtggggcgccgacac
catcatggacctgtccacgggcccacatccacgagacgcgcgagtggtcctgcgcaactccgcccgtccc
cgtgggcccagctcccctaccagcgctggagaaggtggacggcatcgcgagacatccgcaagtacgggaggt
gttccgcgagacgctgatcgagcaggccgagcagggcgtggactacttcacgatccacgcgggctgctgct
45 gcgctacatccccctgaccgccaagcgctgacgggcatcgtgtcccgcggcggtccatccacgcgaagtg
gtgcctggcctaccacaaggagaacttcgcctacgagcactgggacgacatcctggacatctgcaaccagta
cgacgtcgccctgtccatcggcgacggcctgcgcccgggtccatctacgacgccaacgacacggccagtt
cgccgagctgctgaccagggcgagctgacgcgcgcgctgggagaaggacgtgcaggtgatgaacgaggg
ccccggccacgtcccctgcacaagatccccgagaacatcagaagcagctggagtggtgcaacgagggcgcc
cttctacacctgggccccctgacgaccgacatcgcccccgtacgaccacatcacctccgcatcggcgc
50 ggccaacatcggcgccctgggacccgctgctgtgctacgtgacgcccgaaggacacctgggctgcccaa
ccgcgacgacgtgaaggcgggctcatcgccataagatcgccgcccacgcggccgacctggccaagcagca
ccccacgcccaggcgtgggacgacgcgctgtccaaggcgcgcttcgagttccgctggatggaccagttcgc
cctgtccttgaccccatgacggcagatgtccttccacgacgagacgctgcccgcggacggcgcggaaggtcgc
55 ggagaacggctacggctccgcccaggaggccatccgcccaggcatggacgcatgtccgaggagtccaacat
cgccaagaagacgatctccggcgagcagcagggcaggtcgccggcgagatctacctgcccaggtcctacgt
caaggccgcgcagaagTGACAattggcagcagcagctcgatagatcgacacactctggacgctggtcgtg
tgatggactgttccgcccacacttgctgccttgacctgtgaatatccctgcccgttttatcaaacagcctca
gtgtgtttgatctgtgtgtacgcgcttttgcagttgtgactgctgtgtgtatttgcgaataaccacccc
60 agcattcccttccctcgtttcatatcgcttgcatacccaaccgcaacttatctacgctctgtctatccctc
agcgtgctcctgctcctgctcactgcccctgcacagccttgggttgggctccgctgtattctcctggta
ctgcaacctgtaaacagcactgcaatgctgatgcacgggaagtagtgggatgggaacacaaatggaggatc
c

65

SEQ ID NO: 144

Codon-optimized sequence of PmKASII comprising a *C. protothecoides* S106 stearyl-ACP desaturase transit peptide

70

ATGgccaccgcatccactttctcggcgttcaatgccgcgtgcggcgacctgcgtcgctcggcgggctccggg
ccccggcgcccagcagggccctccccgtgcgcggggcgccgcccgcggccgacgccaacccccggccgc

cccagcgccgcgtggtgatcaccggccaggcggtgacccctgggcccagaccatcgagcagttctac
 tctctcctgctggagggcggtgctccggcatctcccagatccagaagttcgacaccaccggctacaccaccacc
 atcgccggcgagatcaagtccttgacgtgacccctacgtgcccgaagcgctgggccaagcgctggacgac
 gtgatcaagtacgtgtacatcgccggcaagcaggccctggagtcggccggcctggccatcgaggccgcccgc
 5 ctggccggcgccggcctggaccccgccctgtgcccgcgtgctgatcggcaccggccatggccggcatgacctcc
 ttgcgcgcggcggtggagggcctgacccgcccggcggtgcgcaagatgaaccccttctgcatcccccttctcc
 atctccaacatggcgccgcatgctggccatggacatcggttcctatgggcccccaactactccatctccacc
 gctgcccaccggcaactactgcatcctggcgccggccgaccacatccggcgggcgacgccaacgtgatg
 10 ctggccggcgccggcggacgcccgcctcatccctccggcatcgccggcttcctcgctgcaagggcctgtcc
 aagcgcaacgacgagcccagcgccgctcccggccctgggacgcccgacgcccgcgaggttcctgatgggcgag
 ggcgccggcggtgctggtgctggaggagctggagcagcccaagcgccgcccggcgccaccatcctggccgagctg
 gtggcgccggcgccgacccctccgacgcccaccacatgaccgagcccagccccaggcgccggcggtgcccctg
 tgcctggagcgcccgctggagcgcccgccctggcccccgagcgcggtgggctacgtgaacgccccagggcacc
 15 tccacccccggcgccgacgtggccgagtaccgcccctccgcccgtgatccccaggactcctgcccgcac
 aactccaccaagtccatgatcgccacactgctggcgccggcgccggcgccgtggaggcggtggccgcccacccag
 gccctgcccacccggtggtgcccaccccaacctgaacctggagaacccccggcggtggaccccggtggtg
 ctggtgggcccccgcaaggagcgccgaggacctggacgtggtgctgtccaactccttcggcttcggcgcc
 cacaactcctgctgatcttcgcaagtacgacgagatggactacaaggaccacgacggcgactacaaggac
 20 cagcatcgactacaaggacgacgacgacaagTGA

SEQ ID NO: 145

Codon-optimized amino acid sequence of PmKASII comprising a C.
 protothecoides S106 stearyl-ACP desaturase transit peptide

MATASTFSAFNARCGDLRRSAGSGPRRPARPLPVRGAAAAADANPARPERRVITGQGVTSLGQTIEQFY
 SSSLLEGVSGISQIQKFDTTGYTTTIAGEIKSLQLDPYVPKRWAKRVDDVIKYVYIAGKQALESAGLPIEAAG
 LAGAGLDPALCGVLIGTAMAGMTSFAAGVEALTRGGVRKMNPFCIPFSISNMGGAMLAMDIGFMGPNYSSIST
 ACATGNYCILGAADHIRRGDANVMLAGGADAAIIPSGIGGFIACKALSKRNDEPERASRPWDADRDGFVMGE
 GAGVLVLEELEHAKRRGATILAEVLGGAATSDAHMTEPDQPQGRGVRLCLERALERARLAPERVGVYVNAHGT
 25 STPAGDVAEYRAIRAVIPQDSLRINSTKSMIGHLLGGAGAVEAVAAIQALRTGWLHPNLLENPAPGVDPVV
 LVGPRKERAEDLDVVLNSNSFGFGGHNSCVIFRKYDEMIDYKDHDGDYKDHDIDYKDDDDK

SEQ ID NO: 146

Codon-optimized *Protetheca moriformis* (UTEX 1435) FAD2 protein-coding
 sequence

ATGccatcaagaccaaccgcccagcccgtggagaagcccccttcaccatcgccaccctgcccgaaggccatc
 cccgcccactgcttcgagcgctccgcccgtgcccctcctccatgtacctggccttcgacatcgccgtgatgtcc
 40 ctgctgtacgtggcctccacctacatcgacccccgccccgtgcccacctgggtgaagtacggcggtgatgtgg
 cccctgtactggttcttccagggcgcccttcggcaccggcggtgtgggtgtgcccacagagtggcgccaccag
 gccttctcctcctcccaggccatcaacgacggcggtgggcccgtggtgttccactcctgctgctggtgcccac
 tactcctggaagcactcccaccgcccaccactccaacaccggctgcctggacaaggacgaggtgttcgtg
 cccccccaccgcccgtggcccacgagggcctggagtgaggagtggtgcccacccgcatggcgatgggcaaggtg
 45 ctggtgacctgacctgggctggcccctgtacctgatgttcaacgtggcctcccggcccctacccccgcttc
 gccaacacttcgacccctggtcccccatcttctccaagcgcgagcgcatcgaggtggtgatctccgacctg
 gccctggtggcgtgctgtccggcctgtccgtgctggcgccaccatgggctgggctggtggtgaagacc
 tacgtggtgcccctacctgatcgtgaacatgtggctggtgctgatcaccctgctgcagcacacccccggcc
 ctgcccactacttcgagaaggactgggactggtgctgcgcggcgccatggccaccgtggacggctccatgggc
 ccccccttcatggacaacatcctgcaccacatctccgaccccacgtgctgcaccacctgttctccaccatc
 50 cccactaccacgcccaggagggcctccgcccacatccgccccatcctgggcaagtactaccagtcgactcc
 cgctgggtggcgccgcccctgtgggaggactggcgcgactgccgctacgtggtgcccgcgcccccgaggac
 gactccgcccgtggttccacaagTAG

SEQ ID NO: 147

Codon-optimized *Protetheca moriformis* (UTEX 1435) amino acid sequence of
Prototheca moriformis FAD2

MAIKTNRQPVEKPPFTIGTLRKAIPAHCFERSALRSSMYLAFDIAVMSLLYVASTYIDPAPVPTWVKYGVMW
 PLYWFFQGAFTGVWVCAHECGHQAFSSSQAINDGVLVFLSHLLVPYYSWKHSHRRHHSNTGCLDKDEVFV
 60 PPHRAVAHEGLEWEWLPIRMGKVLVTLTLGWPLYLMFNVASRPPYPRFANHFDWPSPIFSKRERIEVVISDL
 ALVAVLSGLSVLGRMTGWAWLVKTYVVPYLVNMLVLITLLQHTHPALPHYFEKDWDWLRGAMATVDRSMG
 PPFMDNILHHISDTHVLHHLFSTIPHYHAEESAIRPILGKYYQSDSRWVGRALWEDWRDCRYVVPDAPED
 DSALWFHK

SEQ ID NO: 148

Codon-optimized coding region of *Brassica napus* C18:0-preferring
 thioesterase from pSZ1358

ACTAGTATGCTGAAGCTGTCTTGCAACGTGACCAACAACCTGCACACCTTCTCCTTCTTCTCCGACTCCTCC
 CTGTTTCATCCCCGTGAACCGCCGCACCATCGCCGTGTCTCCGGGCGCGCCTCCAGCTGCGCAAGCCCCGCC
 70 CTGGACCCCCCTGCGCGCCGTGATCTCCGCCGACCAGGGCTCCATCTCCCCCGTGAACCTCTGCACCCCCGCC
 GACCGCCTGCGCGCCGCCGCTGATGGAGGACGGCTACTCCTACAAGGAGAAGTTCATCGTGCCTCCTAC

GAGGTGGGCATCAACAAGACCGCCACCGTGGAGACCATCGCCAACCTGCTGCAGGAGGTGGCCTGCAACCAC
 GTGCAGAAAGTGGCGCTTCTCCACCGACGGCTTCGCCACCACCCTGACCATGCGCAAGCTGCACCTGATCTGG
 GTGACCGCCCGCATGCACATCGAGATCTACAAGTACCCCGCCTGGTCCGACGTGGTGGAGATCGAGACCTGG
 TGCCAGTCCGAGGGCCGCATCGGCACCCGCCGCGACTGGATCCTGCGCGACTCCGCCACCAACGAGGTGATC
 5 GGCCGCGCCACCTCCAAGTGGGTGATGATGAACCAGGACACCCGCCGCTGCAGCGCGTGACCGACGAGGTG
 CGCGACGAGTACCTGGTGTTCGCCCCCGGAGCCCCGCTGGCCTTCCCCGAGGAGAAACAACTCCTCCCTG
 AAGAAGATCCCCAAGCTGGAGGACCCCGCCAGTACTCCATGCTGGAGCTGAAGCCCCGCCGCGCCGACCTG
 GACATGAACCAGCACGTGAACAACGTGACCTACATCGGCTGGGTGCTGGAGTCCATCCCCCAGGAGATCATC
 10 GACACCCACGAGCTGCAGGTGATCACCTGGACTACCGCCGCGAGTGCCAGCAGGACGACATCGTGGACTCC
 CTGACCACCTCCGAGATCCCCGACGACCCCATCTCAAAGTTCACCGGCACCAACGGCTCCGCCATGTCTCTCC
 ATCCAGGGCCACAACGAGTCCCAGTTCCTGCACATGCTGCGCCTGTCCGAGAACGGCCAGGAGATCAACCGC
 GGCCGACCCAGTGGCGCAAGAAGTCTCCCGCATGGATACAAGGACCACGACGGCGACTACAAGGACCAC
 GACATCGACTACAAGGACGACGACGACAAGTGAATCGAT

SEQ ID NO: 149

Amino acid sequence of Brassica napus C18:0-preferring thioesterase
 (Accession No. CAA52070.1)

MLKLSCNVNTNNLHTFSFFSDSSLFIPVNRRTIAVSSSQLRKPALDPLRAVISADQGSISPVNSCTPADRLRA
 20 GRIMEDGYSYKEKFIIVRSYEVGINKTATVETIANLLQEVACNHVQKCGFSTDGFATTLTMRKLHLIWTARM
 HIEIYKYPAWSVDVEIETWCQSEGRIGTRRDWILRDSATNEVIGRATSKWVMNQDTRRLQRTDEVREYIL
 VFCPREPRLAFPEENNSSLKKIPKLEDPAQYSMLELKPRADLDMNQHVNNVTYIGWVLESIPQEIIDTHEL
 QVITLDYRRECQDDIDVSLTTSEIPDDPISKFTGTNGSAMSSIQGHNESQFLHMLRLSENGQEINRGRTQW
 25 RKKSSR

SEQ ID NO: 150

Prototheca moriformis FATA1 allele 1 5' homology donor region

GGAGTCACTGTGCCACTGAGTTCGACTGGTAGCTGAATGGAGTCGCTGCTCCACTAAACGAATTGTCAGCAC
 CGCCAGCCGCGCCGAGGACCCGAGTCATAGCGAGGGTAGTAGCGCGCCATGGCACCGACAGCCTGCTTGCCA
 30 GTACTGGCGTCTCTTCCGCTTCTCTGTGGTCCTCTGCGCGCTCCAGCGCGTGCGCTTTTCCGGTGGATCATG
 CGGTCCGTGGCGCACCGCAGCGGCCGCTGCCCATGCAGCGCCGCTGCTTCCGAACAGTGGCGGTGAGGGCCG
 CACCCGCGGTAGCCGTCCGTCCGGAACCCGCCCAAGAGTTTTGGGAGCAGCTTGAGCCCTGCAAGATGGCGG
 AGGACAAGCGCATCTTCCCTGGAGGAGCACCGGTGCGTGGAGGTCCGGGGCTGACCGGCCGTGCGATTCAACG
 TAATCAATCGCATGATGATCAGAGGACACGAAGTCTTGGTGGCGGTGGCCAGAAACACTGTCCATTGCAAGG
 35 GCATAGGGATGCGTTCCTTCACCTCTCATTTCTCATTTCTGAATCCCTCCCTGCTCACTCTTCTCCTCCTC
 CTCCCCGTTACGCAGCATTCGG

SEQ ID NO: 151

Prototheca moriformis FATA1 allele 1 3' homology donor region

GACAGGGTGGTTGGCTGGATGGGGAAACGCTGGTGCGGGATTCGATCCTGCTGCTTATATCCTCCCTGGAA
 40 GCACACCCACGACTCTGAAGAAGAAAACGTGCACACACACAACCCAAACCGGCCGAATATTTGCTTCTCTTATC
 CCGGTTCCAAGAGAGACTGCGATGCCCCCTCAATCAGCATCCTCCTCCCTGCGCTCAATCTTCCCTGCT
 TGCCTGCGCCCGCGGTGCGCCGTCTGCCCGCCAGTCAGTCACTCCTGCACAGGCCCCCTTGTGCGCAGTGCT
 CCTGTACCCTTTACCGCTCCTTCCATTCTGCGAGGCCCCCTATTGAATGTATTTCGTTGCCTGTGTGGCCAAG
 45 CGGGCTGCTGGGCGCGCCGCGCTCGGGCAGTGCTCGGCGACTTTGGCGGAAGCCGATTGTTCTTCTGTAAGC
 CACGCGCTTGCTGCTTTGGGAAGAGAAGGGGGGGGTACTGAATGGATGAGGAGGAGAAGGAGGGGTATTGG
 TATTATCTGAGTTGGGT

SEQ ID NO: 152

Prototheca moriformis FATA1 allele 2 5' homology donor region

AATGGAGTTCGCTGCTCCACTAATCGAATTGTGAGCACCGCCAGCCGCGCCGAGGACCCGAGTCATAGCGAGGG
 TAGTAGCGCGCCATGGCACCGACGAGCCTGCTTGCCCGTACTGGCGTCTCTTCCGCTTCTCTGTGCTCCTCT
 55 ACGCGCTCCGGCGCGTGCGCTTTTCCGGTGGATCATGCGGTCCGTGGCGCACCGCAGCGGCCGCTGCCCATG
 CAGCGCCGCTGCTTCCGAACAGTGGCTGTGAGGGCCGCACCCGAGTAGCCGTCCGTCCGGAACCCGCCCAA
 GAGTTTTGGGAGCAGCTTGAGCCCTGCAAGATGGCGGAGGACAAGCGCATCTTCTGGAGGAGCACCGGTGC
 GCGGAGGTCCGGGGCTGACCGGCCGTGCGATTCAACGTAATCAATCGCATGATGATCACAGGACGCGACGTC
 TTGGTGGCGGTGGCCAGGGACACTGCCCATTGCACAGGCATAGGAATGCGTTTCTTCTCATTTCTCAGTTTT
 60 CTGAGCCCCCTCCCTCTTCACTCTTCTCCTCCTCCTCCCTCTCACGCAGCATTCGTGG

SEQ ID NO: 153

Prototheca moriformis FATA1 allele 2 3' homology donor region

CACTAGTATCGATTTTGAACAGAGGAGAGGGTGGCTGGTAGTTGCGGGATGGCTGGTGCGCCGTGATCCTG
 CTGCTGCTATTGTCTCCTCCTGCACAAGCCCACCCAGACTCCGAAGAAGAAGAAGAAAACGCGCACACACA
 CAACCCAAACCGCCGAATATTTGCTTCTTATCCCGGGTCCAAGAGAGACGGCGATGCCCCCTCAATCAGC
 65 CTCCTCCTCCCTGCCGCTCCAATCTTCCCTGCTTGATGCGCCGCGAGAGGCTGTCTGCGCGCCCCGTGAG
 TCACTCCCCGTGACAGCGCCTCGTGCTCGGTGCTCCTGTATCCTTTACCGCTCCTTTCTATTCTGCGAGGCCC
 70 CCTGTTGAATGTATTTCGTTGCCGTGTGTGGCCAAGCGCGCTGCTGGGCGCGCCGCGCTCGGGCGGTGCTCGGC

GACTCTGGCGGAAGCCGGTTGTTCTTCTGTAAGCCACGCGCTTGCTGCTTTTGGAAAAGAGGGGGGTTTACT
GAATGGAGGAGGAGCAGGATAATTGGTAGTATCTGAGTTGTTG

5 SEQ ID NO: 154
SAD2 hairpin C

actagtGCGCTGGACGCGGCAGTGGGTGGCCGAGGAGAACCGGCACGGCGACCTGCTGAACAAGTACTGTTG
GCTGACGGGGCGCGTCAACATGCGGGCCGTGGAGGTGACCATCAACAACCTGATCAAGAGCGGCATGAACCC
10 GCAGACGGACAACAACCCCTTACTTGGGCTTCGTCTACACCTCCTTCCAGGAGCGCGCGACCAAGTACAGCCA
CGGCAACACCGCGCGCCTTGCAGGCCGAGCAGTGTGTTTGGAGGTTTGGTTGCCCGTATCGAGGTCTGGTG
GCGCGCATGGGGGAGAAGGCGCCTGTCCCGCTGACCCCCCGGCTACCTCCCGGCACCTTCCAGGGCGCGT
ACGggatccTGCTCGGCCGCAAGGCGCGCGGTGTTGCCGTGGCTGTACTTGGTCGCGCGCTCCTGGAAGGAG
GTGTAGACGAAGCCCAAGTAAGGGTTGTTGTCCGTCTGCGGGTTCATGCCGCTCTTGATCAGGTGTTGATG
15 GTACCTCCACGGCCCGCATGTTGACGCGCCCGCTCAGCCAACAGTACTTGTTCAGCAGGTGCGCGTGCCGG
TTCTCCTCGGCCACCCACTGCCGCGTCCAGCGCaagctt

SEQ ID NO: 155

MSIQFALRAAYIKGTCQRLSGRGAALGLSRDWTGWTLPWCWPASAAATAPPRARHQERAIHLTSGRRRHSA
LASDADERALPSNAPGLVMASQANYFRVRLLEPEEEGELESWSNVNRHTTLLCKPRAML SKLQMRVMVGDRV
20 IVTAIDPVMNTVHAPPFDPLPATRFLVAGEAADMDITVVLNKADLVPEEESAALAEVASWGPVVLSTLTG
RGLQELERQLGSTTAVLAGPSGAGKSSIINALARAARERPSDASVSNVPEEQVVGEDGRALANPPFTLADI
RNAIPKDCFRKSAKSLAYLGLDSITGMAYLAYKINSPWLWPLYWFAQGTMFALFVVGHDGCGHQSFSSTKR
LNDALAWLGALAGTWTWALGVLPMNLNLYLAPYVWLLVTYLHHHGSPDPREEMPWYRGREWSYMRGGLTTID
RDYGLFNKVHHDIGTHVVHH

25 SEQ ID NO: 156

MFWALFVVGHDGCGHQSFSSTSKRLNDAVGLFVHSIIIGVPYHGWRISHRTHHNNHGHVENDESWYPPTESGLKA
MTDMGRQGRFHFPSMLFVYPFYLFWRSPGKTGSHFSPATDLFALWEAPLIRTSNACQLAWLGALAAGTWALG
30 VLPMLNLNLYLAPYVISVAWLDLVTYLHHHGSPDPREEMPWYRGREWSYMRGGLTTIDRDYGLFNKVHHDIGTH
VVHHLFPQIPHYNLCRATKAACKVLGPYYREPERCPLGLLPVHLLAPLLRSLGQDHFVDDAGSVLFYRRAEG
INPWIQKLLPWLGGARRGADAQRDAAQ

SEQ ID NO: 157

Camelina sativa omega-3 FAD7-2

35 MANLVLSECGIRPLPRIYTTPRSNFVSNNKPIFKFRPFTSYKTSSSPLACSRDGFSGKNWSLNVSVPLTTTT
PIVDESPLKEEEEEKQRFDPGAPPPFNLAIRAAIPKHCWVKNPWKSMYSYVLRDVAIVFALAAGASYLNNWI
VWPLYWLAQGTMFALFVLGHDCGHGSFSNNPRLNNVVGHLHSSILVPYHGWRISHRTHHQNHHGHVENDES
40 WHPMSEKIYQSLDKPTRFFRFTLPLVMLAYPFYLVARSFGKKGSHYHPESDLFLPKEKTDVLTSTACWTAMA
ALLICLNFFVGPVQMLKLYGIPYWINVMWLDFTYLYLHHGHGDKLPWYRGKEWSYLRGGLTTLDYGVINN
IHHDIGTHVIHHLFPQIPHYHLVEATEAVKPVLGKYYREPKSGPLPLHLLGILAKSIKEDHYVSDGDDVY
YKADPNMYGEIKVGAD

SEQ ID NO: 158

45 Prototheca moriformis delta 12 desaturase allele 2

50 MAIKTNRQPVEKPPFTIGTLRKAIPAHCFERSALRSSMYLAFDIAVMSLLYVASTYIDPAPVPTWVKYIGIMV
PLYWFFQAGFTGVWVCAHECGHQAFSSSQAINDGVLVHSLLLVPYYSWKHSHRRHHSNTGCLDKDEVFV
PPHRAVAHEGLEWEEWLPIRMGKVLVTLTLGWPLYLMFNVASRPPYPRFANHFDPSPIFSKRERIEVVISDL
ALVAVLSGLSVLGRMTGWAWLVKTYVVPYIMVNMWLVLTLLQHTHPALPHYFEKDWDWLRGAMATVDRSMG
PPFMDSILHHISDTHVLHHLFSTIPHYHAEESA AIRPILGKYYQSDSRWVGRALWEDWRDCRYVVPDAPED
DSALWFFHK

SEQ ID NO: 159

Camelina sativa omega-3 FAD7-1

55 MANLVLSECGIRPLPRIYTTPRSNFVSNNKPIFKFRPLTSYKTSSPLFCSRDGFGRNWSLNVSVPLATTTTP
IVDESPLKEEEEEKQRFDPGAPPPFNLAIRAAIPKHCWVKNPWKSMYSYVLRDVAIVFALAAGAAAYLNNWI
VWPLYWLAQGTMFALFVLGHDCGHGSFSNNPRLNNVVGHLHSSILVPYHGWRISHRTHHQNHHGHVENDES
40 WHPMSEKIYQSLDKPTRFFRFTLPLVMLAYPFYLVARSFGKKGSHYHPESDLFLPKEKTDVLTSTACWTAMA
ALLICLNFFVGPVQMLKLYGIPYWINVMWLDFTYLYLHHGHGDKLPWYRGKEWSYLRGGLTTLDYGVINN
IHHDIGTHVIHHLFPQIPHYHLVEATEAVKPVLGKYYREPKSGPLPLHLLGILAKSIKEDHYVSDGDDVY
YKADPNMYGEIKVGAD

SEQ ID NO: 160
PmFATA-hpB

5 actagtCATTCGGGGCAACGAGGTGGGCCCCCTCGCAGCGGCTGACGATCACGGCGGTGGCCAACATCCTGCA
GGAGGCGGCGGGCAACCACGCGGTGGCCATGTGGGGCCGAGCGTGTGTTTGAGGGTTTTGGTTGCCCGTAT
TGAGGTCTTGGTGGCGCGCATGGGGGAGAAGGCGCCTGTCCCGCTGACCCCCCGGCTACCTCCCGGCACC
TTCCAGGGCGCGTACGg gatccGCTCCGGGCCCCACATGGCCACCGGTGGTTGCCCGCCGCTCCTGCAGGA
TGTTGGCCACCGCCGTGATCGTCAGCCGCTGCGAGGGGCCACCTCGTTGCCCCGAATGaagctt

10 SEQ ID NO: 161
PmFATA-hpC

actagtGGAGGTTTTCGCGACGGACCCGGAGCTGCAGGAGGCGGGTCTCATCTTTGTGATGACGCGCATGCA
GATCCAGATGTACCGCTACCCGCGCTGGGGCGACCTGATGCAGGTGGAGACCTGGTTCCAGAGTGTGTTTGA
GGGTTTTTGGTTGCCCGTATTGAGGTCTTGGTGGCGCGCATGGGGGAGAAGGCGCCTGTCCCGCTGACCCCC
15 CGGCTACCTCCCGGCACCTTCCAGGGCGCGTACGg gatccTCTGGAACCAGGTCTCCACCTGCATCAGGTC
GCCCCAGCGCGGGTAGCGGTACATCTGGATCTGCATGCGCGTCATCACAAAGATGAGACCCGCTCCTGCAG
CTCCGGGTCCGTGCGGAAACCTTCCaagctt

20 SEQ ID NO: 162
PmFATA-hpD

actagtCGGCGGGCAAGCTGGGCGCGCAGCGCGAGTGGGTGCTGCGCGACAAGCTGACCGGCGAGGCGCTGG
GCGCGGCCACCTCGAGCTGGGTTCATGATCAACATCCGCACGCGCCGGCCGTGCCGCATGCCGGGTGTGTTTG
AGGGTTTTTGGTTGCCCGTATCGAGGTCTTGGTGGCGCGCATGGGGGAGAAGGCGCCTGTCCCGCTGACCCCC
25 CCGGCTACCTCCCGGCACCTTCCAGGGCGCGTACGg gatccCCGGCATGCGGCACGGCCGGCGCGTGCAGG
TGTTGATCATGACCCAGCTCGAGGTGGCGCGCGCCAGCGCCTCGCCGGTCAGCTTGTCGCGCAGCACCCACT
CGCGCTGCGCGCCAGCTTGCCCCGCCGaagctt

30 SEQ ID NO: 163
PmFATA-hpE

actagtGTCCGCGTCAAGTCGGCCTTCTTCGCGCGCGAGCCGCGCGCTGGCGCTGCCGCCCCGGGTACAG
CGTGCCAAGCTGCCAACATCGCGACGCCGGCGCGCTGCGCGGGCACCGCCAGGTGCGCGCGCCGACCCGAC
ATGGACATGAACGGGCACGTGAACAACGTGGCCTACCTGGCCTGGTGCCTGGAGTGTGTTTGAGGGTTTTGG
35 TTGCCCGTATTGAGGTCTTGGTGGCGCGCATGGGGGAGAAGGCGCCTGTCCCGCTGACCCCCCGGCTACCC
TCCCGGCACCTTCCAGGGCGCGTACGg gatccTCCAGGCACAGGCCAGGTAGGCCACGTTGTTTACGTGCC
CGTTCATGTCCATGTGCGTGCGGCGCGCGACCTGGCGGTGCCCGCGCAGCGGCGCCGGCGCTCGCGATGTTGG
GCAGCTTGGCACGCGTGACCGCGGGCGGCAGCGCCAGGCGCGGCGGCTCGCGCGCAAGAAGGCCGACTTGA
CGCGGACAagctt

40 SEQ ID NO: 164
PmFATA-hpF

actagtCCGTGCCCCGAGCACGTCTTCAGCGACTACCACCTCTACCAGATGGAGATCGACTTCAAGGCCGAGT
45 GCCACGCGGGGCGACGTATCTCTCCAGGCCGAGCAGATCCCGCCCCAGGAGGCGCTCACGCACAACGGCG
CCGGCCGCAACCCCTCTGCTTCGTCCATAGCATTTCTGCGCGCCGAGACCGAGCGTGTGTTTGAGGGTTTTG
GTTGCCCGTATCGAGGTCTTGGTGGCGCGCATGGGGGAGAAGGCGCCTGTCCCGCTGACCCCCCGGCTACC
CTCCCGGCACCTTCCAGGGCGCGTACGg gatccGCTCGGTCTCGGCGCGCAGAATGCTATGGACGAAGCAGG
AGGGGTTGCGGCGCGCGCGCTTGTGCGTGAGCGCCTCCTGGGGCGGGATCTGCTCGGCCTGGGAGGAGATGA
50 CGTCGCCCCGCTGGCACTCGGCCTTGAAGTCGATCTCCATCTGGTAGAGGTGGTAGTCGCTGAAGACGTGCT
CGGGCACGGAagctt

55 SEQ ID NO: 165
PmFATA-hpG

actagttTCGTCCGCGCGGAACCACATGGTTCGGCCCCCATCGACGCGCCCGCCGCAAGCCGCCCAAGGCGA
GCCACTGAGGACAGGGTGGTTGGCTGGATGGGGAACGCTGGTTCGCGGGATTGATCCTGCTGCTTATATCC
60 TCGTGTGTTTGAGGGTTTTGGTTGCCCGTATTGAGGTCTTGGTGGCGCGCATGGGGGAGAAGGCGCCTGTCC
CGTGACCCCCCGGCTACCTCCCGGCACCTTCCAGGGCGCGTACGg gatccGAGGATATAAGCAGCAGGA
TCGAATCCCGCGACACGCGTTTCCCCATCCAGCCAACCACCTGTCTCAGTGGCTCGCCTTGGGCGGCTTG
GCGGCGGGCGCGTTCGATGGGGGCCGACCATGTGGTTTCGCGCGCGGACGAaagctt

65 SEQ ID NO: 166

Prototheca moriformis (UTEX 1435) glyoxysomal fatty acid beta-oxidation
multifunctional protein

70 ATGTGCAAGGAGCCGAGTCTTTCTACAAGGTGGAGGATGGAGTGGCCATCATCACCATCTCTACCTCCC
ATGAACGCCCTTACCCCCACGCTGATCAAGGACATGTTCAACAACGTCCGTGCGCGCCAGGAGGATGCCAGC
GCCAAGGCCATCGTCATCACCAGTGGCAACGGTTCGCTTCTGCGCGCGGCTTTGACATCAACCAGTTCCAGAAC

