US 20110162259A1

(19) **United States**

(12) **Patent Application Publication**

GAERTNER

(10) **Pub. No.: US 2011/0162259 A1**

(43) **Pub. Date: Jul. 7, 2011**

(54) **PRODUCTION OF FATTY ACID DERIVATIVES**

(75) Inventor: **Alfred GAERTNER**, South San Francisco, CA (US)

(73) Assignee: **LS9, INC.**, South San Francisco, CA (US)

(21) Appl. No.: **12/889,163**

(22) Filed: **Sep. 23, 2010**

Related U.S. Application Data

(60) Provisional application No. 61/245,943, filed on Sep. 25, 2009.

Publication Classification

(51) **Int. Cl.**

C10L 1/19	(2006.01)
C07C 53/126	(2006.01)
C07C 69/003	(2006.01)
C12P 7/64	(2006.01)
C12N 1/00	(2006.01)
C12N 1/21	(2006.01)

(52) **U.S. Cl.** **44/385**; 554/1; 435/134; 435/243; 435/252.3

(57) **ABSTRACT**

Methods and compositions for producing fatty acid derivatives, for example, fatty esters, and commercial fuel compositions comprising fatty acid derivatives are described.

PRODUCTION OF FATTY ACID DERIVATIVES

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims the benefit of U.S. Provisional Application No. 61/245,943, filed Sep. 25, 2009, the contents of which are hereby incorporated in their entirety herein.

BACKGROUND OF THE INVENTION

[0002] Petroleum is a limited, natural resource found in the Earth in liquid, gaseous, or solid forms. Petroleum is primarily composed of hydrocarbons, which are comprised mainly of carbon and hydrogen. It also contains significant amounts of other elements, such as, nitrogen, oxygen, or sulfur, in different forms.

[0003] Petroleum is a valuable resource, but petroleum products are developed at considerable costs, both financial and environmental. First, sources of petroleum must be discovered. Petroleum exploration is an expensive and risky venture. The cost of exploring deep water wells can exceed $100 million. In addition to the economic cost, petroleum exploration carries a high environmental cost. For example, offshore exploration disturbs the surrounding marine environments.

[0004] After a productive well is discovered, the petroleum must be extracted from the Earth at great expense. Even under the best circumstances, only 50% of the petroleum in a well can be extracted. Petroleum extraction also carries an environmental cost. For example, petroleum extraction can result in large seepages of petroleum rising to the surface. Offshore drilling involves dredging the seabed which disrupts or destroys the surrounding marine environment.

[0005] After extraction, petroleum must be transported over great distances from petroleum producing regions to petroleum consuming regions. In addition to the shipping costs, there is also the environmental risk of devastating oil spills.

[0006] In its natural form, crude petroleum extracted from the Earth has few commercial uses. It is a mixture of hydrocarbons (e.g., paraffins (or alkanes), olefins (or alkenes), alkynes, napthenes (or cycloalkanes), aliphatic compounds, aromatic compounds, etc.) of varying length and complexity. In addition, crude petroleum contains other organic compounds (e.g., organic compounds containing nitrogen, oxygen, sulfur, etc.) and impurities (e.g., sulfur, salt, acid, metals, etc.).

[0007] Hence, crude petroleum must be refined and purified before it can be used commercially. Due to its high energy density and its easy transportability, most petroleum is refined into fuels, such as transportation fuels (e.g., gasoline, diesel, aviation fuel, etc.), heating oil, liquefied petroleum gas, etc.

[0008] Crude petroleum is also a primary source of raw materials for producing petrochemicals. The two main classes of raw materials derived from petroleum are short chain olefins (e.g., ethylene and propylene) and aromatics (e.g., benzene and xylene isomers). These raw materials are derived from the longer chain hydrocarbons in crude petroleum by cracking the long chain hydrocarbons at considerable expense using a variety of methods, such as catalytic cracking, steam cracking, or catalytic reforming. These raw materials are used to make petrochemicals, which cannot be directly refined from crude petroleum, such as monomers, solvents, detergents, or adhesives.

[0009] One example of a raw material derived from crude petroleum is ethylene. Ethylene is used to produce petrochemicals such as, polyethylene, ethanol, ethylene oxide, ethylene glycol, polyester, glycol ether, ethoxylate, vinyl acetate, 1,2-dichloroethane, trichloroethylene, tetrachloroethylene, vinyl chloride, and polyvinyl chloride. Another example of a raw material derived from crude petroleum is propylene. Propylene is used to produce isopropyl alcohol, acrylonitrile, polypropylene, propylene oxide, propylene glycol, glycol ethers, butylene, isobutylene, 1,3-butadiene, synthetic elastomers, polyolefins, alpha-olefins, fatty alcohols, acrylic acid, acrylic polymers, allyl chloride, epichlorohydrin, and epoxy resins.

[0010] Petrochemicals can be used to make specialty chemicals, such as plastics, resins, fibers, elastomers, pharmaceuticals, lubricants, or gels. Examples of specialty chemicals which can be produced from petrochemical raw materials are: fatty acids, hydrocarbons (e.g., long chain hydrocarbons, branched chain hydrocarbons, saturated hydrocarbons, unsaturated hydrocarbons, etc.), fatty alcohols, esters, fatty aldehydes, ketones, lubricants, etc.

[0011] Specialty chemicals have many commercial uses. Fatty acids are used commercially as surfactants. Surfactants can be found in detergents and soaps. Fatty acids can also be used as additives in fuels, lubricating oils, paints, lacquers, candles, salad oils, shortenings, cosmetics, and emulsifiers. In addition, fatty acids are used as accelerator activators in rubber products. Fatty acids can also be used as a feedstock to produce methyl esters, amides, amines, acid chlorides, anhydrides, ketene dimers, and peroxy acids and esters.

[0012] Esters have many commercial uses. For example, biodiesel, an alternative fuel, is comprised of esters (e.g., fatty acid methyl ester, fatty acid ethyl esters, etc.). Some low molecular weight esters are volatile with a pleasant odor which makes them useful as fragrances or flavoring agents. In addition, esters are used as solvents for lacquers, paints, and varnishes. Furthermore, some naturally occurring substances, such as waxes, fats, and oils are comprised of esters. Esters are also used as softening agents in resins and plastics, plasticizers, flame retardants, and additives in gasoline and oil. In addition, esters can be used in the manufacture of polymers, films, textiles, dyes, and pharmaceuticals.

[0013] In addition, crude petroleum is a source of lubricants. Lubricants derived petroleum are typically composed of olefins, particularly polyolefins and alpha-olefins. Lubricants can either be refined from crude petroleum or manufactured using the raw materials refined from crude petroleum.

[0014] Obtaining these specialty chemicals from crude petroleum requires a significant financial investment as well as a great deal of energy. It is also an inefficient process because frequently the long chain hydrocarbons in crude petroleum are cracked to produce smaller monomers. These monomers are then used as the raw material to manufacture the more complex specialty chemicals.

[0015] In addition to the problems with exploring, extracting, transporting, and refining petroleum, petroleum is a limited and dwindling resource. One estimate of current world petroleum consumption is 30 billion barrels per year. By some estimates, it is predicted that at current production levels, the world's petroleum reserves could be depleted before the year 2050.

[0016] Finally, the burning of petroleum based fuels releases greenhouse gases (e.g., carbon dioxide) and other forms of air pollution (e.g., carbon monoxide, sulfur dioxide, etc.). As the world's demand for fuel increases, the emission of greenhouse gases and other forms of air pollution also increases. The accumulation of greenhouse gases in the atmosphere leads to an increase in global warming. Hence, in addition to damaging the environment locally (e.g., oil spills, dredging of marine environments, etc.), burning petroleum also damages the environment globally.

[0017] Due to the inherent challenges posed by petroleum, there is a need for a renewable petroleum source which does not need to be explored, extracted, transported over long distances, or substantially refined like petroleum. There is also a need for a renewable petroleum source that can be produced economically. In addition, there is a need for a renewable petroleum source that does not create the type of environmental damage produced by the petroleum industry and the burning of petroleum based fuels. For similar reasons, there is also a need for a renewable source of chemicals that are typically derived from petroleum.

[0018] Renewable energy sources, such as sunlight, water, wind, and biomass, are a potential alternative to petroleum fuels. Biofuel is a biodegradable, clean-burning combustible fuel produced from biomass, and can be made of alkanes and esters. An exemplary biofuel is biodiesel. Biodiesel can be used in most internal combustion diesel engines in either a pure form, which is referred to as "neat" biodiesel, or as a mixture in any concentration with regular petroleum diesel.

[0019] Biodiesel offers advantages compared to petroleum-based diesel, including reduced emissions (e.g., carbon monoxide, sulphur, aromatic hydrocarbons, soot particles) during combustion. Biodiesel also maintains a balanced carbon dioxide cycle because it is based on renewable biological materials. Biodiesel is typically biodegradable, and imparts enhanced safety due to its high flash point and low flammability. Furthermore, biodiesel provides good lubrication properties, thereby reducing wear and tear on engines.

[0020] Current methods of making biodiesel involve transesterification of triacylglycerides from vegetable oil feedstocks, such as rapeseed in Europe, soybean in North America, and palm oil in South East Asia. Industrial-scale biodiesel production is thus geographically and seasonally restricted to areas where vegetable oil feedstocks are produced. The transesterification process leads to a mixture of fatty esters which can be used as biodiesel. However, glycerin is an undesirable byproduct of the transesterification process. To be usable as biodiesel, the fatty esters must be further purified from the heterogeneous product. This increases costs and the amount of energy required for fatty ester production and, ultimately, biodiesel production as well. Furthermore, vegetable oil feedstocks are inefficient sources of energy because they require extensive acreage for cultivation. For example, the yield of biodiesel from rapeseed is only 1300 L/hectare because only the seed oil is used for biodiesel production, while the rest of the rapeseed biomass is discarded. Additionally, cultivating some vegetable oil feedstocks, such as rapeseed and soybean, requires frequent crop rotation to prevent nutrient depletion of the land.

[0021] Thus there is a need for an economically- and energy-efficient biofuel, and methods of making biofuels from renewable energy sources such as biomass.

SUMMARY OF THE INVENTION

[0022] The invention is based, at least in part, on the production of fatty esters, such as fatty esters, including, for example fatty acid methyl esters ("FAME") and fatty acid ethyl esters ("FAEE"), from genetically engineered microorganisms. Accordingly, in one aspect, the invention features a method of producing a fatty ester. The method comprises culturing a host cell in the presence of a carbon source, wherein the host cell is genetically engineered to express or overexpress a gene encoding a thioesterase, a gene encoding an acyl-CoA synthase, and a gene encoding an ester synthase. In some embodiments, the method further comprises isolating the fatty ester.

[0023] In some embodiments, the fatty ester is present in the extracellular environment. In some embodiments, the fatty ester is isolated from the extracellular environment of the host cell. In some embodiments, the fatty ester is spontaneously secreted, partially or completely, from the host cell. In alternative embodiments, the fatty ester is transported into the extracellular environment, optionally with the aid of one or more suitable transport proteins. In other embodiments, the fatty ester is passively transported into the extracellular environment.

[0024] In some embodiments, the method further comprises culturing the host cell in the presence of an alcohol. In certain embodiments, the alcohol is methanol, ethanol, propanol, or butanol. In some embodiments, the alcohol is present at a concentration of about 1 mL/L to about 100 mL/L. For example, the alcohol is present at a concentration of about 1 mL/L or more (e.g., about 1 mL/L or more, about 5 mL/L or more, about 10 mL/L or more). In alternative embodiments, the alcohol is present at a concentration of about 100 mL/L or less (e.g., about 100 mL/L or less, about 90 mL/L or less, about 80 mL/L or less).

[0025] In certain embodiments, the gene encoding a thioesterase is tesA, 'tesA, tesB, fatB, fatB2, fatB3, fatA1, or fatA. In some embodiments, the gene encoding an acyl-CoA synthase is fadD, fadK, BH3103, pfl-4354, EAV15023, fadD1, fadD2, RPC_4074, fadDD35, fadDD22, faa39, the gene encoding the protein of GenBank Accession No. ZP_01644857, or yhfL. In yet other embodiments, the gene encoding an ester synthase is one encoding an enzyme of enzyme classification EC 2.3.1.75 or EC 2.3.1.20, one encoding wax/dgat, a bifunctional ester synthase/acyl-CoA:diacylglycerl acyltransferase from *Simmondsia chinensis, Acinetobacter* sp. ADP1, *Alcanivorax borkumensis, Pseudomonas aeruginosa, Fundibacter jadensis, Arabidopsis thaliana*, or *Alkaligenes eutrophus*, or one encoding AtfA1, AtfA2, ES9, or ES8, or a variant thereof.

[0026] In other embodiments, the host cell is genetically engineered to express, relative to a wild type host cell, a decreased level of at least one of a gene encoding an acyl-CoA dehydrogenase, a gene encoding an outer membrane protein receptor, and a gene encoding a transcriptional regulator of fatty acid biosynthesis. In some embodiments, one or more of a gene encoding an acyl-CoA dehydrogenase, a gene encoding an outer membrane protein receptor, and a gene encoding a transcriptional regulator of fatty acid biosynthesis are functionally deleted.

[0027] In some embodiments, the gene encoding an acyl-CoA dehydrogenase is fadE. In some embodiments, the gene encoding an outer membrane protein receptor encodes an outer membrane ferrichrome transporter, for example, fhuA. Yet in other embodiments, the gene encoding a transcriptional regulator of fatty acid biosynthesis encodes a DNA transcription repressor, for example, fabR.

[0028] In some embodiments, the host cell is genetically engineered to express, relative to a wild type host cell, an attenuated level of at least one of a gene encoding a pyruvate formate lyase, a gene encoding a lactate dehydrogenase, or both. In certain embodiments, one or more of a gene encoding a pyruvate formate lyase, a gene encoding a lactate dehydrogenase, or both, are functionally deleted. In some embodiments, the gene encoding a pyruvate formate lyase is pflB. In certain embodiments, the gene encoding a lactate dehydrogenase is ldhA.

[0029] In some embodiments, the host cell is genetically engineered to express, relative to a wild type host cell, attenuated levels of one or more or all of a gene encoding an acyl-CoA dehydrogenase, a gene encoding a ferrichrome transporter, a gene encoding a pyruvate formate lyase, and a gene encoding a lactate dehydrogenase. In other embodiments, the host cell is engineered such that one or more or all of a gene encoding an endogenous acyl-CoA dehydrogenase, a gene encoding an endogenous ferrichrome transporter, a gene encoding an endogenous pyruvate formate lyase, and a gene encoding an endogenous lactate dehydrogenase are functionally deleted from the host cell.

[0030] In some embodiments, the host cell is cultured in a culture medium comprising an initial concentration of the carbon source of about 2 g/L to about 100 g/L. In other embodiments, the culture medium comprises an initial concentration of about 2 g/L to about 10 g/L of a carbon source, of about 10 g/L to about 20 g/L of a carbon source, of about 20 g/L to about 30 g/L of a carbon source, of about 30 g/L to about 40 g/L of a carbon source, or of about 40 g/L to about 50 g/L of a carbon source.

[0031] In some embodiments, the method further includes the step of monitoring the level of the carbon source in the culture medium. In some embodiments, the method further includes adding a supplemental carbon source to the culture medium when the level of the carbon source in the medium is less than about 0.5 g/L. In some embodiments, supplemental carbon source is added to the culture medium when the level of the carbon source in the medium is less than about 0.4 g/L, less than about 0.3 g/L, less than about 0.2 g/L, or less than about 0.1 g/L.

[0032] In some embodiments, the supplemental carbon source is added to maintain a carbon source level of about 1 g/L to about 25 g/L. In some embodiments, the supplemental carbon source is added to maintain a carbon source level of about 2 g/L or more (e.g., about 2 g/L or more, about 3 g/L or more, about 4 g/L or more). In certain embodiments, the supplemental carbon source is added to maintain a carbon source level of about 5 g/L or less (e.g., about 5 g/L or less, about 4 g/L or less, about 3 g/L or less). In some embodiments, the supplemental carbon source is added to maintain a carbon source level of about 2 g/L to about 5 g/L, of about 5 g/L to about 10 g/L, or of about 10 g/L to about 25 g/L. In some embodiments, the carbon source is glucose.

[0033] In some embodiments, the fatty acid methyl ester is produced at a concentration of about 1 g/L to about 200 g/L. In some embodiments, the fatty acid methyl ester is produced at a concentration of about 1 g/L or more (e.g., about 1 g/L or more, about 10 g/L or more, about 20 g/L or more, about 50 g/L or more, about 100 g/L or more). In some embodiments, the fatty acid methyl ester is produced at a concentration of about 1 g/L to about 170 g/L, of about 1 g/L to about 10 g/L, of about 40 g/L to about 170 g/L, of about 100 g/L to about 170 g/L, of about 10 g/L to about 100 g/L, of about 1 g/L to about 40 g/L, of about 40 g/L to about 100 g/L, or of about 1 g/L to about 100 g/L.

[0034] In some embodiments, the host cell is selected from the group consisting of a mammalian cell, plant cell, insect cell, yeast cell, fungus cell, filamentous fungi cell, and bacterial cell.

[0035] In particular embodiments, the host cell is selected from the genus *Escherichia, Bacillus, Lactobacillus, Rhodococcus, Synechococcus, Synechoystis, Pseudomonas, Aspergillus, Trichoderma, Neurospora, Fusarium, Humicola, Rhizomucor, Kluyveromyces, Pichia, Mucor, Myceliophtora, Penicillium, Phanerochaete, Pleurotus, Trametes, Chrysosporium, Saccharomyces, Stenotrophamonas, Schizosaccharomyces, Yarrowia*, or *Streptomyces*.

[0036] In other embodiments, the host cell is a *Bacillus lentus* cell, a *Bacillus brevis* cell, a *Bacillus stearothermophilus* cell, a *Bacillus licheniformis* cell, a *Bacillus alkalophilus* cell, a *Bacillus coagulans* cell, a *Bacillus circulans* cell, a *Bacillus pumilis* cell, a *Bacillus thuringiensis* cell, a *Bacillus clausii* cell, a *Bacillus megaterium* cell, a *Bacillus subtilis* cell, or a *Bacillus amyloliquefaciens* cell.

[0037] In certain embodiments, the host cell is a *Synechococcus* sp. PCC7002, *Synechococcus elongatus* PCC 7942, *Synechoystis* sp. PCC 6803, *Synechococcus elongatus* PCC6301, *Prochlorococcus marinus* CCMP1986 (MED4), *Anabaena variabilis* ATCC29413, *Nostoc punctiforme* ATCC29133 (PCC73102), *Gloeobacter violaceus* ATCC29082 (PCC7421), *Nostoc* sp. ATCC27893 (PCC7120), *Cyanothece* sp. PCC7425 (29141), *Cyanothece* sp. ATCC51442, or *Synechococcus* sp. ATCC27264 (PCC7002).

[0038] In other embodiments, the host cell is a *Trichoderma koningii* cell, a *Trichoderma viride* cell, a *Trichoderma reesei* cell, a *Trichoderma longibrachiatum* cell, an *Aspergillus awamori* cell, an *Aspergillus fumigates* cell, an *Aspergillus foetidus* cell, an *Aspergillus nidulans* cell, an *Aspergillus niger* cell, an *Aspergillus oryzae* cell, a *Humicola insolens* cell, a *Humicola lanuginose* cell, a *Rhodococcus opacus* cell, a *Rhizomucor miehei* cell, or a *Mucor michei* cell.

[0039] In other embodiments, the host cell is an Actinomycetes cell. In yet other embodiments, the host cell is a *Streptomyces lividans* cell or a *Streptomyces murinus* cell. In other embodiments, the host cell is a *Saccharomyces cerevisiae* cell.

[0040] In yet other embodiments, the host cell is a cell from an eukaryotic plant, algae, *cyanobacterium*, green-sulfur bacterium, green non-sulfur bacterium, purple sulfur bacterium, purple non-sulfur bacterium, extremophile, yeast, fungus, engineered organisms thereof, or a synthetic organism. In some embodiments, the host cell is light dependent or fixes carbon. In some embodiments, the host cell has autotrophic activity. In some embodiments, the host cell has photoautotrophic activity, such as in the presence of light. In certain embodiments, the host cell is a cell from *Arabidopsis thaliana, Panicum virgatums, Miscanthus giganteus, Zea mays, botryococcuse braunii, Chlamydomonas reinhardtii, Dunaliela salina, Thermosynechococcus elongatus, Chlorobium tepidum, Chloroflexus auranticus, Chromatiumm vinosum, Rhodospirillum rubrum, Rhodobacter capsulatus, Rhodopseudomonas palusris, Clostridium ljungdahlii, Clostridiuthermocellum*, or *Pencillium chrysogenum*.

[0041] In certain other embodiments, the host cell is from *Pichia pastories, Saccharomyces cerevisiae, Yarrowia lipoly-*

tica, Schizosaccharomyces pombe, Pseudomonas fluorescens, or *Zymomonas mobilis*. In yet further embodiments, the host cell is a cell from *Synechococcus* sp. PCC 7002, *Synechococcus* sp. PCC 7942, or *Synechocystis* sp. PCC6803.

[0042] In some embodiments, the host cell is a CHO cell, a COS cell, a VERO cell, a BHK cell, a HeLa cell, a Cv1 cell, an MDCK cell, a 293 cell, a 3T3 cell, or a PC12 cell.

[0043] In particular embodiments, the host cell is an *E. coli* cell. In some embodiments, the *E. coli* cell is a strain B, a strain C, a strain K, or a strain W *E. coli* cell.

[0044] In another aspect, the invention features a genetically engineered microorganism capable of producing fatty esters under conditions that allow product production. In some embodiments, the genetically engineered microorganism comprises at least one or more of a gene encoding a thioesterase, a gene encoding an acyl-CoA synthase, and a gene encoding an ester synthase. In certain embodiments, the genetically engineered microorganism comprises a gene encoding a thioesterase, a gene encoding an acyl-CoA synthase, and a gene encoding an ester synthase. In certain embodiments, two or more of the genes encoding a thioesterase, a gene encoding an acyl-CoA synthase, and a gene encoding an ester synthase are linked in a single operon. In certain other embodiments, all three of the genes encoding a thioesterase, a gene encoding an acyl-CoA synthase, and a gene encoding an ester synthase are linked into a single operon.

[0045] In certain embodiment, the genetically engineered microorganism comprises an exogenous control sequence stably incorporated into the genomic DNA of the microorganism upstream of one or more of at least one of a gene encoding a thioesterase, a gene encoding an acyl-CoA synthase, and a gene encoding an ester synthase. In an exemplary embodiment, the microorganism is engineered such that it comprises an exogenous control sequence stably incorporated into the genomic DNA of the microorganism up stream of a gene encoding a thioesterase, an acyl-CoA synthase and a gene encoding an ester synthase. In certain embodiments, the microorganism engineered as such produces an increased level of a fatty ester relative to a wild-type microorganism. In certain embodiments, the exogenous control sequence is, for example, a promoter. Exemplary promoters include, without limitation, a developmentally-regulated, organelle-specific, tissue-specific, inducible, constitutive, or cell-specific promoter.

[0046] In further embodiments, the microorganism is genetically engineered to express, relative to a wild type microorganism, a decreased level of at least one of a gene encoding an acyl-CoA dehydrogenase, a gene encoding an outer membrane protein receptor, and a gene encoding a transcriptional regulator of fatty acid biosynthesis. In certain embodiments, the microorganism is genetically engineered such that at one or more of a gene encoding an acyl-CoA dehydrogenase, a gene encoding an outer membrane protein receptor, and a gene encoding a transcriptional regulator of fatty acid biosynthesis are functionally deleted. In certain embodiments, the gene encoding the acyl-CoA dehydrogenase is fadE. In other embodiments, the gene encoding an outer membrane protein receptor encodes an outer membrane ferrichrome transporter, for example, fhuA. In further embodiments, the gene encoding a transcriptional regulator of fatty acid biosynthesis encodes a DNA transcription repressor, for example, fabR.

[0047] In some embodiments, the microorganism is genetically engineered to express, relative to a wild type microorganism, an attenuated level of a gene encoding a pyruvate formate lyase, a lactate dehydrogenase, or both. In some embodiments, at least one of a gene encoding a pyruvate formate lyase and a gene encoding a lactate dehydrogenase are functionally deleted. In some embodiments, the gene encoding the pyruvate formate lyase is pflB. In other embodiments, the gene encoding the lactate dehydrogenase is ldhA.

[0048] In certain embodiments, the microorganism is genetically engineered to express attenuated levels of one or more or all of an acyl-CoA dehydrogenase gene, an outer membrane ferrichrome transporter gene, a pyruvate formate lyase gene, and a lactate dehydrogenase gene. In other embodiments, one or more or all of an endogenous acyl-CoA dehydrogenase gene, an endogenous outer membrane ferrichrome transporter gene, an endogenous pyruvate formate lyase gene, and an endogenous lactate dehydrogenase gene are functionally deleted from the microorganism.

[0049] In certain embodiments, the genetically engineered microorganism can suitably be selected from a Gram-negative or a Gram-positive bacterium. In some embodiments, the genetically engineered microorganism is selected from an *E. coli, mycobacterium, Nocardia* sp., *Nocardia farcinica, Streptomyces griseus, Salinispora arenicola, Clavibacter michiganenesis, Acinetobacter, Alcanivorax, Alcaligenes, Arabidopsis, Fundibacter, Marinobacter, Mus musculus, Pseudomonas*, or *Simmodsia, Yarrowia, Candida, Rhodotorula, Rhodosporidium, Cryptococcus, Trichosporon*, or *Lipomyces*. In certain embodiments, the genetically engineered microorganism is selected from an *E. coli* strain B, strain C, strain K or strain W. In further embodiments, the genetically engineered microorganism is selected from *Synechococcus* sp. PCC7002, *Synechococcus elongatus* PCC 7942, *Synechoystis* sp. PCC 6803, *Synechococcus elongatus* PCC6301, *Prochlorococcus marinus* CCMP1986 (MED4), *Anabaena variabilis* ATCC29413, *Nostoc punctiforme* ATCC29133 (PCC73102), *Gloeobacter violaceus* ATCC29082 (PCC7421), *Nostoc* sp. ATCC27893 (PCC7120), *Cyanothece* sp. PCC7425 (29141), *Cyanothece* sp. ATCC51442, or *Synechococcus* sp. ATCC27264 (PCC7002).

[0050] In another aspect, the invention features a method of producing a fatty ester comprising culturing a genetically engineered microorganism described herein in the presence of a suitable alcohol substrate. In certain embodiments, the alcohol substrate is ethanol, methanol, propanol, or butanol. In a particular embodiment, the alcohol substrate is a methanol. In further embodiments, the method further comprises culturing the genetically engineered microorganism under conditions that allows it to produce the fatty ester product.

[0051] In any of the aspects described herein, the fatty ester is produced at a yield of about 0.5 g to about 50 g of fatty ester per 100 g of glucose in the culture medium. For example, the fatty ester is produced at a yield of about 0.5 g of fatty ester per 100 g of glucose or more (e.g., about 0.5 g of fatty ester per 100 g of glucose or more, about 2 g of fatty ester per 100 g of glucose or more, about 5 g of fatty ester per 100 g of glucose or more, about 10 g of fatty ester per 100 g of glucose or more, or about 15 g of fatty ester per 100 g of glucose or more). In particular embodiments, the fatty ester is produced at a yield of about 0.5 g to about 40 g of fatty ester per 100 g of glucose, about 0.5 g to about 30 g of fatty ester per 100 g of glucose, about 0.5 g to about 20 g of fatty ester per 100 g of glucose,

about 0.5 g to about 18 g of fatty ester per 100 g of glucose, about 2 g to about 16 g of fatty ester per 100 g of glucose, or about 5 g to about 15 g of fatty ester per 100 g of glucose in the culture medium. In particular embodiments, the fatty ester is produced at a yield of at least 0.5 g of fatty ester, at least 4 g of fatty ester, at least 5 g of fatty ester, at least 10 g of fatty ester, at least 20 g of fatty ester, at least 30 g of fatty ester, at least 40 g of fatty ester, or at least 50 g of fatty ester per 100 g of glucose in the culture medium. In particular embodiments, the fatty ester is produced at a yield of no more than 50 g of fatty ester per 100 g of glucose in the culture medium.

[0052] In some embodiments, the fatty ester is produced at a yield of about 0.5% to about 50% by mass of the glucose in the culture medium. For example, the fatty ester is produced at a yield of about 0.5% or more (e.g., about 0.5% or more, about 2% or more, about 5% or more, or about 10% or more) by mass of the glucose in the culture medium. In particular embodiments, the fatty ester is produced at a yield of about 0.5% to about 40%, about 0.5% to about 30%, about 0.5% to about 20%, about 0.5% to about 10%, about 0.5% to about 5%, or about 0.5% to about 4% by mass of the glucose in the culture medium. In particular embodiments, the fatty ester is produced at a yield of at least about 0.5%, at least about 4%, at least about 5%, at least about 10%, at least about 20%, at least about 30%, at least about 40%, or at least about 50% by mass of glucose in the culture medium. In particular embodiments, the fatty ester is produced at a yield of no more than 50% by mass of glucose in the culture medium.

[0053] In some embodiments, the fatty ester is produced at a yield of about 10% to about 95% by mass of carbon in the carbon source in the culture medium. In particular embodiments, the fatty ester is produced at a yield of about 15% to about 90%, about 20% to about 80%, or about 30% to about 70% by mass of carbon in the carbon source in the culture medium. In particular embodiments, the fatty ester is produced at a yield of at least about 10%, at least about 20%, at least about 30%, at least about 40%, at least about 50%, at least about 60%, at least about 70%, at least about 80%, at least about 90%, or at least about 95% by mass of carbon in the carbon source in the culture medium. In particular embodiments, the fatty ester is produced at a yield of no more than 95% by mass of carbon in the carbon source in the culture medium.

[0054] In some embodiments, the fatty ester is a fatty acid ethyl ester and is produced at a yield of about 0.5 g to about 50 g of fatty acid ethyl ester per 100 g of glucose in the culture medium. For example, the fatty acid ethyl ester and is produced at a yield of about 0.5 g or more (e.g., about 0.5 g or more, about 2 g or more, about 5 g or more, about 10 g or more, about 15 g or more) per 100 g of glucose in the culture medium. In particular embodiments, the fatty acid ethyl ester is produced at a yield of about 0.5 g to about 40 g of fatty acid ethyl ester per 100 g of glucose, about 0.5 g to about 30 g of fatty acid ethyl ester per 100 g of glucose, about 0.5 g to about 20 g of fatty acid ethyl ester per 100 g of glucose, about 0.5 g to about 10 g of fatty acid ethyl ester per 100 g of glucose, about 0.5 g to about 5 g of fatty acid ethyl ester per 100 g of glucose, or about 0.5 g to about 4 g of fatty acid ethyl ester per 100 g of glucose in the culture medium. In particular embodiments, the fatty acid ethyl ester is produced at a yield of at least 0.5 g of fatty acid ethyl ester, at least 4 g of fatty acid ethyl ester, at least 5 g of fatty acid ethyl ester, at least 10 g of fatty acid ethyl ester, at least 20 g of fatty acid ethyl ester, at least 30 g of fatty acid ethyl ester, at least 40 g of fatty acid ethyl ester, or at least 50 g of fatty acid ethyl ester per 100 g of glucose in the culture medium. In particular embodiments, the fatty acid ethyl ester is produced at a yield of no more than 50 g of fatty acid ethyl ester per 100 g of glucose in the culture medium.

[0055] In some embodiments, the fatty acid ethyl ester is produced at a yield of about 0.5% to about 50% by mass of the glucose in the culture medium. For example, the fatty acid ethyl ester is produced at a yield of about 0.5% or more (e.g., of about 0.5% or more, of about 1% or more, of about 2% or more of about 5% or more of about 10% or more) by mass of the glucose in the culture medium. In particular embodiments, the fatty acid ethyl ester is produced at a yield of about 0.5% to about 40%, about 0.5% to about 30%, about 0.5% to about 20%, about 0.5% to about 10%, about 0.5% to about 5%, or about 0.5% to about 4% by mass of the glucose in the culture medium. In particular embodiments, the fatty acid ethyl ester is produced at a yield of at least about 0.5%, at least about 4%, at least about 5%, at least about 10%, at least about 20%, at least about 30%, at least about 40%, or at least about 50% by mass of glucose in the culture medium. In particular embodiments, the fatty acid ethyl ester is produced at a yield of no more than 50% by mass of glucose in the culture medium.

[0056] In some embodiments, the fatty acid ethyl ester is produced at a yield of about 10% to about 95% by mass of carbon in the carbon source in the culture medium. For example, the fatty acid ethyl ester is produced at a yield of about 10% or more (e.g., of about 10% or more, of about 15% or more, of about 20% or more, of about 25% or more) by pass of carbon in the carbon source in the culture medium. In particular embodiments, the fatty acid ethyl ester is produced at a yield of about 15% to about 90%, about 20% to about 80%, or about 30% to about 70% by mass of carbon in the carbon source in the culture medium. In particular embodiments, the fatty acid ethyl ester is produced at a yield of at least about 10%, at least about 20%, at least about 30%, at least about 40%, at least about 50%, at least about 60%, at least about 70%, at least about 80%, at least about 90%, or at least about 95% by mass of carbon in the carbon source in the culture medium. In particular embodiments, the fatty acid ethyl ester is produced at a yield of no more than 95% by mass of carbon in the carbon source in the culture medium.

[0057] In some embodiments, the fatty ester is a fatty acid methyl ester and is produced at a yield of about 0.5 g to about 50 g of fatty acid methyl ester per 100 g of glucose in the culture medium. For example, the fatty ester is a fatty acid methyl ester and is produced at a yield of about 0.5 g or more (e.g., about 0.5 g or more, about 1 g or more, about 2 g or more, about 5 g or more, about 10 g or more) of fatty acid methyl ester per 100 g of glucose in the culture medium. In particular embodiments, the fatty acid methyl ester is produced at a yield of about 0.5 g to about 40 g of fatty acid methyl ester per 100 g of glucose, about 0.5 g to about 30 g of fatty acid methyl ester per 100 g of glucose, about 0.5 g to about 20 g of fatty acid methyl ester per 100 g of glucose, about 0.5 g to about 10 g of fatty acid methyl ester per 100 g of glucose, about 0.5 g to about 5 g of fatty acid methyl ester per 100 g of glucose, or about 0.5 g to about 4 g of fatty acid methyl ester per 100 g of glucose in the culture medium. In particular embodiments, the fatty acid methyl ester is produced at a yield of at least 0.5 g of fatty acid methyl ester, at least 4 g of fatty acid methyl ester, at least 5 g of fatty acid methyl ester, at least 10 g of fatty acid methyl ester, at least 20

g of fatty acid methyl ester, at least 30 g of fatty acid methyl ester, at least 40 g of fatty acid methyl ester, or at least 50 g of fatty acid methyl ester per 100 g of glucose in the culture medium. In particular embodiments, the fatty acid methyl ester is produced at a yield of no more than 50 g of fatty acid methyl ester per 100 g of glucose in the culture medium.

[0058] In some embodiments, the fatty acid methyl ester is produced at a yield of about 0.5% to about 50% by mass of the glucose in the culture medium. For example, fatty acid methyl ester is produced at a yield of about 0.5% or more (e.g., about 0.5% or more, about 1% or more, about 2% or more, about 5% or more, about 10% or more, about 15% or more) by mass of the glucose in the culture medium. In particular embodiments, the fatty acid methyl ester is produced at a yield of about 0.5% to about 40%, about 0.5% to about 30%, about 0.5% to about 20%, about 0.5% to about 10%, about 0.5% to about 5%, or about 0.5% to about 4% by mass of the glucose in the culture medium. In particular embodiments, the fatty acid methyl ester is produced at a yield of at least about 0.5%, at least about 4%, at least about 5%, at least about 10%, at least about 20%, at least about 30%, at least about 40%, or at least about 50% by mass of glucose in the culture medium. In particular embodiments, the fatty acid methyl ester is produced at a yield of no more than 50% by mass of glucose in the culture medium.

[0059] In some embodiments, the fatty acid methyl ester is produced at a yield of about 10% to about 95% by mass of carbon in the carbon source in the culture medium. For example, the fatty acid methyl ester is produced at a yield of about 10% or more (e.g., about 10% or more, about 20% or more, about 25% or more, about 30% or more, about 35% or more, about 40% or more) by mass of carbon in the carbon source in the culture medium. In particular embodiments, the fatty acid methyl ester is produced at a yield of about 15% to about 90%, about 20% to about 80%, or about 30% to about 70% by mass of carbon in the carbon source in the culture medium. In particular embodiments, the fatty acid methyl ester is produced at a yield of at least about 10%, at least about 20%, at least about 30%, at least about 40%, at least about 50%, at least about 60%, at least about 70%, at least about 80%, at least about 90%, or at least about 95% by mass of carbon in the carbon source in the culture medium. In particular embodiments, the fatty acid methyl ester is produced at a yield of no more than 95% by mass of carbon in the carbon source in the culture medium.

[0060] In another aspect, the invention features a fatty ester, such as a fatty acid ester, produced by a method described herein. In some embodiments, the fatty acid ester is a fatty acid methyl ester. In some embodiments, the fatty acid methyl ester is at least about 4, 6, 8, 10, 12, 14, 16, or 18 carbons in length.

[0061] In some embodiments, the fatty ester comprises an A side (i.e., the carbon chain attached to the carboxylate oxygen) and a B side (i.e., the carbon chain comprising the parent carboxylate). In some embodiments, the B side of the fatty acid methyl ester is at least about 4, 6, 8, 10, 12, 14, 16, or 18 carbons in length. In some embodiments, the B side of the fatty acid methyl ester includes a straight chain. In other embodiments, the B side of the fatty acid methyl ester includes a branched chain. In still other embodiments, the B side of the fatty acid methyl ester comprises at least one cyclic moiety. In further embodiments, the fatty ester is selected from methyl dedecanoate, methyl 5-dodecenoate, methyl tetradecanoate, methyl 7-tetradecenoate, methyl hexadecanoate, methyl 9-hexadecenoate, methyl octadecanoate, methyl 11-octadecenoate, or combinations thereof. In yet further embodiments, the fatty ester is selected from ethyl dedecanoate, ethyl 5-dodecenoate, ethyl tetradecanoate, ethyl 7-tetradecenoate, ethyl hexadecanoate, ethyl 9-hexadecenoate, ethyl octadecanoate, ethyl 11-octadecenoate, or combinations thereof.

[0062] In some embodiments, the fatty acid methyl ester is saturated. In other embodiments, the fatty acid methyl ester is unsaturated. In other embodiments, the fatty acid methyl ester is monounsaturated. In certain embodiments, the fatty acid ethyl ester is saturated. In other embodiments, the fatty acid ethyl ester is unsaturated. In other embodiments, the fatty acid ethyl ester is monounsaturated.

[0063] In a further aspect, the invention features a method of producing a biologically-derived diesel fuel of commercial quality according to commercial standards (e.g., ASTM or ANP). In some embodiments, the method comprises fermenting carbohydrates using a genetically modified microorganism described herein. The process provides a direct route for producing fatty esters, for example, fatty acid esters such as fatty acid methyl esters or fatty acid ethyl esters, without the need of producing oils, which are later chemically transesterified with the concomitant production of large quantities of glycerol. The fuel composition thus produced can be utilized as a diesel fuel alone, or be blended with petroleum diesel according to customary proportions, resulting in clean emission profiles and low amounts of impurities and/or undesirable contaminants.

[0064] In another aspect, the invention features a fuel composition, including, for example, a diesel composition, comprising a fatty ester produced by a method or a genetically engineered microorganism described herein. In certain embodiments, the fuel composition further comprises one or more suitable fuel additives.

