with shorter chain length and a higher degree of saturation and without pigments, with greater yield and efficiency. The present invention meets this need.

SUMMARY OF THE INVENTION

[0007] In a first aspect, the present invention refers to a codon-optimized recombinant nucleic acid comprising a fatty acyl-ACP thioesterase coding sequence codon-optimized for expression in *Prototheca*, wherein the coding sequence contains the most or second most preferred codon of Table 1 for at least 60% of the codons of the coding sequence, such that the codon-optimized sequence is more efficiently translated in *Prototheca* than a native sequence encoding the fatty acyl-ACP thioesterase, the fatty acyl-ACP thioesterase having hydrolysis activity towards one or more fatty acyl-ACP substrates, the recombinant nucleic acid further comprising a plastid targeting sequence from microalgae that is in-frame with the codon-optimized fatty acyl-ACP thioesterase coding sequence. In another aspect, the present invention refers to a codon-optimized recombinant nucleic acid comprising a fatty acyl-ACP thioesterase coding sequence codon-optimized for expression in *Prototheca*, wherein the coding sequence contains the most or second most preferred codon of Table 1 for at least 60% of the codons of the coding sequence, such that the codon-optimized sequence is more efficiently translated in *Prototheca* than a native sequence encoding the fatty acyl-ACP thioesterase, the fatty acyl-ACP thioesterase having hydrolysis activity towards one or more fatty acyl-ACP substrates, the recombinant nucleic acid further comprising a plastid targeting sequence in-frame with the codon-optimized fatty acyl-ACP thioesterase coding sequence, the recombinant nucleic acid further comprising first and second flanking nucleic acid sequences endogenous to the nuclear genome of a cell of the genus *Prototheca*, wherein the first and second flanking sequences are each at least 200 nucleotides long and the first and second flanking sequences flank the codon-optimized recombinant nucleic acid for integration of the recombinant nucleic acid into the nuclear genome by homologous recombination with a corresponding genomic sequence. In still another aspect, the present invention refers to a codon-optimized recombinant nucleic acid comprising a fatty acyl-ACP thioesterase coding sequence codon-optimized for expression in *Prototheca*, wherein the coding sequence contains the most or second most preferred codon of Table 1 for at least 60% of the codons of the coding sequence, such that the codon-optimized sequence is more efficiently translated in *Prototheca* than a native sequence encoding the fatty acyl-ACP thioesterase, the fatty acyl-ACP thioesterase having hydrolysis activity towards one or more fatty acyl-ACP substrates, the recombinant nucleic acid further comprising a plastid targeting sequence from microalgae that is in-frame with the codon-optimized fatty acyl-ACP

thioesterase coding sequence, and wherein the amino acid sequence of the fatty acyl-ACP thioesterase comprises an amino acid sequence of a fatty acyl-ACP thioesterase from an organism selected from the group consisting of *Cuphea hookeriana*, *Umbellularia californica*, *Cinnamomum camphora*, *Cuphea palustris* and *Ulmus americana*. In another aspect, the present invention refers to a codon-optimized recombinant nucleic acid comprising a fatty acyl-ACP thioesterase coding sequence and a sucrose invertase coding sequence, each coding sequence codon-optimized for expression in *Prototheca*, wherein each coding sequence contains the most or second most preferred codon of Table 1 for at least 60% of the codons of the coding sequence, such that each codon-optimized sequence is more efficiently translated in *Prototheca* than a native sequence encoding the fatty acyl-ACP thioesterase or the native sequence encoding the sucrose invertase, the fatty acyl-ACP thioesterase having hydrolysis activity towards one or more fatty acyl-ACP substrates, the recombinant nucleic acid further comprising a plastid targeting sequence in-frame with the codon-optimized fatty acyl-ACP thioesterase coding sequence. In another aspect, the present invention refers to a host cell of the genus *Prototheca* comprising a recombinant nucleic acid comprising a fatty acyl-ACP thioesterase coding sequence codon-optimized for expression in *Prototheca*, wherein the coding sequence contains the most or second most preferred codon of Table 1 for at least 60% of the codons of the coding sequence, such that the codon-optimized sequence is more efficiently translated in *Prototheca* than a native sequence encoding the fatty acyl-ACP thioesterase, and wherein a plastid targeting sequence is in-frame with the codon-optimized fatty acyl-ACP thioesterase coding sequence and the fatty acyl-ACP thioesterase has hydrolysis activity towards one or more fatty acyl-ACP substrates. In another aspect, the present invention refers to a host cell of the genus *Prototheca* comprising a codon optimized recombinant nucleic acid comprising a fatty acyl-ACP thioesterase coding sequence codon-optimized for expression in *Prototheca*, wherein the coding sequence contains the most or second most preferred codon of Table 1 for at least 60% of the codons of the coding sequence, such that the codon-optimized sequence is more efficiently translated in *Prototheca* than a native sequence encoding the fatty acyl-ACP thioesterase, the fatty acyl-ACP thioesterase having hydrolysis activity towards a fatty acyl-ACP substrate of chain length C8, C10, C12 or C14, the recombinant nucleic acid further comprising a plastid targeting sequence that is in-frame with the codon-optimized fatty acyl-ACP thioesterase coding sequence. In still another aspect, the present invention refers to a host cell of the genus *Prototheca* comprising the recombinant nucleic acid of the present invention. In one other aspect, the present invention refers to a codon-optimized recombinant