AAGAGCGCCGGCGCSGGCGGGCGGCATCGACATGAGCATCAACAACGCGTTTGTSGAGCTCATCGAGAACGGS
 CGCAAGCCGACGGTGGCCGCGATCGAGCGGCTGGCCCTGGGCGGCGGGTGCAGCTGGCGCTGGCCTGCAAC
 GCGCGCTAGCGGCGCCCGGCACCATCATGGGCCTGCCCCGAGCTGACGCTGGGCATTTTGGCCGGCTTTGGC
 5 GGCACGCAGCGCCTGCCGCGCGTGGTGGGCCTTGAGAAGGGCCTGTCCATGATCCTGACCAGCAAGCCCATC
 AAGGARCCCCGAGGCGCTCAAGCTGGGCCTGCTGGACGAGGTGGTGGCCGCGCCGAGCTGCTGGCGCGCGCC
 AAGGCCCTTTGCGCTGGAGATCGTGMCSGGGCGAGCGCCGCGCGTCAACGCGCTCCAGCGCACGGCAAGCTG
 CCGCGGTACGCGGACGCGCTGCAGATCATCCAGTTTGGCGCGAGCAGACGGCCAAGCGCGCGCCCAACCTG
 CAGCACCCGCTCTTCACGCTGGACGCGATCCAGGCCGGCATCGAGCACGGCGGGCTCAAGGGCCTGCGCGCC
 10 GAGGCCGACGCCTTTGCGCGCTCGGCGGCGCTGCCACGCACGACTCGCTGGTGCACGTCTTCTTCGCGCAG
 CGCTCGACCAAGCGCGTGCAGCGAGTGACCGACGTGCGGGCTCAAGCCACGTGCGGTCAAGAGGATCGCCGTG
 CTGGGCGGCGGGCTCATGGGCTCGGGCATCGCCACCGCCTCGGCGCTGGCGGGCGTSGAGGTGCTGCTCAAG
 GAYATCAARSAYAAAYTTCCSTGAGGGCGGCGCTGGGCGCATCCGCGCCAACCTGGCGTCCAAGGTCAAGCGG
 AAGCAGCTCTCGCAGCAGGCGGCGGACGCCGCGATGGCCGCGTCAAGGGCGTGTCTACCTACGACGACTTC
 15 AAGACSGTCGACATGGTGATCGAGGCGGCCATCGAGAAGGTGGACCTGAAGCAGGACATCTTCGTGGACCTG
 GTCAGGGCCACGCGCCCCGACTGCATCCTGGCCACCAACACGTGACCATCAACATCGAGACGGTGGGCGCC
 AAGATCCCGGGCGACCTGGGCGCGGTGATCGGCGCGCACTTTTTCTCGCCRGCGCACATCATGCCCCCTGCTG
 GAGATCGTGCCTCGGACAAGACCAGCGCGCAGGTATCCTGGACACGCTCGAGTACGGCAGCAAGATCAAG
 AAGACGCGGCTCGTGGTGGGCACTGCACGGGCTTTGCGGTCAACCGCTCTTCTTCCCTACTCCATGGCC
 GCGACCATTTCTTGGACTCGGGGCTCGACCCCTACGCCATCGACGCCGCGCATCACGGCCTTTGGCATGCCC
 20 ATGGGGCCCTTCCGCTGGGCGACCTGGTGGCCTGGACGTCTCGCTCTTCGTGGGCGCCTCCTACCTGCAG
 GACTACGGCGACCGCGTCTACCGCGGCGCCATGGTCCCGCTGCTCAACGAGGCGCGCCGRCTGGGCGAGAAG
 ACCGGCGCGGCTGGTACAAGTTTGACGACCGCGCGCGCGAGCCCGGACCCGGCGCTGGCGCCGCTCCTC
 TCGCCCAAGGACATCGCCGAGTTCATCTTCTCCCGTCTGTCACGAGGCCTGCCGCGTCTGGCCGAGGGC
 ATCGTGGACACGCCCTCCGACCTGGACCTGGACCGGTATGTTCCATGGGCTTCCCGGCTTACCGCGCGGCG
 25 ATCCTCTTCTACGGCGACATCGTGGCGCCAAAGTACGTGCTCGACCGTCTCAACGCGCTGGGCCAAGCAGTAC
 CCGCGCAGGCGGCTTCTTCAAGCCCTGCGACTACCTCCTSCGCGCGCGCAGACCGGCACCAAGCTCGAG
 GCCGCAACAAGCCCGCCTCCAAGATCTGA

SEQ ID NO: 167

Prototheca moriformis (UTEX 1435) glyoxysomal fatty acid beta-oxidation multifunctional protein

MSKEPQSFKVEDGVAIITISYPNMALHPTLIKDMFNNVRRQAEDASAKAIVITGANGRFCAFDINQFQN
 KSAGXGGGIDMSINNAFXELIENXRKPTVAAIERLALGGGCELALACNARVAAPGTIMGLPELTLGILPGFG
 35 GTQRLPRVVGLEKGLSMILTSKPIKXPEALKLGLLDEVVPAPQLLARAKAFALEIVXGQRPRVNALQRTDKL
 PAYADALQIIQFAREQTAKRAPNLQHPLFTLDIAIQAGIEHGGKGLRAEADAFARSAALPTHDSLHVHFFAQ
 RSTKRVRGVTVDVGLKPRAVKRIAVLGGGLMGSGIATASALAGXEVLLKXIXXXFXEGGLGRIRANLASKVKR
 KQLSQQAADAAMARVKGVLTYDDFKXVDMVIEAAIEKVDLQDIFVDLVRATRPDCILATNTSTINIIETVGA
 40 KIPGDLGRVIGAHFFSXAHIMPLLEIVRSDKTSAQVILDTLEYGSKIKKTPVVVGNGCTGFAVNRVFFPYMA
 ATILLDSGLDPYIDAAITAFGMPMPGFRGLDVLGLDVSFLVFGASYLQDYGDRVYRGAMVPLLNEAGXLGEK
 TGAGWYKFDDRRRASPDPALAPLLSPKDIAEFIFFPVNEACRVVAEGIVDKPSDLDIATVMSMGFPAYRGG
 ILFYGDIVGAKYVVDRLNAWAKQYPAQAGFFKPCDYLXRAAQTGKLEAGNKPASKI

SEQ ID NO: 168

Prototheca moriformis (UTEX 1435) Triacylglycerol Lipase (TAGL)

ATGGACAGGGCGTGGATGCTGGACCGCGAGCTGGCGGCCAACCGCAGCCGGGGCCTTGACTGGGGCGTGGCG
 CGGGCGCTGTGCGCCTACGTCTCGGCCTCGTACTGCAACAGCACGAGCCTGGCCGCGTGGAACTGCACGCGC
 50 TGCTCGCCGCGCGGCATCGACGGCAACGCCACGTGGATRGACCCCAACTTTGCGCTGGAGGCGCTGGCCTGG
 GACGAGGCCTGGGACCTGCTGGGCTACGTGGGCTGGAGCGAGGAGATGGGCGCGACCGTGATCGCCTTCCGC
 GGCACCGACTCGCACAGCTACCACAACCTGGGTGGCCAACATGCGCACCTGGCGCACCGACCTCAACCTGACC
 ATGCACGGCGCGCCGGCCAACGCGCTCGTGCACGGCGGCTTCTGGTTCTCCTGGAACGCCAGCTCCCTGGCC
 ACCTCCGTGACCGCGGCGCTGCTGCGCTGACGCGCGGCACGGCGCGCACCCGGTCTACGTGTGCGGGCAC
 55 TCGTGGGCGGCGCGCTGGCCACCATCTGCGCGCTGGAGCTGCGCACCATGCGCCYCKTGCGCGACGTGCAC
 CTGGTCACCTTTGGCAGCCCGCGCGTGGGCAACGCCGTCTTCGCGCGGCTGGTTGAGCGGGCGCATCACCTCG
 CACTGGCGCTTACGCACAACCGCGARATCGTGCCAGCGTGCCGCGCGCCCTACATGGGCTTCTGGCACCTG
 GCGCGCGAGGTCTGGGTGCTGGACAACGCGCGCCTCAACCCGCTCGTGGCGTCTGCGACGGGACGGGGGAG
 GACATGCAGTGCCACAACCTCCATGTGCCACCTGGGCCTCTGCTCCTCCATCGCAGACCACCTCCTCTACATC
 60 TCKGAGATGTACACCCCGAGGCCCATGGGCTGCTGA

SEQ ID NO: 169

Prototheca moriformis (UTEX 1435) Triacylglycerol Lipase (TAGL)

MDRAWMLDRELAANRSRGLDWGVARALSAYVSASYCNSTSLAANCTRCSPRGIDGNATWXDPNFALEALAW
 DEAWDLLGYVGWSEEMGATVIAFRGTDSSHYNWVANMRTWRDLNLTMHGAPANALVHGGFWFSWNASSLA
 65 TSVTAAVLRLQRRHGAHPVYVSGHSLGGALATICALLELRMTXXXRDVHLVTFGSPRVGNVAFAGWFERRITS
 HWRFTHNRIVPSVPPPYMGFWHLAREVWVLDNARLNPLVGVCDGTGEDMQCHNSMCHLGLCSSIADHLLYI
 XEMYTPRPMGC

SEQ ID NO: 170

Prototheca moriformis (UTEX 1435) Fumarate hydratase

ATGAGTACCCGCACGGAGCGCGACACGTTTGGGCCCCCTGGAGGTTCCCTCCGACAGGTAAGTGGGGTGCCAG
 ACGCAGCGTTCCCTGCAAACTTCAAGATCGGCGGCGCCCCGAGCGCATGCCGGAGCCGGTGGTGCGTGCC
 TTTGGCGTGCTCAAGGGCGCGGCGGCCAAGGTGAACATGGACCTGGGCGTGCTGGACAAGACCAAGGGCGAG
 5 GCGATCGTGGCGGCGGCCAAGGAGGTGGCGGAGGGCAAGCTGACGGACCACTTCCCCCTGGTGGTGTGGCAG
 ACGGGGTCCTGGCAGCAGAGCAACATGAACGCCAACGAGGTGATTGCCAACCGCGCGATCGAGATGCTGGGC
 GCGAGCTGGGCGACAAGAGCGTGGTGCACCCCCAACGACCACTGCAACAAGGGCCAGAGCTCCAACGACACC
 TCCCCACGGTGATGCACATCGCGGGCGTGACGGAGATCAGCCGCCGCTGGTGGCGGGTTGCGCCACCTG
 GCGGAGGCGCTGCGCGCCAAGGAGAAGGAGTTTGGCGGCATCATCAAGATCGGGCGCACGCACACGCAGGAC
 10 GCGACGCCGCTGACGCTGGGCCAGGAGTTTACGCGGCTACGCGACGCAGGTGAGTACGGCATCGAGCGCGTG
 CAGGCCGCGCTGCCGCGGCTGAGCATGCTGGCGCAGGGCGGCACCGCGGTGGGCACGGGGCTGAACGCCAAG
 GTCGGCTTTGCGGAGAAGGTGGCCGAGGCGGTGGCCGCCGACACGGGCCTGCCCTTACCACCGCGCCCCAAC
 AAGTTTGAGGCGCTGGCCGCGCACGACGCGGTGGTGGAGGCCAGCGGCGCGCTCAACACCGTGGCTGTGAGC
 CTGATGAAGTTCGCCAACGACGTGCGCTTCTGGGCGAGCGGCCGCGCTGCGGCCTGGGCGAGCTGAGCCTG
 15 CCGGAGAACGAGCCGGGCTCCTCCATCATGCCGGGCAAGGTCAACCCACGCAGTGTGAGGCGCTGACCATG
 GTCTGCGCGCAGGTGATGGGCAACCACGTGGCCGTGACGCTGGGCGGCTCCAACGGCCACTTTGAGCTCAAC
 GTGTTCAAGCCCATGATGATCCGCAACCTGCTGCACTCGGCGCGCCTGCTCGCCGACGCCGCCACCTCCTTC
 ACCGACAACCTGCGTCTGGGCATCAGGGCCAACGAGGGCCGCATCGCGCAGCTCCTGCACGAGTCCCTCATG
 CTCGTACCGCGCTCAACAACCATCGGCTACGACAAGGCCGCCGCTCGCCAAGAAGGCGCACAAAGGAC
 20 GGCAGCAGCTCAAGCAGGCCGCGCTCGAGCTCGGCTACTGCACCGAGCAGGAGTTTGCCGACTGGGTCAAG
 CCCGAGGACATGCTCGCGCCCACGCCCTAG

SEQ ID NO: 171

Prototheca moriformis (UTEX 1435) Fumarate hydratase

MSTRTERDTFGPLEVPSDRYWGAQTQRSLQNFKIGGAPERMPPEPVVRAFGVLKGAAAKVNMDLGVLDKTKGE
 AIVAAAKEVAEGKLTDFPLVWQTGSGTQSNMNEVIANRAIEMLGELGDKSVVHPNDHCNKGQSSNDT
 FPTVMHIAGVTEISRRLVPLRLHLAEALRAKEKEFAGIIKIGRTHQTQDATPLTLGQEFSGYATQVEYGIERV
 QAALPRLSMLAQGGTAVGTGLNAKVGFAEKVAEVAADTGLPFTTAPNKFEEALAHDAVVEASGALNTVAVS
 25 LMKVANDVRFGLSGPRCGLGELSPLPENEPGSSIMPGKVNPTQCEALTMVCAQVMGNHVAHSVGGSNHGFELN
 30 VFKPMIIRNLLHSARLLADAATSFTDNCVVGIRANEGRIAQLLHESLMLVTALNNHIGYDKAAAIKAKHKD
 GSTLKQAALLEGYCTEQEFADWVKPEDMLAPT

SEQ ID NO: 172

Prototheca moriformis (UTEX 1435) succinate semialdehyde dehydrogenase

ATGTCGACCRGTGTCCGAATCCAAGGAGATCCARGTCCGGGACGATGTCCTGAAGCGGCTCAARGACCCATCC
 CTCRTGCACACCGAGTTCGTACATCGGGGGCAGCTGGGTGAACGCTGCCGATGACGACAGGGTGGAGGTGCAC
 GACCCGGCCAGTGGCGAGGTGCTCGCCCGCTGACTCATGCCAAGGCTGTGGAGACCAAGGCGGCCATCGCS
 40 GAGGCGTCTCTGTGTTTGGCATGTGGTCCGCGAAGTCMGGCCWGGAGCGGTCCAACATCCTRCGCCGCTGG
 TTTGAGCTGCTGCGTGAGAATCAGGACGACATRGCGACGCTCATGACCCTGGAGAGCGGCAAGCCCCCTGGSC
 CAGGCCAAGGCCGAGATCGCGAGCGGCTGGGCTCTGTGGAGTGGTTTGGCGAGGAGGCCAAGCGCGTGGAC
 GGCGACATTCTGTGCGAGCCCTTCCCCAACAAAGCGTACTTGGTTCATGCGGCAGCCCATCGGCGTGGTGGG
 GCCATCACGCCCTGGAACCTTCCCTTTTCCATGATCACGCGCAAGATCTCGCCCGCCCTGGCGGCCGCTGC
 45 ACGGTGGTGTGAAGCCGAGCGAGCTGACGCCGCTGACGGCGCTGGCGATGGCCGAGCTGGGGGAGCGGGCC
 GGCATCCCGCTGGGCGTGCTCAACGTGCTGTTGGGCGACGCCAAGTCCATCGGCGACGAGCTCATCAAGAGC
 GACGAGGTGCGCAAGGTGCGCTTACCCGGCTCCACCCGCATCGGCAAGATCCTCATGGCCGGCGCGGCCAAC
 ACGGTCAAGAAGGTCTCCATGGAGCTGGGCGGGAACGCGCCCTACATCGTCTTCCAGACGCCGACCTCAAG
 GCGCGCGCTCCAGGCGGCGGCTCTCCACCGCAACGCGGCGAGACGTGCATCTGCACCAACCGGTG
 50 CTTGTCCACGAAAGCGTGCACGACGAGTTTGTCAAGGAGCTGGTGGCAGCGGTGCAGGAATTACAGGTGGGT
 CACGGCACCCGACCCAGCACACGCACGGCCCGCTCATCCAGCCGGGGGCCGTCGACAAGGTGGAGGAGAAG
 GTGCAGGACGCGGTGGCCAAGGGCGGACGCTGACAGCCGGGGGCAAGCGCCCCACGTTACGCGAGAACAGC
 CCCCTCAACAACGGCTTCTTCTTCGAGCCACCGTCTATCTCCAATGCGACCATCGACATGAAGGTGTTTACAG
 GAGGAGATCTTCGGGCCCCGTGACCCCGCTGTTCAAGTTTCTCGGACGAGGAGGCCGTCAAGTGGCGAAC
 55 AACASQAASSHRNAGQTCICTNRVLVHESVHDEFVKELVAQVEFRLGHGTDPTSTHGPLIQPGAVDKVEEK
 CTCGAGTTCCGCGATGGTCCGCGTGAACGACGTGCTCATCACCAGCTACGTCTCGCCCTTTGGCGGCGTCAAG
 CAGTCGGGCCCTSGGCCGCGAGGAGTCCAAGTACGGCATCGARGAGTACCTGCAGATGAAGTCGGTGTGCATG
 AACCTGGGGTACTGA

SEQ ID NO: 173

Prototheca moriformis (UTEX 1435) succinate semialdehyde dehydrogenase

MSTXSESKEIXVRDDVLKRLXDPSSLXHTESYIGGSWVNAADDDRVEVHDPASGEVLARVTHAKAVETKAAIX
 EASSVFGMWSXKXGXERSNIXRRWFELLRENQDDXATLMTLESGLKPLXQAKAEIASGLGSVEWFEEAKRVD
 GDILCSPFPNKRYLVMRQPIGVVGAITPWNFPFSMITRKISPALAAGCTVVLKPSSELTPLTALAMAELGERA
 65 GIPLGVNLVVGDASKIGDELIKSDEVKVAFTGSTRIGKILMAGAANTVKKVSMELGGNAPYIVFPDADLK
 AAASQAASSHRNAGQTCICTNRVLVHESVHDEFVKELVAQVEFRLGHGTDPTSTHGPLIQPGAVDKVEEK
 VQDAVAKGATVQTGGKRPTFSENSPLNNGFFFEPTVISNATIDMKVFREEIFGPVTPVFKFSSDEEAVKLAN
 NTEYGLAAYLFTKDLARAWKVSEALEFGMVGVDVLTITSYVSPFGGVKQSGXGREESKYGIXEYLMKSVCM
 70 NLGY

SEQ ID NO: 174

Prototheca moriformis (UTEX 1435) Malonyl coenzyme A : acyl carrier protein transacylase

ATGGCCGCTACGCTCAGCATGAAGTCCGTGGCGCTGCAGCGTCCCACCAGGTGTGGGCCTGCACCCCGCACA
GTGGTCCGGGCGCCGCCATGCGCGTCCGGGCTGCAGCCACCGCGCCCGGAGGCGCCCGCGGACCCAC
GCGTTTGTACGACTACAGCCCCGACAGGCCATCCTCTTCCCTGGGCAGGGCGCGCAGAGCGTGGGCTGGCG
GGCGAGCTRGTCAGTCCGTGGTGGCCCAAGGCGGCSACATGTTTGACGAGGCCTCYGCCATCCTKGGCTACGAC
CTGCTCAAGACSTGCGTGGAGGGCCCCAAGGAGCGCCTGGACAGCACGGTGGTGGAGCCAGCCGGCCATCTAC
GTGACGAGCCTGGCGGCCGTCGAGAAGCTGCGCGCCGACGAGGGCGACGAGGCGGTGGCGGCCGTCGACGTG
GCGGCGGGCCTGTCCCTGGGCGAGTACACGGCCCTGGCCTTTGCCGGCGCCTTCAGCTTCGAGGACGGGCTG
CGACTGGTTCGCGCTGCGCGGCCGAGCATGCAGCGCGCGGCCGACGCGGCCGCCAGCGGCATGGTCTCCGTC
ATCGGCCCTCTCGGCCCGCGGACCGAGGCGCTGTGCGAGGCGGCCAACGCCGAGGTCCCGCCGAGCAGGCC
GTGCGCGTGGCCAACTACCTCTGCAACGGCAACTACGCCGTGAGCGGCGGCCTYGAGGGCTGCGCGGCCGTC
GAGCGCCTRGCCAAGGCGCACAAARGCGCGCATGACCGTGCGCCTGGCCGTGGCCGGCGCCTTCCACACGGCC
TACATGCAGCCGGCCGTTGGAGGCGCTGACCGCGGCCCTGGCCAGCACGGACATCGCCGRCGCGCATCCCC
GTCGTCTCCAACGTCGACGCGCGCCGACGCGGACCCGGACACCATCCGCGCCCTGCTGGCCCCGCCAGGTC
ACCGCGCCCGTGCAGTGGGAGACCACGCTCAACACCCTGCTCTCCCGCGGCCTCCAGCGCTCCTACGAGGTC
GGCCCGGCAAGGTCATCGCCGGCATCATCAAGCGCGTCAACAAGAAGCACCCGGTGACCAACATCACTGCC
TGA

SEQ ID NO: 175

Prototheca moriformis (UTEX 1435) malonyl coenzyme A : acyl carrier protein transacylase

MAATLSMKSVLRPRTRCGPAPRTVVVRAPAMRVRAAATAAPEAPADHAFDDYQPRTAILFPGQGAQSVGMA
GEXVKSVPKAXHMFDEAXAIXGYDLLKXCEVEGPKERLDSTVVSQPAIYVTSLLAAVEKLRADEGDEAVAAVDV
AAGLSLGEYALAFAGAFSFEGLRLVALRGASMQRADAAASGMVSVIGLSAAATEALCEANAEPVPEQA
VRVANYLCNGNYAVSGGXEGCAEVERXAKAHXARMTVRLAVAGAFHTAYMQPAVEALTAALASTDIAXPRIP
VVSNDXAPHADPDTIRALLARQVTAPVQWETTLNLTLLSRGLQRSYEVGPGKVIAGIIKRVNKKHPVTNITA

SEQ ID NO: 176

Prototheca moriformis (UTEX 1435) phosphatidate cytidyltransferase

ATGCAGTTGCAGCGAGCGCTCACGGTGCGCCTGCGGAGGAGGGGCTGGCGTCGCAACTCGTTCAACAATGTG
GACGCCGTCTCGCCCAAGCTTGTGCGCAAGWCATCCAATGCCGGGCAGCAGACACAGTCGGAGGACAGGTTT
AGGTCCCTTCAAGGTCAGGACCCTGTCCACAATCGTCTCATCGCCACCTTCATCGGCATCATCGCCACGGGG
CACGTGCCCTCATGCTCATGATCCTGGGCATCCAGTTTCTGATGGTTCMGGGAGCTSTTTGCGCTGGCGAGG
GTSGCGCCCCAGGAGCGCAAGGTCCCCGGCTTCCGGGCGCAGCAGTGGTACTTTTTCTTCGTGGCCGCTTTT
TACCTGTACATCCGCTTCATCAAGACCAACCTCACGGTGGAGCTCTCCTCCTCGGCGCAAACGGCGGCCATG
TTTGGCTGGATCATCCGCCACCACACGCTCCTCTCCTTCGCCCTCTACACGGCCGGGTTTCGTGTGTTTCGTG
CTGCGCCTGAARAAGGACTCTACAGYTACAGTTCAGCAGTACGCCTGGACGCACATGATCCTCATGGTC
ATCTTCCTGCCTTCGTCTTCTTTGTGTACAACTCTTCGAGGGGCTGATCTGGTTCTCTGCTCCCGCGGCC
TTGGTCTGTGTTCAACGACATYCGGCCTACCTGGCAGGCTTTTTCTTTGGGAGGACCCCKCTGATCAAGCTG
TCTCCCAAGAAGACGTGGGAGGGCTTYATCGGCGGCCTGGCGGGCACGGTGGTGGTGGGCTGGTATCTGGCC
AAGTTGCTGTCTCAATTCAACTGGTTCATCTGTCTCCTCGGCGGGACCTGTCCATCATCCAGCCTCTGCACTGT
GACAGCAAGCTGGACATCTACCAGCCAGAGACATTTACCTGACAGATGTCTTGACGCTCTTGCCGGCCGAC
CTGGCCACCGCGGTGGTCACGGGGCTCACCCGCCTCTCGCCCGGCTGGCAGGAAGCTGCGCGGCAGGTGTCC
CTGACCTGCCCTGCCATGCAGCTGCACGCGGTCTGTGCGGACGTTTGCAGAGCTGATCGCACCTTTGGC
GGCTTCTTCGCCTCGGGCTTCAAGCGCGGCTTCAAGATCAAGGACTTTGGCGACAGCATCCCCGGCCACGGC
GGCATGACCGACCGCATGGACTGCCAGGTGGTTCATGTGCTCTTCTCCTWCATCTACCTCACGTCTACGTG
GCGCCCGCGGGCCCCACCGTCGGCAGCGTCTTGGCGGCCCGCGTGAAGCTGGCGCCGCTGGARCACTGGAC
CTGTTTGAGCGCATGGCCAACGTGCTGGTGGGCTCCAAGGTCTGTCCCCGGCCATCGGCGAGACCATCGCC
GCCACCGCGCGCCGAAGCTCCAGACGGGCGCCCTTTAG

SEQ ID NO: 177

Prototheca moriformis (UTEX 1435) phosphatidate cytidyltransferase

MQLQRALTIVRLRRRGWRRNSFNNVDAVSPKLVRKXSNAGQQTQSEDRFRSFKVRTLSTIVLIATFIGIIATG
HVPLMLMILGIQFLMVXEXFALARXAPQERKVPGFRAQQWYFFVAAFYLYIRFIKTNLTVELSSSAQTAAAM
FGWIIIRHHTLSFALYTAGFVSFVLRXLXKGLYXQFQYAWTHMILMVIFLPSSFFVYNLFEGLIWFLLPAA
LVVNDXAAAYLAGFFFRGTXLILKSPKKTWEGXIGGLAGTVVVGWYLAKLLSQFNWFICPRRDLIIQPLHC
DSKLDIYQPETFYLTDLVLTLLPADLATAVVTGLTRLSPGWQEAARQVSLTCLPMQLHAVVLATFASLIAPFG
GFFASGFKRGFKIKDFGDSIPGHGGMTDRMDCQVMSVFSXIYLTSTYVAPAGPTVGSVLAHAVKLAPLXQLD
LFERMANVLVGSKVLSPAIGETIAATARRKLQTGAL

SEQ ID NO: 178

Prototheca moriformis (UTEX 1435) 1-acyl-sn-glycerol-3-phosphate acyltransferase, putative

ATGTTGCTGCCCTTTGAGCTCATGTTTCGACCGCTACAGGCGGCATATTCTGAACAAGATCAACCAGTTTTGG
GCGCGGGTGGCCGCGTGGCCGTTTTACGGCGTCATCGTGAGYGGCCGCGAGAACCTGCCRCCTCCCGCGCAA

GCCGTGATCTATGTGCGCAACCACCAGAGCTACCTGGACATCACGGCGCTGCTGCACATGGGCGCMGACTTC
AAGTACGTCTCCAAGGCCGCCCTTTTCATGGTGCCTTTCTGGGCTGGGCCATGTTCTCACAGGGCACGTC
CGGCTCGTRCGAGACGACCGCGAGTCGCAGCAAAAATGCTTGAAACAGTGCGACACCCTCCTCCGCAACGGC
ACCTCCGTGCGCTTTTTTCCCCGAGGGCACGCGGAGCCAGGATGSSAAGATGCATGCCTTCAAAAAGGGCGCC
5 TTTATGGTGGCGGCCAGGGCGCGGTGAACGTCGTGCCCGTCACGATCGAGGGCTCCGGCGACCTGATGCCC
AGCCGGAACGAGTACCTGATGACCTACGGCGACATTCGGGTGCGCATCCACCGCCATCGCTCGCGCGGC
CGGTGCTGGAGTCGCTGCGCCTGGCGGCGCAGGAGGCGGTGGCGTCCGSGCTGCCCCGAGCGCCTGATCGGG
GACGACATCCGGCCGCGCKAGCTGGAGGCGCCCACGCCGCGCCGACAGAGGCGACAGAGGCGAGAGAGGCG
10 CCGCTGCACGAACAGCGCTGA

SEQ ID NO: 179

Prototheca moriformis 1-acyl-sn-glycerol-3-phosphate acyltransferase,
putative

MLLPFELMFDRYRRHILNKINQFWARVAAPFYGVIVXGRENLXPPGEAVIYVANHQSYLDITALLHMGXDF
KYVSKAALFMVPFLGWAMFLTGHVRLXRDDRESQQKCLKQCDTLLRNGTSVAFFPEGTRSQDXKMHAFKKGA
FMVAARARVNVPTIEGSGDLMPSRNEYLMTYGDIRVRIHPPIASRGRSLESRLAAQEAVASXLPERLIG
DDIRPRXLEAPTAPTEATEAREAPLHEQR

SEQ ID NO: 180

Prototheca moriformis (UTEX 1435) glycerophosphodiester
phosphodiesterase

ATGTCCAAGGCTTCTTCCACCATCGTCTCCGCTCCTGCGGGCCCGCAGGAGTCGCCCCGCTCGCTACAGTACC
25 GAGTCGACCGAGATCCGGAGCTCYCCCCGTGTTTGGACGCCTTTTGGACGTCAAGAGCAACGTTGTGCGCCCTT
GGAGGGCACAGAGGYCTGGGCGCCAACTACTGGGACAAGCCGGCGCAGGGCTCGCTCCGCCGCTTCAGCGTG
CGAGAGAACACCATCCCCCTCCTTCTCAAGTCGGTTGCGGCCGGCGCCAGCTTCATCGAGTTTGACGTGCAG
GTCACGCGGGACGGCGTGCCTGTGATCTGGCAGCAAACTATGTGGTAACGGGCACGCCCCGATTGCTTTGTG
AACCGGCTGATCAGCGACCTGAGCTACGACGAGTTTGTGAGCCTGGTGCCGGCCAGCGCCGAYCTGGCGCGC
30 CAGAAGGCGGGCCAGGTGGCGGGCCTGCGCGCGGAGGACGTGGAGGGCAGCACGCGCTTGCTGCGCGCGCTG
GTGGGCGACGAGCCRGTTGGACCCSGGCAACCCGACCGTGGCGGGGTGGGAGGCCGACTCGGAGTCGCGCCTG
CCCAGCTTGGCCGAGCTGTTTCGACGCCATCCCCCTCGCACGTGCGCTTTGACATCGAGATCAAGGTGCGGACC
AGCAGCCGCACGTTGCACACGCCGCTGCGCGAGGTGGAGCGCCTGCTGTCCGCCATCCTGACGGCCGTGGAC
GCGGCGCAGGAGGCGGCCGAGGCCGCGGGCCGCCGCGGCGCGAGATCCTGTACAGCTCCTTCGACCCGGAC
35 GTCTGCGCGGAGCTGGCGCAGCGCCGCCCGACGCGGCCGTGTCATGTTCTGTGCGGGCGGGCGGGCAGTATGCG
CACTCGGACCCCCGCGCACCTCCTTTGCGGCCGCCATCGACTGGGCGCGCGGCCCAACCTGGCCGGCGTG
ATCTTCCACGCGGGCCGCTGCGCACCGCGCTCGAGGCGGTGCTGCGCGACGCGCTGGCCGCGGGCCTCCAG
GTCATGACCTACGGCCAGGACAACACCGACGCGCGGTACGTGGAGGAGCAGTACGCTCGCGGCGTCCGCGGC
40 GTCATCGTGGACGACGTGAGTCTGGTCTGGACGCGCTCGTCCAGCGCGGCGTTCGCGCGGCGCTCGGTGCGG
CTGGCCGCGGCGAGGGCAAGCAGAACGGCGTGC CGGACTCGGAGAAGGTGAAGGCCTCGCCCCGCGCGCCTG
CCGGCCCTGGCCGCCGCTGCTAA

SEQ ID NO: 181

Prototheca moriformis (UTEX 1435) glycerophosphodiester
phosphodiesterase

MSKASSTIVSAPAGPQESPARYSTESTEIRSXPLFGRLLDVKSNVVALGGHRLGANYWDKPAQGSLLRRFSV
RENTIPSFLKSVAAGASFIEFDVQVTRDGVPIWHDNYVVTGTPDCFNRLISDLSYDEFVSLVPASAXLAR
QKAGQVAGLRAEDVEGSTRLRLALVGDEXVDXGNPTVAGWEADSESRLPSLAELFAAIPSHVAFDIEIKVAT
50 SSRTVHTPLREVERLLSAILTAVDAAQEAEEAAGRPRREILYSSFDPDVCAELAQRDPDAVMFLSGGGQYA
HSDPRRTSFAAAIDWARGANLAGVIFHAGRLRTALEAVVRDALAAGLQVMTYQDNTDAGYVEEQYARGVRG
VIVDDVSSVVDALVQRGVARRSVALAAGEGKQNGVPDSEKVKASPARLPALAAAC

SEQ ID NO: 182

Prototheca moriformis (UTEX 1435) diacylglycerol acyltransferase type 2

ATGACGGAGGCCGTCTCGGACAGGGATGCTCTSCAACCGAGCTGGGCTTCGTSCAGAAGCTGGGCATGTAC
TTTGCGCTGCCAGTTTGGATGATAGGCCCTCGTGGTGGGCGCCCTCTGGCTMCCCGTGACCGTGGTGTGCCTC
60 TTCATCTTCCCAAAGCTGGCGACATTGAGTCTGGGGCTGCTGCTGGTGGCGATGCTCACCCCCCTGTGCTG
CCCTGCCCCAAGCCGCTGGCCGCTTCTGCGCTACTGCACCACCGCCGCGGCCGAGTACTACCCCGTGCGS
TTCATCTACGAGGACAAGGAGGAGATGGAGGCCACCAAGGGCCCCGTCATCATCGGGTACGAGCCCCACAGC
GTCATGCCRCAGGCCATCTCCATGTTTRGCCGAGTACCCGCACCCCGCCGTGGTTCGGCCCCGCTGCGCAAGGCG
CGCGTGCTGGCKTCCAGCACCGGCTTCTGGACGCGGGCATGCGGCACCTGTGGTGGTGGTGGGCACGCGC
CCCGTGAGCAAGCCCTCCTTCTCGCCAGCTGCGCAAGCAGCGCTCCGTGCGCCCTCTGCCCGGGCGGCGTG
65 CAGGAGTGCTCTACATGGCCACGGCAAGGAGTCTGCTACCTCCGCAAACGCTTTGGATTTGTCAAACCTG
GCCATCCAGACCGGCACCCCGCTGGTCCCGTCTTCCGCTTTGGCCAGACGGAGACCTACACCTTTGTGCGG
CCCTTCATCGACTGGGAGACGCGCCTGCTGCCGCGGTCCAAGTACTTTTCGCTGGTGGCGCGCATGGGCTAC
GTGCCCATGATCTTCTTCGGCCACCTGGGCACGGCCATGCCCAAGCGGTGCCCATCCACATCGTCATCGGC
CGGCCCATCGAGGTGCCGACGAGGACGAGCCCGACCCGGCCACGGTGCAGCAGTACCTCGACAAGTTTCATC
70 GACGCCATGCAGGCAATGTTTGAAGAAGCAAGGCCGACGCCGCTACCCCAACCTGACGCTCGAGATCCAC
TGA

SEQ ID NO: 183

Prototheca moriformis (UTEX 1435) diacylglycerol acyltransferase type 2

5 MTEAVSDRDAXQPELGFXQKLGMYPALPVMMIGLVVGLWXPVTVVCLFIFPKLATLSLGLLLVAMLTPLSL
PCPKPLARFLAYCTTAAAEYYPVXFYIEDKEEMEATKGPVIGYEPHSMXQAIISMXAEPHPAVVGPLRKA
RVLXSSTGFWTPGMRHLWWLGTTRPVSKPSFLAQLRKQRSVALCPGGVQECLYMAHGKEVVYLRKRFGFVKL
AIQTGTPLVPVFAFGQTETTYTFVRPFIDWETRLLRPSKYFSLVRRMGYVPMIFFGHLGTAMPKRVPPIHIVIG
10 RPIEVPQQDEPDPAIVQQYLDKFDAMQAMFEKHKADAGYPNLTLEIH

