Innovation, Sciences et Développement économique Canada
Office de la Propriété Intellectuelle du Canada

Innovation, Science and Economic Development Canada
Canadian Intellectual Property Office

CA 2740037 C 2019/06/18

(11)(21) 2 740 037

(12) BREVET CANADIEN
CANADIAN PATENT

(13) C

(86) Date de dépôt PCT/PCT Filing Date: 2009/10/07
(87) Date publication PCT/PCT Publication Date: 2010/06/03
(45) Date de délivrance/Issue Date: 2019/06/18
(85) Entrée phase nationale/National Entry: 2011/04/08
(86) N° demande PCT/PCT Application No.: US 2009/059903
(87) N° publication PCT/PCT Publication No.: 2010/062480
(30) Priorité/Priority: 2008/10/28 (US61/109,131)

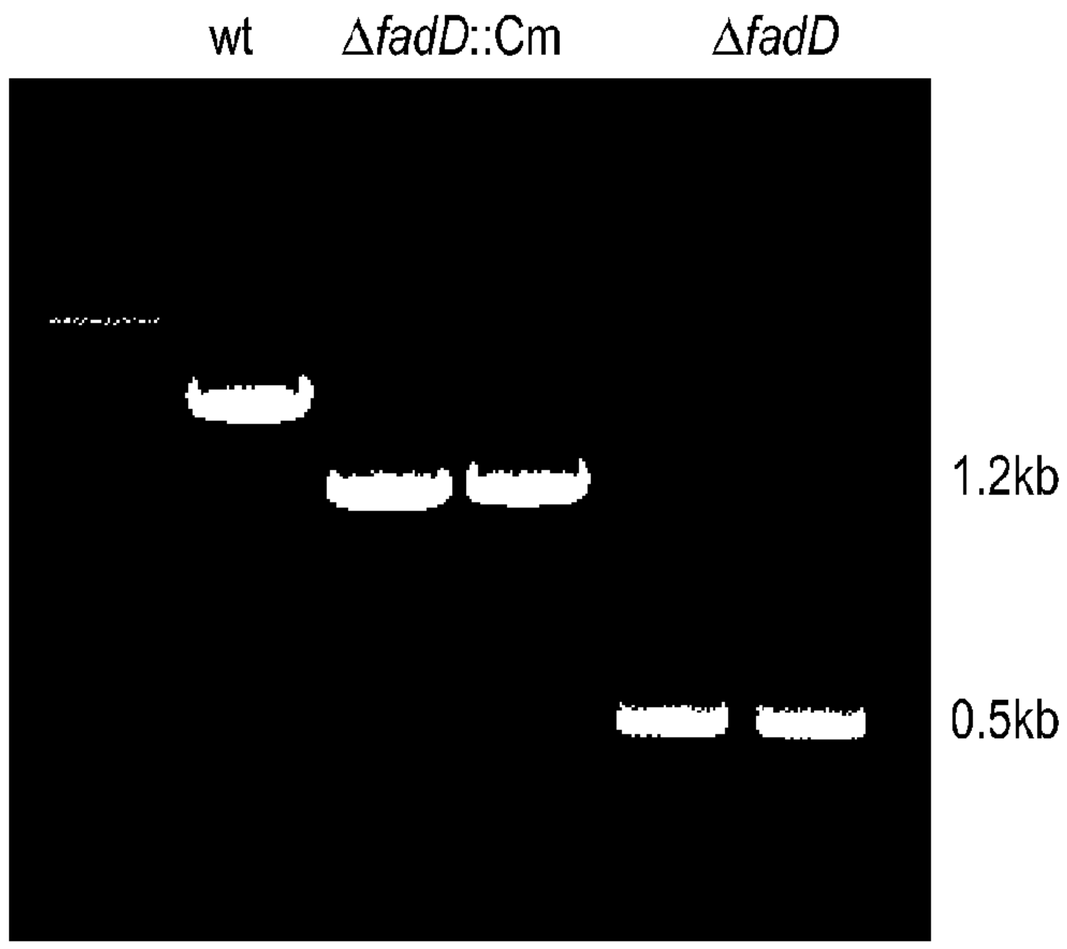
(51) Cl.Int./Int.Cl. *C12P 7/04* (2006.01), *C07C 29/00* (2006.01), *C12N 15/52* (2006.01), *C12N 15/53* (2006.01), *C12N 15/54* (2006.01), *C12N 15/55* (2006.01), *C12N 15/63* (2006.01), *C12N 9/00* (2006.01), *C12N 9/02* (2006.01), *C07C 31/125* (2006.01), *C07C 33/02* (2006.01), *C12N 9/04* (2006.01), *C12N 9/10* (2006.01), *C12N 9/16* (2006.01)

(72) Inventeur/Inventor:
HU, ZHIHAO, US

(73) Propriétaire/Owner:
REG LIFE SCIENCES, LLC, US

(74) Agent: BERESKIN & PARR LLP/S.E.N.C.R.L.,S.R.L.

(54) Titre : PROCEDES PERMETTANT DE PRODUIRE UN ALCOOL GRAS DANS UNE CELLULE HOTE
(54) Title: METHODS FOR PRODUCING A FATTY ALCOHOL IN A HOST CELL



(57) Abrégé/Abstract:
Methods and compositions for producing a fatty alcohol in a host cell are provided. In one embodiment, a method of producing a fatty alcohol in a host cell is provided, the method comprising expressing in a host cell a nucleic acid molecule encoding a polypeptide having thioesterase activity; a nucleic acid molecule encoding a polypeptide having alcohol dehydrogenase activity; and a nucleic acid molecule encoding a polypeptide having carboxylic acid reductase activity, and culturing the host cell in culture media containing a carbon source under conditions wherein a fatty alcohol is produced.

ABSTRACT

Methods and compositions for producing a fatty alcohol in a host cell are provided. In one embodiment, a method of producing a fatty alcohol in a host cell is provided, the method comprising expressing in a host cell a nucleic acid molecule encoding a polypeptide having thioesterase activity; a nucleic acid molecule encoding a polypeptide having alcohol dehydrogenase activity; and a nucleic acid molecule encoding a polypeptide having carboxylic acid reductase activity, and culturing the host cell in culture media containing a carbon source under conditions wherein a fatty alcohol is produced.

DEMANDES OU BREVETS VOLUMINEUX

LA PRÉSENTE PARTIE DE CETTE DEMANDE OU CE BREVETS COMPREND PLUS D'UN TOME.

CECI EST LE TOME __1___ DE __3 __

NOTE: Pour les tomes additionels, veillez contacter le Bureau Canadien des Brevets.

JUMBO APPLICATIONS / PATENTS

THIS SECTION OF THE APPLICATION / PATENT CONTAINS MORE THAN ONE VOLUME.

THIS IS VOLUME __1___ OF __3___

NOTE: For additional volumes please contact the Canadian Patent Office.

METHODS FOR PRODUCING A FATTY ALCOHOL IN A HOST CELL

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. Moreover, there is no guarantee that these wells will contain petroleum. It is estimated that only 40% of drilled wells lead to productive wells generating commercial hydrocarbons. 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. During primary recovery, the natural pressure underground is sufficient to extract about 20% of the petroleum in the well. As this natural pressure falls, secondary recovery methods are employed, if economical. Generally, secondary recovery involves increasing the well's pressure by, for example, water injection, natural gas injection, or gas lift. Using secondary recovery methods, an additional 5% to 15% of petroleum is recovered. Once secondary recovery methods are exhausted, tertiary recovery methods can be used, if economical. Tertiary methods involve reducing the viscosity of the petroleum to make it easier to extract. Using tertiary recovery methods, an additional 5% to 15% of petroleum is

recovered. Hence, 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. Moreover, offshore drilling involves dredging the seabed which disrupts or destroys the surrounding marine environment.

[0005] Since petroleum deposits are not found uniformly throughout the Earth, 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 longer chain hydrocarbons in crude petroleum by cracking it 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. An additional example of a raw material is propylene, which 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] These petrochemicals can then be used to make specialty chemicals, such as plastics, resins, fibers, elastomers, pharmaceuticals, lubricants, or gels. Particular specialty chemicals that can be produced from petrochemical raw materials are fatty acids, hydrocarbons (*e.g.*, long chain, branched chain, saturated, unsaturated, *etc.*), fatty alcohols, esters, fatty aldehydes, ketones, lubricants, *etc.*

[0011] Fatty alcohols have many commercial uses. Worldwide annual sales of fatty alcohols and their derivatives are in excess of US$1 billion. The shorter chain fatty alcohols are used in the cosmetic and food industries as emulsifiers, emollients, and thickeners. Due to their amphiphilic nature, fatty alcohols behave as nonionic surfactants, which are useful in personal care and household products, for example, detergents. In addition, fatty alcohols are used in waxes, gums, resins, pharmaceutical salves and lotions, lubricating oil additives, textile antistatic and finishing agents, plasticizers, cosmetics, industrial solvents, and solvents for fats.

[0012] Aldehydes are used to produce many specialty chemicals. For example, aldehydes are used to produce polymers, resins (*e.g.*, Bakelite™), dyes, flavorings, plasticizers, perfumes, pharmaceuticals, and other chemicals. Some are used as solvents, preservatives, or disinfectants. Some natural and synthetic compounds, such as vitamins and hormones, are aldehydes. In addition, many sugars contain aldehyde groups.

[0013] 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.

[0014] In addition to the problems with exploring, extracting, transporting, and refining petroleum, petroleum is a limited and dwindling resource. One estimate of 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.

[0015] 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 can lead to an increase 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.

[0016] Due to the inherent challenges posed by petroleum, there is a need for a renewable petroleum source that 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 which can be produced economically without creating 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 which are typically derived from petroleum.

[0017] One method of producing renewable petroleum is by engineering microorganisms to produce renewable petroleum products. Some microorganisms have a natural ability to produce chemicals. For example, yeast has been used for centuries to produce ethanol (*e.g.*, beer, wine, *etc.*). In recent years, through the development of advanced biotechnologies, it is possible to metabolically engineer an organism to produce bioproducts that were never previously produced. Products, such as chemicals, derived from these cellular activities are known as bioproducts. Fuels produced these cellular activities are known as biofuels. Biofuels are a renewable alternative fuel to petroleum based fuels. Biofuels can be substituted for any petroleum based fuel (*e.g.*, gasoline, diesel, aviation fuel, heating oil, *etc.*). Biofuels can be derived from renewable sources, such as plant matter, animal matter, or even waste products. These renewable sources are collectively known as biomass. One advantage of biofuels over petroleum based fuels is that they do not require expensive and risky exploration or extraction. In addition, biofuels can be locally produced. Hence, they do not require transportation over long distances. Moreover, biofuels can be made directly without the need for expensive and energy intensive

refining as is needed with refining crude petroleum. In other circumstances, the biofuel may require a limited and cost-effective level of refining. Furthermore, the use of biofuels improves the environment by reducing the amount of environmentally harmful emissions (*e.g.*, green house gases, air pollution, *etc.*) released during combustion. For example, biofuels maintain a balanced carbon cycle because biofuels are produced from biomass, a renewable, natural resource. While the burning of biofuels will release carbon (*e.g.*, as carbon dioxide), this carbon will be recycled during the production of biomass (*e.g.*, the cultivation of crops), thereby balancing the carbon cycle unlike petroleum based fuels.

[0018] For similar reasons, biologically derived chemicals offer the same advantages as biofuels over petroleum based fuels. Biologically derived chemicals are a renewable alternative to petrochemicals. Biologically derived chemicals, such as hydrocarbons (*e.g.*, alkanes, alkenes, or alkynes), fatty alcohols, esters, fatty acids, fatty aldehydes, and ketones are superior to petrochemicals because they are produced directly without extensive refining. Unlike petrochemicals, biologically derived chemicals do not need to be refined like crude petroleum to recover raw materials which must then be further processed to make more complex petrochemicals. Biologically derived chemicals are directly converted from biomass to the desired chemical product.

SUMMARY OF THE INVENTION

[0019] The invention is based, at least in part, on the identification of genes that encode fatty aldehyde biosynthetic polypeptides and fatty alcohol biosynthetic polypeptides, which can be used to produce fatty aldehydes that can subsequently be converted into fatty alcohols. Accordingly, in one aspect, the invention features a method of making a fatty alcohol. The method includes expressing in a host cell a gene encoding a fatty aldehyde biosynthetic polypeptide comprising the amino acid sequence of SEQ ID NO:18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 264, 266, 268, 270, or 272, or a variant thereof. In some embodiments, the method further includes isolating the fatty alcohol from the host cell. In some embodiments, the fatty alcohol is present in the extracellular environment. In certain embodiments, the fatty

alcohol is isolated from the extracellular environment of the host cell. In some embodiments, the fatty alcohol is secreted from the host cell. In alternative embodiments, the fatty alcohol is transported into the extracellular environment. In other embodiments, the fatty alcohol is passively transported into the extracellular environment.

[0020] In some embodiments, the fatty aldehyde biosynthetic polypeptide comprises the amino acid sequence of SEQ ID NO:18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 264, 266, 268, 270, or 272 with one or more amino acid substitutions, additions, insertions, or deletions, and the polypeptide has carboxylic acid reductase activity. In some embodiments, the polypeptide has fatty acid reductase activity.

[0021] In some embodiments, the polypeptide comprises one or more of the following conservative amino acid substitutions: replacement of an aliphatic amino acid, such as alanine, valine, leucine, and isoleucine, with another aliphatic amino acid; replacement of a serine with a threonine; replacement of a threonine with a serine; replacement of an acidic residue, such as aspartic acid and glutamic acid, with another acidic residue; replacement of a residue bearing an amide group, such as asparagine and glutamine, with another residue bearing an amide group; exchange of a basic residue, such as lysine and arginine, with another basic residue; and replacement of an aromatic residue, such as phenylalanine and tyrosine, with another aromatic residue. In some embodiments, the polypeptide has about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 30, 40, 50, 60, 70, 80, 90, 100, or more amino acid substitutions, additions, insertions, or deletions. In some embodiments, the polypeptide has carboxylic acid reductase activity. In some embodiments, the polypeptide has fatty acid reductase activity.

[0022] In some embodiments, the method further includes modifying the expression of a gene encoding a fatty acid synthase in the host cell. In certain embodiments, modifying the expression of a gene encoding a fatty acid synthase includes expressing a gene encoding a fatty acid synthase in the host cell and/or increasing the expression or activity of an endogenous fatty acid synthase in the host cell. In alternate embodiments, modifying the expression of a gene encoding a fatty

acid synthase includes attenuating a gene encoding a fatty acid synthase in the host cell and/or decreasing the expression or activity of an endogenous fatty acid synthase in the host cell. In some embodiments, the fatty acid synthase is a thioesterase. In particular embodiments, the thioesterase is encoded by *tesA*, *tesA* without leader sequence, *tesB*, *fatB*, *fatB2*, *fatB3*, *fatA*, *or fatA1*.

[0023] In some embodiments, the method further includes expressing a gene encoding a fatty alcohol biosynthetic polypeptide in the host cell. In particular embodiments, the fatty alcohol biosynthetic polypeptide comprises the amino acid sequence of SEQ ID NO:94, 96, 98, 100, 102, 104, 106, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, or 194, or a variant thereof.

[0024] In other embodiments, the host cell is genetically engineered to express an attenuated level of a fatty acid degradation enzyme relative to a wild type host cell. In some embodiments, the host cell is genetically engineered to express an attenuated level of an acyl-CoA synthase relative to a wild type host cell. In particular embodiments, the host cell expresses an attenuated level of an acyl-CoA synthase encoded by *fadD*, *fadK*, *BH3103*, *yhfL*, *Pfl-4354*, *EAV15023*, *fadD1*, *fadD2*, *RPC_4074*, *fadDD35*, *fadDD22*, *faa3p* or the gene encoding the protein ZP_01644857. In certain embodiments, the genetically engineered host cell comprises a knockout of one or more genes encoding a fatty acid degradation enzyme, such as the aforementioned acyl-CoA synthase genes.

[0025] In yet other embodiments, the host cell is genetically engineered to express an attenuated level of a dehydratase/isomerase enzyme, such as an enzyme encoded by *fabA* or by a gene listed in Figure 15. In some embodiments, the host cell comprises a knockout of *fabA* or a gene listed in Figure 15. In other embodiments, the host cell is genetically engineered to express an attenuated level of a ketoacyl-ACP synthase, such as an enzyme encoded by *fabB* or by a gene listed in Figure 16. In certain embodiments, the host cell comprises a knockout of *fabB* or a gene listed in Figure 16. In yet other embodiments, the host cell is genetically engineered to express a modified level of a gene encoding a desaturase enzyme, such as *desA*.

[0026] In some embodiments, the polypeptide is from a bacterium, a plant, an insect, a yeast, a fungus, or a mammal.

[0027] In certain embodiments, the polypeptide is from a mammalian cell, plant cell, insect cell, yeast cell, fungus cell, filamentous fungi cell, bacterial cell, or any other organism described herein. In some embodiments, the bacterium is a mycobacterium selected from the group consisting of *Mycobacterium smegmatis*, *Mycobacterium abscessus*, *Mycobacterium avium*, *Mycobacterium bovis*, *Mycobacterium tuberculosis*, *Mycobacterium leprae*, *Mycobacterium marinum*, and *Mycobacterium ulcerans*. In other embodiments, the bacterium is *Nocardia sp.* NRRL 5646, *Nocardia farcinica*, *Streptomyces griseus*, *Salinispora arenicola*, or *Clavibacter michiganenesis*.

[0028] In some embodiments, the method further includes culturing the host cell in the presence of at least one biological substrate for the fatty aldehyde biosynthetic polypeptide.

[0029] In another aspect, the invention features a method of producing a fatty alcohol. The method includes expressing in a host cell a gene encoding a fatty aldehyde biosynthetic polypeptide comprising an amino acid sequence having at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% sequence identity to the amino acid sequence of SEQ ID NO:18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 264, 266, 268, 270, or 272. In some embodiments, the amino acid sequence is the amino acid sequence of SEQ ID NO:18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 264, 266, 268, 270, or 272.

[0030] In some embodiments, the method further includes isolating the fatty alcohol from the host cell. In some embodiments, the fatty alcohol is present in the extracellular environment. In certain embodiments, the fatty alcohol is isolated from the extracellular environment of the host cell. In some embodiments, the fatty alcohol is secreted from the host cell. In alternative embodiments, the fatty alcohol is

transported into the extracellular environment. In other embodiments, the fatty alcohol is passively transported into the extracellular environment.

[0031] In some embodiments, the method further includes modifying the expression of a gene encoding a fatty acid synthase in the host cell. In certain embodiments, modifying the expression of a gene encoding a fatty acid synthase includes expressing a gene encoding a fatty acid synthase in the host cell and/or increasing the expression or activity of an endogenous fatty acid synthase in the host cell. In alternate embodiments, modifying the expression of a gene encoding a fatty acid synthase includes attenuating a gene encoding a fatty acid synthase in the host cell and/or decreasing the expression or activity of an endogenous fatty acid synthase in the host cell. In some embodiments, the fatty acid synthase is a thioesterase. In particular embodiments, the thioesterase is encoded by *tesA*, *tesA* without leader sequence, *tesB*, *fatB*, *fatB2*, *fatB3*, *fatA*, *or fatA1*.

[0032] In some embodiments, the method further includes expressing a gene encoding a fatty alcohol biosynthetic polypeptide in the host cell. In particular embodiments, the fatty alcohol biosynthetic polypeptide comprises the amino acid sequence of SEQ ID NO:94, 96, 98, 100, 102, 104, 106, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, or 194, or a variant thereof

[0033] In other embodiments, the host cell is genetically engineered to express an attenuated level of a fatty acid degradation enzyme relative to a wild type host cell. In some embodiments, the host cell is genetically engineered to express an attenuated level of an acyl-CoA synthase relative to a wild type host cell. In particular embodiments, the host cell expresses an attenuated level of an acyl-CoA synthase encoded by *fadD*, *fadK*, *BH3103*, *yhfL*, *Pfl-4354*, *EAV15023*, *fadD1*, *fadD2*, *RPC_4074*, *fadDD35*, *fadDD22*, *faa3p* or the gene encoding the protein ZP_01644857. In certain embodiments, the genetically engineered host cell comprises a knockout of one or more genes encoding a fatty acid degradation enzyme, such as the aforementioned acyl-CoA synthase genes.

[0034] In yet other embodiments, the host cell is genetically engineered to express an attenuated level of a dehydratase/isomerase enzyme, such as an enzyme encoded

by *fabA* or by a gene listed in Figure 15. In some embodiments, the host cell comprises a knockout of *fabA* or a gene listed in Figure 15. In other embodiments, the host cell is genetically engineered to express an attenuated level of a ketoacyl-ACP synthase, such as an enzyme encoded by *fabB* or by a gene listed in Figure 16. In certain embodiments, the host cell comprises a knockout of *fabB* or a gene listed in Figure 16. In yet other embodiments, the host cell is genetically engineered to express a modified level of a gene encoding a desaturase enzyme, such as *desA*.

[0035] In some embodiments, the polypeptide is from a bacterium, a plant, an insect, a yeast, a fungus, or a mammal.

[0036] In certain embodiments, the polypeptide is from a mammalian cell, plant cell, insect cell, yeast cell, fungus cell, filamentous fungi cell, bacterial cell, or any other organism described herein. In some embodiments, the bacterium is a mycobacterium selected from the group consisting of *Mycobacterium smegmatis*, *Mycobacterium abscessus*, *Mycobacterium avium*, *Mycobacterium bovis*, *Mycobacterium tuberculosis*, *Mycobacterium leprae*, *Mycobacterium marinum*, and *Mycobacterium ulcerans*. In other embodiments, the bacterium is *Nocardia sp.* NRRL 5646, *Nocardia farcinica*, *Streptomyces griseus*, *Salinispora arenicola*, or *Clavibacter michiganenesis*.

[0037] In some embodiments, the method further includes culturing the host cell in the presence of at least one biological substrate for the fatty aldehyde biosynthetic polypeptide.

[0038] In another aspect, the invention features a method of producing a fatty alcohol. The method includes expressing in a host cell a polynucleotide that hybridizes to a complement of the nucleotide sequence of SEQ ID NO:17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 263, 265, 267, 269, or 271, or to a fragment thereof, wherein the polynucleotide encodes a polypeptide having carboxylic acid reductase activity. In some embodiments, the polypeptide has fatty acid reductase activity.

[0039] In some embodiments, the method further includes isolating the fatty alcohol from the host cell. In some embodiments, the fatty alcohol is present in the extracellular environment. In certain embodiments, the fatty alcohol is isolated from

the extracellular environment of the host cell. In some embodiments, the fatty alcohol is secreted from the host cell. In alternative embodiments, the fatty alcohol is transported into the extracellular environment. In other embodiments, the fatty alcohol is passively transported into the extracellular environment.

[0040] In some embodiments, the polynucleotide hybridizes under low stringency, medium stringency, high stringency, or very high stringency conditions, to a complement of the nucleotide sequence of SEQ ID NO:17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 263, 265, 267, 269, or 271, or to a fragment thereof.

[0041] In some embodiments, the method further includes modifying the expression of a gene encoding a fatty acid synthase in the host cell. In certain embodiments, modifying the expression of a gene encoding a fatty acid synthase includes expressing a gene encoding a fatty acid synthase in the host cell and/or increasing the expression or activity of an endogenous fatty acid synthase in the host cell. In alternate embodiments, modifying the expression of a gene encoding a fatty acid synthase includes attenuating a gene encoding a fatty acid synthase in the host cell and/or decreasing the expression or activity of an endogenous fatty acid synthase in the host cell. In some embodiments, the fatty acid synthase is a thioesterase. In particular embodiments, the thioesterase is encoded by *tesA*, *tesA* without leader sequence, *tesB*, *fatB*, *fatB2*, *fatB3*, *fatA*, *or fatA1*.

[0042] In some embodiments, the method further includes expressing a gene encoding a fatty alcohol biosynthetic polypeptide in the host cell. In particular embodiments, the fatty alcohol biosynthetic polypeptide comprises the amino acid sequence of SEQ ID NO:94, 96, 98, 100, 102, 104, 106, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, or 194, or a variant thereof.

[0043] In other embodiments, the host cell is genetically engineered to express an attenuated level of a fatty acid degradation enzyme relative to a wild type host cell. In some embodiments, the host cell is genetically engineered to express an attenuated level of an acyl-CoA synthase relative to a wild type host cell. In particular embodiments, the host cell expresses an attenuated level of an acyl-CoA synthase

encoded by *fadD*, *fadK*, *BH3103*, *yhfL*, *Pfl-4354*, *EAV15023*, *fadD1*, *fadD2*, *RPC_4074*, *fadDD35*, *fadDD22*, *faa3p* or the gene encoding the protein ZP_01644857. In certain embodiments, the genetically engineered host cell comprises a knockout of one or more genes encoding a fatty acid degradation enzyme, such as the aforementioned acyl-CoA synthase genes.

[0044] In yet other embodiments, the host cell is genetically engineered to express an attenuated level of a dehydratase/isomerase enzyme, such as an enzyme encoded by *fabA* or by a gene listed in Figure 15. In some embodiments, the host cell comprises a knockout of *fabA* or a gene listed in Figure 15. In other embodiments, the host cell is genetically engineered to express an attenuated level of a ketoacyl-ACP synthase, such as an enzyme encoded by *fabB* or by a gene listed in Figure 16. In certain embodiments, the host cell comprises a knockout of *fabB* or a gene listed in Figure 16. In yet other embodiments, the host cell is genetically engineered to express a modified level of a gene encoding a desaturase enzyme, such as *desA*.

[0045] In some embodiments, the polynucleotide is from a bacterium, a plant, an insect, a yeast, a fungus, or a mammal.

[0046] In certain embodiments, the polypeptide is from a mammalian cell, plant cell, insect cell, yeast cell, fungus cell, filamentous fungi cell, bacterial cell, or any other organism described herein. In some embodiments, the bacterium is a mycobacterium selected from the group consisting of *Mycobacterium smegmatis*, *Mycobacterium abscessus*, *Mycobacterium avium*, *Mycobacterium bovis*, *Mycobacterium tuberculosis*, *Mycobacterium leprae*, *Mycobacterium marinum*, and *Mycobacterium ulcerans*. In other embodiments, the bacterium is *Nocardia sp.* NRRL 5646, *Nocardia farcinica*, *Streptomyces griseus*, *Salinispora arenicola*, or *Clavibacter michiganenesis*.

[0047] In some embodiments, the method further includes culturing the host cell in the presence of at least one biological substrate for the fatty aldehyde biosynthetic polypeptide.

[0048] In another aspect, the invention features a method of producing a fatty alcohol. The method comprises expressing in a host cell a gene encoding a fatty aldehyde biosynthetic polypeptide comprising the amino acid of SEQ ID NO:16, or a variant thereof. In some embodiments, the method further includes isolating the fatty

alcohol from the host cell. In some embodiments, the fatty alcohol is present in the extracellular environment. In certain embodiments, the fatty alcohol is isolated from the extracellular environment of the host cell. In some embodiments, the fatty alcohol is secreted from the host cell. In alternative embodiments, the fatty alcohol is transported into the extracellular environment. In other embodiments, the fatty alcohol is passively transported into the extracellular environment.

[0049] In some embodiments, the polypeptide comprises the amino acid sequence of SEQ ID NO:16 with one or more amino acid substitutions, additions, insertions, or deletions, wherein the polypeptide has carboxylic acid reductase activity. In some embodiments, the polypeptide has fatty acid reductase activity.

[0050] In some embodiments, the polypeptide comprises one or more of the following conservative amino acid substitutions: replacement of an aliphatic amino acid, such as alanine, valine, leucine, and isoleucine, with another aliphatic amino acid; replacement of a serine with a threonine; replacement of a threonine with a serine; replacement of an acidic residue, such as aspartic acid and glutamic acid, with another acidic residue; replacement of a residue bearing an amide group, such as asparagine and glutamine, with another residue bearing an amide group; exchange of a basic residue, such as lysine and arginine, with another basic residue; and replacement of an aromatic residue, such as phenylalanine and tyrosine, with another aromatic residue. In some embodiments, the polypeptide has about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 30, 40, 50, 60, 70, 80, 90, 100, or more amino acid substitutions, additions, insertions, or deletions. In some embodiments, the polypeptide has carboxylic acid reductase activity. In some embodiments, the polypeptide has fatty acid reductase activity.

[0051] In some embodiments, the method further includes modifying the expression of a gene encoding a fatty acid synthase in the host cell. In certain embodiments, modifying the expression of a gene encoding a fatty acid synthase includes expressing a gene encoding a fatty acid synthase in the host cell and/or increasing the expression or activity of an endogenous fatty acid synthase in the host cell. In alternate embodiments, modifying the expression of a gene encoding a fatty acid synthase includes attenuating a gene encoding a fatty acid synthase in the host cell and/or decreasing the expression or activity of an endogenous fatty acid synthase

in the host cell. In some embodiments, the fatty acid synthase is a thioesterase. In particular embodiments, the thioesterase is encoded by *tesA*, *tesA* without leader sequence, *tesB*, *fatB*, *fatB2*, *fatB3*, *fatA*, or *fatA1*.

[0052] In some embodiments, the method further includes expressing a gene encoding a fatty alcohol biosynthetic polypeptide in the host cell. In particular embodiments, the fatty alcohol biosynthetic polypeptide comprises the amino acid sequence of SEQ ID NO:94, 96, 98, 100, 102, 104, 106, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, or 194, or a variant thereof

[0053] In other embodiments, the host cell is genetically engineered to express an attenuated level of a fatty acid degradation enzyme relative to a wild type host cell. In some embodiments, the host cell is genetically engineered to express an attenuated level of an acyl-CoA synthase relative to a wild type host cell. In particular embodiments, the host cell expresses an attenuated level of an acyl-CoA synthase encoded by *fadD*, *fadK*, *BH3103*, *yhfL*, *Pfl-4354*, *EAV15023*, *fadD1*, *fadD2*, *RPC_4074*, *fadDD35*, *fadDD22*, *faa3p* or the gene encoding the protein ZP_01644857. In certain embodiments, the genetically engineered host cell comprises a knockout of one or more genes encoding a fatty acid degradation enzyme, such as the aforementioned acyl-CoA synthase genes.

[0054] In yet other embodiments, the host cell is genetically engineered to express an attenuated level of a dehydratase/isomerase enzyme, such as an enzyme encoded by *fabA* or by a gene listed in Figure 15. In some embodiments, the host cell comprises a knockout of *fabA* or a gene listed in Figure 15. In other embodiments, the host cell is genetically engineered to express an attenuated level of a ketoacyl-ACP synthase, such as an enzyme encoded by *fabB* or by a gene listed in Figure 16. In certain embodiments, the host cell comprises a knockout of *fabB* or a gene listed in Figure 16. In yet other embodiments, the host cell is genetically engineered to express a modified level of a gene encoding a desaturase enzyme, such as *desA*.

[0055] In some embodiments, the method further includes culturing the host cell in the presence of at least one biological substrate for the fatty aldehyde biosynthetic polypeptide.

[0056] In another aspect, the invention features a method of producing a fatty alcohol. The method includes expressing in a host cell a gene encoding a fatty aldehyde biosynthetic polypeptide comprising an amino acid sequence having at least about 70% sequence identity to the amino acid sequence of SEQ ID NO:16.

[0057] In some embodiments, the method further includes isolating the fatty alcohol from the host cell. In certain embodiments, the fatty alcohol is isolated from the extracellular environment of the host cell. In some embodiments, the fatty alcohol is secreted from the host cell.

[0058] In some embodiments, the amino acid sequence has at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% sequence identity to the amino acid sequence of SEQ ID NO:16. In some embodiments, the amino acid sequence is SEQ ID NO:16.

[0059] In some embodiments, the method further includes modifying the expression of a gene encoding a fatty acid synthase in the host cell. In certain embodiments, modifying the expression of a gene encoding a fatty acid synthase includes expressing a gene encoding a fatty acid synthase in the host cell and/or increasing the expression or activity of an endogenous fatty acid synthase in the host cell. In alternate embodiments, modifying the expression of a gene encoding a fatty acid synthase includes attenuating a gene encoding a fatty acid synthase in the host cell and/or decreasing the expression or activity of an endogenous fatty acid synthase in the host cell. In some embodiments, the fatty acid synthase is a thioesterase. In particular embodiments, the thioesterase is encoded by *tesA*, *tesA* without leader sequence, *tesB*, *fatB*, *fatB2*, *fatB3*, *fatA*, *or fatA1*.

[0060] In some embodiments, the method further includes expressing a gene encoding a fatty alcohol biosynthetic polypeptide in the host cell. In particular embodiments, the fatty alcohol biosynthetic polypeptide comprises the amino acid sequence of SEQ ID NO:94, 96, 98, 100, 102, 104, 106, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, or 194, or a variant thereof.

[0061] In other embodiments, the host cell is genetically engineered to express an attenuated level of a fatty acid degradation enzyme relative to a wild type host cell. In some embodiments, the host cell is genetically engineered to express an attenuated level of an acyl-CoA synthase relative to a wild type host cell. In particular embodiments, the host cell expresses an attenuated level of an acyl-CoA synthase encoded by *fadD*, *fadK*, *BH3103*, *yhfL*, *Pfl-4354*, *EAV15023*, *fadD1*, *fadD2*, *RPC_4074*, *fadDD35*, *fadDD22*, *faa3p* or the gene encoding the protein ZP_01644857. In certain embodiments, the genetically engineered host cell comprises a knockout of one or more genes encoding a fatty acid degradation enzyme, such as the aforementioned acyl-CoA synthase genes.

[0062] In yet other embodiments, the host cell is genetically engineered to express an attenuated level of a dehydratase/isomerase enzyme, such as an enzyme encoded by *fabA* or by a gene listed in Figure 15. In some embodiments, the host cell comprises a knockout of *fabA* or a gene listed in Figure 15. In other embodiments, the host cell is genetically engineered to express an attenuated level of a ketoacyl-ACP synthase, such as an enzyme encoded by *fabB* or by a gene listed in Figure 16. In certain embodiments, the host cell comprises a knockout of *fabB* or a gene listed in Figure 16. In yet other embodiments, the host cell is genetically engineered to express a modified level of a gene encoding a desaturase enzyme, such as *desA*.

[0063] In some embodiments, the method further includes culturing the host cell in the presence of at least one biological substrate for the fatty aldehyde biosynthetic polypeptide.

[0064] In another aspect, the invention features a method of producing a fatty alcohol. The method includes expressing in a host cell a polynucleotide that hybridizes to a complement of the nucleotide sequence of SEQ ID NO:15, or to a fragment thereof, wherein the polynucleotide encodes a polypeptide having carboxylic acid reductase activity. In some embodiments, the polypeptide has fatty acid reductase activity.

[0065] In some embodiments, the method further includes modifying the expression of a gene encoding a fatty acid synthase in the host cell. In certain embodiments, modifying the expression of a gene encoding a fatty acid synthase includes expressing a gene encoding a fatty acid synthase in the host cell and/or

increasing the expression or activity of an endogenous fatty acid synthase in the host cell. In alternate embodiments, modifying the expression of a gene encoding a fatty acid synthase includes attenuating a gene encoding a fatty acid synthase in the host cell and/or decreasing the expression or activity of an endogenous fatty acid synthase in the host cell. In some embodiments, the fatty acid synthase is a thioesterase. In particular embodiments, the thioesterase is encoded by *tesA*, *tesA* without leader sequence, *tesB*, *fatB*, *fatB2*, *fatB3*, *fatA*, *or fatA1*.

[0066] In some embodiments, the method further includes expressing a gene encoding a fatty alcohol biosynthetic polypeptide in the host cell. In particular embodiments, the fatty alcohol biosynthetic polypeptide comprises the amino acid sequence of SEQ ID NO:94, 96, 98, 100, 102, 104, 106, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, or 194, or a variant thereof.

[0067] In other embodiments, the host cell is genetically engineered to express an attenuated level of a fatty acid degradation enzyme relative to a wild type host cell. In some embodiments, the host cell is genetically engineered to express an attenuated level of an acyl-CoA synthase relative to a wild type host cell. In particular embodiments, the host cell expresses an attenuated level of an acyl-CoA synthase encoded by *fadD*, *fadK*, *BH3103*, *yhfL*, *Pfl-4354*, *EAV15023*, *fadD1*, *fadD2*, *RPC_4074*, *fadDD35*, *fadDD22*, *faa3p* or the gene encoding the protein ZP_01644857. In certain embodiments, the genetically engineered host cell comprises a knockout of one or more genes encoding a fatty acid degradation enzyme, such as the aforementioned acyl-CoA synthase genes.

[0068] In yet other embodiments, the host cell is genetically engineered to express an attenuated level of a dehydratase/isomerase enzyme, such as an enzyme encoded by *fabA* or by a gene listed in Figure 15. In some embodiments, the host cell comprises a knockout of *fabA* or a gene listed in Figure 15. In other embodiments, the host cell is genetically engineered to express an attenuated level of a ketoacyl-ACP synthase, such as an enzyme encoded by *fabB* or by a gene listed in Figure 16. In certain embodiments, the host cell comprises a knockout of *fabB* or a gene listed in

Figure 16. In yet other embodiments, the host cell is genetically engineered to express a modified level of a gene encoding a desaturase enzyme, such as *desA*.

[0069] In some embodiments, the method further includes isolating the fatty alcohol from the host cell. In some embodiments, the fatty alcohol is present in the extracellular environment. In certain embodiments, the fatty alcohol is isolated from the extracellular environment of the host cell. In some embodiments, the fatty alcohol is secreted from the host cell. In alternative embodiments, the fatty alcohol is transported into the extracellular environment. In other embodiments, the fatty alcohol is passively transported into the extracellular environment.

[0070] In some embodiments, the polynucleotide hybridizes under low stringency, medium stringency, high stringency, or very high stringency conditions, to a complement of the nucleotide sequence of SEQ ID NO:15, or to a fragment thereof.

[0071] In some embodiments, the method further includes culturing the host cell in the presence of at least one biological substrate for the fatty aldehyde biosynthetic polypeptide.

[0072] In another aspect, the invention features a method of producing a fatty alcohol. The method includes expressing in a host cell a recombinant vector comprising a fatty aldehyde biosynthetic nucleotide sequence having at least about 70% sequence identity to a nucleotide sequence listed in Figure 8. In some embodiments, the nucleotide sequence has at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% sequence identity to the nucleotide sequence of SEQ ID NO:17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 263, 265, 267, 269, or 271. In some embodiments, the nucleotide sequence is SEQ ID NO:17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 263, 265, 267, 269, or 271.

[0073] In some embodiments, the method further includes isolating the fatty alcohol from the host cell. In some embodiments, the fatty alcohol is present in the extracellular environment. In certain embodiments, the fatty alcohol is isolated from the extracellular environment of the host cell. In some embodiments, the fatty alcohol

is secreted from the host cell. In alternative embodiments, the fatty alcohol is transported into the extracellular environment. In other embodiments, the fatty alcohol is passively transported into the extracellular environment.

[0074] In some embodiments, the recombinant vector further comprises a promoter operably linked to the nucleotide sequence. In certain embodiments, the promoter is a developmentally-regulated, an organelle-specific, a tissue-specific, an inducible, a constitutive, or a cell-specific promoter.

[0075] In other embodiments, the recombinant vector comprises at least one sequence selected from the group consisting of (a) a regulatory sequence operatively coupled to the nucleotide sequence; (b) a selection marker operatively coupled to the nucleotide sequence; (c) a marker sequence operatively coupled to the nucleotide sequence; (d) a purification moiety operatively coupled to the nucleotide sequence; (e) a secretion sequence operatively coupled to the nucleotide sequence; and (f) a targeting sequence operatively coupled to the nucleotide sequence.

[0076] In some embodiments, the recombinant vector is a plasmid.

[0077] In some embodiments, the host cell expresses a polypeptide encoded by the recombinant vector. In some embodiments, the nucleotide sequence is stably incorporated into the genomic DNA of the host cell, and the expression of the nucleotide sequence is under the control of a regulated promoter region.

[0078] In some embodiments, the method further includes modifying the expression of a gene encoding a fatty acid synthase in the host cell. In certain embodiments, modifying the expression of a gene encoding a fatty acid synthase includes expressing a gene encoding a fatty acid synthase in the host cell and/or increasing the expression or activity of an endogenous fatty acid synthase in the host cell. In alternate embodiments, modifying the expression of a gene encoding a fatty acid synthase includes attenuating a gene encoding a fatty acid synthase in the host cell and/or decreasing the expression or activity of an endogenous fatty acid synthase in the host cell. In some embodiments, the fatty acid synthase is a thioesterase. In particular embodiments, the thioesterase is encoded by *tesA*, *tesA* without leader sequence, *tesB*, *fatB*, *fatB2*, *fatB3*, *fatA*, or *fatA1*.

[0079] In some embodiments, the method further includes expressing a gene encoding a fatty alcohol biosynthetic polypeptide in the host cell. In particular

embodiments, the fatty alcohol biosynthetic polypeptide comprises the amino acid sequence of SEQ ID NO:94, 96, 98, 100, 102, 104, 106, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, or 194, or a variant thereof.

[0080] In other embodiments, the host cell is genetically engineered to express an attenuated level of a fatty acid degradation enzyme relative to a wild type host cell. In some embodiments, the host cell is genetically engineered to express an attenuated level of an acyl-CoA synthase relative to a wild type host cell. In particular embodiments, the host cell expresses an attenuated level of an acyl-CoA synthase encoded by *fadD*, *fadK*, *BH3103*, *yhfL*, *Pfl-4354*, *EAV15023*, *fadD1*, *fadD2*, *RPC_4074*, *fadDD35*, *fadDD22*, *faa3p* or the gene encoding the protein ZP_01644857. In certain embodiments, the genetically engineered host cell comprises a knockout of one or more genes encoding a fatty acid degradation enzyme, such as the aforementioned acyl-CoA synthase genes.

[0081] In yet other embodiments, the host cell is genetically engineered to express an attenuated level of a dehydratase/isomerase enzyme, such as an enzyme encoded by *fabA* or by a gene listed in Figure 15. In some embodiments, the host cell comprises a knockout of *fabA* or a gene listed in Figure 15. In other embodiments, the host cell is genetically engineered to express an attenuated level of a ketoacyl-ACP synthase, such as an enzyme encoded by *fabB* or by a gene listed in Figure 16. In certain embodiments, the host cell comprises a knockout of *fabB* or a gene listed in Figure 16. In yet other embodiments, the host cell is genetically engineered to express a modified level of a gene encoding a desaturase enzyme, such as *desA*.

[0082] In some embodiments, the method further includes culturing the host cell in the presence of at least one biological substrate for a fatty aldehyde biosynthetic polypeptide.

[0083] In another aspect, the invention features a method of producing a fatty alcohol. The method includes expressing in a host cell a recombinant vector comprising a fatty aldehyde biosynthetic nucleotide sequence having at least about 70% sequence identity to the nucleotide sequence of SEQ ID NO:15.

[0084] In some embodiments, the method further includes isolating the fatty alcohol from the host cell. In some embodiments, the fatty alcohol is present in the extracellular environment. In certain embodiments, the fatty alcohol is isolated from the extracellular environment of the host cell. In some embodiments, the fatty alcohol is secreted from the host cell. In alternative embodiments, the fatty alcohol is transported into the extracellular environment. In other embodiments, the fatty alcohol is passively transported into the extracellular environment.

[0085] In some embodiments, the nucleotide sequence has at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% sequence identity to the nucleotide sequence of SEQ ID NO:15. In some embodiments, the nucleotide sequence is the nucleotide sequence of SEQ ID NO:15.