nucleic acid comprising a fatty acyl-ACP thioesterase coding sequence codon-optimized for expression in *Prototheca*, wherein the coding sequence contains the most or second most preferred codon of Table 1 for at least 60% of the codons of the coding sequence, such that the codon-optimized sequence is more efficiently translated in *Prototheca* than a native sequence encoding the fatty acyl-ACP thioesterase, the fatty acyl-ACP thioesterase having hydrolysis activity towards one or more fatty acyl-ACP substrates, further comprising a promoter comprising the nucleotide sequence of any of SEQ ID NOS: 91-102. The disclosure provides cells of the genus *Prototheca* comprising an exogenous gene, and in some examples the cell is a strain of the species *Prototheca moriformis, Prototheca krugani, Prototheca stagnora* or *Prototheca zopfii* and in other example the cell has a 23S rRNA sequence with at least 70, 75, 80, 85 or 95% nucleotide identity to one or more of SEQ ID NOs: 11-19. In some cells the exogenous gene is coding sequence and is in operable linkage with a promoter, and in some examples the promoter is from a gene endogenous to a species of the genus *Prototheca*. In further examples the coding sequence encodes a protein selected from the group consisting of a sucrose invertase, a fatty acyl-ACP thioesterase, a fatty acyl-CoA/aldehyde reductase, a fatty acyl-CoA reductase, a fatty aldehyde reductase, a fatty aldehyde decarbonylase, an acyl carrier protein and a protein that imparts resistance to an antibiotic. Some examples of a fatty acyl-ACP thioesterase that has hydrolysis activity towards one or more fatty acyl-ACP substrates of chain length C8, C10, C12 or C14, including acyl-ACP thioesterases with at least 50, 60, 70, 80, or 90% amino acid identity with one or more sequences selected from the group consisting of SEQ ID NOs: 59, 61, 63 and 138-140. In further examples the coding sequence comprises a plastid targeting sequence from microalgae, and in some examples the microalgae is a species of the genus *Prototheca* or *Chlorella* as well as other genera from the family *Chlorellaceae*. In some examples the plastid targeting sequence has at least 20, 25, 35, 45, or 55% amino acid sequence identity to one or more of SEQ ID NOs: 127-133 and is capable of targeting a protein encoded by an exogenous gene not located in the plastid genome to the plastid. In other examples the promoter is upregulated in response to reduction or elimination of nitrogen in the culture media of the cell, such as at least a 3-fold upregulation as determined by transcript abundance in a cell of the genus *Prototheca* when the extracellular environment changes from containing at least 10mM or 5mM nitrogen to containing no nitrogen. In further examples the promoter comprises a segment of 50 or more nucleotides of one of SEQ ID NOs: 91-102. In other examples the

cell has a 23S rRNA sequence with at least 70, 75, 80, 85 or 95% nucleotide identity to one or more of SEQ ID NOs: 11-19. In other examples the exogenous gene is integrated into a chromosome of the cell.