SEQ ID NO: 184

Prototheca moriformis (UTEX 1435) Ketoacyl-CoA Reductase (KCR)

15 ATGGACGCTCAAGGATACCTCGTSAAGGCGTCCGAGTCKCCAGCATGGACGTACCTGGTSCGTGCTGGCCTCG
GCGCTGTTTCGCCATCAAGGTCGTGGGATTYGTCTGACGGTTCTGGGCGGGCTGTACGCCCACTTTCTCCGC
AAGGGAAAAGAACTGCGGCGATATGGGGACTGGGCAGTGGTGAAGTGGCGGACAGAGCGGCATCGGCAAGGCC
TACGCTGAGGCCCCGCGAAGCAAAAGCTGCGCCTGGTGTCTGATCTCGCGGACAGAGTCGCGTCTTGAGGAG
GAGGCCCGGCTGCTGCAGGACAAGTTTGGCGTGGAGGTGAAGATCATCCCCCGGACCTCAGTCCTCGGAC
20 GAGGCCGTGTTTGGCCCGATCGGCAAGGGCCTGGAGGGCCTGGACATTGGCATCTGGTGAACAATGCCGGC
ATGTCTTACCCCCACCCGGAGTACCTGCACCTGGTGCAGACGAGACCCTGGCCAACTCATCAATCTCAAC
GTCGCCACCCCGACCAAGCTGTGCAAGATGGTGTGGGCGGCATGAAGGAGCGCGGCCGCGGCTGGTGGTC
AACGTYGGSAGCGGCGTTCGCGAGCGCCATTCCCTCGGGGGCCCCCTCCTCTCGGCCTACACCGCCAGCAAGGCG
TACGTGGACAGCTGAGCGAGTTCGCTGAACGACGAGTACAAGGAGTTTGGCGTGCAGGTGCAGAACGAGCG
CCGCTGTTTGTGGCCACCAAGATGTCCAAGATCCGCAAGCCGCGCATCGACGCGCCACGCCCCGACCTGG
25 GCGGCCGCCGSGTYCGCGCCATGGGRTTCGAGACGCTGTCTTCCCCTACTGGTTCCACGCGCTGCAGGCG
GCCGTCGTGGAGCGCCTGCCGAGGCCATGATCCGCTACCAGGTGATGCAGATCCACCGCAGCCTGCGCCGC
GCGGCTTACAAGAAGAAGGCGCGCGCTCCGCCGCCGCCCTGGCGGACGAGGGCGTCTCCGCGGAGCCCCAAG
AAGGACCTGTGA

SEQ ID NO: 185

Prototheca moriformis (UTEX 1435) Ketoacyl-CoA Reductase (KCR)

35 MDAQGYLXKASEXPWPTYLXLLASALFAIKVVGXVLTVLGGLYAHFLRKGGKLLRRYGDWAVVTGATDGIGKA
YAEALAKQLRLVLISRTESRLEEEARLLQDKFGVEVKIIPADLSSSDEAVFARIGKGLEGLDIGILVNNAG
MSYPHPEYLHLVDEDTLANLINLVATPTKLCKMVLGGMKERGRGXVNVNXXSGVASAIPSGPLLSAYTASKA
YVDQLSESLNDEYKEFGVQVQNQAPLFAVATKMSKIRKPRIDAPTPTGWAAXXXRAMXFETLSFPYWFHALQA
AVVERLPEAMIRYQVMQIHRSLRRAAYKKKARASAAALADEGVSAEPKDL

SEQ ID NO: 186

Prototheca moriformis (UTEX 1435) glycerol-3-phosphate acyltransferase

45 ATGGAGAGGAGGACCTCGGCCAACCCCTCCGTGGGTCTTAGCAGCTTAAAGCACGTTCCGTTGAGCGCGGTG
GACCTCCCCAATTTTCAGTCCGTTCCTCAAGGGTGGCGCCAGCCCCGATGTCCATCGTCAGATAGAGGAGCTG
GTGGACAAGGCACGCGCAGATCGGTACCCGGCATCGTTGTGCTCCTAGCCGATGTGCTGGACATCAGCGGG
CCGCTGCAGGACGCGGCCTCGGCGATGGTGGACGACTCCTTCCTGCGCTGCTTCACCTCCACCATGGACGAG
CCCTGGAACCTGGAACCTTTTACCTGTTCCTCGCTCTGGGCGCTCGGCGTCTGTCGTRCGCAACRTGATCCTGTT
CCCCGCGCCTCCTCACCATCGTGTGGGACGCTGCTTTCGTGCTGGCCTTTGCCGTGACGGGCTTCCTC
50 CCCAAGGAGAGGCGGCTCGCGGCCGAGCAGCGGTGCGTCCAGTTTCATGGCGCAGGCATTCTGGCCTCCTGG
ACAGGCGTGATCCGGTACCACGGCCCCCGCCCCGTGGCGGCGCCCCAACCGCGTGTGGGTGCGCAACACACG
TCCATGATCGACTACGCGGTGCTGTGCGCCTACTGCCCCGTTTGGCGCCATCATGCAGCTGCACCCSGGCTGG
GTGGGCGTCTTRCAGACGCGCTACCTGGCCTCGTGGGCTGCCTCTGGTTCAACCGCACGCAGGCCAAGGAC
CGCACGCTCGTGGCCAAGCGCATGCGCGAGCAGTGGCCAGCGCGACCAGCACGCCGCTGCTCATCTTCCCC
GAGGGCACCTGCGTCAACAACGAGTACTGCGTCTGTTCCGCAAGGGCGCCTTRGACCTGGGCGCCACCGTC
55 TGCCCCGTSGCCATCAAGTACAACAAGATCTTCGTGCGACGCTTTTGAACAGCAAGCGCCAGTCCTCTCC
GCGCACCTCATGAAGATCATGCGCTCGTGGGCCCCGTGCTGCGACGTCTACTTCTGGAGCCCCAGACGCGG
AGGGAGGGCGAGTCGGTCGAGGCGTTTGGCAACCGCGTGCAGGCGCTGATCGCGCAAAAGGCCCGGCTGAAG
GTCGCGCCCTGGGACGCGTACCTCAAGTACTACAACCTGGCCGAGAAGCACCCGGACCTGATCGAGAACCAG
CGGCGCCAGTACAGCGCCGTATCAAAAAGTARGCCGAGGASTGA

SEQ ID NO: 187

Prototheca moriformis (UTEX 1435) glycerol-3-phosphate acyltransferase

65 MERRTSANPSVGLSSSLKHVPLSAVDLPNFESVPKGAASPDVHRQIEELVDKARADRSPASLSLLADVLDISG
PLQDAASAMVDDSFRLCFTSTMDEPNWNWNYLFPLWALGVVXRNXILFPLRLLTIVLGTLLFVLAFAVTGFL
PKERRLAEEQRCVQFMAQAFVASWTGVIRYHGPRPVAAPNRVWVANHTSMIDYAVLCAYCPFAAIMQLHXGW
VGXXQTRYLASLGLFVFNRTQAKDRTLVAKRMREHVASATSTPLLIIFPEGTCVNNEYCVMFRRKGAXDLGATV
CPXAIKYNKIFVDAFWNSKRQSFSAHLMKIMRSWALVCDVYFLEPQTRREGESVEAFANRVQALIAQKARLK
VAPWDGYLKYNNLAEKHPDLIENQRRQYSAVIKKXAE

SEQ ID NO: 188

Prototheca moriformis (UTEX 1435) Monoglyceride lipase

ATGCCCCCTCCCGAACGCCATAGAGCTCCCGGGCGCCGATCCCAAATCGAGCTCGGTGTACAGCTTTTACGGGG
CACAGCACGAATCTGCAAGTCAACGCCAGGGGCGTTCGCKAGTCCGTGTACGACGARGTCGACCCCGCAACG
TACCTGGGGCCAGAGGGCGGCAGGAGACGGTCTGTAACGCGCAGGGCCTCACGCTGCGGGTTTACTACTGG
5 CCGGCAGAGCAGCCCAAGGCAATCTTGCAGTTTGTGCACGGACATGGCGCGCACGCACTGTTTGAACCTTTG
GAMATCGATGAGGCTGGAAAGCCGCCAAGCTACGCTGGCTGGGTGGCCAAGCTCAATGCCGCGGGCATCAGC
GTGGTTCGCCCCACGACAATCAGGGCTGCGGGCGCTCAGAGGCGGCGCGGGGCTGCGCTTCTACATCGAGTCC
TTTGACGATTTTGTCAAGGACGTGTTGCTTCTCCGAGGGAGCTGGTGAAGCAAGAGGGCTTTGACAGCYTG
10 CCGGTYTTTCTGGGGGGCATCTCACTWGGYGGCTGCATAGCGCTCACCTCCCTGCTCGAAGAGCCCGACCAR
TACCRCGGGCTGTGCCTGCTGGCGCCCATGATCTCGCTSCAAAAGGTGTCCCGCAAGGGCCTYAAYCCCTAC
CTGCGGCGCTGGCGGCGCTGCTGAGCCGCGTCTCCCGTGGGCGCCCATCATCGCGACCGACCGCACCTCC
AACTCTGTCTCCAGCTGCAGTGGGATGCAGACCCACTGACCGCGCACATGAACACGCGCGTCCGCAACGCC
AACAGTACCTCTGCGCGACGGAGCGCGTGACGGCCCGGCTGRCRCGYTGAGAAGCCCTTCATCGTCTTC
15 CACTCGGAGAACGACACAATGTGCGACGCGGACGGCTCCAAGCAGCTCTACGCGAGGCACAGTCCAGCGAC
AAGACCATCCGCTTTGTCAACAGCATGTGGCACGTGCTGGTCAAGGARCCGGGCAACGAGGAGGTCTGTCAG
ASTCTGGTGCAGTGGATCCTCGAMCGGGCGTGA

SEQ ID NO: 189

Prototheca moriformis (UTEX 1435) Monoglyceride lipase

MPLPNAIELPGADPKSSSVYSFQGHSTNLQVNARGVXQSVYDXVDPATYLGPRGRQETVVNAQGLTLRVYYW
PAEQPKAILQFVHGHGAHALFELLXIDEAGKPPTYAGWVAKLNAAGISVVAHDNQGCGRSEAARGLRFYIES
FDDFVKDVLRLRELVKQEGFDSXPFLGGISXXGCIALTSLLEEDPXYXGLCLLAPMISXQKVSRKGXXPY
20 LRPLAALLSRVVPWAPIIATDRTSNXLQLQWDADPLTAHMNTRVRNANEYLCATERVTRARLXXXQKPFIVF
HSENDTMCDDADGSKQLYAEAQSSDKTIRFVNSMWHVLVKXPGNEEVXQXLVQWILXRA

SEQ ID NO: 190

Prototheca moriformis (UTEX 1435) acyl-CoA oxidase

ATGGATGCTCAGAAGCGAGTCCAAGTCTTGTCCGGTCACTTGTCCGAAACTGCGGGCGATGCCGTGGCGGTC
TGCCGGGCGGACACCCTGGCCGCGGGGGTGAGCCCCAAGTTCGACCAGGCCAGGACATGACCGTGTTCCTCC
CCGGCCAACCACGATGCGCTCTTCTCAGCGACCTGCTGACTCCCGAGGAAAAGGATGTGCAAATGCGCGTC
CGCAAGTTTTCGCGAGGAGACCGTTGCGCCCAACGTGATTCCTACTGGGAAAAGGCCGAGTTCCCGCACGCG
35 CTGCGKCCCCAGTTTGGRAAGCTGGGCATCGGCGGCGGGCACAACAAGGGGCGACGGCTGCCCGGCGCTGAGC
ACCATGGCGGCGGCTCGGCGGTGGTGGAGCTGGCGCGCGTGGACGGCTCCATGAGCACYTTTTTCTGGTS
CACACTACCTGGGCGCGATGACGGTGGGCTGCTGGGCTCSGAGGCGCAGAAGCGGGAGCTGCTGCCCGGC
TTCCACRCCTTTGACAAGGTGTGCAGCTGGGCGCTGACGGAGCCGAGCAACGGGTGCGACGCTCGGCGCTG
ACCACGACCGCGCGSCGCGTGGAGGGCGGCTGGCTGCTCMGCGGGCGCAAGCGCTGGATCGGCAACGCGACC
40 TTTGCGGACATCATCATCTGCTGCGCGCGCAGCAGCGTGACGGGGCAGGTCAACGCCTTCATCGTGCGCAGG
GGCGCGCCGGGGCTGACCACGAGCAAGATCGAGAACAAATCGCGCTGCGCTGCGTGCGAGAACGCCGACATC
CTGATGGAGGACGTGTTTGTCCCCGACGCGGACCGCTGCCCGGSGTCAACTCGTTTCGCGGACGCGGCGGGC
ATGCTGGCCATGTCGCGCTGCCTGGTGGCTGGCAGCCGGTGGGCTGGTGGTGGGYGTGTACGACATGGTG
45 CTGCGCTACACGCAGCAGCGCCGCGCAGTTTGGCGCGCCGCTGGCCGCCTTCAGCTGGTGCAGGAGCGGCTG
GCGCGCATGCTGGGCACCATCCAGGCCATGTACCTCATGTGCGCGCGCTGAGCAAGCTCTACGACGAGGGC
CGCATGACGCACGAGCAGGCCAGCCTGGTCAAGGCCTGGACCACCGCGCGCTCGCGCGAGGTGCTCGCGCTC
GGCCGCGAGTGCTTGGGCGGCAACGGCATCGTCGGCGAGTTCTCGTGGCCAAGGCCTTTGTGACGCGCGAG
GCGTATTACAGTATGAAGGGACCTACGACGTCAACGTGCTCGTCGCCGCGCGCGGCTCACCGGCTACGCC
GCCTTCAAGACGCCCGCGCGCGGCAAGAAGGACGMCGCCGCAAGAAGGAGCACTGA

SEQ ID NO: 191

Prototheca moriformis (UTEX 1435) acyl-CoA oxidase

MDAQKRQVLSGHLSETAGDAVAVCRADTLAAGVSPKFDQAQDMTVFPPANHDALFLSDLLTPEEKDVQMRV
RKFAEETVAPNVI PYWEKAEPHALXPEFXKLIGGGHNGHGCPLSTMAAXSAVVELARVDGSMXFFLX
55 HTYLGAMTVGLLGXEAQKRELLPGFHXFDKVCWALTEPSNGSDASALTTTAXRVEGGWLLXGRKRWIGNAT
FADIIIVWARSSVTGQVNAFIVRRGAPGLTTSKIENXIALRCVQNADILMEDVFVPDADRLPXVNSFADAAG
MLAMSRCLVAWQPVGLVVXVYDMVLRYTQRRQFGAPLAQFLVQERLARMLGTIQAMYLMCARLSKLYDEG
RMTHEQASLVKAWTTARSREVVALGRECLGGNGIVGEFLVAKAFVDAEAYTYEGTYDVNVLVAAARGLTGYA
60 AFKTPGRGGKDXAAKKEH

SEQ ID NO: 192

Prototheca moriformis (UTEX 1435) enoyl-CoA hydratase

ATGGCGCCGAGCCCGACCTTCAAGGCACTCAAGGTCTGGATGGACGACGATCTGGTGGGACACATCCAGCTC
65 AATCGTCTCAAGCCGCCAATTCGATGAATGAGGAGATGTGGACGAGCTGCCGAGGCCACAGCGTGGTTG
GAGAGCCAGTCTGCGCGCGTCATWCTGGTCACGGGCGSGGCAAGAACTTTTTCGCGGGCATCGACGTGGCG
AGCCTCGGCTCGACCGTGCTCTCCACGKCGCCGGCTGCGCGGCGCGCTCGGCCTACCGCTTCCGCMCGGGC
CTGACCAMSCTGCAGGAGGCCATGACYGCCCTGGAGCGCGTGCGCTGCCCCACCGTGGCGCTGGTGCACGGY
70 CACTGTGTSGGGGCGGCGTGCACCTCATCACGGCCTGCGACATCCGKTTTGCCACGGCGGACGCCAAGCTS
TGCGTCAAGGAGGTGARCTGGCCGTGATCGCCGACATGGGCACGCTCCAGCGCCTCCAGGCATCGTGGGS
CAKGGCCACGCCCGGAGCTGTGCTGACGGCCCGGTCTTCTCCGGGGAGGAAGCAAAGCAGATGGGTCTC

GCCACCGCGCGCTGCCGAGCCGGGACGAGCTGGAGAGGCACGGCCGCGCGGTGCGCGAGGGCATCGCGGCT
AAATCCCCCGTGGCCGTGGCTGGGACCAAGCAGGTGCTGCTCTACCAAAGGGACCACACCGTCGCCGATGGC
CTGGACTACGTCAACACCTGGAACGCTGGCATGCTGCACTCGGCTGATTTGCGCGAGGTGTTTGCGGCCATG
AAGGAGAGGCGTAAGCCGGTGTTTCAGCAAGCTCTGA

5

SEQ ID NO: 193

Prototheca moriformis (UTEX 1435) enoyl-CoA hydratase

MAPSPTFKALKVWMDDDLVGHIQLNRPQAANSNMNEEMWTELPQATAWLESQSARVXLVTGXGKNFCAGIDVA
SLGSTVLSHXAGCAARSAYRFRXGLTXLQEAMXALERVRCPTVALVHXHCXGAGVDLITACDIXFATADAKX
CVKEVXLAVIADMGTLQRLPGIVXXGHARELSLTARVFSGEEAKQMGMLATAALPSRDELERHGRAVAQGI
KSPVAVAGTKQVLLYQRDHTVADGLDYVNTWNAGMLHSADFAEVFAAMKERRKPVFSL

10

SEQ ID NO: 194

Prototheca moriformis (UTEX 1435) mitochondrial acyl carrier protein
allele 2

ATGGCGTTCTTGCAGCGGACGAGCGCGCTGGTGCGCCAGGGTGTCTGTCTCGCCTCGCCGTGCAGGCCAGC
CCCCCCTGAACAGCGTCCGGGCTTCGCCTCGGCGTCTTACCTGGACAAGAATGAGGTGACGCACCGCGTG
CTGTGATTTGTCAAGAACTTTGACCGCGTCGACGCCGGAAGGTACGGACTCTGCCAACTTCCAGTCCGAC
CTGGGTCTGGACTCACTAGACACCGTGGAGCTGGTCATGGCCCTGGAGGAGGAGTTTGCATCGAGATCCCG
GATGCGGAGGCCGACAAGATCCTCTCCGTCCCGGAGGCGATTTCTACATTGCCGGAACCCCATGGCCAAG
TAG

20

SEQ ID NO: 195

Prototheca moriformis (UTEX 1435) mitochondrial acyl carrier protein
allele 2

MAFLQRTSALVRQGVLSRLAVQASPALNSVRAFASASYLDKNEVTHRVLISIVKNFDRVDAGKVTDSANFQSD
LGLDSLDTVELVMALEEEFAIEIPDAEADKILSVPEAISIYIAANPMAK

30

SEQ ID NO: 196

Prototheca moriformis (UTEX 1435) trans-2-enoyl-CoA reductase

ATGCCACTGGAAATCTTGTCAAGACCAGGAATGGTGCACCTGCCTTTTCCAGAGACGGCGGCGCCATTACT
GTGCATAGCACGAGTGCACGGTGCAGGAGGTCAAGGGCCTCATCGCTCGCGCCAAGAAGTTGAGCCCCGCC
CGCCTGCGCCTGACGTTGCCCGCGCCGGCCGACGCGGCCACGGTGCTGGAGGACAAGAAGCCCTTGGCC
GACTATGGTCTGCACGACGGCGCCAGCCTCGTGCTCAAGGACCTGGGACCTCAGATCGGGTACCAGATGGTC
TTTTCTGGGAGTARTTRGGGCGCGTGGCCATCTACCCCTCTTCTACTTCCTGCCYTCCTTGATTTACGGA
AGGRCGACCGAGCACGTCTTTCGCMCAAAGGCAGCGCTCGCCTTCTGGACCTTRCACTACGSGAAGCGCATC
RTGGAGWCCTTTTTCTGTGCACAYGTTTGGGCACGCCACGATGCCCCGTGCGCAATCTGGTGAAGAACTGCAGC
TATTACTGGAGCTTTGGCGCCTTCATCTCCTACTTTRGTGAACCAACCCGCTCTACMCGGCGCCTCCCGCGCC
CAGACCGCGCTCGCCTTTGTGGCCGCCACCCTCTGCACGCTCTCGAACTTCAAGTGTCATCTGATCCTGAGC
AACCTGCGTGCGCCCGGAGGCAGCGGCTATGTCATTCCGCGTGGCTTCTGTTTGACTACGTACCTGTGCC
AACTACACCGCTGAAATCTGGAGCTGGATTTTCTTCTCCATCGGCACCCAGTGCTTGCCGGCYCTCYTYTTC
ACCGTRGCTGGYGCCGCGCAAATGGCCATCTGGGCCGGGGGCAAGCATTGCCGTCTGAAAAAGCTCTTCGAC
GGCAAGGAGGGCCGCGAGCGCTACCCCAAGCGCTACATCATGRTTCCCCCGCTGTACTGA

35

40

45

SEQ ID NO: 197

Prototheca moriformis (UTEX 1435) trans-2-enoyl-CoA reductase

MPLLEILVKTRNGRPAFSRDGGAITVDSTSATVQEVKGLIARAKKLSPARLRLTLAPAGTRPTVLEDKKPLA
DYGLHDGASLVLKDLGPQIGYQMVFFWEXXGPLAIYPLFYFLXSLIYGRXTEHVFXQKAALAFWTXHYXKRI
XEXFFVHXFGHATMPVRNLVKNSYYWFGAFISYXVNHPLYXAPPAQTAVAFVAATLCTLSNFKCHLILS
NLRAPGGSGYVIPRGFLFDYVTCANYTAEIWSWIFFSIGTQCLPLXFTXAXAAQMAIWAGGKHCRLLKFLD
GKEGRERYPKRYIMXPPLY

55

SEQ ID NO: 198

Prototheca moriformis (UTEX 1435) long-chain-fatty-acid-CoA ligase

ATGGCTTCCAAGTATCCCCGGCTGACCAAGGTCTCCGATGGCGTGCTGCCAACGAGGAGGGCGTCGAGTTT
GGACCCGAATACAGTTGCACATTCATGAAGGAAAAGACGCTGTCCGTGCCCGGGATCAAGTCTCTGGGCGAG
CTCTTTGCGGACTCGGTGTGCGAGCACAAGCGCCGGCCGTGCCTGGGCAAGCGCACCAAGGACGGCTACAGC
TGGCGCACGTACGAGCAGACGGGCCGCGAGATCGCGGCCATGGGCGCGGCCTTTGCCGCGCGCGGCCTGGCC
CAGCAGTCGCGTGTGGGCGGTGTACGGGCCCAACGCGCCCGAGTGGATGATCAGGATGCAGGCGTGCAACCGC
CAGGGCTACTACTGCGTGCCTGTACGACTCKTGGGCGAGACCGCGGTGGAGTTTCATCATCAAGCACGCG
GAGGTGAGCGCGGTGGCGGTGGCCGGGCCCAAGCTGGCRGAGCTGGCCAARGCSCTGCCCAACGTGGCGAGC
CAGGTCAAGACGGTGGTCTTGTGGGACGASGCCGACCGYGC GGCGGTGGAGGCGGTGCAGGCGCTGGGCGTG
GGCGTGTGGCTTCGACGAGTTTGTGAAGCAGGGCGAGGCCGCGGAGGCCGTGCCGCCCGCCAGCGTGGAG
GGCGAGGACCTGTGCACCATCATGTACACGAGCGGCACGACGGGCGACCCCAAGGGCGTCATGCTCACGCAC
CGCGCCGTGATCGCGACYGTGCTGAGCCTGSACGCTTYCTGAAGAGYGTGGACGTGGCGCTGGGCGAGGGC

65

70

GAYGCCATYYTKTCYTTCTGCGCTGGCGCACATCTTYGACCGCGCGGCGGAGGAGCTGATGCTCTACGTG
 GGCGGCAGCATCGGCTACTGGAGCGGCAACGTCAAGGGCCTGCTGGGCGACATCGCGGCGCTGCGCCCGACG
 CTCTTCTGCTCCGTGCCGCGCGTGTGTTGACCGCATCTACTCCTCGGTACGCGGCAAGGTRGAGAGYGGCGGC
 TGGCTCAAGCGCACCATGTTCCACCACGCCTTCAAGACCAAGTTYGGGCGGCTGAAGCGCGGCGTGCCGCGAG
 5 GCCAAGGCCGCGGCATCTGGGACAAGATCGTGTTCAGAAGATCAAGGAGGCGCTGGGCGGCGGCGTGCAAG
 ATCATCGTCTCGGGCGGCGCGCGCTGTGACGCACGTGGAGGAGTTCTGCGCGTCTGCATGTGCTCGCAT
 GTGGTGCAGGGGTACGGGCTGACCGAGACCTGCGCGGCCAGCTTCATCGCGGAGCCGCGCCGACATCCGCCAG
 ATGGGCGACGTCGGCCCCGCGCAGCCGCGGTCAGCTTCCGCTGGAGGGCGTGCCGCGAGATGAAGTACAGC
 10 CCCAGCGCGGACCCCGCGCGTGCGGAGCTGCTCATCCGCGGGCCCGCGCTCTTCTCCGGCTACTACAAGGAC
 CAGGAGAAGACGGACGAGGTGGTCACGCCGAGCGGCTGGTTCACACGCGGCGACATCGCGGAGATCACGCCC
 GCCGCGGCCATCCGCATCATCGACCGCAAGAAGAATCTTCAAGCTCGCGCAGGGCGAGTACGTGGCCGTS
 GAGAAGCTGGAAGAACAGTACAAGATGTGCGCGGCGTGGAGCAGATCTGGGTCTACGGCAACAGCTTCGAG
 AGCGTGCTCGTGGCCGTGGTTCGTGCCCAGCGAGGACAAGGTCAAGGCCACGGCGGCTCCAGCGCCGCGCAG
 15 CTCGCCTCGGACGCGCTTCAAGAAGGCCGTGCTGGACGACCTCACCGCCGCGCCCAAGGCCGACAAGCTC
 AAGGGCTTCGAGATGGTCAAGGGCGTCATCATCGAGCCCGAGCCCTTCAGCGTCGAGAACAACCTGCTCAGC
 CCCACCTTCAAGCTCAAGCGGCCGCGAGCTGCTCGACCATTACCGCACCGAGATCGACGCGCTCTACGCGAGT
 CTGAAGAAGTAG

SEQ ID NO: 199

Prototheca moriformis (UTEX 1435) long-chain-fatty-acid-CoA ligase

MASKYPRLTkVSDGVXANEEGVFEGPEYSCTFMKEKTLSPVGIKSLGELFADSVSQHKRRPCLGKRTKDGY
 WRTYEQTGREIAAMGAFAARGLAQQSRVGVYGPNAPEWMITMQACNRQGYVCPLYDXLGETAVEFIKHA
 EVSAVAVAGPKLXELAXLPNVASQVKTVVXWDXADXAAVEAVQALGVGVFGDFVVKQGEAAEAVPPASVE
 25 GEDLCTIMYTSGTTGDPKGVMLTHRAVIAXLXSLXAXLKXVDVALGEGXAXXXFLPLAHIXDRAAEELMLYV
 GGSIGYWSGNVKGLLGDI AALRPTLFCSVPRVFDRIYSSVTGKXEXGGWLKRTMFHHAFKTKXGRLKRGVPQ
 AKAGGIWDKIVFKKIKEALGGRCKIIVSGGAPLSTHVEEFLRVCMCSHVQGYGLTETCAASFIAEPADIRQ
 MGTVGPPQPAVSFRLEGVPENKYSADPARGELLIRGPALFSGYKYDQEKTEDEVVTPDGFHTGDI AEITP
 AGAIRI IDRKKNIFKLAQGEYVAXEKLNTYKMSPAVEQI WVYGNFSFESVLVAVVVPSEDKVKAHGGSSAAQ
 30 LASDXAFKKAVLDDLTAAAKADKLKGFEMVKGVIIEPEPFSVENNLLTPTFKLRPQLLDHYRTEIDALYAS
 LKK

SEQ ID NO: 200

Prototheca moriformis (UTEX 1435) Enoyl-CoA reductase

ATGGCGTCCYCYCAGCTCCTGTCCAAGTCCCAGGGTCTGGTGGTGCCTGAAGGACGTGAGGGCCGTCCGG
 GTGYCCCTGCCCGCGCCGACCCCYTGCCGCCCCCTCGCGCCGATGCCTACTCGAGCTCCAGGGCGCCTGCC
 TCCATGGTGTGTCGCCCCCAGCGCCGCGGGCTGAGCGTGAAGGCTGCCGCCAGCACCAGCCCTGCCGGCACC
 40 GGCTTGCCCATTTGATCTGCGGGGCAAGAAGGCCTTTATCGCGGGCGTGCAGGACGACGAGGGTTTTGGCTGG
 GCCATRGCCAAGCAGCTGGCTGAGGCGGGCGCGGAGATCTCCCTGGGCGTGTGGGTGCCTGCCCTGAACATT
 TTCGAGAGCAACTACCGCCGCGGCAAGTTTGACGAGAACCAGCAAGCTGGCCAACGGCGGCCTGATGGAGTTT
 GCGCATCTACCCCATGGATGCGGTCTTCGACTCGCCGCGGACGTGCCGAGGACATTGCCAGCAACAAG
 CGCTACGCCGGGAACAAGGGCTGGACGGTGAAGGAGACGCGCCGACAAGGTGCGCGCCGACGTGGGCAAGATC
 45 GACATCCTGGTGCCTCGCTGGCCAACGGGCCCCGAGGTGCAGAAGCCGCTGCTGGAGACCAGCCGCGCGGC
 TACCTCGCCGCSCTGAGCGCCTCCTCCTACTCCCTCATCTCCATGGTCCAGCGCTTTGGCCCGCTCATGAAC
 CCCGGCGGCGCGTTCATCTCGCTGACCTACAACGCCTCCAACCTGGTTCATCCCCGGCTACGGCGGCGGCATG
 AGCACGGCCAAGGCCGCGCTCGAGTCCGACACGCGCGTGTGGCCTACGAGGCCGCGCAAGTACCACGTG
 CGCGTGAACACCATCAGCGCCGGGCGCTCGGTGCGCGCCGCGCAAGGCCATCGGCTTCATCGACGACATG
 50 ATCCGCTACTCGTACGAGAAGCGCCCATCCAGAAGGAGCTCAGCGCCTACGAGGTGCGCGCGCGCCGCC
 TTCTCTGCTCCCCGCTCGCTCGGCGCTCACTGGCCACGTGCTGTTTGTGACAACGGCCTCAACACCATG
 GCCTCGCCCTCGACTCCAAGACCCTCGACCYAGCGAGTGA

SEQ ID NO: 201

Prototheca moriformis (UTEX 1435) Enoyl-CoA reductase

MASXQLLSKSQLVLGALKDVRVRVXLPRAGPXAAPRADAYSSSRAPASMLVAPQRRGLSVKAAASTSPAGT
 GLPIDLRGKKAFFIAGVADDQFGWAXAKQLAEAGAEISLGVVVPALNIFESNYRRGKF DENRKL ANGLMEF
 AH IYPM DAVF DSPADVPEDIA SNKRYAGNKGWTVSETADKVAADV GKIDILVHSLANGPEVQKPLLET SRRG
 60 YLAXLSASSYSLISMVQRFGLMNP GGAVISLTYNASNLVIPGYGGGMSTAKAALES DTRVLAYEAGRK YHV
 RVNTISAGPLGSRAAKAIGFIDDMIRYSYENAPIQKELSA YE VGXAA AFLCSPLASAVTGHV MFVDNGLNTM
 GLALDSKTL DXSE

SEQ ID NO: 202

Prototheca moriformis (UTEX 1435) membrane bound O-acyl transferase domain-containing protein

ATGGCGCAGAACCTCATCCTCTCCATGCCGTGGGAGGAGTGGGCGTGCTCGTCCCTGGGCGTGTCCCTTTCT
 GAGCTGCGATTTGTTGTATCATTTTTTGCCTACGTGCTCGTGTCCGCGTGCTCCGCCACATCCGTGGCACC
 CGGACCCGGCACTGGTTTGCCTCGTCACCGGGTTCCTGCTCATCTACTACCCCTTTGGCTCGGGCGTGATG
 70 CACGCCCTTGTGTCTCCACCR TGGTGTACCTGGCKATGGCGGCAGTGCCCTCCCACTGCGGGACSTGGCC
 TGGCTCATRGCTTTCCGTACCTCATCCTCAACCACGTGCTCCAGGCCAGCGGCCTMAGCTGGAAGGAGGGC

CAYCTGGATTTTCACGGGCGCGCAAATGGTGTGACGTTGAAGCTGATCGCCGTCGGCGTCTGCTACCAGGAC
GGCCGGTCCGGCCAGTTCGTACGCRAGCCACTACCGCCAGTCCATGCGCCTGCCCYCGCTGCCGGGCTCCTG
GAGTACTACAGCTACTGCTTTGCCTACGGCAACCTCCTGGCCGGCCCCCTTTTTCGAGGCGAAAGAGTACTTT
GATTTTCATGGCSCGGCAGGGGAGTGGGACGAGGGCGAGCACGGCCGCTGCCCYGCGGCTGGGCGCCGGC
5 CTGGTCCGCTTCTCAAGGGCATCCTGTGCGCCGCGCTGTGGATGCAGCTGGGCAAGCGCTTACGCGCAGT
CTGCTCGAGTCGGTCTTCTGGACCACGCTGTCCGTCCCGCGACGAATGGCGCTGATCTGGGTTCATCGGCTTT
GCCGCGCGGCTCAAGTACTACTTTGTCTGGTCCGTGGCCGAGTCTGGGCTCATCTGTCSGGCCAGTGCTTC
GCCGGCAAGACCCAAAAGGGCAAGGCGCAGTGGACGCGGTACATCAACACCAAGATCAGGGAGCTGACGTCTG
GGCATCTGGCACGGGCTCTTCCCCGGCTACTGGGCCTTTTTCGGGACGAGCGCGCTCATGTTTCGAGGCGGCC
10 AAGACCATCTACCGCTACGAGCTGGGCTGGCCGGCCTGGCTGCGCGCGGCGCTGCCCTGGCGCGCCCTCAAG
ATGGCGCTCACGGCCTTTGTGCTCAACTATGCCGCCATGTCCTTCTGGTGTCTCACCTGGGCGGACACCTGG
GCCGTCTGGAAGTCGGTGGGCTTCTGGGGCACACGCTGCTGCTCCTGATCCTCATAGTCGGCGCTCGTGCTG
CCGCCCCGCGCGGACACAAGGGCCAGGCCACCACGCGCCCGTGGTCACGCCCCGCCCATCCCGGCCGAG
GGAGAGCTGCCMCAYAAGAAGGCAGAGTGA