[0065] The drawings and examples provided herein are intended solely to illustrate the features of the present invention. They are not intended to be limiting.

DETAILED DESCRIPTION OF THE INVENTION

[0066] Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Although methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, suitable methods and materials are described below. All publications, patent applications, patents, and other references mentioned herein, including GenBank database sequences, are incorporated by reference in their entirety. In case of conflict, the present specification, including definitions, will control. In addition, the materials, methods, and examples are illustrative only and not intended to be limiting.

[0067] Other features and advantages of the invention will be apparent from the following detailed description, and from the claims.

DEFINITIONS

[0068] The articles "a" and "an" are used herein to refer to one or to more than one (i.e., to at least one) of the grammatical object of the article. By way of example, "an element" means one element or more than one element.

[0069] The term "about" is used herein to mean a value ±20% of a given numerical value. Thus, "about 60%" means a value of between 60±(20% of 60) (i.e., between 48 and 72).

[0070] As used herein, the term "attenuate" means to weaken, reduce, or diminish. For example, a polypeptide can be attenuated by modifying the polypeptide to reduce its activity (e.g., by modifying a nucleotide sequence that encodes the polypeptide).

[0071] As used herein, the term "biocrude" refers to a product derived from biomass, biomass derivatives, or other biological sources that, like petroleum crude, can be converted into other fuels. For example, biocrude can be converted into gasoline, diesel, jet fuel, or heating oil. Moreover, biocrude, like petroleum crude, can be converted into other industrially useful chemicals for use in, for example, pharmaceuticals, cosmetics, consumer goods, industrial processes, and the like.

[0072] Biocrude may include, for example, hydrocarbons, hydrocarbon products, fatty acid esters, and/or aliphatic ketones. In a preferred embodiment, biocrude is comprised of hydrocarbons, for example aliphatic (e.g., alkanes, alkenes, alkynes) or aromatic hydrocarbons.

[0073] As used herein, the term "biodiesel" means a biofuel that can be a substitute of diesel, which is derived from petroleum. Biodiesel can be used in internal combustion diesel engines in either a pure form, which is referred to as "neat" biodiesel, or as a mixture in any concentration with petroleum-based diesel.

[0074] In one embodiment, biodiesel can include esters or hydrocarbons, such as aldehydes, alkanes, or alkenes.

[0075] As used herein, the term "biofuel" refers to any fuel derived from biomass, biomass derivatives, or other biological sources. Biofuels can be substituted for petroleum based fuels. For example, biofuels are inclusive of transportation fuels (e.g., gasoline, diesel, jet fuel, etc.), heating fuels, and electricity-generating fuels. Biofuels are a renewable energy source.

[0076] As used herein, the term "biomass" refers to a carbon source derived from biological material. Biomass can be converted into a biofuel. One exemplary source of biomass is plant matter. For example, corn, sugar cane, or switchgrass can be used as biomass. Another non-limiting example of biomass is animal matter, for example cow manure. Biomass also includes waste products from industry, agriculture, forestry, and households. Examples of such waste products that can be used as biomass are fermentation waste, straw, lumber, sewage, garbage, and food leftovers. Biomass also includes sources of carbon, such as carbohydrates (e.g., monosaccharides, disaccharides, or polysaccharides).

[0077] As used herein, the phrase "carbon source" refers to a substrate or compound suitable to be used as a source of carbon for prokaryotic or simple eukaryotic cell growth. Carbon sources can be in various forms, including, but not limited to polymers, carbohydrates, acids, alcohols, aldehydes, ketones, amino acids, peptides, and gases (e.g., CO and CO_2). These include, for example, various monosaccharides, such as glucose, fructose, mannose, and galactose; oligosaccharides, such as fructo-oligosaccharide and galacto-oligosaccharide; polysaccharides such as xylose and arabinose; disaccharides, such as sucrose, maltose, and turanose; cellulosic material, such as methyl cellulose and sodium carboxymethyl cellulose; saturated or unsaturated fatty acid esters, such as succinate, lactate, and acetate; alcohols, such as methanol, ethanol, propanol, or mixtures thereof. The carbon source can also be a product of photosynthesis, including, but not limited to, glucose. A preferred carbon source is biomass. Another preferred carbon source is glucose.

[0078] As used herein, a "cloud point lowering additive" is an additive added to a composition to decrease or lower the cloud point of a solution.

[0079] As used herein, the phrase "cloud point of a fluid" means the temperature at which dissolved solids are no longer completely soluble. Below this temperature, solids begin precipitating as a second phase giving the fluid a cloudy appearance. In the petroleum industry, cloud point refers to the temperature below which a solidified material or other heavy hydrocarbon crystallizes in a crude oil, refined oil, or fuel to form a cloudy appearance. The presence of solidified materials influences the flowing behavior of the fluid, the tendency of the fluid to clog fuel filters, injectors, etc., the accumulation of solidified materials on cold surfaces (e.g., a pipeline or heat exchanger fouling), and the emulsion characteristics of the fluid with water.

[0080] A nucleotide sequence is "complementary" to another nucleotide sequence if each of the bases of the two sequences matches (e.g., is capable of forming Watson Crick base pairs). The term "complementary strand" is used herein interchangeably with the term "complement". The complement of a nucleic acid strand can be the complement of a coding strand or the complement of a non-coding strand.

[0081] The terms "comprising," "having," "including," and "containing" are to be construed as open-ended terms (e.g., meaning "including, but not limited to,") unless otherwise noted.

[0082] As used herein, the term "conditions sufficient to allow expression" means any conditions that allow a host cell to produce a desired product, such as a polypeptide, aldehyde, or alkane described herein. Suitable conditions include, for example, fermentation conditions. Fermentation conditions can comprise many parameters, such as temperature ranges, levels of aeration, and media composition. Each of these conditions, individually and in combination, allows the host cell to grow. Exemplary culture media include broths or gels. Generally, the medium includes a carbon source, such as glucose, fructose, cellulose, or the like, that can be metabolized by a host cell directly. In addition, enzymes can be used in the medium to facilitate the mobilization (e.g., the depolymerization of starch or cellulose to fermentable sugars) and subsequent metabolism of the carbon source.

[0083] To determine if conditions are sufficient to allow expression, a host cell can be cultured, for example, for about 4, 8, 12, 24, 36, or 48 hours. During and/or after culturing, samples can be obtained and analyzed to determine if the conditions allow expression. For example, the host cells in the sample or the medium in which the host cells were grown can be tested for the presence of a desired product. When testing for the presence of a product, assays, such as, but not limited to, TLC, HPLC, GC/FID, GC/MS, LC/MS, MS, can be used.

[0084] It is understood that the polypeptides described herein may have additional conservative or non-essential amino acid substitutions, which do not have a substantial effect on the polypeptide functions. Whether or not a particular substitution will be tolerated (e.g., will not adversely affect desired biological properties, such as decarboxylase activity) can be determined as described in Bowie et al., *Science* (1990) 247:1306 1310. A "conservative amino acid substitution" is one in which the amino acid residue is replaced with an amino acid residue having a similar side chain. Families of

amino acid residues having similar side chains have been defined in the art. These families include amino acids with basic side chains (e.g., lysine, arginine, and histidine), acidic side chains (e.g., aspartic acid and glutamic acid), uncharged polar side chains (e.g., glycine, asparagine, glutamine, serine, threonine, tyrosine, and cysteine), nonpolar side chains (e.g., alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, and tryptophan), beta-branched side chains (e.g., threonine, valine, and isoleucine), and aromatic side chains (e.g., tyrosine, phenylalanine, tryptophan, and histidine).

[0085] As used herein, "conditions that permit product production" refers to any fermentation conditions that allow a production host to produce a desired product, such as acyl-CoA or fatty acid derivatives (e.g., fatty acids, hydrocarbons, fatty alcohols, waxes, or fatty esters). Fermentation conditions usually comprise many parameters. Exemplary conditions include, but are not limited to, temperature ranges, levels of aeration, and media composition. Each of these conditions, individually and/or in combination, allows the production host to grow.

[0086] Exemplary media include broths and/or gels. Generally, a suitable medium includes a carbon source (e.g., glucose, fructose, cellulose, etc.) that can be metabolized by the microorganism directly. In addition, enzymes can be used in the medium to facilitate the mobilization (e.g., the depolymerization of starch or cellulose to fermentable sugars) and subsequent metabolism of the carbon source.

[0087] To determine if the fermentation conditions permit product production, the production host can be cultured for about 4, 8, 12, 24, 36, 48 or 96 hours. During culturing or after culturing, samples can be obtained and analyzed to determine if the fermentation conditions have permitted product production. For example, the production hosts in the sample or the medium in which the production hosts are grown can be tested for the presence of the desired product. Exemplary assays, such as TLC, HPLC, GC/FID, GC/MS, LC/MS, MS, as well as those provided herein, can be used identify and quantify the presence of a product.

[0088] As used herein, "control element" means a transcriptional control element. Control elements include promoters and enhancers. The term "promoter element," "promoter," or "promoter sequence" refers to a DNA sequence that functions as a switch that activates the expression of a gene. If the gene is activated, it is said to be transcribed or participating in transcription. Transcription involves the synthesis of mRNA from the gene. A promoter, therefore, serves as a transcriptional regulatory element and also provides a site for initiation of transcription of the gene into mRNA. Control elements interact specifically with cellular proteins involved in transcription (Maniatis et al., *Science* 236:1237 (1987)).

[0089] As used herein, the term "deletion," or "knockout" means modifying or inactivating a polynucleotide sequence that encodes a target protein in order to reduce or eliminate the function of the target protein. A polynucleotide deletion can be performed by methods well known in the art (See, e.g., Datsenko et al., *Proc. Nat. Acad. Sci. USA,* 97:6640-45 (2000), or International Patent Application Nos. PCT/US2007/011923 and PCT/US2008/058788).

[0090] As used herein, the term "endogenous" means a polynucleotide that is in the cell and was not introduced into the cell using recombinant genetic engineering techniques. For example, a gene that was present in the cell when the cell was originally isolated from nature. A polynucleotide is still considered endogenous if the control sequences, such as a promoter or enhancer sequences which activate transcription or translation, have been altered through recombinant techniques.

[0091] As used herein, the term "ester synthase" means a peptide capable of producing fatty esters. More specifically, an ester synthase is a peptide which converts a thioester to a fatty ester. In a preferred embodiment, the ester synthase converts a thioester (e.g., acyl-CoA) to a fatty ester.

[0092] In an alternate embodiment, an ester synthase uses a thioester and an alcohol as substrates to produce a fatty ester. Ester synthases are capable of using short and long chain thioesters as substrates. In addition, ester synthases are capable of using short and long chain alcohols as substrates.

[0093] Non-limiting examples of ester synthases are wax synthases, wax-ester synthases, acyl CoA:alcohol transacylases, acyltransferases, and fatty acyl-coenzyme A:fatty alcohol acyltransferases. Exemplary ester synthases are classified in enzyme classification number EC 2.3.1.75. A number of these enzymes, as well as other useful enzymes for making the products described herein, have been disclosed in, for example, International Patent Application Nos. PCT/US2007/011923 and PCT/US2008/058788, which are incorporated herein by reference.

[0094] As used herein, the term "fatty acid" means a carboxylic acid having the formula RCOOH. R represents an aliphatic group, preferably an alkyl group. R can comprise between about 4 and about 22 carbon atoms. Fatty acids can be saturated, monounsaturated, or polyunsaturated. In a preferred embodiment, the fatty acid is made from a fatty acid biosynthetic pathway.

[0095] As used herein, the term "fatty acid biosynthetic pathway" means a biosynthetic pathway that produces fatty acids. The fatty acid biosynthetic pathway includes fatty acid enzymes that can be engineered, as described herein, to produce fatty acids, and in some embodiments can be expressed with additional enzymes to produce fatty acids having desired carbon chain characteristics.

[0096] As used herein, the term "fatty acid degradation enzyme" means an enzyme involved in the breakdown or conversion of a fatty acid or fatty acid derivative into another product. A non-limiting example of a fatty acid degradation enzyme is an acyl-CoA synthase. A number of these enzymes, as well as other useful enzymes for making the products described herein, have been disclosed in, for example, International Patent Application Nos. PCT/US2007/011923 and PCT/US2008/058788, which are incorporated herein by reference. Additional examples of fatty acid degradation enzymes are described herein.

[0097] As used herein, the term "fatty acid derivative" means products made in part from the fatty acid biosynthetic pathway of the production host organism. "Fatty acid derivative" also includes products made in part from acyl-ACP or acyl-ACP derivatives. The fatty acid biosynthetic pathway includes fatty acid synthase enzymes which can be engineered as described herein to produce fatty acid derivatives, and in some examples can be expressed with additional enzymes to produce fatty acid derivatives having desired carbon chain characteristics. Exemplary fatty acid derivatives include for example, fatty acids, acyl-CoAs, fatty aldehydes, short and long chain alcohols, hydrocarbons, fatty alcohols, ketones, and esters (e.g., waxes, fatty acid esters, or fatty esters).

[0098] As used herein, the term "fatty acid derivative enzymes" means all enzymes that may be expressed or over-

expressed in the production of fatty acid derivatives. These enzymes are collectively referred to herein as fatty acid derivative enzymes. These enzymes may be part of the fatty acid biosynthetic pathway. Non-limiting examples of fatty acid derivative enzymes include fatty acid synthases, thioesterases, acyl-CoA synthases, acyl-CoA reductases, alcohol dehydrogenases, alcohol acyltransferases, carboxylic acid reductases, fatty alcohol-forming acyl-CoA reductase, ester synthases, aldehyde biosynthetic polypeptides, and alkane biosynthetic polypeptides. Fatty acid derivative enzymes convert a substrate into a fatty acid derivative. In some examples, the substrate may be a fatty acid derivative which the fatty acid derivative enzyme converts into a different fatty acid derivative. A number of these enzymes, as well as other useful enzymes for making the products described herein, have been disclosed in, for example, International Patent Application Nos. PCT/US2007/011923 and PCT/US2008/058788, which are incorporated herein by reference.

[0099] As used herein, "fatty acid enzyme" means any enzyme involved in fatty acid biosynthesis. Fatty acid enzymes can be expressed or overexpressed in host cells to produce fatty acids. Non-limiting examples of fatty acid enzymes include fatty acid synthases and thioesterases. A number of these enzymes, as well as other useful enzymes for making the products described herein, have been disclosed in, for example, International Patent Application Nos. PCT/US2007/011923 and PCT/US2008/058788, which are incorporated herein by reference.

[0100] As used herein, the term "fatty ester" means an ester. In a preferred embodiment, a fatty ester is any ester made from a fatty acid to produce, for example, a fatty acid ester. In one embodiment, a fatty ester contains an A side (i.e., the carbon chain attached to the carboxylate oxygen) and a B side (i.e., the carbon chain comprising the parent carboxylate). In a preferred embodiment, when the fatty ester is derived from the fatty acid biosynthetic pathway, the A side is contributed by an alcohol, and the B side is contributed by a fatty acid. Any alcohol can be used to form the A side of the fatty esters. For example, the alcohol can be derived from the fatty acid biosynthetic pathway. Alternatively, the alcohol can be produced through non-fatty acid biosynthetic pathways. Moreover, the alcohol can be provided exogenously. For example, the alcohol can be supplied in the fermentation broth in instances where the fatty ester is produced by an organism that can also produce the fatty acid. Alternatively, a carboxylic acid, such as a fatty acid or acetic acid, can be supplied exogenously in instances where the fatty ester is produced by an organism that can also produce alcohol.

[0101] The carbon chains comprising the A side or B side can be of any length. In one embodiment, the A side of the ester is at least about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 14, 16, 18 or 20 carbons in length. The B side of the ester is at least about 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, or 26 carbons in length. The A side and/or the B side can be straight or branched chain. The branched chains may have one or more points of branching. In addition, the branched chains may include cyclic branches. Furthermore, the A side and/or B side can be saturated or unsaturated. If unsaturated, the A side and/or B side can have one or more points of unsaturation.

[0102] In one embodiment, the fatty ester is produced biosynthetically. In this embodiment, first the fatty acid is "activated." Non-limiting examples of "activated" fatty acids are acyl-CoA, acyl-ACP, and acyl phosphate. Acyl-CoA can be a direct product of fatty acid biosynthesis or degradation. In addition, acyl-CoA can be synthesized from a free fatty acid, a CoA, or an adenosine nucleotide triphosphate (ATP). An example of an enzyme which produces acyl-CoA is acyl-CoA synthase.

[0103] After the fatty acid is activated, it can be readily transferred to a recipient nucleophile. Exemplary nucleophiles are alcohols, thiols, or phosphates.

[0104] In one embodiment, the fatty ester is a wax. The wax can be derived from a long chain alcohol and a long chain fatty acid. In another embodiment, the fatty ester can be derived from a fatty acyl-thioester and an alcohol. In another embodiment, the fatty ester is a fatty acid thioester, for example fatty acyl Coenzyme A (CoA). In other embodiments, the fatty ester is a fatty acyl panthothenate, an acyl carrier protein (ACP), or a fatty phosphate ester. Fatty esters have many uses. For example, fatty esters can be used as biofuels, surfactants, or formulated into additives that provide lubrication and other benefits to fuels and industrial chemicals.

[0105] As used herein, "fraction of modern carbon" or "f_M" has the same meaning as defined by National Institute of Standards and Technology (NIST) Standard Reference Materials (SRMs) 4990B and 4990C, known as oxalic acids standards HOxI and HOxII, respectively. The fundamental definition relates to 0.95 times the $^{14}C/^{12}C$ isotope ratio HOxI (referenced to AD 1950). This is roughly equivalent to decay-corrected pre-Industrial Revolution wood. For the current living biosphere (e.g., plant material), f_M is approximately 1.1.

[0106] The genes "fhuA" and "tonA", which encode a ferrichrome outer membrane transporter (GenBank Accession No. NP_414692), are used interchangeably herein.

[0107] "Gene knockout", as used herein, refers to a procedure by which a gene encoding a target protein is modified or inactivated so to reduce or eliminate the function of the intact protein. Inactivation of the gene may be performed by general methods such as mutagenesis by UV irradiation or treatment with N-methyl-N'-nitro-N-nitrosoguanidine, site-directed mutagenesis, homologous recombination, insertion-deletion mutagenesis, or "Red-driven integration" (Datsenko et al., *Proc. Natl. Acad. Sci. USA,* 97:6640-45 (2000)). For example, in one embodiment, a construct is introduced into a host cell, such that it is possible to select for homologous recombination events in the host cell. One of skill in the art can readily design a knock-out construct including both positive and negative selection genes for efficiently selecting transfected cells that undergo a homologous recombination event with the construct. The alteration in the host cell may be obtained, for example, by replacing through a single or double crossover recombination a wild type DNA sequence by a DNA sequence containing the alteration. For convenient selection of transformants, the alteration may, for example, be a DNA sequence encoding an antibiotic resistance marker or a gene complementing a possible auxotrophy of the host cell. Mutations include, but are not limited to, deletion-insertion mutations. An example of such an alteration includes a gene disruption, i.e., a perturbation of a gene such that the product that is normally produced from this gene is not produced in a functional form. This could be due to a complete deletion, a deletion and insertion of a selective marker, an insertion of a selective marker, a frameshift mutation, an in-frame deletion, or a point mutation that leads to premature termination. In some instances, the entire mRNA for the gene is absent. In other situations, the amount of mRNA produced varies.

[0108] Calculations of "homology" between two sequences can be performed as follows. The sequences are aligned for optimal comparison purposes (e.g., gaps can be introduced in one or both of a first and a second amino acid or nucleic acid sequence for optimal alignment and non-homologous sequences can be disregarded for comparison purposes). In a preferred embodiment, the length of a reference sequence that is aligned for comparison purposes is at least about 30%, preferably at least about 40%, more preferably at least about 50%, even more preferably at least about 60%, and even more preferably at least about 70%, at least about 80%, at least about 90%, or about 100% of the length of the reference sequence. The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are identical at that position (as used herein, amino acid or nucleic acid "identity" is equivalent to amino acid or nucleic acid "homology"). The percent identity between the two sequences is a function of the number of identical positions shared by the sequences, taking into account the number of gaps and the length of each gap, which need to be introduced for optimal alignment of the two sequences.

[0109] The comparison of sequences and determination of percent homology between two sequences can be accomplished using a mathematical algorithm. In a preferred embodiment, the percent homology between two amino acid sequences is determined using the Needleman and Wunsch, *J. Mol. Biol.* 48:444 453 (1970), algorithm that has been incorporated into the GAP program in the GCG software package, using either a Blossum 62 matrix or a PAM250 matrix, and a gap weight of 16, 14, 12, 10, 8, 6, or 4 and a length weight of 1, 2, 3, 4, 5, or 6. In yet another preferred embodiment, the percent homology between two nucleotide sequences is determined using the GAP program in the GCG software package, using a NWSgapdna. CMP matrix and a gap weight of about 40, 50, 60, 70, or 80 and a length weight of about 1, 2, 3, 4, 5, or 6. A particularly preferred set of parameters (and the one that should be used if the practitioner is uncertain about which parameters should be applied to determine if a molecule is within a homology limitation of the claims) are a Blossum 62 scoring matrix with a gap penalty of 12, a gap extend penalty of 4, and a frameshift gap penalty of 5.

[0110] Other methods for aligning sequences for comparison are well known in the art. Various programs and alignment algorithms are described in, for example, Smith & Waterman, *Adv. Appl. Math.* 2:482 (1981); Pearson & Lipman, *Proc. Natl. Acad. Sci.* USA 85:2444 (1988); Higgins & Sharp, *Gene* 73:237 244 (1988); Higgins & Sharp, *CABIOS* 5:151-153 (1989); Corpet et al., *Nucleic Acids Research* 16:10881-10890 (1988); Huang et al., *CABIOS* 8:155-165, 1992; and Pearson et al., *Methods in Molecular Biology* 24:307-331, 1994. and Altschul et al., *J. Mol. Biol.* 215:403-410, 1990.

[0111] As used herein, a "host cell" is a cell used to produce a product described herein (e.g., an aldehyde or alkane). A host cell can be modified to express or overexpress selected genes or to have attenuated expression of selected genes. Non-limiting examples of host cells include plant, animal, human, bacteria, cyanobacteria, yeast, or filamentous fungi cells.

[0112] As used herein, the term "hybridizes under low stringency, medium stringency, high stringency, or very high stringency conditions" describes conditions for hybridization and washing. Guidance for performing hybridization reactions can be found, for example, in *Current Protocols in Molecular Biology*, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. Aqueous and nonaqueous methods are described in that reference and either method can be used. Specific hybridization conditions referred to herein are as follows: 1) low stringency hybridization conditions in 6× sodium chloride/sodium citrate (SSC) at about 45° C., followed by two washes in 0.2×SSC, 0.1% SDS at least at 50° C. (the temperature of the washes can be increased to 55° C. for low stringency conditions); 2) medium stringency hybridization conditions in 6×SSC at about 45° C., followed by one or more washes in 0.2×SSC, 0.1% SDS at 60° C.; 3) high stringency hybridization conditions in 6×SSC at about 45° C., followed by one or more washes in 0.2×SSC, 0.1% SDS at 65° C.; and preferably 4) very high stringency hybridization conditions are 0.5M sodium phosphate, 7% SDS at 65° C., followed by one or more washes at 0.2×SSC, 1% SDS at 65° C. Very high stringency conditions (4) are the preferred conditions unless otherwise specified.

[0113] The term "isolated" as used herein with respect to nucleic acids, such as DNA or RNA, refers to molecules separated from other DNAs or RNAs, respectively, that are present in the natural source of the nucleic acid. Moreover, an "isolated nucleic acid" includes nucleic acid fragments, such as fragments that are not naturally occurring. The term "isolated" is also used herein to refer to polypeptides, which are isolated from other cellular proteins, and encompasses both purified endogenous polypeptides and recombinant polypeptides. The term "isolated" as used herein also refers to a nucleic acid or polypeptide that is substantially free of cellular material, viral material, or culture medium when produced by recombinant DNA techniques. The term "isolated" as used herein also refers to a nucleic acid or polypeptide that is substantially free of chemical precursors or other chemicals when chemically synthesized.

[0114] As used herein, the "level of expression of a gene in a cell" refers to the level of mRNA, pre-mRNA nascent transcript(s), transcript processing intermediates, mature mRNA (s), and/or degradation products encoded by the gene in the cell.

[0115] As used herein, the term "microorganism" means prokaryotic and eukaryotic microbial species from the domains Archaea, Bacteria and Eucarya, the latter including yeast and filamentous fungi, protozoa, algae, or higher Protista. The term "microbial cell", as used herein, means a cell from a microorganism.

[0116] As used herein, the term "nucleic acid" refers to a polynucleotide, such as deoxyribonucleic acid (DNA), and, where appropriate, ribonucleic acid (RNA). The term also includes analogs of either RNA or DNA made from nucleotide analogs, and, as applicable to the embodiment being described, single (sense or antisense) and double-stranded polynucleotides, ESTs, chromosomes, cDNAs, mRNAs, and rRNAs. The term "nucleic acid" may be used interchangeably with "polynucleotide," "DNA," "nucleic acid molecule," "nucleotide sequence," and/or "gene" unless otherwise indicated herein or otherwise clearly contradicted by context.

[0117] As used herein, the term "operably linked" means that a selected nucleotide sequence (e.g., encoding a polypeptide described herein) is in proximity with a promoter to allow the promoter to regulate expression of the selected nucleotide sequence. In addition, the promoter is located upstream of the

selected nucleotide sequence in terms of the direction of transcription and translation. By "operably linked" is meant that a nucleotide sequence and a regulatory sequence(s) are connected in such a way as to permit gene expression when the appropriate molecules (e.g., transcriptional activator proteins) are bound to the regulatory sequence(s).

[0118] The term "or" is used herein to mean, and is used interchangeably with, the term "and/or," unless context clearly indicates otherwise.

[0119] As used herein, "overexpress" means to express or cause to be expressed or produced a nucleic acid, polypeptide, or hydrocarbon in a cell at a greater concentration than is normally expressed in a corresponding wild-type cell. For example, a polypeptide can be "overexpressed" in a recombinant host cell when the polypeptide is present in a greater concentration in the recombinant host cell compared to its concentration in a non-recombinant host cell of the same species.

[0120] As used herein, "partition coefficient" or "P," is defined as the equilibrium concentration of a compound in an organic phase divided by the concentration at equilibrium in an aqueous phase (e.g., fermentation broth). In one embodiment of a bi-phasic system described herein, the organic phase is formed by the aldehyde or alkane during the production process. However, in some examples, an organic phase can be provided, such as by providing a layer of octane, to facilitate product separation. When describing a two phase system, the partition characteristics of a compound can be described as logP. For example, a compound with a logP of 1 would partition 10:1 to the organic phase. A compound with a logP of −1 would partition 1:10 to the organic phase. By choosing an appropriate fermentation broth and organic phase, an organic fatty acid derivative or product with a high logP value can separate into the organic phase even at very low concentrations in the fermentation vessel.

[0121] As used herein, the term "polypeptide" may be used interchangeably with "protein," "peptide," and/or "enzyme" unless otherwise indicated herein or otherwise clearly contradicted by context.

[0122] As used herein, the term "production host" means a cell used to produce the products disclosed herein. The production host is modified to express, overexpress, attenuate or delete expression of selected polynucleotides. Non-limiting examples of production hosts include plant, algal, animal, human, bacteria, yeast, and filamentous fungi cells.

[0123] As used herein, the term "purify," "purified," or "purification" means the removal or isolation of a molecule from its environment by, for example, isolation or separation. "Substantially purified" molecules are at least about 60% free, preferably at least about 75% free, and more preferably at least about 90% free from other components with which they are associated. As used herein, these terms also refer to the removal of contaminants from a sample. For example, the removal of contaminants can result in an increase in the percentage of a fatty acid derivative or product in a sample. For example, when a fatty acid derivatives or products are produced in a host cell, the fatty acid derivatives or products can be purified by the removal of host cell proteins. After purification, the percentage of fatty acid derivatives or products in the sample is increased.

[0124] The terms "purify," "purified," and "purification" do not require absolute purity. They are relative terms. Thus, for example, when the fatty acid derivatives or products are produced in host cells, a purified fatty acid derivative or product is one that is substantially separated from other cellular components (e.g., nucleic acids, polypeptides, lipids, carbohydrates, or other fatty acid derivatives or products). In another example, a purified fatty acid derivative or purified product preparation is one in which the fatty acid derivative or product is substantially free from contaminants, such as those that might be present following fermentation. In some embodiments, a fatty acid derivative or product is purified when at least about 50% by weight of a sample is composed of the fatty acid derivative or product. In other embodiments, a fatty acid derivative or product is purified when at least about 60%, 70%, 80%, 85%, 90%, 92%, 95%, 98%, or 99% or more by weight of a sample is composed of the fatty acid derivative or product.

[0125] As used herein, the term "recombinant polypeptide" refers to a polypeptide that is produced by recombinant DNA techniques, wherein generally DNA encoding the expressed polypeptide or RNA is inserted into a suitable expression vector and that is in turn used to transform a host cell to produce the polypeptide or RNA.

[0126] As used herein, the term "substantially identical" (or "substantially homologous") is used to refer to a first amino acid or nucleotide sequence that contains a sufficient number of identical or equivalent (e.g., with a similar side chain) amino acid residues (e.g., conserved amino acid substitutions) or nucleotides to a second amino acid or nucleotide sequence such that the first and second amino acid or nucleotide sequences have similar activities.

[0127] As used herein, the term "synthase" means an enzyme which catalyzes a synthesis process. As used herein, the term synthase includes synthases, synthetases, and ligases.

[0128] As used herein, the term "transfection" means the introduction of a nucleic acid (e.g., via an expression vector) into a recipient cell by nucleic acid-mediated gene transfer.

[0129] As used herein, the term "transformation" refers to a process in which a cell's genotype is changed as a result of the cellular uptake of exogenous nucleic acid. This may result in the transformed cell expressing a recombinant form of a RNA or polypeptide. In the case of antisense expression from the transferred gene, the expression of a naturally-occurring form of the polypeptide is disrupted.

[0130] As used herein, the term "transport protein" means a polypeptide that facilitates the movement of one or more compounds in and/or out of a cellular organelle and/or a cell. A number of these proteins, as well as other useful proteins for making the products described herein, have been disclosed in, for example, International Patent Application Nos. PCT/US2007/011923 and PCT/US2008/058788, which are incorporated herein by reference.

[0131] As used herein, a "variant" of polypeptide X refers to a polypeptide having the amino acid sequence of polypeptide X in which one or more amino acid residues is altered. The variant may have conservative changes or non-conservative changes. Guidance in determining which amino acid residues may be substituted, inserted, or deleted without affecting biological activity may be found using computer programs well known in the art, for example, LASERGENE software (DNASTAR).

[0132] The term "variant," when used in the context of a polynucleotide sequence, may encompass a polynucleotide sequence related to that of a gene or the coding sequence thereof. This definition may also include, for example, "allelic," "splice," "species," or "polymorphic" variants. A

splice variant may have significant identity to a reference polynucleotide, but will generally have a greater or fewer number of polynucleotides due to alternative splicing of exons during mRNA processing. The corresponding polypeptide may possess additional functional domains or an absence of domains. Species variants are polynucleotide sequences that vary from one species to another. The resulting polypeptides generally will have significant amino acid identity relative to each other. A polymorphic variant is a variation in the polynucleotide sequence of a particular gene between individuals of a given species.

[0133] As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of useful vector is an episome (i.e., a nucleic acid capable of extra-chromosomal replication). Useful vectors are those capable of autonomous replication and/or expression of nucleic acids to which they are linked. Vectors capable of directing the expression of genes to which they are operatively linked are referred to herein as "expression vectors". In general, expression vectors of utility in recombinant DNA techniques are often in the form of "plasmids," which refer generally to circular double stranded DNA loops that, in their vector form, are not bound to the chromosome. In the present specification, "plasmid" and "vector" are used interchangeably, as the plasmid is the most commonly used form of vector. However, also included are such other forms of expression vectors that serve equivalent functions and that become known in the art subsequently hereto.

[0134] As used herein, the term "wax" means a composition comprised of fatty esters. In a preferred embodiment, the fatty ester in the wax is comprised of medium to long carbon chains. In addition to fatty esters, a wax may comprise other components (e.g., hydrocarbons, sterol esters, aliphatic aldehydes, alcohols, ketones, beta-diketones, triacylglycerols, etc.).

[0135] Throughout the specification, a reference may be made using an abbreviated gene name or polypeptide name, but it is understood that such an abbreviated gene or polypeptide name represents the genus of genes or polypeptides. Such gene names include all genes encoding the same polypeptide and homologous polypeptides having the same physiological function. Polypeptide names include all polypeptides that have the same activity (e.g., that catalyze the same fundamental chemical reaction).

[0136] Unless otherwise indicated, the accession numbers referenced herein are derived from the NCBI database (National Center for Biotechnology Information) maintained by the National Institute of Health, U.S.A. Unless otherwise indicated, the accession numbers are as provided in the database as of October 2009.

[0137] EC numbers are established by the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology (NC-IUBMB) (available at http://www.chem.qmul.ac.uk/iubmb/enzyme/). The EC numbers referenced herein are derived from the KEGG Ligand database, maintained by the Kyoto Encyclopedia of Genes and Genomics, sponsored in part by the University of Tokyo. Unless otherwise indicated, the EC numbers are as provided in the database as of October 2009.

[0138] Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs.

[0139] Although methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, suitable methods and materials are described below. All methods described herein can be performed in any suitable order unless otherwise indicated herein or otherwise clearly contradicted by context.

[0140] Unless otherwise stated, amounts listed in percentage (%) are in weight percent, based on the total weight of the composition.

[0141] All publications, patent applications, patents, and other references mentioned herein are incorporated by reference in their entirety. In case of conflict, the present specification, including definitions, will control. In addition, the materials, methods, and examples are illustrative only and not intended to be limiting.

[0142] Recitation of ranges of values herein are merely intended to serve as a shorthand method of referring individually to each separate value falling within the range, unless otherwise indicated herein, and each separate value is incorporated into the specification as if it were individually recited herein.

[0143] The use of any and all examples, or exemplary language (e.g., "such as") provided herein, is intended merely to better illuminate the invention and does not pose a limitation on the scope of the invention unless otherwise claimed. No language in the specification should be construed as indicating any non-claimed element as essential to the practice of the invention.

[0144] Other features and advantages of the invention will be apparent from the following detailed description and from the claims.

Fatty Esters

[0145] This disclosure relates to the production of fatty esters, such as fatty acid esters including, for example fatty acid methyl esters ("FAME") and fatty acid ethyl esters ("FAEE"), in host cells. In particular embodiments, the methods described herein are used to produce fatty acid methyl esters, which can be used in biodiesels.

[0146] Fatty esters produced by the methods described herein are not limited to esters of any particular length or other characteristics. For example, a microorganism can be genetically engineered to produce any of the fatty esters described in Knothe, *Fuel Processing Technology* 86:1059-1070 (2005), using the teachings provided herein. Such fatty esters can be characterized, for example, by cetane number (CN), viscosity, melting point, and heat of combustion, as described by Knothe.

[0147] Fatty esters that are produced in accordance with the methods, cells, or microorganisms herein comprise, consist essentially of, or consist of the following formula: BCOOA, having an A side and a B side, where the "A side" refers to the carbon chain attached to the carboxylate oxygen of the ester, and the "B side" refers to the carbon chain comprising the parent carboxylate of the ester. The A side is contributed by an alcohol, such as a fatty alcohol, and the B side is contributed by an acid, such as a fatty acid. B is an aliphatic group. In some embodiments, B is a carbon chain. In some embodiments, B comprises a carbon chain that is at least 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, or 30 carbons in length. A comprises at least one carbon atom. In some embodiments, A is an aliphatic group. In some embodiments, A is an alkyl group. In some embodiments, the alkyl group comprises, consists essentially

of, or consists of 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 carbon atoms. In some embodiments, any of the above B groups can be combined with any of the above A groups. In some embodiments, A comprises, consists essentially of, or consists of a carbon chain having a number of carbons selected from the group consisting of 1, 2, 3, 4, and 5 carbon atoms, while B comprises, consists essentially of, or consists of at least 12, 13, 14, 15, 16, 17, 18, 19, or 20 carbon atoms.

[0148] In some embodiments, the fatty esters of the invention comprise a plurality of individual fatty esters. In some embodiments, the methods described herein permit production of a plurality of fatty esters of varied length. In some embodiments, the fatty ester product comprises saturated or unsaturated fatty esters product(s) having a carbon atom content limited to between 5 and 25 carbon atoms. In other words, the invention provides a composition comprising C_5-C_{25} fatty esters (e.g., C_{10}-C_{20} fatty esters, or C_{12}-C_{18} fatty esters).

[0149] In some embodiments, the fatty esters comprise one or more fatty esters having a double bond at one or more points in the carbon chain. Thus, in some embodiments, a 6-, 7-, 8-, 9-, 10-, 11-, 12-, 13-, 14-, 15-, 16-, 17-, 18-, 19-, 20-, 21-, 22-, 23-, 24-, 25-, 26-, 27-, 28-, 29-, or 30-carbon chain can have 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, or 24 double bonds, and 1-24 of the aforesaid double bonds can be located following carbon 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, or 29. In some embodiments, a 1-, 2-, 3-, 4-, or 5-carbon chain for A can have 1, 2, 3, or 4 double bonds and 1-4 of the double bonds can be located following carbon 1, 2, 3, or 4. In some embodiments, any of the above A groups can be combined with any of the above B groups.

[0150] In certain preferred embodiments, the B group can have 12, 13, 14, 15, 16, 17, 18 carbon atoms in a chain. In other embodiments, the A group can have one or two carbon atoms.

[0151] In some preferred embodiments, the B group can have one double bond at one or more points in the carbon chain. In more preferred embodiments, the B group can have one double bond at position 7 of the carbon chain, numbering from the reduced end of the carbon chain. One of ordinary skill in the art will recognize that one end of the B group will have a methyl group, and the other end of the B group will have a carboxyl group (C(=O)O—). The end of the B group which is a methyl group is the reduced end of the carbon chain comprising the B group, thus, the double bond is at carbon 7 counting from the methyl group terminus of the B group (e.g., at between carbons 7 and 8 of the B group). The double bond can have any geometry, thus, the double bond in the B group can be cis or trans.

[0152] In some embodiments, the fatty esters comprise straight chain fatty esters. In some embodiments, the fatty esters comprise branched chain fatty esters. In some embodiments, the fatty esters comprise cyclic moieties.

[0153] In certain preferred embodiments, the fatty esters can be selected from the group consisting of methyl dodecanoate, methyl 5-dodecenoate, methyl tetradecanoate, methyl 7-tetradecenoate, methyl hexadecanoate, methyl 9-hexadecenoate, methyl octadecanoate, methyl 11-octadecenoate, and combinations thereof.

[0154] In some embodiments, the fatty ester composition comprises about 5 wt. % or more methyl dedecanoate. In some embodiments, the fatty ester composition comprises about 25% or more methyl dedecanoate. In some embodiments, the fatty ester composition comprises about 5 wt. % to about 25 wt. % methyl dodecanoate.

[0155] In some embodiments, the fatty ester composition comprises about 10 wt. % or less methyl dodec-7-enoate. In some embodiments, the fatty ester composition comprises about 0 wt. % to about 10 wt. % methyl dodec-7-enoate.

[0156] In some embodiments, the fatty ester composition comprises about 30 wt. % or more methyl tetradecanoate. In some embodiments, the fatty ester composition comprises about 50 wt. % or less methyl tetradecanoate. In some embodiments, the fatty ester composition comprises about 30 wt. % to about 50 wt. % methyl tetradecanoate.