[0008] In additional examples of cells as described herein, the cell is of the genus *Prototheca* and comprises an exogenous fatty acyl-ACP thioesterase gene and a lipid profile

of at least 4% C8-C14 of total lipids of the cell, an amount of C8 that is at least 0.3% of total lipids of the cell, an amount of C10 that is at least 2% of total lipids of the cell, an amount of C12 that is at least 2% of total lipids of the cell, an amount of C14 that is at least 4% of total lipids of the cell, and an amount of C8-C14 that is 10-30%, 20-30%, or at least 10, 20, or 30% of total lipids of the cell. In some examples the cell further comprises an exogenous sucrose invertase gene. In some examples the cell is a strain of the species *Prototheca moriformis, Prototheca krugani, Prototheca stagnora* or *Prototheca zopfii*, and in other example the cell has a 23S rRNA sequence with at least 70, 75, 80, 85 or 95% nucleotide identity to one or more of SEQ ID NOs: 11-19. In other examples the exogenous fatty acyl-ACP thioesterase gene is integrated into a chromosome of the cell. Other examples described herein comprise methods of making triglyceride compositions of a lipid profile of at least 4% C8-C14 w/w or area percent of the triglyceride composition, an amount of C8 that is at least 0.3% w/w or area percent, an amount of C10 that is at least 2% w/w or area percent, an amount of C12 that is at least 2% w/w or area percent, an amount of C14 that is at least 4% w/w or area percent, and an amount of C8-C14 that is 10-30%, 20-30%, or at least 10, 20, or 30% w/w or area percent. The disclosure also comprises methods of making triglyceride compositions comprising cultivating the foregoing cells, wherein the cells also comprise an exogenous gene encoding a sucrose invertase and sucrose is provided as a carbon source. In some examples the sucrose invertase has at least 50, 60, 70, 80, or 90% amino acid identity to one or more of SEQ ID NOs: 3, 20-29 and 90.

[0009] Examples described herein include triglyceride oil compositions as well as cells containing triglyceride oil compositions comprising a lipid profile of at least 4% C8-C14 and one or more of the following attributes: 0.1-0.4 micrograms/ml total carotenoids, less than 0.4 micrograms/ml total carotenoids, less than 0.001 micrograms/ml lycopene; less than 0.02 micrograms/ml beta carotene, less than 0.02 milligrams of chlorophyll per kilogram of oil; 0.40-0.60 milligrams of gamma tocopherol per 100 grams of oil; 0.2-0.5 milligrams of total tocotrienols per gram of oil, less than 0.4 milligrams of total tocotrienols per gram of oil, 4-8 mg per 100 grams of oil of campesterol, and 40-60 mg per 100 grams of oil of stigmasterol. In some examples described herein the triglyceride oil compositions have a lipid profile of at least 4% C8-C14 w/w or area percent of the triglyceride composition, an amount of C8 that is at least 0.3% w/w or area percent, an amount of C10 that is at least 2% w/w or area percent, an amount of C12 that is at least 2% w/w or area percent, an amount of C14 that is at least 4% w/w or area percent, and an amount of C8-C14 that is 10-30%, 20-30%, or at least 10, 20, or 30% w/w or area percent. In other examples the triglyceride oil composition is

blended with at least one other composition selected from the group consisting of soy, rapeseed, canola, palm, palm kernel, coconut, corn, waste vegetable, Chinese tallow, olive, sunflower, cotton seed. chicken fat, beef tallow, porcine tallow, microalgae, macroalgae, *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, jatropha, macadamia, Brazil nuts, avocado, petroleum, or a distillate fraction of any of the preceding oils.

[0010] Methods of the disclosure also include processing the aforementioned oils of by performing one or more chemical reactions from the list consisting of transesterification, hydrogenation, hydrocracking, deoxygenation, isomerization, interesterification, hydroxylation, hydrolysis to yield free fatty acids, and saponification. The disclosure also describes hydrocarbon fuels made from hydrogenation and isomerization of the aforementioned oils and fatty acid alkyl esters made from transesterification of the aforementioned oils. In some examples the hydrocarbon fuel is made from triglyceride isolated from cells of the genus *Prototheca* wherein the ASTM D86 T10-T90 distillation range is at least 25°C. In other examples the fatty acid alkyl ester fuel is made from triglyceride isolated from cells of the genus *Prototheca*, wherein the composition has an ASTM D6751 A1 cold soak time of less than 120 seconds.