SEQ ID NO: 203

Prototheca moriformis (UTEX 1435) membrane bound O-acyl transferase
domain-containing protein

MAQNLILSMPWEESACSSLGLSLSELRFVVSFFAYVLVSAVLRHIRGTRTRHWFALVTGFLLIYYPFGSGVM
HAFVSSTXVYLXMAAVPSHCGXLAWLXAFPYLILNHVLQASGXSWKEGXLDFTGAMVLTCLKLIAGVVCYQD
GRSQSYXSHYRQSMRLPXLPLGLEEYYSYCFAYGNLLAGPFEEAKEYFDFMXRQXEWDEGEHRLPXGLGAG
LVRFLKGILCAALWMQLGKRFSASLLESVFWTTLSVPRMALIWVIGFAARLKYYFVWSVAESGLILXGQCF
AGKTQKGKAQTRYINTKIRELTSGIWHGLFPGYWAFFGTSALMFEEAKTIYRYELGWPAPWLRAALPWRALK
20 MALTAFLVNLNAAAMSLVLTWADTWAVWKSVMGFLGHTLLLLILIVGVVLPPIRRRHKGQATTAPVVTAAIPAE
GELXXKKA

SEQ ID NO: 204

Prototheca moriformis (UTEX 1435) NAD-dependent glycerol-3-phosphate
dehydrogenase

ATGTCCGTCCAACCCATTTCCGAGGAGCTGACCAAGGTGACCGTCATCGGCGCCGGCGCCTGGGGTACCGCC
CTCGCCATCCACTGCGCCCGCAAGGGCCATGACACCATGTTGTGGGCCATGGAGCCGCACGCGGTGGAGGAG
ATCAACACCAAGCACACCAACGAGACTTTCTGAAGAACGTGCCCTGCCTGAGTCGCTCAAGGCGACGGGC
35 GACATCAAGGAGGCCGTGGAGCGCGCGGAGATGATCCTGACCGTCATCCCGACTCCCTTCTTGTTCAGTCTG
ATGAGCGGGATCAAGGAGTACCTGAGGGAGGACCAGTACATCGTGTCTGCACCAAGGGCATTCTGAACGAC
ACCTTGGAGACCCCCGACGGCATCATCCGCCGCGCCCTGCCCGATAACCTCTGCAAGAGGCTGGCGTTCTCTG
TCGGGCCCCCTCCTTTGCGGCCGAGGTGGGCGCGCACTTCCCCACGGCGGTACCATCGCGGCCGAGGACCCG
GCCGTGGCGCAGCGCGTGCAGCAGCTCATGTCCACCAACCGCTTCCGCTGCTACCGCACCACCGACGTCTGTG
40 GGCGTCGAGCTCGCTGGCGCGCTGAAGAACGTGCTGGCCATCGCCTGCGGCATCTCCGACGGCTGCGGCTTT
GGCAACAACGGCCGCGCGGCGCTCATCACGCGCGGGCTGGACGAGATCACGCGCATCGCGGTGGCCAGCGGC
GCCAACCCTGACCTTGGCGGGCTGGCGGGCTGGGCGACATCGTGCTCACCTGCTGCGGCGACCTGAGC
CGCAACCGCACGGTGGCGCTGCGCTGGGCAAGGGTGAGAAGCTGGACCACATCGTCAAGACGCTGGGCGCC
ACGGCCGAGGGCGTGTCTCACCTCGCGCGCGGCTACGACCTCACCAAGAAGCTGGGCATCGACTGCGCCGTC
45 ATCCACGGCATCTACCGCGTCTGAACGAGGAGGCCGAGCCCATGAAGGTCTGGCCGAGACCATGGCGCGC
CCGCTGCGCGCCGAGGTGGCCGAGAGCGTCGCCGAGGCCCGCGCTACCAACGCCCGCAACCTCGCCGAGTAG

SEQ ID NO: 205

Prototheca moriformis (UTEX 1435) NAD-dependent glycerol-3-phosphate
dehydrogenase

MSVQPISEELTKVTVIGAGAWGTALAIHCARKGHDMLWAMEPHAVEEINTKHTNETFLKNVPLPESLKATG
DIKEAVERAEMILTPFLFKSMGIKEYLREDQYIVSCRTKGILNDTLETDPDGIIRRALPDNLCKRLAFL
SGPSFAAEVGRDFPTAVTIAAEDPAVAQRVQQLMSTNRFRCYRTTDDVGVLAGALKNVLAIAACGISDGCFL
55 GNNGRAALITRGLDEITRIAVASGANPLTLAGLAGVGDIVLTCGDLSRNRTVGLRLGKGEKLDHIVKTLGA
TAEGVLTSRAAYDLTKKLIDCAVIHGIYRVVNEEAEPMKVVAETMARPLRAEVAESVAEAAVTNARNLAE

SEQ ID NO: 206

Prototheca moriformis Plastidial Acyl-Carrier Protein (ACP) allele 1

ATGGCGATGTCTATGACCTCCTGCCGCGCCGTCTGCGCCCCCTCGCGCCACCCTCCGGGTGCAGGCTCCCCGC
GTGGCTGTCCGCCCCCTTCCGCGCCAGCGCATGATCTGCCGCGCTGTGGACAAGGCTAGCGTCTGAGCGAC
GTGCGCGTGATCATTGCCGAGCAGCTGGGCACCGATGTGAGAAGGTCAACGCCGACGCCAAGTTTGCCGAC
CTGGGTGCCGACTCCCTGGACACCGTTGAGATCATGATGGCTCTGGAGGAGAAGTTTGACCTGCAGCTCGAT
65 GAGGAGGTGCGGAGAAGATCACACCGTGCAGGAGGCCGCTGACCTCATCTCCGCCAGATCGGCGCTTGA

SEQ ID NO: 207

Prototheca moriformis (UTEX 1435) Plastidial Acyl-Carrier Protein (ACP)
allele 1

MAMSMTSCRAVCAPRATLRVQAPRVAVRPFAQRMICRAVDKASVLSVVRVIAEQLGTDVEKVNADAKFAD
LGADSLDTVEIMMALEEKFDLQLDEEGAEEKITTVQEAADLISAQIGA

SEQ ID NO: 208

5 *Prototheca moriformis* (UTEX 1435) 3-hydroxyacyl-CoA dehydratase

10 ATGTGCGCTGAGGAGCGCCTACCTCACTGTTTATAATGCGTCGCTCGCGCTGGGCTGGGCGTACCTGCTGTGG
CTGAGCGTGTCTGTCCTRGCCGCTGGCGGCAGCCTGTGGGACCTGTGGAAGACTGTGGAGGTGCCTCTCAAG
GTCGTTTACAGACGGCGGCGATCGCGGAAGTTGTGCATGCCAGTGTGGCATTGTCCGCTCGCCCCACTCGTC
15 ACTGCCCTGCAAGTGGCTTCGCGAGTGTTCCTGGTCTGGGGCGTGGTGAATCTGGCTCCAGAGGTGGCGACC
GGCTCGCAGGTGGCGGCCATTCCCATCCCCGGGGTGGCGCGTGGGCTCTCCTTTGCGACCCCTGGTCATC
GCCTGGGCGCTCGAGATCATCCGCTACGGCCACTTTGCCGCCAAGGAGGCGGGCATTGCCAGCAAGCTG
CTGCTCTGGCTTCGCTACACCGGCTTCCTGGTGTCTTACCGCTGGGCGTCTCCAGCGAGCTGACCAGTATC
15 TACCTAGTGGCTCCCTACGTCAAGGAGCGCGGCATCCTGAGCCTGGAGATGCCCAACGCCGCCAACTTTGCC
TTTAGCTACTACGCCGCGCTCTGGATCGTCAGCCTCACCTACATTCCCGKMYWWCMKMGWWRKMGGGTAC
ATGCTCAAGCAGCGGAAAAAGATGCTTGGTGGTGGCGCCAAGGCCAAGAACTGGCTTGA

SEQ ID NO: 209

20 *Prototheca moriformis* (UTEX 1435) 3-hydroxyacyl-CoA dehydratase

MSLR SAYLT VYNAS LALGWAYLLWLSVSVXAAGGSLWDLWKTVEVPLKVVQTAAIAEVVHASVGIVRSPPLV
TALQVASRVFLVWGVVNLAPVATGSQVAAIPIPGVGRVGLSFATLVIAWALSEIIRYGHFAAKEAGIASKL
LLWLRYTGFLVLYPLGVSETMIYLVAPYVKERGILSLEMPNANFAFSYAAALWIVSLTYIPXXXXXXGYM
25 LKQRKMLGGGAKAKLA

SEQ ID NO: 210

25 *Prototheca moriformis* (UTEX 1435) Ketoacyl-ACP Synthase II

30 ATGCAGACCGCGCACCAGCGGCCCCCGACCGAGGGGCACTGCTTCGGTGCAGAGGTGCCCCAGGGCGTCGAGG
CGGGCGGTGCGCCGGGCATGGTCCCGCATCGCGCGCGCGGGCGGCCGCGGCCGACGCAAACCCCGCCCGCCCT
GAGCGCGCGTGGTCATCACGGGCCAGGGCGTGGTGAACAGCCTGGGCCAGACGATCGAGCAGTTTTACAGC
AGCCTGCTGGAGGGCGTGAGCGGCATCTCGCAGATCCAAAAGTTTGACACCACGGGCTACACGACGACGATC
GCGGGCGAGATCAAGTCGCTGCAGCTGGACCCGTACGTGCCAAGCGCTGGGCCAAGCGCGTGGACGACGTC
35 ATCAAGTACGTCTACATCGCGGGCAAGCAGGCGCTGGAGAACGCGGGGCTGCCGATCGAGGCGGCGGGGCTG
GCGGGCGCGGGCTGGACCCGGCGCTGTGCGGCGTGTCTATCGGCACCGCCATGGCGGGCATGACGTCTCTTC
GCGGCGGGCGTGGAGGCGCTGACGCGCGGGCGGCGTGCAGCAAGATGAACCCCTTTGTCATCCCTTCTCCATC
TCCAACATGGGCGGCGCGATGCTGGCGATGGACATCGGCTTCATGGGCCCCAACTACTCCGTCTCCACAGCC
TGCGCGACGTCCAACCTACGCATTTGTGAACGCGGCCAACCACATCCGCAAGGGCGACGCGGACGTATGGTC
40 GTGGGCGGCACCGAGGCCTCCATCGTGGCGTGGGCGTGGGCGGCTTTGTGGCCTGCCGCGCGCTGTCCACG
CGCAACGACGAGCCCAAGCGCGCGAGCCGGCCGTGGGACGAGGGCCGCGACGGCTTTGTGATGGGCGAGGGC
GCGGCCGTGCTGGTCATGGAGTCGCTGGAGCACGCGCAGAAGCGTGGCGCGACCATCCTGGGCGAGTACCTG
GGGCGGCCATGACCTGCGACGCGCACCATGACGAGACCCGACCCCGAGGGCTGGGCGTGAGCACCTGC
45 ATCCGCTGGCGCTCGAGGACGCGGCGTCTCGCCGACGAGGTCAACTACGTCAACGCGCACGCCACCTCC
ACCCTGGTGGGCGACAAGGCCGAGGTGCGCGCGGTCAAGTCGGTCTTTGGCGACATGAAGGGTATCAAGATG
AACGCCACCAAGAGTATGATCGGGCACTGCCTGGGCGCCGCCGCGGCATGGAGGCCGTGCCACGCTCATG
GCCATCCGCACCGGTGGGTGCACCCACCATCAACCACGACAACCCCATCGCCGAGGTGATGGCCTGGAC
GTCGTGCCAACGCCAAGGCCAGCACGACATCAACGTGCCATCTCAAACCTCTTCGGCTTTGGCGGGCAC
AACTCCGTGCTCGCCTTTGCGCCCTTCCGCGAGTAG

SEQ ID NO: 211

50 *Prototheca moriformis* (UTEX 1435) Ketoacyl-ACP Synthase II

55 MQTAHQRPPTGEGHCFGARLPTASRRVRRRAWRSRIARAAAAADANPARPERRVVITGQGVVTSLGQTIEQFYS
SLLEGVSGISQIQKFDTTGYTTTIIAGEIKSLQLDPYVPKRWAKRVDDVIKYVYIAGKQALENAGLP IEAAGL
AGAGLDPALCGVLIGTAMAGMTSFAAGVEALTRGGVRKMNPFICIPFSISNMGGAMLAMDIGFMGPNSVSTA
CATSNYAFVNAANHIRKGDADVMVVGTEASIVPVGLGGFVACRALSTRNDEPKRASRPWDEGRDGFVMGEG
AAVLVMESLEHAQKRATILGEYLGAMTCDAAHMTDPHPGLGVSTCIRLALEDAGVSPDEVNYVNAHATS
TLVGDKAEVRVAVKSVFGDMKGIKMNATKSMIGHCLGAAGGMEAVATLMAIRTGWVHPTINHDPNPIAEVDGLD
60 VVANAKAQHDINVAISNSFGFGGHNSVVAFAFPRE

SEQ ID NO: 212

65 *Prototheca moriformis* (UTEX 1435) sterol 14 desaturase

70 ATGGACGCCTTTGGCGAGCTCATCATCCTGACCGCCTCGCGCACGCTGCTGGGCGCGAGGTGCGCGAGTCC
ATGTTCCGCGAGGTGGCCGACCTCTTCCACGACCTGGACGACGGCATGCGCCCGCTGAGCGTGTCTTCCCC
TACCTCCGACCGCCTACCACAGGCGCGCGACGTGCGCGCGCGCGCTGCAGGACATCTTCAAGCGCGTCT
ATCGCCGCGCGCGCGCTCGGGCGTCAAGGAGGACGACGTGCTGCAGGCGCTGATCGACGCGCGCTACCGC
AAGGTCTACGACGGCCGCGCCACGACGAGGAGATCACGGGCTCCTCATCGCGTGTCTTTCGCCGGC
CAGCACACCAGTCCGTGACCTCCACCTGGACGGGYCTCTACATGATGMCSGGCGACCGAGCGCTACTTCAYG
GTGGTTCGAGGAGGAGCAGCGCGCGTGTGCGCGAGCACGGCGACGCGCTCGACTTTGACGTGCTCAACGGC
ATGGACCGCTGCACCTGGCCATCATGGAGGCCCTGCGCCTGCAGCCGCCCTCGTCTCTCATGCGCTAC

GCCAAGGAGCCCTTCGAGGTTCGTACCTCCGACGGCAAGGCCTTCACCGTGCCCAAGGGCGACATTGTGGCG
 ACCTCGCCCTCCTTCTCCACCGCCTCAAAAACGTGTTCCGGGATCCCAACGACTTCCAYCCCCGAYCGCTTT
 GAAGCGCCCCGCGACGAGGACAAGGCCGTGCCCTTCTCCTACATCGGRTTTGGCGGCGGGCGGCACGGGTGC
 ATGGGCTCGCAGTTCGCCTACCTCCAGATCAAGACCATCTGGAGCGTGCTCCTCCGCAACTTCACCTTTGAG
 CTGGTGGACCCCTTCCCGGAGCCGACTGGGCCTCCATGGTCATCGGGCCCAAGCCCTGCCGCGTGCGCTAC
 ACGCGCCGAAAGACGCCCCCTGGCCTGA

SEQ ID NO: 213

Prototheca moriformis (UTEX 1435) sterol 14 desaturase

MDAFGELIILTASRTLLGREVRESMFREVADLFHDLDDGMRPLSVLFPYLPTAYHRRRDVARRRLQDIFKRV
 IAARRASGVKEDDVLQALIDARYRKVYDGRATTDEEITGLLIALLFAGQHTSSVTSTWTXLYMMXGDQRYFX
 VVEEEQRRVVAEHGDALDFDLNMGMDRLHLAIMEALRLQPPLVLLMRYAKEPFVVTSDGKAFTVPKGDIVA
 TSPFSFHLKLVFRDPNDFXPXRFEPARDEDKAVPFSYIXFGGGRHGCMSQFAYLQIKTIWSVLLRNFTFE
 LVDPFPEPDWASMVIGPKPCRVRYTRRKTPLA

SEQ ID NO: 214

Prototheca moriformis (UTEX 1435) Ketoacyl-ACP Synthase III

ATGCGCATACTGGGCAGCGGTTCCCTCCACCCCTTCCACGGTGCTGAGCAACGCCGATCTGGAGCAGCTGGTG
 GAGACGAACGACGAGTGGATCGTTGCCCGGACTGGCATCCGGCGCCGGCACATCCTGGGCCCCGGGGGAGACG
 CTCACCCACCACTGCACCCCTGGCCTGCCAGCGCGCGCTAGAGATGGCGGGCGTGGACGCCAAGGACGTGGAC
 CTCATCCTGCTGGCCACGTCCAGCCCGGACGACTCCTTTGGCAGCGCCTGTGCTGTGCAGGCCGAGCTCGGG
 GCCAAGACGCGCGCCGCTACGACCTACCGCCGCTGCTCCGGCTTTGTCTATGAGCACGGTCACCGCCACC
 CAATTCTTGAGAACGGGGGTGTACAAGAATCCTGGTCATCGGCGCGGACGCCCTGTCCCGCTACATCGAC
 TGGCGGGACCGGAGCACGTGCATCCTGTTTGGCGACGGCGCCGGCGCTGTCTCCTCCAGCGCGACGATCGC
 GAGGGTGAGGACTCCCTTCTCGGGTTCGACATGCACTCTGATGGGCACGGCCAGAAGCACCTCCACGCCGGC
 TTCATGGGCTGCGGCAACAAGGCCTTCTCGGAGCAGCCCTCCAGCGCCGCGCGCTTTTCGGCAACATGTACATG
 AACGGCAGCGAGGTCTTCAAGTTTGGCGTGCGAGCCGTGCCGCGGGTGGTGGAGGCCTCGCTGAAAATGGCG
 GGGCTGAGCGTGTCCGACATCGACTGGCTGGTCCTGCACCAGGCCAACCACGAATTTTACGGCTGCCGCC
 GATCGGCTGGGTGTCCCCGCTGAGAAAGTGGTGTCCAATGTGGCCGAGTACGGCAACACGTCCGCCGCTCC
 ATCCCCCTCGTGCTGGACGAATCGGTCCGGCAGGGCCTGATCAAGCCGGGCGACATCCTGGGCATGGCAGGG
 TTTGGGGCTGGCCTGTCTGGCGGGGAGCCATCCTCCGATGGGGCTAG

SEQ ID NO: 215

Prototheca moriformis (UTEX 1435) Ketoacyl-ACP Synthase III

MRILGSGSSTPSTVLSNADLEQLVETNDEWIVARTGIRRRHILGPGETLTHHCTLACQRALEMAGVDAKDVD
 LILLATSSPDDSFSGACAVQAEKGAKSAAAYDLTAACSGFVMSTVTATQFLRTGVYKNILVIGADALSRYID
 WRDRSTCILFGDGAGAVLLQRDDREGEDSLLGDFMDHSDGHGQKHLHAGFMGCGNKAFSEQPSAAAFGNMYM
 NGSEVFKFAVRVPRVVEASLKMAGLSVSDIDWLVLHQANQRILTAADRLGVPAEKVVSNAEYGNSTSAAS
 IPLVLDESVRQGLIKPGDILGMAGFGAGLSWAGAILRWG

SEQ ID NO: 216

Prototheca moriformis (UTEX 1435) acyl transferase

ATGAAGGAGCAAGTGGTTATGCCCCCCCCCTGAGGCCCCCACGAATGGGGACCTTGCGGCCCCCTTTGTGCGG
 AAGGACCGGTTTCGGGAGTATGGCATGGCCCCGAGCCCCCGATGGTGAAGGTGGTGCTGGCACTCGAGGCC
 TTGGTGCTCTTGCCGCTGCGCGCGGTGAGCGTGCTCATCCTCGTCTGTACTGGCTCATTTGCAATGCT
 TCGGTCGCACTCCCGCCCAAGTACTGCGCGGCCGTACGGTCACCGCTGGGCGCTTGGTCTGCCGGTGGGCG
 CTCTTCTGCTTCGATTCCACTACATCAAGTGGGTGAACCTGGCGGGCGCGGAGGAGGGCCCCCGCCGGGC
 GGCATTGTTAGCAACCACTGCAGCTACCTGGACATCCTGCTGCACATGTCCGATTCTTCCCCGCCTTTGTG
 GCGCGCCAGTCGACGGCCAAGCTGCCCTTTATCGGCATCATCAGCCAAATTATGAGCTGCCTCTACGTGAAC
 CGCGACCGCTCGGGGCCCAACCAGTGGGTGTGGCCGACCTGGTGAAGCAGCGCATGCGGACGAGGCCGAG
 GGGAAGACCCCGCCGAGTACCGGCCGCTGCTCCTCTTCCCCGAGGGCACCACTCCAACGGCGACTACCTG
 CTTCCCTTCAAGACCGGCGCCTTCTTGGCCGGGGTGCCCGTCCAGCCCGTCTGCTCCTCCACTACCACAGGGGA
 ACCCTGTGGCCACGTGGGAGACGATTCGGGCCAAGTGGCACATCTTCTGATGCTCTGCCACCCCGCCAC
 AAAGTGACCGTGATGAAGTTGCCCGTGTACGTCCCCAATGAGGAGGAAAAGGCCGACCCCAAGCTGTACGCC
 CAAAACGTCCGCAAAGCCATGATGGAGGTGCGCGGGACCAAGGACACGACGGCGGTGTTTGAGGACAAGATG
 CGCTACCTGAACTCCCTGAAGAGAAAAGTACGGCAAGCCTGTGCCTAAGAAAATTGAGTGA

SEQ ID NO: 217

Prototheca moriformis (UTEX 1435) acyl transferase

MKEQVVMPPPEAPTNGDLAAPFVRKDRFGEYGMAPQPPMVKVVLALALVLLPLRAVSVLILVLYWLI
 SVALPPKYCAAIVTAGRLVCRWALFCFGFYIKWVNLAGAEEGPRPGGIVSNHCSYLDILLHMSDSFPAFV
 ARQSTAKLPFIGIISQIMSCLYVNRDRSGPNHVGVADLVKQRMQDEAEGKTPPEYRPLLLFPEGTTNSGDYL
 LPFKTGAFLAGVPVQPVVLHYHRGTLWPTWETIPAKWHIFLMLCHPRHKVTVMKLPVYVPEEEKADPKLYA
 QNVRKAMMEVAGTKD'TTAVFEDKMRYLNSLRKYGKPVPKKIE

SEQ ID NO: 218

Prototheca moriformis ENR1-1(Enoyl-ACP Reductase) 1 cDNA sequence;
version 1

ATGGCGTCCGCACAGCTCCTGTCCAAGTCCCAGGGTCTGGTCGGTGCCCTGAAGGACGTCAGGGCCGTCCGG
GTGACCCCTGCCCCGCGCCGGACCCCGTGCCGCCCCCTCGCGCCGATGCCTACTCGAGCTCCAGGGCGCCTGCC
TCCATGGTGTCTGGCCCCCAGCGCCGCGGGCTGAGCGTGAAGGCTGCCGCCAGCACACAGCCCTGCCGGCACC
GGCTTGCCCATTTGATCTGCGGGGCAAGAAGGCCTTTATCGCGGGCGTCGCCGACGACAGGGTTTTGGCTGG
GCCATCGCCAAGCAGCTGGCTGAGGCGGGCGCGGAGATCTCCCTGGGCGTGTGGGTGCCTGCCCTGAACATT
TTCGAGAGCAACTACCGCCGCGGCAAGTTTGACGAGAACCGCAAGCTGGCCAACGGCGGGCTGATGGAGTTT
GCGCACATCTACCCCATGGATGCGGTCTTCGACTCGCCCCGCGGACGTGCCCGAGGACATTGCCAGCAACAAG
CGCTACGCGGGGAACAAGGGCTGGACGGTGAGCGAGACGGCCGACAAGGTGCGGGCCGACGTGGGCAAGATC
GACATCCTGGTGCATCTCGTGGCCAACGGGCCCCGAGGTGCAGAAGCCGCTGCTGGAGACCAGCCGCGCGGC
TACCTCGCGCGCTGAGCGCCTCCTCCTACTCCCTCATCTCCATGGTCCAGCGCTTTGGCCCGCTCATGAAC
CCCGGCGGCGCCGTATCTCGCTGACCTACAACGCCTCCAACCTGGTCATCCCCGGCTACGGCGGCGGCATG
AGCACGGCCAAGGCCGCGCTCGAGTCCGACACGCGCGTGCTGGCCTACGAGGCCGGCCGCAAGTACCACGTG
CGCGTGAACACCATCAGCGCCGGGCGCTCGGCTCGCGCGCCGCCAAGGCCATCGGCTTCATCGACGACATG
ATCCGCTACTCGTACGAGAACGCGCCCATCCAGAAGGAGCTCAGCGCCTACGAGGTGGCGCGGCGCGCGCC
TTCTCTGCTCCCCGCTCGCCTCGGCCGTCACTGGCCACGTATGTTTGTGACAACGGCCTCAACACCATG
GGCCTCGCCCTCGACTCCAAGACCTCGACCACAGCGAGTGA

SEQ ID NO: 219

Prototheca moriformis ENR1-1(Enoyl-ACP Reductase) 1 cDNA sequence;
version 2

ATGGCGTCCGCACAGCTCCTGTCCAAGTCCCAGGGTCTGGTCGGTGCCCTGAAGGACGTCAGGGCCGTCCGG
GTGACCCCTGCCCCGCGCCGGACCCCGTGCCGCCCCCTCGCGCCGATGCCTACTCGAGCTCCAGGGCGCCTGCC
TCCATGGTGTCTGGCCCCCAGCGCCGCGGGCTGAGCGTGAAGGCTGCCGCCAGCACACAGCCCTGCCGGCACC
GGCTTGCCCATTTGATCTGCGGGGCAAGAAGGCCTTTATCGCGGGCGTCGCCGACGACAGGGTTTTGGCTGG
GCTATCGCCAAGCAGCTGGCTGAGGCGGGCGCGGAGATCTCCCTGGGCGTGTGGGTGCCTGCCCTGAACATT
TTCGAGAGCAACTACCGCCGCGGCAAGTTTGACGAGAACCGCAAGCTGGCCAACGGCGGGCTGATGGAGTTT
GCGCACATCTACCCCATGGATGCGGTCTTCGACTCGCCCCGCGGACGTGCCCGAGGACATTGCCAGCAACAAG
CGCTACGCGGGGAACAAGGGCTGGACGGTGAGCGAGACGGCCGACAAGGTGCGGGCCGACGTGGGCAAGATC
GACATCCTGGTGCATCTCGTGGCCAACGGGCCCCGAGGTGCAGAAGCCGCTGCTGGAGACCAGCCGCGCGGC
TACCTCGCGCGCTGAGCGCCTCCTCCTACTCCCTCATCTCCATGGTCCAGCGCTTTGGCCCGCTCATGAAC
CCCGGCGGCGCCGTATCTCGCTGACCTACAACCTGGTCATCCCCGGCTACGGCGGCGGCATG
AGCACGGCCAAGGCCGCGCTCGAGTCCGACACGCGCGTGCTGGCCTACGAGGCCGGCCGCAAGTACCACGTG
CGCGTGAACACCATCAGCGCCGGGCGCTCGGCTCGCGCGCCGCCAAGGCCATCGGCTTCATCGACGACATG
ATCCGCTACTCGTACGAGAACGCGCCCATCCAGAAGGAGCTCAGCGCCTACGAGGTGGCGCGGCGCGCGCC
TTCTCTGCTCCCCGCTCGCCTCGGCCGTCACTGGCCACGTATGTTTGTGACAACGGCCTCAACACCATG
GGCCTCGCCCTCGACTCCAAGACCTCGACCGCAGCGAGTGA

SEQ ID NO: 220

Prototheca moriformis ENR1-1(Enoyl-ACP Reductase) 1 Protein sequence;
version 1

MASAQLLSKSQLVLGALKDVRVRVTLPRAGPRAAPRADAYSSSRAPASMVLPAPQRRGLSVKAAASTSPAGT
GLPIDLRGKKAFIAGVADDQGFGWAIKQLAEAGAEISLGVWVPALNIFESNYRRGKFDENRKLANGGLMEF
AHIYPMDAVFDSPADVPEDIASNKRYAGNKGWTVSETADKVAADVKGIDILVHSLANGPEVQKPLLETSTRRG
YLAALSASSYSLISMVQRFGLMNPGGAVISLTYNASNLVIPGYGGGMSTAKAALES DTRVLAYEAGRKYHV
RVNTISAGPLGSRAAKAIGFIDDMIRYSYENAPIQKELSAYEVGAAAFLCSPLASAVTGHVMFVDNGLNTM
GLALDSKTLDHSE*

SEQ ID NO: 221

Prototheca moriformis ENR1-1(Enoyl-ACP Reductase) 1 Protein sequence;
version 2

MASAQLLSKSQLVLGALKDVRVRVTLPRAGPRAAPRADAYSSSRAPASMVLPAPQRRGLSVKAAASTSPAGT
GLPIDLRGKKAFIAGVADDQGFGWAIKQLAEAGAEISLGVWVPALNIFESNYRRGKFDENRKLANGGLMEF
AHIYPMDAVFDSPADVPEDIASNKRYAGNKGWTVSETADKVAADVKGIDILVHSLANGPEVQKPLLETSTRRG
YLAALSASSYSLISMVQRFGLMNPGGAVISLTYNASNLVIPGYGGGMSTAKAALES DTRVLAYEAGRKYHV
RVNTISAGPLGSRAAKAIGFIDDMIRYSYENAPIQKELSAYEVGAAAFLCSPLASAVTGHVMFVDNGLNTM
GLALDSKTLDRSE*

SEQ ID NO: 222

Prototheca moriformis Heteromeric acetyl-CoA carboxylase b-CT subunit
ATGGCACTTCCACTTCCGCCACTCCCATCACTTCCAAAACCTTCCAAAATTCAATTTCTTTTTTCTTGG
TTAGATTGGGGAAGATAAACGACTTGAGAAAAAAGAAATTGAAAATAAAATAGAGCAACAAATACCG
AATCTAATAAAGTTAAGTGCACTTAAATATAAGAGAGATGTTAAAGCTGAAAAAGATTTAGGTTTATGGACT
CGTTGTGAAAATTGTGGTGTATTATATATACATAAAACATTTACAAGTAAACAAAAAGTTTGTATAGGATGT

AATTATCATATATTAATGAGCAGTTTTGAACGAATTGCTCGCTTTCTTGACATAGAAAGTTGGACTCCATTA
 AATGAAACAATGTCTTCAGGTAATCCTTTAGGATTTAGTGATCAAAAATCATATACAGAAAGATTAATAGAG
 TCTCAAGAATTAACAGGATTACAGGATGCAGTTCAAACAGGAAGTGGCGTTATAGAAGGTATTCCTTTAGCA
 CTGGAATGATGGATTTCCATTTTATGGGTGGTAGTATGGGTTCTGTTGTTGGAGAAAAATTAACTCGTTTA
 5 ATTTGAATATGCAACAATTAAGGCCCTTTCTTTAGTAATTTTTTGTGCTTCTGGTGGTGTCTCGTATGCAAGAA
 GGTATTTTAAGTTTAATGCAAATGGCTAAAATTTCTGCTGCATTAATCGTCATCAAAATATAGCAAAACTT
 TTATATATCCCAGTTTTTAACATCACCTACTACAGGAGGAGTTACTGCAAGTTTTGCGATGCTAGGAGATTTA
 ATTTATGCAGAACCTCGTGCACTTATTGGTTTTGCTGGACGACGAGTTATTTCTCAAACACTTCAAGAACAA
 10 TTACCAAAAGATTTTCAAACAGCTGAATATTTATTACATCATGGATTAATTGATTTAATTATTTCAAGATTT
 TTCTTAAACAAGCATTAGCTGAAACAGTTATTTTATTTCAACAAGCTCCTTCAAAGAAATAGGATCTATT
 GAATTCGTGAACGTGTTACTTTAACAAGTTGAAGAAGAATTATTAAGACGTTCAATAATCGAGAATGAA
 AACTCTTTTAAAAAATGTTAGAATCGTTTTTATCTTTTATCCAGGAAGAAAAAGCAATTTCAACAGAT
 TTAGCATTTAATGAAGTTCTTGATGAAGCTTTTGATAATTCAATTAATGCTTCTATAACAAAAGACTTCAGT
 TATATGCTTTCTTCAACTAAATTAATTGAAGCTTAA