[0086] In some embodiments, the method further includes modifying the expression of a gene encoding a fatty acid synthase in the host cell. In certain embodiments, modifying the expression of a gene encoding a fatty acid synthase includes expressing a gene encoding a fatty acid synthase in the host cell and/or increasing the expression or activity of an endogenous fatty acid synthase in the host cell. In alternate embodiments, modifying the expression of a gene encoding a fatty acid synthase includes attenuating a gene encoding a fatty acid synthase in the host cell and/or decreasing the expression or activity of an endogenous fatty acid synthase in the host cell. In some embodiments, the fatty acid synthase is a thioesterase. In particular embodiments, the thioesterase is encoded by *tesA*, *tesA* without leader sequence, *tesB*, *fatB*, *fatB2*, *fatB3*, *fatA*, *or fatA1*.

[0087] In some embodiments, the method further includes expressing a gene encoding a fatty alcohol biosynthetic polypeptide in the host cell. In particular embodiments, the fatty alcohol biosynthetic polypeptide comprises the amino acid sequence of SEQ ID NO:94, 96, 98, 100, 102, 104, 106, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, or 194, or a variant thereof.

[0088] In other embodiments, the host cell is genetically engineered to express an attenuated level of a fatty acid degradation enzyme relative to a wild type host cell. In some embodiments, the host cell is genetically engineered to express an attenuated level of an acyl-CoA synthase relative to a wild type host cell. In particular embodiments, the host cell expresses an attenuated level of an acyl-CoA synthase encoded by *fadD*, *fadK*, *BH3103*, *yhfL*, *Pfl-4354*, *EAV15023*, *fadD1*, *fadD2*, *RPC_4074*, *fadDD35*, *fadDD22*, *faa3p* or the gene encoding the protein ZP_01644857. In certain embodiments, the genetically engineered host cell comprises a knockout of one or more genes encoding a fatty acid degradation enzyme, such as the aforementioned acyl-CoA synthase genes.

[0089] In yet other embodiments, the host cell is genetically engineered to express an attenuated level of a dehydratase/isomerase enzyme, such as an enzyme encoded by *fabA* or by a gene listed in Figure 15. In some embodiments, the host cell comprises a knockout of *fabA* or a gene listed in Figure 15. In other embodiments, the host cell is genetically engineered to express an attenuated level of a ketoacyl-ACP synthase, such as an enzyme encoded by *fabB* or by a gene listed in Figure 16. In certain embodiments, the host cell comprises a knockout of *fabB* or a gene listed in Figure 16. In yet other embodiments, the host cell is genetically engineered to express a modified level of a gene encoding a desaturase enzyme, such as *desA*.

[0090] In some embodiments, the recombinant vector further comprises a promoter operably linked to the nucleotide sequence. In certain embodiments, the promoter is a developmentally-regulated, an organelle-specific, a tissue-specific, an inducible, a constitutive, or a cell-specific promoter.

[0091] In other embodiments, the recombinant vector comprises at least one sequence selected from the group consisting of (a) a regulatory sequence operatively coupled to the nucleotide sequence; (b) a selection marker operatively coupled to the nucleotide sequence; (c) a marker sequence operatively coupled to the nucleotide sequence; (d) a purification moiety operatively coupled to the nucleotide sequence; (e) a secretion sequence operatively coupled to the nucleotide sequence; and (f) a targeting sequence operatively coupled to the nucleotide sequence.

[0092] In some embodiments, the recombinant vector is a plasmid.

[0093] In some embodiments, the host cell expresses a polypeptide encoded by the recombinant vector. In some embodiments, the nucleotide sequence is stably incorporated into the genomic DNA of the host cell, and the expression of the nucleotide sequence is under the control of a regulated promoter region.

[0094] In some embodiments, the method further includes culturing the host cell in the presence of at least one biological substrate for a fatty aldehyde biosynthetic polypeptide.

[0095] In another aspect, the invention features a method of producing a fatty alcohol. The method includes expressing in a host cell a gene encoding a fatty aldehyde biosynthetic polypeptide comprising (i) SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, and SEQ ID NO:10; (ii) SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, or SEQ ID NO:14; and/or (iii) SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:10, and SEQ ID NO:11; wherein the polypeptide has carboxylic acid reductase activity. In some embodiments, the polypeptide has fatty acid reductase activity.

[0096] In some embodiments, the method further includes isolating the fatty alcohol from the host cell. In some embodiments, the fatty alcohol is present in the extracellular environment. In certain embodiments, the fatty alcohol is isolated from the extracellular environment of the host cell. In some embodiments, the fatty alcohol is secreted from the host cell. In alternative embodiments, the fatty alcohol is transported into the extracellular environment. In other embodiments, the fatty alcohol is passively transported into the extracellular environment.

[0097] In some embodiments, the polypeptide is about 1,000 amino acids to about 2,000 amino acids in length. In certain embodiments, the polypeptide is about 1,000 amino acids in length, about 1,050 amino acids in length, about 1,100 amino acids in length, about 1,150 amino acids in length, about 1,200 amino acids in length, about 1,250 amino acids in length, about 1,300 amino acids in length, about 1,400 amino acids in length, about 1,500 amino acids in length, about 1,600 amino acids in length, about 1,700 amino acids in length, about 1,800 amino acids in length, about 1,900 amino acids in length, or about 2,000 amino acids in length. In other embodiments, the polypeptide is up to about 2,000 amino acids in length, up to about 1,900 amino acids in length, up to about 1,800 amino acids in length, up to about 1,700 amino acids in length, up to about 1,600 amino acids in length, up to about 1,500 amino acids in

length, up to about 1,400 amino acids in length, up to about 1,300 amino acids in length, up to about 1,250 amino acids in length, up to about 1,200 amino acids in length, up to about 1,150 amino acids in length, up to about 1,100 amino acids in length, up to about 1,050 amino acids in length, or up to about 1,000 amino acids in length. In other embodiments, the polypeptide is more than about 1,000 amino acids in length, more than about 1,050 amino acids in length, more than about 1,100 amino acids in length, more than about 1,150 amino acids in length, more than about 1,200 amino acids in length, more than about 1,250 amino acids in length, more than about 1,300 amino acids in length, more than about 1,400 amino acids in length, more than about 1,500 amino acids in length, more than about 1,600 amino acids in length, more than about 1,700 amino acids in length, more than about 1,800 amino acids in length, more than about 1,900 amino acids in length, or about 2,000 amino acids in length.

[0098] In some embodiments, the method further includes modifying the expression of a gene encoding a fatty acid synthase in the host cell. In certain embodiments, modifying the expression of a gene encoding a fatty acid synthase includes expressing a gene encoding a fatty acid synthase in the host cell and/or increasing the expression or activity of an endogenous fatty acid synthase in the host cell. In alternate embodiments, modifying the expression of a gene encoding a fatty acid synthase includes attenuating a gene encoding a fatty acid synthase in the host cell and/or decreasing the expression or activity of an endogenous fatty acid synthase in the host cell. In some embodiments, the fatty acid synthase is a thioesterase. In particular embodiments, the thioesterase is encoded by *tesA*, *tesA* without leader sequence, *tesB*, *fatB*, *fatB2*, *fatB3*, *fatA*, *or fatA1*.

[0099] In some embodiments, the method further includes expressing a gene encoding a fatty alcohol biosynthetic polypeptide in the host cell. In particular embodiments, the fatty alcohol biosynthetic polypeptide comprises the amino acid sequence of SEQ ID NO:94, 96, 98, 100, 102, 104, 106, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, or 194, or a variant thereof.

[0100] In other embodiments, the host cell is genetically engineered to express an attenuated level of a fatty acid degradation enzyme relative to a wild type host cell. In

some embodiments, the host cell is genetically engineered to express an attenuated level of an acyl-CoA synthase relative to a wild type host cell. In particular embodiments, the host cell expresses an attenuated level of an acyl-CoA synthase encoded by *fadD*, *fadK*, *BH3103*, *yhfL*, *Pfl-4354*, *EAV15023*, *fadD1*, *fadD2*, *RPC_4074*, *fadDD35*, *fadDD22*, *faa3p* or the gene encoding the protein ZP_01644857. In certain embodiments, the genetically engineered host cell comprises a knockout of one or more genes encoding a fatty acid degradation enzyme, such as the aforementioned acyl-CoA synthase genes.

[0101] In yet other embodiments, the host cell is genetically engineered to express an attenuated level of a dehydratase/isomerase enzyme, such as an enzyme encoded by *fabA* or by a gene listed in Figure 15. In some embodiments, the host cell comprises a knockout of *fabA* or a gene listed in Figure 15. In other embodiments, the host cell is genetically engineered to express an attenuated level of a ketoacyl-ACP synthase, such as an enzyme encoded by *fabB* or by a gene listed in Figure 16. In certain embodiments, the host cell comprises a knockout of *fabB* or a gene listed in Figure 16. In yet other embodiments, the host cell is genetically engineered to express a modified level of a gene encoding a desaturase enzyme, such as *desA*.

[0102] In some embodiments, the method further includes culturing the host cell in the presence of at least one biological substrate for the fatty aldehyde biosynthetic polypeptide.

[0103] In another aspect, the invention features a method of making a fatty alcohol. The method includes expressing in a host cell a gene encoding a fatty alcohol biosynthetic polypeptide comprising the amino acid sequence of SEQ ID NO:94, 96, 98, 100, 102, 104, 106, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, or 194, or a variant thereof. In some embodiments, the method further includes isolating the fatty alcohol from the host cell. In some embodiments, the fatty alcohol is present in the extracellular environment. In certain embodiments, the fatty alcohol is isolated from the extracellular environment of the host cell. In some embodiments, the fatty alcohol is secreted from the host cell. In alternative embodiments, the fatty alcohol is

transported into the extracellular environment. In other embodiments, the fatty alcohol is passively transported into the extracellular environment.

[0104] In some embodiments, the fatty alcohol biosynthetic polypeptide comprises the amino acid sequence of SEQ ID NO:94, 96, 98, 100, 102, 104, 106, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, or 194 with one or more amino acid substitutions, additions, insertions, or deletions, and the polypeptide has alcohol dehydrogenase activity.

[0105] In some embodiments, the polypeptide comprises one or more of the following conservative amino acid substitutions: replacement of an aliphatic amino acid, such as alanine, valine, leucine, and isoleucine, with another aliphatic amino acid; replacement of a serine with a threonine; replacement of a threonine with a serine; replacement of an acidic residue, such as aspartic acid and glutamic acid, with another acidic residue; replacement of a residue bearing an amide group, such as asparagine and glutamine, with another residue bearing an amide group; exchange of a basic residue, such as lysine and arginine, with another basic residue; and replacement of an aromatic residue, such as phenylalanine and tyrosine, with another aromatic residue. In some embodiments, the polypeptide has about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 30, 40, 50, 60, 70, 80, 90, 100, or more amino acid substitutions, additions, insertions, or deletions. In some embodiments, the polypeptide has alcohol dehydrogenase activity.

[0106] In some embodiments, the method further includes modifying the expression of a gene encoding a fatty acid synthase in the host cell. In certain embodiments, modifying the expression of a gene encoding a fatty acid synthase includes expressing a gene encoding a fatty acid synthase in the host cell and/or increasing the expression or activity of an endogenous fatty acid synthase in the host cell. In alternate embodiments, modifying the expression of a gene encoding a fatty acid synthase includes attenuating a gene encoding a fatty acid synthase in the host cell and/or decreasing the expression or activity of an endogenous fatty acid synthase in the host cell. In some embodiments, the fatty acid synthase is a thioesterase. In

particular embodiments, the thioesterase is encoded by *tesA*, *tesA* without leader sequence, *tesB*, *fatB*, *fatB2*, *fatB3*, *fatA*, *or fatA1*.

[0107] In some embodiments, the method further includes expressing a gene encoding a fatty aldehyde biosynthetic polypeptide in the host cell. In particular embodiments, the fatty aldehyde biosynthetic polypeptide comprises the amino acid sequence of SEQ ID NO:18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 264, 266, 268, 270, or 272, or a variant thereof.

[0108] In other embodiments, the host cell is genetically engineered to express an attenuated level of a fatty acid degradation enzyme relative to a wild type host cell. In some embodiments, the host cell is genetically engineered to express an attenuated level of an acyl-CoA synthase relative to a wild type host cell. In particular embodiments, the host cell expresses an attenuated level of an acyl-CoA synthase encoded by *fadD*, *fadK*, *BH3103*, *yhfL*, *Pfl-4354*, *EAV15023*, *fadD1*, *fadD2*, *RPC_4074*, *fadDD35*, *fadDD22*, *faa3p* or the gene encoding the protein ZP_01644857. In certain embodiments, the genetically engineered host cell comprises a knockout of one or more genes encoding a fatty acid degradation enzyme, such as the aforementioned acyl-CoA synthase genes.

[0109] In yet other embodiments, the host cell is genetically engineered to express an attenuated level of a dehydratase/isomerase enzyme, such as an enzyme encoded by *fabA*. In other embodiments, the host cell is genetically engineered to express an attenuated level of a ketoacyl-ACP synthase, such as an enzyme encoded by *fabB*. In yet other embodiments, the host cell is genetically engineered to express a modified level of a gene encoding a desaturase enzyme, such as *desA*.

[0110] In some embodiments, the polypeptide is from a bacterium, a plant, an insect, a yeast, a fungus, or a mammal.

[0111] In certain embodiments, the polypeptide is from a bacterium. In some embodiments, the bacterium is a mycobacterium selected from the group consisting of *Mycobacterium smegmatis*, *Mycobacterium abscessus*, *Mycobacterium avium*, *Mycobacterium bovis*, *Mycobacterium tuberculosis*, *Mycobacterium leprae*, *Mycobacterium marinum*, and *Mycobacterium ulcerans*. In other embodiments, the

bacterium is *Nocardia sp.* NRRL 5646, *Nocardia farcinica*, *Streptomyces griseus*, *Salinispora arenicola*, or *Clavibacter michiganenesis*.

[0112] In some embodiments, the method further includes culturing the host cell in the presence of at least one biological substrate for the fatty alcohol biosynthetic polypeptide.

[0113] In another aspect, the invention features a method of producing a fatty alcohol. The method includes expressing in a host cell a gene encoding a fatty alcohol biosynthetic polypeptide comprising an amino acid sequence having at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% sequence identity to the amino acid sequence of SEQ ID NO:94, 96, 98, 100, 102, 104, 106, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, or 194. In some embodiments, the amino acid sequence is SEQ ID NO:94, 96, 98, 100, 102, 104, 106, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, or 194.

[0114] In some embodiments, the method further includes isolating the fatty alcohol from the host cell. In some embodiments, the fatty alcohol is present in the extracellular environment. In certain embodiments, the fatty alcohol is isolated from the extracellular environment of the host cell. In some embodiments, the fatty alcohol is secreted from the host cell. In alternative embodiments, the fatty alcohol is transported into the extracellular environment. In other embodiments, the fatty alcohol is passively transported into the extracellular environment.

[0115] In some embodiments, the method further includes modifying the expression of a gene encoding a fatty acid synthase in the host cell. In certain embodiments, modifying the expression of a gene encoding a fatty acid synthase includes expressing a gene encoding a fatty acid synthase in the host cell and/or increasing the expression or activity of an endogenous fatty acid synthase in the host cell. In alternate embodiments, modifying the expression of a gene encoding a fatty

acid synthase includes attenuating a gene encoding a fatty acid synthase in the host cell and/or decreasing the expression or activity of an endogenous fatty acid synthase in the host cell. In some embodiments, the fatty acid synthase is a thioesterase. In particular embodiments, the thioesterase is encoded by *tesA*, *tesA* without leader sequence, *tesB*, *fatB*, *fatB2*, *fatB3*, *fatA*, *or fatA1*.

[0116] In some embodiments, the method further includes expressing a gene encoding a fatty aldehyde biosynthetic polypeptide in the host cell. In particular embodiments, the fatty aldehyde biosynthetic polypeptide comprises the amino acid sequence of SEQ ID NO:18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 264, 266, 268, 270, or 272, or a variant thereof.

[0117] In other embodiments, the host cell is genetically engineered to express an attenuated level of a fatty acid degradation enzyme relative to a wild type host cell. In some embodiments, the host cell is genetically engineered to express an attenuated level of an acyl-CoA synthase relative to a wild type host cell. In particular embodiments, the host cell expresses an attenuated level of an acyl-CoA synthase encoded by *fadD*, *fadK*, *BH3103*, *yhfL*, *Pfl-4354*, *EAV15023*, *fadD1*, *fadD2*, *RPC_4074*, *fadDD35*, *fadDD22*, *faa3p* or the gene encoding the protein ZP_01644857. In certain embodiments, the genetically engineered host cell comprises a knockout of one or more genes encoding a fatty acid degradation enzyme, such as the aforementioned acyl-CoA synthase genes.

[0118] In yet other embodiments, the host cell is genetically engineered to express an attenuated level of a dehydratase/isomerase enzyme, such as an enzyme encoded by *fabA* or by a gene listed in Figure 15. In some embodiments, the host cell comprises a knockout of *fabA* or a gene listed in Figure 15. In other embodiments, the host cell is genetically engineered to express an attenuated level of a ketoacyl-ACP synthase, such as an enzyme encoded by *fabB* or by a gene listed in Figure 16. In certain embodiments, the host cell comprises a knockout of *fabB* or a gene listed in Figure 16. In yet other embodiments, the host cell is genetically engineered to express a modified level of a gene encoding a desaturase enzyme, such as *desA*.

[0119] In some embodiments, the polypeptide is from a bacterium, a plant, an insect, a yeast, a fungus, or a mammal.

[0120] In certain embodiments, the polypeptide is from a mammalian cell, plant cell, insect cell, yeast cell, fungus cell, filamentous fungi cell, bacterial cell, or any other organism described herein. In some embodiments, the bacterium is a mycobacterium selected from the group consisting of *Mycobacterium smegmatis*, *Mycobacterium abscessus*, *Mycobacterium avium*, *Mycobacterium bovis*, *Mycobacterium tuberculosis*, *Mycobacterium leprae*, *Mycobacterium marinum*, and *Mycobacterium ulcerans*. In other embodiments, the bacterium is *Nocardia sp.* NRRL 5646, *Nocardia farcinica*, *Streptomyces griseus*, *Salinispora arenicola*, or *Clavibacter michiganenesis*.

[0121] In some embodiments, the method further includes culturing the host cell in the presence of at least one biological substrate for the fatty alcohol biosynthetic polypeptide.

[0122] In another aspect, the invention features a method of producing a fatty alcohol. The method includes expressing in a host cell a polynucleotide that hybridizes to a complement of the nucleotide sequence of SEQ ID NO:93, 95, 97, 99, 101, 103, 105, 107, 109, 111, 113, 115, 117, 119, 121, 123, 125, 127, 129, 131, 133, 135, 137, 139, 141, 143, 145, 147, 149, 151, 153, 155, 157, 159, 161, 163, 165, 167, 169, 171, 173, 175, 177, 179, 181, 183, 185, 187, 189, 191, or 193, or to a fragment thereof, wherein the polynucleotide encodes a polypeptide having alcohol dehydrogenase activity.

[0123] In some embodiments, the method further includes isolating the fatty alcohol from the host cell. In some embodiments, the fatty alcohol is present in the extracellular environment. In certain embodiments, the fatty alcohol is isolated from the extracellular environment of the host cell. In some embodiments, the fatty alcohol is secreted from the host cell. In alternative embodiments, the fatty alcohol is transported into the extracellular environment. In other embodiments, the fatty alcohol is passively transported into the extracellular environment.

[0124] In some embodiments, the polynucleotide hybridizes under low stringency, medium stringency, high stringency, or very high stringency conditions, to a complement of the nucleotide sequence of SEQ ID NO:93, 95, 97, 99, 101, 103, 105, 107, 109, 111, 113, 115, 117, 119, 121, 123, 125, 127, 129, 131, 133, 135, 137, 139,

141, 143, 145, 147, 149, 151, 153, 155, 157, 159, 161, 163, 165, 167, 169, 171, 173, 175, 177, 179, 181, 183, 185, 187, 189, 191, or 193, or to a fragment thereof.

[0125] In some embodiments, the method further includes modifying the expression of a gene encoding a fatty acid synthase in the host cell. In certain embodiments, modifying the expression of a gene encoding a fatty acid synthase includes expressing a gene encoding a fatty acid synthase in the host cell and/or increasing the expression or activity of an endogenous fatty acid synthase in the host cell. In alternate embodiments, modifying the expression of a gene encoding a fatty acid synthase includes attenuating a gene encoding a fatty acid synthase in the host cell and/or decreasing the expression or activity of an endogenous fatty acid synthase in the host cell. In some embodiments, the fatty acid synthase is a thioesterase. In particular embodiments, the thioesterase is encoded by *tesA*, *tesA* without leader sequence, *tesB*, *fatB*, *fatB2*, *fatB3*, *fatA*, *or fatA1*.

[0126] In some embodiments, the method further includes expressing a gene encoding a fatty aldehyde biosynthetic polypeptide in the host cell. In particular embodiments, the fatty aldehyde biosynthetic polypeptide comprises the amino acid sequence of SEQ ID NO:18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 264, 266, 268, 270, or 272, or a variant thereof.

[0127] In other embodiments, the host cell is genetically engineered to express an attenuated level of a fatty acid degradation enzyme relative to a wild type host cell. In some embodiments, the host cell is genetically engineered to express an attenuated level of an acyl-CoA synthase relative to a wild type host cell. In particular embodiments, the host cell expresses an attenuated level of an acyl-CoA synthase encoded by *fadD*, *fadK*, *BH3103*, *yhfL*, *Pfl-4354*, *EAV15023*, *fadD1*, *fadD2*, *RPC_4074*, *fadDD35*, *fadDD22*, *faa3p* or the gene encoding the protein ZP_01644857. In certain embodiments, the genetically engineered host cell comprises a knockout of one or more genes encoding a fatty acid degradation enzyme, such as the aforementioned acyl-CoA synthase genes.

[0128] In yet other embodiments, the host cell is genetically engineered to express an attenuated level of a dehydratase/isomerase enzyme, such as an enzyme encoded by *fabA* or by a gene listed in Figure 15. In some embodiments, the host cell

comprises a knockout of *fabA* or a gene listed in Figure 15. In other embodiments, the host cell is genetically engineered to express an attenuated level of a ketoacyl-ACP synthase, such as an enzyme encoded by *fabB* or by a gene listed in Figure 16. In certain embodiments, the host cell comprises a knockout of *fabB* or a gene listed in Figure 16. In yet other embodiments, the host cell is genetically engineered to express a modified level of a gene encoding a desaturase enzyme, such as *desA*.

[0129] In some embodiments, the polynucleotide is from a bacterium, a plant, an insect, a yeast, a fungus, or a mammal.

[0130] In certain embodiments, the polypeptide is from a mammalian cell, plant cell, insect cell, yeast cell, fungus cell, filamentous fungi cell, bacterial cell, or any other organism described herein. In some embodiments, the bacterium is a mycobacterium selected from the group consisting of *Mycobacterium smegmatis*, *Mycobacterium abscessus*, *Mycobacterium avium*, *Mycobacterium bovis*, *Mycobacterium tuberculosis*, *Mycobacterium leprae*, *Mycobacterium marinum*, and *Mycobacterium ulcerans*. In other embodiments, the bacterium is *Nocardia sp.* NRRL 5646, *Nocardia farcinica*, *Streptomyces griseus*, *Salinispora arenicola*, or *Clavibacter michiganenesis*.

[0131] In some embodiments, the method further includes culturing the host cell in the presence of at least one biological substrate for the fatty aldehyde biosynthetic polypeptide.

[0132] In another aspect, the invention features a method of producing a fatty alcohol. The method includes expressing in a host cell a recombinant vector comprising a fatty alcohol biosynthetic nucleotide sequence having at least about 70% sequence identity to the nucleotide sequence of SEQ ID NO:93, 95, 97, 99, 101, 103, 105, 107, 109, 111, 113, 115, 117, 119, 121, 123, 125, 127, 129, 131, 133, 135, 137, 139, 141, 143, 145, 147, 149, 151, 153, 155, 157, 159, 161, 163, 165, 167, 169, 171, 173, 175, 177, 179, 181, 183, 185, 187, 189, 191, or 193. In some embodiments, the nucleotide sequence has at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% sequence identity to the nucleotide sequence of SEQ ID NO:93, 95, 97, 99, 101, 103, 105, 107, 109, 111, 113, 115, 117, 119, 121, 123, 125,

127, 129, 131, 133, 135, 137, 139, 141, 143, 145, 147, 149, 151, 153, 155, 157, 159, 161, 163, 165, 167, 169, 171, 173, 175, 177, 179, 181, 183, 185, 187, 189, 191, or 193. In some embodiments, the nucleotide sequence is of SEQ ID NO:93, 95, 97, 99, 101, 103, 105, 107, 109, 111, 113, 115, 117, 119, 121, 123, 125, 127, 129, 131, 133, 135, 137, 139, 141, 143, 145, 147, 149, 151, 153, 155, 157, 159, 161, 163, 165, 167, 169, 171, 173, 175, 177, 179, 181, 183, 185, 187, 189, 191, or 193.

[0133] In some embodiments, the method further includes isolating the fatty alcohol from the host cell. In some embodiments, the fatty alcohol is present in the extracellular environment. In certain embodiments, the fatty alcohol is isolated from the extracellular environment of the host cell. In some embodiments, the fatty alcohol is secreted from the host cell. In alternative embodiments, the fatty alcohol is transported into the extracellular environment. In other embodiments, the fatty alcohol is passively transported into the extracellular environment.

[0134] In some embodiments, the recombinant vector further comprises a promoter operably linked to the nucleotide sequence. In certain embodiments, the promoter is a developmentally-regulated, an organelle-specific, a tissue-specific, an inducible, a constitutive, or a cell-specific promoter.

[0135] In other embodiments, the recombinant vector comprises at least one sequence selected from the group consisting of (a) a regulatory sequence operatively coupled to the nucleotide sequence; (b) a selection marker operatively coupled to the nucleotide sequence; (c) a marker sequence operatively coupled to the nucleotide sequence; (d) a purification moiety operatively coupled to the nucleotide sequence; (e) a secretion sequence operatively coupled to the nucleotide sequence; and (f) a targeting sequence operatively coupled to the nucleotide sequence.

[0136] In some embodiments, the recombinant vector is a plasmid.

[0137] In some embodiments, the host cell expresses a polypeptide encoded by the recombinant vector. In some embodiments, the nucleotide sequence is stably incorporated into the genomic DNA of the host cell, and the expression of the nucleotide sequence is under the control of a regulated promoter region.

[0138] In some embodiments, the method further includes modifying the expression of a gene encoding a fatty acid synthase in the host cell. In certain embodiments, modifying the expression of a gene encoding a fatty acid synthase

includes expressing a gene encoding a fatty acid synthase in the host cell and/or increasing the expression or activity of an endogenous fatty acid synthase in the host cell. In alternate embodiments, modifying the expression of a gene encoding a fatty acid synthase includes attenuating a gene encoding a fatty acid synthase in the host cell and/or decreasing the expression or activity of an endogenous fatty acid synthase in the host cell. In some embodiments, the fatty acid synthase is a thioesterase. In particular embodiments, the thioesterase is encoded by *tesA*, *tesA* without leader sequence, *tesB*, *fatB*, *fatB2*, *fatB3*, *fatA*, *or fatA1*.

[0139] In some embodiments, the method further includes expressing a gene encoding a fatty aldehyde biosynthetic polypeptide in the host cell. In particular embodiments, the fatty aldehyde biosynthetic polypeptide comprises the amino acid sequence of SEQ ID NO:18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 264, 266, 268, 270, or 272, or a variant thereof.

[0140] In other embodiments, the host cell is genetically engineered to express an attenuated level of a fatty acid degradation enzyme relative to a wild type host cell. In some embodiments, the host cell is genetically engineered to express an attenuated level of an acyl-CoA synthase relative to a wild type host cell. In particular embodiments, the host cell expresses an attenuated level of an acyl-CoA synthase encoded by *fadD*, *fadK*, *BH3103*, *yhfL*, *Pfl-4354*, *EAV15023*, *fadD1*, *fadD2*, *RPC_4074*, *fadDD35*, *fadDD22*, *faa3p* or the gene encoding the protein ZP_01644857. In certain embodiments, the genetically engineered host cell comprises a knockout of one or more genes encoding a fatty acid degradation enzyme, such as the aforementioned acyl-CoA synthase genes.

[0141] In yet other embodiments, the host cell is genetically engineered to express an attenuated level of a dehydratase/isomerase enzyme, such as an enzyme encoded by *fabA* or by a gene listed in Figure 15. In some embodiments, the host cell comprises a knockout of *fabA* or a gene listed in Figure 15. In other embodiments, the host cell is genetically engineered to express an attenuated level of a ketoacyl-ACP synthase, such as an enzyme encoded by *fabB* or by a gene listed in Figure 16. In certain embodiments, the host cell comprises a knockout of *fabB* or a gene listed in

Figure 16. In yet other embodiments, the host cell is genetically engineered to express a modified level of a gene encoding a desaturase enzyme, such as *desA*.

[0142] In some embodiments, the method further includes culturing the host cell in the presence of at least one biological substrate for a fatty alcohol biosynthetic polypeptide.

[0143] In any of the aspects of the invention described herein, the host cell can be selected from the group consisting of a mammalian cell, plant cell, insect cell, yeast cell, fungus cell, filamentous fungi cell, and bacterial cell.

[0144] In some embodiments, the host cell is a Gram-positive bacterial cell. In other embodiments, the host cell is a Gram-negative bacterial cell.

[0145] In some embodiments, the host cell is selected from the genus *Escherichia*, *Bacillus*, *Lactobacillus*, *Rhodococcus*, *Pseudomonas*, *Aspergillus*, *Trichoderma*, *Neurospora*, *Fusarium*, *Humicola*, *Rhizomucor*, *Kluyveromyces*, *Pichia*, *Mucor*, *Myceliophtora*, *Penicillium*, *Phanerochaete*, *Pleurotus*, *Trametes*, *Chrysosporium*, *Saccharomyces*, *Stenotrophamonas*, *Schizosaccharomyces*, *Yarrowia*, or *Streptomyces*.

[0146] In certain 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.

[0147] 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.

[0148] In yet other embodiments, the host cell is a *Streptomyces lividans* cell or a *Streptomyces murinus* cell.

[0149] In yet other embodiments, the host cell is an *Actinomycetes* cell.

[0150] In some embodiments, the host cell is a *Saccharomyces cerevisiae* cell.

In some embodiments, the host cell is a *Saccharomyces cerevisiae* cell.

[0151] In particular embodiments, the host cell is a cell from an eukaryotic plant, algae, cyanolacterium, 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 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 some embodiments, the host cell is heterotrophic or mixotrophic in the absence of light. In certain embodiments, the host cell is a cell from *Avabidopsis thaliana*, *Panicum virgatum, Miscanthus giganteus, Zea mays, Botryococcuse braunii, Chlamydomonas reinhardtii, Dunaliela salina, Synechococcus Sp. PCC 7002, Synechococcus Sp. PCC 7942, Synechocystis Sp. PCC 6803, Thermosynechococcus elongates BP-1, Chlorobium tepidum, Chloroflexus auranticus, Chromatiumm vinosum, Rhodospirillum rubrum, Rhodobacter capsulatus, Rhodopseudomonas palusris, Clostridium ljungdahlii, Clostridiuthermocellum, Penicillium chrysogenum, Pichia pastoris, Saccharomyces cerevisiae, Schizosaccharomyces pombe, Pseudomonas fluorescens, or Zymomonas mobilis.*

[0152] In other 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.

[0153] In yet other embodiments, the host cell is an *E. coli* cell. In certain embodiments, the *E. coli* cell is a strain B, a strain C, a strain K, or a strain W *E. coli* cell.

[0154] In another aspect, the invention features a method of producing a fatty alcohol. The method includes contacting a substrate with (i) a fatty alcohol biosynthetic polypeptide comprising the amino acid sequence of SEQ ID NO:94, 96, 98, 100, 102, 104, 106, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, or 194, or a variant thereof, or (ii) a fatty alcohol biosynthetic polypeptide encoded by a nucleotide sequence having at least about 70% identity to the nucleotide sequence of SEQ ID NO:93, 95, 97, 99, 101, 103, 105, 107, 109, 111, 113, 115, 117, 119, 121, 123,

125, 127, 129, 131, 133, 135, 137, 139, 141, 143, 145, 147, 149, 151, 153, 155, 157, 159, 161, 163, 165, 167, 169, 171, 173, 175, 177, 179, 181, 183, 185, 187, 189, 191, or 193, or a variant thereof. In some embodiments, the method further includes purifying the fatty alcohol.

[0155] In some embodiments, the fatty alcohol biosynthetic polypeptide comprises the amino acid sequence of SEQ ID NO:94, 96, 98, 100, 102, 104, 106, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, or 194 with one or more amino acid substitutions, additions, insertions, or deletions, wherein the polypeptide has alcohol dehydrogenase activity.

[0156] In some embodiments, the polypeptide comprises one or more of the following conservative amino acid substitutions: replacement of an aliphatic amino acid, such as alanine, valine, leucine, and isoleucine, with another aliphatic amino acid; replacement of a serine with a threonine; replacement of a threonine with a serine; replacement of an acidic residue, such as aspartic acid and glutamic acid, with another acidic residue; replacement of a residue bearing an amide group, such as asparagine and glutamine, with another residue bearing an amide group; exchange of a basic residue, such as lysine and arginine, with another basic residue; and replacement of an aromatic residue, such as phenylalanine and tyrosine, with another aromatic residue. In some embodiments, the polypeptide has about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 30, 40, 50, 60, 70, 80, 90, 100, or more amino acid substitutions, additions, insertions, or deletions. In some embodiments, the polypeptide has alcohol dehydrogenase activity.

[0157] In some embodiments, the polypeptide has an amino acid sequence that is at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% identical to the amino acid sequence of SEQ ID NO:94, 96, 98, 100, 102, 104, 106, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, or 194. In some embodiments, the

polypeptide has the amino acid sequence of SEQ ID NO:94, 96, 98, 100, 102, 104, 106, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, or 194.

[0158] In some embodiments, the nucleotide sequence has at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% sequence identity to the nucleotide sequence of SEQ ID NO:93, 95, 97, 99, 101, 103, 105, 107, 109, 111, 113, 115, 117, 119, 121, 123, 125, 127, 129, 131, 133, 135, 137, 139, 141, 143, 145, 147, 149, 151, 153, 155, 157, 159, 161, 163, 165, 167, 169, 171, 173, 175, 177, 179, 181, 183, 185, 187, 189, 191, or 193. In some embodiments, the nucleotide sequence is SEQ ID NO:93, 95, 97, 99, 101, 103, 105, 107, 109, 111, 113, 115, 117, 119, 121, 123, 125, 127, 129, 131, 133, 135, 137, 139, 141, 143, 145, 147, 149, 151, 153, 155, 157, 159, 161, 163, 165, 167, 169, 171, 173, 175, 177, 179, 181, 183, 185, 187, 189, 191, or 193.

[0159] In any of the aspects of the invention described herein, the methods can produce fatty alcohols comprising a C_6-C_{26} fatty alcohol. In some embodiments, the fatty alcohol comprises a C_6, C_7, C_8, C_9, C_{10}, C_{11}, C_{12}, C_{13}, C_{14}, C_{15}, C_{16}, C_{17}, C_{18}, C_{19}, C_{20}, C_{21}, C_{22}, C_{23}, C_{24}, C_{25}, or a C_{26} fatty alcohol. In particular embodiments, the fatty alcohol is a C_6, C_8, C_{10}, C_{12}, C_{13}, C_{14}, C_{15}, C_{16}, C_{17}, or C_{18} fatty alcohol. In certain embodiments, the hydroxyl group of the fatty alcohol is in the primary (C_1) position.

[0160] In other embodiments, the fatty alcohol comprises a straight chain fatty alcohol. In other embodiments, the fatty alcohol comprises a branched chain fatty alcohol. In yet other embodiments, the fatty alcohol comprises a cyclic moiety.

[0161] In some embodiments, the fatty alcohol is an unsaturated fatty alcohol. In other embodiments, the fatty alcohol is a monounsaturated fatty alcohol. In certain embodiments, the unsaturated fatty alcohol is a C6:1, C7:1, C8:1, C9:1, C10:1, C11:1, C12:1, C13:1, C14:1, C15:1, C16:1, C17:1, C18:1, C19:1, C20:1, C21:1, C22:1, C23:1, C24:1, C25:1, or a C26:1 unsaturated fatty alcohol. In yet other embodiments, the fatty alcohol is unsaturated at the omega-7 position. In certain embodiments, the unsaturated fatty alcohol comprises a cis double bond.

[0162] In yet other embodiments, the fatty alcohol is a saturated fatty alcohol.

[0163] In any of the aspects of the invention described herein, a substrate for a fatty aldehyde biosynthetic polypeptide can be a fatty acid. In some embodiments, the fatty acid comprises a C_6-C_{26} fatty acid. In some embodiments, the fatty acid comprises a C_6, C_7, C_8, C_9, C_{10}, C_{11}, C_{12}, C_{13}, C_{14}, C_{15}, C_{16}, C_{17}, C_{18}, C_{19}, C_{20}, C_{21}, C_{22}, C_{23}, C_{24}, C_{25}, or a C_{26} fatty acid. In particular embodiments, the fatty acid is a C_6, C_8, C_{10}, C_{12}, C_{13}, C_{14}, C_{15}, C_{16}, C_{17}, or C_{18} fatty acid.

[0164] In other embodiments, the fatty acid comprises a straight chain fatty acid. In other embodiments, the fatty acid comprises a branched chain fatty acid. In yet other embodiments, the fatty acid comprises a cyclic moiety.

[0165] In some embodiments, the fatty acid is an unsaturated fatty acid. In other embodiments, the fatty acid is a monounsaturated fatty acid. In yet other embodiments, the fatty acid is a saturated fatty acid.

[0166] In another aspect, the invention features a genetically engineered microorganism comprising an exogenous control sequence stably incorporated into the genomic DNA of the microorganism upstream of a polynucleotide comprising a nucleotide sequence having at least about 70% sequence identity to the nucleotide sequence of SEQ ID NO:17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 93, 95, 97, 99, 101, 103, 105, 107, 109, 111, 113, 115, 117, 119, 121, 123, 125, 127, 129, 131, 133, 135, 137, 139, 141, 143, 145, 147, 149, 151, 153, 155, 157, 159, 161, 163, 165, 167, 169, 171, 173, 175, 177, 179, 181, 183, 185, 187, 189, 191, 193, 263, 265, 267, 269, or 271, wherein the microorganism produces an increased level of a fatty alcohol relative to a wild-type microorganism.

[0167] In some embodiments, the nucleotide sequence has at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% sequence identity to the nucleotide sequence of SEQ ID NO:17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 93, 95, 97, 99, 101, 103, 105, 107, 109, 111, 113, 115, 117, 119, 121, 123, 125, 127, 129, 131, 133, 135, 137, 139, 141, 143, 145, 147, 149, 151, 153, 155, 157,

159, 161, 163, 165, 167, 169, 171, 173, 175, 177, 179, 181, 183, 185, 187, 189, 191, 193, 263, 265, 267, 269, or 271. In some embodiments, the nucleotide sequence is SEQ ID NO:17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 93, 95, 97, 99, 101, 103, 105, 107, 109, 111, 113, 115, 117, 119, 121, 123, 125, 127, 129, 131, 133, 135, 137, 139, 141, 143, 145, 147, 149, 151, 153, 155, 157, 159, 161, 163, 165, 167, 169, 171, 173, 175, 177, 179, 181, 183, 185, 187, 189, 191, 193, 263, 265, 267, 269, or 271.

[0168] In some embodiments, the polynucleotide is endogenous to the microorganism.

[0169] In other embodiments, the microorganism is genetically engineered to express a modified level of a gene encoding a fatty acid synthase in the host cell. In certain embodiments, the microorganism expresses a recombinant gene encoding a fatty acid synthase or expresses an increased level of an endogenous fatty acid synthase. In alternate embodiments, the microorganism expresses an attenuated level of a gene encoding a fatty acid synthase in the host cell and/or a decreased expression or activity of an endogenous fatty acid synthase. In some embodiments, the fatty acid synthase is a thioesterase. In particular embodiments, the thioesterase is encoded by *tesA*, *tesA* without leader sequence, *tesB*, *fatB*, *fatB2*, *fatB3*, *fatA*, *or fatA1*.

[0170] In other embodiments, the microorganism is genetically engineered to express an attenuated level of a fatty acid degradation enzyme relative to a wild type microorganism. In some embodiments, the microorganism expresses an attenuated level of an acyl-CoA synthase relative to a wild type microorganism. In particular embodiments, the microorganism expresses an attenuated level of an acyl-CoA synthase encoded by *fadD*, *fadK*, *BH3103*, *yhfL*, *Pfl-4354*, *EAV15023*, *fadD1*, *fadD2*, *RPC_4074*, *fadDD35*, *fadDD22*, *faa3p* or the gene encoding the protein ZP_01644857. In certain embodiments, the microorganism comprises a knockout of one or more genes encoding a fatty acid degradation enzyme, such as the aforementioned acyl-CoA synthase genes.

[0171] In yet other embodiments, the microorganism is genetically engineered to express an attenuated level of a dehydratase/isomerase enzyme, such as an enzyme encoded by *fabA* or by a gene listed in Figure 15. In some embodiments, the microorganism comprises a knockout of *fabA* or a gene listed in Figure 15. In other

embodiments, the microorganism is genetically engineered to express an attenuated level of a ketoacyl-ACP synthase, such as an enzyme encoded by *fabB* or by a gene listed in Figure 16. In certain embodiments, the microorganism comprises a knockout of *fabB* or a gene listed in Figure 16. In yet other embodiments, the microorganism is genetically engineered to express a modified level of a gene encoding a desaturase enzyme, such as *desA*.

[0172] In some embodiments, the microorganism is a bacterium. In certain embodiments, the bacterium is a Gram-negative or a Gram-positive bacterium.

[0173] In some embodiments, the microorganism is a mycobacterium selected from the group consisting of *Mycobacterium smegmatis*, *Mycobacterium abscessus*, *Mycobacterium avium*, *Mycobacterium bovis*, *Mycobacterium tuberculosis*, *Mycobacterium leprae*, *Mycobacterium marinum*, and *Mycobacterium ulcerans*.

[0174] In other embodiments, the microorganism is *Nocardia sp.* NRRL 5646, *Nocardia farcinica*, *Streptomyces griseus*, *Salinispora arenicola*, or *Clavibacter michiganenesis*.

[0175] In another aspect, the invention features a fatty alcohol produced by any of the methods or any of the microorganisms described herein, or a surfactant comprising a fatty alcohol produced by any of the methods or any of the microorganisms described herein.

[0176] In some embodiments, the fatty alcohol has a $\delta^{13}C$ of about -15.4 or greater. In certain embodiments, the fatty alcohol has a $\delta^{13}C$ of about -15.4 to about -10.9, or of about -13.92 to about -13.84.

[0177] In some embodiments, the fatty alcohol has an $f_M{}^{14}C$ of at least about 1.003. In certain embodiments, the fatty alcohol has an $f_M{}^{14}C$ of at least about 1.01 or at least about 1.5. In some embodiments, the fatty alcohol has an $f_M{}^{14}C$ of about 1.111 to about 1.124.