[0011] The disclosure also describes a composition comprising (a) polysaccharide comprising one or more monosaccharides from the list consisting of 20-30 mole percent galactose; 55-65 mole percent glucose; and 5-15 mole percent mannose; (b) protein; and (c) DNA comprising a 23S rRNA sequence with at least 70, 75, 80, 85 or 95% nucleotide identity to one or more of SEQ ID NOs: 11-19; and (d) an exogenous gene. In some examples the exogenous gene is selected from a sucrose invertase and a fatty acyl-ACP thioesterase, and in further examples the composition further comprises lipid with a lipid profile of at least 4% C8-C14. In other examples the composition is formulated for consumption as an animal feed.

[0012] The disclosure describes recombinant nucleic acids encoding promoters that are upregulated in response to reduction or elimination of nitrogen in the culture media of a cell of the genus *Prototheca*, such as at least a 3-fold upregulation as determined by transcript abundance when the extracellular environment changes from containing at least 10mM or 5mM nitrogen to containing no nitrogen. In some examples the recombinant nucleic acid comprises a segment of 50 or more nucleotides of one of SEQ ID NOs: 91-102. The

disclosure also describes nucleic acid vectors comprising an expression cassette comprising (a) a promoter that is active in a cell of the genus *Prototheca*; and (b) a coding sequence in operable linkage with the promoter wherein the coding sequence contains the most or second most preferred codons of Table 1 for at least 20, 30, 40, 50, 60, or 80% of the codons of the coding sequence. In some vectors the coding sequence comprises a plastid targeting sequence in-frame with a fatty acyl-ACP thioesterase, including thioesterase that have hydrolysis activity towards one or more fatty acyl-ACP substrates of chain length C8, C10, C12 or C14. Some vectors include plastid targeting sequences that encode peptides that are capable of targeting a protein to the plastid of a cell of the genus *Prototheca*, including those from microalgae and those wherein the plastid targeting sequence has at least 20, 25, 35, 45, or 55% amino acid sequence identity to one or more of SEQ ID NOs. 127-133 and is capable of targeting a protein to the plastid of a cell of the genus *Prototheca*. Additional vectors described herein comprise nucleic acid sequences endogenous to the nuclear genome of a cell of the genus *Prototheca*, wherein the sequence is at least 200 nucleotides long, and some vectors comprise first and second nucleic acid sequences endogenous to the nuclear genome of a cell of the genus *Prototheca*, wherein the first and second sequences (a) are each at least 200 nucleotides long; (b) flank the expression cassette; and (c) are located on the same *Prototheca* chromosome no more than 5, 10, 15, 20, and 50kB apart.

[0013] The disclosure also describes a recombinant nucleic acid with at least 80, 90, 95 or 98% nucleotide identity to one or both of SEQ ID NOs: 134-135 and a recombinant nucleic acid encoding a protein with at least 80, 90, 95 or 98% amino acid identity to one or both of SEQ ID NOs: 136-137.

[0014] The disclosure also describes methods of producing triglyceride compositions, comprising (a) culturing a population of cells of the genus *Prototheca* in the presence of a fixed carbon source, wherein: (i) the cells contain an exogenous gene; (ii) the cells accumulate at least 10, 20, 30, 40, 60, or 70% of their dry cell weight as lipid; and (iii) the fixed carbon source is selected from the group consisting of sorghum and depolymerized cellulosic material; and (b) isolating lipid components from the cultured microorganisms. In some examples the the fixed carbon source is depolymerized cellulosic material selected from the group consisting of corn stover, *Miscanthus*, forage sorghum, sugar beet pulp and sugar cane bagasse, optionally that has been subjected to washing with water prior to the culturing step. In some methods the fixed carbon source is depolymerized cellulosic material and the glucose level of the depolymerized cellulosic material is concentrated to a level of at least 300 g/liter, at least 400 g/liter, at least 500 g/liter, or at least 600 g/liter of prior to the