SEQ ID NO: 223

Prototheca moriformis Heteromeric acetyl-CoA carboxylase b-CT subunit
 MALPLPPLPSLPKLPKIQPFPSWLDWGDKDRLEKKKEIENKIEQQIPNLIKLSALKYKRDVKAEDLGLWT
 RCENCGVIIYIKHLQVNKKVCIGCNHILMSSFERIARFLDIESWTPNETMSSGNPLGFSQKSYTERLIE
 20 SQELTGLQDAVQTGTGVIEGIPLALGMMDFHFMGSGMSVVGKLRILIEYATIKGLSLVIFCASGGARMQE
 GILSLMQMAKISAAALNRHQNIAKLLYIPVLTSPPTGGVTASFAMLDLIYAEPRALIGFAGRRVISQTLQEQ
 LPKDFQTAEYLLHHGLIDLIIISRFLLKQALAEVILFQQAPSKEIGSIEFRERVTLTKVEEELLRRSIIENE
 NSFKKMLESFLSFIPGRKKAISTDLAFNEVLDEAFDNLNASITKDFSYMLSSTKLIEA*

SEQ ID NO: 224

Prototheca moriformis Glycerol-3-phosphate acyltransferase
 (GPAT) nucleotide

TCGGCACCTGGATGAGGGCCTGGACGACATGCCAGCCCATGTCAAGGTTGCTGCCCCGTTGGGCATGACAGTGC
 AAGGGAAATAGGTCATCTTTCACACGCTCACTATAATCAATGAAACATCTACTTTCATCATATTGCAGTCTT
 CTTTGATGGGAGTCGGTCTAACAATCCCCATCAGTCACGGCCGGCTGGCACTAGGTACCTGGCAAGGAGTCT
 ACTTGAATGAGCACAGGGATTGCGGGGGCGCACGCCATCTGATTGTGACGCTGCAGGGGTGCTCTTCCTGAT
 GAGGGTTGGGCCCCATGGCAGCTCGGCGCAATCTGGATCTGATCCCATGCTGGGGCTTTAAACTGCGTTCC
 35 GACATGCGACAGGGACAATGCGCGCCCCACTGACGGCCTTTCATAGCCGTGAGCATAAACAAGATAACCACT
 TGACAGAGTTGCGGGTTGCAAGGGGCTCAAGGCCTCTGGGGGCGCGCCGCTCTTCGCCATGCAAACTTGGC
 TAGCACCCGCACCGTGCCCGCGGCGTGGGACCCGATCGGGCCGCTCATGGGGCGATCAGATATCATGCCCA
 GCGCATCAATCAAGTTTTGCGGCTTCAAGGAGGGGCGCCCGGGCTCTGGGCCCCCCCCGGGCTGTGGCGTC
 TGCCAAGACGTCCCCTGGGGAGCTGCGCCTGCCCTTTAGCCACGTCCGCTCGGAGGAAGAGCTGTTGCGCAT
 40 TTTGCGGTCCGGCGTGGCGCATGGCCAGCTGCCCGCGTGCATTCTGAACCGCGCCTCGGAGCTCATGGAGTC
 GTACACGGCATCCATGAAGCCGGCGGCGTTAGCGACCCGAGCGCTTACACGGTCAAGATAATGGCGGAGCA
 CTTTGGTCTGTCCATGCTGCAATGGCCCATCGGCACAAGTTTGATTGCGCGCACAAAGCGGGTGCTGGAGCC
 GTACGACTACTACTCTTTGGCCAGCGCTACATCCAGGGCCTCATCGACATGCAGCGCTCCTGCCTGGGCCA
 45 CGTGGACCGCTTCGCCGAGATCCAAGCGCAGCTGGACAGAGGCGAGAACGTGGTGCTGCTCGCAAACCA
 GACGGAGGCCGACCCCGCGGTGTGGGCGCTGCTGCTGGAGTCGCGCTTCCCGAAGCTGGCGACGGAGGTGGC
 GTACGTGGCGGGCGACCGCGTGATCACCGACCCGCTCTGCATCCCCTTTTTCGCTGGGCGCAACCTCTACTG
 CGTGCACTCCAAGAAGCACCTGGACGACGTCCCGGAGCTCAAGGCGGAGAAGGCCGCGACAAACCGCCAGAC
 CCTGCGCGTCATGGCCCGGACCTGGCCCGGGCGGGGTGCTCCTCTGGGTTGCGCCCTCGGGCGCGGGA
 50 CCGCGCGGTCTGCCCAGACGCGGCGAGACCTGCCCAGACCCCTTTGACCCCGCGGTGCTGGAGCTCATGCG
 GGCCCTGCTCGCCAAGGCCGGCCGGCCGGGGCACCTCTGGCCCTGGCCATGTACTCTGCCAGGGTGATGCC
 CCCGCCGACGACCATTTGAGAAGGACATTGGCGAACGGCGGGTCTGTCGGCTTTGCCGGCTGCGGTCTCTCCC
 CGGACCCGAACCTCGATGTGGCAGCCCTGGTGCCTCCCAACGTGACGGACCGTGCGCAGCGGCAGCAGCTGCT
 CTCGGATGCGGCGTTTGGTCCATCTCTCACAGGTACGACCGCTGTTGAAAGCGGTGCATGGGCGAGCTGAG
 TGGTGCGGAAGCCGAGGCGTACACCCAGCTCAGATCCTAGTTGCTGCAAGTTTCCCTCGCGCTGGG
 55 CGGCGGTGTGCCACCCAAGCGTTGGGGACCTGGGTTTTCCAGGCTTGGCCTCCACCGCGGCTCAAGACTGG
 CTCTCGACCTCGGCGACCCCGGCTTTCCCTGTGCTCCAATGTGTGCGGCACTCTGAGCGACGACCAACCC
 AGTGATGCCTCAGTTTTTCACTAGCTCTTGCGCATAGATAGCCCCCTCTGAAACCTGTTTCATGCCTTCAACC
 TCTTCCAAGAGTCACCTGCAACCTTTTTTGGCACCAAAAAAA

SEQ ID NO: 225

Prototheca moriformis Glycerol-3-phosphate acyltransferase (GPAT),
 protein

MQTLASTRTVPAASGTDRAAHGAIRYHAQRIQSRFAASRRGARALGPPRAVASAKTSPGELRLPFSHVRSEE
 65 ELFAILRSGVAHGQLPASILNRASELMESYTASMKAGGVSDPSAYTVKIMAEHFGLSMLQMAHRHKFDSAHK
 RVLEPYDYYSFGQRYIQGLIDMQRSCLGHVDRFAEIQQLDRGENVLLANHQTEADPAVWALLLESRFPKL
 ATEVAYVAGDRVITDPLCIPFSLGRNLYCVHSHKHLDDVPELKA EKAATNRQTLRVMARDLAAGGVLLWVAP
 SGGDRRAVCPSTGETLPDFDPAVVELMRALLAKAGRPGLHWPLAMYSARVMPPPTTIEKDIGERRVVGFA
 70 CGLSPGPELDVAALVPPNVTDRARQQLSDAAFESISHEYDRLLKAVHGQLSGAEAEAYTQPKLES*

SEQ ID NO: 226

Prototheca moriformis PAP, nucleotide

5 GGCCGCGCTGAAGCCCTCCAAGGGCAGGCTGGTCCCCGTGGCCAACCCGGCCTGGAGCGCGCTCATCAACTC
CCACTGGCACCAGTACCTTGCCATGGTTCTGGGCTACCTCGTGTGCCTCTACAGCGACCGGGCCACGCCCTT
TCGCCGCGCCATCTACCACGAGACCGACGCCGAGCTCTGGAGGTATAGCTACCCGCTGCTCCCGAACCAGGT
CCCCGCCAGGCGGTGGTGCCTCTGCTTACCGCCCCCTTTGCCGTATCCTGCCCTTTACTTTGCGGG
CCGCATCTCGCGCATCGTGGCGCACACAGGCCCTGCTCCAGGCCCTTTGCGGCCGTCTGTATCACTGGCCTTAT
10 CACCAACCTCATCAAGCTGAATGTGGCGCGTCCCCGGCCGACTTTGTGGCGCGGTGCTGGCCCCACGGCGC
CGCGCGCGCCTTACCGACGTGGGCGTGCCATCTGCGCGCCCGACGCGATCGACTGGGTGGAGGGCCTCAA
GTCTTTCCTCCGGGCACACCTCCTGGGCCGCTCGGGCATGGCCTACCTCACGCTCTGGCTGGTGGGCGC
GCTGCGCGTTTTCTCCGGCGCGCGCGGCCGACCAGCATCGCGCCGCGCTGTGCCCGCTGTGCCTGGCCGC
CTGGATCGGCATGACGCGCATCCAGGACCGGTGGCACACACGGAGGACGTCATGGTCGGCTTCAGCCTGGG
15 CAGCACGATTGCCTACCTCTGCTACCGCGCGGTGCACTGCCCCGTCAACGGACCGCACGCGGGCGCGCTCAC
CGCGCTGGCCCTGCCGGACGAGGGGCGCGGCCGCGGGCCAGCATCAGCCGCGAGCCGAGCTTTGCGCAGAT
GGCGGCCGCGCGCGCGGTGGGCGTGCCCTGCAGCCCGTGCCTACAACCCCTCGTGGAAACACGCCCGA
AGAGACCGTCTAGGACGGGGGGTGGGGCGAGAGGATGTCAAGGCTGCTGTGGTGTGCTGTTTCTTTATGCC
CTGGCCAGTCTGCCAGTGCATGCACCACAATTTTGACGACTGTACCGCTAAACACGAGCTCCCCCTGCCAGG
20 TGACCTCGATGCGTGTGGCAAAAGAAACGTGTATTTGTGCTTCCCGCCGTGCCGTGCGCGGGCAGGTTT
TTCTACCGCCTCCTTCTGGAATGTGTGGGCCGCTCGATCGAGGGCAGTGGCCCCCTGCCATTCTTTATTT
TTAGAATTTGATCCTCAG

SEQ ID NO: 227

Prototheca moriformis PAP, protein

25 MVLGYLVCLYSDRATPFRRAIYHETDAELWRYSYPLLPNQVPAQAVVLLCFTAPFAVILPFYFAGRISRIVA
HHGLLQAFAAVITGLITNLIKLNVARPRPDFVARCWPFGAARAFTDVGVPICAPDAIDWVEGLKSFPSGHT
SWAASGMAYLTWLVLGALRVFSGAARPTSIAAALCPLCLAAWIGMTRIQRWHHTEDVMVGFSLGSTIAYLC
YRAVHCPVNGPHAGALTALALPDEGPGAGPSISREPSFAQMAAAGVGVPLQPVRYNPLVETTPETV*

30 SEQ ID NO: 228

Prototheca moriformis DGAT1-1, nucleotide

35 CTGGAGCTGTGCTGGCCGCTGCTGGCGCTGCTGGCGCTGGCCGACGCCGCGGCCAGCGAGTGGGTCTCTTC
GCGCTCAACCTGCTCAACACGAGCGCGATCCTGGGCGTCCCCGCGGCCGTATCCACTACACACACTCTGAG
CTCTTGCCCGGCTTTGCGCTCTGCACCGTCTCGGTATCCTCTGGCTCAAGACCGTCTCCTATGCGCACTGC
AACTGGGACCTGCGCATGGCGCACCCGCGCGGGCGAGCTGCGCGACGGCGAGCGCGGACGCGCTGGGTCCCG
CGGACTGCGCCGCGCAGCTCAAGTACCCGGAGAACCTCACGCTCCGGAACCTGGGATATTTCTCGCGGCC
CCCACGCTGTGCTACCAGCTGAGCTACCCGCGGTGCGAGCGGTTCCGGCTCAAGTGGCTGCTGCGCCGCGTG
40 CTCATGTTTGCCCTCACCATCACGCTGATGATGTTTCATGATCGAGCAGTACGTGAACCCGCTCATCCACAAC
AGCCTGCAGCCGATGATGAGCATGGACTGGCTGCGCCTGTTTCGAGCGCGTGCTCAAGCTGTGCTGCCAAC
CTCTACCTCTGGCTGCTCATGTTCTACGCGGGCTTCCACCTCTGGCTCAACATCGCCGCGGAGCTGACCATG
TTTGGCGACCGGAGTTCTACAAGGAGTGGTGGAACGCGACCAACCATCGGCGAGTACTGGCGCCTCTGGAAC
CAGCCCGTGACCACTGGATGCTGCGCCACTGCTACTTCCCCTGCGTGCGCCACGGCGTGCCCAAGGTCTGG
45 GCAGGCATCGTCTTCTTCTCGTCAGTGCCGTCTTCCACGAGTGGATCGTCGCGCTGCCCCCTGCACATGGTC
AAGGGCTGGGCC

SEQ ID NO: 229

Prototheca moriformis DGAT1-1, protein

50 MELCWPLALLALLADAAASEWVVFALNLLNTSAILGVPAAVIHYTHSELLPGFALCTVSVILWLKTVSYAHC
NWDLRMAHRAGELRDGERGSAWVPRDCAAQLKYPENLTLRLNGYFLAAPTLCYQLSYPRSERFRLKWLRLRV
LMFALTITLMMFMIEQYVNPLIHNSLQPMMSMDWRLRFLERVLKLSLPNLYLWLLMFYAGFHLWLNIAAELTM
FGDREFYKEWWNATTIGEYWRLLWNQPVHQWMLRHCFPCVRHGVKPVWAGIVVFFVSAVFHEWIVALPLHMV
55 KGWA

SEQ ID NO: 230

Prototheca moriformis DGAT1-2, nucleotide

60 CTGGAGCTGTGCTGGCCGCTGCTGGCGCTGCTGGCGCTGGCCGACGCCGCGGCCAGCGAGTGGGTCTCTTC
GCTTTGAACCTGGTCAACACGAGCGCGATCCTGGGCGTCCCCGCGCGGTATCCACTACACGCACTCTGAG
CTGCTGCCCGGCTTTGCGCTCTGCACCGTGTGCGTATCCTCTGGCTCAAGACCGTGTCTACGCGCACTGC
AACTGGGACCTGCGCATGGCGCACCCGCGGGCGAGCTGCGCGACGGCGAGCGCGACAGCGCCTGGGTCCCG
CGGACTGCACCGCGCAGCTCAAGTACCCGGAGAACCTCACGCTGCGCAACCTGGGCTACTTTTTGGCGGGC
65 CCCACGCTGTGCTACCAGCTGAGCTACCCGCGGTGCGAGCGGTTCCGGCTCAAGTGGCTGCTGCGCCGCGTG
CTCATGTTTGCGCTCACCATCACGCTGATGATGTTTCATGATCGAGCAGTACGTGAACCCGCTGATCCACAAC
AGCCTGCAGCCGATGATGAGCATGGACTGGCTGCGCCTCTTCGAGCGCGTGCTCAAGCTGTGCTGGCCAAC
CTCTACCTCTGGCTGCTCATGTTCTACGCGGGCTTCCACCTCTGGCTCAACATCGCCGCGGAGCTGACCATG
TTTGGCGACCGGAGTTCTACAAGGAGTGGTGGAACGCGACCAACCGTCCGCGAGTACTGGCGCCTCTGGAAC
70 CAGCCCGTGACCACTGGATGCTGCGACACTGCTACTTCCCCTGCGTGCGCCACGGCGTGCCCAAGGTCTGG
GCCGGCATCGTCTTCTTCTCGTCAGCGCCGTCTTCCACGAGTGGATCGTCGCGCTGCCCCCTGCACATGGTC
AAGGGCTGGGCC

SEQ ID NO: 231

Prototheca moriformis DGAT1-2, protein

5 MELCWPLLLALLALADAAASEWVVFALNLVNTSAILGVPAAVIHYTHSELLPGFALCTVSVILWLKTVSYAHC
 NWDLRMAHRAGELRDGERDSAWVPRDCTAQLKYPENLTLRNLGYFLAAPTLCYQLSYPRSERFRLKWLLRRV
 LMFALTTITLMMFMIEQYVNPLIHNSLQPMMSMDWLRFLFERVLKLSLPNLYLWLLMFYAGFHLWLNIAAELTM
 FGDREFYKEWWNATTVEYWRWNQPVHQWMLRHCFPCVRHGVKPVWAGIVVFFVSAVFHEWIVALPLHVMV
 KGWA

SEQ ID NO: 232

Prototheca moriformis DGAT2, nucleotide

15 ATGCCCCGATTACGAAATATTGCTACTGCGTTGCTGATCGGGCCCCGCTGCCCGTGCCTGCACCATGGGGCAC
 GCCATTGACGTTTCGACCCGTTGAACCGGACACACATTTTTTAAATTTAATTCATCTCAGTAGTCGTGCAAT
 GACGGAGGCCGTCTCGGACAGGGATGCTCTGCAACCGGAGCTGGGCTTCGTGCAGAAGCTGGGCATGTACTT
 TGCGCTGCCAGTTTGGATGATAGGCCTCGTGGTGGGCGCCCTCTGGCTTCCCGTGACCGTGGTGTGCCTCTT
 CATCTTCCCAAAGCTGGCGACATTGAGTCTGGGGCTGCTGCTGGTGGCGATGCTCACCCCCCTGTCTGCTGCC
 CTGCCCCAAGCCGCTGGCCCCGCTTCTGGCCTACTGCACCACCGCCGCGGCCGAGTACTACCCCGTGCCTT
 20 CATCTACGAGGACAAGGAGGAGATGGAGGCCACCAAGGGCCCCGTATCATCGGGTACGAGCCCCACAGCGT
 CATGCCTCAGGCCATCTCCATGTTTGGCGAGTACCCGCACCCCGCCGTGGTTCGGCCCCGCTGCGCAAGGCGCG
 CGTGCTGGCATCCAGCACCGGCTTCTGGACGCCGGGCATGCGGCACCTGTGGTGGTGGTGGGCACGCGCCC
 CGTGAGCAAGCCCTCCTTCTCGCCAGCTGCGCAAGCAGCGCTCCGTGCGCCTCTGCCCCGGGCGCGTGCAG
 GGAGTGCCTTACATCGGCCACGGCAAGGAGTCTGCTACCTCCGCAAACGCTTTGGATTGTCAAGTACGA
 25 GTTGTGTGTCAGCAAAGGGGGGTGGAGAGATGCTCTCCAATCTCACCATCCAATCTACTCTTCTCATGAC
 TCCCGTTTCATGTACCGTACGCGCAGACTGGCCATCCAGACCGGCACCCCGCTGGTCCCGGTCTTCGCCTTT
 GGCCAGACGGAGACCTACACCTTTGTGCGGCCCTTCATCGACTGGGAGACGCGCCTGCTGCCGCGGTCCAAG
 TACTTTTCGCTGGTGCGCCGCATGGGCTACGTGCCCATGATCTTCTTCGGCCACCTGGGCACGGCCATGCCC
 AAGCGCGTGCCCATCCACATCGTCATCGGCCGGCCCATCGAGGTGCCGCAGCAGGACGAGCCCCGACCCGGCC
 30 ACGGTGCAGCAGTACCTCGACAAGTTTCATCGACGCCATGCAGGCAATGTTTGAGAAGCACAAAGCCGACGCC
 GGCTACCCCAACCTGACGCTCGAGATCCACTGA

SEQ ID NO: 233

Prototheca moriformis DGAT2, protein

35 MPRLRNIAALLIGPAARACTMGHAIDVRPVEPDTHFLNLIHLSSRRNDGGRLGQGC SATGAGLRAEAGHVL
 CAASLDDRPRGGRPLASRDGRVPLHLPKAGDIESGAAAGGDAHPPVAALPQAAGPLPGLLHRRGRVLPRL
 40 HLRGQGGDGGHQGPRHHRVRAPQRHASGHLHVC RVPAPRRGRPAQAQGARAGIQHRLLDAGHAAPVVVAGHAP
 REQALLPRPAAQAALRRPLPGRRAGVPLHGPRQGGRLPPQTLWICQVRVCVQQRGVERCPPI SPIQSTLPHD
 SRSCHRTRRLAIQTGTPLVPVFAFGQTETTYTFVRPFIDWETRLLR SKYFSLVRRMGYVPMIFFGHLGTAMP
 KRVPIHIVIGRPIEVPQQDEPD PATVQQYLDKFIDAMQAMFEKHKADAGYPNLTLLEIH*

SEQ ID NO: 234

Prototheca moriformis KASIII, nucleotide

45 ATGTGCACAGCTCGGGCAATGGCCGCGGGGCCGCCCTCTCGGTGTTTTGCTCCCAGCTCAACCCAGGCAGCC
 TGCGTCCGCCCGGTTAACCCCTCGCGAAGGGTCTGTGCCCGGGCGGGCATGCGCATGCTGGGCAGCGGTTCC
 50 TCCACCCCTTCCACGGTGTGAGCAACGCCGATCTGGAGCAGCTGGTGGAGACGAACGACGAGTGGATCGTT
 GCCCGGACTGGCATCCGGCGCCGGCACATCCTGGGCCCTGGGGAGACGCTCACCCACCACTGCACCCCTGGCC
 TGCCAGCGCGCGCTAGAGATGGCGGGCGTGGACGCCAAGGACGTGGACCTCATCCTGCTGGCCACGTCCAGC
 CCGGACGACTCCTTTGGCAGCGCCTGTGCTGTGCAGGCCGAGCTCGGGGCCAAGAGCGCGGCCGCGCTACGAC
 55 CTCACCGCCGCTGTCTCCGGCTTTGTGTCATGAGCAGGTACCCGCCACCCAGTTCTTGAGAACGGGGGTGTAC
 AAGAACATCCTGGTTCATCGGCGCGGACGCCCTGTCCCGCTACATCGACTGGCGGGACCGGAGCACCTGCATC
 CTGTTTGGCGACGGCGCCGGCGCCGTCTCTCCAGCGCGACGATCGCGAGGGCGAGGACTCCCTTCTCGGG
 TTCGACATGCACTCTGGTACGCTGCAATGGTGAAGCGGGGCATCCATCCAACCCCTCCACACACCGNNNNN
 NNNNNCGCGGCCGCTACGACCTCACCGCCGCTGTCTCCGGCTTTGTGTCATGAGCAGGGTACCGCCACCCAA
 60 TTCTTGAGAACGGGGGTGTACAAGAATCTTGGTTCATCGGCGGGACGCCCTGTCCCGCTACATCGACTGG
 CGGACCGGAGCACGTGCATCCTGTTTGGCGACGGCGCCGGCGCTGTCTCTCCAGCGCGACGATCGCGAG
 GGTGAGGACTCCCTTCTCGGGTTCGACATGCACTCTGGTACGCTGCAATGGTGGGGGGTGGACGAAGAGGGG
 ACGAATTGGAGCCATCAAGCTGCCCATGCGTGTGTACTGTGCTCTTCCACGGGCCACACCTAGCATCCATC
 CAACCTTTCCACCCACGCTCTTCAGATGGGCACGGCCAGAAGCACCTCCACGCCGCTTCATGGGCTGCGGC
 65 AACAAAGGCCCTTCTCGGAGCAGCCCTCCAGCGCCGCCGCTTTCGGCAACATGTACATGAACGGCAGCGAGGTC
 TTCAAGTTTGGCGTGGAGCCGTGCCGCGGGTGGTGGAGGCTCGCTGAAAAAGGCGGGGCTGAGCGTGTCC
 GACATCGACTGGCTGGTCTGTCACAGGCTCAACCAAGCAATTTTACGGCTGCCGCCGATCGGCTGGGTGTC
 CCCGCTGGTGGCGACGCATGCGGTGCTGCGGCTGGGGATGCGCTGTTTCGACTGTGTGTACAGCGCCATTTT
 GCGGACCCGGGCGCTCTGTGTGGTGGGATGCTCAACGCTCAGTCCCGATTTCTTAAGAAAGTGGTGTCTAAT
 70 GTGGCCGAGTACGGCAACACCTCGGCCGCGTCCATCCCCCTCGTGCTGGACGAATCGGTCCGGCAGGGCCTG
 ATCAAGCCGGGCGACATCCTGGGCATGGCAGGGTTTGGGGCTGGCCTGTCTGGGCGGGAGCCATCTCCGA
 TGGGGC

SEQ ID NO: 235

Prototheca moriformis KASIII, protein

5 MLGSGSSSTPSTVLSNADLEQLVETNDEWIVARTGIRRRHILGPGETLTHHCTLACQRALEMAGVDAKDVDLI
LLATSSPDDSFSGSACAVQAELEGAKSAAAYDLTAACSGFVMSTVTATQFLRTGVYKNILVIGADALSRYIDWR
DRSTCILFGDGAGAVLLQRDDREGEDSLLGDFMHSDELEPSSPCVLCCALPWATPSIHPTLPHT????AAA
YDLTAACSGFVMSTVTATQFLRTGVYKNILVIGADALSRYIDWRDRSTCILFGDGAGAVLLQRDDREGEDSL
10 LGDFMHSDDHGQKHLHAGFMGCGNKAFSEQPSAAAFGNMYMNGSEVFKFAVRVPRVVEASLKKAGLSVSD
IDWLVLHQANQRILTAADRLGVPAEKVVSNAEYGN TSAAS I PLVLDESVRQGLIKPGDILGMAGFGAGLS
WAGAILRWG

SEQ ID NO: 236

15 Prototheca moriformis Choline Kinase, nucleotide

ATGGCGGGCGTCCACACAGCTCTCTGGAGCTCGATCCGCTGTGAGCAGAGCCGAGCTCGAGGTGGGGATC
AGGGCGGTCTGCCCGCATCTGCTGCCGGGCTGGAAGGGCCTTGCCGACGATGCGATAGAGATACCACCATT
20 ACGGGCGGGATCAGCAATGCGCTGTACAAGGCGACCCGACCGTGGACGGGGACCTGGGACCGGTCTCTGC
CGGGTGTACGGCCTCAACACCGATCACTTTATCGAGCGGAAGAAGGAGGTGGCCATCATGCAGATTGTGAAC
GAGCACGATTTGGCACCGAGGTGCTGGGGACGTTTGCAGCGGGCCGGCTGGAGGCCTTTTTGCCTCAAAAG
TGCTTGGAGCCGGAGCARCTCAAGGACCCCCAGATCTCRGCGGCCATTGCCCGGGCCATGGCCCGTTTCCAC
GCCGTCCCRGGGCGGGCCCCAAGACGGCGTCCACGCCGTTTGGCCGCATCCACAAGTGGCTGGACATGGCG
25 AGCGGTTTACCTGGGACGACCCGCGAGAAGCGGGAGGGGTTTGAGGCGTTTGACCTCTCGGAGCTGCGGCRG
CAGGTGGAGGAGGTGGAGGCGCTCTGCRGCGCRGTGSGSGGCCCATCGTCTTYGGGCACAACGATCTCCTG
GCCGGCAACGTGATGGCCTCGGAGGCCTTTCTGCAGCGCGCGCAGCGCGGCGAGCGGCCGGACCCGGACKCG
GACGTCCGCGCGAGCGCCGAGAGCGACCTGACCTTTATCGACTTTGAGTATGCCGACTGGACGCCGCGAGGA
TTCGACTGGGGCAACCACCTTTAACGAGTGATTTGCGGGCTTCGACGGCAACTACGACAACCTACCAACCCCC
30 GCCGAGGCGGCCGTGTTTGTGCGGGCTACATGGCGGCGGAGCGCGGCTGCGAGGRRAGCGCGGTCCCGGAC
GCGGACGTGGACCGSGCGGTGRTCGAGGCGGAYCTGTTTGCCTGGCYAGCCACCAKTA CTGGGGCGCCTGG
TGCTTCYTGCAGGCGCARTGGTCMAARCTK

SEQ ID NO: 237

35 Prototheca moriformis Choline Kinase, protein

MAGVPHSSLELDPLSSRAELEVGIRAVCRDLLPGWKGLADDAIEITITGGISNALYKATPTVDGDLGPVLC
RVYGLNTDHFIERKKEVAIMQIVNEHGFGEVLGTGATGRLEAFLPQKCLEPEQLKDPQISAAIARAMARFH
AVPGPGPKTASTPFGRIRHKWLDMASGFTWDDPQKREGFEAFDLSELR?QVEEVEALC?AVGGPIVFGHNDLL
40 AGNVMASEAFLQRAQRGERPDPD?DVRASAESDLTFIDFEYADWTPRGFDWGNHFNE*FAGFDGNYDNYPTP
AEA AVFVRAYMAAERGCE?SAVPDADVDRV?EADLFALASH?YWGAWCFLQAQWSKL

SEQ ID NO: 238

Prototheca moriformis Pyruvate Dehydrogenase, Alpha subunit, nucleotide

45 ATGTCGTCAGTCGTGATCAGCTCCAGGACGGGCTGCCTCCCCCGCGCCGCGGGCGGGGCCGGCCCGCCCCAG
CTCGCCTGCTCGTTCCGGCCGGGGTCCCACCCGCCCGGGCGGGCGCCCCCGGGCCCCCGCCCGCGCGCGGC
GTGGCCCCGCTGCTCGCGTGGCTACCCCGCTCAAGGATGCGCCCGCGCAGCCCTCCAGCAAGCCCTCGATC
AGCGCCGAGGTGGCCAAGGAGCTCTACCGGGACATGGTGCTTGGTCGGGAGTTTGAGGAGATGTGCGCGCAG
50 ATGTACTACCGCGGCAAGATGTTTGGATTGCTGCACCTCTACTCGGGCCAGGAGGCCGTGTCGACAGGCGTG
ATCAAGGGCTGCCTTCGCAAGGACGACTACATCTGCTCCACGTACCGGGACCACGTGCACGCGCTGAGCAAG
GGCGTGGCGGCGCGCAAGGTGATGGCGGAGCTGTTTGGCAAGCGTACGGGCGTGTGCCGCGGCCAGGGTGGG
TCCATGCACATGTTTGACGCGGAGCACGGGCTGCTGGGCGGGTACGCCTTCATCGGCGAGGGTATCCCCGTG
GGCCTGGGCGCGGCCCTTCTCGGTGGCCTACCGCGGCAACGTGCTCGGCGAGGACGGCGCGGACAGGTGTCT
55 GTCAACTTCTTCGGGGACGGCACCTGCAACGTGGGCGAGTTCTACGAGTCGCTCAACATGAGCCAGCAGCTGTAC
AAGCTGCCCTGCATCTTTGTGGTGGAGAACAACCTGTGGGCCATCGGCATGAACCACCCGCGGTCCACGGGG
CCCACGCTGGGCGACCAGGAGCCCTGGATCTACAAGAAGGGACCCGCCTTTGGCATGCCGGGCGTGCGCGTG
GACGGCATGGACGTGCTCAAGGTGCGGGAGGTGCGCGTGGAGGCCGTGGAGCGCGCGCGCCGCGGCGAGGGG
CCCACGCTGATCGAGGCCGAGACCTACCGCTTCCGCGGGCACTCCCTGGCCGACCCGGACGAGCTGCGCAGC
AAGGAGGAGAAGGCCAAGTACGCGGCGCGCGACCCCATCCCGCAGCTCAAGCGCTTCATGCTCGACAAGGGA
60 CTGGCCACCGAGGAGCTGCGCGACCTCGAGGCCGGCGTCCGCCCGAGGTGGAGGAGTCGGTCACCTTT
GCCGAGGCCAGCCCCAAGCCCGACATGTCCAGCTCCTGGAAAACGTGTTTGGCGACCCCAAGGGC

SEQ ID NO: 239

65 Prototheca moriformis Pyruvate Dehydrogenase, Alpha subunit, protein

MSSVVISSRTGCLPRAAAGPGRPQLACSFPGSHPPRAGAPGPRPRRGVAPLSAVATPLKDAPAQPSSKPSI
SAEVAKELYRDMVLGREFEEMCAQMYRGMFGFVHLYSGQEAVSTGVIKGCLRKDDYICSTYRDHVHALSK
GVAARKVMAELFGKRTGVCRGQGGSMHMFDAEHLLGGYAFIGEGIPVGLGAAFSVAYRRNVLGEDGADQVS
VNFFGDGTCNVGQFYESLNMA SLYKLP C I FVVENNLWAI GMNHP RSTGPTLGDQEPWIYKKGPAFGMPGVRV
70 DGMDVLKVREVALEAVERARRGEGPTLIEAETYRFRGHSLADPDELR SKEEKAKYAARDPIQLKRFMLDKG
LATEEELRDLEAGVAAEVEESVTF AEASPKPDMSQLLENVFADPKG*

SEQ ID NO: 240

Prototheca moriformis Pyruvate Dehydrogenase, Beta subunit, nucleotide

5 ATGCAGAGCGCGCTCAAAACCGTGACCAACCCCGTGTGCGGGGTCTCGCCCCGCCCCGCCCCGGTTCGGGTCC
 CGGGTGCCCCGCGCTCCATGTGGCCCCGCGCCAGGTCTTCCGAGGTGGGCCCCCTGCGTGCCAAGTCGGCCAAG
 AAGGAGATGATGATGTGGGAGGCCCTGCGCGAGGGCCTGGACGAGGAGATGGAGCGCGACCCCCAAAGTCTGC
 CTCATCGGCGAGGACGTGGGTCACTACGGCGGTTCTGTACAAGGTGTCTGACGGCCTGCACAAAAAGTACGGC
 10 GACATGCGCCTGCTGGACACGCCCATCTGCGAGAACGGCTTCTTGGGCATGGGCGTGGGCGCGGCGATGACG
 GGCTTGCAGCGCCGTGGTGGAGGGCATGAACATGGGCTTCTGCTGCTGGCCTTCAACCAGATCAGCAACAAC
 TGCGGCATGCTGCACTACACGAGCGCGGGCAGTTCAGGTGCCCATGGTCTGTGCGCGGGCCCCGGCGGCGTG
 GGGCGCCAGCTGGGCGCCGAGCACTCGCAGCGGCTGGAGTCTTCTCCAGTCCATCCCCGGCGTGCAGCTG
 GTGGCGGTGTCCACGCCCAAGAACGCCAAGGCCCTCTTCAAGGCCGCCATCCGCTCGGACAACCCCATCATC
 15 TTCTTCGAACACGTGCTCCTGTACAACATCAAGGGCGAGGTGGAGGGCCCCGACCACGTGCAGAGCCTGGAG
 CGCGCCGAGGTCTGTGCGCCCGGGCGCGGACGTGACGGTGCTCTGCTACTCGCGCATGCGCTACGTGGCCATG
 CAGGCCGTGCAGCAGCTGGAGCGCGAGGGCTACGACCCCGAGGTTCATCGACCTCATCTCCCTCAAGCCCTTC
 GACATGGAGACCATCGCGCGCTCCGTGCGCAAGACGCGGCGCGTTCATCATCGTGGAGGAGTGCATGAAGACC
 GGCGGCATCGCGCCCTCGCTCTCCGCCGTTCATCCACAGTCCCTCTTCGACGACCTCGACCAAGGTCATC
 20 CGCCTCTCCTCCCAGGAGCTGCCCCACCGCCTACGCCTACGAAGTGGAGGCAGCAACCATCGTCCAGCCCCGAA
 AAGGTCATCGCCCGCGTGAAAAAGGTCTGTCGGCAGCCCCCTCCAAGGCCCTCAGCTACGCGTGA