[0157] In some embodiments, the fatty ester composition comprises about 10 wt. % or less methyl tetradec-7-enoate. In some embodiments, the fatty ester composition comprises about 0 wt. % to about 10 wt. % methyl tetradec-7-enoate.

[0158] In some embodiments, the fatty ester composition comprises about 15 wt. % or less methyl hexadecanoate. In some embodiments, the fatty ester composition comprises about 0 wt. % to about 15 wt. % methyl hexadecanoate.

[0159] In some embodiments, the fatty ester composition comprises about 10 wt. % or more methyl hexadec-7-enoate. In some embodiments, the fatty ester composition comprises about 40 wt. % or less methyl hexadec-7-enoate. In some embodiments, the fatty ester composition comprises about 10 wt. % to about 40 wt. % methyl hexadec-7-enoate.

[0160] In some embodiments, the fatty ester composition comprises about 15 wt. % or less methyl octadec-7-enoate. In some embodiments, the fatty ester composition comprises about 0 wt. % to about 15 wt. % methyl octadec-7-enoate.

[0161] These exemplary fatty ester products, having characteristic features of A side and/or B side, can be prepared and produced using substrates having similar or the same features. Accordingly, each step within a biosynthetic pathway that leads to the production of a fatty acid derivative can be modified to produce or overproduce a substrate leading to a fatty alcohol and/or a fatty acid. For example, known genes involved in the fatty acid biosynthetic pathway or the fatty alcohol pathway can be expressed, overexpressed, or attenuated in host cells to produce a desired substrate (see, e.g., WO2008/119082, the disclosure of which is incorporated by reference herein).

Synthesis of Substrates

[0162] Fatty acid synthase (FAS) is a group of polypeptides that catalyze the initiation and elongation of acyl chains (Marrakchi et al., *Biochemical Society,* 30:1050-1055 (2002)). The acyl carrier protein (ACP) along with the enzymes in the FAS pathway control the length, degree of saturation, and branching of the fatty acid derivatives produced. The fatty acid biosynthetic pathway involves the precursors acetyl-CoA and malonyl-CoA. The steps in this pathway are catalyzed by enzymes of the fatty acid biosynthesis (fab) and acetyl-CoA carboxylase (acc) gene families (see, e.g., Heath et al., *Prog. Lipid Res.* 40(6):467-97 (2001)).

[0163] Host cells can be engineered to express fatty acid derivative substrates by recombinantly expressing or overexpressing one or more fatty acid synthase genes, such as acetyl-CoA and/or malonyl-CoA synthase genes. For example, to increase acetyl-CoA production, one or more of the following genes can be expressed in a host cell: pdh (a multienzyme complex comprising aceEF (which encodes the E1p dehydrogenase component, the E2p dihydrolipoamide acyltransferase component of the pyruvate and 2-oxoglut-

arate dehydrogenase complexes, and lpd), panK, fabH, fabB, fabD, fabG, acpP, and fabF. Exemplary GenBank accession numbers for these genes are: pdh (BAB34380, AAC73227, AAC73226), panK (also known as CoA, AAC76952), aceEF (AAC73227, AAC73226), fabH (AAC74175), fabB (P0A953), fabD (AAC74176), fabG (AAC74177), acpP (AAC74178), fabF (AAC74179). Additionally, the expression levels of fadE, gpsA, ldhA, pflb, adhE, pta, poxB, ackA, and/or ackB can be attenuated or knocked-out in an engineered host cell by transformation with conditionally replicative or non-replicative plasmids containing null or deletion mutations of the corresponding genes or by substituting promoter or enhancer sequences. Exemplary GenBank accession numbers for these genes are: fadE (AAC73325), gspA (AAC76632), ldhA (AAC74462), pflb (AAC73989), adhE (AAC74323), pta (AAC75357), poxB (AAC73958), ackA (AAC75356), and ackB (BAB81430). The resulting host cells will have increased acetyl-CoA production levels when grown in an appropriate environment.

[0164] Malonyl-CoA overexpression can be affected, for example, by introducing one or more or all subunits of a four-subunit protein accABCD (e.g., accession number AAC73296, EC 6.4.1.2) into a host cell. Fatty acids can be further produced in host cells by introducing into the host cell a DNA sequence encoding a lipase (e.g., accession numbers CAA89087, CAA98876).

[0165] In addition, inhibiting PlsB can lead to an increase in the levels of long chain acyl-ACP, which will inhibit early steps in the pathway (e.g., the expression of accABCD, fabH, or fabI). The plsB (e.g., accession number AAC77011) D311E mutation can be used to increase the amount of available fatty acids.

[0166] In addition, a host cell can be engineered to overexpress a sfa gene (suppressor of fabA, e.g., accession number AAN79592) to increase production of monounsaturated fatty acids (Rock et al., *J. Bacteriology* 178:5382-5387 (1996)).

[0167] The chain length of a fatty acid derivative substrate can be selected for by modifying the expression of a thioesterase, which influences the chain length of fatty acids produced. Hence, host cells can be engineered to express, overexpress, have attenuated expression, or not to express one or more selected thioesterases to increase the production of a preferred fatty acid derivative substrate. For example, C_{10} fatty acids can be produced by expressing a thioesterase that has a preference for producing C_{10} fatty acids and attenuating thioesterases that have a preference for producing fatty acids other than C_{10} fatty acids (e.g., a thioesterase which prefers to produce C_{14} fatty acids). This would result in a relatively homogeneous population of fatty acids that have a carbon chain length of, for example, 10. In other instances, C_{14} fatty acids can be produced by attenuating endogenous thioesterases that produce non-C_{14} fatty acids and expressing the thioesterases that use C_{14}-ACP. In some situations, C_{12} fatty acids can be produced by expressing thioesterases that use C_{12}-ACP, while in parallel, attenuating thioesterases that produce non-C_{12} fatty acids. Acetyl-CoA, malonyl-CoA, and fatty acid overproduction can be verified using methods known in the art, for example, by using radioactive precursors, HPLC, or GC-MS subsequent to cell lysis. Non-limiting examples of thioesterases that can be used in the methods described herein are listed in Table 1.

TABLE 1

Thioesterases

Accession Number	Source Organism	Gene
AAC73596	*E. coli*	tesA without leader sequence
AAC73555	*E. coli*	tesB
Q41635, AAA34215	*Umbellularia californica*	fatB
AAC49269	*Cuphea hookeriana*	fatB2
Q39513; AAC72881	*Cuphea hookeriana*	fatB3
Q39473, AAC49151	*Cinnamonum camphorum*	fatB
CAA85388	*Arabidopsis thaliana*	fatB [M141T]*
NP 189147; NP 193041	*Arabidopsis thaliana*	fatA
CAC39106	*Bradyrhiizobium japonicum*	fatA
AAC72883	*Cuphea hookeriana*	fatA
AAL79361	*Helianthus annus*	fatA1

*Mayer et al., *BMC Plant Biology* 7: 1-11 (2007)

[0168] In other instances, fatty esters are produced in a host cell that contains a naturally occurring mutation that results in an increased level of fatty acids in the host cell. In some instances, the host cell is genetically engineered to increase the level of fatty acids in the host cell relative to a corresponding wild-type host cell. For example, the host cell can be genetically engineered to express a reduced or attenuated level of an acyl-CoA synthase relative to a corresponding wild-type host cell. In a particular embodiment, the level of expression of one or more genes (e.g., an acyl-CoA synthase gene) can be eliminated by genetically engineering a "knock out" host cell, wherein the one or more genes are deleted.

[0169] Any known acyl-CoA synthase gene can be reduced or knocked out in a host cell. Non-limiting examples of acyl-CoA synthase genes include fadD, fadK, BH3103, yhfL, Pfl-4354, EAV15023, fadD1, fadD2, RPC_4074, fadDD35, fadDD22, faa3p or the gene encoding the protein ZP_01644857. Specific examples of acyl-CoA synthase genes include fadDD35 from *M. tuberculosis* H37Rv [NP_217021], fadDD22 from *M. tuberculosis* H37Rv [NP_217464], fadD from *E. coli* [NP_416319], fadK from *E. coli* [YP_416216], fadD from *Acinetobacter* sp. ADP1 [YP_045024], fadD from *Haemophilus influenza* RdkW20 [NP_438551], fadD from *Rhodopseudomonas palustris* Bis B18 [YP_533919], BH3101 from *Bacillus halodurans* C-125 [NP_243969], Pfl-4354 from *Pseudomonas fluorescens* Pfo-1 [YP_350082], EAV15023 from *Comamonas testosterone* KF-1 [ZP_01520072], yhfL from *B. subtilis* [NP_388908], fadD1 from *P. aeruginosa* PAO1 [NP_251989], fadD1 from *Ralstonia solanacearum* GM1 1000 [NP_520978], fadD2 from *P. aeruginosa* PAO1 [NP_251990], the gene encoding the protein ZP_01644857 from *Stenotrophomonas maltophilia* R551-3, faa3p from *Saccharomyces cerevisiae* [NP_012257], faa1p from *Saccharomyces cerevisiae* [NP_014962], lcfA from *Bacillus subtilis* [CAA99571], or those described in Shockey et al., *Plant. Physiol.* 129:1710-1722 (2002); Caviglia et al., *J. Biol. Chem.* 279:1163-1169 (2004); Knoll et al., *J. Biol. Chem.* 269(23):16348-56 (1994); Johnson et al., *J. Biol. Chem.* 269: 18037-18046 (1994); and Black et al., *J. Biol. Chem.* 267: 25513-25520 (1992).

Formation of Branched Fatty Esters

[0170] Fatty esters can be produced that contain branch points by using branched fatty acid derivatives as substrates or precursors. For example, although *E. coli* naturally produces straight chain fatty acids (sFAs), *E. coli* can be engineered to produce branched chain fatty acids (brFAs) by introducing and expressing or overexpressing genes that pro-

vide branched precursors in the *E. coli* (e.g., bkd, ilv, icm, and fab gene families). Additionally, a host cell can be engineered to express or overexpress one or more genes encoding one or more proteins for the elongation of brFAs (e.g., ACP, FabF, etc.) and/or to delete or attenuate the corresponding host cell genes that normally lead to sFAs.

[0171] The first step in forming brFAs is the production of the corresponding α-keto acids by a branched-chain amino acid aminotransferase. Host cells may endogenously include genes encoding such enzymes or such genes can be recombinantly introduced. *E. coli*, for example, endogenously expresses such an enzyme, IlvE (EC 2.6.1.42; GenBank accession YP_026247). In some host cells, a heterologous branched-chain amino acid aminotransferase may not be expressed. However, *E. coli* IlvE or any other branched-chain amino acid aminotransferase (e.g., IlvE from *Lactococcus lactis* (GenBank accession AAF34406), IlvE from *Pseudomonas putida* (GenBank accession NP_745648), or IlvE from *Streptomyces coelicolor* (GenBank accession NP_629657)), if not endogenous, can be introduced.

[0172] In another embodiment, the production of α-keto acids can be achieved by using the methods described in Atsumi et al., *Nature* 451:86-89 (2008). For example, 2-ketoisovalerate can be produced by overexpressing the genes encoding IlvI, IlvH, IlvC, or IlvD. In another example, 2-keto-3-methyl-valerate can be produced by overexpressing the genes encoding IlvA and IlvI, IlvH (or AlsS of *Bacillus subtilis*), IlvC, IlvD, or their corresponding homologs. In a further embodiment, 2-keto-4-methyl-pentanoate can be produced by overexpressing the genes encoding IlvI, IlvH, IlvC, IlvD and LeuA, LeuB, LeuC, LeuD, or their corresponding homologs.

[0173] The second step is the oxidative decarboxylation of the α-keto acids to the corresponding branched-chain acyl-CoA. This reaction can be catalyzed by a branched-chain α-keto acid dehydrogenase complex (bkd; EC 1.2.4.4.) (Denoya et al., *J. Bacteria* 177:3504 (1995)), which consists of E1a/13 (decarboxylase), E2 (dihydrolipoyl transacylase), and E3 (dihydrolipoyl dehydrogenase) subunits. These branched-chain α-keto acid dehydrogenase complexes are similar to pyruvate and α-ketoglutarate dehydrogenase complexes. Any microorganism that possesses brFAs and/or grows on branched-chain amino acids can be used as a source to isolate bkd genes for expression in host cells, for example, *E. coli*. Furthermore, *E. coli* has the E3 component as part of its pyruvate dehydrogenase complex (lpd, EC 1.8.1.4, GenBank accession NP_414658). Thus, it may be sufficient to express only the E1α/β and E2 bkd genes. Table 2 lists non-limiting examples of bkd genes from several microorganisms that can be recombinantly introduced and expressed in a host cell to provide branched-chain acyl-CoA precursors.

TABLE 2

Bkd genes from selected microorganisms

Organism	Gene	GenBank Accession #
Streptomyces coelicolor	bkdA1 (E1α)	NP_628006
	bkdB1 (E1β)	NP_628005
	bkdC1 (E2)	NP_628004
Streptomyces coelicolor	bkdA2 (E1α)	NP_733618
	bkdB2 (E1β)	NP_628019
	bkdC2 (E2)	NP_628018

TABLE 2-continued

Bkd genes from selected microorganisms

Organism	Gene	GenBank Accession #
Streptomyces avermitilis	bkdA (E1a)	BAC72074
	bkdB (E1b)	BAC72075
	bkdC (E2)	BAC72076
Streptomyces avermitilis	bkdF (E1α)	BAC72088
	bkdG (E1β)	BAC72089
	bkdH (E2)	BAC72090
Bacillus subtilis	bkdAA (E1α)	NP_390288
	bkdAB (E1β)	NP_390288
	bkdB (E2)	NP_390288
Pseudomonas putida	bkdA1 (E1α)	AAA65614
	bkdA2 (E1β)	AAA65615
	bkdC (E2)	AAA65617

[0174] In another example, isobutyryl-CoA can be made in a host cell, for example in *E. coli*, through the coexpression of a crotonyl-CoA reductase (Ccr, EC 1.6.5.5, 1.1.1.1) and isobutyryl-CoA mutase (large subunit IcmA, EC 5.4.99.2; small subunit IcmB, EC 5.4.99.2) (Han and Reynolds, *J. Bacteriol.* 179:5157 (1997)). Crotonyl-CoA is an intermediate in fatty acid biosynthesis in *E. coli* and other microorganisms. Non-limiting examples of ccr and icm genes from selected microorganisms are listed in Table 3.

TABLE 3

Ccr and icm genes from selected microorganisms

Organism	Gene	GenBank Accession #
Streptomyces coelicolor	ccr	NP_630556
	icmA	NP_629554
	icmB	NP_630904
Streptomyces cinnamonensis	ccr	AAD53915
	icmA	AAC08713
	icmB	AJ246005

[0175] In addition to expression of the bkd genes, the initiation of brFA biosynthesis utilizes β-ketoacyl-acyl-carrier-protein synthase III (FabH, EC 2.3.1.41) with specificity for branched chain acyl-CoAs (Li et al., *J. Bacteriol.* 187:3795-3799 (2005)). Non-limiting examples of such FabH enzymes are listed in Table 4. fabH genes that are involved in fatty acid biosynthesis of any brFA-containing microorganism can be expressed in a host cell. The Bkd and FabH enzymes from host cells that do not naturally make brFA may not support brFA production. Therefore, bkd and fabH can be expressed recombinantly. Vectors containing the bkd and fabH genes can be inserted into such a host cell. Similarly, the endogenous level of Bkd and FabH production may not be sufficient to produce brFA. In this case, they can be overexpressed. Additionally, other components of the fatty acid biosynthesis pathway can be expressed or overexpressed, such as acyl carrier proteins (ACPs) and β-ketoacyl-acyl-carrier-protein synthase II (fabF, EC 2.3.1.41) (non-limiting examples of FabH, ACP, and fabF genes from select microorganisms are listed in Table 4). In addition to expressing these genes, some genes in the endogenous fatty acid biosynthesis pathway can be attenuated in the host cell (e.g., the *E. coli* genes fabH (GenBank accession #NP_415609) and/or fabF (GenBank accession #NP_415613)).

TABLE 4

FabH, ACP and fabF genes from selected microorganisms with brFAs

Organism	Gene	GenBank Accession #
Streptomyces coelicolor	fabH1	NP_626634
	acp	NP_626635
	fabF	NP_626636
Streptomyces avermitilis	fabH3	NP_823466
	fabC3 (acp)	NP_823467
	fabF	NP_823468
Bacillus subtilis	fabH_A	NP_389015
	fabH_B	NP_388898
	acp	NP_389474
	fabF	NP_389016
Stenotrophomonas maltophilia	SmalDRAFT_0818 (fabH)	ZP_01643059
	SmalDRAFT_0821 (acp)	ZP_01643063
	SmalDRAFT_0822 (fabF)	ZP_01643064
Legionella pneumophila	fabH	YP_123672
	acp	YP_123675
	fabF	YP_123676

Formation of Cyclic Fatty Esters

[0176] Cyclic fatty esters can be produced by using cyclic fatty acid derivatives as substrates. To produce cyclic fatty acid derivative substrates, genes that provide cyclic precursors (e.g., the ans, chc, and plm gene families) can be introduced into the host cell and expressed to allow initiation of fatty acid biosynthesis from cyclic precursors. For example, to convert a host cell, such as *E. coli*, into one capable of synthesizing w-cyclic fatty acids (cyFA), a gene that provides the cyclic precursor cyclohexylcarbonyl-CoA (CHC-CoA) (Cropp et al., *Nature Biotech.* 18:980-983 (2000)) can be introduced and expressed in the host cell. Non-limiting examples of genes that provide CHC-CoA in *E. coli* include: ansJ, ansK, ansL, chcA, and ansM from the ansatrienin gene cluster of *Streptomyces collinus* (Chen et al., *Eur. J. Biochem.* 261: 98-107 (1999)) or plmJ, plmK, plmL, chcA, and plmM from the phoslactomycin B gene cluster of *Streptomyces* sp. HK803 (Palaniappan et al., *J. Biol. Chem.* 278:35552-35557 (2003)) together with the chcB gene (Patton et al., *Biochem.* 39:7595-7604 (2000)) from *S. collinus, S. avermitilis*, or *S. coelicolor* (see Table 5). Furthermore, the genes listed in Table 4, which tend to have broad substrate specificity, can then be expressed to allow initiation and elongation of co-cyclic fatty acids. Alternatively, the homologous genes can be isolated from microorganisms that make cyFA and expressed in a host cell (e.g., *E. coli*).

TABLE 5

Genes for the synthesis of CHC-CoA

Organism	Gene	GenBank Accession #
Streptomyces collinus	ansJK	U72144*
	ansL	
	chcA	
	ansM	
	chcB	AF268489
Streptomyces sp. HK803	pmlJK	AAQ84158
	pmlL	AAQ84159
	chcA	AAQ84160
	pmlM	AAQ84161

TABLE 5-continued

Genes for the synthesis of CHC-CoA

Organism	Gene	GenBank Accession #
Streptomyces coelicolor	chcB/caiD	NP_629292
Streptomyces avermitilis	chcB/caiD	NP_629292

*Only chcA is annotated in GenBank entry U72144, ansJKLM are according to Chen et al. (*Eur. J. Biochem.* 261: 98-107 (1999)).

[0177] The genes listed in Table 4 (fabH, acp, and fabF) allow initiation and elongation of co-cyclic fatty acids because they have broad substrate specificity. If the coexpression of any of these genes with the genes listed in Table 5 does not yield cyFA, then fabH, acp, and/or fabF homologs from microorganisms that make cyFAs (e.g., those listed in Table 6) can be isolated (e.g., by using degenerate PCR primers or heterologous DNA sequence probes) and coexpressed.

TABLE 6

Non-limiting examples of microorganisms that contain ω-cyclic fatty acids

Organism	Reference
Curtobacterium pusillum	ATCC19096
Alicyclobacillus acidoterrestris	ATCC49025
Alicyclobacillus acidocaldarius	ATCC27009
*Alicyclobacillus cycloheptanicus**	Moore, *J. Org. Chem.* 62: pp. 2173, 1997

*Uses cycloheptylcarbonyl-CoA and not cyclohexylcarbonyl-CoA as precursor for cyFA biosynthesis.

Fatty Ester Saturation Levels

[0178] The degree of saturation in fatty acids can be controlled by regulating the degree of saturation of fatty acid intermediates. For example, the sfa, gns, and fab families of genes can be expressed, overexpressed, or expressed at reduced levels, to control the saturation of fatty acids. A number of those genes have been described in, for example, WO 2008/119082, the disclosure of which is incorporated by reference. Non-limiting examples of genes in these gene families include GenBank Accession No: AAN79592, AAC44390, ABD18647.1, AAC74076.1, BAA16180, AAF98273, AAU39821, or DDA05501.

[0179] For example, host cells can be engineered to produce unsaturated fatty acids by engineering the production host to overexpress fabB or by growing the production host at low temperatures (e.g., less than 37° C.). FabB has preference to cis-δ3decenoyl-ACP and results in unsaturated fatty acid production in *E. coli*. Overexpression of fabB results in the production of a significant percentage of unsaturated fatty acids (de Mendoza et al., *J. Biol. Chem.* 258:2098-2101 (1983)). In some embodiments, an endogenous fabB gene may be modified such that it is overexpressed. In some other embodiments, a heterologous fabB gene may be inserted into and expressed in host cells not naturally having the gene. These unsaturated fatty acids can then be used as intermediates in host cells that are engineered to produce fatty acid derivatives, such as fatty esters.

[0180] In other instances, a repressor of fatty acid biosynthesis, for example, fabR (GenBank accession NP_418398), can be deleted, which will also result in increased unsaturated fatty acid production in *E. coli* (Zhang et al., *J. Biol. Chem.*

277:15558, (2002)). Similar deletions may be made in other host cells. A further increase in unsaturated fatty acids may be achieved, for example, by overexpressing fabM (trans-2, cis-3-decenoyl-ACP isomerase, GenBank accession DAA05501) and controlled expression of fabK (trans-2-enoyl-ACP reductase II, GenBank accession NP_357969) from *Streptococcus pneumoniae* (Marrakchi et al., *J. Biol. Chem.* 277: 44809 (2002)), while deleting *E. coli* fabI (trans-2-enoyl-ACP reductase, GenBank accession NP_415804). In some examples, the endogenous fabF gene can be attenuated, thus increasing the percentage of palmitoleate (C16:1) produced.

[0181] In yet other examples, host cells can be engineered to produce saturated fatty acids by reducing the expression of an sfa, gns, or fab gene.

[0182] For example, a host cell can be engineered to express a decreased level of fabA and/or fabB. In some instances, the host cell can be grown in the presence of unsaturated fatty acids. In other instances, the host cell can be further engineered to express or overexpress a gene encoding a desaturase enzyme. One non-limiting example of a desaturase is *B. subtilis* DesA (AF037430). Other genes encoding desaturase enzymes are known in the art and can be used in the host cells and methods described herein, such as desaturases that use acyl-ACP, such as hexadecanoyl-ACP or octadecanoyl-ACP.

Ester Synthases

[0183] Fatty esters can be synthesized by acyl-CoA:fatty alcohol acyltransferase, which conjugates an alcohol to a fatty acyl-CoA via an ester linkage. Ester synthases and encoding genes are known from the jojoba plant and the bacterium *Acinetobacter* sp. strain ADP1 (formerly *Acinetobacter calcoaceticus* ADP1). The bacterial ester synthase is a bifunctional enzyme, exhibiting ester synthase activity and the ability to form triacylglycerols from diacylglycerol substrates and fatty acyl-CoAs (acyl-CoA:diglycerol acyltransferase (DGAT) activity). The gene wax/dgat encodes both ester synthase and DGAT. See Cheng et al., *J. Biol. Chem.* 279(36):37798-37807 (2004); Kalscheuer and Steinbuchel, *J. Biol. Chem.* 278:8075-8082 (2003).

[0184] Other ester synthases (EC 2.3.1.20, 2.3.1.75) are known in the art, and exemplary GenBank Accession Numbers include, without limitation, NP_190765, AAA16514, AAF19262, AAX48018, AAO17391, or AAD38041. Methods to identify ester synthase activity are provided in, e.g., U.S. Pat. No. 7,118,896, which is herein incorporated by reference in its entirety. Other ester synthases include bifunctional ester synthase/acyl-CoA:diacylglycerol acyltransferases, nonlimiting examples of which include the multienzyme complexes from *Simmondsia chinensis, Acinetobacter* sp. strain ADP1 (formerly *Acinetobacter calcoaceticus* ADP1), *Alcanivorax borkumensis, Pseudomonas aeruginosa, Fundibacter jadensis, Arabidopsis thaliana*, or *Alcaligenes eutrophus* (later renamed *Ralstonia eutropha*). In one embodiment, the fatty acid elongases, acyl-CoA reductases or wax synthases can be from a multienzyme complex from *Alcaligenes eutrophus* (later renamed *Ralstonia eutropha*) or other organisms known in the literature to produce esters, such as wax or fatty esters. Additional sources of heterologous DNA sequence encoding ester synthesis proteins that can be used in fatty ester production include, but are not limited to, *Mortierella alpina* (e.g., ATCC 32222), *Cryptococcus curvatus* (also referred to as *Apiotricum curvatum*), *Alcanivorax jadensis* (for example T9T, DSM 12718, or ATCC 700854), *Acinetobacter* sp. HO1-N, (e.g., ATCC 14987), *Rhodococcus opacus* (e.g., PD630, DSMZ 44193), and the ester synthases from *Marinobacter hydrocarbonoclastics* (e.g., DSM 8798). In certain embodiments, the ester synthase is selected from the group consisting of: AtfA1 (an ester synthase derived from *Alcanivorax borkumensis* SK2, GenBank Accession No. YP_694462), AtfA2 (another ester synthase derived from *Alcanivorax borkumensis* SK2, GenBank Accession No. YP_693524), ES9 (an ester synthase from *Marinobacter hydrocarbonoclasticus* DSM 8798, GenBank Accession No. ABO21021), ES8 (another ester synthase derived from *Marinobacter hydrocarbonoclasticus* DSM 8798, GenBank Accession No. ABO21020), and variants thereof. In a particular embodiment, the gene encoding the ester synthase or a suitable variant is overexpressed.

[0185] Additional ester synthases that can be used are described in the PCT patent application entitled "Production of Fatty Acid Derivatives", filed concurrently herewith, under Attorney Docket No. 2001235.150WO2.

Genetic Engineering of Host Cells to Produce Fatty Esters

[0186] Various host cells can be used to produce fatty esters, as described herein. A host cell can be any prokaryotic or eukaryotic cell. For example, a polypeptide described herein can be expressed in bacterial cells (such as *E. coli*), insect cells, yeast, or mammalian cells (such as Chinese hamster ovary cells (CHO) cells, COS cells, VERO cells, BHK cells, HeLa cells, Cv1 cells, MDCK cells, 293 cells, 3T3 cells, or PC12 cells). Other exemplary host cells include cells from the members of the genus *Escherichia, Bacillus, Lactobacillus, Rhodococcus, Pseudomonas, Aspergillus, Trichoderma, Neurospora, Fusarium, Humicola, Rhizomucor, Kluyveromyces, Pichia, Mucor, Myceliophtora, Penicillium, Phanerochaete, Pleurotus, Trametes, Chrysosporium, Saccharomyces, Schizosaccharomyces, Yarrowia*, or *Streptomyces*. Yet other exemplary host cells can be a *Bacillus lentus* cell, a *Bacillus brevis* cell, a *Bacillus stearothermophilus* cell, a *Bacillus licheniformis* cell, a *Bacillus alkalophilus* cell, a *Bacillus coagulans* cell, a *Bacillus circulans* cell, a *Bacillus pumilis* cell, a *Bacillus thuringiensis* cell, a *Bacillus clausii* cell, a *Bacillus megaterium* cell, a *Bacillus subtilis* cell, a *Bacillus amyloliquefaciens* cell, a *Trichoderma koningii* cell, a *Trichoderma viride* cell, a *Trichoderma reesei* cell, a *Trichoderma longibrachiatum* cell, an *Aspergillus awamori* cell, an *Aspergillus fumigates* cell, an *Aspergillus foetidus* cell, an *Aspergillus nidulans* cell, an *Aspergillus niger* cell, an *Aspergillus oryzae* cell, a *Humicola insolens* cell, a *Humicola lanuginose* cell, a *Rhizomucor miehei* cell, a *Mucor michei* cell, a *Streptomyces lividans* cell, a *Streptomyces murinus* cell, or an Actinomycetes cell. Other host cells are cyanobacterial host cells.

[0187] In certain embodiments, the host cell is an Actinomycetes, *Saccharomyces cerevisiae, Candida lipolytica* (or *Yarrowia lipolytica*), *E. coli, Arthrobacter* AK 19, *Rhodotorula glutinims, Acintobacter* sp. M-1 cell, or a cell from other oleaginous microorganisms.

[0188] In particular embodiments, the host cell is a cell from an eukaryotic plant, algae, *cyanobacterium*, green-sulfur bacterium, green non-sulfur bacterium, purple sulfur bacterium, purple non-sulfur bacterium, extremophile, yeast, fungus, engineered organisms thereof, or a synthetic organism. In some embodiments, the host cell is light dependent or fixes carbon. In some embodiments, the host cell has

autotrophic activity. In some embodiments, the host cell has photoautotrophic activity, such as in the presence of light. In certain embodiments, the host cell is a cell from *Arabidopsis thaliana, Panicum virgatums, Miscanthus giganteus, Zea mays, botryococcuse braunii, Chlamydomonas reinhardtii, Dunaliela salina, Thermosynechococcus elongatus, Chlorobium tepidum, Chloroflexus auranticus, Chromatiumm vinosum, Rhodospirillum rubrum, Rhodobacter capsulatus, Rhodopseudomonas palusris, Clostridium ljungdahlii, Clostridiuthermocellum*, or *Pencillium chrysogenum*. In certain other embodiments, the host cell is from *Pichia pastories, Saccharomyces cerevisiae, Yarrowia lipolytica, Schizosaccharomyces pombe, Pseudomonas fluorescens*, or *Zymomonas mobilis*. In yet further embodiments, the host cell is a cell from *Synechococcus* sp. PCC 7002, *Synechococcus elongatus*. PCC 7942, or *Synechocystis* sp. PCC6803.

[0189] In a preferred embodiment, the host cell is an *E. coli* cell, a *Saccharomyces cerevisiae* cell, or a *Bacillus subtilis* cell. In a more preferred embodiment, the host cell is from *E. coli* strains B, C, K, or W. Other suitable host cells are known to those skilled in the art.

[0190] Various methods well known in the art can be used to genetically engineer host cells to produce fatty esters. The methods can include the use of vectors, preferably expression vectors, containing a nucleic acid encoding a fatty acid biosynthetic polypeptide described herein, polypeptide variant, or a fragment thereof. Those skilled in the art will appreciate a variety of viral vectors (for example, retroviral vectors, lentiviral vectors, adenoviral vectors, and adeno-associated viral vectors) and non-viral vectors can be used in the methods described herein.

[0191] The recombinant expression vectors described herein include a nucleic acid described herein in a form suitable for expression of the nucleic acid in a host cell. The recombinant expression vectors can include one or more control sequences, selected on the basis of the host cell to be used for expression. The control sequence is operably linked to the nucleic acid sequence to be expressed. Such control sequences are described, for example, in Goeddel, Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego, Calif. (1990). Control sequences include those that direct constitutive expression of a nucleotide sequence in many types of host cells and those that direct expression of the nucleotide sequence only in certain host cells (e.g., tissue-specific regulatory sequences). It will be appreciated by those skilled in the art that the design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression of protein desired, etc. The expression vectors described herein can be introduced into host cells to produce polypeptides, including fusion polypeptides, encoded by the nucleic acids as described herein.

[0192] Recombinant expression vectors can be designed for expression of a fatty acid biosynthetic polypeptide or variant in prokaryotic or eukaryotic cells (e.g., bacterial cells, such as *E. coli*, insect cells (e.g., using baculovirus expression vectors), yeast cells, or mammalian cells). Suitable host cells are discussed further in Goeddel, Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego, Calif. (1990). Alternatively, the recombinant expression vector can be transcribed and translated in vitro, for example, by using T7 promoter regulatory sequences and T7 polymerase.

[0193] Expression of polypeptides in prokaryotes, for example, *E. coli*, is most often carried out with vectors containing constitutive or inducible promoters directing the expression of either fusion or non-fusion polypeptides. Fusion vectors add a number of amino acids to a polypeptide encoded therein, usually to the amino terminus of the recombinant polypeptide. Such fusion vectors typically serve three purposes: (1) to increase expression of the recombinant polypeptide; (2) to increase the solubility of the recombinant polypeptide; and (3) to aid in the purification of the recombinant polypeptide by acting as a ligand in affinity purification. Often, in fusion expression vectors, a proteolytic cleavage site is introduced at the junction of the fusion moiety and the recombinant polypeptide. This enables separation of the recombinant polypeptide from the fusion moiety after purification of the fusion polypeptide. Examples of such enzymes, and their cognate recognition sequences, include Factor Xa, thrombin, and enterokinase. Exemplary fusion expression vectors include pGEX (Pharmacia, Piscataway, N.J.; Smith et al., *Gene* 67:31-40 (1988)), pMAL (New England Biolabs, Inc., Ipswich, Mass.), and pRITS (Pharmacia, Piscataway, N.J.), which fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant polypeptide.

[0194] Examples of inducible, non-fusion *E. coli* expression vectors include pTrc (Amann et al., *Gene* 69:301-315 (1988)) and pET 11d (Studier et al., Gene Expression Technology Methods in Enzymology 185, Academic Press, San Diego, Calif. 60-89 (1990)). Target gene expression from the pTrc vector relies on host RNA polymerase transcription from a hybrid tip-lac fusion promoter. Target gene expression from the pET 11d vector relies on transcription from a T7 gn10-lac fusion promoter mediated by a coexpressed viral RNA polymerase (T7 gn1). This viral polymerase is supplied by host strains BL21(DE3) or HMS174(DE3) from a resident 2, prophage harboring a T7 gn1 gene under the transcriptional control of the lacUV 5 promoter.

[0195] One strategy to maximize recombinant polypeptide expression is to express the polypeptide in a host cell with an impaired capacity to proteolytically cleave the recombinant polypeptide (see, Gottesman, Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego, Calif. 119-128 (1990)). Another strategy is to alter the nucleic acid sequence to be inserted into an expression vector so that the individual codons for each amino acid are those preferentially utilized in the host cell (Wada et al., *Nucleic Acids Res.* 20:2111-2118 (1992)). Such alteration of nucleic acid sequences can be carried out by standard DNA synthesis techniques.

[0196] In another embodiment, the host cell is a yeast cell. In this embodiment, the expression vector is a yeast expression vector. Examples of vectors for expression in yeast *S. cerevisiae* include pYepSec1 (Baldari et al., *EMBO J.* 6:229-234 (1987)), pMFa (Kurjan et al., *Cell* 30:933-943 (1982)), pJRY88 (Schultz et al., *Gene* 54:113-123 (1987)), pYES2 (Invitrogen Corporation, San Diego, Calif.), and picZ (Invitrogen Corp, San Diego, Calif.).

[0197] Alternatively, a polypeptide described herein can be expressed in insect cells using baculovirus expression vectors. Baculovirus vectors available for expression of proteins in cultured insect cells (e.g., Sf9 cells) include, for example, the pAc series (Smith et al., *Mol. Cell Biol.* 3:2156-2165 (1983)) and the pVL series (Lucklow et al., *Virology* 170:31-39 (1989)).

[0198] In yet another embodiment, the nucleic acids described herein can be expressed in mammalian cells using a mammalian expression vector. Examples of mammalian expression vectors include pCDM8 (Seed, *Nature* 329:840 (1987)) and pMT2PC (Kaufman et al., *EMBO J.* 6:187-195 (1987)). When used in mammalian cells, the expression vector's control functions can be provided by viral regulatory elements. For example, commonly used promoters are derived from polyoma, Adenovirus 2, cytomegalovirus, and Simian Virus 40. Other suitable expression systems for both prokaryotic and eukaryotic cells are described in chapters 16 and 17 of Sambrook et al., eds., Molecular Cloning: A Laboratory Manual. 2nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1989).

[0199] Vectors can be introduced into prokaryotic or eukaryotic cells via conventional transformation or transfection techniques. As used herein, the terms "transformation" and "transfection" refer to a variety of art-recognized techniques for introducing foreign nucleic acid (e.g., DNA) into a host cell, including calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, lipofection, or electroporation. Suitable methods for transforming or transfecting host cells can be found in, for example, Sambrook et al. supra.

[0200] For stable transformation of bacterial cells, it is known that, depending upon the expression vector and transformation technique used, only a small fraction of cells will take-up and replicate the expression vector. In order to identify and select these transformants, a gene that encodes a selectable marker (e.g., resistance to antibiotics) can be introduced into the host cells along with the gene of interest. Selectable markers include those that confer resistance to drugs, such as ampicillin, kanamycin, chloramphenicol, carbenicillin, spectinomycin, or tetracycline. Nucleic acids encoding a selectable marker can be introduced into a host cell on the same vector as that encoding a polypeptide described herein or can be introduced on a separate vector. Cells stably transfected with the introduced nucleic acid can be identified by drug selection (e.g., cells that have incorporated the selectable marker gene will survive, while the other cells die).

[0201] For stable transfection of mammalian cells, it is known that, depending upon the expression vector and transfection technique used, only a small fraction of cells may integrate the foreign DNA into their genome. In order to identify and select these integrants, a gene that encodes a selectable marker (e.g., resistance to antibiotics) can be introduced into the host cells along with the gene of interest. Preferred selectable markers include those which confer resistance to drugs, such as G418, hygromycin, and methotrexate. Nucleic acids encoding a selectable marker can be introduced into a host cell on the same vector as that encoding a polypeptide described herein or can be introduced on a separate vector. Cells stably transfected with the introduced nucleic acid can be identified by drug selection (e.g., cells that have incorporated the selectable marker gene will survive, while the other cells die).

Transport Proteins

[0202] Transport proteins can export polypeptides and organic compounds (e.g., fatty esters) out of a host cell. Many transport and efflux proteins serve to excrete a wide variety of compounds and can be naturally modified to be selective for particular types of fatty esters.

[0203] Non-limiting examples of suitable transport proteins are ATP-Binding Cassette (ABC) transport proteins, efflux proteins, and fatty acid transporter proteins (FATP). Additional non-limiting examples of suitable transport proteins include the ABC transport proteins from organisms such as *Caenorhabditis elegans, Arabidopsis thalania, Alkaligenes eutrophus*, and *Rhodococcus erythropolis*. Exemplary ABC transport proteins that can be used have been described in, for example, WO 08/119,082, the disclosure of which is incorporated herein by reference. Exemplary GenBank Accession numbers include, without limitation, CER5 [GenBank Accession No: At1g51500, AY734542, At3g21090, or At1g51460], AtMRP5 [GenBank Accession No: NP_171908], AmiS2 [GenBank Accession No. JC5491], or AtPGP1 [GenBank Accession No. NP_181228]. Host cells can also be chosen for their endogenous ability to secrete organic compounds. The efficiency of organic compound production and secretion into the host cell environment (e.g., culture medium, fermentation broth) can be expressed as a ratio of intracellular product to extracellular product. In some examples, the ratio can be about 5:1, 4:1, 3:1, 2:1, 1:1, 1:2, 1:3, 1:4, or 1:5.

[0204] In certain embodiments, the fatty esters produced by the host cells herein are present in the extracellular environment. In some embodiments, the fatty esters are isolated from the extracellular environment of the host cell. In some other embodiments, the fatty esters are spontaneously secreted, partially or completely, from the host cell. In further embodiments, the fatty esters are transported into the extracellular environment, optionally with the aid of one or more suitable transport proteins as described herein. In other embodiments, the fatty esters are passively transported into the extracellular environment.