culturing step and is fed to the culture over time as the cells grow and accumulate lipid. In some methods the exogenous gene encodes a fatty acyl-ACP thioesterase that has hydrolysis activity towards one or more fatty acyl-ACP substrates of chain length C8, C10, C12 or C14, and in some methods the triglyceride has a lipid profile of at least 4% C8-C14 and one or more of the following attributes: 0.1-0.4 micrograms/ml total carotenoids; less than 0.02 milligrams of chlorophyll per kilogram of oil; 0.40-0.60 milligrams of gamma tocopherol per 100 grams of oil; 0.2-0.5 milligrams of total tocotrienols per gram of oil, 4-8 mg per 100 grams of oil of campesterol, and 40-60 mg per 100 grams of oil of stigmasterol.

[0015] Further methods described herein include producing a triglyceride composition, comprising: (a) culturing a population of microorganisms in the presence of depolymerized cellulosic material, wherein: (i) the depolymerized cellulosic material is subjected to washing with water prior to the culturing step; (ii) the cells accumulate at least 10, 20, 30, 40, 60, or 70% of their dry cell weight as lipid; and (iii) the depolymerized cellulosic material is concentrated to at least 300, 400, 500, or 600 g/liter of glucose prior to the cultivation step; (iv) the microorganisms are cultured in a fed-batch reaction in which depolymerized cellulosic material of at least 300, 400, 500, or 600 g/liter of glucose is fed to the microorganisms; and (b) isolating lipid components from the cultured microorganisms. In some examples the the fixed carbon source is depolymerized cellulosic material selected from the group consisting of corn stover, *Miscanthus*, forage sorghum, sugar beet pulp and sugar cane bagasse. In further examples the microorganisms are a species of the genus *Prototheca* and contain an exogenous gene, including a fatty acyl-ACP thioesterase that has hydrolysis activity towards one or more fatty acyl-ACP substrates of chain length C8, C10, C12 or C14. A further method described herein comprises manufacturing triglyceride oil comprising cultivating a cell that has a 23S rRNA sequence with at least 90 or 96% nucleotide identity to SEQ ID NO: 30 in the presence of sucrose as a carbon source.

[0016] The disclosure also describes methods of manufacturing a chemical comprising performing one or more chemical reactions from the list consisting of transesterification, hydrogenation, hydrocracking, deoxygenation, isomerization, interesterification, hydroxylation, hydrolysis, and saponification on a triglyceride oil, wherein the oil has a lipid profile of at least 4% C8-C14 and one or more of the following attributes: 0.1-0.4 micrograms/ml total carotenoids; less than 0.02 milligrams of chlorophyll per kilogram of oil; 0.10-0.60 milligrams of gamma tocopherol per 100 grams of oil; 0.1-0.5 milligrams of total tocotrienols per gram of oil, 1-8 mg per 100 grams of oil of campesterol, and 10-60 mg per 100 grams of oil of stigmasterol. Some methods are performed by manufacturing the oil by

cultivating a cell of the genus *Prototheca* that comprises an exogenous fatty acyl-ACP thioesterase gene that encodes a fatty acyl-ACP thioesterase having hydrolysis activity towards one or more fatty acyl-ACP substrates of chain length C8, C10, C12 or C14. In some methods the hydrolysis reaction is selected from the group consisting of saponification, acid hydrolysis, alkaline hydrolysis, enzymatic hydrolysis, catalytic hydrolysis, and hot-compressed water hydrolysis, including a catalytic hydrolysis reaction wherein the oil is split into glycerol and fatty acids. In further methods the fatty acids undergo an amination reaction to produce fatty nitrogen compounds or an ozonolysis reaction to produce mono- and dibasic-acids. In some examples the oil undergoes a triglyceride splitting method selected from the group consisting of enzymatic splitting and pressure splitting. In some methods a condensation reaction follows the hydrolysis reaction. Other methods include performing a hydroprocessing reaction on the oil, optionally wherein the product of the hydroprocessing reaction undergoes a deoxygenation reaction or a condensation reaction prior to or simultaneous with the hydroprocessing reaction. Some methods additionally include a gas removal reaction. Additioanl methods include processing the aforementioned oils by performing a deoxygenation reaction selected from the group consisting of: a hydrogenolysis reaction, hydrogenation, a consecutive hydrogenation-hydrogenolysis reaction, a consecutive hydrogenolysis-hydrogenation reaction, and a combined hydrogenation-hydrogenolysis reaction. In some methods a condensation reaction follows the deoxygenation reaction. Other methods include performing an esterification reaction on the aforementioned oils, optionally an interestification reaction or a transesterification reaction. Other methods include performing a hydroxylation reaction on the aforementioned oils, optionally wherein a condensation reaction follows the hydroxylation reaction.