[0178] In any of the aspects described herein, a fatty alcohol is produced at a yield of about 25 mg/L, about 50 mg/L, about 75 mg/L, about 100 mg/L, about 125 mg/L, about 150 mg/L, about 175 mg/L, about 200 mg/L, about 225 mg/L, about 250 mg/L, about 275 mg/L, about 300 mg/L, about 325 mg/L, about 350 mg/L, about 375 mg/L, about 400 mg/L, about 425 mg/L, about 450 mg/L, about 475 mg/L, about 500 mg/L, about 525 mg/L, about 550 mg/L, about 575 mg/L, about 600 mg/L, about 625 mg/L,

about 650 mg/L, about 675 mg/L, about 700 mg/L, about 725 mg/L, about 750 mg/L, about 775 mg/L, about 800 mg/L, about 825 mg/L, about 850 mg/L, about 875 mg/L, about 900 mg/L, about 925 mg/L, about 950 mg/L, about 975 mg/L, about 1000 g/L, about 1050 mg/L, about 1075 mg/L, about 1100 mg/L, about 1125 mg/L, about 1150 mg/L, about 1175 mg/L, about 1200 mg/L, about 1225 mg/L, about 1250 mg/L, about 1275 mg/L, about 1300 mg/L, about 1325 mg/L, about 1350 mg/L, about 1375 mg/L, about 1400 mg/L, about 1425 mg/L, about 1450 mg/L, about 1475 mg/L, about 1500 mg/L, about 1525 mg/L, about 1550 mg/L, about 1575 mg/L, about 1600 mg/L, about 1625 mg/L, about 1650 mg/L, about 1675 mg/L, about 1700 mg/L, about 1725 mg/L, about 1750 mg/L, about 1775 mg/L, about 1800 mg/L, about 1825 mg/L, about 1850 mg/L, about 1875 mg/L, about 1900 mg/L, about 1925 mg/L, about 1950 mg/L, about 1975 mg/L, about 2000 mg/L, or more.

[0179] In another aspect, the invention features a method of making a fatty alcohol described herein. The method includes culturing a host cell described herein in a medium having a low level of iron, under conditions sufficient to produce a fatty alcohol, as described herein. In particular embodiments, the medium contains less than about 500 μM iron, less than about 400 μM iron, less than about 300 μM iron, less than about 200 μM iron, less than about 150 μM iron, less than about 100 μM iron, less than about 90 μM iron, less than about 80 μM iron, less than about 70 μM iron, less than about 60μM iron, less than about 50 μM iron, less than about 40 μM iron, less than about 30 μM iron, less than about 20 μM iron, less than about 10 μM iron, or less than about 5 μM iron. In certain embodiments, the medium does not contain iron.

[0180] In any of the aspects described herein, a fatty alcohol is produced in a host cell or a microorganism described herein from a carbon source.

[0181] The following figures are presented for the purpose of illustration only, and are not intended to be limiting.

BRIEF DESCRIPTION OF THE DRAWINGS

[0182] FIG. 1 is a graphic representation of fatty alcohols produced by recombinant *E. coli* strains transformed with various plasmids.

[0183] FIG. 2 is a graphic representation of two GC/MS traces of organic compounds produced by recombinant *E. coli* strains transformed with various plasmids.

[0184] FIG. 3 is a schematic of a new pathway for fatty alcohol production.

[0185] FIG. 4 is a representation of a gel of PCR products from MG1655 wild type cells, Δ*fadD*::cm cells, and Δ*fadD* cells.

[0186] FIG. 5A is a GC/MS trace of fatty alcohol production in MG1655(DE3, Δ*fadD*) /pETDUet-1-tesA+ pHZ1.140B cells. FIG. 5B is a GC/MS trace of fatty alcohol production in MG16655(DE3, Δ*fadD*, *yjgB*::kan)/ pETDUet-1-tesA+ pHZ1.140B cells. FIG. 5C is a GC/MS trace of fatty alcohol production in MG16655(DE3, Δ*fadD*, *yjgB*::kan)/pDF1+ pHZ1.140B cells. The arrows in FIG. 5A, FIG. 5B, and FIG. 5C indicate the absence of C12:0 fatty aldehydes.

[0187] FIG. 6 is a listing of the nucleotide sequence and the corresponding amino acid sequence of *Nocardia sp.* NRRL 5646 *car* gene.

[0188] FIG. 7 is a listing of amino acid sequence motifs for CAR homologs.

[0189] FIG. 8 is a listing of nucleotide and amino acid sequences of *car* homolog genes.

[0190] FIG. 9 is a table identifying exemplary genes that can be expressed, overexpressed, or attenuated to increase production of particular substrates.

[0191] FIG. 10 is a listing of nucleotide and amino acid sequences of alcohol dehydrogenase genes.

[0192] FIG. 11 is a graphic representation of fatty alcohol production in various deletion mutants of *E. coli*.

[0193] FIG. 12 is a graphic representation of fatty alcohol production in various deletion mutants of *E. coli*.

[0194] FIG. 13 is a GC/MS trace of saturated fatty alcohol production in *E. coli*.

[0195] FIG. 14A is a graphic representation of fatty alcohol production in various Hu9 culture media. FIG. 14B is a graphic representation of fatty alcohol production in various Hu9 culture media.

[0196] FIG. 15 is a listing of nucleotide and amino acid sequences of *fabA* related genes.

[0197] FIG. 16 is a listing of nucleotide and amino acid sequences of *fabB* related genes.

[0198] FIG. 17 is a listing of additional nucleotide and amino acid sequences of the disclosure.

DETAILED DESCRIPTION OF THE INVENTION

[0199] 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.

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.

[0200] Other features and advantages of the invention will be apparent from the following detailed description, and from the claims.

Definitions

[0201] 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).

[0202] 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 2008.

[0203] 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 2008.

[0204] 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.

[0205] 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 70).

[0206] As used herein, the term "alcohol dehydrogenase" (EC 1.1.1.*) is a peptide capable of catalyzing the conversion of a fatty aldehyde to an alcohol (*e.g.*, fatty alcohol). Additionally, one of ordinary skill in the art will appreciate that some alcohol dehydrogenases will catalyze other reactions as well. For example, some alcohol dehydrogenases will accept other substrates in addition to fatty aldehydes. Such non-specific alcohol dehydrogenases are, therefore, also included in this definition. Nucleic acid sequences encoding alcohol dehydrogenases are known in the art, and such alcohol dehydrogenases are publicly available. Exemplary GenBank Accession Numbers are provided in Figure 9.

[0207] 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).

[0208] 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. Biodiesel can include esters or hydrocarbons, such as alcohols.

[0209] As used herein, the term "biofuel" refers to any fuel derived from biomass. 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.

[0210] As used herein, the term "biomass" refers to any biological material from which a carbon source is derived. In some instances, a biomass is processed into a carbon source, which is suitable for bioconversion. In other instances, the biomass may not require further processing into a carbon source. The carbon source can be converted into a biofuel. One exemplary source of biomass is plant matter or vegetation. For example, corn, sugar cane, or switchgrass can be used as biomass. Another non-limiting example of biomass is metabolic wastes, such as animal matter, for example cow manure. In addition, biomass may include algae and other marine plants. 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, ensilage, straw, lumber, sewage, garbage, cellulosic urban waste, and food leftovers. Biomass also includes sources of carbon, such as carbohydrates (*e.g.*, monosaccharides, disaccharides, or polysaccharides).

[0211] 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 ethanol, methanol, and glycerol, 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.

[0212] A nucleotide sequence is "complementary" to another nucleotide sequence if each of the bases of the two sequences matches (*i.e.*, 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.

[0213] 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.

[0214] 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.

[0215] 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 or fatty aldehyde 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.

[0216] 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.

[0217] 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 (*i.e.*, will not adversely affect desired biological properties, such as carboxylic acid reductase 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, histidine), acidic side chains (*e.g.*, aspartic acid, glutamic acid), uncharged polar side chains (*e.g.*, glycine, asparagine, glutamine, serine, threonine, tyrosine, cysteine), nonpolar side chains (*e.g.*, alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), beta-branched side chains (*e.g.*, threonine, valine, isoleucine), and aromatic side chains (*e.g.*, tyrosine, phenylalanine, tryptophan, histidine).

[0218] 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).

[0219] 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.

[0220] 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 synthases 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.

[0221] 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 nonlimiting example of a fatty acid degradation enzyme is an acyl-CoA synthase. Additional examples of fatty acid degradation enzymes are described herein.

[0222] 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-CoA, fatty aldehyde, short and long chain alcohols, hydrocarbons, fatty alcohols, and esters (*e.g.*, waxes, fatty acid esters, or fatty esters).

[0223] As used herein, the term "fatty acid derivative enzyme" means any enzyme that may be expressed or overexpressed in the production of fatty acid derivatives. 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, fatty alcohol-forming acyl-CoA reductases, fatty acid (carboxylic acid) reductases, acyl-ACP reductases, fatty acid hydroxylases, acyl-CoA desaturases, acyl-ACP desaturases, acyl-CoA oxidases, acyl-CoA dehydrogenases, ester synthases, and alkane biosynthetic polypeptides, *etc.* Fatty acid derivative enzymes can convert a substrate into a fatty acid derivative. In some examples, the substrate may be a fatty acid derivative that the fatty acid derivative enzyme converts into a different fatty acid derivative.

[0224] As used herein, "fatty acid enzyme" means any enzyme involved in fatty acid biosynthesis. Fatty acid enzymes can be modified in host cells to produce fatty

acids. Non-limiting examples of fatty acid enzymes include fatty acid synthases and thioesterases. Additional examples of fatty acid enzymes are described herein.

[0225] As used herein, "fatty acid synthase" means any enzyme involved in fatty acid biosynthesis. Fatty acid synthases can be expressed or overexpressed in host cells to produce fatty acids. A non-limiting example of a fatty acid synthase is a thioesterase. Additional examples of fatty acid synthases are described herein.

[0226] As used herein, "fatty aldehyde" means an aldehyde having the formula RCHO characterized by an unsaturated carbonyl group (C=O). In a preferred embodiment, the fatty aldehyde is any aldehyde made from a fatty acid or fatty acid derivative. In one embodiment, the R group is at least about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 carbons in length.

[0227] R 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.

[0228] Furthermore, R can be saturated or unsaturated. If unsaturated, the R can have one or more points of unsaturation.

[0229] In one embodiment, the fatty aldehyde is produced biosynthetically.

[0230] Fatty aldehydes have many uses. For example, fatty aldehydes can be used to produce many specialty chemicals. For example, fatty aldehydes are used to produce polymers, resins, dyes, flavorings, plasticizers, perfumes, pharmaceuticals, and other chemicals. Some are used as solvents, preservatives, or disinfectants. Some natural and synthetic compounds, such as vitamins and hormones, are aldehydes.

[0231] The terms "fatty aldehyde biosynthetic polypeptide", "carboxylic acid reductase", and "CAR" are used interchangeably herein.

[0232] As used herein, "fatty alcohol" means an alcohol having the formula ROH. In a preferred embodiment, the fatty alcohol is any alcohol made from a fatty acid or fatty acid derivative. In one embodiment, the R group is at least about 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 carbons in length.

[0233] R 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.

[0234] Furthermore, R can be saturated or unsaturated. If unsaturated, the R can have one or more points of unsaturation.

[0235] In one embodiment, the fatty alcohol is produced biosynthetically.

[0236] Fatty alcohols have many uses. For example, fatty alcohols can be used to produce many specialty chemicals. For example, fatty alcohols are used as a biofuel; as solvents for fats, waxes, gums, and resins; in pharmaceutical salves, emolients and lotions; as lubricating-oil additives; in detergents and emulsifiers; as textile antistatic and finishing agents; as plasticizers; as nonionic surfactants; and in cosmetics, for examples as thickeners.

[0237] 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.

[0238] "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.

[0239] 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.

[0240] 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 (1970), J. Mol. Biol. 48:444 453, 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 40, 50, 60, 70, or 80 and a length weight of 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.

[0241] As used herein, a "host cell" is a cell used to produce a product described herein (*e.g.*, a fatty alcohol described herein). 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, yeast, or filamentous fungi cells.

[0242] 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 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 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by two washes in 0.2X 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 6X SSC at about 45°C, followed by one or more washes in 0.2X SSC, 0.1% SDS at 60°C; 3) high stringency hybridization conditions in 6X SSC at about 45°C, followed by one or more washes in 0.2.X 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 0.2X SSC, 1% SDS at 65°C. Very high stringency conditions (4) are the preferred conditions unless otherwise specified.

[0243] 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, by an "isolated

nucleic acid" is meant to include nucleic acid fragments, which are not naturally occurring as fragments and would not be found in the natural state. The term "isolated" is also used herein to refer to polypeptides, which are isolated from other cellular proteins and is meant to encompass both purified and recombinant polypeptides. The term "isolated" as used herein also refers to a nucleic acid or peptide 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 peptide that is substantially free of chemical precursors or other chemicals when chemically synthesized. The term "isolated", as used herein with respect to products, such as fatty alcohols, refers to products that are isolated from cellular components, cell culture media, or chemical or synthetic precursors.

[0244] 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 degradation products encoded by the gene in the cell.

[0245] 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 terms "microbial cells" (*i.e.*, cells from microbes) and "microbes" are used interchangeably and refer to cells or small organisms that can only be seen with the aid of a microscope.

[0246] As used herein, the term "nucleic acid" refers to polynucleotides, such as deoxyribonucleic acid (DNA), and, where appropriate, ribonucleic acid (RNA). The term should also be understood to include, as equivalents, 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.

[0247] As used herein, the term "operably linked" means that 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 DNA. 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).

[0248] The term "or" is used herein to mean, and is used interchangeably with, the term "and/or," unless context clearly indicates otherwise.

[0249] As used herein, "overexpress" means to express or cause to be expressed 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.

[0250] 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 fatty aldehyde 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, a fatty aldehyde with a high logP value can separate into the organic phase even at very low concentrations in the fermentation vessel.

[0251] 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 fatty alcohol in a sample. For example, when fatty alcohols are produced in a host cell, the fatty alcohols can be purified by the removal of host cell proteins. After purification, the percentage of fatty alcohols in the sample is increased.

[0252] The terms "purify," "purified," and "purification" do not require absolute purity. They are relative terms. Thus, for example, when fatty alcohols are produced in host cells, a purified fatty alcohol is one that is substantially separated from other cellular components (*e.g.*, nucleic acids, polypeptides, lipids, carbohydrates, or other hydrocarbons). In another example, a purified fatty alcohol preparation is one in which the fatty alcohol is substantially free from contaminants, such as those that might be present following fermentation. In some embodiments, a fatty alcohol is purified when at least about 50% by weight of a sample is composed of the fatty alcohol. In other embodiments, a fatty alcohol 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 alcohol.

[0253] As used herein, the term "recombinant polypeptide" refers to a polypeptide that is produced by recombinant DNA techniques, wherein generally DNA encoding the expressed protein 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.

[0254] 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, *e.g.*, conserved amino acid substitutions) amino acid residues or nucleotides to a second amino acid or nucleotide sequence such that the first and second amino acid or nucleotide sequences have similar activities.

[0255] 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.

[0256] 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.

[0257] As used herein, "transformation" refers to a process in which a cell's genotype is changed as a result of the cellular uptake of exogenous DNA or RNA. This may result in the transformed cell expressing a recombinant form of an 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.

[0258] As used herein, a "transport protein" is a polypeptide that facilitates the movement of one or more compounds in and/or out of a cellular organelle and/or a cell.

[0259] As used herein, a "variant" of polypeptide X refers to a polypeptide having the amino acid sequence of peptide X in which one or more amino acid residues is altered. The variant may have conservative changes or nonconservative 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).

[0260] 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.

[0261] 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.

[0262] The invention is based, at least in part, on the discovery of a new pathway for fatty alcohol biosynthesis in *E. coli* that utilize, in part, genes that encode fatty aldehyde biosynthetic polypeptides. The fatty alcohols can be produced by a biosynthetic pathway depicted in Figure 3. In this pathway, a fatty acid is first activated by ATP and then reduced by a carboxylic acid reductase (CAR)-like enzyme to generate a fatty aldehyde. The fatty aldehyde can then be further reduced into a fatty alcohol by an alcohol dehydrogenase(s), such as *alrAadp1* or *yjgB*. As demonstrated herein, *yjgB* may be the presumed alcohol dehydrogenase, whose substrates includes fatty aldehydes, for example fatty aldehydes with carbon chain lengths from C_{10} to C_{18}.

Fatty Aldehyde Biosynthetic Genes, Fatty Alcohol Biosynthetic Genes, and Variants

[0263] The methods described herein can be used to produce fatty alcohols, for example, from fatty aldehydes. In some instances, a fatty aldehyde is produced by expressing a fatty aldehyde biosynthetic gene, for example, a carboxylic acid reductase gene (*car* gene), having a nucleotide sequence listed in Figures 6 and 8, as well as polynucleotide variants thereof. In some instances, the fatty aldehyde biosynthetic gene encodes one or more of the amino acid motifs depicted in Figure 7. For example, the gene can encode a polypeptide comprising SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, and SEQ ID NO:10; SEQ ID NO:11; SEQ ID NO:12; SEQ ID NO:13; SEQ ID NO:14; and/or SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:10, and SEQ ID NO:11. SEQ ID NO:7 includes a reductase domain; SEQ ID NO:8 and SEQ ID NO:14 include a NADP binding domain; SEQ ID NO:9 includes a phosphopantetheine attachment site; and SEQ ID NO:10 includes an AMP binding domain.

[0264] In other instances, a fatty alcohol is produced by expressing a fatty alcohol biosynthetic gene, for example, having a nucleotide sequence listed in Figure 10, or a variant thereof.

[0265] Variants can be naturally occurring or created *in vitro*. In particular, such variants can be created using genetic engineering techniques, such as site directed mutagenesis, random chemical mutagenesis, Exonuclease III deletion procedures, or standard cloning techniques. Alternatively, such variants, fragments, analogs, or derivatives can be created using chemical synthesis or modification procedures.

[0266] Methods of making variants are well known in the art. These include procedures in which nucleic acid sequences obtained from natural isolates are modified to generate nucleic acids that encode polypeptides having characteristics that enhance their value in industrial or laboratory applications. In such procedures, a large number of variant sequences having one or more nucleotide differences with respect to the sequence obtained from the natural isolate are generated and characterized. Typically, these nucleotide differences result in amino acid changes with respect to the polypeptides encoded by the nucleic acids from the natural isolates.

[0267] For example, variants can be created using error prone PCR (see, *e.g.*, Leung *et al.*, *Technique* 1:11-15, 1989; and Caldwell *et al.*, *PCR Methods Applic.* 2:28-33, 1992). In error prone PCR, PCR is performed under conditions where the copying fidelity of the DNA polymerase is low, such that a high rate of point mutations is obtained along the entire length of the PCR product. Briefly, in such procedures, nucleic acids to be mutagenized (*e.g.*, a fatty aldehyde biosynthetic polynucleotide sequence), are mixed with PCR primers, reaction buffer, $MgCl_2$, $MnCl_2$, Taq polymerase, and an appropriate concentration of dNTPs for achieving a high rate of point mutation along the entire length of the PCR product. For example, the reaction can be performed using 20 fmoles of nucleic acid to be mutagenized (*e.g.*, a fatty aldehyde biosynthetic polynucleotide sequence), 30 pmole of each PCR primer, a reaction buffer comprising 50 mM KCl, 10 mM Tris HCl (pH 8.3), and 0.01% gelatin, 7 mM $MgCl_2$, 0.5 mM $MnCl_2$, 5 units of Taq polymerase, 0.2 mM dGTP, 0.2 mM dATP, 1 mM dCTP, and 1 mM dTTP. PCR can be performed for 30 cycles of 94°C for 1 min, 45°C for 1 min, and 72°C for 1 min. However, it will be appreciated that these parameters can be varied as appropriate. The mutagenized nucleic acids are then cloned into an appropriate vector and the activities of the polypeptides encoded by the mutagenized nucleic acids are evaluated.

[0268] Variants can also be created using oligonucleotide directed mutagenesis to generate site-specific mutations in any cloned DNA of interest. Oligonucleotide mutagenesis is described in, for example, Reidhaar-Olson *et al.*, *Science* 241:53-57, 1988. Briefly, in such procedures a plurality of double stranded oligonucleotides bearing one or more mutations to be introduced into the cloned DNA are synthesized and inserted into the cloned DNA to be mutagenized (*e.g.*, a fatty aldehyde biosynthetic polynucleotide sequence). Clones containing the mutagenized DNA are recovered, and the activities of the polypeptides they encode are assessed.

[0269] Another method for generating variants is assembly PCR. Assembly PCR involves the assembly of a PCR product from a mixture of small DNA fragments. A large number of different PCR reactions occur in parallel in the same vial, with the products of one reaction priming the products of another reaction. Assembly PCR is described in, for example, U.S. Pat. No. 5,965,408.

[0270] Still another method of generating variants is sexual PCR mutagenesis. In sexual PCR mutagenesis, forced homologous recombination occurs between DNA molecules of different, but highly related, DNA sequence *in vitro* as a result of random fragmentation of the DNA molecule based on sequence homology. This is followed by fixation of the crossover by primer extension in a PCR reaction. Sexual PCR mutagenesis is described in, for example, Stemmer, *PNAS, USA* 91:10747-10751, 1994.

[0271] Variants can also be created by *in vivo* mutagenesis. In some embodiments, random mutations in a nucleic acid sequence are generated by propagating the sequence in a bacterial strain, such as an *E. coli* strain, which carries mutations in one or more of the DNA repair pathways. Such "mutator" strains have a higher random mutation rate than that of a wild-type strain. Propagating a DNA sequence (*e.g.*, a fatty aldehyde biosynthetic polynucleotide sequence) in one of these strains will eventually generate random mutations within the DNA. Mutator strains suitable for use for *in vivo* mutagenesis are described in, for example, PCT Publication No. WO 91/16427.

[0272] Variants can also be generated using cassette mutagenesis. In cassette mutagenesis, a small region of a double stranded DNA molecule is replaced with a synthetic oligonucleotide "cassette" that differs from the native sequence. The

oligonucleotide often contains a completely and/or partially randomized native sequence.

[0273] Recursive ensemble mutagenesis can also be used to generate variants. Recursive ensemble mutagenesis is an algorithm for protein engineering (*i.e.*, protein mutagenesis) developed to produce diverse populations of phenotypically related mutants whose members differ in amino acid sequence. This method uses a feedback mechanism to control successive rounds of combinatorial cassette mutagenesis. Recursive ensemble mutagenesis is described in, for example, Arkin *et al.*, *PNAS, USA* 89:7811-7815, 1992.

[0274] In some embodiments, variants are created using exponential ensemble mutagenesis. Exponential ensemble mutagenesis is a process for generating combinatorial libraries with a high percentage of unique and functional mutants, wherein small groups of residues are randomized in parallel to identify, at each altered position, amino acids which lead to functional proteins. Exponential ensemble mutagenesis is described in, for example, Delegrave *et al.*, *Biotech. Res.* 11:1548-1552, 1993. Random and site-directed mutagenesis are described in, for example, Arnold, *Curr. Opin. Biotech.* 4:450-455, 1993.

[0275] In some embodiments, variants are created using shuffling procedures wherein portions of a plurality of nucleic acids that encode distinct polypeptides are fused together to create chimeric nucleic acid sequences that encode chimeric polypeptides as described in, for example, U.S. Pat. Nos. 5,965,408 and 5,939,250.

[0276] Polynucleotide variants also include nucleic acid analogs. Nucleic acid analogs can be modified at the base moiety, sugar moiety, or phosphate backbone to improve, for example, stability, hybridization, or solubility of the nucleic acid. Modifications at the base moiety include deoxyuridine for deoxythymidine and 5-methyl-2'-deoxycytidine or 5-bromo-2'-doxycytidine for deoxycytidine. Modifications of the sugar moiety include modification of the 2' hydroxyl of the ribose sugar to form 2'-O-methyl or 2'-O-allyl sugars. The deoxyribose phosphate backbone can be modified to produce morpholino nucleic acids, in which each base moiety is linked to a six-membered, morpholino ring, or peptide nucleic acids, in which the deoxyphosphate backbone is replaced by a pseudopeptide backbone and the four bases are retained. (See, *e.g.*, Summerton *et al.*, *Antisense Nucleic Acid Drug*

Dev. (1997) 7:187-195; and Hyrup *et al.*, *Bioorgan. Med. Chem.* (1996) 4:5-23.) In addition, the deoxyphosphate backbone can be replaced with, for example, a phosphorothioate or phosphorodithioate backbone, a phosphoroamidite, or an alkyl phosphotriester backbone.

[0277] Any polynucleotide sequence encoding a homolog listed in Figures 6 and 8, or a variant thereof, can be used as a fatty aldehyde biosynthetic polynucleotide in the methods described herein. Any polynucleotide sequence listed in Figure 10, or a variant, can be used as a fatty alcohol biosynthetic polynucleotide in the methods described herein.

Fatty Aldehyde Biosynthetic Polypeptides, Fatty Alcohol Biosynthetic Polypeptide, and Variants

[0278] The methods described herein can also be used to produce fatty alcohols, for example, from fatty aldehydes. In some instances, the fatty aldehyde is produced by a fatty aldehyde biosynthetic polypeptide having an amino acid sequence listed in Figures 6 and 8, as well as polypeptide variants thereof. In some instances, a fatty aldehyde biosynthetic polypeptide is one that includes one or more of the amino acid motifs depicted in Figure 7. For example, the polypeptide can include the amino acid sequences of SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, and SEQ ID NO:10. In other situations, the polypeptide includes one or more of SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, and SEQ ID NO:14. In yet other instances, the polypeptide includes the amino acid sequences of SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:10, and SEQ ID NO:11. SEQ ID NO:7 includes a reductase domain; SEQ ID NO:8 and SEQ ID NO:14 include a NADP binding domain; SEQ ID NO:9 includes a phosphopantetheine attachment site; and SEQ ID NO:10 includes an AMP binding domain.

[0279] In other instances, the methods described herein can be used to produce fatty alcohols using a fatty alcohol biosynthetic polypeptide having an amino acid sequence listed in Figure 10, as well as polypeptide variants thereof.

[0280] Biosynthetic polypeptide variants can be variants in which one or more amino acid residues are substituted with a conserved or non-conserved amino acid residue (preferably a conserved amino acid residue). Such substituted amino acid residue may or may not be one encoded by the genetic code.

[0281] Conservative substitutions are those that substitute a given amino acid in a polypeptide by another amino acid of similar characteristics. Typical conservative substitutions are the following replacements: replacement of an aliphatic amino acid, such as alanine, valine, leucine, and isoleucine, with another aliphatic amino acid; replacement of a serine with a threonine or vice versa; replacement of an acidic residue, such as aspartic acid and glutamic acid, with another acidic residue; replacement of a residue bearing an amide group, such as asparagine and glutamine, with another residue bearing an amide group; exchange of a basic residue, such as lysine and arginine, with another basic residue; and replacement of an aromatic residue, such as phenylalanine and tyrosine, with another aromatic residue.

[0282] Other polypeptide variants are those in which one or more amino acid residues include a substituent group. Still other polypeptide variants are those in which the polypeptide is associated with another compound, such as a compound to increase the half-life of the polypeptide (*e.g.*, polyethylene glycol).

[0283] Additional polypeptide variants are those in which additional amino acids are fused to the polypeptide, such as a leader sequence, a secretory sequence, a proprotein sequence, or a sequence which facilitates purification, enrichment, or stabilization of the polypeptide.

[0284] In some instances, the polypeptide variants retain the same biological function as a polypeptide having an amino acid sequence listed in Figures 6 and 8 (*e.g.*, retain fatty aldehyde biosynthetic activity, such as carboxylic acid or fatty acid reductase activity), or listed in Figure 10 (*e.g.*, retain fatty alcohol biosynthetic activity, such as fatty alcohol dehydrogenase activity) and have amino acid sequences substantially identical thereto.

[0285] In other instances, the polypeptide variants have at least about 50%, at least about 55%, at least about 60%, at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, or more than about 95% homology to an amino acid sequence listed in Figures 6, 8, and/or 10. In another embodiment, the polypeptide variants include a fragment comprising at least about 5, 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, or 150 consecutive amino acids thereof.

[0286] The polypeptide variants or fragments thereof can be obtained by isolating nucleic acids encoding them using techniques described herein or by expressing synthetic nucleic acids encoding them. Alternatively, polypeptide variants or fragments thereof can be obtained through biochemical enrichment or purification procedures. The sequence of polypeptide variants or fragments can be determined by proteolytic digestion, gel electrophoresis, and/or microsequencing. The sequence of the polypeptide variants or fragments can then be compared to an amino acid sequence listed in Figures 6, 8, and/or 10 using any of the programs described herein.

[0287] The polypeptide variants and fragments thereof can be assayed for fatty aldehyde-producing activity and/or fatty alcohol-producing activity using routine methods. For example, the polypeptide variants or fragment can be contacted with a substrate (*e.g.*, a fatty acid, a fatty acid derivative substrate, or other substrate described herein) under conditions that allow the polypeptide variant to function. A decrease in the level of the substrate or an increase in the level of a fatty aldehyde can be measured to determine fatty aldehyde-producing activity. A decrease in the level of the substrate or an increase in the level of a fatty alcohol can be measured to determine fatty alcohol-producing activity.

Antibodies to Biosynthetic Polypeptides

[0288] The fatty aldehyde biosynthetic polypeptides described herein can also be used to produce antibodies directed against fatty aldehyde biosynthetic polypeptides. Such antibodies can be used, for example, to detect the expression of a fatty aldehyde biosynthetic polypeptide or fatty alcohol biosynthetic polypeptide using methods known in the art. The antibody can be, for example, a polyclonal antibody; a monoclonal antibody or antigen binding fragment thereof; a modified antibody such as a chimeric antibody, reshaped antibody, humanized antibody, or fragment thereof (*e.g.*, Fab', Fab, $F(ab')_2$); or a biosynthetic antibody, for example, a single chain antibody, single domain antibody (DAB), Fv, single chain Fv (scFv), or the like.

[0289] Methods of making and using polyclonal and monoclonal antibodies are described, for example, in Harlow *et al.*, Using Antibodies: A Laboratory Manual: Portable Protocol I. Cold Spring Harbor Laboratory (December 1, 1998). Methods for making modified antibodies and antibody fragments (*e.g.*, chimeric antibodies,

reshaped antibodies, humanized antibodies, or fragments thereof, *e.g.*, Fab', Fab, $F(ab')_2$ fragments); or biosynthetic antibodies (*e.g.*, single chain antibodies, single domain antibodies (DABs), Fv, single chain Fv (scFv), and the like), are known in the art and can be found, for example, in Zola, Monoclonal Antibodies: Preparation and Use of Monoclonal Antibodies and Engineered Antibody Derivatives, Springer Verlag (December 15, 2000; 1st edition).

Substrates

[0290] The compositions and methods described herein can be used to produce fatty alcohols, for example, from fatty aldehydes, which themselves can be produced from an appropriate substrate. While not wishing to be bound by theory, it is believed that the fatty aldehyde biosynthetic polypeptides described herein produce fatty aldehydes from substrates via a reduction mechanism. In some instances, the substrate is a fatty acid derivative (*e.g.*, a fatty acid), and a fatty aldehyde having particular branching patterns and carbon chain length can be produced from a fatty acid derivative having those characteristics that would result in a particular fatty aldehyde. Through an additional reaction mechanism, the fatty aldehyde can be converted into the desired fatty alcohol (*e.g.*, by a fatty alcohol biosynthetic polypeptide described herein).

[0291] Accordingly, each step within a biosynthetic pathway that leads to the production of a fatty acid derivative substrate can be modified to produce or overproduce the substrate of interest. For example, known genes involved in the fatty acid biosynthetic pathway or the fatty aldehyde pathway can be expressed, overexpressed, or attenuated in host cells to produce a desired substrate (see, *e.g.*, PCT/US08/058788). Exemplary genes are provided in Figure 9.

Synthesis of Substrates

[0292] 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)).

[0293] 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-oxoglutarate 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.

[0294] Malonyl-CoA overexpression can be affected by introducing *accABCD* (*e.g.*, accession number AAC73296, EC 6.4.1.2) into a host cell. Fatty acids can be further overexpressed in host cells by introducing into the host cell a DNA sequence encoding a lipase (*e.g.*, accession numbers CAA89087, CAA98876).

[0295] 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.*, accABCD, fabH, and fabI). The *plsB* (*e.g.*, accession number AAC77011) D311E mutation can be used to increase the amount of available fatty acids.

[0296] 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).

[0297] The chain length of a fatty acid derivative substrate can be selected for by modifying the expression of selected thioesterases. Thioesterase 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 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 and 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 california*	*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

[0298] In other instances, a fatty aldehyde biosynthetic polypeptide, variant, or a fragment thereof, is expressed 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 level of an acyl-CoA synthase relative to a corresponding wild-type host cell. In one embodiment, the level of expression of one or more genes (*e.g.*, an acyl-CoA synthase gene) is reduced by genetically engineering a "knock out" host cell.

[0299] 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 Alcohols

[0300] Fatty alcohols can be produced from fatty aldehydes that contain branch points by using branched fatty acid derivatives as substrates for a fatty aldehyde biosynthetic polypeptide described herein. 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 provide 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 genes encoding 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.

[0301] 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.

[0302] 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.

[0303] 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. Bacteriol.* 177:3504, 1995), which consists of E1α/β (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 *E1 α/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) *bkdB1* (E1β) *bkdC1* (E2)	NP_628006 NP_628005 NP_638004
Streptomyces coelicolor	*bkdA2* (E1α) *bkdB2* (E1β) *bkdC2* (E2)	NP_733618 NP_628019 NP_628018
Streptomyces avermitilis	*bkdA* (E1a) *bkdB* (E1b) *bkdC* (E2)	BAC72074 BAC72075 BAC72076
Streptomyces avermitilis	*bkdF* (E1α) *bkdG* (E1β) *bkdH* (E2)	BAC72088 BAC72089 BAC72090
Bacillus subtilis	*bkdAA* (E1α) *bkdAB* (E1β) *bkdB* (E2)	NP_390288 NP_390288 NP_390288
Pseudomonas putida	*bkdA1* (E1α) *bkdA2* (E1β) *bkdC* (E2)	AAA65614 AAA65615 AAA65617

[0304] 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

[0305] 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 candidates 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 Alcohols

[0306] Cyclic fatty alcohols can be produced from cyclic fatty aldehydes using cyclic fatty acid derivatives as substrates for a fatty aldehyde biosynthetic polypeptide described herein. 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 ω-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). The genes listed in Table 4 can then be expressed to allow initiation and elongation of ω-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
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).

[0307] The genes listed in Table 4 (*fabH*, *acp*, and *fabF*) allow initiation and elongation of ω-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 Alcohol Saturation Levels

[0308] The degree of saturation in fatty acids (which can then be converted into fatty aldehydes and then fatty alcohols as described herein) 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. Figure 9 lists non-limiting examples of genes in these gene families that may be used in the methods and host cells described herein.

[0309] 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). The gene *fabB* 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 aldehydes.

[0310] 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.

[0311] In yet other examples, host cells can be engineered to produce saturated fatty acids by reducing the expression of an *sfa*, *gns*, and/or *fab* gene.

[0312] In some instances, a host cell can be engineered to express an attenuated level of a dehydratase/isomerase and/or a ketoacyl-ACP synthase. 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 nonlimiting 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. The saturated fatty acids can be used to produce fatty acid derivatives, such as fatty aldehydes, and subsequently saturated fatty alcohols, as described herein.

Production of Fatty Alcohols

[0313] A fatty aldehyde described herein can be converted into a fatty alcohol by an alcohol dehydrogenase. In some examples, a gene encoding a fatty aldehyde biosynthetic polypeptide described herein can be expressed in a host cell that expresses an endogenous alcohol dehydrogenase capable of converting a fatty aldehyde produced by the fatty aldehyde biosynthetic polypeptide into a corresponding fatty alcohol. In other instances, a gene encoding a fatty alcohol biosynthetic polypeptide described herein, such as an amino acid sequence listed in Figure 10 or a variant thereof, can be expressed in a host cell. Exemplary fatty alcohol biosynthetic genes include, but are not limited to, AlrA of *Acenitobacter sp.* M-1 or AlrA homologs; and endogenous *E. coli* alcohol dehydrogenases such as DkgA (NP_417485), DkgB (NP_414743), YjgB, (AAC77226), YdjL (AAC74846), YdjJ (NP_416288), AdhP (NP_415995), YhdH (NP_417719), YahK (NP_414859), YphC (AAC75598), and YqhD (Q46856). In other instances, a gene encoding a fatty

alcohol biosynthetic polypeptide can be co-expressed in a host cell with a gene encoding a fatty aldehyde biosynthetic polypeptide described herein.

Genetic Engineering of Host Cells to Produce Fatty Alcohols

[0314] Various host cells can be used to produce fatty alcohols, as described herein. A host cell can be any prokaryotic or eukaryotic cell. For example, a gene encoding a polypeptide described herein (*e.g.*, a fatty aldehyde biosynthetic polypeptide and/or a fatty alcohol biosynthetic polypeptide) 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.

[0315] 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.

[0316] Additional host cells that can be used in the methods described herein are described in WO2009/111513 and WO2009/111672.

[0317] Various methods well known in the art can be used to genetically engineer host cells to produce fatty alcohols. The methods can include the use of vectors, preferably expression vectors, containing a nucleic acid encoding a fatty aldehyde biosynthetic polypeptide and/or a fatty alcohol 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.

[0318] 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.

[0319] Recombinant expression vectors can be designed for expression of a gene encoding a fatty aldehyde biosynthetic polypeptide (or variant) and/or a gene encoding a fatty alcohol biosynthetic polypeptide 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.

[0320] Expression of genes encoding 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 Biotech Inc; Smith *et al.*, *Gene* (1988) 67:31-40), pMAL (New England Biolabs, Beverly, 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.

[0321] Examples of inducible, non-fusion *E. coli* expression vectors include pTrc (Amann *et al.*, *Gene* (1988) 69:301-315) and pET 11d (Studier *et al.*, Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego, Calif. (1990) 60-89). Target gene expression from the pTrc vector relies on host RNA polymerase transcription from a hybrid trp-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 λ prophage harboring a T7 gn1 gene under the transcriptional control of the lacUV 5 promoter.

[0322] 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. (1990) 119-128). 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.* (1992) 20:2111-2118). Such alteration of nucleic acid sequences can be carried out by standard DNA synthesis techniques.

[0323] 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.* (1987) 6:229-234), pMFa (Kurjan *et al.*, *Cell* (1982) 30:933-943), pJRY88 (Schultz *et al.*, *Gene* (1987) 54:113-123), pYES2 (Invitrogen Corporation, San Diego, Calif.), and picZ (Invitrogen Corp, San Diego, Calif.).

[0324] 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.* (1983) 3:2156-2165) and the pVL series (Lucklow *et al.*, *Virology* (1989) 170:31-39).

[0325] 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* (1987) 329:840) and pMT2PC (Kaufman *et al.*, *EMBO J.* (1987) 6:187-195). 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.

[0326] 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).

[0327] 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, 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).

[0328] 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

[0329] Transport proteins can export polypeptides and organic compounds (*e.g.*, fatty alcohols) 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 hydrocarbons.

[0330] 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 are listed in Figure 9 (*e.g.*, CER5, AtMRP5, AmiS2, and AtPGP1). 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.

Fermentation

[0331] The production and isolation of fatty alcohols 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 hydrocarbon products.

[0332] 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.

[0333] 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 translesion synthesis and also serve as a DNA sequence damage checkpoint. The *umuDC* gene products include UmuC, UmuD, umuD', UmuD'$_2$C, 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 aldehyde 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.

[0334] The percentage of input carbons converted to fatty alcohols can be a cost driver. The more efficient the process is (*i.e.*, the higher the percentage of input carbons converted to fatty alcohols), 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 alcohols 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.

[0335] 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. Patent 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.

[0336] 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. The expression of one or more NADH:NADPH transhydrogenases converts the NADH produced in glycolysis to NADPH, which can enhance the production of fatty alcohols.

[0337] 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, or 10 L; fermented; and induced to express desired fatty aldehyde biosynthetic genes and/or an alcohol dehydrogenase 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 10 L, 100 L, 1000 L, 10,000 L, 100,000 L, 1,000,000 L or larger; fermented; and induced to express desired fatty aldehyde biosynthetic genes and/or alcohol dehydrogenase genes based on the specific genes encoded in the appropriate plasmids or incorporated into the host cell's genome.

[0338] For example, a suitable production host, such as *E. coli* cells, harboring plasmids containing the desired genes or having the genes integrated in its chromosome can be incubated in a suitable reactor, for example a 1 L reactor, for 20 hours 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 alcohol After incubation, the spent media can be extracted and the organic phase can be examined for the presence of fatty alcohols using GC-MS.

[0339] 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 alcohols to rise to the surface and undergo a spontaneous phase separation or precipitation. The fatty alcohol 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.

Producing Fatty Alcohols using Cell-free Methods

[0340] In some methods described herein, a fatty alcohol can be produced using a purified polypeptide (*e.g.*, a fatty alcohol biosynthetic polypeptide) described herein and a substrate (*e.g.*, fatty aldehyde), produced, for example, by a method described herein. For example, a host cell can be engineered to express a fatty alcohol biosynthetic polypeptide or variant as described herein. The host cell can be cultured under conditions suitable to allow expression of the polypeptide. Cell free extracts can then be generated using known methods. For example, the host cells can be lysed using detergents or by sonication. The expressed polypeptides can be purified using known methods. After obtaining the cell free extracts, substrates described herein can be added to the cell free extracts and maintained under conditions to allow conversion of the substrates (*e.g.*, fatty aldehydes) to fatty alcohols. The fatty alcohols can then be separated and purified using known techniques.

[0341] In some instances, a fatty aldehyde described herein can be converted into a fatty alcohol by contacting the fatty aldehyde with a fatty alcohol biosynthetic polypeptide listed in Figure 10, or a variant thereof.

Post-Production Processing

[0342] The fatty alcohols produced during fermentation can be separated from the fermentation media. Any known technique for separating fatty alcohols from aqueous media can be used. One exemplary separation process is a two phase (bi-phasic) separation process. This process involves fermenting the genetically engineered host cells under conditions sufficient to produce a fatty alcohols, allowing the fatty alcohol 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.

[0343] Bi-phasic separation uses the relative immiscibility of fatty alcohols 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 alcohol being produced has a high logP value, the

fatty alcohol can separate into the organic phase, even at very low concentrations, in the fermentation vessel.

[0344] The fatty alcohols produced by the methods described herein can be relatively immiscible in the fermentation broth, as well as in the cytoplasm. Therefore, the fatty alcohol 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 alcohol on cellular function and can allow the host cell to produce more product.

[0345] 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 alcohols 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.

[0346] In some embodiments, the fatty alcohols 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 alcohols produced using methods described herein can contain between about 65% and about 85% carbon; or between about 10% and about 15% hydrogen.

Surfactant and Detergent Compositions and Bioproducts

[0347] The fatty alcohols described herein can be used as or converted into a surfactant or detergent composition. One of ordinary skill in the art will appreciate that, depending upon the intended purpose of the surfactant or detergent, different fatty alcohols can be produced and used. For example, when the fatty alcohols described herein are used as a feedstock for surfactant or detergent production, one of ordinary skill in the art will appreciate that the characteristics of the fatty alcohol feedstock will affect the characteristics of the surfactant or detergent produced.

Hence, the characteristics of the surfactant or detergent product can be selected for by producing particular fatty alcohols for use as a feedstock.

[0348] Bioproducts (*e.g.*, fatty alcohols) comprising biologically produced organic compounds, particularly fatty alcohols biologically produced using the fatty acid biosynthetic pathway, 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. Patent No. 7,169,588.

[0349] 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 instant materials may be followed in commerce on the basis of their unique carbon isotope profile.

[0350] 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.

[0351] 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.

[0352] 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.

[0353] 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.

[0354] 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, ‰, and are calculated as follows:

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

[0355] 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.

[0356] The compositions described herein include bioproducts produced by any of the methods described herein. 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.