Fermentation

[0205] The production and isolation of fatty esters can be enhanced by employing beneficial fermentation techniques. One method for maximizing production while reducing costs is increasing the percentage of the carbon source that is converted to fatty ester products.

[0206] During normal cellular lifecycles, carbon is used in cellular functions, such as producing lipids, saccharides, proteins, organic acids, and nucleic acids. Reducing the amount of carbon necessary for growth-related activities can increase the efficiency of carbon source conversion to product. This can be achieved by, for example, first growing host cells to a desired density (for example, a density achieved at the peak of the log phase of growth). At such a point, replication checkpoint genes can be harnessed to stop the growth of cells. Specifically, quorum sensing mechanisms (reviewed in Camilli et al., *Science* 311:1113, (2006); Venturi, *FEMS Microbio. Rev.* 30:274-291 (2006); and Reading et al., *FEMS Microbiol. Lett.* 254:1-11 (2006)) can be used to activate checkpoint genes, such as p53, p21, or other checkpoint genes.

[0207] Genes that can be activated to stop cell replication and growth in *E. coli* include umuDC genes. The overexpression of umuDC genes stops the progression from stationary phase to exponential growth (Murli et al., *J. of Bact.* 182:1127 (2000)). UmuC is a DNA polymerase that can carry out translesion synthesis over non-coding lesions—the mechanistic basis of most UV and chemical mutagenesis. The umuDC gene products are involved in the process of transle-

sion synthesis and also serve as a DNA sequence damage checkpoint. The umuDC gene products include UmuC, UmuD, umuD', $UmuD'_2C$, $UmuD'_2$, and $UmuD_2$. Simultaneously, product-producing genes can be activated, thus minimizing the need for replication and maintenance pathways to be used while a fatty ester is being made. Host cells can also be engineered to express umuC and umuD from *E. coli* in pBAD24 under the prpBCDE promoter system through de novo synthesis of this gene with the appropriate end-product production genes.

[0208] The percentage of input carbons converted to fatty esters can be a cost driver. The more efficient the process is (i.e., the higher the percentage of input carbons converted to fatty esters), the less expensive the process will be. For oxygen-containing carbon sources (e.g., glucose and other carbohydrate based sources), the oxygen must be released in the form of carbon dioxide. For every 2 oxygen atoms released, a carbon atom is also released leading to a maximal theoretical metabolic efficiency of approximately 34% (w/w) (for fatty acid derived products). This figure, however, changes for other organic compounds and carbon sources. Typical efficiencies in the literature are approximately less than 5%. Host cells engineered to produce fatty esters can have greater than about 1, 3, 5, 10, 15, 20, 25, and 30% efficiency. In one example, host cells can exhibit an efficiency of about 10% to about 25%. In other examples, such host cells can exhibit an efficiency of about 25% to about 30%. In other examples, host cells can exhibit greater than 30% efficiency.

[0209] The host cell can be additionally engineered to express recombinant cellulosomes, such as those described in PCT application number PCT/US2007/003736. These cellulosomes can allow the host cell to use cellulosic material as a carbon source. For example, the host cell can be additionally engineered to express invertases (EC 3.2.1.26) so that sucrose can be used as a carbon source. Similarly, the host cell can be engineered using the teachings described in U.S. Pat. Nos. 5,000,000; 5,028,539; 5,424,202; 5,482,846; and 5,602,030; so that the host cell can assimilate carbon efficiently and use cellulosic materials as carbon sources.

[0210] In one example, the fermentation chamber can enclose a fermentation that is undergoing a continuous reduction. In this instance, a stable reductive environment can be created. The electron balance can be maintained by the release of carbon dioxide (in gaseous form). Efforts to augment the NAD/H and NADP/H balance can also facilitate in stabilizing the electron balance. The availability of intracellular NADPH can also be enhanced by engineering the host cell to express an NADH:NADPH transhydrogenase or an NADPH:NADH transhydrogenase. The expression of one or more NADH:NADPH transhydrogenases converts the NADH produced in glycolysis to NADPH, which can enhance the production of fatty acids.

[0211] For small scale production, the engineered host cells can be grown in batches of, for example, about 100 mL, 500 mL, 1 L, 2 L, 5 L, 10 L or 25 L; fermented and induced to express desired biosynthetic genes based on the specific genes encoded in the appropriate plasmids. For large scale production, the engineered host cells can be grown in batches of about 50 L, 100 L, 1000 L, 10,000 L, 100,000 L, 1,000,000 L or larger; fermented and induced to express desired biosynthetic genes based on the specific genes encoded in the appropriate plasmids or incorporated into the host cell's genome.

[0212] For example, a suitable production host, such as *E. coli* cells, harboring plasmids containing the desired biosynthetic genes or having the biosynthetic genes integrated in its chromosome can be incubated in a suitable reactor, for example a 1 L reactor, for 20 h at 37° C. in M9 medium supplemented with 2% glucose, carbenicillin, and chloramphenicol. When the OD_{600} of the culture reaches 0.9, the production host can be induced with IPTG to activate the engineered gene systems for fatty ester production. After incubation, the spent media can be extracted and the organic phase can be examined for the presence of fatty esters using GC-MS.

[0213] In some instances, after the first hour of induction, aliquots of no more than about 10% of the total cell volume can be removed each hour and allowed to sit without agitation to allow the fatty esters to rise to the surface and undergo a spontaneous phase separation or precipitation. The fatty ester component can then be collected, and the aqueous phase returned to the reaction chamber. The reaction chamber can be operated continuously. When the OD_{600} drops below 0.6, the cells can be replaced with a new batch grown from a seed culture.

Glucose

[0214] In some instances, the methods disclosed herein are performed using glucose as a carbon source. In certain instances, microorganisms are grown in a culture medium containing an initial glucose concentration of about 2 g/L to about 100 g/L, such as about 5 g/L to about 20 g/L. In some instances, the glucose concentration of the culture medium decreases from the initial glucose concentration as the microorganisms consume the glucose, and a concentration of about 0 g/L to about 5 g/L glucose is maintained in the culture medium during the fatty ester production process. In certain instances, glucose is fed to the microorganisms in a solution of about 50% to about 65% wt/wt glucose.

[0215] In some instances, the feed rate of glucose is set to match the cells' growth rate to avoid excess accumulation of glucose (i.e., >0% glucose) in the fermentor. In other instances, and a low concentration of excess glucose (e.g., about 2 g/L to about 5 g/L) is maintained.

[0216] In certain instances, fatty esters can be produced from carbohydrates other than glucose, including but not limited to fructose, hydrolyzed sucrose, hydrolyzed molasses and glycerol.

Post-Production Processing

[0217] The fatty esters produced during fermentation can be separated from the fermentation media. Any known technique for separating fatty esters from aqueous media can be used. One exemplary separation process is a two phase (biphasic) separation process. This process involves fermenting the genetically engineered host cells under conditions sufficient to produce a fatty ester, allowing the fatty ester to collect in an organic phase, and separating the organic phase from the aqueous fermentation broth. This method can be practiced in both a batch and continuous fermentation processes.

[0218] Bi-phasic separation uses the relative immiscibility of fatty esters to facilitate separation. Immiscible refers to the relative inability of a compound to dissolve in water and is defined by the compound's partition coefficient. One of ordinary skill in the art will appreciate that by choosing a fermentation broth and organic phase, such that the fatty ester being

produced has a high logP value, the fatty ester can separate into the organic phase, even at very low concentrations, in the fermentation vessel.

[0219] The fatty esters produced by the methods described herein can be relatively immiscible in the fermentation broth, as well as in the cytoplasm. Therefore, the fatty ester can collect in an organic phase either intracellularly or extracellularly. The collection of the products in the organic phase can lessen the impact of the fatty ester on cellular function and can allow the host cell to produce more product.

[0220] The methods described herein can result in the production of homogeneous compounds wherein at least about 60%, 70%, 80%, 90%, or 95% of the fatty esters produced will have carbon chain lengths that vary by less than about 6 carbons, less than about 4 carbons, or less than about 2 carbons. These compounds can also be produced with a relatively uniform degree of saturation. These compounds can be used directly as fuels, fuel additives, starting materials for production of other chemical compounds (e.g., polymers, surfactants, plastics, textiles, solvents, adhesives, etc.), or personal care additives. These compounds can also be used as feedstock for subsequent reactions, for example, hydrogenation, catalytic cracking (e.g., via hydrogenation, pyrolisis, or both), to make other products.

[0221] In some embodiments, the fatty esters produced using methods described herein can contain between about 50% and about 90% carbon; or between about 5% and about 25% hydrogen. In other embodiments, the fatty esters produced using methods described herein can contain between about 65% and about 85% carbon; or between about 10% and about 15% hydrogen.

Bioproducts

[0222] Bioproducts (e.g., the fatty esters produced in accordance with the present disclosure) comprising biologically produced organic compounds, and in particular, the fatty esters biologically produced using the fatty acid biosynthetic pathway herein, have not been produced from renewable sources and, as such, are new compositions of matter. These new bioproducts can be distinguished from organic compounds derived from petrochemical carbon on the basis of dual carbon-isotopic fingerprinting or ^{14}C dating. Additionally, the specific source of biosourced carbon (e.g., glucose vs. glycerol) can be determined by dual carbon-isotopic fingerprinting (see, e.g., U.S. Pat. No. 7,169,588, which is herein incorporated by reference).

[0223] The ability to distinguish bioproducts from petroleum based organic compounds is beneficial in tracking these materials in commerce. For example, organic compounds or chemicals comprising both biologically based and petroleum based carbon isotope profiles may be distinguished from organic compounds and chemicals made only of petroleum based materials. Hence, the bioproducts herein can be followed or tracked in commerce on the basis of their unique carbon isotope profile.

[0224] Bioproducts can be distinguished from petroleum based organic compounds by comparing the stable carbon isotope ratio ($^{13}C/^{12}C$) in each fuel. The $^{13}C/^{12}C$ ratio in a given bioproduct is a consequence of the $^{13}C/^{12}C$ ratio in atmospheric carbon dioxide at the time the carbon dioxide is fixed. It also reflects the precise metabolic pathway. Regional variations also occur. Petroleum, C_3 plants (the broadleaf), C_4 plants (the grasses), and marine carbonates all show significant differences in $^{13}C/^{12}C$ and the corresponding $\delta^{13}C$ values. Furthermore, lipid matter of C_3 and C_4 plants analyze differently than materials derived from the carbohydrate components of the same plants as a consequence of the metabolic pathway.

[0225] Within the precision of measurement, ^{13}C shows large variations due to isotopic fractionation effects, the most significant of which for bioproducts is the photosynthetic mechanism. The major cause of differences in the carbon isotope ratio in plants is closely associated with differences in the pathway of photosynthetic carbon metabolism in the plants, particularly the reaction occurring during the primary carboxylation (i.e., the initial fixation of atmospheric CO_2). Two large classes of vegetation are those that incorporate the "C_3" (or Calvin-Benson) photosynthetic cycle and those that incorporate the "C_4" (or Hatch-Slack) photosynthetic cycle.

[0226] In C_3 plants, the primary CO_2 fixation or carboxylation reaction involves the enzyme ribulose-1,5-diphosphate carboxylase, and the first stable product is a 3-carbon compound. C_3 plants, such as hardwoods and conifers, are dominant in the temperate climate zones.

[0227] In C_4 plants, an additional carboxylation reaction involving another enzyme, phosphoenol-pyruvate carboxylase, is the primary carboxylation reaction. The first stable carbon compound is a 4-carbon acid that is subsequently decarboxylated. The CO_2 thus released is refixed by the C_3 cycle. Examples of C_4 plants are tropical grasses, corn, and sugar cane.

[0228] Both C_4 and C_3 plants exhibit a range of $^{13}C/^{12}C$ isotopic ratios, but typical values are about −7 to about −13 per mil for C_4 plants and about −19 to about −27 per mil for C_3 plants (see, e.g., Stuiver et al., *Radiocarbon* 19:355 (1977)). Coal and petroleum fall generally in this latter range. The ^{13}C measurement scale was originally defined by a zero set by Pee Dee Belemnite (PDB) limestone, where values are given in parts per thousand deviations from this material. The "$\delta^{13}C$" values are expressed in parts per thousand (per mil), abbreviated, %0, and are calculated as follows:

$$\delta^{13}C(‰)=[(^{13}C/^{12}C)_{sample}-(^{13}C/^{12}C)_{standard}]/(^{13}C/^{12}C)_{standard}\times 1000$$

[0229] Since the PDB reference material (RM) has been exhausted, a series of alternative RMs have been developed in cooperation with the IAEA, USGS, NIST, and other selected international isotope laboratories. Notations for the per mil deviations from PDB is $\delta^{13}C$. Measurements are made on CO_2 by high precision stable ratio mass spectrometry (IRMS) on molecular ions of masses 44, 45, and 46.

[0230] The compositions described herein include bioproducts produced by any of the methods described herein, including, for example, the fatty ester products. Specifically, the bioproduct can have a $\delta^{13}C$ of about −28 or greater, about −27 or greater, −20 or greater, −18 or greater, −15 or greater, −13 or greater, −10 or greater, or −8 or greater. For example, the bioproduct can have a $\delta^{13}C$ of about −30 to about −15, about −27 to about −19, about −25 to about −21, about −15 to about −5, about −13 to about −7, or about −13 to about −10. In other instances, the bioproduct can have a $\delta^{13}C$ of about −10, −11, −12, or −12.3.

[0231] Bioproducts, including the bioproducts produced in accordance with the disclosure herein, can also be distinguished from petroleum based organic compounds by comparing the amount of ^{14}C in each compound. Because ^{14}C has a nuclear half life of 5730 years, petroleum based fuels containing "older" carbon can be distinguished from bioproducts which contain "newer" carbon (see, e.g., Currie, "Source

Apportionment of Atmospheric Particles", *Characterization of Environmental Particles*, J. Buffle and H. P. van Leeuwen, Eds., 1 of Vol. I of the IUPAC Environmental Analytical Chemistry Series (Lewis Publishers, Inc) 3-74, (1992)).

[0232] The basic assumption in radiocarbon dating is that the constancy of ^{14}C concentration in the atmosphere leads to the constancy of ^{14}C in living organisms. However, because of atmospheric nuclear testing since 1950 and the burning of fossil fuel since 1850, ^{14}C has acquired a second, geochemical time characteristic. Its concentration in atmospheric CO_2, and hence in the living biosphere, approximately doubled at the peak of nuclear testing, in the mid-1960s. It has since been gradually returning to the steady-state cosmogenic (atmospheric) baseline isotope rate ($^{14}C/^{12}C$) of about 1.2×10^{-12}, with an approximate relaxation "half-life" of 7-10 years. (This latter half-life must not be taken literally; rather, one must use the detailed atmospheric nuclear input/decay function to trace the variation of atmospheric and biospheric ^{14}C since the onset of the nuclear age.)

[0233] It is this latter biospheric ^{14}C time characteristic that holds out the promise of annual dating of recent biospheric carbon. ^{14}C can be measured by accelerator mass spectrometry (AMS), with results given in units of "fraction of modern carbon" (f_M). f_M is defined by National Institute of Standards and Technology (NIST) Standard Reference Materials (SRMs) 4990B and 4990C. As used herein, "fraction of modern carbon" or "f_M" has the same meaning as defined by National Institute of Standards and Technology (NIST) Standard Reference Materials (SRMs) 4990B and 4990C, known as oxalic acids standards HOxI and HOxII, respectively. The fundamental definition relates to 0.95 times the $^{14}C/^{12}C$ isotope ratio HOxI (referenced to AD 1950). This is roughly equivalent to decay-corrected pre-Industrial Revolution wood. For the current living biosphere (plant material), f_M is approximately 1.1.

[0234] The invention provides a bioproduct comprising one or more fatty esters, which can have an f_M ^{14}C of at least about 1. For example, the bioproduct of the invention can have an f_M ^{14}C of at least about 1.01, an f_M ^{14}C of about 1 to about 1.5, an f_M ^{14}C of about 1.04 to about 1.18, or an f_M ^{14}C of about 1.111 to about 1.124.

[0235] Another measurement of ^{14}C is known as the percent of modern carbon (pMC). For an archaeologist or geologist using ^{14}C dates, AD 1950 equals "zero years old". This also represents 100 pMC. "Bomb carbon" in the atmosphere reached almost twice the normal level in 1963 at the peak of thermo-nuclear weapons. Its distribution within the atmosphere has been approximated since its appearance, showing values that are greater than 100 pMC for plants and animals living since AD 1950. It has gradually decreased over time with today's value being near 107.5 pMC. This means that a fresh biomass material, such as corn, would give a ^{14}C signature near 107.5 pMC. Petroleum based compounds will have a pMC value of zero. Combining fossil carbon with present day carbon will result in a dilution of the present day pMC content. By presuming 107.5 pMC represents the ^{14}C content of present day biomass materials and 0 pMC represents the ^{14}C content of petroleum based products, the measured pMC value for that material will reflect the proportions of the two component types. For example, a material derived 100% from present day soybeans would give a radiocarbon signature near 107.5 pMC. If that material was diluted 50% with petroleum based products, it would give a radiocarbon signature of approximately 54 pMC.

[0236] A biologically based carbon content is derived by assigning "100%" equal to 107.5 pMC and "0%" equal to 0 pMC. For example, a sample measuring 99 pMC will give an equivalent biologically based carbon content of 93%. This value is referred to as the mean biologically based carbon result and assumes all the components within the analyzed material originated either from present day biological material or petroleum based material.

[0237] A bioproduct comprising one or more fatty esters as described herein can have a pMC of at least about 50, 60, 70, 75, 80, 85, 90, 95, 96, 97, 98, 99, or 100. In other instances, a bioproduct described herein can have a pMC of between about 50 and about 100; about 60 and about 100; about 70 and about 100; about 80 and about 100; about 85 and about 100; about 87 and about 98; or about 90 and about 95. In yet other instances, a bioproduct described herein can have a pMC of about 90, 91, 92, 93, 94, or 94.2.

[0238] Fatty esters produced by the methods described herein can be used as biofuels. For example, any fatty acid methyl ester described herein can be used solely or as a component of biodiesel.

[0239] The invention is further illustrated by the following examples. The examples are provided for illustrative purposes only. They are not to be construed as limiting the scope or content of the invention in any way.

EXAMPLES

Example 1

Production of *E. coli* MG1655 ΔfadE

[0240] This example describes the construction of a genetically engineered microorganism wherein the expression of a fatty acid degradation enzyme is attenuated.

[0241] The fadE gene of *E. coli* MG1655 was deleted using the Lambda Red (also known as the Red-Driven Integration) system described in Datsenko et al., *Proc. Natl. Acad. Sci. USA* 97: 6640-6645 (2000), with the following modifications.

[0242] Two primers were used to create the deletion:

Del-fadE-F:
(SEQ ID NO: 1)
5'-AAAAACAGCAACAATGTGAGCTTTGTTGTAATTATATTGTAAACAT
ATTGATTCCGGGGATCCGTCGACC-3'

Del-fadE-R:
(SEQ ID NO: 2)
5'-AAACGGAGCCTTTCGGCTCCGTTATTCATTTACGCGGCTTCAACTT
TCCTGTAGGCTGGAGCTGCTTC-3'

[0243] The Del-fadE-F and Del-fadE-R primers were used to amplify the Kanamycin resistance (Km^R) cassette from plasmid pKD13 (as described in Datsenko et al., supra) by PCR. The PCR product was then used to transform electrocompetent *E. coli* MG1655 cells containing pKD46 (described in Datsenko et al., supra). These cells had been previously induced with arabinose for 3-4 h. Following a 3-h outgrowth in a Super Optimal Broth with Catabolite repression (SOC) medium at 37° C., the cells were plated on Luria agar plates containing 50 μg/mL Kanamycin. Resistant colonies were identified and isolated after an overnight incubation at 37° C. Disruption of the fadE gene was confirmed in select colonies using PCR amplification with primers fadE-L2 and fadE-R1, which were designed to flank the fadE gene:

(SEQ ID NO: 3)
fadE-L2 5'-CGGGCAGGTGCTATGACCAGGAC-3'

(SEQ ID NO: 4)
fadE-R1 5'-CGCGGCGTTGACCGGCAGCCTGG-3'

[0244] After the fadE deletion was confirmed, a single colony was used to remove the Km^R marker, using the pCP20 plasmid as described in Datsenko et al., supra. The resulting MG1655 *E. coli* strain with the fadE gene deleted and the Km^R marker removed was named *E. coli* MG1655 ΔfadE, or *E. coli* MG1655 D1.

Example 2

Production of *E. coli* MG1655 ΔfadE ΔfhuA

[0245] This example describes the construction of a genetically engineered microorganism in which the expression of a fatty acid degradation enzyme and an outer membrane protein receptor are attenuated.

[0246] The fhuA (also known as tonA) gene of *E. coli* MG1655, which encodes a ferrichrome outer membrane transporter (GenBank Accession No. NP_414692), was deleted from strain *E. coli* MG1655 D1 of Example 1 using the Lambda Red system described in Datsenko et al., supra, but with the following modifications.

[0247] Two primers were used to create the deletion:

Del-fhuA-F:
(SEQ ID NO: 5)
5'-ATCATTCTCGTTTACGTTATCATTCACTTTACATCAGAGATATACC
AATGATTCCGGGGATCCGTCGACC-3';

Del-fhuA-R:
(SEQ ID NO: 6)
5'-GCACGGAAATCCGTGCCCCAAAAGAGAAATTAGAAACGGAAGGTT
GCGGTTGTAGGCTGGAGCTGCTTC-3'

[0248] The Del-fhuA-F and Del-fhuA-R primers were used to amplify the Km^R cassette from plasmid pKD13 by PCR. The PCR product obtained was used to transform the electrocompetent *E. coli* MG1655 D1 cells containing pKD46 (see Example 1). These cells had been previously induced with arabinose for 3-4 h. Following a 3-h outgrowth in SOC medium at 37° C., the cells were plated on Luria agar plates containing 50 μg/mL Kanamycin. Kanamycin resistant colonies were identified and isolated after an overnight incubation at 37° C. Disruption of the fhuA gene was confirmed in select colonies by PCR amplification with primers fhuA-verF and fhuA-verR, which were designed to flank the fhuA gene.

[0249] Confirmation of the deletion was performed using the following primers:

fhuA-verF:
5'-CAACAGCAACCTGCTCAGCAA-3' (SEQ ID NO: 7)

fhuA-verR:
5'-AAGCTGGAGCAGCAAAGCGTT-3' (SEQ ID NO: 8)

[0250] After the fhuA deletion was confirmed, a single colony was used to remove the Km^R marker, using the pCP20 plasmid as described in Datsenko et al., supra. The resulting MG1655 *E. coli* strain having the fadE and fhuA gene deletions was named *E. coli* MG1655 ΔfadE ΔfhuA, or *E. coli* MG1655 DV2.

Example 3

Production of *E. coli* MG1655 ΔfadE ΔfhuA ΔpflB ΔldhA

[0251] This example describes the construction of a genetically engineered microorganism in which the expression of an acyl-CoA dehydrogenase, an outer membrane protein receptor, a pyruvate formate lyase and a lactate dehydrogenase are attenuated.

[0252] The pflB gene of *E. coli* MG1655, which encodes a pyruvate formate lyase (GenBank Accession No. AAC73989), was deleted from *E. coli* MG1655 DV2 (see, Example 2) using the Lambda Red System according to Datsenko et al., supra, but with the following modifications:

[0253] The primers used to create the deletion strain were:

Del-pflB-F:
(SEQ ID NO: 33)
5'-GCCGCAGCCTGATGGACAAAGCGTTCATTATGGTGCTGCCGGT
CGCGATGATTCCGGGGATCCGTCGACC-3'

Del-pflB-R:
(SEQ ID NO: 34)
5'-ATCTTCAACGGTAACTTCTTTACCGCCATGCGTGTCCCAGGTG
TCTGTAGGCTGGAGCTGCTTCG-3'

[0254] The Del-pflB-F and Del-pflB-R primers were used to amplify the Kanamycin resistance (Km^R) cassette from plasmid pKD13 by PCR. The PCR product was then used to transform electrocompetent *E. coli* MG1655 DV2 cells (see Example 2).

[0255] In parallel, the ldhA gene of *E. coli* MG1655, which encodes a lactate dehydrogenase, specifically an NAD-linked fermentative D-lactate dehydrogenase (see, e.g., Mat-Jan et al., *J. Bacteriol.* 171(1):342-8 (1989); Bunch et al., *Microbiol.* 143(1):187-95 (1997) (GenBank Accession No. AAC74462) was also deleted from *E. coli* MG1655 DV2 (see, Example 2) using the Lambda Red System according to Datsenko et al., supra, but with the following modifications.

[0256] Two primers were used to create the deletion:

Del-ldhA-F:
(SEQ ID NO: 35)
5'-CTCCCCTGGAATGCAGGGGAGCGGCAAGATTAAACCAGTTCGT
TCG GGCAGTGTAGGCTGGAGCTGCTTCG-3'

Del-ldhA-R:
(SEQ ID NO: 36)
5'-TATTTTTAGTAGCTTAAATGTGATTCAACATCACTGGAGAAAG
TC TTATGCATATGAATATCCTCCTTAGTTCC-3'

[0257] The Del-ldhA-F and Del-ldhA-R primers were used to amplify the chloramphenicol acetyltransferase resistance (Cm^R) cassette from plasmid pKD3 (see, Datsenko et al., supra) by PCR. The PCR product was also used to transform electrocompetent *E. coli* MG1655 DV2 cells (see, Example 2).

[0258] The *E. coli* MG1655 DV2 (see Example 2) cells had been previously induced with arabinose for about 3-4 h. Fol-

lowing a 3-h outgrowth in SOC medium at 37° C., the cells were plated on Luria agar plates containing 50 μg/mL of kanamycin and 30 μg/mL chloramphenicol. Colonies that were resistant to both kanamycin and chloramphenicol were identified and isolated after an overnight incubation at 37° C. Disruption of the pflB gene was confirmed using primers flanking the *E. coli* pflB gene, and disruption of the ldhA gene was verified using primers flanking the *E. coli* ldhA gene.

[0259] Confirmation of the deletion of pflB was performed using the following primers:

```
pflB-verF:
5'-GGACTAAACGTCCTACAAAC-3'     (SEQ ID NO: 37)

PflB-verR:
5'-TTCATCTGTTTGAGATCGAG-3'     (SEQ ID NO: 38)
```

[0260] Confirmation of the deletion of ldhA gene was performed using the following primers:

```
                                        (SEQ ID NO: 39)
ldhA-verF: 5'-CCCGAGCGGTAGCCAGATGCCCGCCAGCG-3'

                                        (SEQ ID NO: 40)
ldhA-verR: 5'-GCTGCGGGTTAGCGCACATCATACGGGTC-3'
```

[0261] After the deletions were confirmed, a single colony was used to remove the Km^R and Cm^R markers in accordance with the method described by Datsenko et al., supra. The resultant MG1655 *E. coli* strain having fadE, fhuA, pflB and ldhA gene deletions was named *E. coli* MG1655 ΔfadE_ΔfhuA_ΔpflB_ΔldhA, or *E. coli* MG1655 DV4.

Example 4

Production of *E. coli* MG1655 ΔfadE, ΔfhuA, $lacI_q$-Ptrc-'tesA-fadD

[0262] This example describes the construction of a genetically engineered microorganism in which nucleotide sequences encoding a thioesterase, and an acyl-CoA synthase are integrated into the microorganism's chromosome, under the control of a promoter.

[0263] Plasmid pGRG25 (GenBank Accession No. DQ460223) (SEQ ID NO:9) was purified and subject to restriction digestions by NotI and AvrII (New England Biolabs, Inc., Ipswich, Mass.). In parallel, a plasmid pACYC-'tesA-fadD, which contained the $lacI_q$, Ptrc-'tesA-fadD cassette was constructed as follows:

[0264] 'tesA is a nucleotide sequence comprising a leaderless *E. coli* tesA gene (GenBank entry AAC73596, refseq accession U00096.2). 'tesA encodes an *E. coli* thioesterase (EC 3.1.1.5, 3.1.2.-) in which the first twenty-five amino acids were deleted and the amino acid in position 26, alanine, was replaced with methionine. That methionine then became the first amino acid of 'tesA. See Cho et al., *J. Biol. Chem.*, 270:4216-4219 (1995).

[0265] *E. coli* fadD (GenBank entry AAC74875; REFSEQ: accession U00096.2) encodes an acyl-CoA synthase.

Construction of the 'tesA Plasmid

[0266] 'tesA was amplified from a pETDuet-1-'tesA plasmid constructed as described below. (see also, e.g., WO 2007/136762 A2, which is incorporated by reference). The 'tesA gene was cloned into an NdeI/AvrII digested pETDuet-1 plasmid (Novagen, Madison, Wis.).

Construction of the fadD Plasmid

[0267] The fadD gene was amplified from a pHZ1.61 plasmid constructed as described below. A fadD gene was cloned into a pCDFDuet-1 plasmid (Novagen, Madison, Wis.) under the control of a T7 promoter, generating a pHZ1.61 plasmid containing the following nucleotide sequence:

(SEQ ID NO: 10)

```
GGGGAATTGTGAGCGGATAACAATTCCCCTGTAGAAATAATTTTGTTT

AACTTTAATAAGGAGATATACCATGGTGAAGAAGGTTTGGCTTAACCG

TTATCCCGCGGACGTTCCGACGGAGATCAACCCTGACCGTTATCAATC

TCTGGTAGATATGTTTGAGCAGTCGGTCGCGCGCTACGCCGATCAACC

TGCGTTTGTGAATATGGGGGAGGTAATGACCTTCCGCAAGCTGGAAGA

ACGCAGTCGCGCGTTTGCCGCTTATTTGCAACAAGGGTTGGGGCTGAA

GAAAGGCGATCGCGTTGCGTTGATGATGCCTAATTTATTGCAATATCC

GGTGGCGCTGTTTGGCATTTTGCGTGCCGGGATGATCGTCGTAAACGT

TAACCCGTTGTATACCCCGCGTGAGCTTGAGCATCAGCTTAACGATAG

CGGCGCATCGGCGATTGTTATCGTGTCTAACTTTGCTCACACACTGGA

AAAAGTGGTTGATAAAACCGCCGTTCAGCACGTAATTCTGACCCGTAT

GGGCGATCAGCTATCTACGGCAAAAGGCACGGTAGTCAATTTCGTTGT

TAAATACATCAAGCGTTTGGTGCCGAAATACCATCTGCCAGATGCCAT

TTCATTTCGTAGCGCACTGCATAACGGCTACCGGATGCAGTACGTCAA

ACCCGAACTGGTGCCGGAAGATTTAGCTTTTCTGCAATACACCGGCGG

CACCACTGGTGTGGCGAAAGGCGCGATGCTGACTCACCGCAATATGCT

GGCGAACCTGGAACAGGTTAACGCGACCTATGGTCCGCTGTTGCATCC

GGGCAAAGAGCTGGTGGTGACGGCGCTGCCGCTGTATCACATTTTTGC

CCTGACCATTAACTGCCTGCTGTTTATCGAACTGGGTGGGCAGAACCT

GCTTATCACTAACCCGCGCGATATTCCAGGGTTGGTAAAAGAGTTAGC

GAAATATCCGTTTACCGCTATCACGGGCGTTAACACCTTGTTCAATGC

GTTGCTGAACAATAAAGAGTTCCAGCAGCTGGATTTCTCCAGTCTGCA

TCTTTCCGCAGGCGGAGGGATGCCAGTGCAGCAAGTGGTGGCAGAGCG

TTGGGTGAAACTGACAGGACAGTATCTGCTGGAAGGCTATGGCCTTAC

CGAGTGTGCGCCGCTGGTCAGCGTTAACCCATATGATATTGATTATCA

TAGTGGTAGCATCGGTTTGCCGGTGCCGTCGACGGAAGCCAAACTGGT

GGATGATGATGATAATGAAGTACCACCGGGTCAACCGGGTGAGCTTTG

TGTCAAAGGACCGCAGGTGATGCTGGGTTACTGGCAGCGTCCGGATGC

TACAGATGAGATCATCAAAAATGGCTGGTTACACACCGGCGACATCGC

GGTGATGGATGAAGAAGGGTTCCTGCGCATTGTCGATCGTAAAAAAGA

CATGATTCTGGTTTCCGGTTTTAACGTCTATCCCAACGAGATTGAAGA

TGTCGTCATGCAGCATCCTGGCGTACAGGAAGTCGCGGCTGTTGGCGT

ACCTTCCGGCTCCAGTGGTGAAGCGGTGAAAATCTTCGTAGTGAAAAA

AGATCCATCGCTTACCGAAGAGTCACTGGTGACCTTTTGCCGCCGTCA

GCTCACGGGCTACAAAGTACCGAAGCTGGTGGAGTTTCGTGATGAGTT
```

-continued

ACCGAAATCTAACGTCGGAAAAATTTTGCGACGAGAATTACGTGACGA

AGCGCGCGGCAAAGTGGACAATAAAGCCTGAAAGCTTGCGGCCGCATA

ATGCTTAAGTCGAACAGAAAGTAATCGTATTGTACACGGCCGCATAAT

CGAAATTAATACGACTCACTATAGGGGAATTGTGAGCGGATAACAATT

CCCCATCTTAGTATATTAGTTAAGTATAAGAAGGAGATATACATATGC

GCCCATTACATCCGATTGATTTTATATTCCTGTCACTAGAAAAAAGAC

AACAGCCTATGCATGTAGGTGGTTTATTTTTGTTTCAGATTCCTGATA

ACGCCCCAGACACCTTTATTCAAGATCTGGTGAATGATATCCGGATAT

CAAAATCAATCCCTGTTCCACCATTCAACAATAAACTGAATGGGCTTT

TTTGGGATGAAGATGAAGAGTTTGATTTAGATCATCATTTTCGTCATA

TTGCACTGCCTCATCCTGGTCGTATTCGTGAATTGCTTATTTATATTT

CACAAGAGCACAGTACGCTGCTAGATCGGGCAAAGCCCTTGTGGACCT

GCAATATTATTGAAGGAATTGAAGGCAATCGTTTTGCCATGTACTTCA

AAATTCACCATGCGATGGTCGATGGCGTTGCTGGTATGCGGTTAATTG

AAAAATCACTCTCCCATGATGTAACAGAAAAAAGTATCGTGCCACCTT

GGTGTGTTGAGGGAAAACGTGCAAAGCGCTTAAGAGAACCTAAAACAG

GTAAAATTAAGAAAATCATGTCTGGTATTAAGAGTCAGCTTCAGGCGA

CACCCACAGTCATTCAAGAGCTTTCTCAGACAGTATTTAAAGATATTG

GACGTAATCCTGATCATGTTTCAAGCTTTCAGGCGCCTTGTTCTATTT

TGAATCAGCGTGTGAGCTCATCGCGACGTTTTGCAGCACAGTCTTTTG

ACCTAGATCGTTTTCGTAATATTGCCAAATCGTTGAATGTGACCATTA

ATGATGTTGTACTAGCGGTATGTTCTGGTGCATTACGTGCGTATTTGA

TGAGTCATAATAGTTTGCCTTCAAAACCATTAATTGCCATGGTTCCAG

CCTCTATTCGCAATGACGATTCAGATGTCAGCAACCGTATTACGATGA

TTCTGGCAAATTTGGCAACCCACAAAGATGATCCTTTACAACGTCTTG

AAATTATCCGCCGTAGTGTTCAAAACTCAAAGCAACGCTTCAAACGTA

TGACCAGCGATCAGATTCTAAATTATAGTGCTGTCGTATATGGCCCTG

CAGGACTCAACATAATTTCTGGCATGATGCCAAAACGCCAAGCCTTCA

ATCTGGTTATTTCCAATGTGCCTGGCCCAAGAGAGCCACTTTACTGGA

ATGGTGCCAAACTTGATGCACTCTACCCAGCTTCAATTGTATTAGACG

GTCAAGCATTGAATATTACAATGACCAGTTATTTAGATAAACTTGAAG

TTGGTTTGATTGCATGCCGTAATGCATTGCCAAGAATGCAGAATTTAC

TGACACATTTAGAAGAAGAAATTCAACTATTTGAAGGCGTAATTGCAA

AGCAGGAAGATATTAAAACAGCCAATTAAAAACAATAAACTTGATTTT

TTAATTTATCAGATAAAACTAAAGGGCTAAATTAGCCCTCCTAGGCTG

CTGCCACCGCTGAGCAATAACTAGCATAACCCCTTGGGGCCTCTAAAC

GGGTCTTGAGGGGTTTTTTGCTGAAACCTCAGGCATTTGAGAAGCACA

CGGTCACACTGCTTCCGGTAGTCAATAAACCGGTAAACCAGCAATAGA

-continued

CATAAGCGGCTATTTAACGACCCTGCCCTGAACCGACGACCGGGTCAT

CGTGGCCGGATCTTGCGGCCCCTCGGCTTGAACGAATTGTTAGACATT

ATTTGCCGACTACCTTGGTGATCTCGCCTTTCACGTAGTGGACAAATT

CTTCCAACTGATCTGCGCGCGAGGCCAAGCGATCTTCTTCTTGTCCAA

GATAAGCCTGTCTAGCTTCAAGTATGACGGGCTGATACTGGGCCGGCA

GGCGCTCCATTGCCCAGTCGGCAGCGACATCCTTCGGCGCGATTTTGC

CGGTTACTGCGCTGTACCAAATGCGGGACAACGTAAGCACTACATTTC

GCTCATCGCCAGCCCAGTCGGGCGGCGAGTTCCATAGCGTTAAGGTTT

CATTTAGCGCCTCAAATAGATCCTGTTCAGGAACCGGATCAAAGAGTT

CCTCCGCCGCTGGACCTACCAAGGCAACGCTATGTTCTCTTGCTTTTG

TCAGCAAGATAGCCAGATCAATGTCGATCGTGGCTGGCTCGAAGATAC

CTGCAAGAATGTCATTGCGCTGCCATTCTCCAAATTGCAGTTCGCGCT

TAGCTGGATAACGCCACGGAATGATGTCGTCGTGCACAACAATGGTGA

CTTCTACAGCGCGGAGAATCTCGCTCTCTCCAGGGGAAGCCGAAGTTT

CCAAAAGGTCGTTGATCAAAGCTCGCCGCGTTGTTTCATCAAGCCTTA

CGGTCACCGTAACCAGCAAATCAATATCACTGTGTGGCTTCAGGCCGC

CATCCACTGCGGAGCCGTACAAATGTACGGCCAGCAACGTCGGTTCGA

GATGGCGCTCGATGACGCCAACTACCTCTGATAGTTGAGTCGATACTT

CGGCGATCACCGCTTCCCTCATACTCTTCCTTTTTCAATATTATTGAA

GCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTA

TTTAGAAAAATAAACAAATAGCTAGCTCACTCGGTCGCTACGCTCCGG

GCGTGAGACTGCGGCGGGCGCTGCGGACACATACAAAGTTACCCACAG

ATTCCGTGGATAAGCAGGGGACTAACATGTGAGGCAAAACAGCAGGGC

CGCGCCGGTGGCGTTTTTCCATAGGCTCCGCCCCTCCTGCCAGAGTTCA

CATAAACAGACGCTTTTCCGGTGCATCTGTGGGAGCCGTGAGGCTCAA

CCATGAATCTGACAGTACGGGCGAAACCCGACAGGACTTAAAGATCCC

CACCGTTTCCGGCGGGTCGCTCCCTCTTGCGCTCTCCTGTTCCGACCC

TGCCGTTTACCGGATACCTGTTCCGCCTTTCTCCCTTACGGGAAGTGT

GGCGCTTTCTCATAGCTCACACACTGGTATCTCGGCTCGGTGTAGGTC

GTTCGCTCCAAGCTGGGCTGTAAGCAAGAACTCCCCGTTCAGCCCGAC

TGCTGCGCCTTATCCGGTAACTGTTCACTTGAGTCCAACCCGGAAAAG

CACGGTAAAACGCCACTGGCAGCAGCCATTGGTAACTGGGAGTTCGCA

GAGGATTTGTTTAGCTAAACACGCGGTTGCTCTTGAAGTGTGCGCCAA

AGTCCGGCTACACTGGAAGGACAGATTTGGTTGCTGTGCTCTGCGAAA

GCCAGTTACCACGGTTAAGCAGTTCCCCAACTGACTTAACCTTCGATC

AAACCACCTCCCCAGGTGGTTTTTTCGTTTACAGGGCAAAAGATTACG

CGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACTGAA

CCGCTCTAGATTTCAGTGCAATTTATCTCTTCAAATGTAGCACCTGAA

GTCAGCCCCATACGATATAAGTTGTAATTCTCATGTTAGTCATGCCCC

-continued

GCGCCCACCGGAAGGAGCTGACTGGGTTGAAGGCTCTCAAGGGCATCG

GTCGAGATCCCGGTGCCTAATGAGTGAGCTAACTTACATTAATTGCGT

TGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGC

ATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGC

GCCAGGGTGGTTTTTCTTTTCACCAGTGAGACGGGCAACAGCTGATTG

CCCTTCACCGCCTGGCCCTGAGAGAGTTGCAGCAAGCGGTCCACGCTG

GTTTGCCCCAGCAGGCGAAAATCCTGTTTGATGGTGGTTAACGGCGGG

ATATAACATGAGCTGTCTTCGGTATCGTCGTATCCCACTACCGAGATG

TCCGCACCAACGCGCAGCCCGGACTCGGTAATGGCGCGCATTGCGCCC

AGCGCCATCTGATCGTTGGCAACCAGCATCGCAGTGGGAACGATGCCC

TCATTCAGCATTTGCATGGTTTGTTGAAAACCGGACATGGCACTCCAG

TCGCCTTCCCGTTCCGCTATCGGCTGAATTTGATTGCGAGTGAGATAT

TTATGCCAGCCAGCCAGACGCAGACGCGCCGAGACAGAACTTAATGGG

CCCGCTAACAGCGCGATTTGCTGGTGACCCAATGCGACCAGATGCTCC

ACGCCCAGTCGCGTACCGTCTTCATGGGAGAAAATAATACTGTTGATG

GGTGTCTGGTCAGAGACATCAAGAAATAACGCCGGAACATTAGTGCAG

GCAGCTTCCACAGCAATGGCATCCTGGTCATCCAGCGGATAGTTAATG

ATCAGCCCACTGACGCGTTGCGCGAGAAGATTGTGCACCGCCGCTTTA

CAGGCTTCGACGCCGCTTCGTTCTACCATCGACACCACCACGCTGGCA

CCCAGTTGATCGGCGCGAGATTTAATCGCCGCGACAATTTGCGACGGC

GCGTGCAGGGCCAGACTGGAGGTGGCAACGCCAATCAGCAACGACTGT

TTGCCCGCCAGTTGTTGTGCCACGCGGTTGGGAATGTAATTCAGCTCC

GCCATCGCCGCTTCCACTTTTTCCCGCGTTTTCGCAGAAACGTGGCTG

GCCTGGTTCACCACGCGGGAAACGGTCTGATAAGAGACACCGGCATAC

TCTGCGACATCGTATAACGTTACTGGTTTCACATTCACCACCCTGAAT

TGACTCTCTTCCGGGCGCTATCATGCCATACCGCGAAAGGTTTTGCGC

CATTCGATGGTGTCCGGGATCTCGACGCTCTCCCTTATGCGACTCCTG

CATTAGGAAATTAATACGACTCACTATA

Construction of pACYC-Ptrc Plasmid Containing 'tesA and fadD

[0268] A pACYC-Ptrc vector having the following sequence was used to construct a pACYC-Ptrc-'tesA-fadD plasmid. The nucleotide sequence of the pACYC-Ptrc vector is as follows:

(SEQ ID NO: 11)

ACTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAG

AATTATGCAGTGCTGCCATAACCATGAGTGATAACACTGCGGCCAACT

TACTTCTGACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTGC

ACAACATGGGGGATCATGTAACTCGCCTTGATCGTTGGGAACCGGAGC

TGAATGAAGCCATACCAAACGACGAGCGTGACACCACGATGCCTGCAG

-continued

CAATGGCAACAACGTTGCGCAAACTATTAACTGGCGAACTACTTACTC

TAGCTTCCCGGCAACAATTAATAGACTGGATGGAGGCGGATAAAGTTG

CAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTATTGCTG

ATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTGCAGCAC

TGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGG

GGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCTGAGATAG

GTGCCTCACTGATTAAGCATTGGTAACTGTCAGACCAAGTTTACTCAT

ATATACTTTAGATTGATTTAAAACTTCATTTTTAATTTAAAAGGATCT

AGGTGAAGATCCTTTTTGATAATCTCATGACCAAAATCCCTTAACGTG

AGTTTTCGTTCCACTGAGCGTCAGACCCCTTAATAAGATGATCTTCTT

GAGATCGTTTTGGTCTGCGCGTAATCTCTTGCTCTGAAAACGAAAAAA

CCGCCTTGCAGGGCGGTTTTTCGAAGGTTCTCTGAGCTACCAACTCTT

TGAACCGAGGTAACTGGCTTGGAGGAGCGCAGTCACCAAAACTTGTCC

TTTCAGTTTAGCCTTAACCGGCGCATGACTTCAAGACTAACTCCTCTA

AATCAATTACCAGTGGCTGCTGCCAGTGGTGCTTTTGCATGTCTTTCC

GGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGAC

TGAACGGGGGGTTCGTGCATACAGTCCAGCTTGGAGCGAACTGCCTAC

CCGGAACTGAGTGTCAGGCGTGGAATGAGACAAACGCGGCCATAACAG

CGGAATGACACCGGTAAACCGAAAGGCAGGAACAGGAGAGCGCACGAG

GGAGCCGCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTT

TCGCCACCACTGATTTGAGCGTCAGATTTCGTGATGCTTGTCAGGGGG

GCGGAGCCTATGGAAAAACGGCTTTGCCGCGGCCCTCTCACTTCCCTG

TTAAGTATCTTCCTGGCATCTTCCAGGAAATCTCCGCCCCGTTCGTAA

GCCATTTCCGCTCGCCGCAGTCGAACGACCGAGCGTAGCGAGTCAGTG

AGCGAGGAAGCGGAATATATCCTGTATCACATATTCTGCTGACGCACC

GGTGCAGCCTTTTTTCTCCTGCCACATGAAGCACTTCACTGACACCCT

CATCAGTGCCAACATAGTAAGCCAGTATACACTCCGCTAGCGCTGAGG

TCTGCCTCGTGAAGAAGGTGTTGCTGACTCATACCAGGCCTGAATCGC

CCCATCATCCAGCCAGAAAGTGAGGGAGCCACGGTTGATGAGAGCTTT

GTTGTAGGTGGACCAGTTGGTGATTTTGAACTTTTGCTTTGCCACGGA

ACGGTCTGCGTTGTCGGGAAGATGCGTGATCTGATCCTTCAACTCAGC

AAAAGTTCGATTTATTCAACAAAGCCACGTTGTGTCTCAAAATCTCTG

ATGTTACATTGCACAAGATAAAAATATATCATCATGAACAATAAAACT

GTCTGCTTACATAAACAGTAATACAAGGGGTGTTATGAGCCATATTCA

ACGGGAAACGTCTTGCTCGAGGCCGCGATTAAATTCCAACATGGATGC

TGATTTATATGGGTATAAATGGGCTCGCGATAATGTCGGGCAATCAGG

TGCGACAATCTATCGATTGTATGGGAAGCCCGATGCGCCAGAGTTGTT

TCTGAAACATGGCAAAGGTAGCGTTGCCAATGATGTTACAGATGAGAT

-continued

GGTCAGACTAAACTGGCTGACGGAATTTATGCCTCTTCCGACCATCAA

GCATTTTATCCGTACTCCTGATGATGCATGGTTACTCACCACTGCGAT

CCCCGGGAAAACAGCATTCCAGGTATTAGAAGAATATCCTGATTCAGG

TGAAAATATTGTTGATGCGCTGGCAGTGTTCCTGCGCCGGTTGCATTC

GATTCCTGTTTGTAATTGTCCTTTTAACAGCGATCGCGTATTTCGTCT

CGCTCAGGCGCAATCACGAATGAATAACGGTTTGGTTGATGCGAGTGA

TTTTGATGACGAGCGTAATGGCTGGCCTGTTGAACAAGTCTGGAAAGA

AATGCATAAGCTTTTGCCATTCTCACCGGATTCAGTCGTCACTCATGG

TGATTTCTCACTTGATAACCTTATTTTTGACGAGGGGAAATTAATAGG

TTGTATTGATGTTGGACGAGTCGGAATCGCAGACCGATACCAGGATCT

TGCCATCCTATGGAACTGCCTCGGTGAGTTTTCTCCTTCATTACAGAA

ACGGCTTTTTCAAAAATATGGTATTGATAATCCTGATATGAATAAATT

GCAGTTTCATTTGATGCTCGATGAGTTTTTCTAATCAGAATTGGTTAA

TTGGTTGTAACACTGGCAGAGCATTACGCTGACTTGACGGGACGGCGG

CTTTGTTGAATAAATCGAACTTTTGCTGAGTTGAAGGATCAGATCACG

CATCTTCCCGACAACGCAGACCGTTCCGTGGCAAAGCAAAAGTTCAAA

ATCACCAACTGGTCCACCTACAACAAAGCTCTCATCAACCGTGGCTCC

CTCACTTTCTGGCTGGATGATGGGGCGATTCAGGCCTGGTATGAGTCA

GCAACACCTTCTTCACGAGGCAGACCTCAGCGCTCAAAGATGCAGGGG

TAAAAGCTAACCGCATCTTTACCGACAAGGCATCCGGCAGTTCAACAG

ATCGGGAAGGGCTGGATTTGCTGAGGATGAAGGTGGAGGAAGGTGATG

TCATTCTGGTGAAGAAGCTCGACCGTCTTGGCCGCGACACCGCCGACA

TGATCCAACTGATAAAAGAGTTTGATGCTCAGGGTGTAGCGGTTCGGT

TTATTGACGACGGGATCAGTACCGACGGTGATATGGGGCAAATGGTGG

TCACCATCCTGTCGGCTGTGGCACAGGCTGAACGCCGGAGGATCCTAG

AGCGCACGAATGAGGGCCGACAGGAAGCAAAGCTGAAAGGAATCAAAT

TTGGCCGCAGGCGTACCGTGGACAGGAACGTCGTGCTGACGCTTCATC

AGAAGGGCACTGGTGCAACGGAAATTGCTCATCAGCTCAGTATTGCCC

GCTCCACGGTTTATAAAATTCTTGAAGACGAAAGGGCCTCGTGATACG

CCTATTTTTATAGGTTAATGTCATGATAATAATGGTTTCTTAGACGTC

TTAATTAATCAGGAGAGCGTTCACCGACAAACAACAGATAAAACGAAA

GGCCCAGTCTTTCGACTGAGCCTTTCGTTTTATTTGATGCCTGGCAGT

TCCCTACTCTCGCATGGGGAGACCCCACACTACCATCGGCGCTACGGC

GTTTCACTTCTGAGTTCGGCATGGGGTCAGGTGGGACCACCGCGCTAC

TGCCGCCAGGCAAATTCTGTTTTATCAGACCGCTTCTGCGTTCTGATT

TAATCTGTATCAGGCTGAAAATCTTCTCTCATCCGCCAAAACAGCCAA

GCTGGAGACCGTTTAAACTCAATGATGATGATGATGATGGTCGACGGC

GCTATTCAGATCCTCTTCTGAGATGAGTTTTTGTTCGGGCCCAAGCTT

CGAATTCCCATATGGTACCAGCTGCAGATCTCGAGCTCGGATCCATGG

-continued

TTTATTCCTCCTTATTTAATCGATACATTAATATATACCTCTTTAATT

TTTAATAATAAAGTTAATCGATAATTCCGGTCGAGTGCCCACACAGAT

TGTCTGATAAATTGTTAAAGAGCAGTGCCGCTTCGCTTTTTCTCAGCG

GCGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACATT

ATACGAGCCGGATGATTAATTGTCAACAGCTCATTTCAGAATATTTGC

CAGAACCGTTATGATGTCGGCGCAAAAAACATTATCCAGAACGGGAGT

GCGCCTTGAGCGACACGAATTATGCAGTGATTTACGACCTGCACAGCC

ATACCACAGCTTCCGATGGCTGCCTGACGCCAGAAGCATTGGTGCACC

GTGCAGTCGATGATAAGCTGTCAAACCAGATCAATTCGCGCTAACTCA

CATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGT

CGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTT

TGCGTATTGGGCGCCAGGGTGGTTTTTCTTTTCACCAGTGAGACGGGC

AACAGCTGATTGCCCTTCACCGCCTGGCCCTGAGAGAGTTGCAGCAAG

CGGTCCACGCTGGTTTGCCCCAGCAGGCGAAAATCCTGTTTGATGGTG

GTTGACGGCGGGATATAACATGAGCTGTCTTCGGTATCGTCGTATCCC

ACTACCGAGATATCCGCACCAACGCGCAGCCCGGACTCGGTAATGGCG

CGCATTGCGCCCAGCGCCATCTGATCGTTGGCAACCAGCATCGCAGTG

GGAACGATGCCCTCATTCAGCATTTGCATGGTTTGTTGAAAACCGGAC

ATGGCACTCCAGTCGCCTTCCCGTTCCGCTATCGGCTGAATTTGATTG

CGAGTGAGATATTTATGCCAGCCAGCCAGACGCAGACGCGCCGAGACA

GAACTTAATGGGCCCGCTAACAGCGCGATTTGCTGGTGACCCAATGCG

ACCAGATGCTCCACGCCCAGTCGCGTACCGTCTTCATGGGAGAAAATA

ATACTGTTGATGGGTGTCTGGTCAGAGACATCAAGAAATAACGCCGGA

ACATTAGTGCAGGCAGCTTCCACAGCAATGGCATCCTGGTCATCCAGC

GGATAGTTAATGATCAGCCCACTGACGCGTTGCGCGAGAAGATTGTGC

ACCGCCGCTTTACAGGCTTCGACGCCGCTTCGTTCTACCATCGACACC

ACCACGCTGGCACCCAGTTGATCGGCGCGAGATTTAATCGCCGCGACA

ATTTGCGACGGCGCGTGCAGGGCCAGACTGGAGGTGGCAACGCCAATC

AGCAACGACTGTTTGCCCGCCAGTTGTTGTGCCACGCGGTTGGGAATG

TAATTCAGCTCCGCCATCGCCGCTTCCACTTTTTCCCGCGTTTTCGCA

GAAACGTGGCTGGCCTGGTTCACCACGCGGGAAACGGTCTGATAAGAG

ACACCGGCATACTCTGCGACATCGTATAACGTTACTGGTTTCACATTC

ACCACCCTGAATTGACTCTCTTCCGGGCGCTATCATGCCATACCGCGA

AAGGTTTTGCACCATTCGATGGTGTCAACGTAAATGCATGCCGCTTCG

CCTTCGCGCGCGAATTGATCTGCTGCCTCGCGCGTTTCGGTGATGACG

GTGAAAACCTCTGACACATGCAGCTCCCGGAGACGGTCACAGCTTGTC

TGTAAGCGGATGCCGGGAGCAGACAAGCCCGTCAGGGCGCGTCAGCGG

GTGTTGGCGGGGCCGGCCTCG

[0269] The 'tesA, and fadD genes were amplified using high fidelity Phusion™ polymerase (New England Biolabs, Inc., Ipswich, Mass.), with the following primers from their respective plasmids, pETDuet-1-'tesA and pHZ1.61:

'tesAForward-
(SEQ ID NO: 12)
5'-CTCTAGAAATAATTTAACTTTAAGTAGGAGAUAGGTACCCATGGC
GGACACGTTATTGAT-3'

'tesAReverse-
(SEQ ID NO: 13)
5'-CTTCGAATTCCATTTAAATTATTTCTAGAGTCATTATGAGTCATG
ATTTACTAAAGGC-3'

fadDForward-
(SEQ ID NO: 14)
5'-CTCTAGAAATAATTTTAGTTAAGTATAAGAAGGAGATATACCATG
GTGAAGAAGGTTTGGCTTAA-3'

fadDReverse-
(SEQ ID NO: 15)
5'-CTTCGAATTCCATTTAAATTATTTCTAGAGTTATCAGGCTTTATT
GTCCAC-3'

Insertion of 'tesA into pACYC-Ptrc Plasmid

[0270] Using NcoI and EcoRI sites on both the insert and vector, the 'tesA PCR product amplified from pETDuet-1-'tesA was cloned into the initial position of pACYC-Ptrc vector (SEQ ID NO:11). A T4 DNA ligase (New England Biolabs, Ipswich, Mass.) was then used to ligate the pACYC-Ptrc vector and 'tesA, producing a pACYC-Ptrc-'tesA plasmid. Following overnight ligation, the DNA product was transformed into TOP 10® One Shot® cells (Invitrogen, Carlsbad, Calif.). The insertion of 'tesA into the pACYC-Ptrc vector was confirmed by restriction digestion. An SwaI restriction site as well as overlapping fragments for In-Fusion™ cloning (Clontech, Mountain View, Calif.) were also created as a result at the 3'-end of the 'tesA insert.

Construction of pACYC-Ptrc-'tesA-fadD

[0271] The pACYC-Ptrc-'tesA plasmid was then subject to an overnight restriction digestion by SwaI. fadD amplified the plasmid pHZ1.61 (described above) was cloned downstream from the 'tesA gene using the In-Fusion™ PCR Cloning System (Clontech, Menlo Park, Calif.). The insertion of fadD was verified with restriction digestion. The insertion of fadD destroys the SwaI site following the 'tesA gene, but recreates a new SwaI site at the 3' end of fadD.

[0272] The pACYC-Ptrc-'tesA-fadD plasmid was used as a template to generate a Ptrc-'tesA-fadD cassette. The following primers were used to obtain the cassette:

IFF:
(SEQ ID NO: 16)
5'-GGGTCAATAGCGGCCGCCAATTCGCGCGCGAAGGCG-3'

IFR:
(SEQ ID NO: 17)
5'-TGGCGCGCCTCCTAGGGCATTACGCTGACTTGACGGG-3'

[0273] This cassette was subsequently cloned into the NotI and AvrII restriction sites of pGRG25 (described above) using the Infusion™ PCR cloning system (Clontech, Mountain View, Calif.), creating the Tn7tesfad plasmid (SEQ ID NO:18), wherein the $lacI_q$, Ptrc-'tesA-fadD genes were flanked by the left and right Tn7 ends.

[0274] The plasmid Tn7tesfad was electroporated into strain *E. coli* MG1655 DV2 (Example 2, above) using a protocol described by McKenzie et al., *BMC Microbiology* 6:39 (2006). After electroporation, Ampicillin-resistant cells were selected by growth in an LB medium containing 0.1% glucose and 100 μg/mL carbenicillin at 32° C. overnight. This was followed by selection of plasmids comprising the Tn7-transposition fractions, using the growth of cells on an LB plus 0.1% arabinose plates overnight at 32° C. Single colonies were selected and streaked onto new LB medium plates with and without Ampicillin, and they were grown overnight at 42° C. to cure of Tn7tesfad plasmid. Thus, the $lacI_q$, Ptrc-'tesA-fadD genes were integrated into the attTn7 site on the *E. coli* MG1655 chromosome located between the pstS and glmS genes. Integration of these genes was confirmed by PCR and sequencing using the following primers:

attTn7.A:
5'-GATGCTGGTGGCGAAGCTGT-3' (SEQ ID NO: 19)

attTn7.C:
5'-GTTGCGACGGTGGTACGCATAAC-3' (SEQ ID NO: 20)

[0275] The resulting strain was given the name *E. coli* MG1655 DAM1.

Example 5

Production of *E. coli* MG1655 DAM1/pDS57

[0276] A plasmid pDS57 (SEQ ID NO:41) was prepared as described below.

[0277] An ester synthase gene encoding an ester synthase ES9 from *Marinobacter hydrocarbonoclasticus* DSM8789 (GenBank Accession No. ABO21021) was synthesized by DNA2.0 (Menlo Park, Calif.). The synthesized gene was then cloned into a pCOLADuet-1 plasmid (EMD Chemicals, Inc., Gibbstown, N.J.) to form a pHZ1.97-ES9 construct. The internal BspHI restriction site of the ester synthase gene was then removed by site-directed mutagenesis, using the QuikChange™ Multi Kit (Stratagene, Carlsbad, Calif.) and the primer:

(SEQ ID NO: 21)
ES9BspF: 5'-CCCAGATCAGTTTTATGATTGCCTCGCTGG-3'

[0278] This primer introduced a silent mutation into the ester synthase gene. The resulting plasmid was called pDS32.

[0279] pDS32 was then used as a template to amplify the ester synthase gene using the following primers:

ES9BspH-Forward:
5'-ATCATGAAACGTCTCGGAAC-3' (SEQ ID NO: 22)

ES9Xho-Reverse:
5'-CCTCGAGTTACTTGCGGGTTCGGGCGCG-3' (SEQ ID NO: 23)

[0280] The PCR product was subject to restriction digestions with BspHI and XhoI. This digestion fragment was then ligated into a pDS23 plasmid (as described below) that had been digested with NcoI and XhoI, to form a plasmid pDS33. ES9 (SEQ ID NO:24).

Construction of pDS23

[0281] A Pspc promoter (SEQ ID NO:25) was obtained by PCR amplification, using Phusion™ Polymerase (New England Biolabs, Inc., Ipswich, Mass.) from *E. coli* MG1655 chromosomal DNA. The following primers were used:

```
PspcIFF:
                                              (SEQ ID NO: 26)
5'-AAAGGATGTCGCAAACGCTGTTTCAGTACACTCTCTCAATAC-3'

PspcIFR:
                                              (SEQ ID NO: 27)
5'-GAGCTCGGATCCATGGTTTAGTGCTCCGCTAATG-3'
```

[0282] The PCR fragment was then used to replace the $lacI_q$ and Ptrc promoter sequences of a plasmid OP-80 (SEQ ID NO:28), which was constructed as described below:

Construction of Plasmid OP-80:

[0283] A commercial vector pCL1920 (see, Lerner, et al., *Nucleic Acids Res.* 18:4631 (1990)), carrying a strong transcriptional promoter, was used as the starting point. The pCL1920 vector was digested with AflII and sfoI (New England Biolabs, Ipswich, Mass.). Three DNA fragments were produced, among which, a 3737-bp fragment was gel-purified using a gel-purification kit (Qiagen, Inc., Valencia, Calif.).

[0284] In parallel, a DNA fragment comprising the Ptrc promoter and the lacI sequences was obtained from a plasmid pTrcHis2 (Invitrogen, Carlsbad, Calif.) using the following primers:

```
                                              (SEQ ID NO: 29)
LF302: 5'-ATATGACGTCGGCATCCGCTTACAGACA-3'

                                              (SEQ ID NO: 30)
LF303: 5'-AATTCTTAAGTCAGGAGAGCGTTCACCGACAA-3'
```

[0285] These primers also introduced the restriction sites for ZraI and AflII.

[0286] The PCR product was purified using a PCR-purification kit (Qiagen, Inc., Valencia, Calif.) and digested with ZraI and AflII. The digestion product was gel-purified and ligated with the 3737-bp fragment (described above). The ligation mixture was then transformed into TOP 10® chemically competent cells (Invitrogen, Carlsbad, Calif.). The transformants were selected on Luria agar plates containing 100 μg/mL spectinomycin during overnight incubation. Resistant colonies were identified, and plasmids within these colonies were purified, and verified with restriction digestion and sequencing. One plasmid produced this way was retained, and given the name of OP-80 (SEQ ID NO:28).

[0287] The PCR fragment comprising the Pspc promoter (described above) was cloned into the BseRI and NcoI restriction sites of OP-80 using the InFusion™ Cloning Kit (Clontech, Menlo Park, Calif.). The resulting plasmid was given the name pDS22. pDS22 still possessed a lacZ gene sequence downstream of the multiple cloning site. The lacZ sequence was removed with PCR employing the following primers:

```
pCLlacDF:
5'-GAATTCCACCCGCTGACGAGCTTA-3'  (SEQ ID NO: 31)

pCLEcoR:
5'-CGAATTCCCATATGGTACCAG-3'     (SEQ ID NO: 32)
```

The PCR product was subject to restriction digestion by EcoRI. The digested product was subsequently self-ligated to form a plasmid named pDS23, which did not contain $lacI_q$, lacZ or promoter Ptrc sequence.

[0288] The plasmid pDS33.ES9 (SEQ ID NO:24) (described above) was again digested with BspHI and XhoI. After digestion, the fragment was ligated with an OP-80 plasmid (described above) that had been previously linearized using NcoI/XhoI restriction digestions.

[0289] The ligation product was transformed into TOP 10® One Shot chemically competent cells (Invitrogen, Carlsbad, Calif.). Cells were then plated on LB plates containing 100 μg/mL spectinomycin, and incubated overnight at 37° C. After overnight growth, several colonies were purified and the sequence of the inserts verified. The plasmid was given the name pDS57 (SEQ ID NO:41).

[0290] *E. coli* DAM1 strain was made electrocompetent using standard methods. The competent cells were then transformed with plasmid pDS57 and plated on LB plates containing 100 μg/mL of spectinomycin, and incubated overnight at 37° C. Resistant colonies were purified and the presence of the pDS57 plasmid was confirmed using restriction digestion and sequencing. The resulting construct was given the name *E. coli* DAM1/pDS57.

Example 6

Production of Biodiesel by Fermentation

[0291] This example demonstrates processes to produce a fatty ester composition using the genetically modified microorganisms described herein. A fermentation and recovery process was used to produce biodiesel of commercial grade quality by fermentation of carbohydrates. The fermentation process produced a mix of fatty acid methyl esters (FAME) and fatty acid ethyl esters (FAEE) for use as a biodiesel using the genetically engineered microorganisms described in Examples 1-5.

Fermentation

[0292] The *E. coli* MG1655 DAM1/pDS57 cells (see, Example 4, supra) was taken from a frozen stock and grown in a defined media consisting of: 4.54 g/L of K_2HPO_4 trihydrate, 4 g/L of $(NH_4)_2SO_4$, 0.15 g/L of $MgSO_4$ heptahydrate, 20 g/L of glucose, 200 mM of Bis-Tris buffer (pH 7.2), 1.25 mL/L of a first trace mineral solution and 1.25 mL/L of a vitamin solution. The first trace metals solution was composed of 27 g/L of $FeCl_3.6H_2O$, 2 g/L of $ZnCl_2.4H_2O$, 2 g/L of $CaCl_2.6H_2O$, 2 g/L of $Na_2MoO_4.2H_2O$, 1.9 g/L of $CuSO_4.5H_2O$, 0.5 g/L of H_3BO_3, and 100 mL/L of concentrated HCl. The trace vitamin solution was composed of 0.42 g/L of riboflavin, 5.4 g/L of pantothenic acid, 6 g/L of niacin, 1.4 g/L of pyridoxine, 0.06 g/L of biotin, and 0.04 g/L of folic acid.

[0293] 50 mL of cultures of DAM1/pDS57 cells were grown overnight and subsequently used to inoculate 1 L of fermentation medium containing: 0.5 g/L $(NH_4)_2SO_4$, 2.0 g/L KH_2PO_4, 0.15 g/L $MgSO_4$ heptahydrate, 0.034 g/L ferric citrate, 2.5 g/L bacto casamino acids, 10 g/L glucose, 1.25 mL/L of a second trace mineral solution and 1.25 mL/L of a vitamin solution, in a fermentor with temperature, pH, agitation, aeration and dissolved oxygen control. The second trace mineral solution contained 2 g/L of $ZnCl_2.4H_2O$, 2 g/L of $CaCl_2.6H_2O$, 2 g/L of $Na_2MoO_4.2H_2O$, 1.9 g/L of $CuSO_4.5H_2O$, 0.5 g/L of H_3BO_3, and 100 mL/L of concentrated HCl. The trace vitamin solution was as described above.

[0294] The feed to the fermentor contained 600 g/L glucose, 3.9 g/L $MgSO_4$ heptahydrate, 1.6 g/L KH_2PO_4, 2.5 g/L casamino acids, 0.05 g/L ferric citrate, 20 mL/L of the second trace mineral solution and 2 mL/L of the trace vitamin solution.

[0295] The preferred conditions for the fermentation were 32° C., pH 6.8 and dissolved oxygen (DO) equal to 30% of saturation. The pH was maintained by addition of NH_4OH, which also served as nitrogen source for cell growth. The glucose level in the culture was monitored using methods known to those skilled in the art. When the initial glucose was almost consumed, a feed consisting of 600 g/L glucose, 3.9 g/L $MgSO_4$ heptahydrate, 1.6 g/L KH_2PO_4, 2.5 g/L casamino acids, 0.05 g/L ferric citrate, 20 mL/L of the second trace mineral solution and 2 mL/L of the trace vitamin solution was supplied to the fermentor. The feed rate was set up to allow for a cell growth rate of 0.3 h^{-1}, up to a maximum of 10 g glucose/L/h, at which point it was fixed. This rate was maintained for the remaining duration of the fermentation as long as glucose did not accumulate in the fermentor. By avoiding glucose accumulation, it was possible to reduce or eliminate the formation of by-products such as acetate, formate and ethanol, which are commonly produced by *E. coli*. In the early phases of the growth, the production of fatty esters was induced by the addition of 1 mM IPTG and 20 mL/L of a pure methanol or a pure ethanol. The fermentation was continued for a period of 3 days. Methanol or ethanol was added several times during the run to replenish what was consumed by the cells for the production of fatty esters, but mostly to replace the alcohol lost by evaporation in the off-gas. The additions were targeted to maintain the concentration of methanol or ethanol in the fermentation broth between 10 and 30 mL/L, to allow efficient production while avoiding inhibition of cell growth.

[0296] The progression of the fermentation was followed by measurements of OD_{600} (optical density at 600 nm), glucose consumption, and ester production.

Analysis

[0297] Glucose consumption throughout the fermentation was analyzed by High Pressure Liquid Chromatography (HPLC). The HPLC analysis was performed according to methods commonly used for some sugars and organic acids in the art, using the following conditions: Agilent HPLC 1200 Series with Refractive Index detector; Column. Aminex HPX-87H, 300 mm×7.8 mm; column temperature: 350° C.; mobile phase: 0.01M H_2SO_4 (aqueous); flow rate: 0.6 mL/m; injection volume: 20 μL.

[0298] The production of fatty acid methyl and ethyl esters was analyzed by gas chromatography with flame ionization detector (GC-FID). Samples from fermentation broth were extracted with ethyl acetate in a ratio of 1:1 vol/vol. After strong vortexing, the samples were centrifuged, and the organic phase was analyzed by gas chromatography (GC). The analysis conditions were as follows: instrument: Trace GC Ultra, Thermo Electron Corporation with Flame ionization detector (FID) detector; column: DB-1 (1% diphenyl siloxane; 99% dimethyl siloxane) CO1 UFM 1/0.1/5 01 DET from Thermo Electron Corporation, phase pH 5, FT: 0.4 μm, length 5 m, id: 0.1 mm; inlet conditions: 250° C. splitless, 3.8 m 1/25 split method used depending upon sample concentration with split flow of 75 mL/m; carrier gas, flow rate: Helium, 3.0 mL/m; block temperature: 330° C.; oven temperature: 0.5 m hold at 50° C., 100° C./m to 330° C., 0.5 m hold at 330° C.; detector temperature: 300° C.; injection volume: 2 μL; run time/flow rate: 6.3 m/3.0 mL/m (splitless method), 3.8 m/1.5 mL/m (split 1/25 method), 3.04 m/1.2 mL/m (split 1/50 method).

Recovery

[0299] After fermentation, the fatty ester composition can be separated from the fermentation broth by several different methods well known in the art. After the completion of the fermentation, the broth was centrifuged to separate a first light phase containing the methylesters from a first heavy phase consisting of water, salts and the microbial biomass. The first light phase was centrifuged a second time to recover the biodiesel and to separate a second light phase (which consisted of a mixture of ethers) from a second heavy phase.

[0300] Centrifugation was performed in disk-stacked continuous centrifuges of pilot scale capacity (fixed centrifugal force about 10,000 g) with flows from about 1 to about 5 LPM. The same centrifuge was used each time for the first and second steps. Normal adjustments known in the art to centrifugation configuration and conditions (gravity ring size, back pressure in outlets, flow) were undertaken in each case to achieve the most favorable separation in terms of recovery efficiency and cleanness of the product. For the first centrifugation step, the fermentation broth was sent directly from the fermentor to the centrifuge without any physical or chemical adjustments.

[0301] In instances where it was more difficult to break the emulsion and obtain clear oil, additional pretreatments were applied to the light phase to help with the separation during the second centrifugation step. These treatments consisted of the following: heating to 60 to 80° C., adjusting the pH to 2.0 to 2.5 with sulfuric acid, and addition of demulsifiers (ARB-8285 (Baker Hughes, Houston, Tex.), less than 1% of the emulsion/light phase volume). The temperature was held for 1 to 2 h before the second centrifugation.

Polishing

[0302] The composition obtained from the harvesting step had characteristics that were very close to the regulatory standards for biodiesel. The inherent properties of this composition, as well as the properties related to purity, met the regulatory standards for biodiesel, including cetane number, kinematic viscosity, flash point, oxidation stability, copper corrosion, free and total glycerin, methanol, phosphorous, sulfate, K^+ and Na^+ content. However, the composition obtained from the harvesting step can, in certain embodiments, be subjected to optional minor further purification steps to eliminate other impurities.

[0303] Specifically, the following polishing steps were performed on the clarified composition: (1) a lime wash step; (2) an acid wash step; (3) a water wash step; and (4) an absorption treatment step. In the lime wash step, 1 part by volume of a lime slurry (5% lime slurry w/w) was mixed with 5 parts of raw oil at room temperature for 5 m. The mixture was then centrifuged at 10,000 g for 15 m. The lime-washed oil was then mixed with a dilute sulfuric acid (5% v/v) in the ratio of 1 part oil to 0.75 parts diluted sulfuric acid. The mixture was centrifuged at 10,000 g for 15 m. The acid-washed oil was then mixed water in the ratio of 1 part oil plus 1 part water. The mixture was then centrifuged at 10,000 g for 15 m. The oil was separated, and the water-washed oil was then dried at 80° C. for 1 h in the rotovap. This dried oil was heated to 90° C.,

and added to an absorbent, Magnesol™ D60 (The Dallas Group, Inc., Whitehouse, N.J.) (1% w/v), and mixed for 1 h. Magnesol™ D60 was removed by filtration through 0.2 micron filters.

Results

[0304] The *E. coli* MG1655 DAM1/pDS57 strain was grown according to the protocol above, in 2-L and 5-L fermentors. Representative results obtained are presented in the Table 7. The yield is expressed as the grams of product obtained per 100 g of carbon source used.

TABLE 7

Parameter	Methanol	Ethanol
FAME concentration (g/L)	44.8	—
FFA concentration (g/L)	0.6	4.9
FAEE Concentration (g/L)	1.1	8.1
Yield of Fatty Ester on glucose (%)	16.1	5.1
Yield of All Fatty Species on glucose (%)	16.7	8.1

Example 7

Performance Profile

[0305] This example illustrates the performance profile of the fatty ester composition produced using the genetically modified microorganism produced and processed in accordance with Example 6. The fatty ester composition was subject to the analyses as a diesel fuel, without blending with petroleum diesel. The analyses were performed by the School of Chemistry, Federal University of Rio De Janeiro, Brazil. The results of the analysis are set forth in Table 8.

TABLE 8

Properties	Methods	Results	ANP No. 7 Standard
Aspect	Visual	Lipid, without impurities (24° C.)	LII
Specific mass @20° C.	ASTM D4052	875.3 kg/m^3	850-900 kg/m^3
Kinematic Viscosity @40° C.	ASTM D445	3.55 mm^2/s	3-6 mm^2/s
Water content, max	EN 12937	450 mg/kg	≦500 mg/kg
Total contamination	EN 12662	3.2 mg/kg	≦24 mg/kg
Flash point, min	ASTM D93	128° C.	≧100° C.
Ester content, min	EN 14103	97.5 wt. %	≧96.5 wt. %
Carbon residue, max	ASTM D4530	N/D	≦0.05 wt. %
Sulfated ash, max	ASTM D874	0.01 wt. %	≦0.02 wt. %
Total sulfur, max	EN 20884	15.0 mg/kg	≦50 mg/kg
Na + K, max	EN 14108 EN 14109	1.6 mg/kg	≦5 mg/kg
Ca + Mg, max	NBR 15556	0.3 mg/kg	≦5 mg/kg
Phosphorus, max	ASTM D4951	0.9 mg/kg	≦10 mg/kg
Copper corrosion, 3 h @50° C., max	ASTM D130	1	1
Cold filter plugging point, max	ASTM D6371	−3° C.	≦19° C.
Acid value, max	ASTM D664	0.5 mg KOH/g	≦0.5 mg KOH/g
Free glycerol, max	ASTM D6584	0.02 wt. %	≦0.02 wt. %
Total glycerol, max	ASTM D6584	0.03 wt. %	≦0.25 wt. %
Monoacylglycerol	ASTM D6584	0.02 wt. %	Report
Diacylglycerol	ASTM D6584	0 wt. %	Report
Triacylglycerol	ASTM D6584	0 wt. %	Report
Methanol or Ethanol, max	EN 14110	0.01 wt. %	≦0.2 wt. %
Iodine value	EN 14111	64.92 g/100 g	Report
Oxidation stability @100° C., min	EN 14112	11.0 h	≧6 h

[0306] In parallel, a sample obtained from the same process was provided to Gorge Analytical in Hood River, Oreg. for testing and the results are listed below in Table 9:

TABLE 9

Analysis	Method	Result	Pass/Fail
Cloud Point	ASTM D2500	−4° C.	n/a
Simulated distillation - T90	ASTM D2887	358° C.	Pass
Sulfur by UVF	ASTM D5453	10.7 ppm	Pass
Karl Fischer Moisture-Coulometric	ASTM D6304	0.0542 wt. %	n/a
Total Acid Number	ASTM D664	0.14 mg KOH/g	Pass
Ca + Mg	EN 14538 - Ca, Mg	<2 ppm	Pass
Na + K	EN 14538 - Na, K	2.2 ppm	Pass

OTHER EMBODIMENTS

[0307] It is to be understood that while the invention has been described in conjunction with the detailed description thereof, the foregoing description is intended to illustrate and not limit the scope of the invention, which is defined by the scope of the appended claims. Other aspects, advantages, and modifications are within the scope of the following claims.