BRIEF DESCRIPTION OF THE DRAWINGS

[0017] Figures 1 and 2 illustrate the growth curves of *Prototheca* species and *Chlorella luteoviridis* strain SAG 2214 grown on sorghum as the carbon source.

[0018] Figure 3 shows time course growth of SAG 2214 on glucose and sucrose.

[0019] Figure 4 shows maps of the cassettes using in *Prototheca* transformations, as described in Example 3.

[0020] Figure 5 shows the results of Southern blot analysis on three transformants of UTEX strain 1435, as described in Example 3.

[0021] Figure 6 shows a schematic of the codon optimized and non-codon optimized *suc2* (yeast sucrose invertase (yInv)) transgene construct. The relevant restriction cloning sites are indicated and arrows indicate the direction of transcription.

WHAT IS CLAIMED IS:

1. A codon-optimized recombinant nucleic acid comprising a fatty acyl-ACP thioesterase coding sequence codon-optimized for expression in *Prototheca*, wherein the coding sequence contains the most or second most preferred codon of Table 1 for at least 60% of the codons of the coding sequence, such that the codon-optimized sequence is more efficiently translated in *Prototheca* than a native sequence encoding the fatty acyl-ACP thioesterase, the fatty acyl-ACP thioesterase having hydrolysis activity towards one or more fatty acyl-ACP substrates, the recombinant nucleic acid further comprising a plastid targeting sequence from microalgae that is in-frame with the codon-optimized fatty acyl-ACP thioesterase coding sequence.

2. The codon-optimized recombinant nucleic acid of claim 1, wherein the plastid targeting sequence is a sequence selected from the group consisting of SEQ ID NOs: 127-130.

3. A codon-optimized recombinant nucleic acid comprising a fatty acyl-ACP thioesterase coding sequence codon-optimized for expression in *Prototheca*, wherein the coding sequence contains the most or second most preferred codon of Table 1 for at least 60% of the codons of the coding sequence, such that the codon-optimized sequence is more efficiently translated in *Prototheca* than a native sequence encoding the fatty acyl-ACP thioesterase, the fatty acyl-ACP thioesterase having hydrolysis activity towards one or more fatty acyl-ACP substrates, the recombinant nucleic acid further comprising a plastid targeting sequence in-frame with the codon-optimized fatty acyl-ACP thioesterase coding sequence, the recombinant nucleic acid further comprising first and second flanking nucleic acid sequences endogenous to the nuclear genome of a cell of the genus *Prototheca*, wherein the first and second flanking sequences are each at least 200 nucleotides long and the first and second flanking sequences flank the codon-optimized recombinant nucleic acid for integration of the recombinant nucleic acid into the nuclear genome by homologous recombination with a corresponding genomic sequence.

4. The codon-optimized recombinant nucleic acid of claim 3, wherein the first and second flanking sequences are found on the same *Prototheca* chromosome and are located no more than 20 kB apart from each other.

5. A codon-optimized recombinant nucleic acid comprising a fatty acyl-ACP thioesterase coding sequence codon-optimized for expression in *Prototheca*, wherein the coding sequence

contains the most or second most preferred codon of Table 1 for at least 60% of the codons of the coding sequence, such that the codon-optimized sequence is more efficiently translated in *Prototheca* than a native sequence encoding the fatty acyl-ACP thioesterase, the fatty acyl-ACP thioesterase having hydrolysis activity towards one or more fatty acyl-ACP substrates, the recombinant nucleic acid further comprising a plastid targeting sequence from microalgae that is in-frame with the codon-optimized fatty acyl-ACP thioesterase coding sequence, and wherein the amino acid sequence of the fatty acyl-ACP thioesterase comprises an amino acid sequence of a fatty acyl-ACP thioesterase from an organism selected from the group consisting of *Cuphea hookeriana, Umbellularia californica, Cinnamomum camphora, Cuphea palustris* and *Ulmus americana*.