[0357] Bioproducts 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) (1992) 3-74).

[0358] 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.)

[0359] 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.

[0360] The compositions described herein include bioproducts that can have an f_M ^{14}C of at least about 1. For example, the bioproduct 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.

[0361] 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.

[0362] 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.

[0363] A bioproduct 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.

[0364] Fuel additives are used to enhance the performance of a fuel or engine. For example, fuel additives can be used to alter the freezing/gelling point, cloud point, lubricity, viscosity, oxidative stability, ignition quality, octane level, and/or flash point. In the United States, all fuel additives must be registered with Environmental Protection Agency. The names of fuel additives and the companies that sell the fuel additives are publicly available by contacting the EPA or by viewing the agency's website. One of ordinary skill in the art will appreciate that the fatty alcohol-based biofuels described herein can be mixed with one or more fuel additives to impart a desired quality.

[0365] The fatty alcohol-based surfactants and/or detergents described herein can be mixed with other surfactants and/or detergents well known in the art.

[0366] In some examples, the mixture can include at least about 10%, 15%, 20%, 30%, 40%, 50%, or 60% by weight of the fatty alcohol. In other examples, a surfactant or detergent composition can be made that includes at least about 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 85%, 90% or 95% of a fatty alcohol that includes a carbon chain that is 8, 10, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21 or 22 carbons in length. Such surfactant or detergent compositions can additionally include at least one additive selected from a surfactant; a microemulsion; at least about 5%, 10%, 15%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 85%, 90%, or 95% of surfactant or detergent from nonmicrobial sources such as plant oils or petroleum.

[0367] 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

Identification of Carboxylic Acid Reductase (CAR) Homologs

[0368] The carboxylic acid reductase (CAR) from *Nocardia sp.* strain NRRL 5646 can reduce carboxylic acids (e.g., fatty acids) into their corresponding aldehydes without utilizing separate activating enzymes, such as acyl-CoA synthases (Li *et al.*, *J. Bacteriol.* 179:3482-3487, 1997; He *et al.*, *Appl. Environ. Microbiol.* 70:1874-1881, 2004)).

[0369] A BLAST search using the NRRL 5646 CAR amino acid sequence (Genpept accession AAR91681) (SEQ ID NO:16) as the query sequence identified approximately 20 homologous sequences. Three homologs, listed in Table 7, were evaluated for their ability to convert fatty acids into fatty aldehydes *in vivo* when expressed in *E. coli*.

[0370] At the nucleotide sequence level, *carA* (SEQ ID NO:19), *carB* (SEQ ID NO:21), and *fadD9* (SEQ ID NO:17) demonstrated 62.6%, 49.4%, and 60.5% homology, respectively, to the *car* gene (AY495697) of *Nocardia sp.* NRRL 5646 (SEQ ID NO:15). At the amino acid level, CARA (SEQ ID NO:20), CARB (SEQ ID NO:22), and FadD9 (SEQ ID NO:18) demonstrated 62.4%, 59.1% and 60.7% identity, respectively, to CAR of *Nocardia sp.* NRRL 5646 (SEQ ID NO:16).

Table 7: CAR-like Protein and the corresponding coding sequences.

Genpept accession	Locus_tag	Annotation in GenBank	Gene name
NP_217106	Rv 2590	Probable fatty-acid-CoA ligase (FadD9)	*fadD9*
ABK75684	MSMEG 2956	NAD dependent epimerase/dehydratase family protein	*carA*
YP_889972.1	MSMEG 5739	NAD dependent epimerase/dehydratase family protein	*carB*

EXAMPLE 2

Identification of Alcohol Dehydrogenase Genes

Reverse Engineering

[0371] *E. coli* contains at least one enzyme that catalyzes the reversible oxidoreduction of fatty aldehydes and fatty alcohols (i.e. fatty aldehyde reductase / alcohol dehydrogenase). Reverse engineering was used to identify such fatty aldehyde reductases / fatty alcohol dehydrogenases in *E. coli* MG1655 cells expressing the acyl-ACP reductase YP_400611 from *Synechococcus elongatus* (Synpcc7942_1594) (SEQ ID NO:196). Four 3 mL LB cultures were grown overnight at 37°C, and 55 µL of stationary phase cultures were used to inoculate four independent 5.5 mL of LB. Those 5.5 mL cultures were then grown to an OD_{600} of 0.8-1.0 and were then used to inoculate a corresponding number of 2 L baffled shakeflasks, each with 500 mL Hu-9 minimal media. 20 hrs after induction the cells were pelleted at 4,000 x g for 20 min. The cell pellet was resuspended in 30 mL of 100 mM phosphate buffer at pH 7.2 with 1x Bacterial Protease Arrest (G Biosciences). The cells were lysed in a french press at 15,000 psi with two passes through the instrument. The cell debris was then removed by centrifuging at 10,000 x g for 20 mins. The cell lysate was loaded onto two HiTrapQ™ columns (GE Healthcare) connected in series. The following buffers were used to elute proteins: (A) 50 mM Tris, pH 7.5 and (B) 50 mM Tris, pH 7.5 with 1 M NaCl. A gradient from 0 % B to 100% B was run over 5 column volumes at a flow rate of 3 mL/min while 4 mL fractions were collected.

[0372] The fractions were assayed for alcohol dehydrogenase activity by taking 190 µL of a protein fraction and adding 5 µL of a 20 mM NADPH (Sigma) solution and 5 µL of a 20 mM dodecanal (Fluka) solution in DMSO. The reactions were incubated at 37°C for 1 hr. They were then extracted with 100 µL of ethyl acetate and analyzed for dodecanol via GC/MS. Fractions eluting around 350 mM NaCl contained alcohol dehydrogenase activity.

[0373] Fractions containing alcohol dehydrogenase activity were pooled and loaded onto a 1 mL ResourceQ™ column (GE Healthcare). The same conditions used for the HiTrapQ™ column were used, except 0.5 mL fractions were collected. Protein fractions demonstrating alcohol dehydrogenase activity were then pooled and

concentrated using Amicon™ (Millipore) protein concentrators (10,000 kDa cutoffs) to a volume of 1 mL. The solution was then loaded onto a HiPrep™ 200 size exclusion column (GE Healthcare). A buffer solution containing 50 mM Tris, pH 7.5, and 150 mM NaCl was run through the column at a rate of 0.3 mL per min. 2 mL fractions were collected. Two protein fractions contained alcohol dehydrogenase activity. These two fractions, plus fractions before and after these two fractions, were loaded onto a polyacrylamide gel and stained with SimplySafe Coomassie stain (Invitrogen).

[0374] Comparing the bands in the active and inactive fractions, one protein band appeared in the active fractions that was not seen in the inactive fraction. This protein band was cut from gel and submitted to the Stanford Mass Spectroscopy Facility for LC/MS/MS protein sequencing. One of the proteins identified in this analysis was YahK.

[0375] To verify that YahK was indeed an alcohol dehydrogenase, *yahK* was knocked out in *E. coli* MG1655(DE3, *ΔfadD, ΔyjgB)* (control strain) (described in Example 4). The *yahK* knock-out strain MG1655(DE3, *ΔfadD, Δyjg,B ΔyahK)* was constructed using the lambda red system (described in Example 4) with the following primers:

yahK_F
(CATATCAGGCGTTGCCAAATACACATAGCTAATCAGGAGTAAAC
ACAATG) (SEQ ID NO:197) and
yahK_R
(AATCGCACACTAACAGACTGAAAAAATTAATAAATACCCTGTGG
TTTAAC) (SEQ ID NO:198).

This *ΔyahK* strain and the control strain, both expressing the acyl-ACP reductase YP_400611, were cultured under conditions described above. Cell free lysates were made from both strains, and each lysate was assayed for alcohol dehydrogenase activity as discussed above.

[0376] The *ΔyahK* strain did not convert dodecanal to dodecanol, while the wild type strain had this activity. For additional verification, each lysate was run on a HiTrapQ™ column as described above. The wild type lysate had alcohol

dehydrogenase activity in fractions eluting around 350 mM NaCl, while the *ΔyahK* lysate had no alcohol dehydrogenase activity in this region.

Bioinformatics

[0377] It was reasoned that possible alcohol dehydrogenases in *E. coli* were members of four protein families: Zn-dependent alcohol dehydrogenases (Pfam 00107 and 08240), Fe-dependent alcohol dehydrogenases (Pfam 00465), aldo-keto reductases (Pfam 00248) and short-chain dehydrogenases (Pfam 00106) (Pfam = protein family according to "pfam.sanger.ac.uk"). Using the Pfam motifs, all members of these four protein families in *E .coli* were identified (listed in Figure 10). From this list, the following 8 candidates were chosen for experimental analysis: *yahK*, *yjgB*, *adhP*, *dkgA*, *dkgB*, *yhdH*, *ydjL*, and *yqhD*.

[0378] To determine if these genes could reduce fatty aldehydes to fatty alcohols, these 8 genes were cloned into a pET-Duet vector along with *E. coli 'tesA*. These genes were then transformed into *E. coli* (DE3) MG1655 *ΔyjgBΔyahK* cells. Next 3 mL overnight starter cultures were grown in LB with carbanecillin (100 mg/L) at 37 °C. A control strain lacking a candidate alcohol dehydrogenase was also included in the experiment. 1 mL of each overnight culture was used to inoculate 50 mL of fresh LB with carbanecillin. The cultures were shaken at 37 °C until reaching an OD_{600} of 0.8-1. The cultures were then transferred to 18 °C, induced with 1 mM IPTG, and shaken overnight.

[0379] Cell free lysates were prepared by centrifuging the cultures at 4,000 x g for 20 mins. The cultures were then resuspended in 1 mL of Bugbuster (Novagen) and gently shaken at room temperature for 5 min. The cell debris was removed by spinning at 15,000 x g for 10 min. The resulting lysates were assayed for alcohol dehydrogenase activity by mixing 88 μL of lysate, 2 μL of 40 mM cis-11-hexadecenal in DMSO, and 10 μL of 20 mM NADPH. The samples were incubated at 37 °C for 30 min. and were then extracted with 100 μL of ethyl acetate. The extracts were analyzed using GC/MS.

[0380] All proteins showed significantly better conversion of cis-11-hexadecenal to cis-11-hexadecanol as compared with the 'TesA only control (see Table 8). These

results were confirmed in assays using dodecanal instead of cis-11-hexadecenal as the substrate (see Table 8).

[0381] To investigate how these enzymes contribute to fatty alcohol dehydrogenase activity in *E. coli* under production conditions, first the *yjgB yahK* double knock-out strain in MG1655(DE3, Δ*fadD*) (described above) was tested by transforming it with a plasmid expressing acyl-ACP reductase YP_400611 and analyzing fatty aldehyde and fatty alcohol titers. The test strain also contained a plasmid expressing a decarbonylase. This double knock-out mutant showed slightly higher fatty aldehyde titers in several experiments (see, *e.g.*, Figure 11), confirming that these two putative alcohol dehydrogenases contribute to fatty alcohol dehydrogenase activity in *E. coli* under production conditions (see also Example 4 for similar results from a MG1655(DE3, Δ*fadD Δyjg B)* strain). Next, two additional genes, *yncB* and *ydjA*, were deleted in the *yjgB yahK* double mutant. *YdjA*, which is not a member of the four protein families mentioned above, demonstrated slightly elevated fatty aldehyde levels (see Figure 11), indicating that it may also contribute to fatty alcohol dehydrogenase activity in *E. coli* under production conditions.

[0382] Additionally, the active fatty alcohol dehydrogenases from Table 8 were also deleted in MG1655 (DE3, Δ*fadD*, *Δyjg,B ΔyahK*) and tested as described above. Several of these deletion strains showed slightly elevated fatty aldehyde levels, suggesting that these may also contribute to fatty alcohol dehydrogenase activity in *E. coli* under production conditions (see Figure 12).

Table 8: Overexpression of putative fatty alcohol dehydrogenase genes

	GC/MS Assay		NADPH assay
	% conversion to corresponding alcohol		initial rate (slope)
substrate	dodecanal	*cis* 11-hexadecenal	*cis* 11-hexadecenal
Overexpression:			
none	9	12	0.2
YjgB	54	89	24.8
YahK	47	87	28.3
AdhP	52	45	4.1
YdjL	51	23	0.14
YhdH	59	74	13.7
YqhD	55	23	7.3
yafB (dkgB)	52	65	9.4
YqhE (dkgA)	45	50	9.6

EXAMPLE 3

Expression of CAR Homologs and Alcohol Dehydrogenase in *E. coli*

A. CAR Plasmid Construction

[0383] Three *E. coli* expression plasmids were constructed to express the genes encoding the CAR homologs listed in Table 7. First, *fadD9* was amplified from genomic DNA of *Mycobacterium tuberculosis* H37Rv (obtained from The University of British Columbia, and Vancouver, BC Canada) using the primers fadD9F and FadDR (see Table 9). The PCR product was first cloned into PCR-blunt (Invitrogen) and then released as an *Nde*I-*Avr*II fragment. The *Nde*I-*Avr*II fragment was then cloned between the *Nde*I and *Avr*II sites of pACYCDuet-1 (Novogen) to generate pACYCDuet-1-fadD9.

[0384] The *carA* gene was amplified from the genomic DNA of *Mycobacterium smegmatis* MC2 155 (obtained from the ATCC (ATCC 23037D-5)) using primers CARMCaF and CARMCaR (see Table 9). The *carB* gene was amplified from the genomic DNA of *Mycobacterium smegmatis* MC2 155 (obtained from the ATCC (ATCC 23037D-5)) using primers CARMCbF and CARMCbR (see Table 9). Each PCR product was first cloned into PCR-blunt and then released as an *Nde*I-*Avr*II fragment. Each of the two fragments was then subcloned between the *Nde*I and *Avr*II

sites of pACYCDuet-1 (Novogen) to generate pACYCDuet-1-carA and pACYCDuet-1-carB.

Table 9. Primers used to amplify genes encoding CAR homologs

fadD9F	cat ATGTCGATCAACGATCAGCGACTGAC (SEQ ID NO:1)
fadD9R	cctagg TCACAGCAGCCCGAGCAGTC (SEQ ID NO:2)
CARMCaF	cat ATGACGATCGAAACGCG (SEQ ID NO:3)
CARMCaR	cctagg TTACAGCAATCCGAGCATCT (SEQ ID NO:4)
CARMCbF	cat ATGACCAGCGATGTTCAC (SEQ ID NO:5)
CARMCbR	cctagg TCAGATCAGACCGAACTCACG (SEQ ID NO:6)

B. Alcohol Dehydrogenase Plasmid Construction

[0385] The plasmid pETDuet-1-'tesA-yjgB carries *'tesA* and *yjgB* (a putative alcohol dehydrogenase; GenBank accession number, NP_418690; GenPept accession number AAC77226) from the *E. coli* K12 strain.

[0386] The gene *yjgB* (GenBank accession number, NP_418690) was amplified from the genomic DNA of *E. coli* K-12 using the following primers.

[0387] The *yjgB* insert was generated by PCR using the following primers:

NcoI YjgB forward:

aatccTGGCATCGATGATAAAAAGCTATGCCGCAAAAG (SEQ ID NO:199)

HindIII YjgB reverse:

ataaagctTTCAAAAATCGGCTTTCAACACCACGCGG (SEQ ID NO:200)

The PCR product was then subcloned into the *Nco*I and *Hind*III sites of pETDuet-1-'tesA to generate pETDuet-1-'tesA-yjgB.

[0388] The plasmid pETDuet-1-'tesA-alrAadp1 carries *'tesA* and *alrAadp1* (GenPept accession number CAG70248.1) from *Acinetobacter baylyi* ADP1.

[0389] The gene *alrAadp1* was amplified from the genomic DNA of *Acinetobacter baylyi* ADP1 by a two-step PCR procedure. The first set of PCR reactions eliminated an internal *Nco*I site at bp 632-636 with the following primer pairs:

ADP1 Alr mut1 reverse:

5'-GACCACGTGATCGGCCCCCATAGCTTTGAGCTCATC (SEQ ID NO:201)

ADP1 Alr1 mut1 forward:

5'-GATGAGCTCAAAGCTATGGGGGCCGATCACGTGGTC (SEQ ID NO:202)

The PCR products were then isolated, purified using the Qiagen gel extraction kit, and used as inputs for a second PCR reaction with the following primers to produce full-length *AlrAadp1* with a C→T mutation at position 633:

*Nco*I ADP1 Alr1 forward:

5'-AATACCATGGCAACAACTAATGTGATTCATGCTTATGCTGCA (SEQ ID NO:203)

*Hind*III ADP1 Alr1 reverse:

5'-ATAAAAGCTTTTAAAAATCGGCTTTAAGTACAATCCGATAAC (SEQ ID NO:204)

The plasmid pETDuet-1-'tesA-alrAadp1 was prepared by inserting the *alrAadp1* gene (gene locus-tag= "ACIAD3612"), a homolog of *Acinetobacter baylyi* ADP1, into the *Nco*I and *Hind*III sites of pETDuet-1-'tesA.

B. Evaluation of Fatty Aldehyde and Fatty Alcohol Production

[0390] In order to evaluate the affect of carboxylic acid reductases and alcohol dehydrogenases on the production of fatty alcohols, various combinations of the prepared plasmids were transformed in the *E. coli* strain C41 (DE3, Δ*fadE*) (described in PCT/US08/058788).

[0391] For example, the plasmid pACYCDuet-1-carA, encoding the CAR homolog carA, was co-transformed with pETDuet-1-'tesA-alrAadp1 (see, *e.g.*, Figure 1).

[0392] The plasmid pACYCDuet-1-carB, encoding the CAR homolog carB, was co-transformed with pETDuet-1-'tesA. In addition, pACYCDuet-1-carB was also

separately co-transformed with pETDuet-1-'tesA-yjgB and pETDuet-1-'tesA-alrAadp1. As a control, pACYCDuet-1-carB was co-transformed with the empty vector pETDuet-1 (see, *e.g.*, Figure 1).

[0393] The plasmid pACYCDuet-1-fadD9, encoding the CAR homolog fadD9, was co-transformed with pETDuet-1-'tesA. In addition, pACYCDuet-1- fadD9 was also separately co-transformed with pETDuet-1-'tesA-yjgB and pETDuet-1-'tesA-alrAadp1. As a control, pACYCDuet-1- fadD9was co-transformed with the empty vector pETDuet-1 (see, *e.g.*, Figure 1).

[0394] As an additional control, pETDuet-1-'tesA-yjgB was co-transformed with the empty vector pACYCDuet-1.

[0395] The *E. coli* transformants were grown in 3 mL of LB medium supplemented with carbenicillin (100 mg/L) and chloramphenicol (34 mg/L) at 37 °C. After overnight growth, 15 μL of culture was transferred into 2 mL of fresh LB medium supplemented with carbenicillin and chloramphenicol. After 3.5 hours of growth, 2 mL of culture were transferred into a 125 mL flask containing 20 mL of M9 medium with 2% glucose and with carbenicillin and chloramphenicol. When the OD_{600} of the culture reached 0.9, 1 mM of IPTG was added to each flask. After 20 hours of growth at 37 °C, 20 mL of ethyl acetate (with 1% of acetic acid, v/v) was added to each flask to extract the fatty alcohols produced during the fermentation. The crude ethyl acetate extract was directly analyzed with GC/MS as described herein.

[0396] The measured retention times were 6.79 minutes for cis-5-dodecen-1-ol, 6.868 minutes for 1-dodecanol, 8.058 minutes for cis-7-tetradecen-1-ol, 8.19 minutes for 1-tetradecanol, 9.208 minutes for cis-9-hexadecen-1-ol, 9.30 minutes for 1-hexadecanol, and 10.209 minutes for cis-11-octadecen-1-ol.

[0397] The co-expression of the leaderless tesA and any of the three *car* genes in *E. coli* resulted in high titers of fatty alcohols and detectable fatty aldehyde production (Figures 1, 2, 5). The expression of *carA* or *carB* with the leaderless *tesA* and *alrAadp1* resulted in fatty alcohol titers of greater than 700 mg/L and reduced fatty aldehyde production. Likewise, *fadD9* co-expressed with the leaderless *tesA* and *alrAadp1* produced over 300 mg/L of fatty alcohol. When expressed without the leaderless *tesA*, neither *carB* nor *fadD9* produced more than 10 mg/L of fatty alcohols

(possibly resulting from the accumulation of free fatty acids in the cell due to endogenous *tesA*). Taken together, this data indicates that fatty acids are the substrates for these CAR homologs and that overexpression of a thioesterase, such as *'tesA* (to release fatty acids from acyl-ACP), achieves significant production of fatty alcohols.

[0398] In one fermentation, *E. coli* strain C41 (DE3, Δ*fadE*) co-transformed with pACYCDuet-1- carB+ pETDuet-1-tesA produced an average of 695 mg/L of fatty alcohols and 120 mg/L of fatty aldehydes. The presence of large amounts of fatty aldehydes is consistent with CAR being an aldehyde-generating, fatty acid reductase (AFAR). This mechanism is different from alcohol-generating fatty acyl-CoA reductases (FAR), represented by JjFAR, and fatty acyl-CoA reductases, represented by Acr1.

[0399] The production of fatty alcohols from fatty aldehydes in the *E. coli* strains described above may have been catalyzed by an endogenous alcohol dehydrogenase(s). *E. coli* produces an alcohol dehydrogenase(s) (*e.g.*, *yjgB*) capable of converting fatty aldehydes of various chain-length into fatty alcohols (Naccarato *et al.*, *Lipids* 9: 419-428 (1974); Reiser *et al.*, *J. Bacteriol.* 179: 2969-2975 (1997); Venkitasubramanian *et al.*, *J. Biol. Chem*. 282:478-485 (2007)).

[0400] Therefore, alcohols dehydrogenases may also play a role in the fatty alcohol biosynthetic pathway in addition to carboxylic acid reductases. For example, expression of either *yjgB* or *alrAadp1* with *carB* and the leaderless *tesA* significantly reduced the accumulation of fatty aldehydes, compared to strains that did not overexpress *yjgB* or *alrAadp1* (Figure 2).

[0401] Following the fermentations where pACYCDuet-1-carB was transformed in *E. coli* strain C41 (DE3, Δ*fadE*), a white, round, disk-like deposit was observed at the bottom center of the flasks used for fatty alcohol production with recombinant *E. coli* strains. In contrast, no such deposits were observed at the bottom of the control flasks that did not express *car* homologs. GC/MS analysis of the deposit dissolved in ethyl acetate (with 1% of acetic acid, v/v) revealed that the deposit was a fatty alcohol deposit.

C. Types of Fatty Alcohols Produced by Different CAR Homologs

[0402] Depending upon the CAR homolog expressed in *E. coli* strain C41 (DE3, Δ*fadE*), different mixtures of fatty alcohols were produced. Different compositions of fatty alcohols were observed among the three CAR homologs evaluated (see Table 10). FadD9 produced more C_{12} fatty alcohols relative to other fatty alcohols with carbon chain lengths greater than 12. Both CarA and CarB produced a wider range in chain length of fatty alcohols than was observed when expressing FadD9.

Table 10. Acyl-composition of fatty alcohols produced by recombinant *E. coli* strains

Expressed with TesA* and AlrAadp1	Acyl-composition of fatty alcohols (%)						
	C10:0	C12	C14:1	C14:0	C16:1	C16:0	C18:1
CarA	trace	38	13	27	16	4	3
FadD9	trace	63	14	16	7	trace	trace
CarB	trace	32	11	41	12	trace	trace

* the leaderless TesA. C12, including C12:0 and C12:1 fatty alcohol.

D. Quantification and Identification of Fatty Alcohols

[0403] GC-MS was performed using an Agilent 5975B MSD system equipped with a 30m×0.25mm (0.10μm film) DB-5 column. The column temperature was 3 min isothermal at 100°C. The column was programmed to rise from 100 °C to 320 °C at a rate of 20 °C/min. When the final temperature was reached, the column remained isothermal for 5 minutes at 320°C. The injection volume was 1 μL. The carrier gas, helium, was released at 1.3 mL/min. The mass spectrometer was equipped with an electron impact ionization source. The ionization source temperature was set at 300 °C.

[0404] Prior to quantification, various alcohols were identified using two methods. First, the GC retention time of each compound was compared to the retention time of a known standards, such as cetyl alcohol, dodecanol, tetradecanol, octadecanol, and cis-9-octadecenol. Second, identification of each compound was confirmed by matching the compound's mass spectrum to a standard's mass spectrum in the mass spectra library (*e.g.,* C12:0, C12:1, C13:0, C14:0, C14:1, C15:0. C16:0, C16:1, C17:0, C18:0 and C18:1 alcohols).

EXAMPLE 4

Production of Fatty Alcohol by Heterologous Expression of CAR Homologs in *E. coli* MG1655 (DE3, *ΔfadD*)

Construction of fadD deletion strain

[0405] The *fadD* gene of *E. coli* MG1655 was deleted using the lambda red system (Datsenko et al., 2000, Proc. Natl. Acad. Sci. USA. 97: 6640-6645) as follows: The chloramphenicol acetyltransferase gene from pKD3 was amplified with the primers fad1 (5`-TAACCGGCGTCTGACGACTGACTTAACGCTCAGGCTTTATTGTCCACT TTGTGTAGGCTGGAGCTGCTTCG-3`) (SEQ ID NO:205), and fad2 (5`-CATTTGGGGTTGCGATGACGACGAACACGCATTTTAGAGGTGAAGAA TTGCATATGAATATCCTCCTTTAGTTCC-3`) (SEQ ID NO:206).
This PCR product was electroporated into *E. coli* MG1655 (pKD46). The cells were plated on L-chloramphenicol (30 μg/mL)(L-Cm) and grown overnight at 37 °C. Individual colonies were picked on to another L-Cm plate and grown at 42 °C. These colonies were then patched to L-Cm and L- carbenicillin (100 mg/mL) (L-Cb) plates and grown at 37 °C overnight. Colonies that were Cm^R and Cb^S were evaluated further by PCR to ensure the PCR product inserted at the correct site. PCR verification was performed on colony lysates of these bacteria using the primers fadF (5`- CGTCCGTGGTAATCATTTGG-3`) (SEQ ID NO:207) and fadR (5`-TCGCAACCTTTTCGTTGG-3`) (SEQ ID NO:208). Expected size of the *ΔfadD*::Cm deletion was about 1200 bp (Figure 4). The chloramphenicol resistance gene was eliminated using a FLP helper plasmid as described in Datsenko *et al.*, *Proc. Natl. Acad. Sci. USA* 97:6640-6645 (2000). PCR verification of the deletion was performed with primers fadF and fadR (Figure 4). The MG1655 *ΔfadD* strain was unable to grow on M9 + oleate agar plates (oleate as carbon source). It was also unable to grow in M9 + oleate liquid media. The growth defect was complemented by an *E. coli fadD* gene supplied in trans (in pCL1920-Ptrc).

Construction of MG1655(DE3, ΔfadD) strain

[0406] To generate a T7-responsive strain, the λDE3 Lysogenization Kit (Novagen) was utilized, which is designed for site-specific integration of λDE3

prophage into an *E. coli* host chromosome, such that the lysogenized host can be used to express target genes cloned in T7 expression vectors. λDE3 is a recombinant phage carrying the cloned gene for T7 RNA polymerase under *lacUV5* control. Briefly, the host strain was cultured in LB supplemented with 0.2% maltose, 10 mM $MgSO_4$, and antibiotics at 37 °C to an OD_{600} of 0.5. Next, 10^8 pfu λDE3, 10^8 pfu Helper Phage, and 10^8 pfu Selection Phage were incubated with 10 μL host cells. The host/phage mixture was incubated at 37 °C for 20 min to allow phage to adsorb to host. Finally, the mixture was pipeted onto an LB plate supplemented with antibiotics. The mixture was spread evenly using plating beads, and the plates were inverted plates and incubated at 37 °C overnight.

[0407] λDE3 lysogen candidates were evaluated by their ability to support the growth of the T7 Tester Phage. T7 Tester Phage is a T7 phage deletion mutant that is completely defective unless active T7 RNA polymerase is provided by the host cell. The T7 Tester Phage makes very large plaques on authentic λDE3 lysogens in the presence of IPTG, while much smaller plaques are observed in the absence of inducer. The relative size of the plaques in the absence of IPTG is an indication of the basal level expression of T7 RNA polymerase in the lysogen, and can vary widely between different host cell backgrounds.

[0408] The following procedure was used to determine the presence of DE3 lysogeny. First, candidate colonies were grown in LB supplemented with 0.2% maltose, 10 mM $MgSO_4$, and antibiotics at 37 °C to an OD_{600} of 0.5. An aliquot of T7 Tester Phage was then diluted in 1X Phage Dilution Buffer to a titer of 2×10^3 pfu/mL. In duplicate tubes, 100 μL host cells were mixed with 100 μL diluted phage. The host/phage mixture was incubated at room temperature for 10 min to allow phage to adsorb to host. Next, 3 mL of molten top agarose was added to each tube containing host and phage. The contents of one duplicate were plated onto an LB plate and the other duplicate onto an LB plate supplemented with 0.4 mM IPTG (isopropyl-b-thiogalactopyranoside) to evaluate induction of T7 RNA polymerase. Plates were allowed to sit undisturbed for 5 min until the top agarose hardened. The plates were then inverted at 30 °C overnight.

Construction of MG1655(DE3, ΔfadD, yjgB::kan) strain

[0409] The *yjgB* knockout strain, MG1655(DE3, Δ*fadD*, *yjgB*::kan), was constructed by using the following lambda red system (Datsenko *et al.*, *Proc. Natl. Acad. Sci. USA* 97:6640-6645 (2000)):

[0410] The kanamycin resistant gene from pKD13 was amplified with the primers yjgBRn (5`-GCGCCTCAGATCAGCGCTGCGAATGATTTTCAAAAATCGGCTTTCAA CACTGTAGGCTGGAGCTGCTTCG-3`) (SEQ ID NO:209), and yjgBFn (5`-CTGCCATGCTCTACACTTCCCAAACAACACCAGAGAAGGACCAAAAA ATGATTCCGGGGATCCGTCGACC-3`) (SEQ ID NO:210). The PCR product was then electroporated into *E. coli* MG1655(DE3, Δ*fadD*)/pKD46. The cells were plated on kanamycin (50 μg/mL) (L-Kan) and grown overnight at 37 °C. Individual colonies were picked on to another L-Kan plate and grown at 42 °C. These colonies were then patched to L-Kan and carbenicillin (100 mg/mL) (L-Cb) plates and grown at 37 °C overnight. Colonies that were kan^R and Cb^S were evaluated further by PCR to ensure the PCR product was inserted at the correct site. PCR verification was performed on colony lysates of these bacteria using the primers BF (5`-gtgctggcgataCGACAAAACA-3`) (SEQ ID NO:211) and BR (5`-CCCCGCCCTGCCATGCTCTACAC-3`) (SEQ ID NO:212). The expected size of the *yjgB*::kan knockout was about 1450 bp.

Evaluation of FadD on fatty alcohol production using MG1655(DE3, ΔfadD) strain

[0411] In Example 3, a *fadE* deletion strain was used for fatty aldehyde and fatty alcohol production from 'TesA, CAR homologs, and endogenous alcohol dehydrogenase(s) in *E. coli*. To demonstrate that CAR homologs used fatty acids instead of acyl-CoA as a substrate, the gene encoding for acyl-CoA synthase in *E. coli* (*fadD*) was deleted so that the fatty acids produced were not activated with CoA. *E. coli* strain MG1655(DE3, Δ*fadD*) was transformed with pETDuet-1-'tesA and pACYCDuet-1-carB . The transformants were evaluated for fatty alcohol production using the methods described herein. These transformants produced about 360 mg/L

of fatty alcohols (dodecanol, dodecenol, tetredecanol, tetredecenol, cetyl, hexadecenol, and octadecenol).

YjgB is an alcohol dehydrogenase

[0412] To confirm that YjgB was an alcohol dehydrogenase responsible for converting fatty aldehydes into their corresponding fatty alcohols, pETDuet-1-'tesA and pACYCDuet-1-fadD9 were co-transformed into either MG1655(DE3, Δ*fadD*) or MG1655(DE3, Δ*fadD*, *yjgB*::kan). At the same time, MG1655(DE3, Δ*fadD*, *yjgB*::kan) was transformed with both pETDuet-1-'tesA-yjgB and pACYCDuet-1-fadD9.

[0413] The *E. coli* transformants were grown in 3 mL of LB medium supplemented with carbenicillin (100 mg/L) and chloramphenicol (34 mg/L) at 37 °C. After overnight growth, 15 μL of culture was transferred into 2 mL of fresh LB medium supplemented with carbenicillin and chloramphenicol. After 3.5 hrs of growth, 2 mL of culture was transferred into a 125 mL flask containing 20 mL of M9 medium with 2% glucose, carbenicillin, and chloramphenicol. When the OD_{600} of the culture reached 0.9, 1 mM of IPTG was added to each flask. After 20 hrs of growth at 37 °C, 20 mL of ethyl acetate (with 1% of acetic acid, v/v) was added to each flask to extract the fatty alcohols produced during the fermentation. The crude ethyl acetate extract was directly analyzed with GC/MS as described herein.

[0414] The *yjgB* knockout strain resulted in significant accumulation of dodecanal and a lower fatty alcohol titer (Figure 5). The expression of *yjgB* from plasmid pETDuet-1-'tesA-yjgB in the *yjgB* knockout strain effectively removed the accumulation of dodecanal (Figure 5). The data shows that YjgB was involved in converting dodecanal into dodecanol and that there may be other alcohol dehydrogenase(s) present in *E. coli* to convert other aldehydes into alcohols. Dodecanal accumulated in the *yjgB* knockout strain, but it was not observed in either the wild-type strain (MG1655(DE3, Δ*fadD*)) or the *yjgB* knockout strain with the *yjgB* expression plasmid. The arrows (in Figure 5) indicate the GC trace of dodecanal (C12:0 aldehyde).

EXAMPLE 5

Production of Saturated Fatty Alcohols in *E. coli*

[0415] Fatty alcohols for commercial uses are saturated. However, *E. coli* typically has a certain amount (about 20-25%) of unsaturated fatty acids in its membrane to maintain fluidity. An *E. coli* strain was engineered that was able to produce exclusively saturated fatty acids in a medium not supplemented with unsaturated fatty acid or cyclopropane-fatty acid and was able to produce saturated fatty alcohols.

[0416] Two enzymes, a dehydratase/isomerase and a ketoacylsynthase I (KASI), encoded by *fabA* and *fabB*, respectively, are involved in unsaturated fatty acid biosynthesis. Usually, an *E. coli* strain lacking either FabA or FabB does not survive without supplementation of unsaturated fatty acids, such as oleate. To overcome this, the *fabB* gene was knocked out of an *E. coli* host strain, and the strain was able to grow without unsaturated fatty acid supplementation by genetically engineering the cells to express a recombinant desaturase gene (AF037430, encoding DesA) from *Bacillus subtilis*. Although the first generation of the strain expressing *desA* required oleate for normal growth, subsequent plating of the strain on L Agar plates several times resulted in a strain that did not require oleate for growth.

Materials

[0417] *E. coli* JWC280 cells (described in Campbell *et al.*, *Mol. Microbiol.* 47:793-805 (2003)) and *E. coli* GRT23 cells (described in Morgan-Kiss *et al.*, *Arch. Microbiol.* 190:427-437 (2008)) were obtained from John Cronan.

Plasmid construction

[0418] The *desA* gene (also referred to as *Δ5 des*) was amplified with primers delta5Fn and delta5Rn (listed in Table 11) from the genomic DNA of *Bacillus subtilis str. 168* and digested with *Avr*II and *Eco*RI. The *desA* gene was then cloned into pET-21(a), which had been linearized with *Avr*II-*Eco*RI, to produce pET-21a-Δ5. The *desA* gene was then removed as an *Nde*I-*Eco*RI fragment from pET-21a-Δ5 and inserted between the *Nde*I and *Eco*RI sites of OP180, a pACYC derived plasmid carrying a trc promoter. The resultant plasmid was named pACYC-Δ5.

[0419] A *desA_kan* gene cassette was cloned between the *Avr*II-*BamH*I sites of CDFDuet-1. A *kan* gene cassette was produced by *EcoR*I and *BamH*I digestion of a PCR product that was amplified with primers kanF and kanR (see Table 11) from pKD13 as the template (pKD13 was obtained from The Coli Genetic Stock Center, Yale University, and is described in Datsenko *et al.*, *Proc. Natl. Acad. Sci. USA* 97:6640-6645 (2000)). The amplified *desA* gene (described above) was digested with *Avr*II and *EcoR*I. The *Avr*II-*EcoR*I fragment of the *desA* gene and the *EcoR*I-*BamH*I fragment of the *kan* gene were then inserted between the *Avr*II-*BamH*I sites of pCDFDuet-1 (from EMD Chemicals, Gibbstown, NJ) to produce a plasmid that was named pCDFDuct-1-Δ5-kan.

[0420] A p84.17fabBΔ5kan plasmid was constructed to replace *fabB* with the desA_kan cassette by several subcloning steps. First, a DNA fragment (L-fabB) flanking the upstream region of *fabB* was amplified with primers fabBLF and fabBLR (see Table 11), and a DNA fragment (R-fabB) flanking the downstream region of *fabB* was amplified with primers fabBRF and fabBRR (see Table 11) from *E. coli* MG1655 genomic DNA. Second, L-fabB was digested with *Xba*I and *Bgl*II, and R-fabB was digested with *Not*I and *Bgl*II. The digested L-fabB and R-fabB fragments were purified from agarose gel and were ligated with *Xba*I-*Not*I linearized pKOV. The resultant plasmid was designated pHZ1.186. Next, the desA_kan gene cassette was removed from pCDFDuet-1-Δ5-kan as an *Avr*II-*BamH*I fragment and was inserted between the *Avr*II and *Bgl*II sites of pHZ1.186, resulting in the desA_kan gene cassette being sandwiched by L-fabB and R-fabB. Finally, the L-fabB-desA_kan-R-fabB fragment was amplified with fabBLF and fabBRR (see Table 11) from pHZ1.186 and cloned into the two *Pvu*II sites of pMOD-4-MCS (Epicentre Biotechnologies, Madison, WI). The final plasmid was designated p84.17fabB.

[0421] DNA spanning from about 1 kb upstream to about 1 kb downstream of fabB::cm was amplified from the genome of GRT23 cells using the primers fabBup and fabBdowm (see Table 11). The amplified DNA fragment was then digested with *Pvu*II and inserted between the two *Pvu*II sites of pMOD -4-MCS. The resulting plasmid was designated p84.15.

[0422] The genes encoding a thioesterase ('TesA) and a fatty acid reductase (CarB) were cloned as an operon, and the operon was placed under the trc promoter and pCL1920 vector. The final plasmid was named pCL-Ptrc-carB_'tesA (the sequence is listed in Figure 17 as SEQ ID NO:213).

Table 11: Primer sequences

Primer ID	Sequence
delta5Fn	TTTT CCTAGG ATG ACT GAA CAA ACC A (SEQ ID NO:214)
delta5Rn	TTTT GAATTC TTA TCA TTG TGA AAG CCAGAA (SEQ ID NO:215)
kanF	TTTT GAATTC TGT AGG CTG GAG CTG CTTCG (SEQ ID NO:216)
kanR	ATTCCG GGG ATC CGT CGA CC (SEQ ID NO:217)
fabBLF	TTTT CTA GAA ATA GCG CCA GCG ACA (SEQ ID NO:218)
fabBLR	TTTT AGA TCT TAG CCC TAG GCC AGT AAT CAC TGC ACG (SEQ ID NO:219)
fabBRF	TTTT AGA TCT AGC TTC GGC TTC GGC G (SEQ ID NO:220)
fabBRR	TTTT GCG GCC GCG CCC ATC CTT TGC TGG C (SEQ ID NO:221)
fabBup	ACG ACA AAT GCG CCG C (SEQ ID NO:222)
fabBdown	ATC CGC GCA ATA AAG C (SEQ ID NO:223)

Strain construction

[0423] An *E. coli* MG1655 (ΔfadEΔfhuAfabB::cm)/pACYC-Δ5 strain was constructed by transforming p84.15fabB into MG1655 (ΔfadEΔfhuA)/pACYC-Δ5. Plasmid p84.17fabB was transformed into MG1655 (ΔfadEΔfhuAfabB::cm)/pACYC-Δ5 to produce MG1655 (ΔfadEΔfhuAfabB::desA_kan)/pACYC-Δ5. After each transformation, the

transformant mix was plated onto L agar plates supplemented with 1 mM IPTG and appropriate antibiotics (17 mg/L of chloramphenicol or 50 mg/L of kanamycin).

[0424] MG1655 (ΔfadEΔfhuAfabB::desA_kan)/pACYC-Δ5 grew normally in L Broth supplemented with oleate (potassium salt, 50mg/L). Cells were plated onto L agar plates supplemented with 50 mg/L of oleate and incubated at 37 °C for 2 days. Colonies were then patched onto L Agar plates, supplemented with 50 mg/L of oleate and 100 mg/L of carbenicillin. One of the colonies, which lost resistance to carbenicillin but retained kanamycin resistance, was streaked onto an L agar plate supplemented with 50 mg/L of kanamycin, but no oleate. One of the colonies was selected from the plate and was designated ALC119A.

ALC119A with a fatty alcohol pathway produced almost exclusive saturated fatty alcohol

[0425] Plasmid pCL-Ptrc-carB_'tesA was transformed into the ACL119A strain. Three transformants of ALC119A/pCL-Ptrc-carB_'tesA were grown in 3 mL of L broth with 100 mg/L of spectinomycin in a 37 °C shaker overnight. 15 μL of the overnight culture were transferred into 2 mL of fresh L broth with 100 mg/L of spectinomycin and 2 μL of 70% potassium oleate. The fresh inoculation was placed in a 37 °C shaker for about 3 hrs. The 2 mL culture was then transferred into 20 mL of V9 medium (Hu-9 medium without ferric chloride) in a 125 mL baffle flask. When the OD_{600} of the culture reached about 0.9, 1 mM of IPTG was added to each flask. After 20 hrs of growth at 37 °C, 20 mL of ethyl acetate (with 1% of acetic acid, v/v) was added to each flask to extract the fatty alcohols produced during the fermentation. The crude ethyl acetate extract was directly analyzed with GC/MS as described in WO 2008/119082. Cetyl alcohol was used as a reference for quantification of fatty alcohol.

[0426] As shown in Figure 13, the ALC119A/pCL-Ptrc-carB_'tesA strain produced almost exclusively saturated fatty alcohols, including dodecanol, tetradecanol and hexadecanol.

EXAMPLE 6

Production of Fatty Alcohols in the Cyanobacterium *Synechococcus sp.* PCC7002

[0427] This example describes the use of photoautotrophic bacteria to produce fatty alcohols from carbon dioxide (instead of glucose) using the *carB-'tesA-yahK* pathway. First, a vector is constructed for homologous recombination into the *Synechococcus* sp. PCC7002 plasmid pAQ1 (genbank accession NC_0050525) using 500 bp homologous regions corresponding to positions 3301-3800 and 3801-4300 of pAQ1. As a selectable marker, a spectinomycin resistance cassette containing the aminoglycoside 3' adenylyltransferase, *aad*, promoter, gene and terminator (from plasmid pCL1920), is added between the homologous regions. For gene expression, the promoter and ribosome binding site of aminoglycoside phosphotransferase, *aph* (from plasmid pACYC177), is added followed by the unique cloning sites *Nde*I and *EcoR*I for insertion of a heterologous gene or operon. This complete integration cassette is constructed by gene synthesis and cloned into pUC19 for maintenance and delivery. The resulting plasmid, pLS9-7002, allows (i) cloning and expression of a foreign gene, and (ii) delivery and stable *in vivo* integration into *Synechococcus* sp. PCC7002 plasmid pAQ1.