SEQ ID NO: 241

Prototheca moriformis Pyruvate Dehydrogenase, Beta subunit, protein

25 MQSALKTVTNPVSGVSPRPAPVGSRVPGVHVARRQVFRGGPLRAKSAKKEMMMWEALREGLDEEMERDPKVC
 LIGEDVGHYGGSYKVSYGLHKKYGDMLLDTPICENGFLGMGVGAAMTGLRPVVEGMNMGFLLAFNQISNN
 CGMLHYTSGGQFKVPMVVRGPGGVGRQLGAHSQRLESYFQSIIPGVQLVAVSTPKNAKALFKAAIRSDNP II
 FFEHVLLYNIKGEVEGPDHVQSLERAEEVVRPGADVTVLCSYRMRYVAMQAVQQLEREGYDPEVIDLISLKPF
 30 DMETIARSVRKTRRVIIVEECMKTGGIGASLSAVIHESLFDDLDHEVIRLSSQDVPTAYAYELEAATIVQPE
 KVIAAVKKVVGSPSKALSYA*

SEQ ID NO: 242

Prototheca moriformis Pyruvate Dehydrogenase, DLAT E2 subunit, nucleotide

35 ATGCCGCGCTCTCGCCACGATGTCCAGGGCAACCTGGTGGAGTGGACGGTGAAGGAGGGGCAGGAGGTC
 GCGCCCGCGACTCCCTGGCCGAGGTGAGACGGACAAGGCCACCATGACGTGGGAGAATCAGGATGACGGC
 TTTGTGGCCAAGATCCTGGTGGCCGCGCGCCAGGGCATCGAGGTGGGCACGCCCCGTCTGGTACTGGTC
 40 GAGGACGAGGCGGACGTGGCCGCCTTCAAGGACTACAAGCCCCAGGGCCAGGCAGCGCGCGCGGGGACAAT
 AAGGAGGAGGGTCCGAGAGCAAGGCCCGCGCGCAAGGCCGAGAAGGCGAGTGCGGCGTCTGGCGTGGCG
 AGGAACGACCACCTGCTTGGCCCCGCGGCCCGGTGCTGCTGCAGGGCGTGGGGCTGTGGTTCGAGGACGTG
 ACGCCCACGGGGCCGACGGCGTGTGACCAAGGGCGACGTCTGGCGGCGATCGAGAGCGGCGCCAAGCCC
 GCGCCGAAGAAGGAGGAGGCGAAGAAGAGCGAGAAGGTGGTTCGCGAAGAAGGAGGCGGCAACCCCCGCGCAG
 45 GAGGCCGCGGCGCGCGCTCGGGCGCGCGCGCGCGCCAGCGCGGCCAGTACACGGACGTGCCCCAACTCG
 CAGATTTCGAAGATCATCGCCAGCCGCCTGTCCGAGTCCAAGTCGACCATCCCGCACCTCTACCTCTCCGCG
 GACGTGGACCTGACCGCGCTGGCGGCGCTGCGCGCCGCGCTCAAGGCGCAGGGCCTGAAGATCTCGGTCAAC
 GACTGCGTGGTCAAGGCGGCGCGGGCGCGCTGGCGGCCGTCCCCGCGGCCAACTTCCACTGGGACGACGCC
 50 GAGGAGCGGGCGCGGGCGGTGCGGAGCGAGGGCGTGGACATCTCCATCGCCGTGGCCACGGACAAGGGCCTC
 ATCACGCCGCTGCTGCGCGGCGCGACGGCAAGGCCTTAACCGCCGTGGCGTCCGAGATCCGGGAGCTGGCG
 GGCCGCGCGCGCAAGAACAAGCTGCGGCGGAGGAGTTCAGGGCGGCTCCTTCTCCATCAGCAATCTGGGC
 ATGTTTGGCATCGACAGCTTCTGCGCCATCATCAACCCGCCCCAGGCCTGCATCATGGCCGTGGCGCGGGG
 CGCGAGGAGACCGTGTCTGTTGACGGCGGCCCGCCAGACCAAGACCATCATGACGGTACCCTCTCCGCGGAT
 55 CACAGGCTCTACGACGGTGAATTCGCTCGGCGCTCTGGAGTCTTCAAGGCCAAGATCGAGGCGCCCTAC
 ACGTCTCTCTGGCTTGA

SEQ ID NO: 243

Prototheca moriformis Pyruvate Dehydrogenase, DLAT E2 subunit, protein

60 MPALSPTMSQGNLVEWTVKEGQEVAPGDSLAEVETDKATMTWENQDDGFVAKILVPAGAQQIEVGTPVLVLV
 EDEADVAAFVDYKPQGGAAAAAGDNKEEGSESKAPAAKA EKASAASGVARNHLLGPAARLLLQGVGLSVEDV
 TPTGPHGVVTKGDVLAIESGAKPAPKKEEAKKSEKVVAKKEAATPAQEAAPSGGRARRRQRQYTDVPNS
 QIRKIIASRLSESKSTIPHLVLSADVDLTALAALRAALKAQGLKISVND CVVKAAGALAAVPAANFHWDDA
 EERARAVRSEGVDISIAVATDKGLITPVVRGADGKALTAVASEIRELAGRARKNKL RPPEEFQGGSFISNLG
 65 MFGIDSFCAIINPPQACIMAVGAGREETVLVDGRPQTKTMTVTL SADHRVYDGEIASALLESFKAKIEAPY
 TLLLA*

SEQ ID NO: 244

Prototheca moriformis Acyl-CoA Binding Protein 1, nucleotide

70

ATGTCTGAGCGACTTGGAGCCCCAAGTTCAAAAAGGCCGTGTGGCTGATCCGCAACGGACCCCCGGCCGAGGGC
GCCACCACCGAGACCAAGCTGCTCTACTACGCGCACTACAAGCAGGCCACGGAGGGCGATGTACGGGGTCTG
CAGCCCTGGCGGGCGCAGTTTCGAGGCCCCGGGCCAAGTGGGACGCCTGGGCCAAGCTCAAGGGCATGAGYCGG
GAGGAAGCTATGCAGAAATACATCGACCTCATCAGCAAGTCCAGACCCGACTGGGAGAACAACCTGTCTCTC
AAGGACTACCAGGAGGAGTAG

SEQ ID NO: 245

Prototheca moriformis Acyl-CoA Binding Protein 1, protein

MSSDLEPKFKKAVWLIRNGPPAEGATTETKLLYYAHYKQATEGDTVGSQPWRAQFEARAKWDAWAKLKGMRS
EEAMQKYIDLISKSRPDWENNPAKLDYQEE*

SEQ ID NO: 246

Prototheca moriformis Acyl-CoA Binding Protein 2, nucleotide
(transcript)

GCTCGGCCTCCTCCACCTCGCCCTCCTTCTCCTCCTTGGCCTCCTCATCACAAGTTCAATTTTTGATCACTT
GCATGTATACGTTGTTTCGATACACGACATCGGCTCTGGGTCCCTCGAGAGTTCACTGATCCGTCCCCTTG
CGGGCCGCACGGTGCGATGAGCGCCGCGCGAGGTGCAGGAGGAGGAGGAGGACCATTTCGAGGCGGCGGC
CGAGCAACTGACCGTGACCATCGCGACGGGCACGAGACTGCCGGACGCCTCCTGTCTGGAGCTGTACGCCCT
GTACAAGCAGGGCTCGGAGGGGCCCTGCTCCAGGCCCAAGCCGTCTTCTTGGATCCCAAGGGCCGAACCAA
ATGGAAGGCCTGGCAGACCTGGGCGCCATGCCCGCGATGAGGCCAGGGACAAGTACGTCTGTCGCGGTGCG
GCGACTGCTGGGCGGCGCGCCTGAGGGCCAGCCGACCGCGCGCCTACGGCAGCACGGGCCCCGGCGGTGAG
CACCCTGGCCCGCCCTCGGCGAGGAGGCGGGGACGCGGTGGGCGAGGGGGCCGACGACGTGGGGCCCCCT
CCACGCGCGCTGGCCGATGGAGACGCGCGCGGTTTCGCGGCCCGCGTCTGCCGAGGCCACGACCGGCCAGGC
CGAGCGGCGGGACGCGGAGGGGGGACCCCCCTCCACATCGCCGCGGATGCCGGCAACGTCTGAGGGCCATTGA
CCTCCTGAGAGCCAAGGGAGTCGACCTGGATCTGAAAGATCAGGACGGCATGACGGCGCTGCACTATGCGGC
GCTGGCGAGCCAGCGAGAGGCGTACGAGGCCCTTGCTTGAAGCGGGCGCGGATAGCAGTGTCCGAGACGCCGA
GGGCCAGTCGCCCCCGACCTGGCGCCCCCGGACTGGAACCGCGCGTGACGGGTGGGTTGTATGCCAAAACA
TTTGAGACTGGAGTTTCCCAATCGGTGACTGGCCCCGCGCTTCTGTCCAAATTTCTAGACAATTTCTAGAT
CAGAGAGAGACGGCTCGGGTGCGGCTGCAACAGACTGCGCTTTCTGTAAGCAAGGCACTCTCGACAGATGGC
GCTCCCGATGCCGGGCGCGAGCGCGCATCGCTACCCTGCGCGCTGCGCTTCTCCTCCTCCTCGTGCAGG
AGGCGGAGCCGCGCGCGCGGATGACGTTGTGCATGGTGGCGCGATGGTCATGGGCCCGATGCCGCC

SEQ ID NO: 247

Prototheca moriformis Acyl-CoA Binding Protein 2, protein

LGLLHLALLLLLGLLITSSIFDHLHVYVVSIIHDIGSGSLESSLIRPTCGPHGAMSAAREVQEEEDHFEEAA
EQLTVTIATGTRLPDASLLELYALYKQSGEGPCSRPKPSFLDPKGRTKWKAWHDLGAMPRDEARDKYVAVR
RLLGGAPEGQPQPAAYGSTGPAVSTPARALGEEAGDAVGEAGDDVGPLHAALADGDAARFAAAVAEATTGQA
ERRDAEGGTPLHIAADAGNVEAIDLLRAKGVLDLKDQDGMTALHYAALASQREAYEALLEAGADSSVRDAE
QSPADLAPPDWNRA*

SEQ ID NO: 248

Prototheca moriformis Malic Enzyme, nucleotide

CGAGCCCTGGGGCCGTGCACCCTTTGGAGCAGCCAGGTGCAGCGACACCCTTTGGTACGCAGCCTGGCCACA
GGCCCCGTGACTTTTACCTAGACTACATCTGATTGAATTTTATGTATGCTTGTCTGTAATCGCGCATCTCCCG
GTGTGCCCCATTGCTTGCAGTAGGCCTTCTCGATCGCTTACCCGGCTAGCAGCGTCTGGGTGCGGTGGATCG
CCTCGGACGACCCCAACTTCCCTTACAAGCCCCGAGGTGAACGCCACGAGCGTGGTGCACCACCGCGGCA
TGGAGCTGCTGCAGAACCCGGTGTTCACAAGGGCACCGCCTACTCTGTGGCCGAGCGGGAGCGCCTGGGCG
TGCGCGGTCTGCTCCCGCCGCGAGGTGCTGCCCCTGTGCGGCCAGATGGCGCGCGTCTATGGACCGGTACTGGC
ACGGCTCCGACTTCATCAGCCCCGAGGAGATCGAGTCGGGCGGCATCACGCACGAGCACACCCGCAAGTGGC
TCGTCCTGACCGAGCTCCAGGACCGGAATGAGACGCTGTTTTACCGCTGCCTGATAGAAAACCTTTGTGGAGA
TGGCGCCGATTCTGTACAGCCAACGGTGGGCTGGGCGTGCCTGAACTACCACAAGCTCTTCCGCCGCCCGC
GCGGCATGTACTTCTCGGCCTTTGACCACGGCAAGATGTGACCATGGTGCACAACCTGGCCGCACGAGGAGG
TGGACGCGATCGTGGTCACGACGGCTCGCGCATCCTGGGGCTGGGCGACCTCGGCTGCAACGGGCTGGGCA
TCCCCGTGGGCAAGCTGGACTTGTACGTGGCCGCGGGCGGCTTCCACCCCTCGCGCGTGTGCCCTGCGTGG
TCGACGTGGGCACCAACAACGAGGCCCTGCGCGCCGACCCGAGTACCTGGGCGTCAAGCAGGCGCGCCTCG
TGGCGGACGCCTACTACTCCTGCTCGACGAGTTTGTGACGGCCGTATGCTGCGCTGGCCGACGCGCGTGC
TCCAGTTCGAGGACTTTGCCATGGAGCACGCGCGGCCACTGCTGGCGCGCTACCGCAACCACACACCGTCT
TCAACGACGACATCCAGGGCACCGCCTGCGTCTGCGCTGGCGGGGCTCTACGGCGCGCTGCGCGTGTGGGCC
GCGGGCCCGCGACCTACCGGCCTGCGCATCGCCGTTGTGGGCGCCGCGAGTGCCGGCATGGGCGTCTGTGT
CCATGATCGCGAAAGGCATGGAGCAGCACGGGCTGACGCCGAGGAGGCGGCGGACAACCTTTTGGGTGCTGG
ACGCCGACGGGCTGATTACGGAGGAGCGCGCAACCTGCGGGCCACGTCCGGCGCTTTGCGCGCTCCGACG
AGCGGCGCACCGACGGCGAGCGCCTGCTGGAGACCATCCGCGCGTCAAGCCGACCTGCCTCATCGGCCTCT
CGGGCGCGGGGCGCCTCTTCAATTAAGGACGTGCTCCAGGCCATGGCGGAGGTGACGCGCGCGCCCATCATCT
TCCCCATGAGCAACCCCATCGCCAAGATGGAGTGCACGCACAGGAGGCCATCAAGCACACCAAGGGCAAGG
CGGTGTTTGCCTCTGGGAGCCCGCAGGGCCCGGTGGAGTGGGAGGGCCACTCGCACGCCCCCTCGCAGGCCA
ACAACATGTACATCTTCCCCGGCATCGCGCTGGGGGCTGGCTGGCGCGGTGGGCGACGATCACGACCGCA
TGCTCATGGCCTCGGCGGAGGCCCTGGCCCAGTGCACCACGCCCCGCGGAGCTCGCCCTCGGCATGGTCTACC

CCAGCATGACCCGCATCAGGGAAATTTGCTGAGGGTGGCCGAAGCGGTGATCGAGGCCGCCGGCGCCTTGG
 CCCAGAACGAGCGCTGCTCAAGGCAAAGGCCGAGGGGCCGAGGCCCTGCGGGACTACATCCAGGCGCACA
 TGTACCACCCAGAGTACACAACCTTGGTCTACAAAGAAAGCCGGTAAATTGCCGAGCTTGCTTGAACGTCTG
 ATGTTTTTGGCCATAATGAGGCAGCCTCTGGCCAGCGCTGCCCTTGCTTTCCCTTGCTGGTCTTTTTAAAG
 TTGGTTTCGCTCCCTATGGTGTGTTTGTGCGATGTGTAGAAAGT

SEQ ID NO: 249
 Prototheca moriformis Malic Enzyme, protein

LEVNATSVVHHRGMELLQNPVFNKGTAYSVAERERLGVRLPQVLPVSRQMARVMDRYWHGSDFI SPEE I
 ESGGITHEHTRKWLVLTELQDRNETLFYRCLIEFNVEAPILYTPTVWACLNYHKLFRPRGRMYFSAFDHG
 KMSTMVHNWPHHEEVDIVVTDGSRILGLDGLCNGLGI PVGKLDLYVAAGGFHPSRVLPCVVDVGTNNEALR
 ADPQYLGVKQARLVGDAYYSLLEDFVQAVMLRPHAVLQFEDFAMEHAAPLLARYRNHHTVFNDDIQGTACV
 ALAGLYGALRVLGRGPADLTGLRIAVVGAGSAGMGVVSIMIAKMEQHGLTPEEAADNFWVLADGLITEERA
 NLPahVRRFARSDERGTDERLLETIRRVKPTCLIGLSGAGRLFIKDVQLQAMAEVTPRPIIFPMSNPIAKME
 CTHQEAIKHTKGKAVFASGSPQGPVEWEGHSHAPSQANNMYIFPGIALGAWLARSGITDRMLMASAEALAQ
 CTTPAELALGMVYPSMTRIreisLRVAEAVIEAAGALAQNERLLKAKAEGPEALRDYIQAHMYHPEYTTLVY
 KESR*

SEQ ID NO: 250
 Prototheca moriformis ATP:Citrate Lyase Subunit A, nucleotide
 (transcript)

TTTTTTTTAGGGTATCCAGACTTCAACTGCTAGGTTGATCAGTGCATAATACAGTGGTGACCACAGCTGTGC
 AGGATGCTCTAGGGGGCACCGTATCCAGGCGTTCTAGCGAATCGACAATTGCACGAATAGAATCTATCATGC
 GCGATTGCGGGCAGGTGGGGGCGCATGAACAGCTTCGCCAGCCCTTGACGGGAGAAAAAGCTACATGCGCTT
 CTTGAGACGCGCATCTACAGACGCGCATCCAGTCCTGTACATCTTCAGATGCGCCTCGTGAGACGCGTGGAT
 CCAGACACGCACCCCTGCCTTCAGGCGGCGCTGGAGGCCTTGGCGTGGCGGGCGTGGCCTCGGCCAGGTGG
 TCGATGGCGCGCTTGCAGATGCTGGTTCATGGTGGTCTCGGGGCCGTAGACCTCGATGTGATGCCCCGTCTCC
 TGGCCAGCTTGCGCATGGCCTCCAGGCCGGCCCTGTAGTTGGGCCCGCCGCGGCGCACGAAGAGCTTGAGG
 CGCGCGCGCGGATGGCGCCCTCGCGCTCGCGGATGGCGGCGATGATGCCGCGGAAGGTGGCGGCGACGTGC
 GTGAAGTTGGCGATGCCGCCGCCGATGAGCAGCGCGCGCGCGCGCGCGTGGCGCGGGCGGTGGCGCAGTCG
 AGCAGGCGCAGCGCTAGGCGTGGTCTCGGCCGCGGAGGGCCCGCCGAGTACTCGGCGTAGTTGCCCAGC
 TCCGCGGCGCGCGCGAGGTGCGCCGACCGTGTCCGCGTAGATGACGCTGGCGCGCGCGCGCGCGGCGACCATGGTC
 CACACGCGTCCGGACGGTTGAGGATGCTCAGCTTCAGGCTCGCGCCCGTGGCCTCGTCCATGCCCCGCCACG
 GCGCCTCGGCCGCGCTCAGGCGGCGGCCAAAGGGCAGCGGGAAGTCCAGGCCCTCGCCGCCCACTTCTTG
 GCGGAGCGGAAGCCCGCGGTGTGCTCCAGCTCGCGCGCATGTCCAGCGGGAAGGGGCGCCCGTCCGGCCCC
 AGCGTGAAGGGTTTCATCTCCATCAGCGTGAAGTCCAGGTCCACGTACACCGCGTAGGGCCCCCGCGATGAAG
 GCCTCCAGCCCCGGGCCGACCTCCAGCGGCAGCGTGGCCACCAGCGGCGCCAGCTTCTCGCCGGACAGCTCC
 TCCTCCACGCCCACGCTGATCGTGCGCACAAAGTCCCAGTTGTCTCTCGATCTCCATGCCGCCCGCCTCCGAG
 AAGGAGATGTCCGCGCCCCGTGCGCGTGGCTGGATGCAAAGGTAGTACTCCTCCTCGTGCGGCACAAAACGGC
 TCGATGATGAAGCAGTCGATCACACCCTTGACCGTGGCCATGTGATGACCTTGCCCATGCGTGGCCGATA
 AACTCCTCAGCGCCGGCCAAGTCCAGCTTGAGGCCCCAGAGGTGCTGCTTGCCGCGCTTTCCAAACAGACAA
 TCGGGCTTGACCACCAGCGGGACCTGCGTCAACCAGGGGTGCTCGCCCACTCGCCAAAGTTGGTGTCC
 TTTTGTGACCTGGGCCACCCTGATGGGCAGGTCCAGGCCCGCCAGACGCGCCATGTGCGTCTTCAGCAGCGTC
 TTGGAGTTGTACTCCCTGATCTTTTTCGCTGCCATGGCGATCTAATGCGTCCCTGGTCTTGGTGGAGCCGGA
 GAGGCTGGACAGAAAGCGATCGCGGGGTGCTGGCGGGACGTCCGCGGACAACGAAGCGCGGCCCCAAGCCG
 ACCAGCACGCACCCGTCAAGCCTCACGGATGCCGAGCACGACGCGATTCTGTAGCTTTGGAGCGGTAGGTAG
 CCTTGAATCCAGCAATACGGCCGTCTGGAACACTGAGCGGCTTGTGAGCGACCTCGCGCCTATGCAATAAGG
 GGTGGTTTCGAGTGTGACGAGGCTCTCCTAGCCTCGGGCACAGTGGGAATTCCTGGCCGCCAAAAAGGTC
 GCGAGCAGCCCCAGTGGGCGCTTTCAGTGATGTACAGAGCAGCCCCAGCGTTTATCTTACATTTTACTTGGT
 GACAGTGGATGGGTGTTTGAATGAAGTTACCTGTGTCTGATCACTTATGATGACACCAGTGTGGTTGGCA
 CGGCGAATGCATATGGATGGCATGCAGAAAGTACACCTCCGTTTTTGTCTGCTGGCACAAATTCCTACTGC
 AAAGTCTCATGTTTATTCTGTGCAGTCCCTTCTAGTTATAGACTCGGTCCAATTACGAACAATTTGCATGG
 GATGGGTGCTCAACTCATTCCTGCCAGATACCACAAACAGACCCCAACATAGCCAGACTCATCTATTGTTG
 CAAGTCAAATCAAGCTGCTCACAAAAGTATGATTATGGTGGGAGCCTGTGCTTGTCTCAAAGAATGTGATCAC
 GAATGGCGCAGCCCTGCAGCATTTCCCGTGAATATATATGTATATAAACGAACACATCCAGGTATTGCCTGC
 AGCGTTATTCTATTTTAGATTGATCAAGAAAGAGTAAGCTTTTGTGTGTTGTTG

SEQ ID NO: 251
 Prototheca moriformis ATP:Citrate Lyase Subunit A, protein

MARKKIREYNSKTLKTHMARLAGLDLPIRVAQVKKDTNFGELVGEHPWLTQVPLVVKPDCLFGKRGKHDLV
 GLKLDLAGAEFIGQRMGKVIDMGTVKGVIDCFIIEFPVPHEEEYLLCIQGTTRTGADISFSEAGGMEIEDNW
 DLVRTI
 SVGVEEELSGEKLAPLVATLPLEVRPGLEAFIAGAYAVYVDLDFTLMEMNPFTLPGDGRPFPLDMRGELDDT
 AGFRSAKWGGEGLEFPLPFGRRLSAAEAAGVAGMDEGTGASLKLILNPSGRVWMTMVAGGGASVIYADTVGD
 LGAAAEELGNAYEYSGGPSAAETHAYALRLLD CATAGADGRARALLIGGGIANFTDVAATFRGIIAAIREREG
 AIRGARLKL FVRRGGPNYRAGLEAMRKLQETGIDIEVYGPETMTSICKRAIDHLAEADAADAKASSAA*

SEQ ID NO: 252

Prototheca moriformis ATP:Citrate Lyase Subunit B, nucleotide
(transcript)

5 GCCGGGGGCTGCGCCCTCATGCTTCCCCTGGCCCTCATTGTGAACAACCTCAGTCCTCACCGTCGCTACCCG
ACCTTCTGGTGGTGAGGGGCTGCAGCGCCTGTGTTATCTTACCGTGGTGGGTCCACATGTCTTGCAGCAAC
AGTGGCCAGCGAGACACGCGAGTGCCACGTTTTAGATATTTATGCCATCTTTCTTTGTTGAGTGAGCATCC
10 ACGCCAATAGCTCTCTTCAGGTGCGAGGCTTTTCATTATATCATTTTGGGATTGCGATGCTGGTTCCTGTGTG
CATTTACAATCCAGGTGCGACAATTTCTTCCCACTGGCAAACCTGCACTCATCGCTACCCACCAACCGTGGT
TGCCTTTTTGCCTCCCATTTGAGCCATTGCCAATATATAAAATATAATACCAACCCCTACGTCATCATATCATT
15 TATTCATGTGCAAACCCCTCGCTCGTACCAATGACAATGAGAAAAAAGGTGGCGTCCGCGAGTGTCAAGGC
AAAGGGGGCGGCGGTTTGCCACGACCACGGGGGACGCCTTTCGAATGCCGCGTCAGCATGGTGAGCGGGAGG
CATTAGCGAGGATTGGCCAAAAGCGGTGCTGTTCCGGCCTCCAGCCCGTACCAGCCCAAGCCCAATCGCC
AAGGCCTCTACAGCCACACGAAAGAGGTGCTACAGGCGCTCAGAGCGACCGCTCGATCACGATGCCCACGT
20 CCCCACGCACGCGGGGCGCTACGGCCAGGCGGGCCAGCTCTTCAGCAAGCAGACGCAGGCCATCTTCTACA
ACTGGAAGCAACTGCCAGTGCAGCGCATGCTGGACTTCGACTTTCTGTGCGGTGCGACCCTTCCCTCTGTGG
CATGCATTGTCCAGCCCGGCGAGCCAGCAAGGGTTCCAAAAGGTCTTCTTCGGCCCGGAGGAGGTGGCCATCC
CGCAGTACGGCTCGATCGCCGAGGCGGTGGAGCAGCACCCGCGCGCGGACGTGCTGATCAACATGTGAGTT
25 TCCGGAGCGCTTTGAGTCGTGCTGGAGGCGCTGCAGCAGCCGACGATCCGACGCGTGGCCATCATCGCCG
AGGGCGTGCCCGAGCGACACCAAGAAGCTGATCGCCGTGGCCAGCGCGCGAACAAGGTATCCTGGGCC
CGGCCACCGTGGGCGGCGTGCAGGCGGGCGCCTTCAAGGTGCGGACGCGGCGGCGACGCTGGACAACATCG
TCGCTGCAAGCTGTACCGCCCGGCGAGCGTGGGCTTTGTGAGCAAGTCGGGCGGCATGAGCAACGAGCTGT
30 ACAACGTGATCGCGCGCGCGGCGGACGGCATCTACGAGGGCATCGCCATCGGCGGCGACGCCTACCCGGGT
CGACGCTGTGCGACCACTGCCTGCGCTACAGCACATCCCCGGCGTCAAGATGATCGTGTGCTGGGCGAGA
TCGGCGGGCGCCACCGTCGCCGACGCCATCGGCCTGCTCTGGTTCAAGCGCCGCTCGCCCTCGTGGCCCT
35 GGGTGTGCGGCGAGTGCAGCGAGCTGTTCCAGACCGAGGTCCAGTTTGGGCACGCGGGCGCCAAGAGCGGCG
GCGCCGACGAGTCCGCGCAGGCCAAGAAGCGGCGCTGCGCGCGGCGGGCGGCGTGTGCCCCGACTCGTTG
AGGACCTGGAGCCGACCATCAAGCGCGTCTACGACGAGCTGGTGGGCGCGGGCGCGATCGTGGCCGCGCCG
TGCCGCGCGCCCTCGGTGCCGAGGACCTCGCGGCGGCCAAGCGCGCGGGGCGCGTGCAGCTGCCACGC
40 ACATCGTCTCTCCATCTGCGACGACCGCGGCGAGGAGCCGACCTTCAACGGAGAGTCCATGAGCGCGCTCA
TGGAGCGCGCGCCACCGTCGCCGACGCCATCGGCCTGCTCTGGTTCAAGCGCCGCTCGCCCTCGTGGCCCT
CGCGCTTCATCGAGATGTGCGTGTGCTGCGCCGACCATGGTCCCTGCGTCTCGGCGCGACACAACACCA
TCGTACCCGCGCGCGCGGGGAAGGACCTCATCTCTCTCTCGTCTCGGGCTGCTCACCATCGGGCGCGCT
TTGGCGGCGCCATCGACGCGCGCGCGCAGCACTTCAAGGAGGCGGTGAGAAAGGCTGGGAGCCCGACGAGT
45 TCGTGGAGATCATGAAGCGCGCGGCGTGCAGCTGCGGCGCATCGGCCACCGCATCAAGTCCAAGGACAACC
GCGACAAGCGCGTGGAGCTGCTGCAGCGCTACGCGCGGAACGCTTCCCGTCCACCCGCTTCTCTGGACTACG
CCATCGAGGTGCGAGGCTACACCCGTGAAAAGGCGCAACCTGGTGCTCAACGTGCGACGCGTGCATCGGCG
50 CGCTCTCTCGACCTCCTGCACCTCGGTGCGCATGTTTGCCAAAGCCTGAGATCGACGACATCGTCCAGATCG
GCTACCTAAACGGCTCTTCTGCTCTCGCGCGCGCCATCGGGCTCATCGGGCACTGCCTCGACCAGAAGCGCC
TGGGGCAGCCCTCTACCGCCACCCCTGGGACGACGTCTCTACACC

SEQ ID NO: 253

Prototheca moriformis ATP:Citrate Lyase Subunit B, protein

45 MPTSPRTPGRYGQAGQLFSKQTQAIFYNWKQLPVQRMDFDLFCGRITLPSVACIVQPGSQQGFQKVFFGPPEE
VAIPQYGSIAEAVEQHPRADVLINMSSFRSAFESSLEALQQPTIIRTVAIIEGVPERDTKKLIIVAQRANKV
ILGPATVGGVQAGAFKVADAAGTLDNIVACKLYRPGSVGFVSKSGMSNELYNVIARAADGIYEGIAIGGDA
YPGSTLSDHCLRYQHPIGVKMIIVLGEIGRDEYSLVKALEAGAITKPVVAVVSGTCATLFTQTEVQFGHAGA
KSGGADESQAQKNAALRAAGAVPDSFEDLEPTIKRVDDELVGAGAIVPAPLPPPSVPEDLAAAKRAGRVR
50 VPTHIVSSICDDRGEPTFNGESMSALMERGATVADAIGLLWFKRRLPPYATRFIEMCVVLCADHGPCVSGA
HNTIVTARAGKDLISLVSGLLTIGPRFGGAIDGAAQHFKEAVEKGWEPDEFVEIMKRRGVRVPGIGHRIKS
KDNDRKRVELLQRYARERFPSTRFLDYAIEVEGYTLQKAANLVNVDGCIGALFLDLLHSVGMFAKPEIDDI
VQIGYLNGLFVLARAIGLIGHCLDQKRLGQPLYRHPWDDVLYT

55 SEQ ID NO: 254

Prototheca moriformis LEC2, nucleotide (transcript)

60 GAGTTTCAAGTTGTGCTCTCCACTTCCGCGCGCATTGCGACGAATTCTCACGCCCTTTGTGCAGGTGTCCAG
TTCCACATCGGGAGGCATGGGGGGCCCGCCACCTCGGCGCGTCAAACGGCGTCAGGTAGGGCGGGCAGCTG
GCACTCTTGGCCGATGACCCGTACCGAGCAGTACTGCCATCGCTCGGGTCGGAGACGGTCGGCACCGCCCT
GATCGCGAGTCTCCAAATGATCAAACCTTGAGACACGATGGACGTGCGCGCGGTGATATCGCCCC
65 AGACAACGCCGCGCTGGCCGCGAGGCCGGGTGCGGTGTCGACGCTCCGCGCAGGCGCCAGACGACTGGGTG
CCGGTGGCAGCGAGGACGAAGAAAGGGATGTACTGGACCTCAAGCTTCCAGGAGTGGGCGGGACGCCTCTGT
CGCCCTGGAATGCGGTGATCCTGGAGGACAGCCATCGACGAAGGACGACGCACGTGCCACGCAACACGATG
AGAGTCCACGCCGACGGCTTCCCGCCGCAAGATCTGAAGAGGCTCCGATTTAGAGTTCCATGGCCAATGGT
70 TTTCTTGTACAAATGATTTACATGCAGTACTGTTGCAACACACCTCTCGACACCGGGGATATCGCCTCGCAG
GGCAATCGAAGGCCGTCAAATCTGGCGGTAGTCTGGGGAAGTGACCTGCCCCGGGGGCAATCCGATCCAT
CAGCGAGGTGCGCAGGACAGGCACGTCTGCCGTATATTTGAGGGCTTCCCTAGGGGTGAGTCACTTGTCT
CGGCTTCATTTTGCAGGTTTTACCAAAAACCTGCTTGGCATTTTTCTTAACAGGGCATCTTGAGGTGCGAC
ATACCTGATACCGACGAGCTCGAGCCAACCTTCGCTCGCTGCGGCATTTTTGTAACGGTCGACACGAGC
75 ATTGCGGCAGCGATAATGTCCAATACTACGATCAAAGTACTGTGCTACAAGTCTCTGACGCACAGCGATGTG
ACGACGGAGATTGCCAAGTCTGGGCGACTGGTCTGCCGAGGGCCAGGTACGGCCCTGCCTGCAGCGCCTC