6. A codon-optimized recombinant nucleic acid comprising a fatty acyl-ACP thioesterase coding sequence and a sucrose invertase coding sequence, each coding sequence codon-optimized for expression in *Prototheca*, wherein each coding sequence contains the most or second most preferred codon of Table 1 for at least 60% of the codons of the coding sequence, such that each codon-optimized sequence is more efficiently translated in *Prototheca* than a native sequence encoding the fatty acyl-ACP thioesterase or the native sequence encoding the sucrose invertase, the fatty acyl-ACP thioesterase having hydrolysis activity towards one or more fatty acyl-ACP substrates, the recombinant nucleic acid further comprising a plastid targeting sequence in-frame with the codon-optimized fatty acyl-ACP thioesterase coding sequence.

7. A host cell of the genus *Prototheca* comprising a recombinant nucleic acid comprising a fatty acyl-ACP thioesterase coding sequence codon-optimized for expression in *Prototheca*, wherein the coding sequence contains the most or second most preferred codon of Table 1 for at least 60% of the codons of the coding sequence, such that the codon-optimized sequence is more efficiently translated in *Prototheca* than a native sequence encoding the fatty acyl-ACP thioesterase, and wherein a plastid targeting sequence is in-frame with the codon-optimized fatty acyl-ACP thioesterase coding sequence and the fatty acyl-ACP thioesterase has hydrolysis activity towards one or more fatty acyl-ACP substrates.

8. A host cell of the genus *Prototheca* comprising a codon optimized recombinant nucleic acid comprising a fatty acyl-ACP thioesterase coding sequence codon-optimized for expression in *Prototheca*, wherein the coding sequence contains the most or second most preferred codon of Table 1 for at least 60% of the codons of the coding sequence, such that

the codon-optimized sequence is more efficiently translated in *Prototheca* than a native sequence encoding the fatty acyl-ACP thioesterase, the fatty acyl-ACP thioesterase having hydrolysis activity towards a fatty acyl-ACP substrate of chain length C8, C10, C12 or C14, the recombinant nucleic acid further comprising a plastid targeting sequence that is in-frame with the codon-optimized fatty acyl-ACP thioesterase coding sequence.

9. A host cell of the genus *Prototheca* comprising the recombinant nucleic acid of claim 1.

10. A host cell of the genus *Prototheca* comprising the recombinant nucleic acid of claim 2.

11. A host cell of the genus *Prototheca* comprising the recombinant nucleic acid of claim 3.

12. A host cell of the genus *Prototheca* comprising the recombinant nucleic acid of claim 4.

13. A host cell of the genus *Prototheca* comprising the recombinant nucleic acid of claim 5.

14. A host cell of the genus *Prototheca* comprising the recombinant nucleic acid of claim 6.

15. A codon-optimized recombinant nucleic acid comprising a fatty acyl-ACP thioesterase coding sequence codon-optimized for expression in *Prototheca*, wherein the coding sequence contains the most or second most preferred codon of Table 1 for at least 60% of the codons of the coding sequence, such that the codon-optimized sequence is more efficiently translated in *Prototheca* than a native sequence encoding the fatty acyl-ACP thioesterase, the fatty acyl-ACP thioesterase having hydrolysis activity towards one or more fatty acyl-ACP substrates, further comprising a promoter comprising the nucleotide sequence of any of SEQ ID NOS: 91-102.

16. The codon-optimized recombinant nucleic acid of claim 6, wherein each coding sequence contains the most preferred codon of Table 1 for at least 60% of the codons of each coding sequence.

17. The codon-optimized recombinant nucleic acid of claim 6, wherein each coding sequence contains the most preferred codon of Table 1 for at least 80% of the codons of each coding sequence.