[0428] The fatty alcohol pathway for expression in *Synechococcus* sp. PCC7002 is constructed as follows. The *carB-'tesA* operon from pCL-Ptrc-carB-'tesA (described in Example 4) is extended by adding *yahK* downstream of *'tesA* and then cloning into the *Nde*I and *EcoR*I sites of pLS9-7002 downstream of the *aph* promoter and ribosome binding site. The resulting plasmid is transformed into *Synechococcus* sp. PCC7002 as described by Stevens *et al.* (*Proc. Natl. Acad. Sci. U.S.A.* 77:6052-6056 (1980)). Stable integrants are selected for on ATCC 1047 medium supplemented with 15 μg/mL spectinomycin. 1 L of ATCC 1047 medium contains 40 mg $MgSO_4$ x 7 H_2O, 20 mg $CaCl_2$ x 2 H_2O, 750 mg $NaNO_3$, 2 mg K_2HPO_4, 3.0 mg citric acid, 3.0 mg ferric ammonium citrate, 0.5 mg EDTA, 20 mg Na_2CO_3, 2.86 mg H_3BO_3, 1.81 mg $MnCl_2$, 0.22 mg $ZnSO_4$, 0.04 mg Na_2MoO_4, 0.08 mg $CuSO_4$, 0.05 mg $Co(NO_3)_2$, 0.02 mg vitamin B12, 10 g agar, and 750 mL sea water. Spectinomycin resistant colonies are restreaked several times on ATCC medium 1047 with spectinomycin and tested for isogenic intergration of the *carB-'tesA-yahK* operon by PCR with primers pAQ1-U (atgtctgacaaggggtttgacccct) (SEQ ID NO:224)

and pAQ1-D (gcacatccttatccaattgctctag) (SEQ ID NO:225). Complete isogenic *carB-'tesA-yahK* integrants are then grown in 50 mL liquid ATCC 1047 medium with spectinomycin in 500 mL shake flasks with appropriate aeration and illumination at 30 °C for five to seven days. Culture aliquots are extracted at various time points with an equal volume of ethyl acetate and the extracts are analyzed for fatty alcohol production as described in Example 3. Fatty alcohols are produced.

EXAMPLE 7

Production of Fatty Alcohols in the Cyanobacterium *Synechococcus elongatus* PCC7942

[0429] This example describes a second method of using photoautotrophic bacteria to produce fatty alcohols from carbon dioxide (instead of glucose) using the *carB-'tesA-yahK* pathway. First, a vector is constructed for homologous recombination into the *Synechococcus* elongatus PCC7942 genome (genbank accession CP_000100) using 800 bp homologous regions corresponding to positions 2577844-2578659 and 2578660-2579467 of CP_000100. This chromosomal location is known as neutral site one (NS1) (Mackey *et al.*, *Meth. Mol. Biol.* 362:115-129 (2007)). As a selectable marker, a spectinomycin resistance cassette containing the aminoglycoside 3' adenylyltransferase, *aad*, promoter, gene and terminator (from plasmid pCL1920), is added between the homologous regions. Aditionally, the unique cloning sites *Nde*I and *EcoR*I are added for insertion of a heterologous gene or operon. This integration cassette is constructed by gene synthesis and cloned into pUC19 for maintenance and delivery. The resulting plasmid, pLS9-7942_NS1, allows (i) cloning and expression of a foreign gene and (ii) delivery and stable *in vivo* integration into the *Synechococcus elongatus* PCC7942 genome.

[0430] The complete *carB-'tesA-yahK* operon (described in Example 6), including its *p*trc promoter and ribosome binding site, is cloned into the *Nde*I or *EcoR*I site of pLS9-7942_NS1. The resulting plasmid is transformed into *S. elongatus* PCC7942 as described by Mackey *et al.*, *Meth. Mol. Biol.* 362:115-129 (2007). Stable integrants are selected for on BG-11 medium supplemented with 4 μg/mL spectinomycin. 1 L of BG-11 medium contains 75 mg $MgSO_4$ x 7 H_2O, 36 mg $CaCl_2$ x 2 H_2O, 1.5 g $NaNO_3$, 40 mg K_2HPO_4, 6.0 mg citric acid, 6.0 mg ferric ammonium

citrate, 1.0 mg EDTA, 20 mg Na_2CO_3, 2.86 mg H_3BO_3, 1.81 mg $MnCl_2$, 0.22 mg $ZnSO_4$, 0.04 mg Na_2MoO_4, 0.08 mg $CuSO_4$, 0.05 mg $Co(NO_3)_2$, and 10 g agar. Spectinomycin resistant colonies are restreaked several times on BG-11 medium with spectinomycin and tested for isogenic integration of the *carB-'tesA-yahK* operon by PCR with primers NS1-U (gatcaaacaggtgcagcagcaactt) (SEQ ID NO:226) and NS1-D (attcttgacaagcgatcgcggtcac) (SEQ ID NO:227). Complete isogenic *carB-'tesA-yahK* integrants are then grown in 50 mL liquid BG-11 medium with spectinomycin in 500 mL shake flasks with appropriate aeration and illumination at 30 °C up to seven days. Culture aliquots are extracted at various time points with an equal volume of ethyl acetate and the extracts are analyzed for fatty alcohol production as described in Example 3. Fatty alcohols are produced.

EXAMPLE 8

Malonyl-CoA-Independent Production of Fatty Alcohols in *E. coli*

[0431] Certain protists such as *Euglena gracilis* are capable of malonyl-CoA independent fatty acid biosynthesis. The biosynthetic machinery for this pathway is located in the mitochondria and is thought to reverse the direction of β-oxidation by using acetyl-CoA as priming as well as elongating substrates to produce C_8 to C_{18} fatty acids (Inui *et al.*, *Eur. J. Biochem.* 142:121-126 (1984)). The enzymes involved are trans-2-enoyl-CoA reductases (TER), which catalyze the irreversible reduction of trans-2-enoyl-CoA to acyl-CoA and thereby drive the otherwise reversible pathway in the reductive direction (while the opposite is true for β-oxidation, where the irreversible acyl-CoA dehydrogenase, FadE, drives the reaction in the oxidative direction). One TER gene from *E. gracilis* as well as other eukaryotic and prokaryotic homologs are known (Hoffmeister *et al.*, *J. Biol. Chem.* 280:4329-4338 (2005); Tucci *et al.*, *FEBS Lett.* 581:1561-1566 (2007)). The only known TER enzyme from *E. gracilis* has been shown *in vitro* to reduce trans-2-butenoyl-CoA (C4) and trans-2-hexenoyl-CoA (C6) to the respective acyl-CoAs, (longer-chain trans-2-enoyl-CoAs have not been tested). Currently, very little is known about the other pathway enzymes in *E. gracilis*.

[0432] A pathway that creates a flux exclusively from acetyl-CoA precursors to acyl-CoA (as in *Euglena gracilis* mitochondria) can be engineered in *E. coli* using

different sets of enzymes with the following four enzymatic activities: (i) non-decarboxylating, condensing thiolase, (ii) 3-ketoacyl-CoA reductase (or 3-hydroxyacyl-CoA dehydrogenase), (iii) 3-hydroxyacyl-CoA hydratase (or enoyl-CoA dehydratase) and (iv) *trans*-2-enoyl-CoA reductase. All four enzymes can have sufficiently relaxed chain lengths specificity to allow synthesis of acyl-CoAs with longer chain length, *e.g.*, C_{12} or C_{14}.

[0433] A plasmid encoding all four activities is constructed as follows. A synthetic operon of *E. coli fadA* (YP_026272) (shown in Figure 17 as SEQ ID NO:229) (non-decarboxylating thiolase) and *fadB* (NP_418288) (shown in Figure 17 as SEQ ID NO:231) (3-hydroxyacyl-CoA dehydrogenase and enoyl-CoA dehydratase) and *E. gracilis ter* (Q5EU90) (shown in Figure 17 as SEQ ID NO:228) (trans-2-enoyl-CoA reductase, codon optimized without its 5' sequence encoding a transit peptide) is constructed and cloned downstream of a *p*trc promoter into a pACYC plasmid with a carbenicillin or chloramphenicol resistence gene. Alternatively, instead of the *E. coli fadA* and *fadB* genes, the E. coli *fadI* (NP_416844) (shown in Figure 17 as SEQ ID NO:223) and *fadJ* (NP_416843) (shown in Figure 17 as SEQ ID NO:235) genes (or the corresponding orthologs from other organisms) are used. As an alternative to the *E. gracilis ter* gene, the corresponding orthologs from other organisms or the *E. coli fabI* (NP_415804) (shown in Figure 17 as SEQ ID NO:237) gene are used. Although FabI normally reduces trans-2-enoyl-ACPs, it is also active with trans-2-enoyl-CoAs (Bergler *et al.*, *J. Biol. Chem.* 269:5493-5496 (1993)).

[0434] The pACYC-*p*trc_*fadAB-ter* plasmid or the pACYC-*p*trc_*fadAB-fabI* plasmid is cotransformed with the pCL-*p*trc_*carB-'tesA* plasmid (described in Example 4) into an *E. coli* Δ*fadE* strain. These strains are cultured, extracted and analyzed for fatty alcohol production as described in Example 3. The two different strains produce fatty alcohols with different chain length distribution.

[0435] As these strains express 'TesA, a portion of the fatty alcohols produced are derived from malonyl-CoA dependent acyl-ACP precursors. 'TesA efficiently hydrolyzes acyl-ACPs when overexpressed in *E. coli*, although it has higher specific activity for acyl-CoAs as compared to acyl-ACPs. To increase the proportion or exclusively produce fatty alcohols derived from the malonyl-CoA independent

pathway, alternative thioesterases that have lower hydrolytic activity towards acyl-ACPs are used instead of 'TesA. One example is *E. coli* TesB (NP_414986), which prefers acyl-CoAs over acyl-ACPs (Spencer *et al.*, *J. Biol. Chem.* 253:5922-5926 (1978)) and when overexpressed in *E. coli* does not hydrolyze acyl ACPs (Zheng *et al.*, *App. Environ. Microbiol.* 70:3807-3813 (2004)). In alternative methods, orthologs of TesA and TesB or thioesterases from other protein families that hydrolyze acyl-CoAs with high efficiency while hydrolyzing acyl-ACPs with low efficiency are used.

[0436] In one method, a pCL-*p*trc_*carB-'tesB* plasmid is constructed as described in Example 4 by replacing the *'tesA* gene with the *tesB* gene (NP_414986) (shown in Figure 17 as SEQ ID NO:239). The plasmid is cotransformed with the pACYC-*p*trc_*fadAB-ter* plasmid or the pACYC-*p*trc_*fadAB-fabI* plasmid into an *E. coli* Δ*fadE* strain. These strains are cultured, extracted and analyzed for fatty alcohol production as described in Example 3.

[0437] In another method, the pCL-ptrc_*carB-'tesA* plasmid is replaced with a pCL-ptrc_*acr1* plasmid, which expresses the acyl-CoA reductase Acr1 from *Acinetobacter baylyi* ADP1 (YP_047869) (shown in Figure 17 as SEQ ID NO:241). This reductase specifically reduces acyl-CoAs but not acyl-ACPs to the corresponding fatty alcohols (Reiser *et al.*, *J. Bacteriol.* 179:2969-2975 (1997)). The plasmid is cotransformed with the pACYC-*p*trc_*fadAB-ter* plasmid or the pACYC-*p*trc_*fadAB-fabI* plasmid into an *E. coli* Δ*fadE* strain. These strains are cultured, extracted and analyzed for fatty alcohol production as described in Example 3. The strains produce fatty alcohols independent of of malonyl-CoA.

EXAMPLE 9

Identification of Iron as an Inhibitor of Fatty Alcohol Production

[0438] Hu9 medium is a known fermentation medium, which contains 6 g/L Na_2HPO_4, 3 g/L KH_2PO_4, 0.5 g/L NaCl, 1 g/L NH_4Cl, 0.1 mM $CaCl_2$, 1 mM $MgSO_4$, 15 g/L agar, 10 mM glucose, 50 mg/L Uracil, and trace minerals containing 100 μM $FeCl_3$, 500 μM $ZnCl_2$, 200 μM Na_2Mo_4, 200 μM $CuSO_4$, 200 μM H_3BO_3. However, it was observed that the production of fatty alcohols was completely reduced when recombinant *E. coli* strains, otherwise capable of producing fatty alcohols, were

grown in Hu9 medium. As described in detail below, the inability of *E. coli* strains to produce fatty alcohols in various incomplete Hu9 media was measured, and it was found that the recombinant bacteria were incapable of producing fatty alcohols when iron was present in the medium. However, the addition of iron did not inhibit the growth of the bacteria.

[0439] In order to identify the component(s) involved in the inhibition of fatty alcohol production, different versions of incomplete Hu9 medium were made, some of which lacked a dispensable ingredient, and then the production of fatty alcohol was evaluated.

[0440] In the first step the following media were made: complete Hu9 medium, incomplete Hu9 medium lacking uracil, and incomplete Hu9 medium lacking trace elements. K6 cells (a recombinant bacterial strain C41 (DE3, Δ*fadE*) carrying pACYCDuet-1-carB, encoding the CAR homolog carB and pETDuet-1-'tesA) were cultured in 2 mL of LB containing appropriate antibiotics. After reaching an OD of 1.0, the 2 mL cultures were scaled up in 125 mL shake flasks (containing one of the Hu9 media described above) to a volume 22 mL. The cultures were induced by adding IPTG to a final concentration of 1 mM. After growing them for 20 hrs at 37 °C, 22 mL of ethyl acetate (with 1% of acetic acid, v/v) was added to each flask to extract the fatty alcohols produced during the fermentation. The crude ethyl acetate extract was directly analyzed with GC/MS and the total fatty alcohol titers were quantified.

[0441] As depicted in Figure 14A, the fatty alcohol production was inhibited to a great extent by the addition of trace elements as compared to the addition of uracil to the incomplete Hu9 medium. This indicated that the inhibitory component(s) was a part of trace mineral solution.

[0442] In order to find out which trace element was responsible for the fatty alcohol production inhibition, the following Hu9 media were made: complete Hu9 medium; Hu9 lacking $FeCl_3$; Hu9 lacking $ZnCl_2$; Hu9 lacking Na_2Mo_4; Hu9 lacking $CuSO_4$; and Hu9 lacking H_3BO_3. The fatty alcohol production of K6 cells grown in these different Hu9 media was evaluated using the method described above.

[0443] As shown in Figure 14B, fatty alcohol production was inhibited mainly by the addition of iron to the medium. Thus, by eliminating or reducing the presence of

iron (*e.g.*, ferric citrate, ferric chloride, or ferrous sulfate) in the culture medium, fatty alcohols can be produced.

DEMANDES OU BREVETS VOLUMINEUX

LA PRÉSENTE PARTIE DE CETTE DEMANDE OU CE BREVETS COMPREND PLUS D'UN TOME.

CECI EST LE TOME __1___ DE __3___

NOTE: Pour les tomes additionels, veillez contacter le Bureau Canadien des Brevets.

JUMBO APPLICATIONS / PATENTS

THIS SECTION OF THE APPLICATION / PATENT CONTAINS MORE THAN ONE VOLUME.

THIS IS VOLUME __1___ OF __3___

NOTE: For additional volumes please contact the Canadian Patent Office.

CLAIMS:

1. A method of producing a fatty alcohol in a host cell, the method comprising:

(a) expressing in a host cell a nucleic acid molecule encoding a polypeptide having thioesterase activity; and

(b) further expressing in the host cell a nucleic acid molecule encoding a polypeptide having alcohol dehydrogenase activity; and

(c) further expressing in the host cell an exogenous nucleic acid molecule encoding a polypeptide having carboxylic acid reductase activity and comprising a sequence having at least 80% sequence identity to the full length amino acid sequence of SEQ ID NO:22; and

(d) culturing the host cell in culture media containing a carbon source under conditions wherein a fatty alcohol is produced.

2. The method of claim 1, where the thioesterase activity releases a fatty acid from an acyl-ACP and the carboxylic acid reductase converts the fatty acid to an aldehyde.

3. The method of claim 1 or 2, wherein the polypeptide having alcohol dehydrogenase activity converts a fatty aldehyde to a fatty alcohol.

4. The method of any one of claims 1 to 3, wherein the host cell is genetically engineered to express a reduced level of at least one fatty acid degradation enzyme relative to a wild type host cell.

5. The method of claim 4, wherein the at least one fatty acid degradation enzyme has acyl-CoA synthetase activity.

6. The method of any one of claims 1 to 5, wherein the host cell is selected from the group consisting of a plant cell, insect cell, fungus cell, and bacterial cell.

7. The method of any one of claims 1 to 6, wherein the host cell is a yeast cell or filamentous fungal cell.

8. The method of any one of claims 1 to 7, wherein the fatty alcohol is isolated from the host cell culture.

9. The method of any one of claims 1 to 8, wherein the fatty alcohol comprises a C6-C26 fatty alcohol.

10. The method of any one of claims 1 to 8, wherein the fatty alcohol is a C6, C8, C10, C12, C13, C14, C15, C16, C17, or C18 fatty alcohol.

11. The method of any one of claims 1 to 10, wherein the fatty alcohol is an unsaturated fatty alcohol.

12. The method of claim 11, wherein the unsaturated fatty alcohol is C10:1, C12:1, C14:1, C16:1, or C18:1.

13. The method of claim 12, wherein the fatty alcohol is unsaturated at the omega-7 position.

14. The method of claim 11, wherein the unsaturated fatty alcohol comprises a cis double bond.

15. The method of any one of claims 1-10, wherein the fatty alcohol is a saturated fatty alcohol.

16. The method of any one of claims 1-15, further comprising culturing the host cell in the presence of at least one biological substrate for the polypeptide set forth in SEQ ID NO:22.

17. The method of claim 16, wherein the at least one substrate is a fatty acid.

18. The method of any one of claims 1 to 17, wherein the fatty alcohol is produced by way of a malonyl-CoA-independent pathway.

19. The method of any one of claims 1 to 18, where the carbon source is a carbohydrate.

20. The method of claim 19, wherein the carbohydrate comprises a monosaccharide, disaccharide or polysaccharide.

21. A method of producing a fatty alcohol in a host cell, the method comprising: culturing the host cell in a culture media containing a carbon source, wherein the host cell comprises polynucleotide sequences encoding:

(a) a first polypeptide having thioesterase activity effective to release a fatty acid from acyl-ACP, and

(b) a second polypeptide having carboxylic acid reductase activity and comprising an amino acid sequence

(i) having at least 95% sequence identity to the full length amino acid sequence of SEQ ID NO:22, and

(ii) having activity effective to convert a fatty acid to a fatty aldehyde, wherein the second polypeptide is expressed from an exogenous nucleic acid molecule, and

(c) a third polypeptide, the third polypeptide having alcohol dehydrogenase activity effective to convert a fatty aldehyde to a fatty alcohol.

22. The method of claim 21, wherein one or more of the polynucleotide sequences are plasmid sequences.

23. The method of claim 21, wherein one or more of the polynucleotide sequences are integrated into the genome of the host cell.

24. The method of any one of claims 21 to 23, wherein the second polypeptide comprises the amino acid sequence of SEQ ID NO:22.

25. The method of any one of claims 21 to 24, wherein the first polypeptide is a TesA polypeptide.

26. The method of any one of claims 21 to 25, wherein the fatty alcohol comprises C12 and C14 fatty alcohols.

27. The method of claim 26, wherein the C12 and C14 fatty alcohols are primary alcohols.

28. The method of claim 26 or 27, wherein the fatty alcohol has a C12 to C14 ratio of approximately 3 to 1.

29. The method of any one of claims 21 to 28, wherein the fatty alcohol comprises an unsaturated fatty alcohol.

30. The method of claim 29, wherein the unsaturated fatty alcohol comprises a cis double bond.

31. The method of claim 29, wherein the fatty alcohol is unsaturated at the omega-7 position.

32. The method of any one of claims 21 to 31, wherein the fatty alcohol further comprises a C6, C8, C10, C11, C13, C15, C16, or C18 fatty alcohol.

33. The method of any one of claims 21 to 32, wherein the fatty alcohol is produced at a level of at least 700 mg/L.

34. The method of any one of claims 21 to 33, where the carbon source is a carbohydrate.

35. The method of claim 34, wherein the carbohydrate comprises a monosaccharide, disaccharide or polysaccharide.

36. The method of claim 34, where the carbon source is a glycerol-containing carbon source.

37. The method of any one of claims 21 to 36, wherein acyl-CoA dehydrogenase expression or activity is reduced in the host cell as compared to a wild-type host cell.

38. The method of any one of claims 21 to 36, wherein acyl-CoA dehydrogenase expression or activity is deleted or reduced in the host cell as compared to a wild-type host cell.

39. The method of any one of claims 21 to 36, wherein 3-ketoacyl-CoA thiolase expression or activity is deleted or reduced in the host cell as compared to a wild-type host cell.

40. The method of any one of claims 21 to 36, wherein fadB expression or activity is deleted or reduced in the host cell as compared to a wild-type host cell.

41. The method of any one of claims 21 to 36, wherein beta-ketoacyl-CoA thiolase expression or activity is deleted or reduced in the host cell as compared to a wild-type host cell.

42. The method of any one of claims 21 to 36, wherein fadJ expression or activity is deleted or reduced in the host cell as compared to a wild-type host cell.

43. A host cell genetically engineered to express

(a) an exogenous polypeptide having the amino acid sequence of SEQ ID NO:22,

(b) a polypeptide having thioesterase activity, and

(c) a polypeptide having alcohol dehydrogenase activity.

44. The cell of claim 43, wherein the genetically engineered host cell produces a fatty alcohol when cultured in the presence of a carbon source.

45. A method of making a fatty alcohol, comprising:

(a) culturing a host cell genetically engineered to express an exogenous polypeptide having the amino acid sequence of SEQ ID NO:22 in a culture medium comprising glycerol; and

(b) fermenting the genetically engineered host cell of part (a) in a fermentation media, wherein a fatty alcohol comprising C12 and C14 fatty alcohols is produced by the genetically engineered host cell,

wherein the genetically engineered host cell is further engineered to express an exogenous polynucleotide encoding a polypeptide having thioesterase activity and an exogenous polynucleotide encoding a polypeptide having alcohol dehydrogenase activity.

46. The method of claim 45, wherein the fatty alcohol is separated from the fermentation media.

47. The method of claim 45 or 46, wherein the thioesterase has activity effective to release a fatty acid from acyl-ACP.

48. The method of claim 47, wherein the culture medium further comprises a fatty acid.

49. The method of claim 47, wherein, acyl-CoA dehydrogenase expression or activity is reduced in the genetically engineered host cell as compared to a wild-type host cell.

50. The method of claim 47, wherein acyl-CoA dehydrogenase expression or activity is deleted or reduced in the genetically engineered host cell as compared to a wild-type host cell.

51. The method of claim 47, wherein 3-ketoacyl-CoA thiolase expression or activity is deleted or reduced in the genetically engineered host cell as compared to a wild-type host cell.

52. The method of claim 47, wherein fadB expression or activity is deleted or reduced in the genetically engineered host cell as compared to a wild-type host cell.

53. The method of claim 47, wherein beta-ketoacyl-CoA thiolase expression or activity is deleted or reduced in the genetically engineered host cell as compared to a wild-type host cell.

54. The method of claim 47, wherein fadJ expression or activity is deleted or reduced in the genetically engineered host cell as compared to a wild-type host cell.

55. The method of any one of claims 1 to 30, wherein the alcohol dehydrogenase comprises an amino acid sequence having at least 80% sequence identity to the full-length amino acid sequence of SEQ ID NO:95.

56. Use of the fatty alcohol produced by the method of any one of claims 1 to 42 or 45 to 55, for preparation of a surfactant.

57. A microbial cell for the production of fatty alcohols, said cell genetically engineered to express a polypeptide having carboxylic acid reductase activity; a polypeptide having thioesterase activity; and a polypeptide having alcohol dehydrogenase activity, wherein said polypeptide having carboxylic acid reductase activity comprises at least 80% sequence identity to the full length amino acid sequence of SEQ ID NO:22.

58. The microbial cell of claim 57, wherein said polypeptide having carboxylic acid reductase activity is CarB.

59. The microbial cell of claim 58, wherein said CarB comprises the amino acid sequence of SEQ ID NO: 22.

60. The microbial cell of any one of claims 58 to 59, wherein said thioesterase is encoded by a nucleic acid molecule selected from the group consisting of *tesA*, *tesA* without leader sequence, *tesB*, *fatB*, *fatB2*, *fatB3*, *fatA*, and *fatA1*.

61. The microbial cell of claim 60, wherein said thioesterase nucleic acid molecule is encoded by tesA without leader sequence.

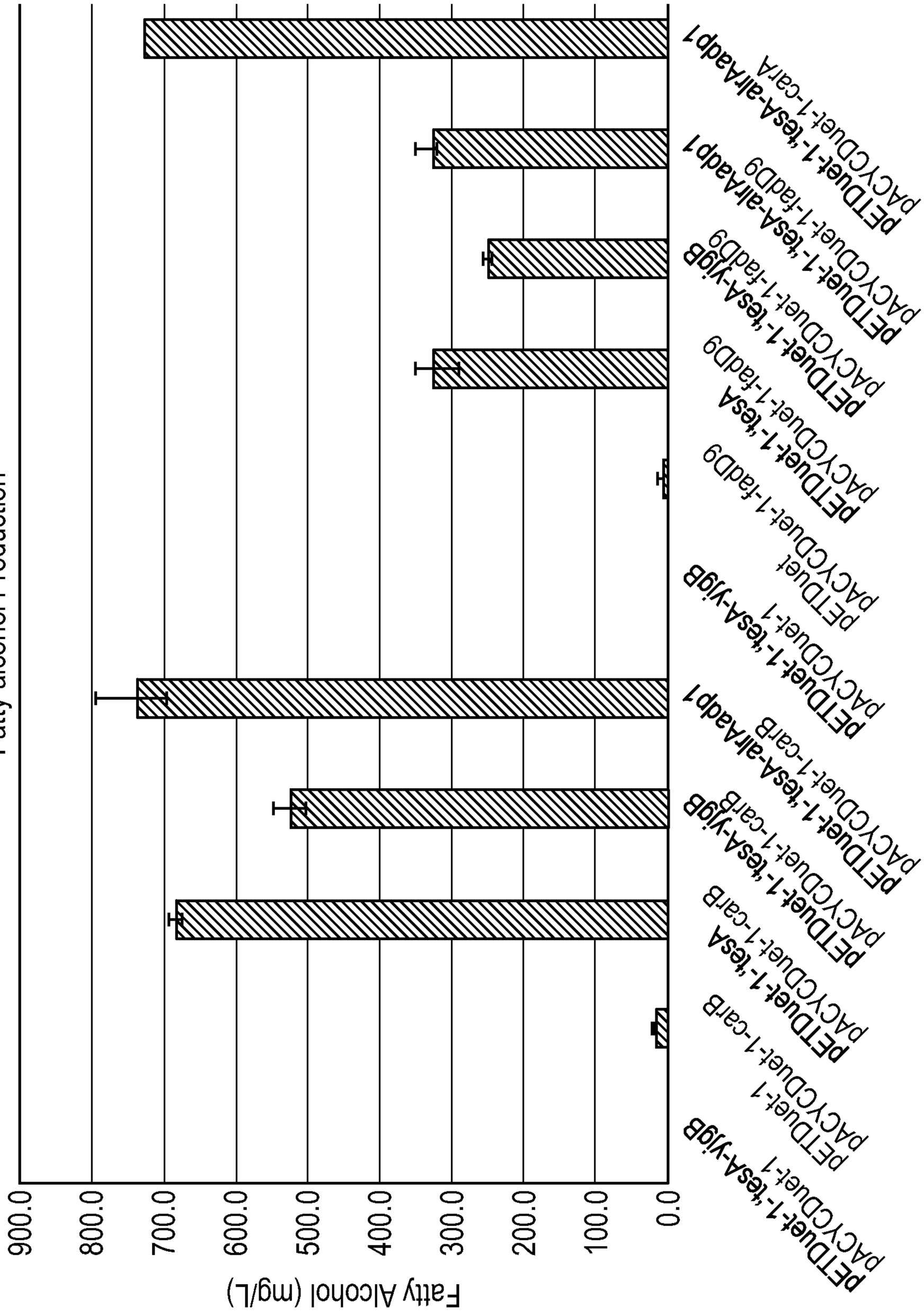
62. The microbial cell of any one of claims 57 to 61, wherein said alcohol dehydrogenase is encoded by a nucleic acid molecule selected from the group consisting of *alrAadp1*, *yahK*, *yjgB*, *adhP*, *dkgA*, *dkgB*, *yhdH*, *ydjL*, and *yqhD*.

63. The microbial cell of claim 62, wherein said *alrAadp1* encodes AlrA from *Acinetobacter baylyi* ADP1.

64. The microbial cell of claim 62, wherein said *yjgB* encodes YjgB from *Escherichia coli.*

65. The microbial cell of any one of claims 58 to 64, wherein said microbial cell is *Escherichia coli.*

FIG. 1


Fatty alcohol Production
Fatty Alcohol (mg/L)
900.0
800.0
700.0
600.0
500.0
400.0
300.0
200.0
100.0
0.0
pETDuet-1-tesA-altAadp7 pACYCDuet-1-carA
pETDuet-1-tesA-altAadp7 pACYCDuet-1-fadD9
pETDuet-1-tesA-yjgB pACYCDuet-1-fadD9
pETDuet-1-tesA pACYCDuet-1-fadD9
pETDuet-1-fadD9 pACYCDuet
pETDuet-1-tesA-yjgB pACYCDuet-1
pETDuet-1-tesA-altAadp7 pACYCDuet-1-carB
pETDuet-1-tesA-yjgB pACYCDuet-1-carB
pETDuet-1-tesA pACYCDuet-1-carB
pETDuet-1-carB pETDuet-1 pACYCDuet-1
pETDuet-1-tesA-yjgB pETDuet-1 pACYCDuet-1

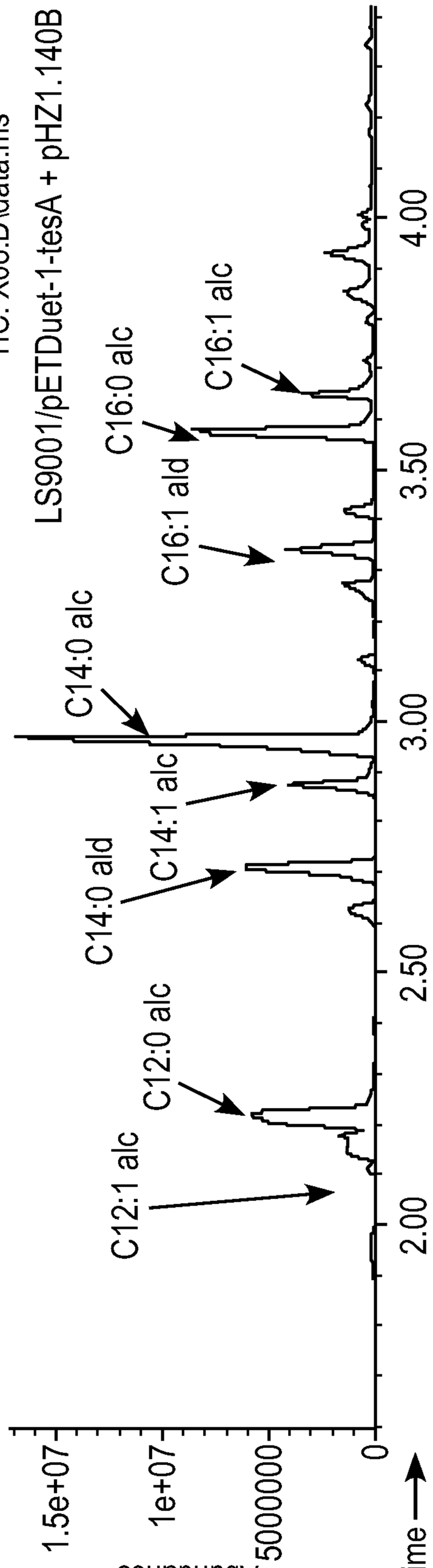

TIC: X08.D\data.ms
LS9001/pETDuet-1-tesA + pHZ1.140B
C14:0 alc
C16:0 alc
C16:1 ald
C16:1 alc
C14:0 ald
C14:1 alc
C12:0 alc
C12:1 alc
4.00
3.50
3.00
2.50
2.00
1.5e+07
1e+07
5000000
0
Abundance
Time

FIG. 2A

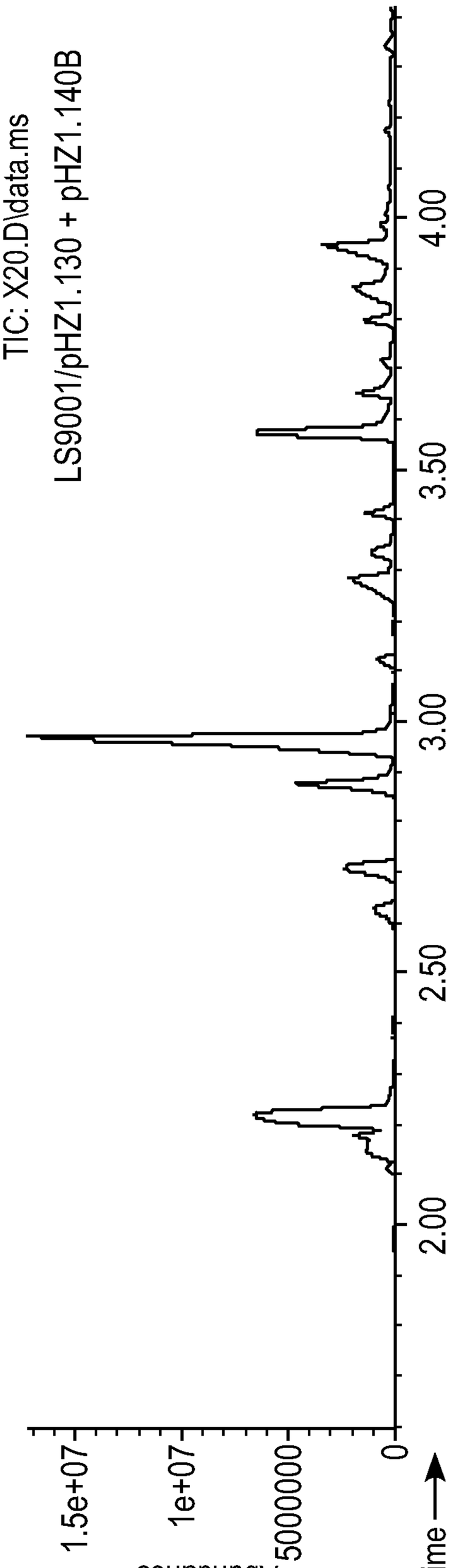

TIC: X20.D\data.ms
LS9001/pHZ1.130 + pHZ1.140B
4.00
3.50
3.00
2.50
2.00
1.5e+07
1e+07
5000000
0
Abundance
Time

FIG. 2B

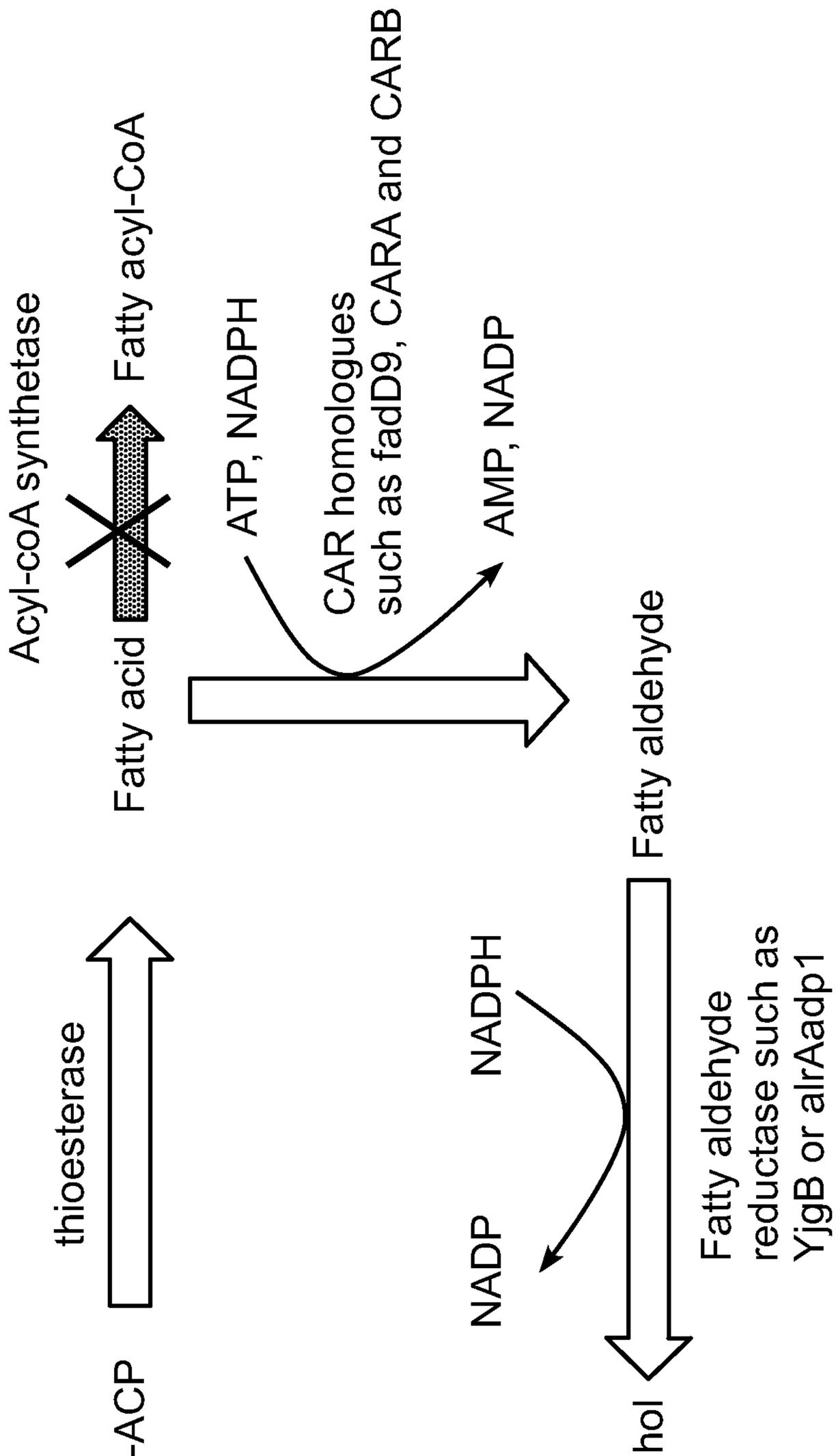

Fatty acyl-ACP
thioesterase
Fatty acid
Acyl-coA synthetase
Fatty acyl-CoA
ATP, NADPH
CAR homologues
such as fadD9, CARA and CARB
AMP, NADP
aldehyde
NADP
NADPH
Fatty aldehyde
reductase such as
YjgB or alrAadp1
Fatty alcohol