CTGGCCCTCGGCAGCGCCGGCGGAGCCCCAGGCAAGGGTCGAGCGGGATCTCTGTGCGGTGCCGTCCCCCTTG
 TTCATGCACGACATGGAGGGCAAGGTCTGGAACCTTTCAGCTGAAGACATGGCACAACGTTGTGGCCTCTGGA
 GGCCGGGCAACCTACGTGTTGGAGAATACTGCTGGTTTTGTGTCAGAAAACAAGCTTCAACCGGGAAGCTGG
 CTGGCTATTTGCGAGCAGGACGACCGCCTGGTCTCCGACCCGACATCGTCAAGGGCCGCGATTTCCTGCTC
 5 CCGTGGTGGCGCCCGGACACGGTCAAGAAGCCTCTCCCCAGTTTTGGCGCCTGGTCTCTCAAGAGT
 TCGGACCTCTCGGCCACGACATGACCTGGTGGCTGGGCGACGACACGCCCTCCAGCAGGAGGACAAG
 GCCTCCGCGCGTCCCCCACCATCCTGTGCGGTGCCGTGGTGGCGCCGCGCCCATGTACGGCGCCATCAGC
 GTCCAGCAGGTGCCGCGTTGGGCTGGGCGCTGAGCATGGTGGCGACACAGCCGCGCGGCCAGCCGTGCTGG
 10 TAACCCCGGGTCGCCCCCTCCGGCGTGTCTTCAAGCACACCCCTGCCCTTCCCAAGCCCGCGCGGCCCGCTC
 CGCGCGGCTTCTCCCCCCCCGAACCCACCAACAACGTGACCTCGACGGCCACGCCCGTGGGGCTCGGGACCT
 TGGCACCCCTTCAGTCGCACTACCCCGTCCCCACCAATAACAGTGTGCTTGGAAATAAACCGGCAGCGTC
 CTCAGCCGACAGCTCACCCTCACCCTGTTCCCTTACGCGCCCGCGGGACGCCCTCGTCCCGGCAAGC
 CCGGGCGCAGTTGCCACCCCTCGATGTCTCACAGTGTACTCGCCCTTTTCTCGATCGGGCTCCCAAGCTT
 15 GTCAATTTTGGCGCGCGGCCCTCTCTTGCACTGTGGCACCGGACCCAACTCTCATTATTTTTGTTGCA
 CTTGAACCTGCACACCCGACCTTGGCCCCACCGGCACCTTGCAGCAAGCCAGAGTGTCCCGCAAATTCATT
 CTTGGTGGCGTTCGGCCGACTTCTGCCCAAAGCGCGCGGCCCGCCCTCGTATAGATACTGCTCTTTTCTCTTA
 CTGTGAGATCTCCCTCGGTACCCGGGTGCGGCTGCCAGGCGCGCAGCAGGCCGCCAGGTGGCGTAGGCC
 GACCAGCGCCAGAGTGTGAGCAGGTAGAAAAAGTTGGCGTTAGGCCGCCAAGATCCAGCCACACATG
 20 GGCACAGCGCGTGCCTGCCAGCAGCGTCATCCACAGGCACTGCCAGGCGTGGGGTGGCGCCCTCTCCG
 CCGCAGCACGGCCGCGCTCGTCGCCATGCACAGCGCCAGGCCA

SEQ ID NO: 255

Prototheca moriformis Leafy cotyledon2 LEC2, protein

MSNTTIKVLCKSLTHSDVTTEIAKSGRLVLPRAQVTACLQRLALGSAGGAPGKGRAGSLCGAVPLFMHDM
 EGKVNWFQLKTHNVVASGGRATYVLENTAGFVHENKLQPGSWLAICEQDDRLVLRDIDVKGDRDFFLPVVAP
 ATTVKKPSSPSFAPWSSKSSDLSAASMTLVSLGHDTTPSSSEDKASAPSTILSVPLVRPRPMYGAISVQQVP
 RLGWALSMVPHQPRGQPCW*

SEQ ID NO: 256

Prototheca moriformis DGK/SpiK, nucleotides (transcript)

GGTTTGAGAGGAACAATTTGGGGATGGGCGGTGGCACTGGGATGAGGCACTGGGATGAGGCGCTCATTGTCA
 AAGAATAAAACGGAATGGACGAATCGGTGAGGGCCGCGACACCGGCCAGGGCCCAGGAGGCAGGCGTTTCA
 35 AGGCGGGGAATGGGGCCCCAGCTAGTTTCGCGCTCGTCAGCCCTCCACGCCTCTGGCAAACACGTGATCA
 CTCCGCGGTGGAGCTGCGCGTTGATCGTGCGGCTTTGAATGAGCTCCCATCGACGTTCCAGTGGCTCTCCT
 GGCCAAACCGCTTCGATGTGCACCGCGTGGGCGGGCAGGACCTGGATGTAGGACCCGTGCTGCCCGCCGCT
 CGAGGCCGATGTGCGACATCCTGAGCAGAACTTGAGGTACTGGAGGCGGCTGCACCGCCTGATGAGCACCA
 40 GGTGCAGCAGGCCGTGCTCAGGTGACCATACTTGGCGACGCCCTTTTTCGACTTGTGAGAGCGGCAGGGCA
 TGATGACGAGCATGATGCCTGCAAACTCCTGCTCCGGGAGGTGCACAACTCTGCTCGCCGCGCGCTCCCAGA
 GCTCCTCGGCCGAGGCGGTGAACCCCGCGCGCGCTCGCGGCCCTCGCGCCCGCGCGCGCGCAGAACTCGC
 AGTTGGCGGAGGCGTGGCTGAAGCTGGCCGACCGCTGAACCGCGCGCGCGCGCGCGCGCGCGCGCGCG
 CGGCGTAGGCCCTGTTTCGCGCGGAGCATCTTGGCGCCACAGCTCGTACCGCGCGGGCCCCAGCCAGCGCA
 45 GGCGCTCCGACTCGGCCATGAGGTGCGCCATGAAGCCGTAGCTGACCATGCAGGTGCGAACTCGGGCAGGC
 CCGGCTCCCGCTCCAGACGCAGCACATCCAGCGCCACGGCGTGGCCAGCACGACGTGCACCGCGGCCGTGA
 AGGCCGAACGCGTGCCGTTGAGCGTGCAGGCCACGGCGTGGTGTGCGGCGCGGGATGTGGCCGATGCGGA
 ACTGCGCCCCCGCGCGACGGCCGCGCACCCGGCCGCGCGCAGGCCAGCACACCGTTGACGATCTCGTGGA
 50 AGAGCCCGTCCCCGCCACGGCCACCGCGCTGAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG
 TGTGCGGCTCGGCGCCAGCGGGGACAGGGAACGGCGCTCCGCGCCCGCGTGCAGCGTGCCTCTCGGCGCCG
 CGCCGATCACCTGGCGCACGATGGTGACGGCGTGTGCGGCTCCTGCGTCTCCACCACCGCGACCTTGATCC
 CGGCCTTTTCGGAAGACGGGCGCCACGGTGGAGCTCCAGGTGAGGGCGCCTGGCGCGAGCCGCCACCGGGT
 TCACGATCACAGCAGCCGCGCGGGCGCCAGGTCTGCGCCCGCATGCCGGCTTGACGCTGGGCGTGCCACG
 55 CACGTAGCAAGTCTCTGTGAGACAGGCTCACTTGGCGCGGATGCCAGTACGAGGGGTGGAGGGCG
 CGCGCCGAAAGGTGAAGAGCGTGAGGGCCAGGCTCTTCTGAGCCAGGAACACCACCTGGCGACGAGCGCG
 AGGGCTTGACCTCCGCGCGATGATCTCTTGACGCGGACGCACTCTGCACGCTGCCGAAGCCTGGCAAGA
 GGTGACGCAGCCGTCCATCTCAAACGCCCAGCTCAAGCCGCCCTCGCTGCCCGCCAGATCCAGACTGACTT
 60 GCGCCGCGACGCCGTCCACCTGCATCGGCACGCCCTGCAGCTCCGCTCGCTGAGGCGCTTGGGAAAGGAGG
 AGGACCGCGATATCAGGGGGTGATGAGTGGGGGTGCTTGGGGTGCGCCATCAGGCTGGGCGACGCATTGG
 TACCTCGTGTCCGATGCGGAGAAGACAGCGATCCCCCGGGGAAGAGGGCAGGCTCACGGAATCGTGTCCA
 65 TGTCCGATGACGGGCGTGAAGCACCAGCTCGGTGCGAGAACGTCACGGGGAGCTCCGCGCATGTCCCACT
 ACCGACCCAAACCGCACGCAGGGAGACGGTCCGTGAGGCATGAACGCGATGCGCATAAGGAAGCAACAACCAT
 GCGTGTGTGGCAATCGAACGGCGAAATGGGCCGATTCTTGGCAAAGGGGGGAGGCTGGCGAGGGACGCACAC
 CGCGTCTTCCGTGCGCATCACGTGGATCTGTGTTTATGCATCGGGATGCCTTACATGCAACATTGGCGCC
 70 GAGGCAGGCGCAATGAGCCGAGCAACAGTGTACGCACGCCAGGAATCAATTATACTGACTAGACATTTCTG
 CCATGAACCTTTCTGACTCGCTCTCCACGGCAGCAGGCGCTGGCGGGGCGGGAGCGGGGGCGGGAGGGGGG
 GGGCGCGCGGTGGCGCTCGTGTGCGGCATACCCCGACTGGCGGGGGGGGGGGCGCGCGCAACGTCGT
 CCGTGGGCGGCGGTCCCGGGACCGCGCCCGCTAGCGTGTGATGCGCGTGGTACGCGCCCGCTGATCGG
 GCGCGGGCGGCGGCACCGCCCCGTACTGCGCGCGGTAGGCGTACGGGTGGCGGGCGCCAGGGTAGGCGCGT
 ACGGCTGGGCTGGGGCTCGGCGCGGGGTACGCACCCGGCGGGCGGGGCGCGCGCGGTACGCCCCGCGCG
 GCGGCGGCGGCGGGCGGGTAGGCGCGCGGTAGGGCGGTGGTGGGGCGCGGTACCCCGGGCGCGTA

SEQ ID NO: 257

Prototheca moriformis DGK/SpiK, protein

MQVDGVPAQVSLDLAGSEGLSWAFEMDGCVNLLPGFGQRAESLPLQEIIAAEVKPSLRRQWCSWLEKAWA
 LTLFTFRRAPSNPSYWHPRQVTLVSTDEELLRAWHAQLQAGMRAQTWRPRRLLVIVNPVGGSRQAPSTWSST
 VAPVFRKAGIKVAVVETQEPQHAVTIVRQVIGGGAESDADAGAERRSLSPAGADRDKPRRSRSPSRRAFDGV
 VAVGGDGLFHEIVNGVLGLRAAGCAAVAPGAQFRIGHIPAGSTDAVACTLNGTRSAFTAAVHVVLGDAVALD
 VLRLEPGLPEFATCMVSYGFMGDLMAESERLRWLGPARYELVGAKMLAANRAYAARVSYLAPAAGPGRAS
 FSHACLANCEFCARGGREGRDGAAGFTASAEELWERRRAEFVHLPEQEFAGIMLVIMPCRSDKSKKGVAKYG
 H LSDGLLHLVLIIRCSRLQYLKFLLRMSHIGLEAGGQHGSYIQVLPAAHVHIEAVGQESHWNVDGELIQSRT
 INAQLHRGVIDVFARGVEG*

SEQ ID NO: 258

Prototheca moriformis Ketoacyl-ACP Reductase (KAR) nucleotide

TGGAGTTCAGACGTGTGCTCTTCCGATCTCGCGTGGTGAGCAAGCCTGCGCCGACGTGTTTGGTTTCTTCTT
 CATGTTGTTGCTCGGAGGCGCCGTCGATTAGTTTGCCACCTGACCACCGTGCCACAGTCCCCGACCAGCTC
 ACTGGCACTTTACTGCATGCTCATGCCGCCGTCGATGCTGAGGACCTGGCCGGTGATGTAGGCGGCGGCCGG
 GTCGGTGAGGAGAACTTGACCAGCCCGCGCAGCTGGTCCGGCGTGCCGTAGCTGTTGAGCGGGATCTTGGC
 GAGGATGGCCTTCTCGACCTCGGGGTTGAGCAGCCCGTCATCTCCGACTCGATGAAGCCCGCGCGATGGC
 GTTGACGAGCACCTTGCGGCCGGCCAGCTCGCGCGGAGGGAGCGCGTCATGCCGATGACGCCGCCCTTGGA
 GGAGGCGTAGTTGACCTGGCCCGCGTTGCCCATGAGCCCCACCACGGAGGTGATGTTGACGATGCGCCCGCG
 CCGCGCCTTGCGCATGACCTTGCGCCGCGGCCCTGGCTCGCGTAGAACACGCCGGACAGGTTGACGTCGATGAC
 CTGCTGCCACTGCTCGGGCTTCATGCGCAGCAGCAGCGTGTGCGCGTGATGCCCGCGTTGTTGACCAGCGC
 GTCCAGCGGGCCCCACCGGTCCACCGCCGCTTGATCAGCGCCTGCGCGTCTCGATCTTGACAGGTCCCC
 CTTGACCACGATCGCCTCGCCGCCAGAGGCCTCGATCTGGTGCGCCACCTCCTCCGCGGCCCGCTGGACGC
 CGCATAGTTGACCAGCACCTTGCCGCCCGCGCGCCAGCTTGAGCGCGATCGCGCGGCCAATGCCGCGCGA
 GGAGCCCGTCACCACCGTCACGCCCTGCTCCAGCTCCTTCTGGAGGGAGACGCCGTTGGGCGACGAGGCCAC
 GGAGGCAGCGATCCATGCCGATCAACAGCCGGCTTTTTCAGAACCTTAATCATGCAGAGCTCCGTGATGAACG
 TGCCGGCGACCTTTTCGGCGGGATGTGCCCTTAGGGCCTCCCCCGCCTGCCGGCCGCTCGCCCGCGCGAC
 ACGCGCTCGCGTCCCGGTGCGCGGCCAGCTGACGGTCAAGGTGCTGGTCAACTATGCGGCGTCCAGCGGGGCG
 CGGAGGAGGTGGCGCACAGATCGAGGCCTCTGGCGGCGAGGCGATCGTGGTCAAGGGGGACCTGTCCAAGA
 TCGAGGACGCGCAGGCGCTGATCAAGGCGGCGGTGGACCGGTGGGGCCGCGTGGACGCGCTGGTCAACAACG
 CGGGCATCACGCGGACACGCTGCTGCTGCGCATGAAGCCCGAGCAGTGGCAGCAGGTTCATCGACGTCAACC
 TGTCCGGCGTGTCTACGCGAGCCAGGCGCGGCCAAGGTTCATGGCCAAGGCGCGGCGCGCATCGTCA
 ACATCACCTCCGTGGTGGGCTCATGGGCAACGCGGGCCAGGTCAACTACGCCTCCTCCAAGGCCGGCGTCA
 TCGGCATGACGCGCTCCCTCGCGCGGAGCTGGCCGGCCGCAAGGTGCTCGTCAACGCCATCGCGCCGGGCT
 TCATCGAGTCGAGATGACGGGCGTGTCAACCCCGAGGTGAGAAAGCCATCCTCGCCAAGATCCCGCTCA
 ACAGCTACGGCACGCCCGACCAGGTGCGCGGGCTGGTCAAGTTCCTCCTCACCGACCCGGCCGCGCCTACA
 TCACCGGCCAGGTCTCAGCATCGACGGCGCATGAGCATGAGTAAAGTGCCAGTGAGCTGGTGGGGGACT
 GTGGGCAGGTGGTCAAGGTGGCAAACTAATCGACGGCGCCTCCGAGCAACAACATGAAGAAGAAACCAACA
 CGTCGGCGCAGGCTTGCTCACCACGCGAGATCGGAAGAGCACACGTCTGAACTCCA

SEQ ID NO: 259

Prototheca moriformis Ketoacyl-ACP Reductase (KAR) protein

MQSSVMNVPATLSAGCALRASPRLPAAAPQAHAAGRRSPGQLQVTRAAASVASSPNGVSLQKELEQGVTVVT
 GSSRGIGRAIALKLGAAGKVLVNYAASSGAAEEVAHQIEASGGEAIVVKGDLSKIEDAQALIKAAVDRWGR
 VDALVNNAGITRDTLRLMKPEQWQQVIDVNLGVSFYASQAAAKVMAKARRGRIVNITSVVGLMGNAGQVNY
 ASSKAGVIGMTRSLARELAGRKVLVNAIAPGFIESEMTGVLNPEVEKAILAKIPLNSYGTDPQVAGLVKFL
 TDPAAAYITGQVLSIDGGMSMQ*

SEQ ID NO: 260

Prototheca moriformis Lipoate Synthase (LS) transcript

ACTCAATCGACCCGCACGCGAGTACGAAAAGATTCCATGGTGCGAAACGTTGAATGCCATAGTTTGTAAACGG
 GCCAGGCGTACGCGTCCGGGTGATTTGAGTTGACGTCCCGTAGCCCCGCTGGGCGGTACCTGAGTGAATTT
 GGAGCGAGCCGATCCGACGACGGGTACCTGTATTGAAACACCCGTATCATTACGGGAGTTTGTATTGC
 CAAACTTTTATAATTGAACCTGCCGACAGTACATTCGCTTGGAAGCATGTCGTCTCGGATCGACATGGGC
 GTGTGACGTCGCGGAGGTGGGCAACCCAATGACTGAGCGATCAGCCGCCCACCAATTGCGGGCGACCGCT
 GCACGCGTGCCAAGACGGCATGCACCTTCAAATCCGCCGCAAGGCATCGCGGGCCAGCGGCGATGCGCTCG
 GCGGCGTTTGACGAGCTGGTGGCGCTGGCCCGGGCCAAAGGACCCACCCCTGCCCCTGCCAAGCGCAAGCCG
 GCCAGCCCGCGCCCCAAGCCCCCAAGGCGCCCTGGCTGCGGCAGCGCGCCCCCAGGGCGAGCGCTACGAG
 TACCTCAAGCAGCAGCTGACGGGCTGAAGCTGGCGACCGTGTGCCAGGAGGCGCAGTGCCCCAACATCGGC
 GAGTGTGGTTCGGGCGGCGCCGAGGGGATCGGACGGCCACCATCATGGTGTCTGGGTGACACCTGCACGCGC
 GGCTGCCGCTTCTGCGCTGTCAAGACGTGCGGTGCGCCGCGCCGCGGACGCGGACGAGCCCGAGAACACG
 GCCGCGGCGGTGGCCGACTGGGGCGTGGGTACGTGGTGTGACCTCGGTGACCGCGCAGACCTGCCCGAC
 GGCGGCGCGGCTACTTTGCGGCCACGGTGCAGCGCATCAAGCGCCGGCGCCCCGAGGTCTGGTTCGAGTGC
 CTGACCCCGACTTTTTCGGGGGACGAGGCGCTCGTCCCGAGCTGGCGCGGTGCGGGCTAGACGTCTTTGCG
 CACAACGTGGAGACGGTGGCGCGCTGCAGCGGCGCGTGCCTGACCACCGCGCCGGCTACGAGCAGACGCTG

GCCGTGCTGCGCGCGGCCAAGCGCGCCTGCCCCCTCGCTGCTACCAAGTCCTCCATCATGCTCGGGCTCGGC
GAGCGCGACGAGGAGGTGAGCAGACCATGCGCGACCTGCGGGGGCGGGCGTGGACATCCTGACCCTGGGG
CAGTACCTGCAGCCCACGCACCGCCACCTGGAGGTCCAGGACATGGTCCACCCAGACGCCTTTGACCACTGG
CGAGAGGTTCGAGAAGCAATGGGGTTCCGGTACGTGGCCTCGGGTCCGCTGGTGCGGTCCTCGTACAGAGCC
5 GGCGAGTTTTACGTGGCAGCCATGCTCAGGGAAGGAAAGCGCGGCGGGGCGCAGTTGTGCGCCTGA
GGCGCGGTGGCAGTCTCTTTGAGGCTTTTGGGTGCGCAATCTCAAACACGTGTGTCTGTCTGCTGCGCGGGACC
GAGAGGCATCATATTCGCGACGAGCGTGCATCATCCGGATGCAAACACCTATTCTCGTATTCTAATCTTGTG
AATTATAGGCTGGCGTCACCAGCTGCTGAATCATGCGTTTTTGGAGGCACCGCCCCCTGAATATTTGGCGGGCC
10 ACTGCAACGCTGTAAGAATCTTCTTGCCCTTTTACCCGGAAGTTCAAATATCTCTGCGTAAACAGCAGGCGT
AGAAACAGAAACGAGCGACGCATGGTAGCGGGAGGAACGAAACAAGGCGCGCTGCTCAGGGGCAAGGGGCGAT
TCGTATCACAGATTCGCGCCAGCTCCCAATTTCCCAACCACTGCTAGATCTGGCGCGCGCCGCGGCTGTC
GAGGCGGACCACCTGACTGGAGTTGGTCTCCAGCCCGCGTGGCGGAGAAAGGCCTCGGACATGGCGCGCGC
GACGCGCTCGCCGGTGGCCGGGTCCGGCACGACCGCCACGGCC

SEQ ID NO: 261

Prototheca moriformis Lipoate Synthase (LS) protein

MSSRIDMGVSTSGRVGNPMTERSAAHQLRATAARVPRRHAPFKSAARHRGPAAMRSAAFDELVALARAKDPT
LPLPKRKPASPRPKPPKAPWLRQAPQGERYEYLKQQLTGLKLATVCQEAQCPNIGECWSGGAEGIGTATIM
VLGDTCTRGCRFCAVKTSRAPPPDADEPENTAAAVADWGVGVVLTSDRDDLDPDGAGHFAATVRAIKRR
RPEVLVECLTPDFSGDEALVAELARCLDVFAHNVTVARLQRRVRDHRAGYEQTLAVLRAAKRACPSLLTK
SSIMLGLGERDEEVEQTMRLRGAGVDILTGLQYLQPTRHLEVQDMVHPDAFDHWREVGEAMGFRYVASGP
LVRSSYRAGEFYVAAMLREGKRGGAGAQLSP*

SEQ ID NO: 262

Prototheca moriformis Homomeric Acetyl-CoA Carboxylase, Nucleotide

ATGACGGTGGCCAATCCCCCGGAAGCCCCGTTCGACAGCGAGGGTTCTCGCTGGCGCCCCGACAATGGGTCC
AGCAAGCCCACCAAGCTGAGCTCCACCCGGTCCTTGCTGTCCATCTCCTACCGGGAGCTCTCGCGTTCCAAC
CAGCAGTCGCGGTCTGTTCACTTCGCGGGGGGTGCCAGGGAGGACGAGCGTTTCGGATGAGCTGGAGCGCCGC
ATCCTCGAGTGGCAGGGCGATCGCGCCATCCACAGCGTGTGGTGGCCAACAACGGTCTGGCGGGCGGTCAAG
TTCATCCGGTTCGATCCGGTCTGTGGTCTGACAAGACGTTTGGGAACGAGCGTGCGGTGAAGCTGATCGCGATG
GCGACGCCGAGGACATGCGCGCGGACGCGGAGCACATCCGCATGGCGGACCAGTTTGTGGAGGTCCCCGGC
35 GGCAAGAACGTGCAGAACTACGCCAACGTGGGCTGATCACCTCGGTGGCGGTGCGCACCGGGGTGGACGCG
GTCCACTGCGGGGGGAGATCCGGTTCTGTGCGCACCGGGGTGGACGCGGTGTGGCCCGGTGGGGCCACGCG
TCCGAGTTCCCGGAGCTGCCCGAGTCTGCGGCGCCACGCCCTCGCAGATCCGGTTCTGTGGGCCCGCCGCG
GGGCCCATGGCGGGCCCTGGGCGACAAGGTGGGCTCCACCATCTGGCGCAGGCGGGCGGGCGTGCCACGCTG
GCCTGGTCCGGCTCGGGCGTGAGCATCGCCTACGCGGACTGCCCGCGCGGCGAGATCCCGCCCGAGGTCTAC
40 CGCCGCGCCTGCATCGACTCGCTGGAGGCGGCGCTGGCCTGCTGCGAGCGCATCGGCTACCCGGTTCATGCTC
AAGGCCAGCTGGGGCGGCGGCGGCAAGGGCATCCGCAAGGTGCTCAGCGCCGACGAGGTCAAGCTGGCCCTAC
ACCCAGGTGTCGGCGGAGTCCCGGCTCGCCGTGTTTCGCCATGAAGCTCGCGCCGAGTCCGCGCCACTG
GAGGTCCAGCTGCTCTGTGACGCGCACGGCCAGGTCTGCTCGCTCTACTCGCGCGACTGCTCCGTGCAGCGC
CGCCACCAGAAGGTGGTGGAGGAGGGGCGGTCACCGCGCGCGCGCCGAGGTGCTGGAGGGCATGGAGCGC
45 TGCGCGCGCTCCCTGGCGCGCGCCGTGGGTACGTGCGCGCCGCCACGGTCGAGTACCTCTACATGGTCGAG
ACGCGGGAGTACTGCTTCCCTGGAGCTCAACCCGCGCCTGCAGGTGGAGCACCCGGTGACGGAGATGATCACG
GGCGTGAACATCCCCGCCGCGCAGCTGCTCGTCGCGCGCGGCGTGGCCCTGCACCGCATCCCCGACATCCGC
AGGCTGTACCGCGCGCCGTGGCCGAGGACGGCCCCATCGACTTTGAGGACGACGCGCGCGCTGCCGCC
AACGGGCACGTGCTGGCGGTGCGCGTGACGGCCGAGAACGCGCACGACGGCTTCAAGCCCACCTCGGGCGCG
50 ATCGAGGAGATCAGCTTCCGCTCCACGCCCCGACGTCTGGGGCTACTTCTCGGTCAAGGGCGGCGAGCGCCGTG
CACGAGTTCGCGGACAGCCAGTTCGGCCACCTGTTGCGCGCGCGGCGACTCGCGCGAGGCGGCGCGTGGCGG
ATGGTGGTGGCGCTCAAGGAGATCAAGATCAGGGGCGAGATCCAGACGCCCGTTCGACTACGTGCGCGCATG
ATCCAGACCGACGACTTCTTGGGCAACCGCCACCATACGGGCTGGCTGGACGCGCGCATCGCGCGCGAGTAC
GGCGCGAGCGCCCTGGTTCTTGGTCTGCTCGCTCATCGCGGCGCGGCTGCTGCGCGCGACGTCGCGCGTGAAC
55 GAGCGCGCGGCGCCCTTTTGGACCACCTACGCAAGGGCCAGCTCCCGCCAGCGAGGTCCCGCTCACGCGC
GTGGCCGAGGAGTTTGTGGTGCACGGCACCAAGTACGCGGTGACGTACGCGCACGGGGCCGAGGCCCTAC
CGCGTCGCGCTGCGCGACCCCGAGGGCGACGGCGCCCCCGGGCACGCGCGCCGCGAGAGCCTCGCCTCCGCG
GCCGGCGCGGCCAGCTCCGTGACGTATCGCGCGCACGCTGCAGGACGGCGGCGCTGCTGCTGACAGGTGCGA
CGGGCGCGTGCACGTGTCACAGCGAGGAGGCGCTGGGCACGCGCCTGGTTCATCGACTCGGCGCCAGCTG
60 CCTGCTCAGCAACGACGACACCCCTCGCAGCTCGTGCCGTGACGACGGGCAAGCTCGTGCGCCACCTCGT
CGCCGAGGGCGAGCACGTGCGCGCGGACGAGCCCTACGCCGAGGTGAGGTGATGAAGATGATGATGACGCT
GCTCGCGCCCCGCGGCGCGCGGTGCGCTGGGAGGTGGCCGAGGGCGCGCGCTGGCGCGGGGCGCTGCTGC
TCGCGCGCCTGGAGGAAGCTGGGACGACCCCTCCCGCCGTGCGCCCGCGAGCCCTTCCGCGGCGCCTTCCCG
ACCTGGGCGCGCCCTCGCCGACTTTGACGGCCTGGCCGCGCGCTACCGCTCCGCGCTCGAGGCCGCGCAGG
65 ACGTGTCTGACGGCTTCGAGGGCGACGTGCCCCGCGTGTGCGCGCGCTGCTGGCCGCGCTGGACGACCCCG
CGCTTCCCTGCTGCTGGACGAGGCGCTGGCGGTGCTGCGCGCGGCTGCCGCGCGCTGCCGCGCGCTGGCGCG
GCCTGGAGGCCGCGGCTGGCCGAGGGCGCGCGGAGATGGCGGGCGAGCTCTCGGACGACGAGCAGGACGCC
GCGCTGGCGGCGCCGGGTGCGCGCGCAAGGGAGGGGCTCCACGCGCGCTTCTGGTCAAGCGCCGCGACCC
CGGCGGCGCGCCTGCAGCCGCGCGTCTTCGAGGAGGGCGCGCGCGCGCGCTGGCCTCGCCGCTCTCTG
70 CTCGAGCGCTACCTGGACGTGGAGGAGCGCTTCGAGTCCGGCGGCGCCAAGACGGACAGGAGGTTCATCGAC
GGCTGCGCGGCACGCACAGCGCCGACCCGGGCAAGGTGCTGGAGATCGTGGTTCGCGCACCAAGTTCGGCCGCG

CCGCGCGCCGACCTGGTGCACCGCCTGCTCAACGCGCTGGTGTGCCCTCGCCCGAGGCCTACCGCCCCATC
 CTGCGCCGCTGGCCGCGCTGACGGGCGCCGGCAGCGCCGCGCCCGCGCAGCGCGCGCGCCGCTGCTGGAG
 CACAGCCTGCTGGGCGAGCTGCGCGTGTGGTGTGCTGCGCGCGCNGTCCGGCCNTGGACATGGTTTCAGCGAC
 GCCGCGCTGCGCGGGGCTGTGCCTGGGCGAGTCGCGGGGCGAGCCCGTCTCGCCGCGCGCGGAGCTGGCCAGC
 5 GCGCTCGCGGCCGAGTCTCGGTGCGCTCGCCACCTCGCGGCGGTCCACGCTCGCGGAGGGGCTCTACGTC
 GGCTGGGCAACCTGGCGGTGCGGCCGCGAGCAGCGTGGAGGCGCGCATGGCCATGCTGGTTCGAGGCCCCC
 GCCGCGGTGCGACGACGCGCTGGCCACGCTGCTGGACCACCCGACCCGGTGCTCCAGCGCCGCGCTGTCC
 ACCTACGTCCGCGCATCTACTTCCCGGGCGTGTGCACGAGCCGAGGTGCTCGGGCGTGGGCCGCGCG
 10 CGGTCCGTCCGCGTCTGGGCGACGCGGGGCGCGCCCGGCGCGATGGCGCCGCGGCCAGCAGGACCAGCATG
 GCCGGGCTGCGCGCCGGGACGCTGCACGTGGTCTGACCGCCGAGGGCGCCGCGCGTGCAGCTGGACGCC
 GCCGCGCAGGCCGCGCTGGGCACGCTTGACGTCTCCGGTACGTGCGGCCGAGACACCAGCAGCAGCCGGGG
 CTGCGGACGCCCTCGTGTGCGGGCAGTGCCGCGCCGCGGTGTGCGCGCTGCGGCCGCGCTGCGCGAG
 GCCGGCTACAGCGCCGTCTCCTTCTCACCAGCGCGGCGGCGTGCAGCCCCCTGCGCGTGTGTTTTACGAC
 15 GGCACGTGAGCGGCAGCGTGCCGCGCCGACTCCTCGTCCACCCAACTGGCCATCGACCGGTGGTCCC
 TGGACCCCGTGTCTCGGCACGGTCGAGCCGCCGACGGCCGAGGCGTGGAGCTGGCCTGTCTGGCCGGCCGCG
 CGGGCGCCTCGCACGACGCTCGCGCAACCGTGGCACATGTGGACCGTGGCCGAGCGCGCCGGCAAGCGCAG
 CGCGGTGTGCGCCGACCTTTTTGCGCGGCGCGGTGCGCTCGTGGGCCGCGCGCGCTGTGAGCGCGGC
 CTACGCGGGAACCGCCGCTGTGCGGCGCGCGGTGCGCCGAGCTGGAGCAGCGTGGAGGCGCGCCGCC
 20 GCCGAGCTGGAGCGGCTGGGCAAGGGCCGCGTGGGCGGCGCCTCGTCCGACTGGACGACCGGTGCTTGTG
 GTGTGATGCTGCTGCGCTGGGCGCGTCCGAGCCGCGCGAAGGGAAAGGCGCGCTGCGCGCGCGCTGGC
 CGCGCCGCGGCCATCGCCTCGCGCCACGTGGCGGCCCTGCGCCGCGCGGCGTGGCGCAGTGGGAGGTCTGC
 CTGCGCACCGGCTCGGGCGGCGCCTCGCACCGCGAGGGCGGCTGGCGCGTGGTGTCTCTCGCCACGGGG
 CACGAGGCCGCGAGGCCCTTGTGGAGCTTACCGGAGGCGCGGACGGCACGCTGCGCGCGGTCAACCCC
 25 AGCCTGCGCGCCGCGCGGCGCGTGGACGCGCTCCGTGTGCTGCGCCCTACCCGAGCTGCGCGCGCTGCGCG
 CAGCGCCGCGTGCAGGCTCGGCGGCACAAGACCAGTACGCGTACGACTTCCCGGCCGTATTTCGAGGACGCG
 CTGCGCTCGATCTGGCTGCAGCGCGCGTGGAGCTGGGCGCCACCGGGTGGAGGCCGCGCGGAGGACCTG
 CTGCCGCCCGGACCGCCTGGTCACGGCCGAGCTGGTGTGAGACACGAGGAGGCGATCTACGAAAACGGC
 GCCGCGCACATTCGCCTGACCGACCGCGCGCCCTCGATGAACGACCTCGGCGTGTGGCCTGGCGCCTGACG
 30 CTGGCCACGCCCCGAGTGCCCGCGCGCGCGCGTGTGGCCATCGCCAACGACATACCTACAACCTCCGGC
 TCCTTCGCCCCGCGCGAGGACGCTTCTTCAAGGCGCGCGTACCCGAGTACGCGTGGCGGCGCGCTGCCGCTG
 GTCTACCTGGCGGCCAACTCCGGCGCGCGCTCGGGTGGCCGAGGAGGTCAAGCGCTGCGTGTGGAG
 TGGAGCGTGGCGGCGACCCGACCAAGGGCCACAAGTACCTCTACCTGGACGACGAGGACTACCGCTCCATC
 ACGGGCCGCGCGCGGCGCGCACGCTGCCCCGTAGCTGCTCGGCCAAGGTGCGCGCCGACGGGCGCACGCGG
 35 CACGTGCTGCACGACGTATCGGGCTCGAGGACGGGCTGGGCGTGCAGTGCCTGAGCGGGTCCGGGGCCATC
 GCCGGCGCCTTTGCGCGCGCCTTCCGCGAGGGCTTACCGTCACGCTCGTGTGGGGCGCACCGTGGGCGATC
 GCGCCTACCTACCTCGCGCGCTGGGCGGCGCTGCGTGCAGCGCCGCGAGCAGCCATCATCTCACCGGCTT
 GCCGCGCTCAACAAGCTGCTGGGGCGCGAGGTGTACACCTCRACGACGAGCTGGGCGGCCCGCGCGTATG
 GCGGCCAACGGCGTACGCCACCACGTGCTGGACGACGACCTGCAGGGCGTGCACGCCGTGCTGCGCTGGCTG
 40 GCCTACACGCGCCGCGCGCTGGGCGAGCTGCCGCCACGCTGCGCGCGGCCGACCCCGTSGACCGCGCGTG
 ACCTACGCGCCGCGCGGAGAACGAGAAGCTGGACCCGCGCCTGGCCGTGGCGGGCGGCGACGCGCCGAGCCC
 GTGACGGGCCCTTTCGACCGCGCGAGCTGGACCGAGGCGCAGGCGGGCTGGGCACAGACCGTGGTCAAGGG
 CGCGCTACCTCGCGCGCTGGGCGGCGTGGGCGTGGGCGTGGAGGTGAACGCGGTGCTGCTGCATCCCC
 45 GCCGACCCGGGCGATGCCGACTCGGCCGAGCGCACCATCCCGCAGGCGGGCCAGGTCTGGTTCCCGGACTCA
 GCGCTCAAGACCGCGCAGGCGATCGAGGAGTTTGGCCTGGAGGGCCTGCCCTCTTCATCTGGCCAACTGG
 CGCGGCTTCTCGGGCGGGCAGCGCGACCTGTTTCAGGGCGTGTGTCAGGCGGGCAGCCAGATCGTGGAGATG
 CTCCGCACCTACCGCGCCCGGTSACCGTCTACCTGGGCCCGGCTGCGAGCTGCGCGGCGGCGCCTGGGTG
 50 TGCTCGACTCCCATCAACCCGGCCAGCATCGAGATGTACGCCGACCCACCGCGCAGGCGCGCGCTGCTCGAG
 CCGCAGGGTGTCTGGAGATCAAGTTTCAGGACCCGAGCTCCTCGCCGCGATGCACCGGTGGACGAAAAG
 ATCATCGCGCTCAAGTCGGACGACTCGCCAGCGCGTGGCGGCCATCAAGGCGCGAGAAAGCGAGCTGCTG
 CCCGTCTACAGCCAGGTGCGGCACAGTTTCGCGCAGATGCACGACGGCCCCGTGCGCATGCTCGCCAAGGGC
 GTGCTGCGCGGCATCGTGCCCTGGTTCGGCCGCGCGCGCCTTCTGGCCACGCGCCTGCGCCGGCGCCTGGCC
 55 GAGGAGGCGTGTTCGCCAGATCGCGGCGGCCGACGCGTCCGTGAGACGCGCGACGCGTGGCCATGCTG
 CGTCTTGGTTCTCAGCTCGCCGCCACGGGCGGCGCGCCCGGCGCGCCCGGCGCGCTCGGCGCGCTGCTC
 AAGGAGCGGTGKVLSDADEVKLAYTQVCEVPGSPVFAMKLAPQSRHLEVQLLCDAHGQVCSLYSRDCSVQR
 CTGGACTGGAGCGAGGCCGAGGCCGCGCCTCGCGCGTGCAGCTCAAGAGCCTGCGCGTCAACGTC
 60 GCCATGCGCAGCGTCGACAGACTGTGCCAGACCCCGAGGGGACCGCCGGGCTCGTCAAGGGACTGGACGAG
 GCCATCAAGTCAACCCGTCCTCCTCTGCTCGCTCGCTGGTCAAGCCGTAG