SEQUENCE LISTING

<160> NUMBER OF SEQ ID NOS: 42

<210> SEQ ID NO 1
<211> LENGTH: 70
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 1

aaaaacagca acaatgtgag ctttgttgta attatattgt aaacatattg attccgggga 60

tccgtcgacc 70

<210> SEQ ID NO 2
<211> LENGTH: 68
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 2

aaacggagcc tttcggctcc gttattcatt tacgcggctt caactttcct gtaggctgga 60

gctgcttc 68

<210> SEQ ID NO 3
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 3

cgggcaggtg ctatgaccag gac 23

<210> SEQ ID NO 4
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 4

cgcggcgttg accggcagcc tgg 23

<210> SEQ ID NO 5
<211> LENGTH: 70
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 5

atcattctcg tttacgttat cattcacttt acatcagaga tataccaatg attccgggga 60

tccgtcgacc 70

-continued

<210> SEQ ID NO 6
<211> LENGTH: 69
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 6

gcacggaaat ccgtgcccca aaagagaaat tagaaacgga aggttgcggt tgtaggctgg 60

agctgcttc 69

<210> SEQ ID NO 7
<211> LENGTH: 21
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 7

caacagcaac ctgctcagca a 21

<210> SEQ ID NO 8
<211> LENGTH: 21
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 8

aagctggagc agcaaagcgt t 21

<210> SEQ ID NO 9
<211> LENGTH: 5180
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 9

gggtcaatag cggccgccaa ttcgcgcgcg aaggcgaagc ggcatgcatt tacgttgaca 60

ccatcgaatg gtgcaaaacc tttcgcggta tggcatgata gcgcccggaa gagagtcaat 120

tcagggtggt gaatgtgaaa ccagtaacgt tatacgatgt cgcagagtat gccggtgtct 180

cttatcagac cgtttcccgc gtggtgaacc aggccagcca cgtttctgcg aaaacgcggg 240

aaaaagtgga agcggcgatg gcggagctga attacattcc caaccgcgtg gcacaacaac 300

tggcgggcaa acagtcgttg ctgattggcg ttgccacctc cagtctggcc ctgcacgcgc 360

cgtcgcaaat tgtcgcggcg attaaatctc gcgccgatca actgggtgcc agcgtggtgg 420

tgtcgatggt agaacgaagc ggcgtcgaag cctgtaaagc ggcggtgcac aatcttctcg 480

cgcaacgcgt cagtgggctg atcattaact atccgctgga tgaccaggat gccattgctg 540

tggaagctgc ctgcactaat gttccggcgt tatttcttga tgtctctgac cagacaccca 600

tcaacagtat tattttctcc catgaagacg gtacgcgact gggcgtggag catctggtcg 660

cattgggtca ccagcaaatc gcgctgttag cgggcccatt aagttctgtc tcggcgcgtc 720

tgcgtctggc tggctggcat aaatatctca ctcgcaatca aattcagccg atagcggaac 780

-continued

gggaaggcga ctggagtgcc atgtccggtt ttcaacaaac catgcaaatg ctgaatgagg 840

gcatcgttcc cactgcgatg ctggttgcca acgatcagat ggcgctgggc gcaatgcgcg 900

ccattaccga gtccgggctg cgcgttggtg cggatatctc ggtagtggga tacgacgata 960

ccgaagacag ctcatgttat atcccgccgt caaccaccat caaacaggat tttcgcctgc 1020

tggggcaaac cagcgtggac cgcttgctgc aactctctca gggccaggcg gtgaagggca 1080

atcagctgtt gcccgtctca ctggtgaaaa gaaaaaccac cctggcgccc aatacgcaaa 1140

ccgcctctcc ccgcgcgttg gccgattcat taatgcagct ggcacgacag gtttcccgac 1200

tggaaagcgg gcagtgagcg caacgcaatt aatgtgagtt agcgcgaatt gatctggttt 1260

gacagcttat catcgactgc acggtgcacc aatgcttctg gcgtcaggca gccatcggaa 1320

gctgtggtat ggctgtgcag gtcgtaaatc actgcataat tcgtgtcgct caaggcgcac 1380

tcccgttctg gataatgttt tttgcgccga catcataacg gttctggcaa atattctgaa 1440

atgagctgtt gacaattaat catccggctc gtataatgtg tggaattgtg agcggataac 1500

aatttcacac aggaaacagc gccgctgaga aaaagcgaag cggcactgct ctttaacaat 1560

ttatcagaca atctgtgtgg gcactcgacc ggaattatcg attaacttta ttattaaaaa 1620

ttaaagaggt atatattaat gtatcgatta aataaggagg aataaaccat ggcggacacg 1680

ttattgattc tgggtgatag cctgagcgcc gggtatcgaa tgtctgccag cgcggcctgg 1740

cctgccttgt tgaatgataa gtggcagagt aaaacgtcgg tagttaatgc cagcatcagc 1800

ggcgacacct cgcaacaagg actggcgcgc cttccggctc tgctgaaaca gcatcagccg 1860

cgttgggtgc tggttgaact gggcggcaat gacggtttgc gtggttttca gccacagcaa 1920

accgagcaaa cgctgcgcca gattttgcag gatgtcaaag ccgccaacgc tgaaccattg 1980

ttaatgcaaa tacgtctgcc tgcaaactat ggtcgccgtt ataatgaagc ctttagcgcc 2040

atttacccca aactcgccaa agagtttgat gttccgctgc tgccctttt tatggaagag 2100

gtctacctca agccacaatg gatgcaggat gacggtattc atcccaaccg cgacgcccag 2160

ccgtttattg ccgactggat ggcgaagcag ttgcagcctt tagtaaatca tgactcataa 2220

tgactctaga aataatttta gttaagtata agaaggagat ataccatggt gaagaaggtt 2280

tggcttaacc gttatcccgc ggacgttccg acggagatca accctgaccg ttatcaatct 2340

ctggtagata tgtttgagca gtcggtcgcg cgctacgccg atcaacctgc gtttgtgaat 2400

atgggggagg taatgacctt ccgcaagctg gaagaacgca gtcgcgcgtt tgccgcttat 2460

ttgcaacaag ggttggggct gaagaaaggc gatcgcgttg cgttgatgat gcctaattta 2520

ttgcaatatc cggtggcgct gtttggcatt ttgcgtgccg ggatgatcgt cgtaaacgtt 2580

aacccgttgt ataccccgcg tgagcttgag catcagctta acgatagcgg cgcatcggcg 2640

attgttatcg tgtctaactt tgctcacaca ctggaaaaag tggttgataa aaccgccgtt 2700

cagcacgtaa ttctgacccg tatgggcgat cagctatcta cggcaaaagg cacggtagtc 2760

aatttcgttg ttaaatacat caagcgtttg gtgccgaaat accatctgcc agatgccatt 2820

tcatttcgta gcgcactgca taacggctac cggatgcagt acgtcaaacc cgaactggtg 2880

ccggaagatt tagcttttct gcaatacacc ggcggcacca ctggtgtggc gaaaggcgcg 2940

atgctgactc accgcaatat gctggcgaac ctggaacagg ttaacgcgac ctatggtccg 3000

ctgttgcatc cgggcaaaga gctggtggtg acggcgctgc cgctgtatca catttttgcc 3060

-continued

ctgaccatta actgcctgct gtttatcgaa ctgggtgggc agaacctgct tatcactaac 3120

ccgcgcgata ttccagggtt ggtaaaagag ttagcgaaat atccgtttac cgctatcacg 3180

ggcgttaaca ccttgttcaa tgcgttgctg aacaataaag agttccagca gctggatttc 3240

tccagtctgc atctttccgc aggcggaggg atgccagtgc agcaagtggt ggcagagcgt 3300

tgggtgaaac tgacaggaca gtatctgctg gaaggctatg gccttaccga gtgtgcgccg 3360

ctggtcagcg ttaacccata tgatattgat tatcatagtg gtagcatcgg tttgccggtg 3420

ccgtcgacgg aagccaaact ggtggatgat gatgataatg aagtaccacc gggtcaaccg 3480

ggtgagcttt gtgtcaaagg accgcaggtg atgctgggtt actggcagcg tccggatgct 3540

acagatgaga tcatcaaaaa tggctggtta cacaccggcg acatcgcggt gatggatgaa 3600

gaagggttcc tgcgcattgt cgatcgtaaa aaagacatga ttctggtttc cggttttaac 3660

gtctatccca acgagattga agatgtcgtc atgcagcatc ctggcgtaca ggaagtcgcg 3720

gctgttggcg taccttccgg ctccagtggt gaagcggtga aaatcttcgt agtgaaaaaa 3780

gatccatcgc ttaccgaaga gtcactggtg acctttttgcc gccgtcagct cacgggctac 3840

aaagtaccga agctggtgga gtttcgtgat gagttaccga aatctaacgt cggaaaaatt 3900

ttgcgacgag aattacgtga cgaagcgcgc ggcaaagtgg acaataaagc ctgataactc 3960

tagaaataat ttaaatggaa ttcgaagctt gggcccgaac aaaaactcat ctcagaagag 4020

gatctgaata gcgccgtcga ccatcatcat catcatcatt gagtttaaac ggtctccagc 4080

ttggctgttt tggcggatga gagaagattt tcagcctgat acagattaaa tcagaacgca 4140

gaagcggtct gataaaacag aatttgcctg gcggcagtag cgcggtggtc ccacctgacc 4200

ccatgccgaa ctcagaagtg aaacgccgta gcgccgatgg tagtgtgggg tctccccatg 4260

cgagagtagg gaactgccag gcatcaaata aaacgaaagg ctcagtcgaa agactgggcc 4320

tttcgtttta tctgttgttt gtcggtgaac gctctcctga ttaattaaga cgtctaagaa 4380

accattatta tcatgacatt aacctataaa aataggcgta tcacgaggcc ctttcgtctt 4440

caagaatttt ataaaccgtg gagcgggcaa tactgagctg atgagcaatt tccgttgcac 4500

cagtgccctt ctgatgaagc gtcagcacga cgttcctgtc cacggtacgc ctgcggccaa 4560

atttgattcc tttcagcttt gcttcctgtc ggccctcatt cgtgcgctct aggatcctcc 4620

ggcgttcagc ctgtgccaca gccgacagga tggtgaccac catttgcccc atatcaccgt 4680

cggtactgat cccgtcgtca ataaaccgaa ccgctacacc ctgagcatca aactctttta 4740

tcagttggat catgtcggcg gtgtcgcggc caagacggtc gagcttcttc accagaatga 4800

catcaccttc ctccaccttc atcctcagca aatccagccc ttcccgatct gttgaactgc 4860

cggatgcctt gtcggtaaag atgcggttag cttttacccc tgcatctttg agcgctgagg 4920

tctgcctcgt gaagaaggtg ttgctgactc ataccaggcc tgaatcgccc catcatccag 4980

ccagaaagtg agggagccac ggttgatgag agctttgttg taggtggacc agttggtgat 5040

tttgaacttt tgctttgcca cggaacggtc tgcgttgtcg ggaagatgcg tgatctgatc 5100

cttcaactca gcaaaagttc gatttattca acaaagccgc cgtcccgtca agtcagcgta 5160

atgccctagg aggcgcgcca 5180

<210> SEQ ID NO 10
<211> LENGTH: 6700

-continued

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

<400> SEQUENCE: 10

ggggaattgt gagcggataa caattcccct gtagaaataa ttttgtttaa ctttaataag 60

gagatatacc atggtgaaga aggtttggct taaccgttat cccgcggacg ttccgacgga 120

gatcaaccct gaccgttatc aatctctggt agatatgttt gagcagtcgg tcgcgcgcta 180

cgccgatcaa cctgcgtttg tgaatatggg ggaggtaatg accttccgca agctggaaga 240

acgcagtcgc gcgtttgccg cttatttgca acaagggttg gggctgaaga aaggcgatcg 300

cgttgcgttg atgatgccta atttattgca atatccggtg gcgctgtttg gcattttgcg 360

tgccgggatg atcgtcgtaa acgttaaccc gttgtatacc ccgcgtgagc ttgagcatca 420

gcttaacgat agcggcgcat cggcgattgt tatcgtgtct aactttgctc acacactgga 480

aaaagtggtt gataaaaccg ccgttcagca cgtaattctg acccgtatgg gcgatcagct 540

atctacggca aaaggcacgg tagtcaattt cgttgttaaa tacatcaagc gtttggtgcc 600

gaaataccat ctgccagatg ccatttcatt tcgtagcgca ctgcataacg gctaccggat 660

gcagtacgtc aaacccgaac tggtgccgga agatttagct tttctgcaat acaccggcgg 720

caccactggt gtggcgaaag gcgcgatgct gactcaccgc aatatgctgg cgaacctgga 780

acaggttaac gcgacctatg gtccgctgtt gcatccgggc aaagagctgg tggtgacggc 840

gctgccgctg tatcacattt ttgccctgac cattaactgc ctgctgttta tcgaactggg 900

tgggcagaac ctgcttatca ctaacccgcg cgatattcca gggttggtaa aagagttagc 960

gaaatatccg tttaccgcta tcacgggcgt taacaccttg ttcaatgcgt tgctgaacaa 1020

taaagagttc cagcagctgg atttctccag tctgcatctt tccgcaggcg gagggatgcc 1080

agtgcagcaa gtggtggcag agcgttgggt gaaactgaca ggacagtatc tgctggaagg 1140

ctatggcctt accgagtgtg cgccgctggt cagcgttaac ccatatgata ttgattatca 1200

tagtggtagc atcggtttgc cggtgccgtc gacggaagcc aaactggtgg atgatgatga 1260

taatgaagta ccaccgggtc aaccgggtga gctttgtgtc aaaggaccgc aggtgatgct 1320

gggttactgg cagcgtccgg atgctacaga tgagatcatc aaaaatggct ggttacacac 1380

cggcgacatc gcggtgatgg atgaagaagg gttcctgcgc attgtcgatc gtaaaaaaga 1440

catgattctg gtttccggtt ttaacgtcta tcccaacgag attgaagatg tcgtcatgca 1500

gcatcctggc gtacaggaag tcgcggctgt tggcgtacct tccggctcca gtggtgaagc 1560

ggtgaaaatc ttcgtagtga aaaaagatcc atcgcttacc gaagagtcac tggtgacctt 1620

ttgccgccgt cagctcacgg gctacaaagt accgaagctg gtggagtttc gtgatgagtt 1680

accgaaatct aacgtcggaa aaattttgcg acgagaatta cgtgacgaag cgcgcggcaa 1740

agtggacaat aaagcctgaa agcttgcggc cgcataatgc ttaagtcgaa cagaaagtaa 1800

tcgtattgta cacggccgca taatcgaaat taatacgact cactataggg gaattgtgag 1860

cggataacaa ttccccatct tagtatatta gttaagtata agaaggagat atacatatgc 1920

gcccattaca tccgattgat tttatattcc tgtcactaga aaaaagacaa cagcctatgc 1980

atgtaggtgg tttatttttg tttcagattc ctgataacgc cccagacacc tttattcaag 2040

-continued

atctggtgaa tgatatccgg atatcaaaat caatccctgt tccaccattc aacaataaac 2100

tgaatgggct tttttgggat gaagatgaag agtttgattt agatcatcat tttcgtcata 2160

ttgcactgcc tcatcctggt cgtattcgtg aattgcttat ttatatttca caagagcaca 2220

gtacgctgct agatcgggca aagcccttgt ggacctgcaa tattattgaa ggaattgaag 2280

gcaatcgttt tgccatgtac ttcaaaattc accatgcgat ggtcgatggc gttgctggta 2340

tgcggttaat tgaaaaatca ctctcccatg atgtaacaga aaaaagtatc gtgccacctt 2400

ggtgtgttga gggaaaacgt gcaaagcgct taagagaacc taaaacaggt aaaattaaga 2460

aaatcatgtc tggtattaag agtcagcttc aggcgacacc cacagtcatt caagagcttt 2520

ctcagacagt atttaaagat attggacgta atcctgatca tgtttcaagc tttcaggcgc 2580

cttgttctat tttgaatcag cgtgtgagct catcgcgacg ttttgcagca cagtcttttg 2640

acctagatcg ttttcgtaat attgccaaat cgttgaatgt gaccattaat gatgttgtac 2700

tagcggtatg ttctggtgca ttacgtgcgt atttgatgag tcataatagt ttgccttcaa 2760

aaccattaat tgccatggtt ccagcctcta ttcgcaatga cgattcagat gtcagcaacc 2820

gtattacgat gattctggca aatttggcaa cccacaaaga tgatccttta caacgtcttg 2880

aaattatccg ccgtagtgtt caaaactcaa agcaacgctt caaacgtatg accagcgatc 2940

agattctaaa ttatagtgct gtcgtatatg gccctgcagg actcaacata atttctggca 3000

tgatgccaaa acgccaagcc ttcaatctgg ttatttccaa tgtgcctggc ccaagagagc 3060

cactttactg gaatggtgcc aaacttgatg cactctaccc agcttcaatt gtattagacg 3120

gtcaagcatt gaatattaca atgaccagtt atttagataa acttgaagtt ggtttgattg 3180

catgccgtaa tgcattgcca agaatgcaga atttactgac acatttagaa gaagaaattc 3240

aactatttga aggcgtaatt gcaaagcagg aagatattaa aacagccaat taaaaacaat 3300

aaacttgatt ttttaattta tcagataaaa ctaaagggct aaattagccc tcctaggctg 3360

ctgccaccgc tgagcaataa ctagcataac cccttggggc ctctaaacgg gtcttgaggg 3420

gttttttgct gaaacctcag gcatttgaga agcacacggt cacactgctt ccggtagtca 3480

ataaaccggt aaaccagcaa tagacataag cggctattta acgaccctgc cctgaaccga 3540

cgaccgggtc atcgtggccg gatcttgcgg cccctcggct tgaacgaatt gttagacatt 3600

atttgccgac taccttggtg atctcgcctt tcacgtagtg gacaaattct tccaactgat 3660

ctgcgcgcga ggccaagcga tcttcttctt gtccaagata agcctgtcta gcttcaagta 3720

tgacgggctg atactgggcc ggcaggcgct ccattgccca gtcggcagcg acatccttcg 3780

gcgcgatttt gccggttact gcgctgtacc aaatgcggga caacgtaagc actacatttc 3840

gctcatcgcc agcccagtcg ggcggcgagt tccatagcgt taaggtttca tttagcgcct 3900

caaatagatc ctgttcagga accggatcaa agagttcctc cgccgctgga cctaccaagg 3960

caacgctatg ttctcttgct tttgtcagca agatagccag atcaatgtcg atcgtggctg 4020

gctcgaagat acctgcaaga atgtcattgc gctgccattc tccaaattgc agttcgcgct 4080

tagctggata acgccacgga atgatgtcgt cgtgcacaac aatggtgact tctacagcgc 4140

ggagaatctc gctctctcca ggggaagccg aagtttccaa aaggtcgttg atcaaagctc 4200

gccgcgttgt ttcatcaagc cttacggtca ccgtaaccag caaatcaata tcactgtgtg 4260

gcttcaggcc gccatccact gcggagccgt acaaatgtac ggccagcaac gtcggttcga 4320

-continued

gatggcgctc gatgacgcca actacctctg atagttgagt cgatacttcg gcgatcaccg 4380

cttccctcat actcttcctt tttcaatatt attgaagcat ttatcagggt tattgtctca 4440

tgagcggata catatttgaa tgtatttaga aaaataaaca aatagctagc tcactcggtc 4500

gctacgctcc gggcgtgaga ctgcggcggg cgctgcggac acatacaaag ttacccacag 4560

attccgtgga taagcagggg actaacatgt gaggcaaaac agcagggccg cgccggtggc 4620

gtttttccat aggctccgcc ctcctgccag agttcacata aacagacgct tttccggtgc 4680

atctgtggga gccgtgaggc tcaaccatga atctgacagt acgggcgaaa cccgacagga 4740

cttaaagatc cccaccgttt ccggcgggtc gctccctctt gcgctctcct gttccgaccc 4800

tgccgtttac cggatacctg ttccgccttt ctcccttacg ggaagtgtgg cgctttctca 4860

tagctcacac actggtatct cggctcggtg taggtcgttc gctccaagct gggctgtaag 4920

caagaactcc ccgttcagcc cgactgctgc gccttatccg gtaactgttc acttgagtcc 4980

aacccggaaa agcacggtaa aacgccactg gcagcagcca ttggtaactg ggagttcgca 5040

gaggatttgt ttagctaaac acgcggttgc tcttgaagtg tgcgccaaag tccggctaca 5100

ctggaaggac agatttggtt gctgtgctct gcgaaagcca gttaccacgg ttaagcagtt 5160

ccccaactga cttaaccttc gatcaaacca cctccccagg tggttttttc gtttacaggg 5220

caaaagatta cgcgcagaaa aaaaggatct caagaagatc ctttgatctt ttctactgaa 5280

ccgctctaga tttcagtgca atttatctct tcaaatgtag cacctgaagt cagccccata 5340

cgatataagt tgtaattctc atgttagtca tgccccgcgc ccaccggaag gagctgactg 5400

ggttgaaggc tctcaagggc atcggtcgag atcccggtgc ctaatgagtg agctaactta 5460

cattaattgc gttgcgctca ctgcccgctt tccagtcggg aaacctgtcg tgccagctgc 5520

attaatgaat cggccaacgc gcggggagag gcggtttgcg tattgggcgc cagggtggtt 5580

tttcttttca ccagtgagac gggcaacagc tgattgccct tcaccgcctg gccctgagag 5640

agttgcagca agcggtccac gctggtttgc cccagcaggc gaaaatcctg tttgatggtg 5700

gttaacggcg ggatataaca tgagctgtct tcggtatcgt cgtatcccac taccgagatg 5760

tccgcaccaa cgcgcagccc ggactcggta atggcgcgca ttgcgcccag cgccatctga 5820

tcgttggcaa ccagcatcgc agtgggaacg atgccctcat tcagcatttg catggtttgt 5880

tgaaaaccgg acatggcact ccagtcgcct tcccgttccg ctatcggctg aatttgattg 5940

cgagtgagat atttatgcca gccagccaga cgcagacgcg ccgagacaga acttaatggg 6000

cccgctaaca gcgcgatttg ctggtgaccc aatgcgacca gatgctccac gcccagtcgc 6060

gtaccgtctt catgggagaa aataatactg ttgatgggtg tctggtcaga gacatcaaga 6120

aataacgccg gaacattagt gcaggcagct tccacagcaa tggcatcctg gtcatccagc 6180

ggatagttaa tgatcagccc actgacgcgt tgcgcgagaa gattgtgcac cgccgcttta 6240

caggcttcga cgccgcttcg ttctaccatc gacaccacca cgctggcacc cagttgatcg 6300

gcgcgagatt taatcgccgc gacaatttgc gacggcgcgt gcagggccag actggaggtg 6360

gcaacgccaa tcagcaacga ctgtttgccc gccagttgtt gtgccacgcg gttgggaatg 6420

taattcagct ccgccatcgc cgcttccact ttttcccgcg ttttcgcaga aacgtggctg 6480

gcctggttca ccacgcggga aacggtctga taagagacac cggcatactc tgcgacatcg 6540

tataacgtta ctggtttcac attcaccacc ctgaattgac tctcttccgg gcgctatcat 6600

-continued

gccataccgc gaaaggtttt gcgccattcg atggtgtccg ggatctcgac gctctccctt 6660

atgcgactcc tgcattagga aattaatacg actcactata 6700

<210> SEQ ID NO 11
<211> LENGTH: 5733
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 11

actcaccagt cacagaaaag catcttacgg atggcatgac agtaagagaa ttatgcagtg 60

ctgccataac catgagtgat aacactgcgg ccaacttact tctgacaacg atcggaggac 120

cgaaggagct aaccgctttt ttgcacaaca tgggggatca tgtaactcgc cttgatcgtt 180

gggaaccgga gctgaatgaa gccataccaa acgacgagcg tgacaccacg atgcctgcag 240

caatggcaac aacgttgcgc aaactattaa ctggcgaact acttactcta gcttcccggc 300

aacaattaat agactggatg gaggcggata aagttgcagg accacttctg cgctcggccc 360

ttccggctgg ctggtttatt gctgataaat ctggagccgg tgagcgtggg tctcgcggta 420

tcattgcagc actggggcca gatggtaagc cctcccgtat cgtagttatc tacacgacgg 480

ggagtcaggc aactatggat gaacgaaata gacagatcgc tgagataggt gcctcactga 540

ttaagcattg gtaactgtca gaccaagttt actcatatat actttagatt gatttaaaac 600

ttcattttta atttaaaagg atctaggtga agatcctttt tgataatctc atgaccaaaa 660

tcccttaacg tgagttttcg ttccactgag cgtcagaccc cttaataaga tgatcttctt 720

gagatcgttt tggtctgcgc gtaatctctt gctctgaaaa cgaaaaaacc gccttgcagg 780

gcggtttttc gaaggttctc tgagctacca actctttgaa ccgaggtaac tggcttggag 840

gagcgcagtc accaaaactt gtcctttcag tttagcctta accggcgcat gacttcaaga 900

ctaactcctc taaatcaatt accagtggct gctgccagtg gtgcttttgc atgtctttcc 960

gggttggact caagacgata gttaccggat aaggcgcagc ggtcggactg aacgggggt 1020

tcgtgcatac agtccagctt ggagcgaact gcctacccgg aactgagtgt caggcgtgga 1080

atgagacaaa cgcggccata acagcggaat gacaccggta aaccgaaagg caggaacagg 1140

agagcgcacg agggagccgc caggggaaa cgcctggtat ctttatagtc ctgtcgggtt 1200

tcgccaccac tgatttgagc gtcagatttc gtgatgcttg tcaggggggc ggagcctatg 1260

gaaaaacggc tttgccgcgg ccctctcact tccctgttaa gtatcttcct ggcatcttcc 1320

aggaaatctc cgccccgttc gtaagccatt tccgctcgcc gcagtcgaac gaccgagcgt 1380

agcgagtcag tgagcgagga agcggaatat atcctgtatc acatattctg ctgacgcacc 1440

ggtgcagcct tttttctcct gccacatgaa gcacttcact gacaccctca tcagtgccaa 1500

catagtaagc cagtatacac tccgctagcg ctgaggtctg cctcgtgaag aaggtgttgc 1560

tgactcatac caggcctgaa tcgccccatc atccagccag aaagtgaggg agccacggtt 1620

gatgagagct ttgttgtagg tggaccagtt ggtgattttg aacttttgct ttgccacgga 1680

acggtctgcg ttgtcgggaa gatgcgtgat ctgatccttc aactcagcaa aagttcgatt 1740

tattcaacaa agccacgttg tgtctcaaaa tctctgatgt tacattgcac aagataaaaa 1800

tatatcatca tgaacaataa aactgtctgc ttacataaac agtaatacaa ggggtgttat 1860

-continued

gagccatatt caacgggaaa cgtcttgctc gaggccgcga ttaaattcca acatggatgc 1920

tgatttatat gggtataaat gggctcgcga taatgtcggg caatcaggtg cgacaatcta 1980

tcgattgtat gggaagcccg atgcgccaga gttgtttctg aaacatggca aaggtagcgt 2040

tgccaatgat gttacagatg agatggtcag actaaactgg ctgacggaat ttatgcctct 2100

tccgaccatc aagcatttta tccgtactcc tgatgatgca tggttactca ccactgcgat 2160

ccccgggaaa acagcattcc aggtattaga agaatatcct gattcaggtg aaaatattgt 2220

tgatgcgctg gcagtgttcc tgcgccggtt gcattcgatt cctgtttgta attgtccttt 2280

taacagcgat cgcgtatttc gtctcgctca ggcgcaatca cgaatgaata acggtttggt 2340

tgatgcgagt gattttgatg acgagcgtaa tggctggcct gttgaacaag tctggaaaga 2400

aatgcataag cttttgccat tctcaccgga ttcagtcgtc actcatggtg atttctcact 2460

tgataacctt atttttgacg aggggaaatt aataggttgt attgatgttg gacgagtcgg 2520

aatcgcagac cgataccagg atcttgccat cctatggaac tgcctcggtg agttttctcc 2580

ttcattacag aaacggcttt ttcaaaaata tggtattgat aatcctgata tgaataaatt 2640

gcagtttcat ttgatgctcg atgagttttt ctaatcagaa ttggttaatt ggttgtaaca 2700

ctggcagagc attacgctga cttgacggga cggcggcttt gttgaataaa tcgaactttt 2760

gctgagttga aggatcagat cacgcatctt cccgacaacg cagaccgttc cgtggcaaag 2820

caaaagttca aaatcaccaa ctggtccacc tacaacaaag ctctcatcaa ccgtggctcc 2880

ctcactttct ggctggatga tggggcgatt caggcctggt atgagtcagc aacaccttct 2940

tcacgaggca gacctcagcg ctcaaagatg caggggtaaa agctaaccgc atctttaccg 3000

acaaggcatc cggcagttca acagatcggg aagggctgga tttgctgagg atgaaggtgg 3060

aggaaggtga tgtcattctg gtgaagaagc tcgaccgtct tggccgcgac accgccgaca 3120

tgatccaact gataaaagag tttgatgctc agggtgtagc ggttcggttt attgacgacg 3180

ggatcagtac cgacggtgat atggggcaaa tggtggtcac catcctgtcg gctgtggcac 3240

aggctgaacg ccggaggatc ctagagcgca cgaatgaggg ccgacaggaa gcaaagctga 3300

aaggaatcaa atttggccgc aggcgtaccg tggacaggaa cgtcgtgctg acgcttcatc 3360

agaagggcac tggtgcaacg gaaattgctc atcagctcag tattgcccgc tccacggttt 3420

ataaaattct tgaagacgaa agggcctcgt gatacgccta tttttatagg ttaatgtcat 3480

gataataatg gtttcttaga cgtcttaatt aatcaggaga gcgttcaccg acaaacaaca 3540

gataaaacga aaggcccagt ctttcgactg agcctttcgt tttatttgat gcctggcagt 3600

tccctactct cgcatgggga gaccccacac taccatcggc gctacggcgt ttcacttctg 3660

agttcggcat ggggtcaggt gggaccaccg cgctactgcc gccaggcaaa ttctgtttta 3720

tcagaccgct tctgcgttct gatttaatct gtatcaggct gaaaatcttc tctcatccgc 3780

caaaacagcc aagctggaga ccgtttaaac tcaatgatga tgatgatgat ggtcgacggc 3840

gctattcaga tcctcttctg agatgagttt ttgttcgggc ccaagcttcg aattcccata 3900

tggtaccagc tgcagatctc gagctcggat ccatggttta ttcctcctta tttaatcgat 3960

acattaatat atacctcttt aatttttaat aataaagtta atcgataatt ccggtcgagt 4020

gcccacacag attgtctgat aaattgttaa agagcagtgc cgcttcgctt tttctcagcg 4080

gcgctgtttc ctgtgtgaaa ttgttatccg ctcacaattc cacacattat acgagccgga 4140

-continued

tgattaattg tcaacagctc atttcagaat atttgccaga accgttatga tgtcggcgca 4200

aaaaacatta tccagaacgg gagtgcgcct tgagcgacac gaattatgca gtgatttacg 4260

acctgcacag ccataccaca gcttccgatg gctgcctgac gccagaagca ttggtgcacc 4320

gtgcagtcga tgataagctg tcaaaccaga tcaattcgcg ctaactcaca ttaattgcgt 4380

tgcgctcact gcccgctttc cagtcgggaa acctgtcgtg ccagctgcat taatgaatcg 4440

gccaacgcgc ggggagaggc ggtttgcgta ttgggcgcca gggtggtttt tcttttcacc 4500

agtgagacgg gcaacagctg attgcccttc accgcctggc cctgagagag ttgcagcaag 4560

cggtccacgc tggtttgccc cagcaggcga aaatcctgtt tgatggtggt tgacggcggg 4620

atataacatg agctgtcttc ggtatcgtcg tatcccacta ccgagatatc cgcaccaacg 4680

cgcagcccgg actcggtaat ggcgcgcatt gcgcccagcg ccatctgatc gttggcaacc 4740

agcatcgcag tgggaacgat gccctcattc agcatttgca tggtttgttg aaaaccggac 4800

atggcactcc agtcgccttc ccgttccgct atcggctgaa tttgattgcg agtgagatat 4860

ttatgccagc cagccagacg cagacgcgcc gagacagaac ttaatgggcc cgctaacagc 4920

gcgatttgct ggtgacccaa tgcgaccaga tgctccacgc ccagtcgcgt accgtcttca 4980

tgggagaaaa taatactgtt gatgggtgtc tggtcagaga catcaagaaa taacgccgga 5040

acattagtgc aggcagcttc cacagcaatg gcatcctggt catccagcgg atagttaatg 5100

atcagcccac tgacgcgttg cgcgagaaga ttgtgcaccg ccgctttaca ggcttcgacg 5160

ccgcttcgtt ctaccatcga caccaccacg ctggcaccca gttgatcggc gcgagattta 5220

atcgccgcga caatttgcga cggcgcgtgc agggccagac tggaggtggc aacgccaatc 5280

agcaacgact gtttgcccgc cagttgttgt gccacgcggt tgggaatgta attcagctcc 5340

gccatcgccg cttccacttt ttcccgcgtt ttcgcagaaa cgtggctggc ctggttcacc 5400

acgcgggaaa cggtctgata agagacaccg gcatactctg cgacatcgta taacgttact 5460

ggtttcacat tcaccaccct gaattgactc tcttccgggc gctatcatgc cataccgcga 5520

aaggttttgc accattcgat ggtgtcaacg taaatgcatg ccgcttcgcc ttcgcgcgcg 5580

aattgatctg ctgcctcgcg cgtttcggtg atgacggtga aaacctctga cacatgcagc 5640

tcccggagac ggtcacagct tgtctgtaag cggatgccgg gagcagacaa gcccgtcagg 5700

gcgcgtcagc gggtgttggc ggggccggcc tcg 5733

<210> SEQ ID NO 12
<211> LENGTH: 60
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer
<220> FEATURE:
<223> OTHER INFORMATION: Description of Combined DNA/RNA Molecule: Synthetic primer

<400> SEQUENCE: 12

ctctagaaat aatttaactt taagtaggag auaggtaccc atggcggaca cgttattgat 60

<210> SEQ ID NO 13
<211> LENGTH: 58
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence

-continued

<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 13

cttcgaattc catttaaatt atttctagag tcattatgag tcatgattta ctaaaggc 58

<210> SEQ ID NO 14
<211> LENGTH: 65
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 14

ctctagaaat aattttagtt aagtataaga aggagatata ccatggtgaa gaaggtttgg 60

cttaa 65

<210> SEQ ID NO 15
<211> LENGTH: 51
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 15

cttcgaattc catttaaatt atttctagag ttatcaggct ttattgtcca c 51

<210> SEQ ID NO 16
<211> LENGTH: 36
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 16

gggtcaatag cggccgccaa ttcgcgcgcg aaggcg 36

<210> SEQ ID NO 17
<211> LENGTH: 37
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 17

tggcgcgcct cctagggcat tacgctgact tgacggg 37

<210> SEQ ID NO 18
<211> LENGTH: 17683
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 18