FIG. 3

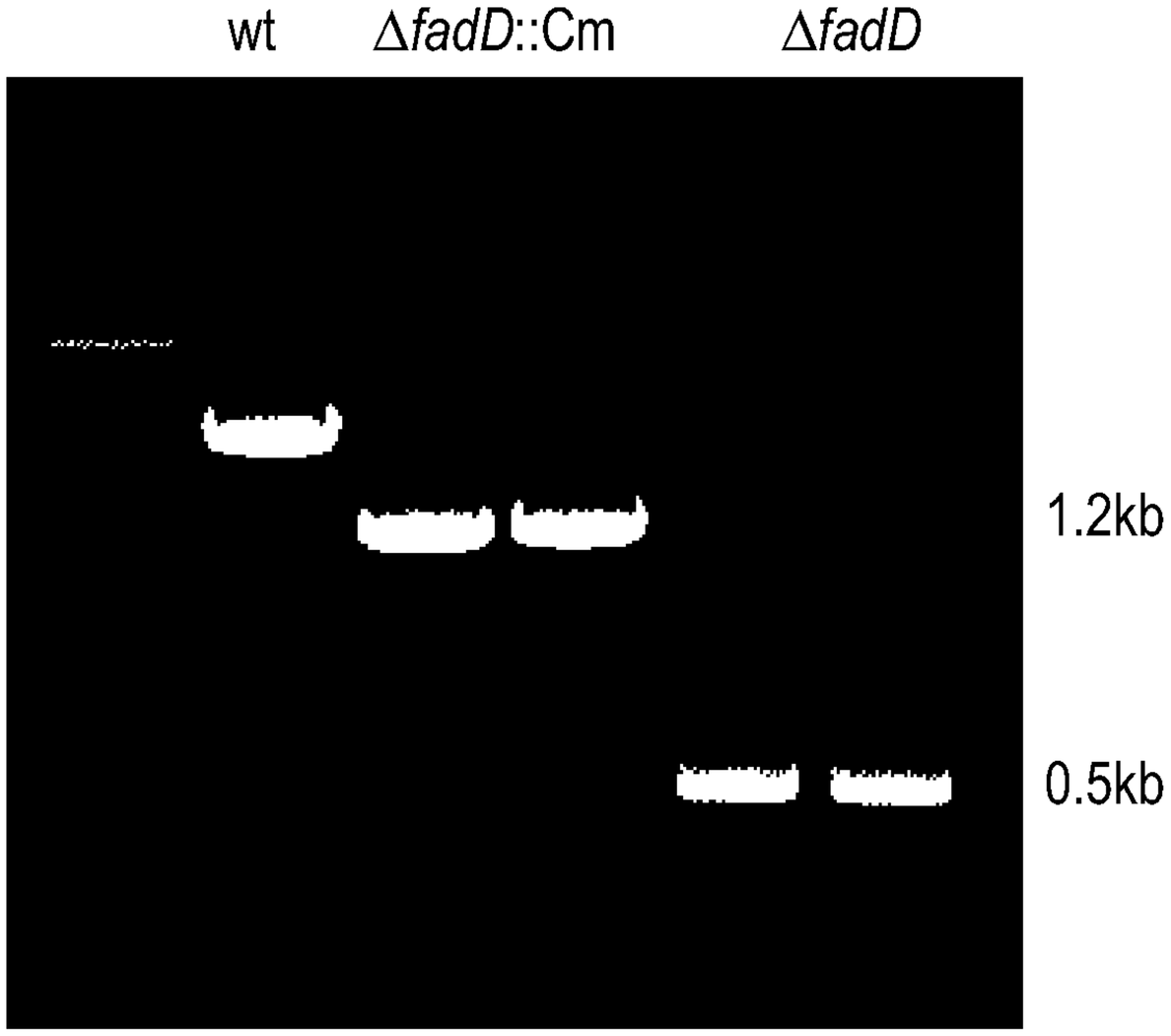


FIG. 4

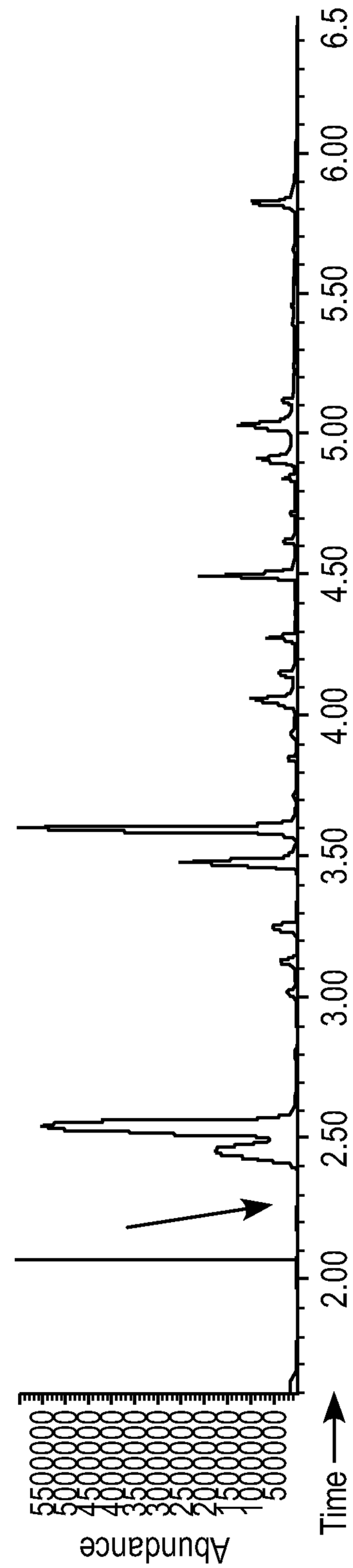


FIG. 5A

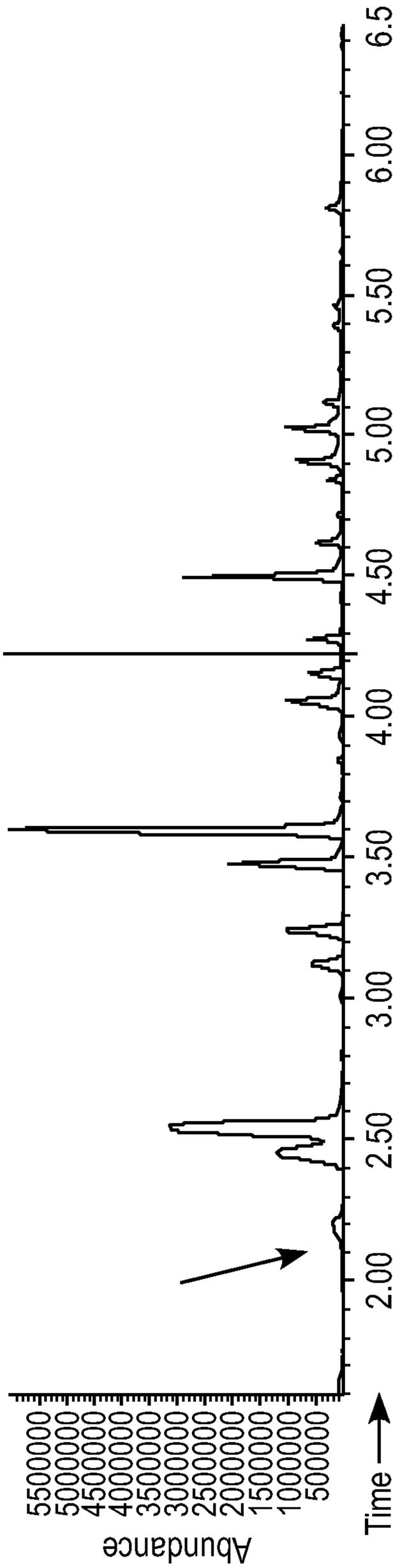


FIG. 5B

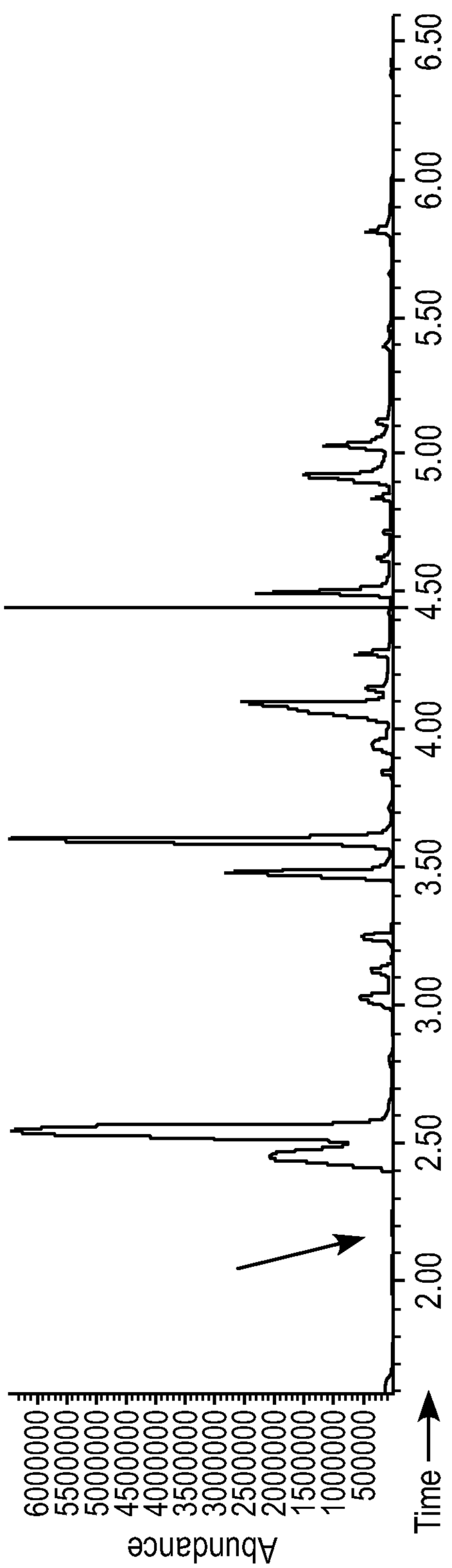


FIG. 5C

FIG. 6

AAR91681.1

Nucleotide sequence (SEQ ID NO:15)

>gi|40796034:488-4012 Nocardia sp. NRRL 5646 ATP/NADPH-dependent carboxylic acid reductase (car) gene, complete cds

ATGGCAGTGGATTCACCGGATGAGCGGCTACAGCGCCGCATTGCACAGT
TGTTTGCAGAAGATGAGCAGGTCAAGGCCGCACGTCCGCTCGAAGCGGT
GAGCGCGGCGGTGAGCGCGCCCGGTATGCGGCTGGCGCAGATCGCCGCC
ACTGTTATGGCGGGTTACGCCGACCGCCCGGCCGCCGGGCAGCGTGCGT
TCGAACTGAACACCGACGACGCGACGGGCCGCACCTCGCTGCGGTTACT
TCCCCGATTCGAGACCATCACCTATCGCGAACTGTGGCAGCGAGTCGGC
GAGGTTGCCGCGGCCTGGCATCATGATCCCGAGAACCCCTTGCGCGCAG
GTGATTTCGTCGCCCTGCTCGGCTTCACCAGCATCGACTACGCCACCCT
CGACCTGGCCGATATCCACCTCGGCGCGGTTACCGTGCCGTTGCAGGCC
AGCGCGGCGGTGTCCCAGCTGATCGCTATCCTCACCGAGACTTCGCCGC
GGCTGCTCGCCTCGACCCCGGAGCACCTCGATGCGGCGGTCGAGTGCCT
ACTCGCGGGCACCACACCGGAACGACTGGTGGTCTTCGACTACCACCCC
GAGGACGACGACCAGCGTGCGGCCTTCGAATCCGCCCGCCGCCGCCTTG
CCGACGCGGGCAGCTTGGTGATCGTCGAAACGCTCGATGCCGTGCGTGC
CCGGGGCCGCGACTTACCGGCCGCGCCACTGTTCGTTCCCGACACCGAC
GACGACCCGCTGGCCCTGCTGATCTACACCTCCGGCAGCACCGGAACGC
CGAAGGGCGCGATGTACACCAATCGGTTGGCCGCCACGATGTGGCAGGG
GAACTCGATGCTGCAGGGGAACTCGCAACGGGTCGGGATCAATCTCAAC
TACATGCCGATGAGCCACATCGCCGGTCGCATATCGCTGTTCGGCGTGC
TCGCTCGCGGTGGCACCGCATACTTCGCGGCCAAGAGCGACATGTCGAC
ACTGTTCGAAGACATCGGCTTGGTACGTCCCACCGAGATCTTCTTCGTC
CCGCGCGTGTGCGACATGGTCTTCCAGCGCTATCAGAGCGAGCTGGACC
GGCGCTCGGTGGCGGGCGCCGACCTGGACACGCTCGATCGGGAAGTGAA
AGCCGACCTCCGGCAGAACTACCTCGGTGGGCGCTTCCTGGTGGCGGTC
GTCGGCAGCGCGCCGCTGGCCGCGGAGATGAAGACGTTCATGGAGTCCG
TCCTCGATCTGCCACTGCACGACGGGTACGGGTCGACCGAGGCGGGCGC
AAGCGTGCTGCTCGACAACCAGATCCAGCGGCCGCCGGTGCTCGATTAC
AAGCTCGTCGACGTGCCCGAACTGGGTTACTTCCGCACCGACCGGCCGC
ATCCGCGCGGTGAGCTGTTGTTGAAGGCGGAGACCACGATTCCGGGCTA
CTACAAGCGGCCCGAGGTCACCGCGGAGATCTTCGACGAGGACGGCTTC
TACAAGACC

FIG. 6 (Cont)

GGCGATATCGTGGCCGAGCTCGAGCACGATCGGCTGGTCTATGTCGACC
GTCGCAACAATGTGCTCAAACTGTCGCAGGGCGAGTTCGTGACCGTCGC
CCATCTCGAGGCCGTGTTCGCCAGCAGCCCGCTGATCCGGCAGATCTTC
ATCTACGGCAGCAGCGAACGTTCCTATCTGCTCGCGGTGATCGTCCCCA
CCGACGACGCGCTGCGCGGCCGCGACACCGCCACCTTGAAATCGGCACT
GGCCGAATCGATTCAGCGCATCGCCAAGGACGCGAACCTGCAGCCCTAC
GAGATTCCGCGCGATTTCCTGATCGAGACCGAGCCGTTCACCATCGCCA
ACGGACTGCTCTCCGGCATCGCGAAGCTGCTGCGCCCCAATCTGAAGGA
ACGCTACGGCGCTCAGCTGGAGCAGATGTACACCGATCTCGCGACAGGC
CAGGCCGATGAGCTGCTCGCCCTGCGCCGCGAAGCCGCCGACCTGCCGG
TGCTCGAAACCGTCAGCCGGGCAGCGAAAGCGATGCTCGGCGTCGCCTC
CGCCGATATGCGTCCCGACGCGCACTTCACCGACCTGGGCGGCGATTCC
CTTTCCGCGCTGTCGTTCTCGAACCTGCTGCACGAGATCTTCGGGGTCG
AGGTGCCGGTGGGTGTCGTCGTCAGCCCGGCGAACGAGCTGCGCGATCT
GGCGAATTACATTGAGGCGGAACGCAACTCGGGCGCGAAGCGTCCCACC
TTCACCTCGGTGCACGGCGGCGGTTCCGAGATCCGCGCCGCCGATCTGA
CCCTCGACAAGTTCATCGATGCCCGCACCCTGGCCGCCGCCGACAGCAT
TCCGCACGCGCCGGTGCCAGCGCAGACGGTGCTGCTGACCGGCGCGAAC
GGCTACCTCGGCCGGTTCCTGTGCCTGGAATGGCTGGAGCGGCTGGACA
AGACGGGTGGCACGCTGATCTGCGTCGTGCGCGGTAGTGACGCGGCCGC
GGCCCGTAAACGGCTGGACTCGGCGTTCGACAGCGGCGATCCCGGCCTG
CTCGAGCACTACCAGCAACTGGCCGCACGGACCCTGGAAGTCCTCGCCG
GTGATATCGGCGACCCGAATCTCGGTCTGGACGACGCGACTTGGCAGCG
GTTGGCCGAAACCGTCGACCTGATCGTCCATCCCGCCGCGTTGGTCAAC
CACGTCCTTCCCTACACCCAGCTGTTCGGCCCCAATGTCGTCGGCACCG
CCGAAATCGTCCGGTTGGCGATCACGGCGCGGCGCAAGCCGGTCACCTA
CCTGTCGACCGTCGGAGTGGCCGACCAGGTCGACCCGGCGGAGTATCAG
GAGGACAGCGACGTCCGCGAGATGAGCGCGGTGCGCGTCGTGCGCGAGA
GTTACGCCAACGGCTACGGCAACAGCAAGTGGGCGGGGGAGGTCCTGCT
GCGCGAAGCACACGATCTGTGTGGCTTGCCGGTCGCGGTGTTCCGTTCG
GACATGATCCTGGCGCACAGCCGGTACGCGGGTCAGCTCAACGTCCAGG
ACGTGTTCACCCGGCTGATCCTCAGCCTGGTCGCCACCGGCATCGCGCC
GTACTCGTTCTACCGAACCGACGCGGACGGCAACCGGCAGCGGGCCCAC
TATGACGGCTTGCCGGCGGACTTCACGGCGGCGGCGATCACCGCGCTCG
GCATCCAAGCCACCGAAGGCTTCCGGACCTACGACGTGCTCAATCCGTA
CGACGATGGCATCTCCCTCGATGAATTCGTCGACTGGCTCGTCGAATCC
GGCCACCCGATCCAGCGCATCACCGACTACAGCGACTGGTTCCACCGTT
TCGAGACGGCGATCCGCGCGCTGCCGGAAAAGCAACGCCAGGCCTCGGT
GCTGCCGTTGCTGGACGCCTACCGCAACCCCTGCCCGGCGGTCCGCGGC
GCGATACTCCCGGCCAAGGAGTTCCAAGCGGCGGTGCAAACAGCCAAAA
TCGGTCCGGAACAGGACATCCCGCATTTG

FIG. 6 (Cont)

TCCGCGCCACTGATCGATAAGTACGTCAGCGATCTGGAACTGCTTCAGC
TGCTCTGA

Amino acid sequence (SEQ ID NO:16)

>gi|40796035|gb|AAR91681.1| ATP/NADPH-dependent carboxylic acid reductase [Nocardia sp. NRRL 5646]

MAVDSPDERLQRRIAQLFAEDEQVKAARPLEAVSAAVSAPGMRLAQIAA
TVMAGYADRPAAGQRAFELNTDDATGRTSLRLLPRFETITYRELWQRVG
EVAAAWHHDPENPLRAGDFVALLGFTSIDYATLDLADIHLGAVTVPLQA
SAAVSQLIAILTETSPRLLASTPEHLDAAVECLLAGTTPERLVVFDYHP
EDDDQRAAFESARRRLADAGSLVIVETLDAVRARGRDLPAAPLFVPDTD
DDPLALLIYTSGSTGTPKGAMYTNRLAATMWQGNSMLQGNSQRVGINLN
YMPMSHIAGRISLFGVLARGGTAYFAAKSDMSTLFEDIGLVRPTEIFFV
PRVCDMVFQRYQSELDRRSVAGADLDTLDREVKADLRQNYLGGRFLVAV
VGSAPLAAEMKTFMESVLDLPLHDGYGSTEAGASVLLDNQIQRPPVLDY
KLVDVPELGYFRTDRPHPRGELLLKAETTIPGYYKRPEVTAEIFDEDGF
YKTGDIVAELEHDRLVYVDRRNNVLKLSQGEFVTVAHLEAVFASSPLIR
QIFIYGSSERSYLLAVIVPTDDALRGRDTATLKSALAESIQRIAKDANL
QPYEIPRDFLIETEPFTIANGLLSGIAKLLRPNLKERYGAQLEQMYTDL
ATGQADELLALRREAADLPVLETVSRAAKAMLGVASADMRPDAHFTDLG
GDSLSALSFSNLLHEIFGVEVPVGVVVSPANELRDLANYIEAERNSGAK
RPTFTSVHGGGSEIRAADLTLDKFIDARTLAAADSIPHAPVPAQTVLLT
GANGYLGRFLCLEWLERLDKTGGTLICVVRGSDAAAARKRLDSAFDSGD
PGLLEHYQQLAARTLEVLAGDIGDPNLGLDDATWQRLAETVDLIVHPAA
LVNHVLPYTQLFGPNVVGTAEIVRLAITARRKPVTYLSTVGVADQVDPA
EYQEDSDVREMSAVRVVRESYANGYGNSKWAGEVLLREAHDLCGLPVAV
FRSDMILAHSRYAGQLNVQDVFTRLILSLVATGIAPYSFYRTDADGNRQ
RAHYDGLPADFTAAAITALGIQATEGFRTYDVLNPYDDGISLDEFVDWL
VESGHPIQRITDYSDWFHRFETAIRALPEKQRQASVLPLLDA
YRNPCPAVRGAILPAKEFQAAVQTAKIGPEQDIPHLSAPLIDKYVSDLE
LLQLL

FIG. 7

Motif 1

-G-Y-X-X-S/A/T-K-W/L (SEQ ID NO:7); and

-G-X-X-G-X-L-G (SEQ ID NO:8); and

-L/V/I-G-G-D-S-X-X-A (SEQ ID NO:9); and

-[LIVMFY]-{E}-{VES}-[STG]-[STAG]-G-[ST]-[STEIA]-[SG]-X-[PASLIVM]-[KR] (SEQ ID NO:10), where {X} stands for any amino acid except X and [X_1X_2] stands for X_1 or X_2

Motif 2

RTVLLX_1GAX_2GX_3LGRX_4LX_5LX_6WL (SEQ ID NO:11)

where X_1 is S or T;

X_2 is T or N;

X_3 is F or W;

X_4 is F or Y;

X_5 is A or T; and

X_6 is E or Q

Motif 3

LXXGXXGXLGXXLXLXWLXR (SEQ ID NO:12)

Motif 4

WAXEVLLR (SEQ ID NO:13), where X can be any amino acid; or

LXXGXXGXLGXXLXX$_1$XX$_2$LX$_3$R (SEQ ID NO:14), where

X_1 is Leu or Ile;

X_2 is Trp or Leu; and

FIG. 7 (Cont)

X_3 varies between 13 amino acids or 14 amino acids

Motif 5

-G-Y-X-X-S/A/T-K-W/L (SEQ ID NO:7); and

-L/V/I-G-G-D-S-X-X-A (SEQ ID NO:9); and

-[LIVMFY]-{E}-{VES}-[STG]-[STAG]-G-[ST]-[STEIA]-[SG]-X-[PASLIVM]-[KR] (SEQ ID NO:10), where {X} stands for any amino acid except X and [X_1X_2] stands for X_1 or X_2; and

RTVLLX_1GAX_2GX_3LGRX_4LX_5LX_6WL (SEQ ID NO:11), where

X_1 is S or T;

X_2 is T or N;

X_3 is F or W;

X_4 is F or Y;

X_5 is A or T; and

X_6 is E or Q

FIG. 8

NP 217106 (FADD9)

Nucleotide sequence (SEQ ID NO:17)

>gi|57116681:2917871-2921377 Mycobacterium tuberculosis H37Rv, complete genome

ATGTCGATCAACGATCAGCGACTGACACGCCGCGTCGAGGACCTATACGCC
AGCGACGCCCAGTTCGCCGCCGCCAGTCCCAACGAGGCGATCACCCAGGCG
ATCGACCAGCCCGGGGTCGCGCTTCCACAGCTCATCCGTATGGTCATGGAG
GGCTACGCCGATCGGCCGGCACTCGGCCAGCGTGCGCTCCGCTTCGTCACC
GACCCCGACAGCGGCCGCACCATGGTCGAGCTACTGCCGCGGTTCGAGACC
ATCACCTACCGCGAACTGTGGGCCCGCGCCGGCACATTGGCCACCGCGTTG
AGCGCTGAGCCCGCGATCCGGCCGGGCGACCGGGTTTGCGTGCTGGGCTTC
AACAGCGTCGACTACACAACCATCGACATCGCGCTGATCCGGTTGGGCGCC
GTGTCGGTTCCACTGCAGACCAGTGCGCCGGTCACCGGGTTGCGCCCGATC
GTCACCGAGACCGAGCCGACGATGATCGCCACCAGCATCGACAATCTTGGC
GACGCCGTCGAAGTGCTGGCCGGTCACGCCCCGGCCCGGCTGGTCGTATTC
GATTACCACGGCAAGGTTGACACCCACCGCGAGGCCGTCGAAGCCGCCCGA
GCTCGGTTGGCCGGCTCGGTGACCATCGACACACTTGCCGAACTGATCGAA
CGCGGCAGGGCGCTGCCGGCCACACCCATTGCCGACAGCGCCGACGACGCG
CTGGCGCTGCTGATTTACACCTCGGGTAGTACCGGCGCACCCAAAGGCGCC
ATGTATCGCGAGAGCCAGGTGATGAGCTTCTGGCGCAAGTCGAGTGGCTGG
TTCGAGCCGAGCGGTTACCCCTCGATCACGCTGAACTTCATGCCGATGAGC
CACGTCGGGGGCCGTCAGGTGCTCTACGGGACGCTTTCCAACGGCGGTACC
GCCTACTTCGTCGCCAAGAGCGACCTGTCGACGCTGTTCGAGGACCTCGCC
CTGGTGCGGCCCACAGAATTGTGCTTCGTGCCGCGCATCTGGGACATGGTG
TTCGCAGAGTTCCACAGCGAGGTCGACCGCCGCTTGGTGGACGGCGCCGAT
CGAGCGGCGCTGGAAGCGCAGGTGAAGGCCGAGCTGCGGGAGAACGTGCTC
GGCGGACGGTTTGTCATGGCGCTGACCGGTTCCGCGCCGATCTCCGCTGAG
ATGACGGCGTGGGTCGAGTCCCTGCTGGCCGACGTGCATTTGGTGGAGGGT
TACGGCTCCACCGAGGCCGGGATGGTCCTGAACGACGGCATGGTGCGGCGC
CCCGCGGTGATCGACTACAAGCTGGTCGACGTGCCCGAGCTGGGCTACTTC
GGCACCGATCAGCCCTACCCCCGGGGCGAGCTGCTGGTCAAGACGCAAACC
ATGTTCCCCGGCTACTACCAGCGCCCGGATGTCACCGCCGAGGTGTTCGAC
CCCGACGGCTTCTACCGGACCGGGGACATCATGGCCAAAGTAGGCCCCGAC
CAGTTCGTCTACCTCGACCGCCGCAACAACGTGCTAAAGCTCTCCCAGGGC
GAGTTCATCGCCGTGTCGAAGCTCGAGGCGGTGTTCGGCGACAGCCCGCTG
GTCCGACAGATCTTCATCTACGGCAACAGTGCCCGGGCCTACCCGCTGGCG

FIG. 8 (Cont.)

GTGGTTGTCCCGTCCGGGGACGCGCTTTCTCGCCATGGCATCGAGAATCTC
AAGCCCGTGATCAGCGAGTCCCTGCAGGAGGTAGCGAGGGCGGCCGGCCTG
CAATCCTACGAGATTCCACGCGACTTCATCATCGAAACCACGCCGTTCACC
CTGGAGAACGGCCTGCTCACCGGCATCCGCAAGCTGGCACGCCCGCAGTTG
AAGAAGTTCTATGGCGAACGTCTCGAGCGGCTCTATACCGAGCTGGCCGAT
AGCCAATCCAACGAGCTGCGCGAGCTGCGGCAAAGCGGTCCCGATGCGCCG
GTGCTTCCGACGCTGTGCCGTGCCGCGGCTGCGTTGCTGGGCTCTACCGCT
GCGGATGTGCGGCCGGACGCGCACTTCGCCGACCTGGGTGGTGACTCGCTC
TCGGCGCTGTCGTTGGCCAACCTGCTGCACGAGATCTTCGGCGTCGACGTG
CCGGTGGGTGTCATTGTCAGCCCGGCAAGCGACCTGCGGGCCCTGGCCGAC
CACATCGAAGCAGCGCGCACCGGCGTCAGGCGACCCAGCTTCGCCTCGATA
CACGGTCGCTCCGCGACGGAAGTGCACGCCAGCGACCTCACGCTGGACAAG
TTCATCGACGCTGCCACCCTGGCCGCAGCCCCGAACCTGCCGGCACCGAGC
GCCCAAGTGCGCACCGTACTGCTGACCGGCGCCACCGGCTTTTTGGGTCGC
TACCTGGCGCTGGAATGGCTCGACCGCATGGACCTGGTCAACGGCAAGCTG
ATCTGCCTGGTCCGCGCCAGATCCGACGAGGAAGCACAAGCCCGGCTGGAC
GCGACGTTCGATAGCGGCGACCCGTATTTGGTGCGGCACTACCGCGAATTG
GGCGCCGGCCGCCTCGAGGTGCTCGCCGGCGACAAGGGCGAGGCCGACCTG
GGCCTGGACCGGGTCACCTGGCAGCGGCTAGCCGACACGGTGGACCTGATC
GTGGACCCCGCGGCCCTGGTCAACCACGTGCTGCCGTATAGCCAGCTGTTC
GGCCCAAACGCGGCGGGCACCGCCGAGTTGCTTCGGCTGGCGCTGACCGGC
AAGCGCAAGCCATACATCTACACCTCGACGATCGCCGTGGGCGAGCAGATC
CCGCCGGAGGCGTTCACCGAGGACGCCGACATCCGGGCCATCAGCCCGACC
CGCAGGATCGACGACAGCTACGCCAACGGCTACGCGAACAGCAAGTGGGCC
GGCGAGGTGCTGCTGCGCGAAGCTCACGAGCAGTGCGGCCTGCCGGTGACG
GTCTTCCGCTGCGACATGATCCTGGCCGACACCAGCTATACCGGTCAGCTC
AACCTGCCGGACATGTTCACCCGGCTGATGCTGAGCCTGGCCGCTACCGGC
ATCGCACCCGGTTCGTTCTATGAGCTGGATGCGCACGGCAATCGGCAACGC
GCCCACTATGACGGCTTGCCGGTCGAATTCGTCGCAGAAGCCATTTGCACC
CTTGGGACACATAGCCCGGACCGTTTTGTCACCTACCACGTGATGAACCCC
TACGACGACGGCATCGGGCTGGACGAGTTCGTCGACTGGCTCAACTCCCCA
ACTAGCGGGTCCGGTTGCACGATCCAGCGGATCGCCGACTACGGCGAGTGG
CTGCAGCGGTTCGAGACTTCGCTGCGTGCCTTGCCGGATCGCCAGCGCCAC
GCCTCGCTGCTGCCCTTGCTGCACAACTACCGAGAGCCTGCAAAGCCGATA
TGCGGGTCAATCGCGCCCACCGACCAGTTCCGCGCTGCCGTCCAAGAAGCG
AAAATCGGTCCGGACAAAGACATTCCGCACCTCACGGCGGCGATCATCGCG
AAGTACATCAGCAACCTGCGACTGCTCGGGCTGCTGTGA

FIG. 8 (Cont.)

Amino acid sequence (SEQ ID NO:18)

>gi|15609727|ref|NP_217106.1| fatty-acid-CoA ligase [Mycobacterium tuberculosis H37Rv]

MSINDQRLTRRVEDLYASDAQFAAASPNEAITQAIDQPGVALPQLIRMVME
GYADRPALGQRALRFVTDPDSGRTMVELLPRFETITYRELWARAGTLATAL
SAEPAIRPGDRVCVLGFNSVDYTTIDIALIRLGAVSVPLQTSAPVTGLRPI
VTETEPTMIATSIDNLGDAVEVLAGHAPARLVVFDYHGKVDTHREAVEAAR
ARLAGSVTIDTLAELIERGRALPATPIADSADDALALLIYTSGSTGAPKGA
MYRESQVMSFWRKSSGWFEPSGYPSITLNFMPMSHVGGRQVLYGTLSNGGT
AYFVAKSDLSTLFEDLALVRPTELCFVPRIWDMVFAEFHSEVDRRLVDGAD
RAALEAQVKAELRENVLGGRFVMALTGSAPISAEMTAWVESLLADVHLVEG
YGSTEAGMVLNDGMVRRPAVIDYKLVDVPELGYFGTDQPYPRGELLVKTQT
MFPGYYQRPDVTAEVFDPDGFYRTGDIMAKVGPDQFVYLDRRNNVLKLSQG
EFIAVSKLEAVFGDSPLVRQIFIYGNSARAYPLAVVVPSGDALSRHGIENL
KPVISESLQEVARAAGLQSYEIPRDFIIETTPFTLENGLLTGIRKLARPQL
KKFYGERLERLYTELADSQSNELRELRQSGPDAPVLPTLCRAAAAALLGSTA
ADVRPDAHFADLGGDSLSALSLANLLHEIFGVDVPVGVIVSPASDLRALAD
HIEAARTGVRRPSFASIHGRSATEVHASDLTLDKFIDAATLAAAPNLPAPS
AQVRTVLLTGATGFLGRYLALEWLDRMDLVNGKLICLVRARSDEEAQARLD
ATFDSGDPYLVRHYRELGAGRLEVLAGDKGEADLGLDRVTWQRLADTVDLI
VDPAALVNHVLPYSQLFGPNAAGTAELLRLALTGKRKPYIYTSTIAVGEQI
PPEAFTEDADIRAISPTRRIDDSYANGYANSKWAGEVLLREAHEQCGLPVT
VFRCDMILADTSYTGQLNLPDMFTRLMLSLAATGIAPGSFYELDAHGNRQR
AHYDGLPVEFVAEAICTLGTHSPDRFVTYHVMNPYDDGIGLDEFVDWLNSP
TSGSGCTIQRIADYGEWLQRFETSLRALPDRQRHASLLPLLHNYREPAKPI
CGSIAPTDQFRAAVQEAKIGPDKDIPHLTAAIIAKYISNLRLLGLL

ABK75684 (CARA)

Nucleotide sequence (SEQ ID NO:19)

>gi|118168627:3015785-3019291 Mycobacterium smegmatis str. MC2 155, complete genome

TTACAGCAATCCGAGCATCTGCAGGTTGCTGATGTACTTGACGATCACGTC
GGCCGTGACGTGCGGAATGTCCTTGTCGGGGCCGATCTTCGCGTCCTGCAC

FIG. 8 (Cont.)

CGCGGCACGGAACCGGTCGGTGGGTGCCATGGCACCGCACACGGGCGGTGA
GGGCTGCTGATAGTTGTGCAGCAGCGGCAGCAGCGAGGCCTGACGTTGCCG
TTCCGGCAGGGCCCGCAGTGCGGTTTCGAACCGGCTCAGCCAGGTGGCGTA
GTCGTCGACGCGGTGCACGGGGTAGCCGGCCTCGATCAGCCAGTCCACGTA
CTCGTCGAGGCCGATGCCGTCGTCGTACGGGTTCATCACGTGGAACGTCTC
GAATCCGTCGGTGACCTGCGAGCCGATGGTGGAGATCGCCTCGGCGATGAA
CTCCACGGGCAGCCCGTCGTAGTGGGCGCGCTGCCGGTTGCCGTCCGCATC
GAGTTCGTAGAACGAACCGGGCGCGATGCCGGTCGCCACGAGGCTCAGCAT
CAGGCGGGTGAACATGTCCGGCAGGTTCAGCTGACCCGAGTAGGTCGTGTC
GGCCAGGATCATGTCGCAGCGGAACACCGAGACCGGCAGACCACACCAGTC
GTGCGCCTCCCGCAGCAGGACCTCGCCGGCCCACTTGCTGTTGCCGTAGCC
GTTGGCGTACGAGTCGTCGACCCGGCGCGTCGCGCTGATCTCGCGGATGTC
GGCGTCCTCGACGAACGCCTCGGGGGAGATGCCCTGTCCCACACCGATCGT
CGAGACGTACACGTACGGCTTGATCGTGGTGGTCAGCGCGATCCGGATGAG
TTCGGCGGTGCCGAGCGCATTGGGTCCGAACATCTGGCTGTACGGCAGGAC
GTGATTGACCAGGGCGGCCGGATCGACGATCAGATCGACGGTGTCGGCCAG
TCGCTGCCACGTGTCGTGGTCGAGACCCAGATCGGCCTCGCCCTTGTCACC
GGCGATCACCTCGAGGTGATCGGCTGCCAGCGCGCGGTAGTGCTCGAGCAG
TGTCGCGTCCCCGGTGTCGAACGTGGCGTCCAGACGCGCCCGGGCCTCGTC
GTCGCTGCGGGCGCGCACCAGGCAGATCACCTTGCCGTCCACCAGGTCCAT
GCGCTCCAGCCATTCCAGCGCCAGATAGCGGCCCAGGAACCCGGTGGCGCC
GGTCAGCAGCACGGTGCGGATCTCGGTGCCCGAACGCGGCAGACCCGGCGC
GGCGGACAGGGTCTTGGCGTCGATGAACTTGCCCAGGGCGAGATCACGCGC
GCGCACCTCGGTGGCGTCGCGCCCGTGCACCGACGCGTATGTGGGGCGCTT
GGAGCCGCGCAGTTCGCCCTCGATGTAGGCCGCGACGCCTGCCAGGTCGGT
GGCCGGGCTGACGATGACGCCGACCGGCACGTCGACATCGAAGATCTCGTG
CAACAGGTTCGAGAAGCTCAAGGCCGACAACGAATCTCCACCCAGATCGGT
GAAGTGCGCATCGGACCGCAGATCCGTGACGGAGGCACCGAGCAGTGCGAC
CGCGGCGCGGCTGACGGTCTCGACCACGGGCCGGTCGGCTCCGTTGCGGCG
CAACTCGCGCAACTCGTTGGCCTGCCCCTCGGCCAGGTCGGTGTAGAGCTG
TTCGAGGCGTTCGCCGTAGTGCGCCTTCAGTTTCGGCCGGGCCAGCTTGCG
GATACCGGTCAGCAGGCCGTTCTCCAGCGTGAAAGGTGTTGTCTCGACGAG
GAAGTCACGCGGGATCTCATACGACTGCAATCCGGCGGCTCGTGCCGCGTC
CTGCAGTGAGTCGCTGATGCGCGACTTGAGTTCGTCACCGTCCCAACGTGA
CAGTGCCTCTTCGGTCGGGACCACGACCGCCAGCAGATAGGACCGCGCGCT
GTTGCCGTAGACGTAGATCTGGCGTACCAGGGGGCTGTCGCCGAACACCGC
CTCCAGCTTGGAGACCGTGACGAATTCGCCCTGCGACAGTTTCAGCACGTT
GTTGCGGCGGTCGAGGTATTCGAGATGGTCGGGCCCGAGCTCGGCGACGAT
GTCGCCGGTGCGGTAGTACCCGTCCTCGTCGAACATCTCGGCGGTGATCTC

FIG. 8 (Cont.)

CGGACGCTTGTAGTAGCCGGGGAACATCTGCTCGGACTTGACCAGAAGTTC
GCCGCGCGGGTAGGGCCGGTCCGTGGCGAAGTAGCCGAGATCGGGCACGTC
GACCAGCTTGTAGTCGATGACCGGCGGGCGCTGGATCTGCCCGTCGATGAA
CACCGCGCCGGCCTCGGTGGAGCCGTAGCCCTCCAGCAGATGCATGTCGAG
CAGGTCCTCGACCCAGCTCTTCATCTCCGCCGAGATGGGAGCCGATCCGGT
CAGGGCCGAAACGAATCGCCCGCCGAGCAGTTGGGTGCGGACCTCTTCGAG
GACTGCGGCTTCGGCTCGGTCCTCGGATCCCTCGGCGCGGCGGTTGTCGAG
GCGGCTCTGGTACTCCTGGAACAGCATGTCCCAGATGCGAGGAACGAAGTT
GAGCTGCGTGGGCCGCACGAGGGCGAGGTCCTCCAGGAAGGTGGACAGGTC
GCTGCGTGCGGCGAAGTACGCGGTTCCGCCGCTGGCGAGTGTGCTGCACAG
GATGCCGCGCCCCATGACGTGACTCATGGGCATGAAGTTCAGGGTGATCGA
CGGCATCACGCCGAGGGTCTCGTCCCACCGGGCCTTGGACCCGGCCTGCCA
CATCGTGGCGGTCTTGGACTCGGGGTACATCGCGCCCTTGGGAGTGCCGGT
GCTGCCGGAGGTGTAGATGAGAAGGGTCAGCGGGTCGGCCTCGTCGGGCAC
GTAGAGCGGTGCGTCGGCGAGTGACCGCCCGCGGTCCAGTGCGTCGGTGAT
CGTCTCGACGACGACGCCGGTGCCTGCGAGCTTGCCCTTGGCCGCCTCGAA
CGCCTCACGCTGATCGTCGACCTCGTGGCTGTAGTCGAACACCACCAGTCG
CGACGGCGCGGGCCCGGACTCGACGAGAGCGACTGCGTCGGCGAGGAAGTC
GACGCTCGACGCGATCACCTTGGGCTCGGTCTCGGCGACGATCGGCTGCAG
TTGGGCCACCGGCGCACTGGTCTGCAGCGGTACGGACACGGCGCCGAGTTC
GAGCAGGGCGATGTCGATCGTCGTGTAGTCGACACTGGTGAAACCCAGGAT
GGCCACGCGGTCACCGGCATTCACCGGATGGTTGTGCCAGGCATTGGTCAC
GGCCTGGATCCGGCCTGCGAGCTGACGGTAGGTGATGGTGTCGAAGCGGGG
CAGGAGCTTCGCGGTGGTGCGGCCTTCTTCGTCGGTGACGAACTCGACGGC
GCGCTTGCCCAGCGCAGGGCGGTCCGCATAGCCGGCCAGAATCTGTTTGAC
CGCGGCAGGAAGGCGCAACTCCGGATCGGCGGCAGCCGCGCTGATCGCCTC
GTCGGGACGGGCGGCGGCGAACTGCGGGTCGGTTTCGAACAAGTGGTCAAT
GCGCCGGTTGAAGCGGTCTTCGCGCGTTTCGATCGTCAT

Amino acid sequence (SEQ ID NO:20)

>gi|118174788|gb|ABK75684.1| NAD dependent epimerase/dehydratase family protein [Mycobacterium smegmatis str. MC2 155]

MTIETREDRFNRRIDHLFETDPQFAAARPDEAISAAAADPELRLPAAVKQI
LAGYADRPALGKRAVEFVTDEEGRTTAKLLPRFDTITYRQLAGRIQAVTNA
WHNHPVNAGDRVAILGFTSVDYTTIDIALLELGAVSVPLQTSAPVAQLQPI
VAETEPKVIASSVDFLADAVALVESGPAPSRLVVFDYSHEVDDQREAFEAA

FIG. 8 (Cont.)

KGKLAGTGVVVETITDALDRGRSLADAPLYVPDEADPLTLLIYTSGSTGTP
KGAMYPESKTATMWQAGSKARWDETLGVMPSITLNFMPMSHVMGRGILCST
LASGGTAYFAARSDLSTFLEDLALVRPTQLNFVPRIWDMLFQEYQSRLDNR
RAEGSEDRAEAAVLEEVRTQLLGGRFVSALTGSAPISAEMKSWVEDLLDMH
LLEGYGSTEAGAVFIDGQIQRPPVIDYKLVDVPDLGYFATDRPYPRGELLV
KSEQMFPGYYKRPEITAEMFDEDGYYRTGDIVAELGPDHLEYLDRRNNVLK
LSQGEFVTVSKLEAVFGDSPLVRQIYVYGNSARSYLLAVVVPTEEALSRWD
GDELKSRISDSLQDAARAAGLQSYEIPRDFLVETTPFTLENGLLTGIRKLA
RPKLKAHYGERLEQLYTDLAEGQANELRELRRNGADRPVVETVSRAAVALL
GASVTDLRSDAHFTDLGGDSLSALSFSNLLHEIFDVDVPVGVIVSPATDLA
GVAAYIEGELRGSKRPTYASVHGRDATEVRARDLALGKFIDAKTLSAAPGL
PRSGTEIRTVLLTGATGFLGRYLALEWLERMDLVDGKVICLVRARSDDEAR
ARLDATFDTGDATLLEHYRALAADHLEVIAGDKGEADLGLDHDTWQRLADT
VDLIVDPAALVNHVLPYSQMFGPNALGTAELIRIALTTTIKPYVYVSTIGV
GQGISPEAFVEDADIREISATRRVDDSYANGYGNSKWAGEVLLREAHDWCG
LPVSVFRCDMILADTTYSGQLNLPDMFTRLMLSLVATGIAPGSFYELDADG
NRQRAHYDGLPVEFIAEAISTIGSQVTDGFETFHVMNPYDDGIGLDEYVDW
LIEAGYPVHRVDDYATWLSRFETALRALPERQRQASLLPLLHNYQQPSPPV
CGAMAPTDRFRAAVQDAKIGPDKDIPHVTADVIVKYISNLQMLGLL

YP 889972 (CARB)

Nucleotide sequence (SEQ ID NO:21)

>gi|118467340:5821317-5824838 Mycobacterium
smegmatis str. MC2 155, complete genome

ATGACCAGCGATGTTCACGACGCCACAGACGGCGTCACCGAAACCGCACTC
GACGACGAGCAGTCGACCCGCCGCATCGCCGAGCTGTACGCCACCGATCCC
GAGTTCGCCGCCGCCGCACCGTTGCCCGCCGTGGTCGACGCGGCGCACAAA
CCCGGGCTGCGGCTGGCAGAGATCCTGCAGACCCTGTTCACCGGCTACGGT
GACCGCCCGGCGCTGGGATACCGCGCCCGTGAACTGGCCACCGACGAGGGC
GGGCGCACCGTGACGCGTCTGCTGCCGCGGTTCGACACCCTCACCTACGCC
CAGGTGTGGTCGCGCGTGCAAGCGGTCGCCGCGGCCCTGCGCCACAACTTC
GCGCAGCCGATCTACCCCGGCGACGCCGTCGCGACGATCGGTTTCGCGAGT
CCCGATTACCTGACGCTGGATCTCGTATGCGCCTACCTGGGCCTCGTGAGT
GTTCCGCTGCAGCACAACGCACCGGTCAGCCGGCTCGCCCCGATCCTGGCC
GAGGTCGAACCGCGGATCCTCACCGTGAGCGCCGAATACCTCGACCTCGCA
GTCGAATCCGTGCGGGACGTCAACTCGGTGTCGCAGCTCGTGGTGTTCGAC

FIG. 8 (Cont.)