SEQ ID NO: 263

Prototheca moriformis Homomeric Acetyl-CoA Carboxylase, Protein

MTVANPPEAPFDSESSSLAPDNGSSKPTKLSSTRSLLSISYRELSRSNQSRSFSTRGVPGRTDVSDELERR
 ILEWQGDRAIHSLVANGLAAVKFIRSIRWSYKTFGNRAVKLIAMATPEDMRADAEHIRMADQFVEVPG
 65 GKNVQNYANVGLITSVAVRTGVDVDCGGQIRFVRTGVDVWPWGHASEFPPELPESLGATPSQIRFVGPPA
 GPMAALGDKVGSTILAQAAGVPTLAWSGSGVSIAYADCPGEIPPEVYRRACIDSLEAALACCCERIGYPVML
 KASWGGGKGIKVLSDADEVKLAYTQVCEVPGSPVFAMKLAPQSRHLEVQLLCDAHGQVCSLYSRDCSVQR
 RHQKVVEEGPVTAAPPEVLEGMERCARSLARAVGYVGAATVEYLYMVETREYCFLELNPRLQVEHPVTEMIT
 GVNIPAAQLLVAAGVPLHRIPDIRRLYRAPLAEDGPIDFEDDAARLPNGHVLAVRVTAENAHDFKPTSGA
 70 IEEISFRSTPDVWGYFSVKGSASVHEFADSQFGLHFAFGDSREAAVRAMVVALKEIKIRGEIQTPVDYVARM
 IQTDDFLGNRHHTGWLDAIAAQVGAERPPWFLVVIAGAVLRATRAVNRAAFLDHLRKGQLPPSEVPLTR
 VAEFFVVDGTYAVDVTRTGPQAYRVALRDPEDGAPGHARRESLASAAGAASSVDVIARTLQDGGLLLQVR

RARARAAQRGGGAGHAPGHRLGHVPAQQRARPLAARGRQHGQARAPRRRRGRARARGRALRRGRGHEDDDDA
 ARAPRPARCAGRWPRAPRWRRRACCSRAWKLGPRSPRRAPASPSAAPSPTWARPRPTLTAWPRATAPRSRPRR
 TCSTASRATCPRSSPRCWPRWTTTPRSPSSCWTRRPWSRRGCRPRWPRAWRPRLAEGAAEMAGELSDDEQDA
 5 ALAAPGRRRAKGGASTRASWSSAATRRRACSRASSRRAAGARRWPRLLLLERYLDVEERFESGGAKTDQEV
 GLRGTHSADPGKVL EIVVAHQSA GPRADLVHRLNLALVLPSP EAYRPI LRRLAALTGAGSAAQAQRARLL
 HSLLGELRVLVAARAVRPWTWFSDAALRGLCLGESPGEPVSPRAELASALAAESSVASPTSRRSTLAEGLYV
 GLGNLAAAAASSVEARMAMLVEAPAAVDDALATLLDHPDVLQRRALSTYVRR IYFPGVLHEPQVVGRRAGA
 RSVRVWATRGAPGAMAPRPSRTSMAGLAPGTLHVVLTAEGAAALQLDAAAQAALGTL DVSGYVAPEHQQPG
 10 LPDASVVAASAAA AVSALAPALREAGYSAVSFLT KRGGVEPLRVVFDYDGTSSGSAAAPDSSSTQTGHRPAGP
 WTPCSARSSRRRPRRWSWPVWPAARAPRTT PRATVAHVDRGRARRQAQRGAAPHLFARRGALAGPAGAAERG
 LRGNPVAVAAAALAELEQTVEAAAAELERLGKGRVGGASSRLDAPCLSVLMLAAAGRVRAARRERRRRARAG
 RAAAIASRHVAALRRRAALAQWEVCLRTGSGGASHREGGWVRVVSSPTGHEAGEAFVDVYREAADGTLRAVNP
 SLRAPGPLDQSVLAPYPTLAPLQRRRLTARRHKTYYAYDFPAVFEDALRSIWLQRAVELGATGLEAAREDL
 15 LPPGTAWSRPELVLDTEEA IYENGAAH IRLTDRAPSMNDLGVVAVRLTLATPECPRGRAVVAIANDITYNSG
 SFGPREDAFFKAA TEYALAE RLPVVYLAANS GARVGLAE EVKRCLRV EWSVPGDPTKGHKYLYLDDEYRSI
 TGRAAGRTL PVSCSAKVGADGRTRHVLHDVIGLEDGLGVECLSGSGAIAGAFARAFREGFTVTLVSGRTVGI
 GAYLARLRRCVQRREQPI ILTGFAALNKLLGREVYTSQQQLGGPRVMGANGVSHHVDDDLQGVHAVLRWL
 AYTPARVGLPPTLRAADPVDRRVTYAPAENEKLDPR LAVAGGDAPEPVTGLFDRGSWTEAQAGWAQT VVTG
 20 RARLGGIPVGVVAVEVNAVSLHI PADPGMPDSAERTIPQAGQVWFPSALKTAQAIEEFGLEGLPLFILANW
 RGFSGGQRDLFEGVLQAGSQI VEMLR TYRRPVTYVLGPAASCAAAPGSCSTP INPASIEMYADPTAQGAVLE
 PQGVVEIKFRTPDLLAAMHRLDEKI IALKSDDSPSALAAI KARESELLPVYSQVAHQFAQMHDGPVRMLAKG
 VLRGIVPWSAARAFLATRLRRRLAE EALLRQIAAADASVEHADALAMLRSWFLSSPTGGAPGAPGALGALL
 KETVVAPPDAGEAPLALWQDDLAFLDWSEAEAGASRVALELKS LRVNVAMRSVDRLCQTPEGTAGLVKGLDE
 25 AIKSNPSLLLCLRSLVKP*

SEQ ID NO: 264

Prototheca moriformis Nitrogen Response Regulator (NRR1)

ATGGAGGAGCCGGACCAACAGGACCCTGCGGGCTCCGGGACCACCCAGGCGGAGGCTGAAAAGAGGAGTGCA
 30 TCCCCAAAGCCCGTCAGCGAGGCCACCAGCGACGGCAAGGCCTCCTCCGTGGCCACGGCGGCCGAAGCCAAG
 CCGGAAAGCTCTCCAGGCGAGGATGTGGCCGCCCTCTACCCGGCATCCCGCTCCTGCCGACCTGGGCCGG
 TTGCGGGACAACCGGCGCGACGTGCCCTGACCTGCCAGGTGGTGGGCTGCGGCGAGAGCCTCAGCGGGCAC
 GCCGAGTACTACCGGCGGTACCGCGTCTGCAAGCGCCACCTCAAGGGCTCGGCGCTGATCGTGACGGGGTG
 35 CCGCAGCGATTCTGCCAGCAGTGCAGTTCGCTTCCACCTGCTGGAGGCCTTTGACGGCGAGAAGAGAACTGC
 CGCGCCCGCTGGAGGAGCACAACTCGCGCGGGCGAAAGGT CATGGCGCCGGGGGGCGGCGGCCAGGTGACG
 GGCGCGCGCGACGCGCGGGAGTCGGAGCGGCACGGCCGCGCATGGACCTGCGCGGCGCGGGGCTCGAG
 GTCGCTCGGTGCGCCC

SEQ ID NO: 265

Prototheca moriformis Nitrogen Response Regulator (NRR1), protein

MEEPDQQDPAGSGTTQAEAEKRSASPKPVSEATSDGKASSVATAAEAKPESSPGEDVAALYPGIPVLPDLGR
 45 LRDNRDVP L TCQVVGCGESLSGHA EYRRYRVCKRHLKGSALIVDGVPQRFCQQCGRFHLLEAFDGEKRN
 RARLEEHNRRRKVMAPGGGGQVTGARATRRESERHGRGMDLRGAGLEVARSP

SEQ ID NO: 266

Prototheca moriformis Monoacylglycerol acyltransferase (MGAT1)

CGCAAAGGGTGTT CAGTGGTCGATGGGGACGAACTTGACGTCTCGTAGCCTGGGAAGCTGGGCGCGTAGCG
 50 TGCGAAGAGGTCTCCAGCGAGGCGTAGAACCGCGCGTGGACGTCTCGATCTGGGCCTGCGAGGGGGCCCTC
 TGGCGCCACCTTTGTGCGCGGGAGGATGGGCTTGCCGATCACAACCTGAGTCCTGTGGGGCGGGGAAGGG
 CGACACCTGGTAGCGCCCGCCGATGAGGTAGGGCACGGGGAAGCCAGGCGCTTGTAGGTCCAGCGCAGCAT
 GCCCCGACAGTTGACCAGGTTTTGGAGCGAGTGCAGCTCGCCAGCGCCAGCACCGGCACGAGCGCGGCTC
 55 GCTCTGCAGCGCCAGCCGACAAAGCCCTTGCTGGCTCGGGAACACGCAAACTCCCGCTCGGCGCGCGAT
 GCGGTGCGTGTGCACCAGCTCGGCCCTGGCCACCGGGGACGAGCAGCACGGCGCCCTTCTCGCGCAGCGCGG
 GAGGAAGGATCGCCGCGAGACGACGCGCAGGCCGAGCCAGGCCAGCAGGTGCGCGCAGCAAGGGCACCGCAAA
 CACGATCGAGGCGCAGAGCGTCACGGGGCGGATGCCGGGAGCAGGCGCCGGAAGCCGGGCGAGCGCGCAG
 GTAGCCGGCCGCGCAGCGGTACAGACCGTGCACGGCAAAGCCAAAGACGTGGCGCGCGCTCGGGGTCTACTG
 60 GTCCGCCCTCGCCGTGCGCAGGATCTCCAGCGACCTTACCCAGGCCGAGCACCTCTCCGAGCCGCGCC
 AAACAGTACCCGGAACAGCGGCCACTCCCGCGGCTGTCTCGCAAAGGGCCCCGCCAATCAGGGTGATGCT
 GTAGTACACAAAGGCCGAAGCGAGCAGGAGCGTCAGGGCAATCTCCAGCGGAAAGAGAAGAAAGTAGAGCGC
 AATGATTGTGCACAGCAACTGGACCGGGGCGTAGAGGAGGGTGATGACAAGGGCGTGGAGAGCCGGGTCCG
 TAGGTGCGTGGCGCGCAGCGCGCGCTCCGAGAGCGACGCGTGA GTCCGCGCACGAGCTCCATGGTCCGCAC
 65 GGGGCCCTTGTACAGCAACAGCGAGCCGGCCAGCACAGGTGGGAGAAGACCATGAAGATGCCTGCGGTCTG
 GGCAAAGCGCTGCGCGCTGTGCAGATGCCGATGGTGCTGTGGGCCATGAGGTAGATGATGAGGCTCAGCCC
 CGCGGAGTAGAGGCGCCACCCAGGGCTGCGTGGCCACGAACAGCCGCGCCCTTGAAGGGCTGGAAAAA
 GTGCCATGCGGCGGCGCGGCCCTCATCGCCGCCCTCGAGTCGCTCTCGCTGCGGGGCTCGGACTTGCGATG
 CACGTGCCGAGGTGGCCCCCAGCCCGTAGGTACAAAGGCGGCCATGGTCAAGCATGCGCCGAGAGCGC
 CAGGTACGCCAGCCGACGGGCGCGTCTCCAGCAGGTTGGCGCCAGGTGCAGCAGCGCGGCGCCAGCGC
 70 GACCCAGAGCGACCCATCCCGATCACGACGTACGAGGCCGCGGCGGACGAGCGCAGCGACTTGACGTCCCA
 CAAGACATCGGTGTGGCCCTGCTGGAGACCCGCGAGGATGCGCATGGGCTTGGCCGCGGGGTCAAAGACCAG

CACGGCAGACACCTGGAAGAGCTGGGAGGACACGGCTCCCACGGCCGCGCGGCCACCAGCAGCTCCACCGT
 CAGGCTAGACATCCAATCGTGCCGCGCGGCCGACAGAATGGGCACCACTACGGCGCAGCTCAGGAGCCAAGC
 TGAGAAGGACATGGCCTGGAAGACGAGGGTGCCTACCCCGCCATCGTGCCCCAGCTCCACCCGGGCGCGAC
 CTGGGTCTTGCCCCCTGCCCAGCTGCACGGCGCGGTGCAGCCATCGCGCCCCGCCAACGTAGGTGAGGGGCGT
 5 GGACACCGCGAACACCACCCAGCAGGCCGCCAGGCAAATGACCACGTGGTCGCCGAGAGGACCGACTCCGA
 GAGGGATGGCTCACCAGCTCATAGCAAGGGCCCTCGTCCGTGACACAGACTCATTGATGCACATCCCTG
 CTCCCCGCGTGGGAATGAGGTTGGGAAAAAGATGGGATCGACCTGGGAATGGAAGCGAACTGCCAGGAAGGC
 ATCGTCACCGGGGGTTCCGACAACCATATCGTCATTGATGCATCGGGGCTCCGGGGAGGCGTCTGCCGCCG
 10 GACATGGGAGCGATCAGGCGACCGGAGTGTTTAAGCGAGCACACGACCAGCACCCTCGTTGAGCCTCGAAAG
 TGATGCAAA

SEQ ID NO: 267

Prototheca moriformis Monoacylglycerol acyltransferase (MGAT1), protein

MTSVSHSLSESVLSDHVVICLAACWVVFVAVSTPLTYVGGARWLHRAVQLGRGKTQVRPGWSWGTMGGVRTL
 VFQAMSFSAWLLSCAVVVPILSAARHDWMSLTVELLVAAAAGAVSSQLFQVSAVLVDFPAAKPMRILAGL
 QQGHTDVLWDVKSLSAAASYVVI MGSLWVALGAALLHLGANLLEDAPVRLAYLALSAGCLTMAAFVITYG
 LAGHLRHVHRKSEPGSESDSEGGDEAAPAAWHFFQPFKGGGWVATQALGWALYSAGLSLI IYLMHSTIGI
 VHSARFAQTAGIFMVFSHLVLAGSLLLYKGPVRTMELVRGLSASLSERALRPTDLRLRLSDALVITLLYAP
 20 VQLLCTIIALYFLLFPLEIALTLLASAFVYYSITLIGGPFGETGRREWPLFRDWFGRRSEEVLPWVGSLE
 ILRDGEADQYDPERRHVFGFAVHGLYPLAAGYLPLLPGFRLLPGIRPVTLCASIVFAVPLLRDLLAWLGLR
 VVSRRSFLRALREKGAVALLVPGGQAEVHTHRIARPTREFCVFRSHKGFVRLALQSEALVPVLALGEVDSL
 QNLVNLPGMLRWITYKRLGFPVYLIIGGRYQVSPFPRPTGLRFVIGKPILPATKVAPEGPSQAQIDDVHARFY
 ASLEDLFARYAPSPFGYEDVKFVPIDH*

SEQ ID NO: 268

Prototheca moriformis Cellulase, Endoglucanase (EG1)

CTCCTCCCAGCCACCGTCCACCAAATCCCACCTCTCGCCCGCTCAGACGTTGATGTAGGTGCCGAGATGTC
 30 GGAGCGGCGGAAGATGCCGTACTGCTGCAGGCACTGCTCCCAGAGCCCAGAGGGCAGCTTGCTGGCGCCGGC
 CAGGGCGCCGATCAGGCCCGCGTTGTTCTCCACGCCACGCTAGGCCGCGGTCTGTGCGCGGTGTCCACAAA
 GTTGTCCGAGCTGGCCGTGCCGTACAGAGCGCGCCCGTGAGCGTGTGCGTGTGCGGGTCCGGCGACAGCAG
 ACCGGTCACGCGGTTGCACACCTCTGGCTCGGCCGGGCAGGCCGCGTCTGCGGTCTGCGTGTGCGCTTGGGCGG
 GTTCTTGCCAAAGCCGACGACGAGCGAGCGCCCCGAGTCTGCCCAGCACGTAGCGGATCTGGCTCAGGCCCA
 35 GCACTGGTACTCAAACACGCGCGTTCGACCCGGGGTCTGTTAGCGAGTCCGCGTAGGCCAGCGCCGCAAAGGC
 CGTGTTTATGGTGGAGCCCAGCGCGCGCTCATGGGGTGTAGGCGCGGCCCTTGAGCGTGTAGTTGGCCGC
 CAGGTCCGAGCAGATCCAGGACTTGAGGAACAGCTCCGACTGGCTCTGGAAGGTACGCGCGCCCGTCTCCTG
 CGCCAGCAGCACGTTGGTTCGGCCAGAACACGTTGTCCAGTCAAACGCGTACTTGAAGTCGCTCACCAGCA
 CTCGTACTCCAGGTGGTTCACGTACTGGCTGTACATGTTGGTCAGGTACTGCTGCTCCCCCGTTCGCTTGTA
 40 CATCCAGCTGCTCGACCACGCCAGGTCTGCTGTAACAGTCTGCTCGAGTTGTACACCTTGGAATGTTCCAGTC
 CTGCACCGAGTAGTGGCCCAGGTCTGGTGGAGGCGATGTGCTACACGGCCTTGGCCTTGGTCAGCAGCTCCGT
 CGCGTAGGCCGCGTTCGCCGTGGCGTTGAACACGATGGAGGCCGAGGCCAGCGCCCGCGACCTGGCCGCC
 CAGGTCTGCTCGCGCCCTTGGCCATGTCCACCAGTACGCGGGCCGGTCCGTGGTGTAGTCTCCAGACGGTA
 45 CCAGTTGAGTCTCTCTGCTGGATGTTGCCAATCTGGACAGAATGTAGGTGGTGTGCTCGTCTGCTGTCAC
 CGTCTTGAGCAGGTAGTCCGTGGACCAGCGCAGGTTGTCCAGCAGCGACGCGGAGCTGCCCGCGTCCCCGAT
 GGCTTCGCCGTAGGTCTGGTAGGCCAGGACATGAGCGTTCGCCGTAAAGGCGATGGGCAGCGTGGCCTTGAA
 GTTGCCCGGCCCATGCCCGTCACGTAGCCGCCCTCCAGGTCCAGGTTCGCGCCAGCTTGACCGCTGCGG
 CACCGACTTGGTTCGACGAGGAGGGGCTGATCGTGAAGGCGGCGGCGAGGTTGGGGATGAGCGACGAAGGCGA
 50 GGAGCCCGGGTGTGGGCGACGAGCCCGCCACGGGCGACACACCAGGTTGGCAAACGG

SEQ ID NO: 269

Prototheca moriformis Cellulase, Endoglucanase (EG1), protein

PFANLVWSPVAGSSPSTPGSSPSSLIPTLPPPSTISPSSSTKSVPQAVKLGATWDLEGGYVTGMAAGNFKAT
 55 LPIAFTATLMSWAYQTYGEAIGDAGSSASLLDNLRWSTDYLLKTVQTTSNNTYILSRIGNIQEELNWRLE
 DYTDRPAYVVDMAKGASDLGGQVAAALASASIVFNASGDAAAYATELLTKAKAVYDIASDGLHYSVQDWN
 SKVYNSSTFYDDLAWSSSWMYKATGEQYLTNMYSQYVNHLEYESSVSDFKYAFDWDNVFWPTNVLLAQETG
 GVTFSQSSEFLKSWICSDLAANYTLKGRAYNPMSAPLGSTMNTAFALAYADSLNDPGSTRVFEYQCWGLS
 60 QIRYVLGDSGRSLVVGFGKNPPKRTQDRDAACPAEPEVCNRVTGLLSPDPDTHLTGALVYGTASSDNFVDT
 RDTTAAAYGVENNAGLIGALAGASKLPSGLWEQCLQQYGFRRSDICGTIINV*

SEQ ID NO: 270

P. moriformis KAS II

MQTAHQRPPTTEGHCFGARLPASRRRAVRRRAWRSRIARAAAAADANPARPERRVVITGQGVVTSLGQTIEQFYS
 65 SLLEGVSGISQIQKFDTTGYTTTIIAGEIKSLQLDPYVPKRWAKRVDDVIKYVYIAGKQALESAGLP IEAAGL
 AGAGLDPALCGVLIGTAMAGMTSFAAGVEALTRGGVVRKMNPF CIPFSISNMGGAMLAMDIGFMGPNY SISTA
 CATGNYCILGAADHIRRGDANVMLAGGADA AII PSIGGFIAKALSKRNDEPERASRPWDADRDFVMGEG
 AGVLVLEELEHAKRRGATILAE LVGAATSDAHMTEPD PQRGVRLCLERALERARLAPERVGVYVNAHGTS
 70 TPAGDVAEYRAIRAVIPQDSLRLNSTKSMIGHLLGGAGAVEAVAAIQALRTGWLHPNLNLENPAPGVDPVVL
 VGPRKERAEDLDVVLNSNFGFGGHNSCVIFRKYDE

SEQ ID NO: 271

P. moriformis stearoyl ACP desaturase

5 MASAVTFACAPPRGAVAAPGRRRAASRPLVVRAVASEAPLGVPSPVQRPSPVVYSKLDKQHRLTPERLELVQS
MGQFAEERVLPLVHPVDKLWQPQDFLPDPESPDPFEDQVAELRARAKDLPDEYFVVVLVGDMITEEALPTYMAM
LNTLDGVRDDTGAADHPWARWTRQWVAEENRHGDLLNKYCWLTTGRVNMRAVEVTINNLIKSGMNPQTDNNPY
LGFVYTSFQERATKYSHGNTARLAAEHGDKNLSKICGLIASDEGRHEIAYTRIVDEFFRLDPEGAVAAAYANM
10 MRKQITMPAHLMDDMGHGEANPGRNLFADFSAVAEEKIDVYDAEDYCRILEHLNARWKVDERQVSGQAAADQE
YVLGLPQFRKLAEKTAARKRVARRPVAFSWISGREIMV

SEQ ID NO: 272

Prototheca moriformis d12 desaturase allele 1

15 MAIKTNRQPVEKPPFTIGTLRKAIPAHCFERSALRSSMYLAFDIAVMSLLYVASTYIDPAPVPTWVKYGVMM
PLYWFFQGAFTGVWVCAHECGHQAFSSSQAINDGVLVFNHSLLLVPYYSWKHSHRRHHSNTGCLDKDEVFV
PPHRAVAHEGLEWEEWLPIRMKGVLVTLTLGWPLYLMFNVASRPYPRFANHFDWPSPIFSKRERIEVVISDL
ALVAVLSGLSVLGRMTGWAWLVKTYVVPYLIIVNMWLVLTLLQHTHPALPHYFEKDWDWLRGAMATVDRSMG
20 PPFMDNILHHISDTHVLHHLFSTIPHYHAEESAAIRPILGKYYQSDSRWVGRALWEDWRDCRYVVPDAPED
DSALWFHK

SEQ ID NO: 273

Prototheca moriformis delta 12 desaturase allele 2

25 MAIKTNRQPVEKPPFTIGTLRKAIPAHCFERSALRSSMYLAFDIAVMSLLYVASTYIDPAPVPTWVKYGIMW
PLYWFFQGAFTGVWVCAHECGHQAFSSSQAINDGVLVFNHSLLLVPYYSWKHSHRRHHSNTGCLDKDEVFV
PPHRAVAHEGLEWEEWLPIRMKGVLVTLTLGWPLYLMFNVASRPYPRFANHFDWPSPIFSKRERIEVVISDL
ALVAVLSGLSVLGRMTGWAWLVKTYVVPYMIIVNMWLVLTLLQHTHPALPHYFEKDWDWLRGAMATVDRSMG
30 PPFMDNILHHISDTHVLHHLFSTIPHYHAEESAAIRPILGKYYQSDSRWVGRALWEDWRDCRYVVPDAPED
DSALWFHK

WHAT IS CLAIMED IS:

1. A recombinant nucleic acid comprising a coding sequence that encodes a *Prototheca* lipid biosynthesis protein, provided that the protein is not ketoacyl acyl carrier protein synthase II, stearoyl acyl carrier protein desaturase, or delta 12 fatty acid desaturase.
2. The recombinant nucleic acid of claim 1 wherein the lipid biosynthesis protein is selected from a protein in Table 1.
3. The recombinant nucleic acid of claim 2 wherein the protein has at least one point mutation in comparison to a protein in Table 1.
4. The recombinant nucleic acid of claim 1 wherein the protein has at least 50%, 60%, 70%, 80%, 85%, 90% or 95% sequence identity to a protein in Table 1.
5. The recombinant nucleic acid of any one of claims 1 to 4, wherein the coding sequence is in operable linkage with a promoter.
6. The recombinant nucleic acid of any one of claims 1 to 5, wherein the coding sequence is in operable linkage with an untranslated control element.
7. The recombinant nucleic acid of any one of claims 1 to 6, wherein the coding sequence is in operable linkage with a targeting sequence.
8. The recombinant nucleic acid of claim 7, wherein the targeting sequence is a transit peptide selected from the group of plastidial targeting sequence and mitochondrial targeting sequence.
9. The recombinant nucleic acid of any one of claims 1 to 8, wherein the nucleic acid is a DNA molecule.
10. The recombinant nucleic acid of any one of claims 1 to 9 that encodes a functional *Prototheca* lipid biosynthesis protein.
11. The recombinant nucleic acid of claim 3 wherein the protein is diacylglycerol diacyltransferase (DGAT) having at least one point mutation.
12. The recombinant nucleic acid of any one of claims 1 to 11 that further encodes a sucrose invertase.
13. The recombinant nucleic acid of any of one of claims 1 to 12 that further encodes an inhibitory RNA that suppresses expression of a *Prototheca* lipid biosynthesis gene.
14. An expression cassette comprising a recombinant nucleic acid of any one of claims 1 to 13.

15. A genetically engineered microbial cell transformed with a recombinant nucleic acid of any one of claims 1 to 13.
16. The genetically engineered microbial cell of claim 15, wherein the cell is a cell of the genus *Prototheca* or *Chlorella*.
- 5 17. The genetically engineered microbial cell of claim 16, wherein the cell is a cell of the genus *Prototheca*.
18. The cell of claim 17, wherein said cell comprises both an endogenous lipid biosynthesis gene and one or more exogenous *Prototheca* lipid biosynthesis gene selected from the genes listed in Table 1.
- 10 19. The cell of claim 18, wherein said exogenous gene encodes a lipid biosynthesis protein and wherein the amino acid sequence of the lipid biosynthesis protein is identical to the endogenous lipid biosynthesis protein.
20. The cell of claim 19, wherein said exogenous gene comprises a nucleotide sequence in which the codons of the nucleotide sequence encoding the amino acids of the lipid biosynthesis protein have been altered.
- 15 21. The cell of any one of claims 15 to 20, wherein said exogenous gene encodes a protein with at least 50% amino acid identity to a native *Prototheca* protein.
22. The cell of any of claims 15 to 21, wherein said exogenous gene is in operable linkage with a promoter element that is not a native *Prototheca* promoter.
- 20 23. The cell of any of claims 15 to 22, wherein said exogenous gene is in operable linkage with an untranslated control element that is not a native *Prototheca* untranslated control element.
24. The cell of any of claims 15 to 23 wherein said exogenous gene is in operable linkage with a nucleotide sequence encoding a transit peptide that is not a native *Prototheca* transit peptide.
- 25 25. The cell of claim 24, wherein the transit peptide is a plastidial targeting sequence or a mitochondrial targeting sequence.
26. The cell of any of claims 15 to 25, wherein the cell has a 23S rRNA sequence with at least 75% nucleotide identity to SEQ ID NO: 5.
- 30 27. The *Prototheca* cell of any one of claims 15 to 26, wherein the cell has a fatty acid profile that is at least 10% C8-C14.
28. The *Prototheca* cell of any one of claims 15 to 26, wherein the cell has a fatty acid profile that is at least 50% C12.

29. The *Prototheca* cell of any one of claims 15 to 26, wherein the cell has a fatty acid profile that is at least 70% C18:1.
30. A method for obtaining microbial oil comprising culturing a genetically engineered *Prototheca* cell of any one of claims 15 to 29 under conditions such that oil is produced.
31. A microbial oil produced by the method of claim 30.
32. A genetically engineered cell of the genus *Prototheca* wherein the activity of one or more endogenous lipid biosynthesis gene selected from the genes listed in Table 1 has been attenuated.
33. The cell of claim 32, wherein the activity of the endogenous gene has been attenuated through chromosomal gene deletion, chromosomal gene insertion, frameshift mutation, point mutation, or inhibitory RNA.
34. The genetically engineered cell of claims 32 or 33, further comprising an exogenous *Prototheca* lipid biosynthesis pathway gene selected from the genes listed in Table 1.
35. The genetically engineered cell of any one of claims 32 to 34, wherein one or more allele of an endogenous lipid biosynthesis gene is attenuated.
36. The genetically engineered cell of any one of claims 32 to 35, wherein one allele of the endogenous lipid biosynthesis gene is replaced with an exogenous *Prototheca* lipid biosynthesis pathway gene selected from Table 1 and a selectable marker.
37. The genetically engineered cell of any one of claims 32 to 36 wherein two or more alleles of the endogenous lipid biosynthesis gene are replaced with an exogenous *Prototheca* lipid biosynthesis pathway gene selected from Table 1 and a selectable marker.
38. The genetically engineered cell of any one of claims 15 to 37 wherein the cell has a 23S rRNA sequence with at least 75% nucleotide identity to SEQ ID NO:5.
39. The *Prototheca* cell of any of one of claims 32 to 35 wherein the cell has a fatty acid profile that is at least 10% C8-C14.
40. A method for obtaining microbial oil comprising culturing a genetically engineered *Prototheca* cell of any one of claims 32 to 39 under conditions such that oil is produced.
41. A microbial oil produced by the method of claim 40.