ggccacgatg cgtccggcgt agaggatctg ctcatgtttg acagcttatc atcgatgcat 60

aatgtgcctg tcaaatggac gaagcaggga ttctgcaaac cctatgctac tccgtcaagc 120

-continued

cgtcaattgt ctgattcgtt accaattatg acaacttgac ggctacatca ttcacttttt 180

cttcacaacc ggcacggaac tcgctcgggc tggccccggt gcatttttta aatacccgcg 240

agaaatagag ttgatcgtca aaaccaacat tgcgaccgac ggtggcgata ggcatccggg 300

tggtgctcaa aagcagcttc gcctggctga tacgttggtc ctcgcgccag cttaagacgc 360

taatccctaa ctgctggcgg aaaagatgtg acagacgcga cggcgacaag caaacatgct 420

gtgcgacgct ggcgatatca aaattgctgt ctgccaggtg atcgctgatg tactgacaag 480

cctcgcgtac ccgattatcc atcggtggat ggagcgactc gttaatcgct tccatgcgcc 540

gcagtaacaa ttgctcaagc agatttatcg ccagcagctc cgaatagcgc ccttcccctt 600

gcccggcgtt aatgatttgc ccaaacaggt cgctgaaatg cggctggtgc gcttcatccg 660

ggcgaaagaa ccccgtattg gcaaatattg acggccagtt aagccattca tgccagtagg 720

cgcgcggacg aaagtaaacc cactggtgat accattcgcg agcctccgga tgacgaccgt 780

agtgatgaat ctctcctggc gggaacagca aaatatcacc cggtcggcaa acaaattctc 840

gtccctgatt tttcaccacc ccctgaccgc gaatggtgag attgagaata taacctttca 900

ttcccagcgg tcggtcgata aaaaaatcga gataaccgtt ggcctcaatc ggcgttaaac 960

ccgccaccag atgggcatta aacgagtatc ccggcagcag gggatcattt tgcgcttcag 1020

ccatactttt catactcccg ccattcagag aagaaaccaa ttgtccatat tgcatcagac 1080

attgccgtca ctgcgtcttt tactggctct tctcgctaac caaaccggta accccgctta 1140

ttaaaagcat tctgtaacaa agcgggacca aagccatgac aaaaacgcgt aacaaaagtg 1200

tctataatca cggcagaaaa gtccacattg attatttgca cggcgtcaca ctttgctatg 1260

ccatagcatt tttatccata agattagcgg atcctacctg acgcttttta tcgcaactct 1320

ctactgtttc tccatacccg ttttttgggg ctagcgaatt cgagctcggt acccaagtct 1380

taaactagac agaatagttg taaactgaaa tcagtccagt tatgctgtga aaaagcatac 1440

tggacttttg ttatggctaa agcaaactct tcattttctg aagtgcaaat tgcccgtcgt 1500

attaaagagg ggcgtggcca agggcatggt aaagactata ttccatggct aacagtacaa 1560

gaagttcctt cttcaggtcg ttcccaccgt atttattctc ataagacggg acgagtccat 1620

catttgctat ctgacttaga gcttgctgtt tttctcagtc ttgagtggga gagcagcgtg 1680

ctagatatac gcgagcagtt ccccttatta cctagtgata ccaggcagat tgcaatagat 1740

agtggtatta agcatcctgt tattcgtggt gtagatcagg ttatgtctac tgatttttta 1800

gtggactgca aagatggtcc ttttgagcag tttgctattc aagtcaaacc tgcagcagcc 1860

ttacaagacg agcgtacctt agaaaaacta gaactagagc gtcgctattg gcagcaaaag 1920

caaattcctt ggttcatttt tactgataaa gaaataaatc ccgtagtaaa agaaaatatt 1980

gaatggcttt attcagtgaa aacagaagaa gtttctgcgg agcttttagc acaactatcc 2040

ccattggccc atatcctgca agaaaaagga gatgaaaaca ttatcaatgt ctgtaagcag 2100

gttgatattg cttatgattt ggagttaggc aaaacattga gtgagatacg agccttaacc 2160

gcaaatggtt ttattaagtt caatatttat aagtctttca gggcaaataa gtgtgcagat 2220

ctctgtatta gccaagtagt gaatatggag gagttgcgct atgtggcaaa ttaatgaggt 2280

tgtgctattt gataatgatc cgtatcgcat tttggctata gaggatggcc aagttgtctg 2340

gatgcaaata agcgctgata aaggagttcc acaagctagg gctgagttgt tgctaatgca 2400

-continued

gtatttagat gaaggccgct tagttagaac tgatgaccct tatgtacatc ttgatttaga 2460

agagccgtct gtagattctg tcagcttcca gaagcgcgag gaggattatc gaaaaattct 2520

tcctattatt aatagtaagg atcgtttcga ccctaaagtc agaagcgaac tcgttgagca 2580

tgtggtccaa gaacataagg ttactaaggc tacagtttat aagttgttac gccgttactg 2640

gcagcgtggt caaacgccta atgcattaat tcctgactac aaaaacagcg gtgcaccagg 2700

ggaaagacgt tcagcgacag gaacagcaaa gattggccga gccagagaat atggtaaggg 2760

tgaaggaacc aaggtaacgc ccgagattga acgccttttt aggttgacca tagaaaagca 2820

cctgttaaat caaaaaggta caaagaccac cgttgcctat agacgatttg tggacttgtt 2880

tgctcagtat tttcctcgca ttccccaaga ggattaccca acactacgtc agtttcgtta 2940

tttttatgat cgagaatacc ctaaagctca gcgcttaaag tctagagtta aagcaggggt 3000

atataaaaaa gacgtacgac ccttaagtag tacagccact tctcaggcgt taggccctgg 3060

gagtcgttat gagattgatg ccacgattgc tgatatttat ttagtggatc atcatgatcg 3120

ccaaaaaatc ataggaagac caacgcttta cattgtgatt gatgtgttta gtcggatgat 3180

cacgggcttt tatatcggct ttgaaaatcc gtcttatgtg gtggcgatgc aggcttttgt 3240

aaatgcttgc tctgacaaaa cggccatttg tgcccagcat gatattgaga ttagtagctc 3300

agactggccg tgtgtaggtt tgccagatgt gttgctagcg gaccgtggcg aattaatgag 3360

tcatcaggtc gaagccttag tttctagttt taatgtgcga gtggaaagtg ctccacctag 3420

acgtggcgat gctaaaggca tagtggaaag cacttttaga acactacaag ccgagtttaa 3480

gtcctttgca cctggcattg tagagggcag tcggatcaaa agccatggtg aaacagacta 3540

taggttagat gcatctctgt cggtatttga gttcacacaa attattttgc gtacgatctt 3600

attcagaaat aaccatctgg tgatggataa atacgatcga gatgctgatt ttcctacaga 3660

tttaccgtct attcctgtcc agctatggca atggggtatg cagcatcgta caggtagttt 3720

aagggctgtg gagcaagagc agttgcgagt agcgttactg cctcgccgaa aggtctctat 3780

ttcttcattt ggcgttaatt tgtggggttt gtattactcg gggtcagaga ttctgcgtga 3840

gggttggttg cagcggagca ctgatatagc tagacctcaa catttagaag cggcttatga 3900

cccagtgctg gttgatacga tttatttgtt tccgcaagtt ggcagccgtg tattttggcg 3960

ctgtaatctg acggaacgta gtcggcagtt taaaggtctc tcattttggg aggtttggga 4020

tatacaagca caagaaaaac acaataaagc caatgcgaag caggatgagt taactaaacg 4080

cagggagctt gaggcgttta ttcagcaaac cattcagaaa gcgaataagt taacgcccag 4140

tactactgag cccaaatcaa cacgcattaa gcagattaaa actaataaaa aagaagccgt 4200

gacctcggag cgtaaaaaac gtgcggagca tttgaagcca agctcttcag gtgatgaggc 4260

taaagttatt cctttcaacg cagtggaagc ggatgatcaa gaagattaca gcctacccac 4320

atacgtgcct gaattatttc aggatccacc agaaaaggat gagtcatgag tgctacccgg 4380

attcaagcag tttatcgtga tacgggggta gaggcttatc gtgataatcc ttttatcgag 4440

gccttaccac cattacaaga gtcagtgaat agtgctgcat cactgaaatc ctctttacag 4500

cttacttcct ctgacttgca aaagtcccgt gttatcagag ctcataccat ttgtcgtatt 4560

ccagatgact attttcagcc attaggtacg catttgctac taagtgagcg tatttcggtc 4620

atgattcgag gtggctacgt aggcagaaat cctaaaacag gagatttaca aaagcattta 4680

-continued

caaaatggtt atgagcgtgt tcaaacggga gagttggaga catttcgctt tgaggaggca 4740

cgatctacgg cacaaagctt attgttaatt ggttgttctg gtagtgggaa gacgacctct 4800

cttcatcgta ttctagccac gtatcctcag gtgatttacc atcgtgaact caatgtagag 4860

caggtggtgt atttgaaaat agactgctcg cataatggtt cgctaaaaga aatctgcttg 4920

aatttttca gagcgttgga tcgagccttg ggctcgaact atgagcgtcg ttatggctta 4980

aaacgtcatg gtatagaaac catgttggct ttgatgtcgc aaatagccaa tgcacatgct 5040

ttagggttgt tggttattga tgaaattcag catttaagcc gctctcgttc gggtggatct 5100

caagagatgc tgaacttttt tgtgacgatg gtgaatatta ttggcgtacc agtgatgttg 5160

attggtaccc ctaaagcacg agagattttt gaggctgatt tgcggtctgc acgtagaggg 5220

gcagggtttg gagctatatt ctgggatcct atacaacaaa cgcaacgtgg aaagcccaat 5280

caagagtgga tcgcttttac ggataatctt tggcaattac agcttttaca acgcaaagat 5340

gcgctgttat cggatgaggt ccgtgatgtg tggtatgagc taagccaagg agtgatggac 5400

attgtagtaa aacttttttgt actcgctcag ctccgtgcgc tagctttagg caatgagcgt 5460

attaccgctg gtttattgcg gcaagtgtat caagatgagt taaagcctgt gcaccccatg 5520

ctagaggcat tacgctcggg tatcccagaa cgcattgctc gttattctga tctagtcgtt 5580

cccgagattg ataaacggtt aatccaactt cagctagata tcgcagcgat acaagaacaa 5640

acaccagaag aaaaagccct tcaagagtta gataccgaag atcagcgtca tttatatctg 5700

atgctgaaag aggattacga ttcaagcctg ttaattccca ctattaaaaa agcgtttagc 5760

cagaatccaa cgatgacaag acaaaagtta ctgcctcttg ttttgcagtg gttgatggaa 5820

ggcgaaacgg tagtgtcaga actagaaaag ccctccaaga gtaaaaaggt ttcggctata 5880

aaggtagtca agcccagcga ctgggatagc ttgcctgata cggatttacg ttatatctat 5940

tcacaacgcc aacctgaaaa aaccatgcat gaacggttaa aagggaaagg ggtaatagtg 6000

gatatggcga gcttatttaa acaagcaggt tagccatgag aaactttcct gttccgtact 6060

cgaatgagct gatttatagc actattgcac gggcaggcgt ttatcaaggg attgttagtc 6120

ctaagcagct gttggatgag gtgtatggca accgcaaggt ggtcgctacc ttaggtctgc 6180

cctcgcattt aggtgtgata gcaagacatc tacatcaaac aggacgttac gctgttcagc 6240

agcttattta tgagcatacc ttattccctt tatatgctcc gtttgtaggc aaggagcgcc 6300

gagacgaagc tattcggtta atggagtacc aagcgcaagg tgcggtgcat ttaatgctag 6360

gagtcgctgc ttctagagtt aagagcgata accgctttag atactgccct gattgcgttg 6420

ctcttcagct aaataggtat ggggaagcct tttggcaacg agattggtat ttgcccgctt 6480

tgccatattg tccaaaacac ggtgctttag tcttctttga tagagctgta gatgatcacc 6540

gacatcaatt ttgggctttg ggtcatactg agctgctttc agactacccc aaagactccc 6600

tatctcaatt aacagcacta gctgcttata tagcccctct gttagatgct ccacgagcgc 6660

aagagctttc cccaagcctt gagcagtgga cgctgtttta tcagcgctta gcgcaggatc 6720

tagggctaac caaaagcaag cacattcgtc atgacttggt ggcggagaga gtgaggcaga 6780

cttttagtga tgaggcacta gagaaactgg atttaaagtt ggcagagaac aaggacacgt 6840

gttggctgaa aagtatattc cgtaagcata gaaaagcctt tagttattta cagcatagta 6900

ttgtgtggca agccttattg ccaaaactaa cggttataga agcgctacag caggcaagtg 6960

-continued

ctcttactga gcactctata acgacaagac ctgttagcca gtctgtgcaa cctaactctg 7020

aagatttatc tgttaagcat aaagactggc agcaactagt gcataaatac caaggaatta 7080

aggcggcaag acagtcttta gagggtgggg tgctatacgc ttggctttac cgacatgaca 7140

gggattggct agttcactgg aatcaacagc atcaacaaga gcgtctggca cccgccccta 7200

gagttgattg gaaccaaaga gatcgaattg ctgtacgaca actattaaga atcataaagc 7260

gtctagatag tagccttgat cacccaagag cgacatcgag ctggctgtta aagcaaactc 7320

ctaacggaac ctctcttgca aaaaatctac agaaactgcc tttggtagcg ctttgcttaa 7380

agcgttactc agagagtgtg gaagattatc aaattagacg gattagccaa gcttttatta 7440

agcttaaaca ggaagatgtt gagcttaggc gctggcgatt attaagaagt gcaacgttat 7500

ctaaagagcg gataactgag gaagcacaaa gattcttgga aatggtttat ggggaagagt 7560

gagtggttag gctagctaca tttaatgaca atgtgcaggt tgtacatatt ggtcatttat 7620

tccgtaactc gggtcataag gagtggcgta tttttgtttg gtttaatcca atgcaagaac 7680

ggaaatggac tcgatttact catttgcctt tattaagtcg agctaaggtg gttaacagta 7740

caacaaagca aataaataag gcggatcgtg tgattgagtt tgaagcatcg gatcttcaac 7800

gagccaaaat aatcgatttt cctaatctct cgtcctttgc ttccgtacgc aacaaggatg 7860

gagcgcagag ttcatttatt tacgaagctg aaacaccata tagcaagact cgttatcaca 7920

tcccacagtt agagctagct cggtcattat ttttagggga tcctctagag tcgacctgca 7980

ggcatgcaag cttggctgtt ttggcggatg agagaagatt ttcagcctga tacagattaa 8040

atcagaacgc agaagcggtc tgataaaaca gaatttgcct ggcggcagta gcgcggtggt 8100

cccacctgac cccatgccga actcagaagt gaaacgccgt agcgccgatg gtagtgtggg 8160

gtctccccat gcgagagtag ggaactgcca ggcatcaaat aaaacgaaag gctcagtcga 8220

aagactgggc ctttcgtttt atctgttgtt tgtcggtgaa cgctctcctg agtaggacaa 8280

atccgccggg agcggatttg aacgttgcga agcaacggcc cggagggtgg cgggcaggac 8340

gcccgccata aactgccagg catcaaatta agcagaaggc catcctgacg gatggccttt 8400

ttgcgtttct acaaactctt ttgtttattt ttctaaatac attcaaatat gtatccgctc 8460

atgagacaat aaccctgata aatgcttcaa taatattgaa aaaggaagag tatgagtatt 8520

caacatttcc gtgtcgccct tattcccttt tttgcggcat tttgccttcc tgtttttgct 8580

cacccagaaa cgctggtgaa agtaaaagat gctgaagatc agttgggtgc acgagtgggt 8640

tacatcgaac tggatctcaa cagcggtaag atccttgaga gttttcgccc cgaagaacgt 8700

tttccaatga tgagcacttt taaagttctg ctatgtggcg cggtattatc ccgtgttgac 8760

gccgggcaag agcaactcgg tcgccgcata cactattctc agaatgactt ggttgagtac 8820

tcaccagtca cagaaaagca tcttacggat ggcatgacag taagagaatt atgcagtgct 8880

gccataacca tgagtgataa cactgcggcc aacttacttc tgacaacgat cggaggaccg 8940

aaggagctaa ccgctttttt gcacaacatg ggggatcatg taactcgcct tgatcgttgg 9000

gaaccggagc tgaatgaagc cataccaaac gacgagcgtg acaccacgat gcctgcagca 9060

atggcaacaa cgttgcgcaa actattaact ggcgaactac ttactctagc ttcccggcaa 9120

caattaatag actggatgga ggcggataaa gttgcaggac cacttctgcg ctcggccctt 9180

ccggctggct ggtttattgc tgataaatct ggagccggtg agcgtgggtc tcgcggtatc 9240

-continued

attgcagcac tggggccaga tggtaagccc tcccgtatcg tagttatcta cacgacgggg 9300

agtcaggcaa ctatggatga acgaaataga cagatcgctg agataggtgc ctcactgatt 9360

aagcattggt aactgtcaga ccaagtttac tcatatatac tttagattga tttacgcgcc 9420

ctgtagcggc gcattaagcg cggcgggtgt ggtggttacg cgcagcgtga ccgctacact 9480

tgccagcgcc ctagcgcccg ctcctttcgc tttcttccct tcctttctcg ccacgttcgc 9540

cgccggccag cctcgcagag caggattccc gttgagcacc gccaggtgcg aataagggac 9600

agtgaagaag gaacacccgc tcgcgggtgg gcctacttca cctatcctgc ccggcggcat 9660

caccggcgcc acaggtgcgg ttgctggcgc ctatatcgcc gacatcaccg atggggaaga 9720

tcgggctcgc cacttcgggc tcatgagcgc ttgtttcggc gtgggtatgg tggcaggccc 9780

cgtggccggg ggactgttgg gcgccatctc cttgcatgca ccattccttg cggcggcggt 9840

gctcaacggc ctcaacctac tactgggctg cttcctaatg caggagtcgc ataagggaga 9900

gcgtcgatcc ccgacagtaa gacgggtaag cctgttgatg ataccgctgc cttactgggt 9960

gcattagcca gtctgaatga cctgtcacgg gataatccga agtggtcaga ctggaaaatc 10020

agagggcagg aactgctgaa cagcaaaaag tcagatagca ccacatagca gacccgccat 10080

aaaacgccct gagaagcccg tgacgggctt ttcttgtatt atgggtagtt tccttgcatg 10140

aatccataaa aggcgcctgt agtgccattt acccccattc actgccagag ccgtgagcgc 10200

agcgaactga atgtcacgaa aaagacagcg actcaggtgc ctgatggtcg gagacaaaag 10260

gaatattcag cgatttgccc gagcttgcga gggtgctact taagccttta gggttttaag 10320

gtctgttttg tagaggagca aacagcgttt gcgacatcct tttgtaatac tgcggaactg 10380

actaaagtag tgagttatac acagggctgg gatctattct ttttatcttt ttttattctt 10440

tctttattct ataaattata accacttgaa tataaacaaa aaaaacacac aaaggtctag 10500

cggaatttac agagggtcta gcagaattta caagttttcc agcaaaggtc tagcagaatt 10560

tacagatacc cacaactcaa aggaaaagga ctagtaatta tcattgacta gcccatctca 10620

attggtatag tgattaaaat cacctagacc aattgagatg tatgtctgaa ttagttgttt 10680

tcaaagcaaa tgaactagcg attagtcgct atgacttaac ggagcatgaa accaagctaa 10740

ttttatgctg tgtggcacta ctcaacccca cgattgaaaa ccctacaagg aaagaacgga 10800

cggtatcgtt cacttataac caatacgttc agatgatgaa catcagtagg gaaaatgctt 10860

atggtgtatt agctaaagca accagagagc tgatgacgag aactgtggaa atcaggaatc 10920

ctttggttaa aggctttgag attttccagt ggacaaacta tgccaagttc tcaagcgaaa 10980

aattagaatt agtttttagt gaagagatat tgccttatct tttccagtta aaaaaattca 11040

taaaatataa tctggaacat gttaagtctt ttgaaaacaa atactctatg aggatttatg 11100

agtggttatt aaaagaacta acacaaaaga aaactcacaa ggcaaatata gagattagcc 11160

ttgatgaatt taagttcatg ttaatgcttg aaaataacta ccatgagttt aaaaggctta 11220

accaatgggt tttgaaacca ataagtaaag atttaaacac ttacagcaat atgaaattgg 11280

tggttgataa gcgaggccgc ccgactgata cgttgatttt ccaagttgaa ctagatagac 11340

aaatggatct cgtaaccgaa cttgagaaca accagataaa aatgaatggt gacaaaatac 11400

caacaaccat tacatcagat tcctacctac ataacggact aagaaaaaca ctacacgatg 11460

ctttaactgc aaaaattcag ctcaccagtt ttgaggcaaa atttttgagt gacatgcaaa 11520

-continued

gtaagtatga tctcaatggt tcgttctcat ggctcacgca aaaacaacga accacactag 11580

agaacatact ggctaaatac ggaaggatct gaggttctta tggctcttgt atctatcagt 11640

gaagcatcaa gactaacaaa caaaagtaga acaactgttc accgttacat atcaaaggga 11700

aaactgtcca tatgcacaga tgaaaacggt gtaaaaaaga tagatacatc agagctttta 11760

cgagtttttg gtgcatttaa agctgttcac catgaacaga tcgacaatgt aacagatgaa 11820

cagcatgtaa cacctaatag aacaggtgaa accagtaaaa caaagcaact agaacatgaa 11880

attgaacacc tgagacaact tgttacagct caacagtcac acatagacag cctgaaacag 11940

gcgatgctgc ttatcgaatc aaagctgccg acaacacggg agccagtgac gcctcccgtg 12000

gggaaaaaat catggcaatt ctggaagaaa tagcgctttc agcctgtggg cggacaaaat 12060

agttgggaac tgggaggggt ggaaatggag tttttaagga ttatttaggg aagagtgaca 12120

aaatagatgg gaactgggtg tagcgtcgta agctaatacg aaaattaaaa atgacaaaat 12180

agtttggaac tagatttcac ttatctggtt ggtcgacact agtattaccc tgttatccct 12240

agatttaaat gatatcggat cctagtaagc cacgttttaa ttaatcagat gggtcaatag 12300

cggccgccaa ttcgcgcgcg aaggcgaagc ggcatgcatt tacgttgaca ccatcgaatg 12360

gtgcaaaacc tttcgcggta tggcatgata gcgcccggaa gagagtcaat tcagggtggt 12420

gaatgtgaaa ccagtaacgt tatacgatgt cgcagagtat gccggtgtct cttatcagac 12480

cgtttcccgc gtggtgaacc aggccagcca cgtttctgcg aaaacgcggg aaaaagtgga 12540

agcggcgatg gcggagctga attacattcc caaccgcgtg gcacaacaac tggcgggcaa 12600

acagtcgttg ctgattggcg ttgccacctc cagtctggcc ctgcacgcgc cgtcgcaaat 12660

tgtcgcggcg attaaatctc gcgccgatca actgggtgcc agcgtggtgg tgtcgatggt 12720

agaacgaagc ggcgtcgaag cctgtaaagc ggcggtgcac aatcttctcg cgcaacgcgt 12780

cagtgggctg atcattaact atccgctgga tgaccaggat gccattgctg tggaagctgc 12840

ctgcactaat gttccggcgt tatttcttga tgtctctgac cagacaccca tcaacagtat 12900

tattttctcc catgaagacg gtacgcgact gggcgtggag catctggtcg cattgggtca 12960

ccagcaaatc gcgctgttag cgggcccatt aagttctgtc tcggcgcgtc tgcgtctggc 13020

tggctggcat aaatatctca ctcgcaatca aattcagccg atagcggaac gggaaggcga 13080

ctggagtgcc atgtccggtt ttcaacaaac catgcaaatg ctgaatgagg gcatcgttcc 13140

cactgcgatg ctggttgcca acgatcagat ggcgctgggc gcaatgcgcg ccattaccga 13200

gtccgggctg cgcgttggtg cggatatctc ggtagtggga tacgacgata ccgaagacag 13260

ctcatgttat atcccgccgt caaccaccat caaacaggat tttcgcctgc tggggcaaac 13320

cagcgtggac cgcttgctgc aactctctca gggccaggcg gtgaagggca atcagctgtt 13380

gcccgtctca ctggtgaaaa gaaaaaccac cctggcgccc aatacgcaaa ccgcctctcc 13440

ccgcgcgttg gccgattcat taatgcagct ggcacgacag gtttcccgac tggaaagcgg 13500

gcagtgagcg caacgcaatt aatgtgagtt agcgcgaatt gatctggttt gacagcttat 13560

catcgactgc acggtgcacc aatgcttctg gcgtcaggca gccatcggaa gctgtggtat 13620

ggctgtgcag gtcgtaaatc actgcataat tcgtgtcgct caaggcgcac tcccgttctg 13680

gataatgttt tttgcgccga catcataacg gttctggcaa atattctgaa atgagctgtt 13740

gacaattaat catccggctc gtataatgtg tggaattgtg agcggataac aatttcacac 13800

-continued

aggaaacagc gccgctgaga aaaagcgaag cggcactgct ctttaacaat ttatcagaca 13860

atctgtgtgg gcactcgacc ggaattatcg attaacttta ttattaaaaa ttaaagaggt 13920

atatattaat gtatcgatta aataaggagg aataaaccat ggcggacacg ttattgattc 13980

tgggtgatag cctgagcgcc gggtatcgaa tgtctgccag cgcggcctgg cctgccttgt 14040

tgaatgataa gtggcagagt aaaacgtcgg tagttaatgc cagcatcagc ggcgacacct 14100

cgcaacaagg actggcgcgc cttccggctc tgctgaaaca gcatcagccg cgttgggtgc 14160

tggttgaact gggcggcaat gacggtttgc gtggttttca gccacagcaa accgagcaaa 14220

cgctgcgcca gattttgcag gatgtcaaag ccgccaacgc tgaaccattg ttaatgcaaa 14280

tacgtctgcc tgcaaactat ggtcgccgtt ataatgaagc ctttagcgcc atttacccca 14340

aactcgccaa agagtttgat gttccgctgc tgccctttt tatggaagag gtctacctca 14400

agccacaatg gatgcaggat gacggtattc atcccaaccg cgacgcccag ccgtttattg 14460

ccgactggat ggcgaagcag ttgcagcctt tagtaaatca tgactcataa tgactctaga 14520

aataatttta gttaagtata agaaggagat ataccatggt gaagaaggtt tggcttaacc 14580

gttatcccgc ggacgttccg acggagatca accctgaccg ttatcaatct ctggtagata 14640

tgtttgagca gtcggtcgcg cgctacgccg atcaacctgc gtttgtgaat atgggggagg 14700

taatgacctt ccgcaagctg gaagaacgca gtcgcgcgtt tgccgcttat ttgcaacaag 14760

ggttggggct gaagaaaggc gatcgcgttg cgttgatgat gcctaattta ttgcaatatc 14820

cggtggcgct gtttggcatt ttgcgtgccg ggatgatcgt cgtaaacgtt aacccgttgt 14880

ataccccgcg tgagcttgag catcagctta acgatagcgg cgcatcggcg attgttatcg 14940

tgtctaactt tgctcacaca ctggaaaaag tggttgataa aaccgccgtt cagcacgtaa 15000

ttctgacccg tatgggcgat cagctatcta cggcaaaagg cacggtagtc aatttcgttg 15060

ttaaatacat caagcgtttg gtgccgaaat accatctgcc agatgccatt tcatttcgta 15120

gcgcactgca taacggctac cggatgcagt acgtcaaacc cgaactggtg ccggaagatt 15180

tagcttttct gcaatacacc ggcggcacca ctggtgtggc gaaaggcgcg atgctgactc 15240

accgcaatat gctggcgaac ctggaacagg ttaacgcgac ctatggtccg ctgttgcatc 15300

cgggcaaaga gctggtggtg acggcgctgc cgctgtatca catttttgcc ctgaccatta 15360

actgcctgct gtttatcgaa ctgggtgggc agaacctgct tatcactaac ccgcgcgata 15420

ttccagggtt ggtaaaagag ttagcgaaat atccgtttac cgctatcacg ggcgttaaca 15480

ccttgttcaa tgcgttgctg aacaataaag agttccagca gctggatttc tccagtctgc 15540

atctttccgc aggcggaggg atgccagtgc agcaagtggt ggcagagcgt tgggtgaaac 15600

tgacaggaca gtatctgctg gaaggctatg gccttaccga gtgtgcgccg ctggtcagcg 15660

ttaacccata tgatattgat tatcatagtg gtagcatcgg tttgccggtg ccgtcgacgg 15720

aagccaaact ggtggatgat gatgataatg aagtaccacc gggtcaaccg ggtgagcttt 15780

gtgtcaaagg accgcaggtg atgctgggtt actggcagcg tccggatgct acagatgaga 15840

tcatcaaaaa tggctggtta cacaccggcg acatcgcggt gatggatgaa gaagggttcc 15900

tgcgcattgt cgatcgtaaa aaagacatga ttctggtttc cggttttaac gtctatccca 15960

acgagattga agatgtcgtc atgcagcatc ctggcgtaca ggaagtcgcg gctgttggcg 16020

taccttccgg ctccagtggt gaagcggtga aaatcttcgt agtgaaaaaa gatccatcgc 16080

-continued

ttaccgaaga gtcactggtg acctttgcc gccgtcagct cacgggctac aaagtaccga 16140

agctggtgga gtttcgtgat gagttaccga aatctaacgt cggaaaaatt ttgcgacgag 16200

aattacgtga cgaagcgcgc ggcaaagtgg acaataaagc ctgataactc tagaaataat 16260

ttaaatggaa ttcgaagctt gggcccgaac aaaaactcat ctcagaagag gatctgaata 16320

gcgccgtcga ccatcatcat catcatcatt gagtttaaac ggtctccagc ttggctgttt 16380

tggcggatga gagaagattt tcagcctgat acagattaaa tcagaacgca gaagcggtct 16440

gataaaacag aatttgcctg gcggcagtag cgcggtggtc ccacctgacc ccatgccgaa 16500

ctcagaagtg aaacgccgta gcgccgatgg tagtgtgggg tctccccatg cgagagtagg 16560

gaactgccag gcatcaaata aaacgaaagg ctcagtcgaa agactgggcc tttcgtttta 16620

tctgttgttt gtcggtgaac gctctcctga ttaattaaga cgtctaagaa accattatta 16680

tcatgacatt aacctataaa aataggcgta tcacgaggcc ctttcgtctt caagaatttt 16740

ataaaccgtg gagcgggcaa tactgagctg atgagcaatt tccgttgcac cagtgccctt 16800

ctgatgaagc gtcagcacga cgttcctgtc cacggtacgc ctgcggccaa atttgattcc 16860

tttcagcttt gcttcctgtc ggccctcatt cgtgcgctct aggatcctcc ggcgttcagc 16920

ctgtgccaca gccgacagga tggtgaccac catttgcccc atatcaccgt cggtactgat 16980

cccgtcgtca ataaaccgaa ccgctacacc ctgagcatca aactctttta tcagttggat 17040

catgtcggcg gtgtcgcggc caagacggtc gagcttcttc accagaatga catcaccttc 17100

ctccaccttc atcctcagca aatccagccc ttcccgatct gttgaactgc cggatgcctt 17160

gtcggtaaag atgcggttag cttttacccc tgcatctttg agcgctgagg tctgcctcgt 17220

gaagaaggtg ttgctgactc ataccaggcc tgaatcgccc catcatccag ccagaaagtg 17280

agggagccac ggttgatgag agctttgttg taggtggacc agttggtgat tttgaacttt 17340

tgctttgcca cggaacggtc tgcgttgtcg ggaagatgcg tgatctgatc cttcaactca 17400

gcaaaagttc gatttattca acaaagccgc cgtcccgtca agtcagcgta atgccctagg 17460

aggcgcgcca cggccgcgtc gaccccacgc ccctctttaa tacgacgggc aatttgcact 17520

tcagaaaatg aagagtttgc tttagccata acaaaagtcc agtatgcttt ttcacagcat 17580

aactggactg atttcagttt acaactattc tgtctagttt aagactttat tgtcatagtt 17640

tagatctatt ttgttcagtt taagacttta ttgtccgccc aca 17683

<210> SEQ ID NO 19
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 19

gatgctggtg gcgaagctgt 20

<210> SEQ ID NO 20
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

-continued

<400> SEQUENCE: 20

gttgcgacgg tggtacgcat aac 23

<210> SEQ ID NO 21
<211> LENGTH: 30
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 21

cccagatcag ttttatgatt gcctcgctgg 30

<210> SEQ ID NO 22
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 22

atcatgaaac gtctcggaac 20

<210> SEQ ID NO 23
<211> LENGTH: 28
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 23

cctcgagtta cttgcgggtt cgggcgcg 28

<210> SEQ ID NO 24
<211> LENGTH: 5199
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 24

cactatacca attgagatgg gctagtcaat gataattact agtccttttc ctttgagttg 60

tgggtatctg taaattctgc tagacctttg ctggaaaact tgtaaattct gctagaccct 120

ctgtaaattc cgctagacct ttgtgtgttt tttttgttta tattcaagtg gttataattt 180

atagaataaa gaaagaataa aaaaagataa aaagaataga tcccagccct gtgtataact 240

cactacttta gtcagttccg cagtattaca aaaggatgtc gcaaacgctg tttcagtaca 300

ctctctcaat acgaataaac ggctcagaaa tgagccgttt atttttttcta cccatatcct 360

tgaagcggtg ttataatgcc gcgccctcga tatggggatt tttaacgacc tgattttcgg 420

gtctcagtag tagttgacat tagcggagca ctaaaccatg aaacgtctcg gaaccctgga 480

cgcctcctgg ctggcggttg aatctgaaga caccccgatg catgtgggta cgcttcagat 540

tttctcactg ccggaaggcg caccagaaac cttcctgcgt gacatggtca ctcgaatgaa 600

agaggccggc gatgtggcac caccctgggg atacaaactg gcctggtctg gtttcctcgg 660

-continued

gcgcgtgatc gccccggcct ggaaagtcga taaggatatc gatctggatt atcacgtccg 720

gcactcagcc ctgcctcgcc ccggcgggga gcgcgaactg ggtattctgg tatcccgact 780

gcactctaac cccctggatt tttcccgccc tctttgggaa tgccacgtta ttgaaggcct 840

ggagaataac cgttttgccc tttacaccaa aatgcaccac tcgatgattg acggcatcag 900

cggcgtgcga ctgatgcaga gggtgctcac caccgatccc gaacgctgca atatgccacc 960

gccctggacg gtacgcccac accaacgccg tggtgcaaaa accgacaaag aggccagcgt 1020

gcccgcagcg gtttcccagg caatggacgc cctgaagctc caggcagaca tggcccccag 1080

gctgtggcag gccggcaatc gcctggtgca ttcggttcga cacccggaag acggactgac 1140

cgcgcccttc actggaccgg tttcggtgct caatcaccgg gttaccgcgc agcgacgttt 1200

tgccacccag cattatcaac tggaccggct gaaaaacctg gcccatgctt ccggcggttc 1260

cttgaacgac atcgtgcttt acctgtgtgg caccgcattg cggcgctttc tggctgagca 1320

gaacaatctg ccagacaccc cgctgacggc tggtataccg gtgaatatcc ggccggcaga 1380

cgacgagggt acgggcaccc agatcagttt tatgattgcc tcgctggcca ccgacgaagc 1440

tgatccgttg aaccgcctgc aacagatcaa aacctcgacc cgacgggcca aggagcacct 1500

gcagaaactt ccaaaaagtg ccctgaccca gtacaccatg ctgctgatgt caccctacat 1560

tctgcaattg atgtcaggtc tcggggggag gatgcgacca gtcttcaacg tgaccatttc 1620

caacgtgccc ggcccggaag gcacgctgta ttatgaagga gcccggcttg aggccatgta 1680

tccggtatcg ctaatcgctc acggcggcgc cctgaacatc acctgcctga gctatgccgg 1740

atcgctgaat ttcggtttta ccggctgtcg ggatacgctg ccgagcatgc agaaactggc 1800

ggtttatacc ggtgaagctc tggatgagct ggaatcgctg attctgccac ccaagaagcg 1860

cgcccgaacc cgcaagtaac tcgagatctg cagctggtac catatgggaa ttcacccgct 1920

gacgagctta gtaaagccct cgctagattt taatgcggat gttgcgatta cttcgccaac 1980

tattgcgata acaagaaaaa gccagccttt catgatatat ctcccaattt gtgtagggct 2040

tattatgcac gcttaaaaat aataaaagca gacttgacct gatagtttgg ctgtgagcaa 2100

ttatgtgctt agtgcatcta acgcttgagt taagccgcgc cgcgaagcgg cgtcggcttg 2160

aacgaattgt tagacattat ttgccgacta ccttggtgat ctcgcctttc acgtagtgga 2220

caaattcttc caactgatct gcgcgcgagg ccaagcgatc ttcttcttgt ccaagataag 2280

cctgtctagc ttcaagtatg acgggctgat actgggccgg caggcgctcc attgcccagt 2340

cggcagcgac atccttcggc gcgattttgc cggttactgc gctgtaccaa atgcgggaca 2400

acgtaagcac tacatttcgc tcatcgccag cccagtcggg cggcgagttc catagcgtta 2460

aggtttcatt tagcgcctca aatagatcct gttcaggaac cggatcaaag agttcctccg 2520

ccgctggacc taccaaggca acgctatgtt ctcttgcttt tgtcagcaag atagccagat 2580

caatgtcgat cgtggctggc tcgaagatac ctgcaagaat gtcattgcgc tgccattctc 2640

caaattgcag ttcgcgctta gctggataac gccacggaat gatgtcgtcg tgcacaacaa 2700

tggtgacttc tacagcgcgg agaatctcgc tctctccagg ggaagccgaa gtttccaaaa 2760

ggtcgttgat caaagctcgc cgcgttgttt catcaagcct tacggtcacc gtaaccagca 2820

aatcaatatc actgtgtggc ttcaggccgc catccactgc ggagccgtac aaatgtacgg 2880

ccagcaacgt cggttcgaga tggcgctcga tgacgccaac tacctctgat agttgagtcg 2940

-continued

atacttcggc gatcaccgct tccctcatga tgtttaactt tgttttaggg cgactgccct 3000

gctgcgtaac atcgttgctg ctccataaca tcaaacatcg acccacggcg taacgcgctt 3060

gctgcttgga tgcccgaggc atagactgta ccccaaaaaa acagtcataa caagccatga 3120

aaaccgccac tgcgccgtta ccaccgctgc gttcggtcaa ggttctggac cagttgcgtg 3180

agcgcatacg ctacttgcat tacagcttac gaaccgaaca ggcttatgtc cactgggttc 3240

gtgccttcat ccgtttccac ggtgtgcgtc acccggcaac cttgggcagc agcgaagtcg 3300

aggcatttct gtcctggctg gcgaacgagc gcaaggtttc ggtctccacg catcgtcagg 3360

cattggcggc cttgctgttc ttctacggca aggtgctgtg cacggatctg ccctggcttc 3420

aggagatcgg aagacctcgg ccgtcgcggc gcttgccggt ggtgctgacc ccggatgaag 3480

tggttcgcat cctcggtttt ctggaaggcg agcatcgttt gttcgcccag cttctgtatg 3540

gaacgggcat gcggatcagt gagggtttgc aactgcgggt caaggatctg gatttcgatc 3600

acggcacgat catcgtgcgg gagggcaagg gctccaagga tcgggccttg atgttacccg 3660

agagcttggc acccagcctg cgcgagcagg ggaattaatt cccacgggtt ttgctgcccg 3720

caaacgggct gttctggtgt tgctagtttg ttatcagaat cgcagatccg gcttcagccg 3780

gtttgccggc tgaaagcgct atttcttcca gaattgccat gatttttcc ccacgggagg 3840

cgtcactggc tcccgtgttg tcggcagctt tgattcgata agcagcatcg cctgtttcag 3900

gctgtctatg tgtgactgtt gagctgtaac aagttgtctc aggtgttcaa tttcatgttc 3960

tagttgcttt gttttactgg tttcacctgt tctattaggt gttacatgct gttcatctgt 4020

tacattgtcg atctgttcat ggtgaacagc tttgaatgca ccaaaaactc gtaaaagctc 4080

tgatgtatct atctttttta caccgttttc atctgtgcat atggacagtt ttccctttga 4140

tatgtaacgg tgaacagttg ttctactttt gtttgttagt cttgatgctt cactgataga 4200

tacaagagcc ataagaacct cagatccttc cgtatttagc cagtatgttc tctagtgtgg 4260

ttcgttgttt ttgcgtgagc catgagaacg aaccattgag atcatactta ctttgcatgt 4320

cactcaaaaa ttttgcctca aaactggtga gctgaatttt tgcagttaaa gcatcgtgta 4380

gtgtttttct tagtccgtta tgtaggtagg aatctgatgt aatggttgtt ggtattttgt 4440

caccattcat ttttatctgg ttgttctcaa gttcggttac gagatccatt tgtctatcta 4500

gttcaacttg gaaaatcaac gtatcagtcg ggcggcctcg cttatcaacc accaatttca 4560

tattgctgta agtgtttaaa tctttactta ttggtttcaa aacccattgg ttaagccttt 4620

taaactcatg gtagttattt tcaagcatta acatgaactt aaattcatca aggctaatct 4680

ctatatttgc cttgtgagtt ttcttttgtg ttagttcttt taataaccac tcataaatcc 4740

tcatagagta tttgttttca aaagacttaa catgttccag attatatttt atgaattttt 4800

ttaactggaa aagataaggc aatatctctt cactaaaaac taattctaat ttttcgcttg 4860

agaacttggc atagtttgtc cactggaaaa tctcaaagcc tttaaccaaa ggattcctga 4920

tttccacagt tctcgtcatc agctctctgg ttgctttagc taatacacca taagcatttt 4980

ccctactgat gttcatcatc tgagcgtatt ggttataagt gaacgatacc gtccgttctt 5040

tccttgtagg gttttcaatc gtggggttga gtagtgccac acagcataaa attagcttgg 5100

tttcatgctc cgttaagtca tagcgactaa tcgctagttc atttgctttg aaaacaacta 5160

attcagacat acatctcaat tggtctaggt gattttaat 5199

-continued

<210> SEQ ID NO 25
<211> LENGTH: 169
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 25

gctgtttcag tacactctct caatacgaat aaacggctca gaaatgagcc gtttatttt 60

tctacccata tccttgaagc ggtgttataa tgccgcgccc tcgatatggg gatttttaac 120

gacctgattt tcgggtctca gtagtagttg acattagcgg agcactaaa 169

<210> SEQ ID NO 26
<211> LENGTH: 42
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 26

aaaggatgtc gcaaacgctg tttcagtaca ctctctcaat ac 42

<210> SEQ ID NO 27
<211> LENGTH: 34
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 27

gagctcggat ccatggttta gtgctccgct aatg 34

<210> SEQ ID NO 28
<211> LENGTH: 5903
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 28