CATCACCCCGAGGTCGACGACCACCGCGACGCACTGGCCCGCGCGCGTGAA
CAACTCGCCGGCAAGGGCATCGCCGTCACCACCCTGGACGCGATCGCCGAC
GAGGGCGCCGGGCTGCCGGCCGAACCGATCTACACCGCCGACCATGATCAG
CGCCTCGCGATGATCCTGTACACCTCGGGTTCCACCGGCGCACCCAAGGGT
GCGATGTACACCGAGGCGATGGTGGCGCGGCTGTGGACCATGTCGTTCATC
ACGGGTGACCCCACGCCGGTCATCAACGTCAACTTCATGCCGCTCAACCAC
CTGGGCGGGCGCATCCCCATTTCCACCGCCGTGCAGAACGGTGGAACCAGT
TACTTCGTACCGGAATCCGACATGTCCACGCTGTTCGAGGATCTCGCGCTG
GTGCGCCCGACCGAACTCGGCCTGGTTCCGCGCGTCGCCGACATGCTCTAC
CAGCACCACCTCGCCACCGTCGACCGCCTGGTCACGCAGGGCGCCGACGAA
CTGACCGCCGAGAAGCAGGCCGGTGCCGAACTGCGTGAGCAGGTGCTCGGC
GGACGCGTGATCACCGGATTCGTCAGCACCGCACCGCTGGCCGCGGAGATG
AGGGCGTTCCTCGACATCACCCTGGGCGCACACATCGTCGACGGCTACGGG
CTCACCGAGACCGGCGCCGTGACACGCGACGGTGTGATCGTGCGGCCACCG
GTGATCGACTACAAGCTGATCGACGTTCCCGAACTCGGCTACTTCAGCACC
GACAAGCCCTACCCGCGTGGCGAACTGCTGGTCAGGTCGCAAACGCTGACT
CCCGGGTACTACAAGCGCCCCGAGGTCACCGCGAGCGTCTTCGACCGGGAC
GGCTACTACCACACCGGCGACGTCATGGCCGAGACCGCACCCGACCACCTG
GTGTACGTGGACCGTCGCAACAACGTCCTCAAACTCGCGCAGGGCGAGTTC
GTGGCGGTCGCCAACCTGGAGGCGGTGTTCTCCGGCGCGGCGCTGGTGCGC
CAGATCTTCGTGTACGGCAACAGCGAGCGCAGTTTCCTTCTGGCCGTGGTG
GTCCCGACGCCGGAGGCGCTCGAGCAGTACGATCCGGCCGCGCTCAAGGCC
GCGCTGGCCGACTCGCTGCAGCGCACCGCACGCGACGCCGAACTGCAATCC
TACGAGGTGCCGGCCGATTTCATCGTCGAGACCGAGCCGTTCAGCGCCGCC
AACGGGCTGCTGTCGGGTGTCGGAAAACTGCTGCGGCCCAACCTCAAAGAC
CGCTACGGGCAGCGCCTGGAGCAGATGTACGCCGATATCGCGGCCACGCAG
GCCAACCAGTTGCGCGAACTGCGGCGCGCGGCCGCCACACAACCGGTGATC
GACACCCTCACCCAGGCCGCTGCCACGATCCTCGGCACCGGGAGCGAGGTG
GCATCCGACGCCCACTTCACCGACCTGGGCGGGGATTCCCTGTCGGCGCTG
ACACTTTCGAACCTGCTGAGCGATTTCTTCGGTTTCGAAGTTCCCGTCGGC
ACCATCGTGAACCCGGCCACCAACCTCGCCCAACTCGCCCAGCACATCGAG
GCGCAGCGCACCGCGGGTGACCGCAGGCCGAGTTTCACCACCGTGCACGGC
GCGGACGCCACCGAGATCCGGGCGAGTGAGCTGACCCTGGACAAGTTCATC
GACGCCGAAACGCTCCGGGCCGCACCGGGTCTGCCCAAGGTCACCACCGAG
CCACGGACGGTGTTGCTCTCGGGCGCCAACGGCTGGCTGGGCCGGTTCCTC
ACGTTGCAGTGGCTGGAACGCCTGGCACCTGTCGGCGGCACCCTCATCACG
ATCGTGCGGGGCCGCGACGACGCCGCGGCCCGCGCACGGCTGACCCAGGCC
TACGACACCGATCCCGAGTTGTCCCGCCGCTTCGCCGAGCTGGCCGACCGC
CACCTGCGGGTGGTCGCCGGTGACATCGGCGACCCGAATCTGGGCCTCACA

FIG. 8 (Cont.)

CCCGAGATCTGGCACCGGCTCGCCGCCGAGGTCGACCTGGTGGTGCATCCG
GCAGCGCTGGTCAACCACGTGCTCCCCTACCGGCAGCTGTTCGGCCCCAAC
GTCGTGGGCACGGCCGAGGTGATCAAGCTGGCCCTCACCGAACGGATCAAG
CCCGTCACGTACCTGTCCACCGTGTCGGTGGCCATGGGGATCCCCGACTTC
GAGGAGGACGGCGACATCCGGACCGTGAGCCCGGTGCGCCCGCTCGACGGC
GGATACGCCAACGGCTACGGCAACAGCAAGTGGGCCGGCGAGGTGCTGCTG
CGGGAGGCCCACGATCTGTGCGGGCTGCCCGTGGCGACGTTCCGCTCGGAC
ATGATCCTGGCGCATCCGCGCTACCGCGGTCAGGTCAACGTGCCAGACATG
TTCACGCGACTCCTGTTGAGCCTCTTGATCACCGGCGTCGCGCCGCGGTCG
TTCTACATCGGAGACGGTGAGCGCCCGCGGGCGCACTACCCCGGCCTGACG
GTCGATTTCGTGGCCGAGGCGGTCACGACGCTCGGCGCGCAGCAGCGCGAG
GGATACGTGTCCTACGACGTGATGAACCCGCACGACGACGGGATCTCCCTG
GATGTGTTCGTGGACTGGCTGATCCGGGCGGGCCATCCGATCGACCGGGTC
GACGACTACGACGACTGGGTGCGTCGGTTCGAGACCGCGTTGACCGCGCTT
CCCGAGAAGCGCCGCGCACAGACCGTACTGCCGCTGCTGCACGCGTTCCGC
GCTCCGCAGGCACCGTTGCGCGGCGCACCCGAACCCACGGAGGTGTTCCAC
GCCGCGGTGCGCACCGCGAAGGTGGGCCCGGGAGACATCCCGCACCTCGAC
GAGGCGCTGATCGACAAGTACATACGCGATCTGCGTGAGTTCGGTCTGATC
TGA

Amino acid sequence (SEQ ID NO:22)

>gi|118469671|ref|YP_889972.1| putative long-chain fatty-acid--CoA ligase [Mycobacterium smegmatis str. MC2 155]

MTSDVHDATDGVTETALDDEQSTRRIAELYATDPEFAAAAPLPAVVDAAHK
PGLRLAEILQTLFTGYGDRPALGYRARELATDEGGRTVTRLLPRFDTLTYA
QVWSRVQAVAAALRHNFAQPIYPGDAVATIGFASPDYLTLDLVCAYLGLVS
VPLQHNAPVSRLAPILAEVEPRILTVSAEYLDLAVESVRDVNSVSQLVVFD
HHPEVDDHRDALARAREQLAGKGIAVTTLDAIADEGAGLPAEPIYTADHDQ
RLAMILYTSGSTGAPKGAMYTEAMVARLWTMSFITGDPTPVINVNFMPLNH
LGGRIPISTAVQNGGTSYFVPESDMSTLFEDLALVRPTELGLVPRVADMLY
QHHLATVDRLVTQGADELTAEKQAGAELREQVLGGRVITGFVSTAPLAAEM
RAFLDITLGAHIVDGYGLTETGAVTRDGVIVRPPVIDYKLIDVPELGYFST
DKPYPRGELLVRSQTLTPGYYKRPEVTASVFDRDGYYHTGDVMAETAPDHL
VYVDRRNNVLKLAQGEFVAVANLEAVFSGAALVRQIFVYGNSERSFLLAVV
VPTPEALEQYDPAALKAALADSLQRTARDAELQSYEVPADFIVETEPFSAA
NGLLSGVGKLLRPNLKDRYGQRLEQMYADIAATQANQLRELRRAAATQPVI

FIG. 8 (Cont.)

DTLTQAAATILGTGSEVASDAHFTDLGGDSLSALTLSNLLSDFFGFEVPVG
TIVNPATNLAQLAQHIEAQRTAGDRRPSFTTVHGADATEIRASELTLDKFI
DAETLRAAPGLPKVTTEPRTVLLSGANGWLGRFLTLQWLERLAPVGGTLIT
IVRGRDDAAARARLTQAYDTDPELSRRFAELADRHLRVVAGDIGDPNLGLT
PEIWHRLAAEVDLVVHPAALVNHVLPYRQLFGPNVVGTAEVIKLALTERIK
PVTYLSTVSVAMGIPDFEEDGDIRTVSPVRPLDGGYANGYGNSKWAGEVLL
REAHDLCGLPVATFRSDMILAHPRYRGQVNVPDMFTRLLLSLLITGVAPRS
FYIGDGERPRAHYPGLTVDFVAEAVTTLGAQQREGYVSYDVMNPHDDGISL
DVFVDWLIRAGHPIDRVDDYDDWVRRFETALTALPEKRRAQTVLPLLHAFR
APQAPLRGAPEPTEVFHAAVRTAKVGPGDIPHLDEALIDKYIRDLREFGLI

YP 905678.1

Nucleotide sequence (SEQ ID NO:23)

>uniprot|A0PPD8|A0PPD8_MYCUA Fatty-acid-CoA ligase FadD9

ATGTCGCCAATCACGCGTGAAGAGCGGCTCGAGCGCCGCATCCAGGACCTC
TACGCCAACGACCCGCAGTTCGCCGCCGCCAAACCCGTCACGGCGATCACC
GCAGCAATCGAGCGGCCGGGTCTACCGCTACCCCAGATCATCGAGACCGTC
ATGACCGGATACGCCGATCGGCCGGCTCTCGCTCAGCGCTCGGTCGAATTC
GTGACCGATGCCGGCACCGGCCACACCACGCTGCGACTGCTCCCCCACTTC
GAAACCATCAGCTACGGCGAGCTTTGGGACCGCATCAGCGCACTGGCCGAC
GTGCTCAGCACCGAACAGACGGTGAAACCGAGCGACCGGGTCTGCTTGTTG
GGCTTCAACAGCGTCGACTACGCCACGATCGACATGACTTTGGCGCGGCTG
GGCGCGGTGGCTGTACCACTGCAGACCAGCGCGGCGATAACCCAGCTGCAG
CCGATCGTCGCCGAGACCCAGCCCACCATGATCGCGGCCAGCGTCGACGCA
CTCGCTGACGCCACCGAATTGGCTCTGTCCGGTCAGACCGCTACCCGAGTC
CTGGTGTTCGACCACCACCGGCAGGTTGACGCACACCGCGCAGCGGTCGAA
TCCGCCCGGGAGCGCCTGGCTGGCTCGGCGGTCGTCGAAACCCTGGCCGAG
GCCATCGCGCGCGGCGACGTGCCCCGCGGTGCGTCCGCCGGCTCGGCGCCC
GGCACCGATGTGTCCGACGACTCGCTCGCGCTACTGATCTACACCTCGGGC
AGCACCGGTGCGCCCAAGGGCGCGATGTACCCCCGACGCAACGTTGCGACC
TTCTGGCGCAAGCGCACCTGGTTCGAAGGCGGCTACGAGCCGTCGATCACG
CTGAACTTCATGCCAATGAGCCACGTCATGGGCCGCCAAATCCTGTACGGC
ACGCTGTGCAATGGCGGCACCGCCTACTTCGTGGTGAAAAGCGATCTCTCC
ACCTTGTTCGAAGACCTGGCGCTGGTGCGGCCCACCGAGCTGACCTTCGTG

FIG. 8 (Cont.)

CCGCGCGTGTGGGACATGGTGTTCGACGAGTTTCAGAGTGAGGTCGACCGC
CGCCTGGTCGACGGCGCCGACCGGGTCGCGCTCGAAGCCCAGGTCAAGGCC
GAGATACGCAACGACGTGCTCGGTGGACGGTATACCAGCGCACTGACCGGC
TCCGCCCCGATCTCCGACGAGATGAAGGCGTGGGTCGAGGAGCTGCTCGAC
ATGCATCTGGTCGAGGGCTACGGCTCCACCGAGGCCGGGATGATCCTGATC
GACGGAGCCATTCGGCGCCCGGCGGTACTCGACTACAAGCTGGTCGATGTT
CCCGACCTGGGTTACTTCCTGACCGACCGGCCACATCCGCGGGGCGAGTTG
CTGGTCAAGACCGATAGTTTGTTCCCGGGCTACTACCAGCGAGCCGAAGTC
ACCGCCGACGTGTTCGATGCTGACGGCTTCTACCGGACCGGCGACATCATG
GCCGAGGTCGGCCCCGAACAGTTCGTGTACCTCGACCGCCGCAACAACGTG
TTGAAGCTGTCGCAGGGCGAGTTCGTCACCGTCTCCAAACTCGAAGCGGTG
TTTGGCGACAGCCCACTGGTACGGCAGATCTACATCTACGGCAACAGCGCC
CGTGCCTACCTGTTGGCGGTGATCGTCCCCACCCAGGAGGCGCTGGACGCC
GTGCCTGTCGAGGAGCTCAAGGCGCGGCTGGGCGACTCGCTGCAAGAGGTC
GCAAAGGCCGCCGGCCTGCAGTCCTACGAGATCCCGCGCGACTTCATCATC
GAAACAACACCATGGACGCTGCAGAACGGCCTGCTCACCGGCATCCGCAAG
TTGGCCAGGCCGCAGCTGAAAAAGCATTACGGCGAGCTTCTCGAGCAGATC
TACACGGACCTGGCACACGGCCAGGCCGACGAACTGCGCTCGCTGCGCCAA
AGCGGTGCCGATGCGCCGGTGCTGGTGACGGTGTGCCGCGCGGCGGCCGCG
CTGTTGGGCGGCAGCGCCTCTGACGTCCAGCCCGATGCGCACTTCACCGAT
TTGGGCGGCGACTCGCTGTCGGCGCTGTCGTTCACCAACCTGCTGCACGAG
ATCTTCGACATCGATGTGCCGGTGGGCGTCATCGTCAGCCCCGCCAACGAC
TTGCAGGCCCTGGCCGACTACGTCGAGGCGGCTCGCAAACCCGGCTCGTCA
CGACCGACCTTCGCCTCGGTCCACGGCGCCTCGAATGAGCAGGTCACCGAG
GTGCATGCCGGTGACCTGTCCCTGGACAAATTCATCGATGCCGCAACCCTG
GCCGAAGCTCCCCGGCTGCCCGCCGCAAACACCCAAGTGCGCACCGTGCTG
CTGACCGGCGCCACCGGCTTCCTCGGGCGCTACCTGGCCCTGGAATGGCTG
GAGCGGATGGACCTGGTCGACGGCAAACTGATCTGCCTGGTCCGGGCCAAG
TCCGACACCGAAGCACGGGCGCGGCTGGAAAAGACGTTCGACAGCGGCGCC
CCCGAACTGCTGGCCCACTACCGCGCACTGGCCGGCGACCACCTCGAGGTG
CTCGCCGGTGACAAGGGCGAAGCCGACCTCGGACTGGACCGGCAGACCTGG
CAACGCCTGGCCGACACGGTCGACCTGATCGTGGACCCCGCGGCCCTGGTC
AACCACGTACTGCCATACAGCCAGCTGTTCGGGCCCAACGCGCTGGGCACC
GCCGAGCTGCTGCGGCTCGCGCTCACCTCCAAGATCAAGCCCTACAGCTAC
ACCTCGACAATCGGTGTCGCCGACCAGATCCCGCCGTCGGCGTTCACCGAG
GACGCCGACATCCGGGTCATCAGCGCCACCCCGCGCGGTCGACGACAGCTAC
GCCAATGGCTATTCGAACAGCAAGTGGGCCGGCGAGGTGCTGTTGCGCGAG
GCGCATGTCCTGTGTGGCCTGCCGGTTGCGGTGTTCCGCTGCGACATGATC
CTGGCCGACACCACATGGGCGGGACAGCTCAACGTGCCGGACATGTTCACC

FIG. 8 (Cont.)

CGTATGATCCTGAGCCTGGCGGCCACCGGTATCGCGCCGGGTTCGTTCTAT
GAGCTTGCGGCCGACGGCGCCCGGCAACGCGCCCACTATGACGGTCTGCCC
GTCGAGTTCATCGCCGAGGCGATTTCGACTTTGGGTGCGCAGAGCCAGGAT
GGGTTCCACACGTATCACGTGATGAACCCTTACGACGACGGCATCGGACTC
GACGAGTTCGTCGACTGGCTCAACGAGTCCGGTTGCCCCATCCAGCGCATC
GCTGACTATGGCGACTGGCTGCAGCGCTTCGAAACCGCACTGCGCGCACTG
CCCGATCGGCAGCGGCACAGCTCACTGCTGCCGCTGTTGCACAACTATCGG
CAGCCGGAGCGGCCCGTCCGCGGGTCGATCGCCCCTACCGATCGCTTCCGG
GCAGCGGTGCAAGAGGCCAAGATCGGCCCCGACAAAGACATTCCGCACGTC
GGCGCGCCGATCATCGTGAAGTACGTCAGCGACCTGCGCCTACTCGGCCTG
CTCTGA

Amino acid sequence (SEQ ID NO:24)

>uniprot|A0PPD8|A0PPD8_MYCUA Fatty-acid-CoA ligase FadD9

MSPITREERLERRIQDLYANDPQFAAAKPVTAITAAIERPGLPLPQIIET
VMTGYADRPALAQRSVEFVTDAGTGHTTLRLLPHFETISYGELWDRISAL
ADVLSTEQTVKPSDRVCLLGFNSVDYATIDMTLARLGAVAVPLQTSAAIT
QLQPIVAETQPTMIAASVDALADATELALSGQTATRVLVFDHHRQVDAHR
AAVESARERLAGSAVVETLAEAIARGDVPRGASAGSAPGTDVSDDSLALL
IYTSGSTGAPKGAMYPRRNVATFWRKRTWFEGGYEPSITLNFMPMSHVMG
RQILYGTLCNGGTAYFVVKSDLSTLFEDLALVRPTELTFVPRVWDMVFDE
FQSEVDRRLVDGADRVALEAQVKAEIRNDVLGGRYTSALTGSAPISDEMK
AWVEELLDMHLVEGYGSTEAGMILIDGAIRRPAVLDYKLVDVPDLGYFLT
DRPHPRGELLVKTDSLFPGYYQRAEVTADVFDADGFYRTGDIMAEVGPEQ
FVYLDRRNNVLKLSQGEFVTVSKLEAVFGDSPLVRQIYIYGNSARAYLLA
VIVPTQEALDAVPVEELKARLGDSLQEVAKAAGLQSYEIPRDFIIETTPW
TLQNGLLTGIRKLARPQLKKHYGELLEQIYTDLAHGQADELRSLRQSGAD
APVLVTVCRAAAALLGGSASDVQPDAHFTDLGGDSLSALSFTNLLHEIFD
IDVPVGVIVSPANDLQALADYVEAARKPGSSRPTFASVHGASNEQVTEVH
AGDLSLDKFIDAATLAEAPRLPAANTQVRTVLLTGATGFLGRYLALEWLE
RMDLVDGKLICLVRAKSDTEARARLEKTFDSGAPELLAHYRALAGDHLEV
LAGDKGEADLGLDRQTWQRLADTVDLIVDPAALVNHVLPYSQLFGPNALG
TAELLRLALTSKIKPYSYTSTIGVADQIPPSAFTEDADIRVISATRAVDD
SYANGYSNSKWAGEVLLREAHVLCGLPVAVFRCDMILADTTWAGQLNVPD
MFTRMILSLAATGIAPGSFYELAADGARQRAHYDGLPVEFIAEAISTLGA
QSQDGFHTYHVMNPYDDGIGLDEFVDWLNESGCPIQRIADYGDWLQRFET

FIG. 8 (Cont.)

ALRALPDRQRHSSLLPLLHNYRQPERPVRGSIAPTDRFRAAVQEAKIGPD
KDIPHVGAPIIVKYVSDLRLLGLL

ZP_04027864

Nucleotide sequence (SEQ ID NO:25)

>gi|227980074:564806-568123 Tsukamurella
paurometabola DSM 20162 TpauDRAFT_4083016_Cont3,
whole genome shotgun sequence

ATGTCGATTGAGACGGTGCAGAACGGCGTCCCCGCAGAGGGCTCGGTGCCC
CCCGCCGACCAGCAGACCGAGCGACTGCCGCAGGTGATCGCCAGGATCTTC
GCCCAGTTCGCGGATCGTCCGGCCTTCGCGACCCGCGAGGCGGGGCCGGGG
ACCCCCTACGCCACCGTCTCCTATCGGGAGATCTGGCGGCGCGTCACCGCG
CTGGTGGCCTCCTGGCAGAGCGAAGTGGCTCCGGGAGACTTCGTCGCCATC
CTCGGCTTCACCAGCTCGGACTTCGTGACCGTCGACCTCGCGACCACACTG
CTCGGCGCCCCGAACGTGCCGCTCCAGGCCGGGGCCCCCGCCGCTCGCATC
GCGACCATCCTCGATGAGACCCGGCCGAAGATCCTCGCCGTGAGTGCCGAT
CAGGTCGACCTCGCCCAGGAGGCTCTGGCCGAGTCCGCGGCTACCCCGCGG
GTGGTCGTCTTCGACGGCGAACGCGACGGCTACGAGGGCATCGAGGCGGAC
ATCCTTTCCGGCTCCGCCCTGCCGGCACCGGAGTTCTTCGCGCCCGAGCCC
GGCACCGATCCTCTCGTCACGCTCATCTACACATCCGGCAGCACCGGTACC
CCGAAGGGGGCCATGTACACCGAGCAGTTGGTTCGCGATGCCTGGCTCAAG
GTGGACAGCATCGTCGACATCGACATGCCGGCCGAGTCGCTCCTGCACTTC
CTGCCTATGAGCCATATGTACGGGCGCAACTGGCTGATCGCCGGCCTGGCA
TCGGGCGGGACCGGGTACTTCGCCGGCGCCTCCGATATGTCCACCCTGTTC
GACGACCTCGCCGCCGCCCGGCCCACCGCCATCGGCCTGGTGCCCCGCGTG
TGCGAGCTGATACACCAGCGCTATCTGGCCGTCGAGGCGGACACTGATGCG
GAGACCGCGCGCGTCGAACTGCGTGACCGGGTACTCGGCGGTCGGCTGCAG
GCCGCGATGTGCGGTAGCGCCGCCCTCTCGTCGGAGCTGCAGACCTTCATG
GAGTGGTTGCTCGGAATCGATATCCAGATCGGCTACGGATCCACCGAGGCC
GGTGGTGTCATCCGCGACGGAGTGGTCGTTCGGCCGCCGGTCACGGAGTAC
AAGCTGATCGATGTCCCCGAACTGGGCTACTTCGTCACCGACTCCCCGCAT
CCACGCGGCGAACTCCTGGTCAAGTCGACGCAGTTGATTCCCGGGTACTAC
AACTCCGACAAGCGGATCCGCGACGACGAAGGCTTCTACCGCACCGGCGAT
GTGATGGCCGAGCTGGGACCCGACCGGCTCGAGTACGTCGACCGGCGGAGC
AACGTGATCAAGTTGGCACAGGGAGAGTTCGTGCCGATCGCCCAACTCGAG

FIG. 8 (Cont.)

GCCATCTACGCCGCCGGTCCCGATGTGCACCAGATCTTCCTGTACGGAACC
AGCGAACGCTCCTACCTGATCGGCGTCGTCGTGCCCGCGCCGGGACCCGAC
GGCGAGACCGATGCGCAGACCCGCACCCGCGTACTCGATGGCCTGGCCGCG
ATCGCCCGTGAGAACGATCTCGCTGCCTACGAGGTGCCGCGCGATGTGCTC
ATCGAACGTGATCCCTTCTCTCAGGAGAACGGGCTGCGGTCGGGGATCGGC
AAGCTGGTGCGCCCGGCCCTCATCGCCCGCTACGGTGACCGGTTGCACGAC
CTCTACGCCCAGGCCGACACCCGTCAACGCGAGGGCTTGCGCGCTCTCGAC
GCCTCGGGCCCGATCATCGACACCGTGCTCGGGGCGGCTGCGTTGACGCTC
GGCGCGGATATCGCGGACTTCGACGCCGACACTCGATTCGGCGACCTCGGT
GGCGACTCGTTGTCGGCGCTCTCGCTCGCGACGACGCTCGAAGGCCTCTAC
GACGTGCCCGTCCCCGTGCAGACGATCGTCGGACCGACCGCCACACTCGGC
GGCGTCGCCCGGCACATCGAGAAGGCTCGATCGGGTGGCGTCGCGGCACCG
ACCGCCGACTCGGTGCACGGCGTGGGTGCGAGCGTCGCCCGGGCCACCGAC
CTGACGCTGGAGAAGTTCATCGACCCCGAGCTCCTCGCGCTCGCGCCGACG
CTTCCCGCGGCGACCGGTGAGCCGAACACCGTGCTGCTCACCGGATCCACC
GGCTACCTCGGCCGCTTCCTGCTGCTGGACTGGTTGCGACGGGTCGCTCCG
CACGGCGGCACCGTGATCGCGCTGGTGCGCGGCGCCGACGCCGACGATGCG
CGACGCCGCGTCACGGCCGCGATCGGTGACTCGGATCCTGACCTGACACAA
GAGTTCACGTCACTCGCGGAGCATCACCTCCACGTGATCGCCGGTGACTTC
GGCAGCCCCGCACTCGGACTCGACGATGCCACCTGGAGCGATCTCGCCGGG
CGAGTCGATCACGTGGTGCACTGCGGCGCGCTCGTCAACCACGTGCTGCCC
TACGACCAACTGTTCGGTCCCAATGTGGTGGCCACCGGCGAAGTGGTGCGA
CTCGCACTCACCACGCGCCGCAAGTCCGTGGATTACGTCTCCACGGTGGCT
GTGGTTCCGCAGGATGACGGCCGCGTCCTGGTCGAGGACGACGATGTTCGC
GAGCTCGGCGCCGAACGGCGCATCGGGGCCGATGCCTACGCGAACGGCTAC
GCCGTGAGCAAATGGGCGGGCGAAGTGCTGTTGCATGAGGCAGCCGACCTG
GCGGACCTGCCGGTGCGGGTGTTCCGCTCCGATATGATCTTGGCGCACAGT
CGATTCCACGGACAGTTCAACGAGGTCGACCAGTTCACCCGCCTGCTCCTG
AGTATCGCCGAGACCGGACTGGCGCCGGCGTCGTTCTACACGCCGGATCCG
AGTGGACACCGCCCGCACTACGACGGGCTGCCGGTGGACTTCACCGCCGAA
GCGATCACCACGCTCAGCGCCGCGGGGCGTTCGGGGTACCGGACCTTCCAC
GTGCTCAACGCCAACGATGACGGCGTGAGCCTGGACAGCTTCGTCGACTGG
ATCGCCGCCTCGGGCCGGAGCATCGAACGGATCGACGACTACGACACCTGG
TTCGCCCGGTTCGAGCAGGCGCTCCAGCAGCTCCCCGATGAGGCGCGCCAG
CGGTCGGTGCTGCCCCTGCTGCACGCGGTGCGCGAGCCGGCTCCGGCCGCC
GGGACCTCCGCGCTGTCGGTGGACCGGTTCCGTGGTGCGGTGCGTGAGACC
GGAGTAGGACCGGGGGACATCCCGGTGCTCGATCGCGCCCTGATCGAGAAG
TACCTGCGCGACTTCGAGACCGCGGGCTGGCTCGCGCCCGGTGCGCGCGAC
TGA

FIG. 8 (Cont.)

Amino acid sequence (SEQ ID NO:26)

>gi|227980601|ref|ZP_04027864.1| thioester reductase-like protein [Tsukamurella paurometabola DSM 20162]

MSIETVQNGVPAEGSVPPADQQTERLPQVIARIFAQFADRPAFATREAGPG
TPYATVSYREIWRRVTALVASWQSEVAPGDFVAILGFTSSDFVTVDLATTL
LGAPNVPLQAGAPAARIATILDETRPKILAVSADQVDLAQEALAESAATPR
VVVFDGERDGYEGIEADILSGSALPAPEFFAPEPGTDPLVTLIYTSGSTGT
PKGAMYTEQLVRDAWLKVDSIVDIDMPAESLLHFLPMSHMYGRNWLIAGLA
SGGTGYFAGASDMSTLFDDLAAARPTAIGLVPRVCELIHQRYLAVEADTDA
ETARVELRDRVLGGRLQAAMCGSAALSSELQTFMEWLLGIDIQIGYGSTEA
GGVIRDGVVVRPPVTEYKLIDVPELGYFVTDSPHPRGELLVKSTQLIPGYY
NSDKRIRDDEGFYRTGDVMAELGPDRLEYVDRRSNVIKLAQGEFVPIAQLE
AIYAAGPDVHQIFLYGTSERSYLIGVVVPAPGPDGETDAQTRTRVLDGLAA
IARENDLAAYEVPRDVLIERDPFSQENGLRSGIGKLVRPALIARYGDRLHD
LYAQADTRQREGLRALDASGPIIDTVLGAAALTLGADIADFDADTRFGDLG
GDSLSALSLATTLEGLYDVPVPVQTIVGPTATLGGVARHIEKARSGGVAAP
TADSVHGVGASVARATDLTLEKFIDPELLALAPTLPAATGEPNTVLLTGST
GYLGRFLLLDWLRRVAPHGGTVIALVRGADADDARRRVTAAIGDSDPDLTQ
EFTSLAEHHLHVIAGDFGSPALGLDDATWSDLAGRVDHVVHCGALVNHVLP
YDQLFGPNVVATGEVVRLALTTRRKSVDYVSTVAVVPQDDGRVLVEDDDVR
ELGAERRIGADAYANGYAVSKWAGEVLLHEAADLADLPVRVFRSDMILAHS
RFHGQFNEVDQFTRLLLSIAETGLAPASFYTPDPSGHRPHYDGLPVDFTAE
AITTLSAAGRSGYRTFHVLNANDDGVSLDSFVDWIAASGRSIERIDDYDTW
FARFEQALQQLPDEARQRSVLPLLHAVREPAPAAGTSALSVDRFRGAVRET
GVGPGDIPVLDRALIEKYLRDFETAGWLAPGARD

ZP_05045132

Nucleotide sequence (SEQ ID NO:27)

>gi|254430111:343253-346687 Cyanobium sp. PCC 7001 scf_1106012173546 genomic scaffold, whole genome shotgun sequence

FIG. 8 (Cont.)

GTGAATGAGTCTTCCGCGGACCAGAGTTCCGGCAACGTTTCCGAGGGGTGG
CCTGATGCTTCGGTCACAGCACGGGCCCTGCAGGCTCACCTGCGCTACGAA
CAGATCATCGATGCCATTCTGAGCGGCTACGCCGAGCGCCCTGCTCTGGCG
GAGCGTTCCTACCTGGTGCGGCCGGACCCGAGCACAGGTCAAACGGTGCGT
GTCCACGAGCAGGCCTTCCGCTCGATCAGCTACCGAACCCTGCAGGAACGG
GTTCATGCCCTCACCATGGCCTGGCGCCTTCATCCCGATAGCCCGGTGCAA
GCGGGAGCCTTCGTGGTGCTGGTGGGATTTGCCAGCATCGATTACGCCGTT
CTTGATCTGGCACTGGCCTACACCAAGGGCGTGCCGGTGCCCCTGTCACCG
AACCATTCCAGCGAGGACGATGACGCCATCCTCGGCACAGTCCAGCCCGTC
ACTCTGGCGGTATCGATCAGTGAGTTCTCTGGCTGTGTCGACCTGATCGCC
CGATCGACGTCGATCCGAACTGTGATCGTCTTTGACCTTGACCCTGCCGTC
GACTGCGAGCGCGCCGCACTGGAGAGCGGCATCCGGGCACTCAACGAGAAG
GGGTCAGACGTTGTCGTTCAGACACTGCAGGATCTGATTGACGTTGGGAGA
GACGCAGAGTTCAGCTTCCTGCCGATCCAGGCGCAGGATCAAGATGACCTG
GCACTTCTGATTCACACATCCGGCAGCACAGGCACACCCAAGGGAGCCTGC
ATCTCATCCCGTGCACTGATCAACACCTGGCGCCATGTTTCCGGTCCCTAT
CCAAAAGTGACCGTGGTTCTGGCACCCTTCCACCACATGATGGGACGAGAC
TCGATGATCACGGCATTGGGCGCGGGCGGCACCGCCTACTTCACGCTCAGG
CCTGACCTTTCGACCGTGATTGAAGACATCAGACTGGCACGGCCCACAGGC
CTGGTGCTGTTTCCCCGCCTCTGCGAAGTGATCGAACACCACCTGACTACT
GCGCCGGAGTATTCAGGCAACGAGATCCTCGGAGGCAGACTGCAATCCATT
GTGGTGGCCTCGGCTCCGATCACGCCACGCTTGAAGGCATCCCTGGAGTGC
CTCCTTGGGGTGCCTGTGAGCGAAGGCTACAGCAGCACGGAAACAGCCAGT
GGCGGGCTGGCGATGAATGGACTGCTGAACCGCAACAACATTCTCGCGTAT
CGCCTTCGCGATGTGCCCGAGGCAGGGTATTCAGTGAATGATCGGCCCTTT
CCGCGCGGAGAACTCTGCGTGAAGACCCGCTTCGGTATCTCAGGCTATTTC
AGAAATCCAGAGGCCACTGCAGAGCTGTTCGACGACGATGGCTTCTATTGC
ACCGGTGACATCGTTGAAGAGCGGGCCCCCGATCAAATCGCCATCATCGAC
CGGCGAAAGAATGTCATCAAGCTGGCGCAGGGTGAATATGTCGCTGTGGGC
AGGCTGGAACAGCTTTTCCAGGAAGGTTGTGGTTGCGTGCAGCAGATTCAC
CTCCACGGCGACAGCACAAGGGCCTATCTGCTGGCAGTCGTGGTACCTGAT
CGCAACACCCTTGCACCGCCCGGGTCACGGCAGGCCAGTGAGGCCGAGTTA
AAGGCACGGGTGCGCGAGGAGATTCTCACCTTGGCAAACCAACGGGAGCTG
CGCGGCTTCGAGATCCCTCGAGACCTGATCCTGGCGGAGGAACCCTTCTCC
CAGCAGAACGGTCTGCTGTCGTCCTTGGGTAAGCCGATCCGCCCGGCCATC
CGCGCGCGCTACCGCAGCCGGCTGGAGAGCCTGTATGCCAGCCATGAGGCC
ACCCGAGGCACTGAGCTCGAGGCCATCAGAGCGTCAGCTGGCGCGGTGGAT
GTGGAAACCACCCTGTTGGCGCTGCTGAGCAGCACGCTGGGTGTGGTGTGT
GGGGCTGCCGATCGGCAGACGAGTTTCCGCGAGCTGGGGGGCGACTCCCTG

FIG. 8 (Cont.)

GCCGCTGTGCAGCTGGCGATGGAGATCAAGAAGCAGTTCGGGGTGGGGCTG
GAAGGGAGCCAGATTCTCGGGCCGGGCGGCACGGTGGAAGCGTGGGCGCGG
AGGATCCACACCGCCTCCATCCAGCAGGCCCCGCACCAGCGGGTTGGCAGT
CCCCTCGCCGCCATTCCGGCCGAGGGGTGGCTGAAGCCGGACCACTACAGG
CTGGAGAACCTGATCGGGATTCCCATCGGTACACCCTCAGCCGAGGTGGCC
AGGCCCACAGGCGGGCCCCCTACGGTTCTGCTCACCGGTGCCACCGGTTTT
CTGGGGGGGCGCTTGTGCCTGGAGTGGCTGCAACGGCTGGCTGGCCAGGGG
GGCAGGCTGATCTGTCTGGTGCGCCCCTCGAACAGCCATTCCGCCTGGGAG
CGACTGAGGAACCGCTTCTCCCATCTGGAACCCGAGCAGGTGGCACGCTTC
CGCGAGCTGGCGGGAAGGCATCTGGAGGTGATTCCGGCGGACATCGGAGAG
CCCGGCCTGGGGCTTGAACCGGGTTGCCAGGAGCGGCTCGCCACTGAGGTG
GACGCGATCTGTCACTGCGCAGCGGAAGTGAATCACCGGCTGCCCTATCGC
CACCTCTACCGGCCCAATGTGATCGGCACCGCGGAGATCATTCACCTGGCG
ATCACGACGCGGCTGAAATCGGTGGACTTCATCTCCAGCATCGGGGTGGCT
TCCCTGCCCCGGCGGCCGGGAGGGAGCATCCCGGTGGAGGGCGGCTACGCC
CGGGGCTACTTCGCCAGCAAGTGGGCTTGCGAGCAACTGCTGCGCTCCACC
CATGACTGCACCGGTGTGCCCGTACGGGTGATTCGGCCCAGCCTCATTCTT
CCCGATCGTGTGCTGGCCGGGGAGATGAACCCGGACGATCTGCTTTCAAGA
CTGCTGTACAGCATCCTGGTGACCGGGATCGCCCCCGGGTGCTTTGGGGAG
GAGTCGCAGAACAGTGGACGATCGGGGTTCTCGGTGCAGGGCCTCCCCGTC
GACCAGTTGGCGCAGACCATCCTTGCCCTCGGGGAAGCGCGCACGGAGGGA
TTTCATGTGCTCAACCTCAACGCTGACAGTGGCAGCGGTGTTCCCCTGGAT
GCCATCCTCCAGGACATCGCCGCCAAAGGAATCAGGCTGCGACGGGTGGAG
GGCTATGACCTCTGGCTCGACGCGATCACAACCCGCCTGCGTCGCCTGCCA
GCCGAGCAACGGGCCCGTTCCCTGCTGGATGTGGCGGAAGCCTATGCAGGA
TCAGCAGGCCAGACAACGCAGAGCAGCGGTGAAATGCAGGCGGGCAGCAGC
TCCTGCCCGGAGGAGATCACCAGCCTGCAACCGGACTTCAGTAGGGCCTAC
AGGCGCAAGATCGTGGATGATCTGGCTCGGTGGGGGCTGATCGAGCCTCCA
GGACCCGTGGATCAGTGA

Amino acid sequence (SEQ ID NO:28)

>gi|254431429|ref|ZP_05045132.1| putative long-chain fatty-acid--CoA ligase [Cyanobium sp. PCC 7001]

MNESSADQSSGNVSEGWPDASVTARALQAHLRYEQIIDAILSGYAERPALA
ERSYLVRPDPSTGQTVRVHEQAFRSISYRTLQERVHALTMAWRLHPDSPVQ
AGAFVVLVGFASIDYAVLDLALAYTKGVPVPLSPNHSSEDDDAILGTVQPV

FIG. 8 (Cont.)

TLAVSISEFSGCVDLIARSTSIRTVIVFDLDPAVDCERAALESGIRALNEK
GSDVVVQTLQDLIDVGRDAEFSFLPIQAQDQDDLALLIHTSGSTGTPKGAC
ISSRALINTWRHVSGPYPKVTVVLAPFHHMMGRDSMITALGAGGTAYFTLR
PDLSTVIEDIRLARPTGLVLFPRLCEVIEHHLTTAPEYSGNEILGGRLQSI
VVASAPITPRLKASLECLLGVPVSEGYSSTETASGGLAMNGLLNRNNILAY
RLRDVPEAGYSVNDRPFPRGELCVKTRFGISGYFRNPEATAELFDDDGFYC
TGDIVEERAPDQIAIIDRRKNVIKLAQGEYVAVGRLEQLFQEGCGCVQQIH
LHGDSTRAYLLAVVVPDRNTLAPPGSRQASEAELKARVREEILTLANQREL
RGFEIPRDLILAEEPFSQQNGLLSSLGKPIRPAIRARYRSRLESLYASHEA
TRGTELEAIRASAGAVDVETTLLALLSSTLGVVCGAADRQTSFRELGGDSL
AAVQLAMEIKKQFGVGLEGSQILGPGGTVEAWARRIHTASIQQAPHQRVGS
PLAAIPAEGWLKPDHYRLENLIGIPIGTPSAEVARPTGGPPTVLLTGATGF
LGGRLCLEWLQRLAGQGGRLICLVRPSNSHSAWERLRNRFSHLEPEQVARF
RELAGRHLEVIPADIGEPGLGLEPGCQERLATEVDAICHCAAEVNHRLPYR
HLYRPNVIGTAEIIHLAITTRLKSVDFISSIGVASLPRRPGGSIPVEGGYA
RGYFASKWACEQLLRSTHDCTGVPVRVIRPSLILPDRVLAGEMNPDDLLSR
LLYSILVTGIAPGCFGEESQNSGRSGFSVQGLPVDQLAQTILALGEARTEG
FHVLNLNADSGSGVPLDAILQDIAAKGIRLRRVEGYDLWLDAITTRLRRLP
AEQRARSLLDVAEAYAGSAGQTTQSSGEMQAGSSSCPEEITSLQPDFSRAY
RRKIVDDLARWGLIEPPGPVDQ

YP 882653.1

Nucleotide sequence (SEQ ID NO:29)

>uniprot|A0QIB5|A0QIB5_MYCA1 Putative acyl-CoA dehydrogenase

ATGTCGACTGCCACCCATGACGAACGACTCGACCGTCGCGTCCACGAACTC
ATCGCCACCGACCCGCAATTCGCCGCCGCCCAACCCGACCCGGCGATCACC
GCCGCCCTCGAACAGCCCGGGCTGCGGCTGCCGCAGATCATCCGCACCGTG
CTCGACGGCTACGCCGACCGGCCGGCGCTGGGACAGCGCGTGGTGGAGTTC
GTCACGGACGCCAAGACCGGGCGCACGTCGGCGCAGCTGCTCCCCCGCTTC
GAGACCATTACGTACGGCGAAGTGGCGCAGCGTGTTTCGGCGCTGGGCCGC
GCCCTGTCTGACGACGCGGTGCACCCCGGCGACCGGGTGTGCGTGCTGGGC
TTCAACAGCGTCGACTACGCCACCATCGACATGGCGCTGGGCGCCATCGGC
GCCGTCTCGGTGCCGCTGCAGACCAGCGCGGCAATCAGCTCGCTGCAGCCG
ATCGTGGCCGAGACCGAGCCCACCCTGATCGCGTCCAGCGTGAACCAGCTG

FIG. 8 (Cont.)