cactatacca attgagatgg gctagtcaat gataattact agtccttttc ctttgagttg 60

tgggtatctg taaattctgc tagacctttg ctggaaaact tgtaaattct gctagaccct 120

ctgtaaattc cgctagacct ttgtgtgttt tttttgttta tattcaagtg gttataattt 180

atagaataaa gaaagaataa aaaaagataa aaagaataga tcccagccct gtgtataact 240

cactacttta gtcagttccg cagtattaca aaaggatgtc gcaaacgctg tttgctcctc 300

tacaaaacag accttaaaac cctaaaggcg tcggcatccg cttacagaca agctgtgacc 360

gtctccggga gctgcatgtg tcagaggttt tcaccgtcat caccgaaacg cgcgaggcag 420

cagatcaatt cgcgcgcgaa ggcgaagcgg catgcattta cgttgacacc atcgaatggt 480

gcaaaacctt tcgcggtatg gcatgatagc gcccggaaga gagtcaattc agggtggtga 540

atgtgaaacc agtaacgtta tacgatgtcg cagagtatgc cggtgtctct tatcagaccg 600

tttcccgcgt ggtgaaccag gccagccacg tttctgcgaa aacgcgggaa aaagtggaag 660

-continued

cggcgatggc ggagctgaat tacattccca accgcgtggc acaacaactg gcgggcaaac 720

agtcgttgct gattggcgtt gccacctcca gtctggccct gcacgcgccg tcgcaaattg 780

tcgcggcgat taaatctcgc gccgatcaac tgggtgccag cgtggtggtg tcgatggtag 840

aacgaagcgg cgtcgaagcc tgtaaagcgg cggtgcacaa tcttctcgcg caacgcgtca 900

gtgggctgat cattaactat ccgctggatg accaggatgc cattgctgtg gaagctgcct 960

gcactaatgt tccggcgtta tttcttgatg tctctgacca gacacccatc aacagtatta 1020

ttttctccca tgaagacggt acgcgactgg gcgtggagca tctggtcgca ttgggtcacc 1080

agcaaatcgc gctgttagcg ggcccattaa gttctgtctc ggcgcgtctg cgtctggctg 1140

gctggcataa atatctcact cgcaatcaaa ttcagccgat agcggaacgg gaaggcgact 1200

ggagtgccat gtccggtttt caacaaacca tgcaaatgct gaatgagggc atcgttccca 1260

ctgcgatgct ggttgccaac gatcagatgg cgctgggcgc aatgcgcgcc attaccgagt 1320

ccgggctgcg cgttggtgcg gatatctcgg tagtgggata cgacgatacc gaagacagct 1380

catgttatat cccgccgtta accaccatca aacaggattt tcgcctgctg gggcaaacca 1440

gcgtggaccg cttgctgcaa ctctctcagg gccaggcggt gaagggcaat cagctgttgc 1500

ccgtctcact ggtgaaaaga aaaaccaccc tggcgcccaa tacgcaaacc gcctctcccc 1560

gcgcgttggc cgattcatta atgcagctgg cacgacaggt ttcccgactg gaaagcgggc 1620

agtgagcgca acgcaattaa tgtaagttag cgcgaattga tctggtttga cagcttatca 1680

tcgactgcac ggtgcaccaa tgcttctggc gtcaggcagc catcggaagc tgtggtatgg 1740

ctgtgcaggt cgtaaatcac tgcataattc gtgtcgctca aggcgcactc ccgttctgga 1800

taatgttttt tgcgccgaca tcataacggt tctggcaaat attctgaaat gagctgttga 1860

caattaatca tccggctcgt ataatgtgtg gaattgtgag cggataacaa tttcacacag 1920

gaaacagcgc cgctgagaaa aagcgaagcg gcactgctct ttaacaattt atcagacaat 1980

ctgtgtgggc actcgaccgg aattatcgat taactttatt attaaaaatt aaagaggtat 2040

atattaatgt atcgattaaa taaggaggaa taaaccatgg atccgagctc gagatctgca 2100

gctggtacca tatgggaatt cgaagcttgg gcccgaacaa aaactcatct cagaagagga 2160

tctgaatagc gccgtcgacc atcatcatca tcatcattga gtttaaacgg tctccagctt 2220

ggctgttttg gcggatgaga gaagattttc agcctgatac agattaaatc agaacgcaga 2280

agcggtctga taaaacagaa tttgcctggc ggcagtagcg cggtggtccc acctgacccc 2340

atgccgaact cagaagtgaa acgccgtagc gccgatggta gtgtggggtc tccccatgcg 2400

agagtaggga actgccaggc atcaaataaa acgaaaggct cagtcgaaag actgggcctt 2460

tcgttttatc tgttgtttgt cggtgaacgc tctcctgacg cctgatgcgg tattttctcc 2520

ttacgcatct gtgcggtatt tcacaccgca tatggtgcac tctcagtaca atctgctctg 2580

atgccgcata gttaagccag ccccgacacc cgccaacacc cgctgacgag cttagtaaag 2640

ccctcgctag attttaatgc ggatgttgcg attacttcgc caactattgc gataacaaga 2700

aaaagccagc ctttcatgat atatctccca atttgtgtag ggcttattat gcacgcttaa 2760

aaataataaa agcagacttg acctgatagt ttggctgtga gcaattatgt gcttagtgca 2820

tctaacgctt gagttaagcc gcgccgcgaa gcggcgtcgg cttgaacgaa ttgttagaca 2880

ttatttgccg actaccttgg tgatctcgcc tttcacgtag tggacaaatt cttccaactg 2940

-continued

atctgcgcgc gaggccaagc gatcttcttc ttgtccaaga taagcctgtc tagcttcaag 3000

tatgacgggc tgatactggg ccggcaggcg ctccattgcc cagtcggcag cgacatcctt 3060

cggcgcgatt ttgccggtta ctgcgctgta ccaaatgcgg gacaacgtaa gcactacatt 3120

tcgctcatcg ccagcccagt cgggcggcga gttccatagc gttaaggttt catttagcgc 3180

ctcaaataga tcctgttcag gaaccggatc aaagagttcc tccgccgctg gacctaccaa 3240

ggcaacgcta tgttctcttg cttttgtcag caagatagcc agatcaatgt cgatcgtggc 3300

tggctcgaag atacctgcaa gaatgtcatt gcgctgccat tctccaaatt gcagttcgcg 3360

cttagctgga taacgccacg gaatgatgtc gtcgtgcaca acaatggtga cttctacagc 3420

gcggagaatc tcgctctctc caggggaagc cgaagtttcc aaaaggtcgt tgatcaaagc 3480

tcgccgcgtt gtttcatcaa gccttacggt caccgtaacc agcaaatcaa tatcactgtg 3540

tggcttcagg ccgccatcca ctgcggagcc gtacaaatgt acggccagca acgtcggttc 3600

gagatggcgc tcgatgacgc caactacctc tgatagttga gtcgatactt cggcgatcac 3660

cgcttccctc atgatgttta actttgtttt agggcgactg ccctgctgcg taacatcgtt 3720

gctgctccat aacatcaaac atcgacccac ggcgtaacgc gcttgctgct tggatgcccg 3780

aggcatagac tgtaccccaa aaaaacagtc ataacaagcc atgaaaaccg ccactgcgcc 3840

gttaccaccg ctgcgttcgg tcaaggttct ggaccagttg cgtgagcgca tacgctactt 3900

gcattacagc ttacgaaccg aacaggctta tgtccactgg gttcgtgcct tcatccgttt 3960

ccacggtgtg cgtcacccgg caaccttggg cagcagcgaa gtcgaggcat ttctgtcctg 4020

gctggcgaac gagcgcaagg tttcggtctc cacgcatcgt caggcattgg cggccttgct 4080

gttcttctac ggcaaggtgc tgtgcacgga tctgccctgg cttcaggaga tcggaagacc 4140

tcggccgtcg cggcgcttgc cggtggtgct gaccccggat gaagtggttc gcatcctcgg 4200

ttttctggaa ggcgagcatc gtttgttcgc ccagcttctg tatggaacgg gcatgcggat 4260

cagtgagggt ttgcaactgc gggtcaagga tctggatttc gatcacggca cgatcatcgt 4320

gcgggagggc aagggctcca aggatcgggc cttgatgtta cccgagagct tggcacccag 4380

cctgcgcgag caggggaatt aattcccacg ggttttgctg cccgcaaacg ggctgttctg 4440

gtgttgctag tttgttatca gaatcgcaga tccggcttca gccggtttgc cggctgaaag 4500

cgctatttct tccagaattg ccatgatttt ttccccacgg gaggcgtcac tggctcccgt 4560

gttgtcggca gctttgattc gataagcagc atcgcctgtt tcaggctgtc tatgtgtgac 4620

tgttgagctg taacaagttg tctcaggtgt tcaatttcat gttctagttg ctttgtttta 4680

ctggtttcac ctgttctatt aggtgttaca tgctgttcat ctgttacatt gtcgatctgt 4740

tcatggtgaa cagctttgaa tgcaccaaaa actcgtaaaa gctctgatgt atctatcttt 4800

tttacaccgt tttcatctgt gcatatggac agttttccct ttgatatgta acggtgaaca 4860

gttgttctac ttttgtttgt tagtcttgat gcttcactga tagatacaag agccataaga 4920

acctcagatc cttccgtatt tagccagtat gttctctagt gtggttcgtt gtttttgcgt 4980

gagccatgag aacgaaccat tgagatcata cttactttgc atgtcactca aaaattttgc 5040

ctcaaaactg gtgagctgaa tttttgcagt taaagcatcg tgtagtgttt ttcttagtcc 5100

gttatgtagg taggaatctg atgtaatggt tgttggtatt ttgtcaccat tcatttttat 5160

ctggttgttc tcaagttcgg ttacgagatc catttgtcta tctagttcaa cttggaaaat 5220

-continued

caacgtatca gtcgggcggc ctcgcttatc aaccaccaat ttcatattgc tgtaagtgtt 5280

taaatcttta cttattggtt tcaaaaccca ttggttaagc cttttaaact catggtagtt 5340

attttcaagc attaacatga acttaaattc atcaaggcta atctctatat ttgccttgtg 5400

agttttcttt tgtgttagtt cttttaataa ccactcataa atcctcatag agtatttgtt 5460

ttcaaaagac ttaacatgtt ccagattata ttttatgaat ttttttaact ggaaaagata 5520

aggcaatatc tcttcactaa aaactaattc taatttttcg cttgagaact tggcatagtt 5580

tgtccactgg aaaatctcaa agcctttaac caaaggattc ctgatttcca cagttctcgt 5640

catcagctct ctggttgctt tagctaatac accataagca ttttccctac tgatgttcat 5700

catctgagcg tattggttat aagtgaacga taccgtccgt tctttccttg tagggttttc 5760

aatcgtgggg ttgagtagtg ccacacagca taaaattagc ttggtttcat gctccgttaa 5820

gtcatagcga ctaatcgcta gttcatttgc tttgaaaaca actaattcag acatacatct 5880

caattggtct aggtgatttt aat 5903

<210> SEQ ID NO 29
<211> LENGTH: 28
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 29

atatgacgtc ggcatccgct tacagaca 28

<210> SEQ ID NO 30
<211> LENGTH: 32
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 30

aattcttaag tcaggagagc gttcaccgac aa 32

<210> SEQ ID NO 31
<211> LENGTH: 24
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 31

gaattccacc cgctgacgag ctta 24

<210> SEQ ID NO 32
<211> LENGTH: 21
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 32

cgaattccca tatggtacca g 21

-continued

<210> SEQ ID NO 33
<211> LENGTH: 70
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 33

gccgcagcct gatggacaaa gcgttcatta tggtgctgcc ggtcgcgatg attccgggga 60

tccgtcgacc 70

<210> SEQ ID NO 34
<211> LENGTH: 65
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 34

atcttcaacg gtaacttctt taccgccatg cgtgtcccag gtgtctgtag gctggagctg 60

cttcg 65

<210> SEQ ID NO 35
<211> LENGTH: 71
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 35

ctcccctgga atgcagggga gcggcaagat taaaccagtt cgttcgggca gtgtaggctg 60

gagctgcttc g 71

<210> SEQ ID NO 36
<211> LENGTH: 74
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 36

tatttttagt agcttaaatg tgattcaaca tcactggaga aagtcttatg catatgaata 60

tcctccttag ttcc 74

<210> SEQ ID NO 37
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 37

ggactaaacg tcctacaaac 20

<210> SEQ ID NO 38

-continued

<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 38

ttcatctgtt tgagatcgag 20

<210> SEQ ID NO 39
<211> LENGTH: 29
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 39

cccgagcggt agccagatgc ccgccagcg 29

<210> SEQ ID NO 40
<211> LENGTH: 29
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 40

gctgcgggtt agcgcacatc atacgggtc 29

<210> SEQ ID NO 41
<211> LENGTH: 7314
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 41

cactatacca attgagatgg gctagtcaat gataattact agtccttttc ctttgagttg 60

tgggtatctg taaattctgc tagacctttg ctggaaaact tgtaaattct gctagaccct 120

ctgtaaattc cgctagacct ttgtgtgttt tttttgttta tattcaagtg gttataattt 180

atagaataaa gaaagaataa aaaaagataa aaagaataga tcccagccct gtgtataact 240

cactacttta gtcagttccg cagtattaca aaaggatgtc gcaaacgctg tttgctcctc 300

tacaaaacag accttaaaac cctaaaggcg tcggcatccg cttacagaca agctgtgacc 360

gtctccggga gctgcatgtg tcagaggttt tcaccgtcat caccgaaacg cgcgaggcag 420

cagatcaatt cgcgcgcgaa ggcgaagcgg catgcattta cgttgacacc atcgaatggt 480

gcaaaacctt tcgcggtatg gcatgatagc gcccggaaga gagtcaattc agggtggtga 540

atgtgaaacc agtaacgtta tacgatgtcg cagagtatgc cggtgtctct tatcagaccg 600

tttcccgcgt ggtgaaccag gccagccacg tttctgcgaa aacgcgggaa aaagtggaag 660

cggcgatggc ggagctgaat tacattccca accgcgtggc acaacaactg gcgggcaaac 720

agtcgttgct gattggcgtt gccacctcca gtctggccct gcacgcgccg tcgcaaattg 780

tcgcggcgat taaatctcgc gccgatcaac tgggtgccag cgtggtggtg tcgatggtag 840

-continued

aacgaagcgg cgtcgaagcc tgtaaagcgg cggtgcacaa tcttctcgcg caacgcgtca 900

gtgggctgat cattaactat ccgctggatg accaggatgc cattgctgtg gaagctgcct 960

gcactaatgt tccggcgtta tttcttgatg tctctgacca gacacccatc aacagtatta 1020

ttttctccca tgaagacggt acgcgactgg gcgtggagca tctggtcgca ttgggtcacc 1080

agcaaatcgc gctgttagcg ggcccattaa gttctgtctc ggcgcgtctg cgtctggctg 1140

gctggcataa atatctcact cgcaatcaaa ttcagccgat agcggaacgg gaaggcgact 1200

ggagtgccat gtccggtttt caacaaacca tgcaaatgct gaatgagggc atcgttccca 1260

ctgcgatgct ggttgccaac gatcagatgg cgctgggcgc aatgcgcgcc attaccgagt 1320

ccgggctgcg cgttggtgcg gatatctcgg tagtgggata cgacgatacc gaagacagct 1380

catgttatat cccgccgtta accaccatca aacaggattt tcgcctgctg gggcaaacca 1440

gcgtggaccg cttgctgcaa ctctctcagg gccaggcggt gaagggcaat cagctgttgc 1500

ccgtctcact ggtgaaaaga aaaaccaccc tggcgcccaa tacgcaaacc gcctctcccc 1560

gcgcgttggc cgattcatta atgcagctgg cacgacaggt ttcccgactg gaaagcgggc 1620

agtgagcgca acgcaattaa tgtaagttag cgcgaattga tctggtttga cagcttatca 1680

tcgactgcac ggtgcaccaa tgcttctggc gtcaggcagc catcggaagc tgtggtatgg 1740

ctgtgcaggt cgtaaatcac tgcataattc gtgtcgctca aggcgcactc ccgttctgga 1800

taatgttttt tgcgccgaca tcataacggt tctggcaaat attctgaaat gagctgttga 1860

caattaatca tccggctcgt ataatgtgtg gaattgtgag cggataacaa tttcacacag 1920

gaaacagcgc cgctgagaaa aagcgaagcg gcactgctct ttaacaattt atcagacaat 1980

ctgtgtgggc actcgaccgg aattatcgat taactttatt attaaaaatt aaagaggtat 2040

atattaatgt atcgattaaa taaggaggaa taaaccatga aacgtctcgg aaccctggac 2100

gcctcctggc tggcggttga atctgaagac accccgatgc atgtgggtac gcttcagatt 2160

ttctcactgc cggaaggcgc accagaaacc ttcctgcgtg acatggtcac tcgaatgaaa 2220

gaggccggcg atgtggcacc accctgggga tacaaactgg cctggtctgg tttcctcggg 2280

cgcgtgatcg ccccggcctg gaaagtcgat aaggatatcg atctggatta tcacgtccgg 2340

cactcagccc tgcctcgccc cggcggggag cgcgaactgg gtattctggt atcccgactg 2400

cactctaacc ccctggattt ttcccgccct ctttgggaat gccacgttat tgaaggcctg 2460

gagaataacc gttttgccct ttacaccaaa atgcaccact cgatgattga cggcatcagc 2520

ggcgtgcgac tgatgcagag ggtgctcacc accgatcccg aacgctgcaa tatgccaccg 2580

ccctggacgg tacgcccaca ccaacgccgt ggtgcaaaaa ccgacaaaga ggccagcgtg 2640

cccgcagcgg tttcccaggc aatggacgcc ctgaagctcc aggcagacat ggcccccagg 2700

ctgtggcagg ccggcaatcg cctggtgcat tcggttcgac acccggaaga cggactgacc 2760

gcgcccttca ctggaccggt ttcggtgctc aatcaccggg ttaccgcgca gcgacgtttt 2820

gccacccagc attatcaact ggaccggctg aaaaacctgg cccatgcttc cggcggttcc 2880

ttgaacgaca tcgtgcttta cctgtgtggc accgcattgc ggcgctttct ggctgagcag 2940

aacaatctgc cagacacccc gctgacggct ggtataccgg tgaatatccg gccggcagac 3000

gacgagggta cgggcaccca gatcagtttt atgattgcct cgctggccac cgacgaagct 3060

gatccgttga accgcctgca acagatcaaa acctcgaccc gacgggccaa ggagcacctg 3120

-continued

cagaaacttc caaaaagtgc cctgacccag tacaccatgc tgctgatgtc accctacatt 3180

ctgcaattga tgtcaggtct cgggggggagg atgcgaccag tcttcaacgt gaccatttcc 3240

aacgtgcccg gcccggaagg cacgctgtat tatgaaggag cccggcttga ggccatgtat 3300

ccggtatcgc taatcgctca cggcggcgcc ctgaacatca cctgcctgag ctatgccgga 3360

tcgctgaatt tcggttttac cggctgtcgg gatacgctgc cgagcatgca gaaactggcg 3420

gtttataccg gtgaagctct ggatgagctg gaatcgctga ttctgccacc caagaagcgc 3480

gcccgaaccc gcaagtaact cgagatctgc agctggtacc atatgggaat tcgaagcttg 3540

ggcccgaaca aaaactcatc tcagaagagg atctgaatag cgccgtcgac catcatcatc 3600

atcatcattg agtttaaacg gtctccagct tggctgtttt ggcggatgag agaagatttt 3660

cagcctgata cagattaaat cagaacgcag aagcggtctg ataaaacaga atttgcctgg 3720

cggcagtagc gcggtggtcc cacctgaccc catgccgaac tcagaagtga aacgccgtag 3780

cgccgatggt agtgtggggt ctccccatgc gagagtaggg aactgccagg catcaaataa 3840

aacgaaaggc tcagtcgaaa gactgggcct ttcgttttat ctgttgtttg tcggtgaacg 3900

ctctcctgac gcctgatgcg gtattttctc cttacgcatc tgtgcggtat ttcacaccgc 3960

atatggtgca ctctcagtac aatctgctct gatgccgcat agttaagcca gccccgacac 4020

ccgccaacac ccgctgacga gcttagtaaa gccctcgcta gattttaatg cggatgttgc 4080

gattacttcg ccaactattg cgataacaag aaaaagccag cctttcatga tatatctccc 4140

aatttgtgta gggcttatta tgcacgctta aaaataataa aagcagactt gacctgatag 4200

tttggctgtg agcaattatg tgcttagtgc atctaacgct tgagttaagc cgcgccgcga 4260

agcggcgtcg gcttgaacga attgttagac attatttgcc gactaccttg gtgatctcgc 4320

ctttcacgta gtggacaaat tcttccaact gatctgcgcg cgaggccaag cgatcttctt 4380

cttgtccaag ataagcctgt ctagcttcaa gtatgacggg ctgatactgg gccggcaggc 4440

gctccattgc ccagtcggca gcgacatcct tcggcgcgat tttgccggtt actgcgctgt 4500

accaaatgcg ggacaacgta agcactacat ttcgctcatc gccagcccag tcgggcggcg 4560

agttccatag cgttaaggtt tcatttagcg cctcaaatag atcctgttca ggaaccggat 4620

caaagagttc ctccgccgct ggacctacca aggcaacgct atgttctctt gcttttgtca 4680

gcaagatagc cagatcaatg tcgatcgtgg ctggctcgaa gatacctgca agaatgtcat 4740

tgcgctgcca ttctccaaat tgcagttcgc gcttagctgg ataacgccac ggaatgatgt 4800

cgtcgtgcac aacaatggtg acttctacag cgcggagaat ctcgctctct ccaggggaag 4860

ccgaagtttc caaaaggtcg ttgatcaaag ctcgccgcgt tgtttcatca agccttacgg 4920

tcaccgtaac cagcaaatca atatcactgt gtggcttcag gccgccatcc actgcggagc 4980

cgtacaaatg tacggccagc aacgtcggtt cgagatggcg ctcgatgacg ccaactacct 5040

ctgatagttg agtcgatact tcggcgatca ccgcttccct catgatgttt aactttgttt 5100

tagggcgact gccctgctgc gtaacatcgt tgctgctcca taacatcaaa catcgaccca 5160

cggcgtaacg cgcttgctgc ttggatgccc gaggcataga ctgtacccca aaaaaacagt 5220

cataacaagc catgaaaacc gccactgcgc cgttaccacc gctgcgttcg gtcaaggttc 5280

tggaccagtt gcgtgagcgc atacgctact tgcattacag cttacgaacc gaacaggctt 5340

atgtccactg ggttcgtgcc ttcatccgtt tccacggtgt gcgtcacccg gcaaccttgg 5400

-continued

gcagcagcga agtcgaggca tttctgtcct ggctggcgaa cgagcgcaag gtttcggtct 5460

ccacgcatcg tcaggcattg gcggccttgc tgttcttcta cggcaaggtg ctgtgcacgg 5520

atctgccctg gcttcaggag atcggaagac ctcggccgtc gcggcgcttg ccggtggtgc 5580

tgaccccgga tgaagtggtt cgcatcctcg gttttctgga aggcgagcat cgtttgttcg 5640

cccagcttct gtatggaacg ggcatgcgga tcagtgaggg tttgcaactg cgggtcaagg 5700

atctggattt cgatcacggc acgatcatcg tgcgggaggg caagggctcc aaggatcggg 5760

ccttgatgtt acccgagagc ttggcaccca gcctgcgcga gcaggggaat taattcccac 5820

gggttttgct gcccgcaaac gggctgttct ggtgttgcta gtttgttatc agaatcgcag 5880

atccggcttc agccggtttg ccggctgaaa gcgctatttc ttccagaatt gccatgattt 5940

tttccccacg ggaggcgtca ctggctcccg tgttgtcggc agctttgatt cgataagcag 6000

catcgcctgt ttcaggctgt ctatgtgtga ctgttgagct gtaacaagtt gtctcaggtg 6060

ttcaatttca tgttctagtt gctttgtttt actggtttca cctgttctat taggtgttac 6120

atgctgttca tctgttacat tgtcgatctg ttcatggtga acagctttga atgcaccaaa 6180

aactcgtaaa agctctgatg tatctatctt ttttacaccg ttttcatctg tgcatatgga 6240

cagttttccc tttgatatgt aacggtgaac agttgttcta cttttgtttg ttagtcttga 6300

tgcttcactg atagatacaa gagccataag aacctcagat ccttccgtat ttagccagta 6360

tgttctctag tgtggttcgt tgtttttgcg tgagccatga gaacgaacca ttgagatcat 6420

acttactttg catgtcactc aaaaattttg cctcaaaact ggtgagctga atttttgcag 6480

ttaaagcatc gtgtagtgtt tttcttagtc cgttatgtag gtaggaatct gatgtaatgg 6540

ttgttggtat tttgtcacca ttcattttta tctggttgtt ctcaagttcg gttacgagat 6600

ccatttgtct atctagttca acttggaaaa tcaacgtatc agtcgggcgg cctcgcttat 6660

caaccaccaa tttcatattg ctgtaagtgt ttaaatcttt acttattggt ttcaaaaccc 6720

attggttaag ccttttaaac tcatggtagt tattttcaag cattaacatg aacttaaatt 6780

catcaaggct aatctctata tttgccttgt gagttttctt ttgtgttagt tcttttaata 6840

accactcata aatcctcata gagtatttgt tttcaaaaga cttaacatgt tccagattat 6900

attttatgaa ttttttttaac tggaaaagat aaggcaatat ctcttcacta aaaactaatt 6960

ctaatttttc gcttgagaac ttggcatagt ttgtccactg gaaaatctca aagcctttaa 7020

ccaaaggatt cctgatttcc acagttctcg tcatcagctc tctggttgct ttagctaata 7080

caccataagc attttcccta ctgatgttca tcatctgagc gtattggtta taagtgaacg 7140

ataccgtccg ttctttcctt gtagggtttt caatcgtggg gttgagtagt gccacacagc 7200

ataaaattag cttggtttca tgctccgtta agtcatagcg actaatcgct agttcatttg 7260

ctttgaaaac aactaattca gacatacatc tcaattggtc taggtgattt taat 7314

<210> SEQ ID NO 42
<211> LENGTH: 787
<212> TYPE: PRT
<213> ORGANISM: Stenotrophomonas maltophilia
<220> FEATURE:
<223> OTHER INFORMATION: R551-3

<400> SEQUENCE: 42

Met Thr Met Gly Val Gly Leu Cys Thr Gly Asp Val Ser Ile Gly Arg
1 5 10 15

-continued

```
Gly Gly Gly Arg Ser Arg Arg Gly Trp Val Ser Glu Phe Leu Pro Val
            20                  25                  30

Ala Val Ala Gly Asp Glu Gly Ala Gly Val Ala Ala Ala Thr Gln Trp
        35                  40                  45

Cys Arg Cys Asp Leu Arg Thr Arg Arg Asn Arg Cys Gly Gly Gly Trp
    50                  55                  60

Leu Asp Thr Arg Gly His Cys Leu His Arg Gly Gly Asp Leu Val Val
65                  70                  75                  80

Gln Ala Glu Asp Arg His Gly Val Gly Gln Leu Ala Ser Leu Phe Phe
                85                  90                  95

His Arg Thr Cys Gly Gly Gly Gly Phe Phe His Gln Arg Ser Ile Leu
            100                 105                 110

Leu Gly Gly Phe Val His Leu His His Gly Leu Val Asp Leu Leu Asp
        115                 120                 125

Ala Gly Arg Leu Leu Leu Arg Gly Arg Gly Asp Leu Gly His Asp Val
    130                 135                 140

Gly His Ala Leu His Arg Gly Asp Asp Leu Gly His Gly Ala Ala Gly
145                 150                 155                 160

Leu Val Asp Gln Ala Gly Ala Leu Gly Asp Leu Ala Asp Arg Val Ile
                165                 170                 175

Asp Gln Ala Leu Asp Leu Leu Gly Gly Gly Gly Arg Ala Leu Gly Glu
            180                 185                 190

Gly Ala His Phe Ala Gly Asp His Gly Lys Ala Thr Ala Leu Phe Ala
        195                 200                 205

Gly Thr Cys Gly Phe His Cys Gly Val Gln Arg Gln Asp Val Gly Leu
    210                 215                 220

Glu Gly Asp Ala Ile Asp Asp Ala Asp Asp Leu Gly Asp Leu Leu Arg
225                 230                 235                 240

Arg Gly Phe Asp Gly Arg His Gly Val Asp His Leu Ala Asp His Ala
                245                 250                 255

Ala Thr Leu Arg Gly His Thr Leu Arg Ala Asp Gly Glu Leu Val Gly
            260                 265                 270

Leu Ala Gly Met Leu Gly Val Leu Ala His Gly Gly Gly Gln Leu Leu
        275                 280                 285

His Arg Gly Arg Gly Phe Phe Gln Ile Gly Ser Leu Leu Leu Gly Thr
    290                 295                 300

Ala Arg Gln Val Ala Val Ala Arg Arg Asp Leu Thr Ser Gly Glu Gly
305                 310                 315                 320

Asp Ala Gly Arg Ala Gly Leu Asp Leu Ala Asp Asp Leu Gly Gln Leu
                325                 330                 335

Gly Asp Gly Gly Val Gly Ile Val Thr His Ala Arg Glu His Ala Leu
            340                 345                 350

Val Ile Ala Met His Ala Cys Gly Gln Val Ala Phe Gly Asp Gly Arg
        355                 360                 365

Gln Gln Leu Arg Glu Leu Ala Glu Val Val Ile Ala Asp Arg His His
    370                 375                 380

Arg Val Glu Val Phe Gln His Gln Ala Glu Ile Val Val Glu Ala Leu
385                 390                 395                 400

Arg Val Ala Thr Thr Ala Glu Val Ala Gly Gly Gly Gly Thr Gly Gln
                405                 410                 415
```

-continued

```
Leu Leu Asp Leu Gly Val Asp Ala Ala Lys Val Gly Leu Asp Arg Ile
            420                 425                 430

Asp Gly Gly Gly His His Arg Leu Leu Ala Arg Val Ala Cys Asp Val
        435                 440                 445

Thr Ala Glu Val Ala Asp Cys Val Leu Leu His Asp Leu Gln Tyr Ile
    450                 455                 460

Val Gln Gly Leu His Met Ala Thr Asp Gln Gly Val Gly Phe Leu His
465                 470                 475                 480

His Gln Pro Val Phe Ala Gly Glu Gly Ala Gly Ile Asp Ala Val Ala
                485                 490                 495

Gln Leu Ala Thr Val Val Ala Ala Gly His Phe Ala Leu Ala Ala Asp
            500                 505                 510

His Arg Val Gln Leu Leu Leu His Ala Gly His Arg Leu Gln Gln Ala
        515                 520                 525

Ala Gly Phe Ile Met Arg Leu Arg Ala Asp Met Ala Val Glu Leu Ala
    530                 535                 540

Gly Gly Asp Thr Phe Gly Asp Gly Gly Gly Leu Leu Gln Arg His Gly
545                 550                 555                 560

Asp Ala Val Ala Asp Asp Pro Ala Gln Gly Gln His Asp Gln His Gln
                565                 570                 575

His Ala Ala Gly Glu Gly His Asp Gly Gly Glu His Gln Arg Leu Leu
            580                 585                 590

Leu Asn Val Ile His Val Asp Ala Ala Ala Asp His Pro Val Pro Gly
        595                 600                 605

Cys Glu Gln Ala Gly Val Gly His Leu Leu Asp Val Ala Leu Ala Ala
    610                 615                 620

Arg Leu Gly Pro Ala Val Val Asp Glu Ala Ala Ala Gly Leu Gly Arg
625                 630                 635                 640

Leu Asp Leu Val Val Val Asp Ala His Ala Val Val Gly Ala Glu Val
                645                 650                 655

Leu His Val Leu Ala Asp Gln Val Leu Ala Glu Arg Val His Gln Gln
            660                 665                 670

Ala Ile Ala Gly Val Val Asp Val Val Val Ile Gly Val Val Leu Gly
        675                 680                 685

Ala His His Phe Gln Arg Leu Gln Gly Ser Gly Leu Gly Ser Leu Leu
    690                 695                 700

Ala Glu Leu Ala Gly Ala Gly Gln Ala Met Val Val Leu Glu Asp Ala
705                 710                 715                 720

Ile Ala Gln Phe Asp Leu Gly Leu Gln Arg Gly Leu Ala Gly Leu Gly
                725                 730                 735

Gln Val Gln Val Leu His Ala Gly Gly Asp His Arg Gln Arg Asp His
            740                 745                 750

Ala Glu Arg Asp Glu Gln Arg Gln Gln Val Glu Leu Ala Pro Asp Gly
        755                 760                 765

Glu Val Ala Glu Val Val Leu Pro Ala Met Gln Lys Ile His Arg Ala
    770                 775                 780

Gly Pro Gly
785
```

1. A method of producing a fatty acid derivative, the method comprising culturing a host cell in the presence of a carbon source, wherein the host cell is genetically engineered to overexpress a gene encoding a thioesterase, a gene encoding an acyl-CoA synthase, and a gene encoding an ester synthase, wherein the gene encoding the thioesterase and the gene encoding the acyl-CoA synthase are integrated into the genomic DNA of the host cell.

2. The method of claim **1**, further comprising isolating the fatty acid derivative.

3. The method of claim **1**, further comprising culturing the host cell in the presence of an alcohol.

4. The method of claim **1**, wherein the gene encoding a thioesterase is tesA, 'tesA, fatB, fatB2, fatB3, fatA1, or fatA.

5. The method of claim **1**, wherein the gene encoding an acyl-CoA synthase is fadD, fadK, BH3103, pfl-4354, EAV15023, fadD1, fadD2, RPC_4074, fadDD, faa39, the gene that encodes the protein of GenBank Accession No. ZP_01644875, or yhfL.

6. The method of claim **1**, wherein the gene encoding an ester synthase is obtained from *Acinetobacter* sp., *Alcanivorax borkumensis, Alcaligenes eutrophus, Mortierella alpina, Cryptococcus curvatus, Arabidopsis thaliana, Fundibacter jadensis, Pseudomonas aeruginosa, Rhodococcus opacus, Marinobacter hydrocarbonoclastics, Saccharomyces cerevisiae, Homo sapiens*, or *Simmondis chinensis*.

7. The method of claim **6**, wherein the gene encoding an ester synthase is wax/dgat, a gene encoding a wax synthase, a gene encoding a bifunctional ester synthase/acyl-CoA:diacylglycerol acyltransferase, ES9, ES8, DSM 8798, or AtfA2.

8. The method of claim **1**, wherein the host cell is genetically engineered to express, relative to a wild type host cell, a decreased level of at least one of a gene encoding an acyl-CoA dehydrogenase, a gene encoding an outer membrane protein receptor, and a gene encoding a transcriptional regulator of fatty acid biosynthesis.

9. The method of claim **8**, wherein the gene encoding the outer membrane protein receptor is a gene encoding an outer membrane ferrichrome transporter.

10. The method of claim **8**, wherein the gene encoding the transcriptional regulator of fatty acid biosynthesis encodes a DNA transcriptional repressor.

11. A method of producing a fatty acid derivative the method comprising culturing a host cell in the presence of a carbon source, wherein the host cell is genetically engineered to overexpress a gene encoding a thioesterase, a gene encoding an acyl-CoA synthase, and a gene encoding an ester synthase.

12. The method of claim **11**, wherein the host cell is genetically engineered to express, relative to a wild type host cell, a decreased level of a gene encoding a pyruvate formate lyase, a lactate dehydrogenase, or both.

13. The method of claim **1**, wherein the host cell is selected from the group consisting of a mammalian cell, plant cell, insect cell, yeast cell, fungus cell, filamentous fungi cell, cyanobacterial cell, and bacterial cell.

14. The method of claim **1**, wherein the host cell is an *E. coli* cell.

15. A fatty acid derivative produced by the method of claim **1**.

16. A fatty ester produced by the method of claim **1**.

17. A genetically engineered microorganism comprising at least one of a gene encoding a thioesterase, a gene encoding an acyl-CoA synthase, and a gene encoding an ester synthase, wherein the gene encoding the thioesterase and the gene encoding the acyl-CoA synthase are integrated into the genomic DNA of the microorganism, and wherein the microorganism produces an increased level of a fatty ester relative to a wild-type microorganism.

18. A genetically engineered microorganism comprising an exogenous control sequence stably incorporated into the genomic DNA of the microorganism upstream of a gene encoding a thioesterase, and a gene encoding an acyl-CoA synthase, wherein the microorganism produces an increased level of a fatty ester relative to a wild-type microorganism.

19. The genetically engineered microorganism of claim **18**, wherein the exogenous control sequence is a promoter.

20. The genetically engineered microorganism of claim **19**, wherein the microorganism is genetically engineered to express, relative to a wild type microorganism, a decreased level of at least one of a gene encoding an acyl-CoA dehydrogenase, a gene encoding an outer membrane protein receptor, and a gene encoding a transcriptional regulator of fatty acid biosynthesis.

21. The genetically engineered microorganism of claim **19**, wherein the microorganism is genetically engineered to express, relative to a wild type microorganism, a decreased level of at least one of a gene encoding a pyruvate formate lyase, and a gene encoding a lactate dehydrogenase.

22. The genetically engineered microorganism of claim **20**, wherein the gene encoding the outer membrane protein receptor encodes a ferrichrome outer membrane transporter.

23. The genetically engineered microorganism of claim **20**, wherein the gene encoding the transcriptional regulator encodes a DNA binding transcriptional repressor.

24. The genetically engineered microorganism of claim **17**, selected from a Gram-negative or a Gram-positive bacterium.

25. A method of producing a fatty acid derivative comprising culturing the genetically engineered microorganism of claim **17**, in the presence of an alcohol.

26. A fatty acid derivative produced by the method of claim **25**.

27. The fatty acid derivative of claim **26**, wherein the fatty acid derivative is a fatty ester.

28. A biofuel composition comprising the fatty acid derivative of claim **26**.

29. The biofuel composition of claim **28**, wherein the composition has a cloud point of about 5° C. or lower, of about 0° C. or lower, or of about −4° C. or lower.

30. The biofuel composition of claim **28**, wherein the composition has a simulated distillation T90 of about 360° C. or lower, or of about 358° C. or lower.

31. The biofuel composition of claim **28**, wherein the composition has a sulfur content of about 15 ppm or less, of about 12 ppm or less, or of about 10.7 ppm or less.

32. The biofuel composition of claim **28**, wherein the composition has a Karl Fischer moisture content of about 0.1 wt. % or less, of about 0.08 wt. % or less, or of about 0.06 wt. % or less.

33. The biofuel composition of claim **28**, wherein the composition has a total acid number of about 0.50 mg KOH/g or less, of about 0.20 mg KOH/g or less, or of about 0.15 mg KOH/g or less.

34. The biofuel composition of claim **28**, wherein the composition has a combined calcium and magnesium content of about 5 ppm or less, of about 2 ppm or less, or of about 0.5 ppm or less.

35. The biofuel composition of claim **28**, wherein the composition has a combined sodium and potassium content of about 5 ppm or less, of about 2.5 ppm or less, or of about 1.6 ppm or less.

36. The biofuel composition of claim **28**, wherein the composition has a specific mass at 20° C. of about 850 to about 900 kg/m^3, of about 860 to about 890 kg/m^3, or of about 870 to about 880 kg/m^3.

37. The biofuel composition of claim **28**, wherein the composition has a kinematic viscosity at 40° C. of about 3 to about 6 mm^2/s, of about 3.2 to about 5 mm^2/s, or of about 3.5 to about 4 mm^2/s.

38. The biofuel composition of claim **28**, wherein the composition has a water content of about 500 mg/kg or less, of about 480 mg/kg or less, or of about 450 mg/kg or less.

39. The biofuel composition of claim **28**, wherein the composition has a total contamination level of about 24 mg/kg or less, of about 20 mg/kg or less, of about 10 mg/kg or less, or of about 5 mg/kg or less.

40. The biofuel composition of claim **28**, wherein the composition has a flash point of about 100° C. or higher, of about 110° C. or higher, or of about 120° C. or higher.

41. The biofuel composition of claim **28**, wherein the composition has an ester content of about 96.5 wt. % or more, or of about 97 wt. % or more.

42. The biofuel composition of claim **28**, wherein the composition has a carbon residue number of about 0.05 wt. % or less.

43. The biofuel composition of claim **28**, wherein the composition has a sulfated ash level of 0.02 wt. % or less, or of 0.01 wt. % or less.

44. The biofuel composition of claim **28**, wherein the composition has a phosphorus level of about 10 mg/kg or less, of about 5 mg/kg or less, or of about 1 mg/kg or less.

45. The biofuel composition of claim **28**, wherein the composition has a copper corrosion score of 1.

46. The biofuel composition of claim **28**, wherein the composition has a cold filter plugging point of about 19° C. or lower, of about 10° C. or lower, or of about 0° C. or lower.

47. The biofuel composition of claim **28**, wherein the composition has a free glycerol level of about 0.02 wt. % or less.

48. The biofuel composition of claim **28**, wherein the composition has a total glycerol level of about 0.25 wt. % or less, of about 0.15 wt. % or less, of about 0.10 wt. % or less, or of about 0.05 wt. % or less.

49. The biofuel composition of claim **28**, wherein the composition has a monoacylglycerol level of about 0.02 wt. % or less.

50. The biofuel composition of claim **28**, wherein the composition has a methanol or ethanol level of about 0.2 wt. % or less, of about 0.1 wt. % or less, or of about 0.02 wt. % or less.

51. The biofuel composition of claim **28**, wherein the composition has an iodine number of about 65 g/100 g or less.

52. The biofuel composition of claim **28**, wherein the composition has an oxidation stability at 110° C. of about 6 hours or longer, of about 9 hours or longer, or of about 11 hours or longer.

* * * * *