TCCGACGCGGTGCAGCTGATCACCGGCGCCGAGCAGGCGCCCACCCGGCTG
GTGGTGTTCGACTACCACCCGCAGGTCGACGACCAGCGCGAGGCCGTCCAG
GACGCCGCGGCGCGGCTGTCCGGCACCGGCGTGGCCGTCCAGACGCTGGCC
GAGCTGCTGGAGCGCGGCAAGGACCTGCCCGCCGTCGCGGAGCCGCCCGCC
GACGAGGACTCGCTGGCCCTGCTGATCTACACCTCCGGGTCCACCGGCGCC
CCCAAGGGCGCGATGTACCCGCAGAGCAACGTCGGCAAGATGTGGCGCCGC
GGCAGCAAGAACTGGTTCGGCGAGAGCGCCGCGTCGATCACCCTGAATTTC
ATGCCGATGAGCCACGTGATGGGCCGAAGCATCCTCTACGGCACGCTGGGC
AACGGCGGCACCGCCTACTTCGCCGCCCGCAGCGACCTGTCCACCCTGCTC
GAGGACCTCGAGCTGGTGCGGCCCACCGAGCTCAACTTCGTCCCGCGGATC
TGGGAGACGCTGTACGGCGAATTCCAGCGTCAGGTCGAGCGGCGGCTCTCC
GAGGCCGGGGACGCCGGCGAACGTCGCGCCGTCGAGGCCGAGGTGCTGGCC
GAGCAGCGCCAGTACCTGCTGGGCGGGCGGTTCACCTTCGCGATGACGGGC
TCGGCGCCCATCTCGCCCGAGCTGCGCAACTGGGTCGAGTCGCTGCTCGAA
ATGCACCTGATGGACGGCTACGGCTCCACGGAGGCCGGAATGGTGTTGTTC
GACGGGGAGATTCAGCGCCCGCCGGTGGTCGACTACAAGCTGGTCGACGTG
CCGGACCTGGGCTACTTCAGCACCGACCGGCCGCATCCGCGCGGCGAGCTG
CTGCTGCGCACCGAGAACATGTTCCCGGGCTACTACAAGCGGGCCGAAACC
ACCGCGGGCGTCTTCGACGAGGACGGCTACTACCGCACCGGCGACGTGTTC
GCCGAGATCGCCCCGGACCGGCTGGTCTACGTCGACCGCCGCAACAACGTG
CTCAAGCTGGCGCAGGGCGAATTCGTCACGCTGGCCAAGCTGGAGGCGGTG
TTCGGCAACAGCCCGCTGATCCGCCAGATCTACGTCTACGGCAACAGCGCC
CAGCCCTACCTGCTGGCGGTCGTGGTGCCCACCGAGGAGGCGCTGGCCTCG
GGTGACCCCGAGACGCTCAAGCCCAAGATCGCCGACTCGCTGCAGCAGGTC
GCCAAGGAGGCCGGCCTGCAGTCCTACGAGGTGCCGCGCGACTTCATCATC
GAGACCACCCCGTTCAGCCTGGAAAACGGTCTGCTGACCGGGATCCGGAAG
CTGGCGTGGCCGAAACTGAAGCAGCACTACGGGGAACGGCTGGAGCAGATG
TACGCCGACCTGGCCGCCGGACAGGCCGACGAGCTGGCCGAGCTGCGCCGC
AACGGTGCCCAGGCGCCGGTGTTGCAGACCGTGAGCCGCGCCGCGGGCGCC
ATGCTGGGTTCGGCCGCCTCCGACCTGTCCCCCGACGCCCACTTCACCGAT
CTGGGCGGAGACTCGTTGTCGGCGTTGACATTCGGCAACCTGCTGCGCGAG
ATCTTCGACGTCGACGTGCCGGTGGGCGTGATCGTCAGCCCGGCCAACGAC
CTGGCGGCCATCGCGAGCTACATCGAGGCCGAGCGGCAGGGCAGCAAGCGC
CCGACGTTCGCCTCGGTGCACGGCCGGGACGCGACCGTGGTGCGCGCCGCC
GACCTGACGCTGGACAAGTTCCTCGACGCCGACACGCTGGCCTCCGCGCCG
AACCTGCCCAAGCCGGCCACCGAGGTGCGCACCGTGCTGCTGACCGGCGCC
ACCGGCTTCCTGGGCCGCTACCTGGCCCTGGAATGGCTGGAGCGGATGGAC
ATGGTGGACGGCAAGGTCATCGCCCTGGTCCGGGCCCGCTCCGACGAGGAG
GCACGCGCCCGGCTGGACAAGACCTTCGACAGCGGCGACCCGAAGCTGCTC

FIG. 8 (Cont.)

GCGCACTACCAGCAGCTGGCTGCCGATCACCTGGAGGTCATCGCCGGCGAC
AAGGGCGAGGCCAATCTGGGCCTGCGCCAAGATGTTTGGCAACGGCTGGCC
GACACGGTCGACGTGATCGTCGACCCCGCCGCGCTGGTCAACCACGTGTTG
CCGTACAGCGAGCTGTTCGGGCCCAACGCCCTGGGCACCGCGGAGCTGATC
CGGCTGGCGCTGACGTCCAAGCAGAAGCCGTACACCTACGTGTCCACCATC
GGCGTGGGCGACCAGATCGAGCCGGGCAAGTTCGTCGAGAACGCCGACATC
CGGCAGATGAGCGCCACCCGGGCGATCAACGACAGCTACGCCAACGGCTAC
GGCAACAGCAAGTGGGCCGGCGAGGTGCTGCTGCGCGAGGCGCACGACCTG
TGCGGGCTGCCCGTCGCGGTGTTCCGCTGCGACATGATCCTGGCCGACACC
ACGTATGCCGGGCAGCTCAACCTGCCGGACATGTTCACCCGGCTGATGCTG
AGCCTGGTGGCCACCGGGATCGCGCCCGGCTCGTTCTACGAGCTCGACGCC
GACGGCAACCGGCAGCGGGCGCACTACGACGGCCTGCCGGTCGAGTTCATC
GCCGCGGCGATCTCGACGCTGGGTTCGCAGATCACCGACAGCGACACCGGC
TTCCAGACCTACCACGTGATGAACCCCTACGATGACGGCATCGGTCTGGAC
GAGTACGTCGATTGGCTGGTGGACGCCGGCTATTCGATCGAGCGGATCGCC
GACTACTCCGAATGGCTGCGGCGGTTCGAGACCTCGCTGCGGGCCCTGCCG
GACCGGCAGCGCCAGTACTCGCTGCTGCCGCTGCTGCACAACTACCGCACG
CCCGAGAAGCCGATCAACGGGTCGATAGCTCCCACCGACGTGTTCCGGGCA
GCGGTGCAGGAGGCGAAAATCGGCCCCGACAAAGACATTCCGCACGTGTCG
CCGCCGGTCATCGTCAAGTACATCACCGACCTGCAGCTGCTCGGGCTGCTC
TGA

Amino acid sequence (SEQ ID NO:30)

>uniprot|A0QIB5|A0QIB5_MYCA1 Putative acyl-CoA dehydrogenase

MSTATHDERLDRRVHELIATDPQFAAAQPDPAITAALEQPGLRLPQIIRT
VLDGYADRPALGQRVVEFVTDAKTGRTSAQLLPRFETITYGEVAQRVSAL
GRALSDDAVHPGDRVCVLGFNSVDYATIDMALGAIGAVSVPLQTSAAISS
LQPIVAETEPTLIASSVNQLSDAVQLITGAEQAPTRLVVFDYHPQVDDQR
EAVQDAAARLSGTGVAVQTLAELLERGKDLPAVAEPPADEDSLALLIYTS
GSTGAPKGAMYPQSNVGKMWRRGSKNWFGESAASITLNFMPMSHVMGRSI
LYGTLGNGGTAYFAARSDLSTLLEDLELVRPTELNFVPRIWETLYGEFQR
QVERRLSEAGDAGERRAVEAEVLAEQRQYLLGGRFTFAMTGSAPISPELR
NWVESLLEMHLMDGYGSTEAGMVLFDGEIQRPPVVDYKLVDVPDLGYFST
DRPHPRGELLLRTENMFPGYYKRAETTAGVFDEDGYYRTGDVFAEIAPDR
LVYVDRRNNVLKLAQGEFVTLAKLEAVFGNSPLIRQIYVYGNSAQPYLLA
VVVPTEEALASGDPETLKPKIADSLQQVAKEAGLQSYEVPRDFIIETTPF

FIG. 8 (Cont.)

SLENGLLTGIRKLAWPKLKQHYGERLEQMYADLAAGQADELAELRRNGAQ
APVLQTVSRAAGAMLGSAASDLSPDAHFTDLGGDSLSALTFGNLLREIFD
VDVPVGVIVSPANDLAAIASYIEAERQGSKRPTFASVHGRDATVVRAADL
TLDKFLDADTLASAPNLPKPATEVRTVLLTGATGFLGRYLALEWLERMDM
VDGKVIALVRARSDEEARARLDKTFDSGDPKLLAHYQQLAADHLEVIAGD
KGEANLGLRQDVWQRLADTVDVIVDPAALVNHVLPYSELFGPNALGTAEL
IRLALTSKQKPYTYVSTIGVGDQIEPGKFVENADIRQMSATRAINDSYAN
GYGNSKWAGEVLLREAHDLCGLPVAVFRCDMILADTTYAGQLNLPDMFTR
LMLSLVATGIAPGSFYELDADGNRQRAHYDGLPVEFIAAAISTLGSQITD
SDTGFQTYHVMNPYDDGIGLDEYVDWLVDAGYSIERIADYSEWLRRFETS
LRALPDRQRQYSLLPLLHNYRTPEKPINGSIAPTDVFRAAVQEAKIGPDK
DIPHVSPPVIVKYITDLQLLGLL

YP 887275.1

Nucleotide sequence (SEQ ID NO:31)

>uniprot|A0QWI7|A0QWI7_MYCS2 NAD dependent epimerase/dehydratase family protein

ATGACGATCGAAACGCGCGAAGACCGCTTCAACCGGCGCATTGACCACTTG
TTCGAAACCGACCCGCAGTTCGCCGCCGCCCGTCCCGACGAGGCGATCAGC
GCGGCTGCCGCCGATCCGGAGTTGCGCCTTCCTGCCGCGGTCAAACAGATT
CTGGCCGGCTATGCGGACCGCCCTGCGCTGGGCAAGCGCGCCGTCGAGTTC
GTCACCGACGAAGAAGGCCGCACCACCGCGAAGCTCCTGCCCCGCTTCGAC
ACCATCACCTACCGTCAGCTCGCAGGCCGGATCCAGGCCGTGACCAATGCC
TGGCACAACCATCCGGTGAATGCCGGTGACCGCGTGGCCATCCTGGGTTTC
ACCAGTGTCGACTACACGACGATCGACATCGCCCTGCTCGAACTCGGCGCC
GTGTCCGTACCGCTGCAGACCAGTGCGCCGGTGGCCCAACTGCAGCCGATC
GTCGCCGAGACCGAGCCCAAGGTGATCGCGTCGAGCGTCGACTTCCTCGCC
GACGCAGTCGCTCTCGTCGAGTCCGGGCCCGCGCCGTCGCGACTGGTGGTG
TTCGACTACAGCCACGAGGTCGACGATCAGCGTGAGGCGTTCGAGGCGGCC
AAGGGCAAGCTCGCAGGCACCGGCGTCGTCGTCGAGACGATCACCGACGCA
CTGGACCGCGGGCGGTCACTCGCCGACGCACCGCTCTACGTGCCCGACGAG
GCCGACCCGCTGACCCTTCTCATCTACACCTCCGGCAGCACCGGCACTCCC
AAGGGCGCGATGTACCCCGAGTCCAAGACCGCCACGATGTGGCAGGCCGGG
TCCAAGGCCCGGTGGGACGAGACCCTCGGCGTGATGCCGTCGATCACCCTG
AACTTCATGCCCATGAGTCACGTCATGGGGCGCGGCATCCTGTGCAGCACA

FIG. 8 (Cont.)

CTCGCCAGCGGCGGAACCGCGTACTTCGCCGCACGCAGCGACCTGTCCACC
TTCCTGGAGGACCTCGCCCTCGTGCGGCCCACGCAGCTCAACTTCGTTCCT
CGCATCTGGGACATGCTGTTCCAGGAGTACCAGAGCCGCCTCGACAACCGC
CGCGCCGAGGGATCCGAGGACCGAGCCGAAGCCGCAGTCCTCGAAGAGGTC
CGCACCCAACTGCTCGGCGGGCGATTCGTTTCGGCCCTGACCGGATCGGCT
CCCATCTCGGCGGAGATGAAGAGCTGGGTCGAGGACCTGCTCGACATGCAT
CTGCTGGAGGGCTACGGCTCCACCGAGGCCGGCGCGGTGTTCATCGACGGG
CAGATCCAGCGCCCGCCGGTCATCGACTACAAGCTGGTCGACGTGCCCGAT
CTCGGCTACTTCGCCACGGACCGGCCCTACCCGCGCGGCGAACTTCTGGTC
AAGTCCGAGCAGATGTTCCCCGGCTACTACAAGCGTCCGGAGATCACCGCC
GAGATGTTCGACGAGGACGGGTACTACCGCACCGGCGACATCGTCGCCGAG
CTCGGGCCCGACCATCTCGAATACCTCGACCGCCGCAACAACGTGCTGAAA
CTGTCGCAGGGCGAATTCGTCACGGTCTCCAAGCTGGAGGCGGTGTTCGGC
GACAGCCCCCTGGTACGCCAGATCTACGTCTACGGCAACAGCGCGCGGTCC
TATCTGCTGGCGGTCGTGGTCCCGACCGAAGAGGCACTGTCACGTTGGGAC
GGTGACGAACTCAAGTCGCGCATCAGCGACTCACTGCAGGACGCGGCACGA
GCCGCCGGATTGCAGTCGTATGAGATCCCGCGTGACTTCCTCGTCGAGACA
ACACCTTTCACGCTGGAGAACGGCCTGCTGACCGGTATCCGCAAGCTGGCC
CGGCCGAAACTGAAGGCGCACTACGGCGAACGCCTCGAACAGCTCTACACC
GACCTGGCCGAGGGGCAGGCCAACGAGTTGCGCGAGTTGCGCCGCAACGGA
GCCGACCGGCCCGTGGTCGAGACCGTCAGCCGCGCCGCGGTCGCACTGCTC
GGTGCCTCCGTCACGGATCTGCGGTCCGATGCGCACTTCACCGATCTGGGT
GGAGATTCGTTGTCGGCCTTGAGCTTCTCGAACCTGTTGCACGAGATCTTC
GATGTCGACGTGCCGGTCGGCGTCATCGTCAGCCCGGCCACCGACCTGGCA
GGCGTCGCGGCCTACATCGAGGGCGAACTGCGCGGCTCCAAGCGCCCCACA
TACGCGTCGGTGCACGGGCGCGACGCCACCGAGGTGCGCGCGCGTGATCTC
GCCCTGGGCAAGTTCATCGACGCCAAGACCCTGTCCGCCGCGCCGGGTCTG
CCGCGTTCGGGCACCGAGATCCGCACCGTGCTGCTGACCGGCGCCACCGGG
TTCCTGGGCCGCTATCTGGCGCTGGAATGGCTGGAGCGCATGGACCTGGTG
GACGGCAAGGTGATCTGCCTGGTGCGCGCCCGCAGCGACGACGAGGCCCGG
GCGCGTCTGGACGCCACGTTCGACACCGGGGACGCGACACTGCTCGAGCAC
TACCGCGCGCTGGCAGCCGATCACCTCGAGGTGATCGCCGGTGACAAGGGC
GAGGCCGATCTGGGTCTCGACCACGACACGTGGCAGCGACTGGCCGACACC
GTCGATCTGATCGTCGATCCGGCCGCCCTGGTCAATCACGTCCTGCCGTAC
AGCCAGATGTTCGGACCCAATGCGCTCGGCACCGCCGAACTCATCCGGATC
GCGCTGACCACCACGATCAAGCCGTACGTGTACGTCTCGACGATCGGTGTG
GGACAGGGCATCTCCCCCGAGGCGTTCGTCGAGGACGCCGACATCCGCGAG
ATCAGCGCGACGCGCCGGGTCGACGACTCGTACGCCAACGGCTACGGCAAC
AGCAAGTGGGCCGGCGAGGTCCTGCTGCGGGAGGCGCACGACTGGTGTGGT

FIG. 8 (Cont.)

CTGCCGGTCTCGGTGTTCCGCTGCGACATGATCCTGGCCGACACGACCTAC
TCGGGTCAGCTGAACCTGCCGGACATGTTCACCCGCCTGATGCTGAGCCTC
GTGGCGACCGGCATCGCGCCCGGTTCGTTCTACGAACTCGATGCGGACGGC
AACCGGCAGCGCGCCCACTACGACGGGCTGCCCGTGGAGTTCATCGCCGAG
GCGATCTCCACCATCGGCTCGCAGGTCACCGACGGATTCGAGACGTTCCAC
GTGATGAACCCGTACGACGACGGCATCGGCCTCGACGAGTACGTGGACTGG
CTGATCGAGGCCGGCTACCCCGTGCACCGCGTCGACGACTACGCCACCTGG
CTGAGCCGGTTCGAAACCGCACTGCGGGCCCTGCCGGAACGGCAACGTCAG
GCCTCGCTGCTGCCGCTGCTGCACAACTATCAGCAGCCCTCACCGCCCGTG
TGCGGTGCCATGGCACCCACCGACCGGTTCCGTGCCGCGGTGCAGGACGCG
AAGATCGGCCCCGACAAGGACATTCCGCACGTCACGGCCGACGTGATCGTC
AAGTACATCAGCAACCTGCAGATGCTCGGATTGCTGTAA

Amino acid sequence (SEQ ID NO:32)

>uniprot|A0QWI7|A0QWI7_MYCS2 NAD dependent epimerase/dehydratase family protein

MTIETREDRFNRRIDHLFETDPQFAAARPDEAISAAAADPELRLPAAVKQ
ILAGYADRPALGKRAVEFVTDEEGRTTAKLLPRFDTITYRQLAGRIQAVT
NAWHNHPVNAGDRVAILGFTSVDYTTIDIALLELGAVSVPLQTSAPVAQL
QPIVAETEPKVIASSVDFLADAVALVESGPAPSRLVVFDYSHEVDDQREA
FEAAKGKLAGTGVVVETITDALDRGRSLADAPLYVPDEADPLTLLIYTSG
STGTPKGAMYPESKTATMWQAGSKARWDETLGVMPSITLNFMPMSHVMGR
GILCSTLASGGTAYFAARSDLSTFLEDLALVRPTQLNFVPRIWDMLFQEY
QSRLDNRRAEGSEDRAEAAVLEEVRTQLLGGRFVSALTGSAPISAEMKSW
VEDLLDMHLLEGYGSTEAGAVFIDGQIQRPPVIDYKLVDVPDLGYFATDR
PYPRGELLVKSEQMFPGYYKRPEITAEMFDEDGYYRTGDIVAELGPDHLE
YLDRRNNVLKLSQGEFVTVSKLEAVFGDSPLVRQIYVYGNSARSYLLAVV
VPTEEALSRWDGDELKSRISDSLQDAARAAGLQSYEIPRDFLVETTPFTL
ENGLLTGIRKLARPKLKAHYGERLEQLYTDLAEGQANELRELRRNGADRP
VVETVSRAAVALLGASVTDLRSDAHFTDLGGDSLSALSFSNLLHEIFDVD
VPVGVIVSPATDLAGVAAYIEGELRGSKRPTYASVHGRDATEVRARDLAL
GKFIDAKTLSAAPGLPRSGTEIRTVLLTGATGFLGRYLALEWLERMDLVD
GKVICLVRARSDDEARARLDATFDTGDATLLEHYRALAADHLEVIAGDKG
EADLGLDHDTWQRLADTVDLIVDPAALVNHVLPYSQMFGPNALGTAELIR
IALTTTIKPYVYVSTIGVGQGISPEAFVEDADIREISATRRVDDSYANGY
GNSKWAGEVLLREAHDWCGLPVSVFRCDMILADTTYSGQLNLPDMFTRLM
LSLVATGIAPGSFYELDADGNRQRAHYDGLPVEFIAEAISTIGSQVTDGF
ETFHVMNPYDDGIGLDEYVDWLIEAGYPVHRVDDYATWLSRFETALRALP

FIG. 8 (Cont.)

ERQRQASLLPLLHNYQQPSPPVCGAMAPTDRFRAAVQDAKIGPDKDIPHV
TADVIVKYISNLQMLGLL

ZP_05224908

Nucleotide sequence (SEQ ID NO:33)

>gi|163719654:2489-6013 Mycobacterium intracellulare ATCC 13950 NZ_ABIN01000072, whole genome shotgun sequence

ATGTCGACTGCCATTCATGACGAACACCTCGACCGTCGCATCGAGGAACTC
ATCGCCAACGACCCCCAATTCGCCGCCGCCCGACCGGACCCGGCCATCACC
GCCGCCACCGAAGCGCCCGGGCTGCGGCTGCCGCAGATCATCCGGACCGTG
CTCGACGGCTACGCCGACCGGCCTGCCCTGGCGCAGCGCGTCGTGGAGTTC
GTCACCGACGCCAAGACCGGGCGGACGACGGCCGAGCTGCTCCCCCGTTTC
GAGACCATCACGTATGGCGAACTCGGCGAACGGGTTTCGGCCCTCGGCCGT
GCCTGGGCCGGCGACGCGGTGCGCCCCGGCGACCGCGTCTGCGTGCTCGGC
TTCAACAGCGTTGACTACGCCACCATCGACATCGCGCTGGGCACCATCGGG
GCCGTGTCGGTGCCGCTGCAGACCAGCGCGGCGATCTCCTCGTTGCAGCCG
ATCGTCGCCGAGACCGAGCCCAGCCTGATCGCCTCGAGCGTCAACCAGCTG
CCCGACGCGGTGGAGCTGATCCTGGCCGGCGACCACGTGCCCGGCAAGCTC
GTCGTGTTCGACTACCAGCCCCAGGTCGACGACCAGCGCGAGGCCGTGGAG
GCCGCCGCCGCGCGGTTGGCCGACTCCGGCGTCGCGGTCGAGGCTCTCGCC
GACGTGCTGCGGCGCGGCAAGGACCTGCCGGCCGTCGAGCCGCCGGCGAGC
GACGAGGACTCGCTGGCCCTGCTGATCTACACCTCCGGCAGCACCGGCGCG
CCCAAGGGCGCGATGTACCCGCAGAGCAACGTCGGCAAGATGTGGCGGCGC
GGGAGCAAGAACTGGTTCGGGGAAAGCGCCGCGTCGATCACCCTCAACTTC
ATGCCGATGAGCCACGTCATGGGGCGCGGAATCCTCTACGGCACGCTGGGC
AACGGCGGCACCGCGTACTTCGCCGCCCGCAGCGACCTGTCCACCCTGCTC
GAGGACCTCGAGTTGGTGCGGCCCACCGAGATGAACTTCGTCCCCCGCATC
TGGGAGACGCTGTACGGCGAATTCCAGCGCCAGGTCGAGCGGCGGCTGGCC
GACGGCGATGCGGGCCCGGAGGCCCGCGAGACTGTGGCGGCTGCGGTGTTG
GAAGAACAGCGCCAGTACCTGCTGGGCGGGCGGTTCATCTTCGCGATGACG
GGCTCGGCACCCACCTCGCCGGAGCTCAAGGCGTGGGCCGAGTCGCTCCTG
CAGATGCACCTGATGGACGGCTACGGCTCCACCGAGGCCGGAATGGTGTTG
TTCGACGGGGAGATTCAGCGTCCGCCGGTTATTGATTACAAGCTGGTCGAC
GTTCCGGATCTGGGCTATTTCAGCACCGACCGTCCGCATCCGCGCGGTGAG
TTGTTGCTGCGGACCGAGAACATGTTCCCGGGTTATTACAAGCGGGCCGAG

FIG. 8 (Cont.)

ACCACCGCGAACGTGTTCGACGAGGACGGTTATTACCGCACCGGTGACGTG
TTCGCCGAGATCGCGCCGGACCGGCTGGTGTATGTCGATCGCCGCAACAAC
GTGCTCAAGTTGGCCCAGGGCGAGTTCGTGACGCTGGCCAAGCTGGAGGCG
GTGTTCGGCAACAGCCCGCTGATCCGCCAGATCTACGTTTACGGCAACAGC
TCCCAGCCCTACCTGCTGGCCGTGGTGGTGCCGACCGAGGAAGCGTTGGCG
GACAACGATCTTGAGTCGCTCAAGCCGAAGATCGCCGACTCGCTGCAGAAG
GTCGCCAAGGAGACCGGCCTGCAGTCCTACGAGGTGCCGCGCGACTTCATC
ATCGAGACCACGCCGTTCACCCTGGAAAACGGCCTGCTGACCGGGATCCGC
AAGCTGGCGTGGCCCAAGCTCAAGGCGCACTACGGGGATCGGCTCGAGCAG
ATGTATGCCGAGCTGGCCGCGGGACAGGCCAACGAGTTGGCCGAACTGCGC
CGCAGCGGCGCGGCGGCGCCGGTGGCCCAGACCGTGAGCCGGGCCGCGGCC
GCCCTGCTGGGTGCGACGGCCGGGGATCTGTCCGCAGATGCCCACTTCACC
GATCTTGGTGGAGACTCGTTGTCGGCGTTGACCTTCGGCAACCTGCTGCGC
GAGATCTTCGATGTCGACGTGCCGGTGGGGGTGATCGTCAGCCCGGCCAAC
GACCTGGCGGGGATCGCCGCCTACATCGAGGCCGAGCGGCAGGGCTCCAAG
CGCCCGACGTTCGCCGCCGTGCACGGTCGCGGTGCGACCATGGTGCACGCC
AGTGACCTCACGCTGGACAAGTTCCTCGACGAGGCGACCCTGGCCGCCGCG
CCCAGCCTGCCCAAGCCGGCCACCGAGGTGCGCACCGTGCTGTTGACCGGC
GCGACCGGCTTTTTGGGCCGCTACCTGGCGCTGGACTGGCTCGAGCGGATG
GACATGGTCGACGGCAAGGTCATCGCCCTGGTGCGGGCCCGCACCGATGAG
GAGGCGCGCGCCCGGCTGGACAAGACCTTCGACAGCGGCGACCCCAAACTG
CTGGCGCACTACCAGCGGCTGGCCGCCGACCACCTCGAGGTCATCGCCGGC
GACAAGGGTGAGGCCAACCTCGGCCTGGACCCCCAGACCTGGCAGCGACTG
GCCGAGGAGGTCGACGTCATCGTCGACCCCGCCGCGCTGGTCAACCACGTG
CTGCCCTACAGCGAGCTGTTCGGCCCCAACGCCCTGGGCACCGCGGAGCTG
ATCCGGATCGCGCTGACCTCCAGGCAAAAGCCCTACACCTACGTGTCGACG
ATCGGGGTGGGCGATCAGATCCAGCCAGGTGAGTTCGTCGAGAACGCCGAC
ATCCGCCAGATCAGCGCCACCCGCGAGATCAACGACGGCTACGCCAACGGC
TACGGCAACAGCAAGTGGGCCGGCGAGGTGTTGCTGCGCGAGGCCCACGAC
CTGTGCGGCCTGCCCGTCACGGTGTTCCGCTGCGACATGATCCTGGCCGAC
ACCACCTATGCCGGGCAGCTCAACCTGCCCGACATGTTCACCCGGCTGATG
CTGAGCCTGGTCGCCACCGGTATCGCGCCCGGGTCGTTCTACGAACTGGAC
GCCGACGGCAACCGCCAGCGGGCACACTACGACGGTTTGCCGGTCGAGTTC
ATCGCCGCGGCGATCTCGACGCTGGGGACCCAAATCACCGACAGCGACACG
GGCTTTCAGACCTACCACGTGATGAACCCCTACGACGACGGCATCGGGCTG
GATGAGTACATCGATTGGCTGATCGAGGCCGGGTATTCGATCGAGCGGATC
GCCGATTACTCCGAGTGGCTGCGGCGCTTCGAGACCTCGCTGCGGGCCCTG
CCCGATCGGCAGCGTCAGTACTCGCTGCTGCCGCTGCTGCACAACTACCAG
AAGCCGGAAAAGCCGATCAACGGCTCGATGGCGCCCACCGACGTGTTCCGT

FIG. 8 (Cont.)

GCCGCGGTGCAGGAAGCGAAAATCGGCCCCGACAAAGACATCCCGCACGTC
TCGGCGCCGGTGATCGTCAAGTACATCACCGACCTGGAGTTGCTCGGACTC
CTCTGA

Amino acid sequence (SEQ ID NO:34)

>gi|254819907|ref|ZP_05224908.1| FadD9
[Mycobacterium intracellulare ATCC 13950]

MSTAIHDEHLDRRIEELIANDPQFAAARPDPAITAATEAPGLRLPQIIRTV
LDGYADRPALAQRVVEFVTDAKTGRTTAELLPRFETITYGELGERVSALGR
AWAGDAVRPGDRVCVLGFNSVDYATIDIALGTIGAVSVPLQTSAAISSLQP
IVAETEPSLIASSVNQLPDAVELILAGDHVPGKLVVFDYQPQVDDQREAVE
AAAARLADSGVAVEALADVLRRGKDLPAVEPPASDEDSLALLIYTSGSTGA
PKGAMYPQSNVGKMWRRGSKNWFGESAASITLNFMPMSHVMGRGILYGTLG
NGGTAYFAARSDLSTLLEDLELVRPTEMNFVPRIWETLYGEFQRQVERRLA
DGDAGPEARETVAAAVLEEQRQYLLGGRFIFAMTGSAPTSPELKAWAESLL
QMHLMDGYGSTEAGMVLFDGEIQRPPVIDYKLVDVPDLGYFSTDRPHPRGE
LLLRTENMFPGYYKRAETTANVFDEDGYYRTGDVFAEIAPDRLVYVDRRNN
VLKLAQGEFVTLAKLEAVFGNSPLIRQIYVYGNSSQPYLLAVVVPTEEALA
DNDLESLKPKIADSLQKVAKETGLQSYEVPRDFIIETTPFTLENGLLTGIR
KLAWPKLKAHYGDRLEQMYAELAAGQANELAELRRSGAAAPVAQTVSRAAA
ALLGATAGDLSADAHFTDLGGDSLSALTFGNLLREIFDVDVPVGVIVSPAN
DLAGIAAYIEAERQGSKRPTFAAVHGRGATMVHASDLTLDKFLDEATLAAA
PSLPKPATEVRTVLLTGATGFLGRYLALDWLERMDMVDGKVIALVRARTDE
EARARLDKTFDSGDPKLLAHYQRLAADHLEVIAGDKGEANLGLDPQTWQRL
AEEVDVIVDPAALVNHVLPYSELFGPNALGTAELIRIALTSRQKPYTYVST
IGVGDQIQPGEFVENADIRQISATREINDGYANGYGNSKWAGEVLLREAHD
LCGLPVTVFRCDMILADTTYAGQLNLPDMFTRLMLSLVATGIAPGSFYELD
ADGNRQRAHYDGLPVEFIAAAISTLGTQITDSDTGFQTYHVMNPYDDGIGL
DEYIDWLIEAGYSIERIADYSEWLRRFETSLRALPDRQRQYSLLPLLHNYQ
KPEKPINGSMAPTDVFRAAVQEAKIGPDKDIPHVSAPVIVKYITDLELLGL
L

YP 889972.1

Nucleotide sequence (SEQ ID NO:35)

>uniprot|A0R484|A0R484_MYCS2 Putative long-chain
fatty-acid--CoA ligase

FIG. 8 (Cont.)

ATGACCAGCGATGTTCACGACGCCACAGACGGCGTCACCGAAACCGCACTC
GACGACGAGCAGTCGACCCGCCGCATCGCCGAGCTGTACGCCACCGATCCC
GAGTTCGCCGCCGCCGCACCGTTGCCCGCCGTGGTCGACGCGGCGCACAAA
CCCGGGCTGCGGCTGGCAGAGATCCTGCAGACCCTGTTCACCGGCTACGGT
GACCGCCCGGCGCTGGGATACCGCGCCCGTGAACTGGCCACCGACGAGGGC
GGGCGCACCGTGACGCGTCTGCTGCCGCGGTTCGACACCCTCACCTACGCC
CAGGTGTGGTCGCGCGTGCAAGCGGTCGCCGCGGCCCTGCGCCACAACTTC
GCGCAGCCGATCTACCCCGGCGACGCCGTCGCGACGATCGGTTTCGCGAGT
CCCGATTACCTGACGCTGGATCTCGTATGCGCCTACCTGGGCCTCGTGAGT
GTTCCGCTGCAGCACAACGCACCGGTCAGCCGGCTCGCCCCGATCCTGGCC
GAGGTCGAACCGCGGATCCTCACCGTGAGCGCCGAATACCTCGACCTCGCA
GTCGAATCCGTGCGGGACGTCAACTCGGTGTCGCAGCTCGTGGTGTTCGAC
CATCACCCCGAGGTCGACGACCACCGCGACGCACTGGCCCGCGCGCGTGAA
CAACTCGCCGGCAAGGGCATCGCCGTCACCACCCTGGACGCGATCGCCGAC
GAGGGCGCCGGGCTGCCGGCCGAACCGATCTACACCGCCGACCATGATCAG
CGCCTCGCGATGATCCTGTACACCTCGGGTTCCACCGGCGCACCCAAGGGT
GCGATGTACACCGAGGCGATGGTGGCGCGGCTGTGGACCATGTCGTTCATC
ACGGGTGACCCCACGCCGGTCATCAACGTCAACTTCATGCCGCTCAACCAC
CTGGGCGGGCGCATCCCCATTTCCACCGCCGTGCAGAACGGTGGAACCAGT
TACTTCGTACCGGAATCCGACATGTCCACGCTGTTCGAGGATCTCGCGCTG
GTGCGCCCGACCGAACTCGGCCTGGTTCCGCGCGTCGCCGACATGCTCTAC
CAGCACCACCTCGCCACCGTCGACCGCCTGGTCACGCAGGGCGCCGACGAA
CTGACCGCCGAGAAGCAGGCCGGTGCCGAACTGCGTGAGCAGGTGCTCGGC
GGACGCGTGATCACCGGATTCGTCAGCACCGCACCGCTGGCCGCGGAGATG
AGGGCGTTCCTCGACATCACCCTGGGCGCACACATCGTCGACGGCTACGGG
CTCACCGAGACCGGCGCCGTGACACGCGACGGTGTGATCGTGCGGCCACCG
GTGATCGACTACAAGCTGATCGACGTTCCCGAACTCGGCTACTTCAGCACC
GACAAGCCCTACCCGCGTGGCGAACTGCTGGTCAGGTCGCAAACGCTGACT
CCCGGGTACTACAAGCGCCCCGAGGTCACCGCGAGCGTCTTCGACCGGGAC
GGCTACTACCACACCGGCGACGTCATGGCCGAGACCGCACCCGACCACCTG
GTGTACGTGGACCGTCGCAACAACGTCCTCAAACTCGCGCAGGGCGAGTTC
GTGGCGGTCGCCAACCTGGAGGCGGTGTTCTCCGGCGCGGCGCTGGTGCGC
CAGATCTTCGTGTACGGCAACAGCGAGCGCAGTTTCCTTCTGGCCGTGGTG
GTCCCGACGCCGGAGGCGCTCGAGCAGTACGATCCGGCCGCGCTCAAGGCC
GCGCTGGCCGACTCGCTGCAGCGCACCGCACGCGACGCCGAACTGCAATCC
TACGAGGTGCCGGCCGATTTCATCGTCGAGACCGAGCCGTTCAGCGCCGCC
AACGGGCTGCTGTCGGGTGTCGGAAAACTGCTGCGGCCCAACCTCAAAGAC
CGCTACGGGCAGCGCCTGGAGCAGATGTACGCCGATATCGCGGCCACGCAG
GCCAACCAGTTGCGCGAACTGCGGCGCGCGGCCGCCACACAACCGGTGATC

FIG. 8 (Cont.)

GACACCCTCACCCAGGCCGCTGCCACGATCCTCGGCACCGGGAGCGAGGTG
GCATCCGACGCCCACTTCACCGACCTGGGCGGGGATTCCCTGTCGGCGCTG
ACACTTTCGAACCTGCTGAGCGATTTCTTCGGTTTCGAAGTTCCCGTCGGC
ACCATCGTGAACCCGGCCACCAACCTCGCCCAACTCGCCCAGCACATCGAG
GCGCAGCGCACCGCGGGTGACCGCAGGCCGAGTTTCACCACCGTGCACGGC
GCGGACGCCACCGAGATCCGGGCGAGTGAGCTGACCCTGGACAAGTTCATC
GACGCCGAAACGCTCCGGGCCGCACCGGGTCTGCCCAAGGTCACCACCGAG
CCACGGACGGTGTTGCTCTCGGGCGCCAACGGCTGGCTGGGCCGGTTCCTC
ACGTTGCAGTGGCTGGAACGCCTGGCACCTGTCGGCGGCACCCTCATCACG
ATCGTGCGGGGCCGCGACGACGCCGCGGCCCGCGCACGGCTGACCCAGGCC
TACGACACCGATCCCGAGTTGTCCCGCCGCTTCGCCGAGCTGGCCGACCGC
CACCTGCGGGTGGTCGCCGGTGACATCGGCGACCCGAATCTGGGCCTCACA
CCCGAGATCTGGCACCGGCTCGCCGCCGAGGTCGACCTGGTGGTGCATCCG
GCAGCGCTGGTCAACCACGTGCTCCCCTACCGGCAGCTGTTCGGCCCCAAC
GTCGTGGGCACGGCCGAGGTGATCAAGCTGGCCCTCACCGAACGGATCAAG
CCCGTCACGTACCTGTCCACCGTGTCGGTGGCCATGGGGATCCCCGACTTC
GAGGAGGACGGCGACATCCGGACCGTGAGCCCGGTGCGCCCGCTCGACGGC
GGATACGCCAACGGCTACGGCAACAGCAAGTGGGCCGGCGAGGTGCTGCTG
CGGGAGGCCCACGATCTGTGCGGGCTGCCCGTGGCGACGTTCCGCTCGGAC
ATGATCCTGGCGCATCCGCGCTACCGCGGTCAGGTCAACGTGCCAGACATG
TTCACGCGACTCCTGTTGAGCCTCTTGATCACCGGCGTCGCGCCGCGGTCG
TTCTACATCGGAGACGGTGAGCGCCCGCGGGCGCACTACCCCGGCCTGACG
GTCGATTTCGTGGCCGAGGCGGTCACGACGCTCGGCGCGCAGCAGCGCGAG
GGATACGTGTCCTACGACGTGATGAACCCGCACGACGACGGGATCTCCCTG
GATGTGTTCGTGGACTGGCTGATCCGGGCGGGCCATCCGATCGACCGGGTC
GACGACTACGACGACTGGGTGCGTCGGTTCGAGACCGCGTTGACCGCGCTT
CCCGAGAAGCGCCGCGCACAGACCGTACTGCCGCTGCTGCACGCGTTCCGC
GCTCCGCAGGCACCGTTGCGCGGCGCACCCGAACCCACGGAGGTGTTCCAC
GCCGCGGTGCGCACCGCGAAGGTGGGCCCGGGAGACATCCCGCACCTCGAC
GAGGCGCTGATCGACAAGTACATACGCGATCTGCGTGAGTTCGGTCTGATC
TGA

Amino acid sequence (SEQ ID NO:36)

>uniprot|A0R484|A0R484_MYCS2 Putative long-chain fatty-acid--CoA ligase

MTSDVHDATDGVTETALDDEQSTRRIAELYATDPEFAAAAPLPAVVDAAH
KPGLRLAEILQTLFTGYGDRPALGYRARELATDEGGRTVTRLLPRFDTLT
YAQVWSRVQAVAAALRHNFAQPIYPGDAVATIGFASPDYLTLDLVCAYLG

FIG. 8 (Cont.)

LVSVPLQHNAPVSRLAPILAEVEPRILTVSAEYLDLAVESVRDVNSVSQL
VVFDHHPEVDDHRDALARAREQLAGKGIAVTTLDAIADEGAGLPAEPIYT
ADHDQRLAMILYTSGSTGAPKGAMYTEAMVARLWTMSFITGDPTPVINVN
FMPLNHLGGRIPISTAVQNGGTSYFVPESDMSTLFEDLALVRPTELGLVP
RVADMLYQHHLATVDRLVTQGADELTAEKQAGAELREQVLGGRVITGFVS
TAPLAAEMRAFLDITLGAHIVDGYGLTETGAVTRDGVIVRPPVIDYKLID
VPELGYFSTDKPYPRGELLVRSQTLTPGYYKRPEVTASVFDRDGYYHTGD
VMAETAPDHLVYVDRRNNVLKLAQGEFVAVANLEAVFSGAALVRQIFVYG
NSERSFLLAVVVPTPEALEQYDPAALKAALADSLQRTARDAELQSYEVPA
DFIVETEPFSAANGLLSGVGKLLRPNLKDRYGQRLEQMYADIAATQANQL
RELRRAAATQPVIDTLTQAAATILGTGSEVASDAHFTDLGGDSLSALTLS
NLLSDFFGFEVPVGTIVNPATNLAQLAQHIEAQRTAGDRRPSFTTVHGAD
ATEIRASELTLDKFIDAETLRAAPGLPKVTTEPRTVLLSGANGWLGRFLT
LQWLERLAPVGGTLITIVRGRDDAAARARLTQAYDTDPELSRRFAELADR
HLRVVAGDIGDPNLGLTPEIWHRLAAEVDLVVHPAALVNHVLPYRQLFGP
NVVGTAEVIKLALTERIKPVTYLSTVSVAMGIPDFEEDGDIRTVSPVRPL
DGGYANGYGNSKWAGEVLLREAHDLCGLPVATFRSDMILAHPRYRGQVNV
PDMFTRLLLSLLITGVAPRSFYIGDGERPRAHYPGLTVDFVAEAVTTLGA
QQREGYVSYDVMNPHDDGISLDVFVDWLIRAGHPIDRVDDYDDWVRRFET
ALTALPEKRRAQTVLPLLHAFRAPQAPLRGAPEPTEVFHAAVRTAKVGPG
DIPHLDEALIDKYIRDLREFGLI

ZP_04751860

Nucleotide sequence (SEQ ID NO:37)

>gi|218125542:1370-4894 Mycobacterium kansasii ATCC 12478 NZ_ACBV01000156, whole genome shotgun sequence

Atgtcgactaccactcgtgacgagcgcctcgagcgccgcatcgacaccctg
atccacgacgacgcacagttcgccgccgccaagccggacccggcaatcgcc
gccgccctggaaaagcccggcctgagcctgccagagatcatccagaccgcg
ctgcagggctacgccgaccggccggctctcgggcagcgcgccgtcgagttc
gtcacggacacccagaccggccgcacctcggtgcggctgctgacccgcttc
gagaccatcacctaccgtcagctcggcgaccgggtcggtgcgctggcgcgc
gccctgacccacgactcggtgcacgccggcgaccgggtctgcgtgctgggc
ttcaacagcctcgactacaccaccatcgacatggcgctggcgaaggtcggc

FIG. 8 (Cont.)

gcggtgtcggtgccgctgcagaccagtgcggcggttacccagctgcagccg
atcgtggccgagaccgagcccaccatgatggcggccagtgtgaatcagctg
tcggatgccgtggacgtgctgttgagcggccacctgccggccaagctggtg
gtgttcgactaccaccccgaggtggacgaccagcgcgaggcgctcgacacg
gcccgggagcggttggcggacacagcggtggttgtccagaccctcaaggac
gtcctggaccacggcgcaacgctggcggcggggtcggttgccgagccgctc
gcggcgtcgggcgacaacgactcgctggcgctgctgatctacacctccggc
agcaccggcgcacccaaaggtgcgatgtatcgacagagcaacgtcggcaag
atgtggcgccggtcgagcaagaactggttcgggcccaccgcggcgtcgatc
accctcaacttcatgccgatgagccacatcatgggacgtggagtcctctac
ggcacgctcggcaacggcggcaccgcgtacttcgcggccaggagcgacctc
tcgacgctgctggaagacttgcggctggtgcggcccaccgagttgaacttc
gtgccgcggatctgggagaccttgtacggcgaataccagcgcgcggtcgac
cagcggtcggtcgatcccggtgaacccgccgcccgcgaagccgtcgaagcc
caggtcatggccgagcagcgccaggacctgctgggtgggcgctatatcttc
gcgatgacgggctcggcgcccatgtccccggagctgcggaactgggtggaa
gcgctgctggagatcccactgctcgacggttacggctccaccgaggccggg
atggtcatgttcgacggtgaaattcagcgcccgccggtgatcgactacaag
ctggtcgacgtgcccgatctgggctacttcagcaccgaccagccgtatccg
cggggcgagttgctgctcaagaccgagaacatgttcccaggctattacaag
cggccggaggttacggccagcgtgttcgacgcggacggctactaccggacc
ggagacgtcgtggccgaggtcgctccggaccggctggtgtacgtggaccgc
cgcaacaacgtgctgaagcttgcccagggcgagttcgtgaccgtcgccaag
ctggaggcggtgttcggcaacagcccgctggtgcgccagatctacgtctac
ggcaacagcgcgcatccctacctgttggccgtggtggtgccaacggaagaa
gcgtcggcgggcactgacatagcggccttgaagccgctgatcgccgactcg
ctgcagaccgtcgccaaagaggccggcctgcagtcctacgaggtgccgcgc
gacttcctcatcgagacaacaccattcacgctggagaacggtctgctcacc
ggtatccgcaagctggcgtggccgaagctgaggcagcattacggcgaacgg
ctggagcagctctacacggagctggccgcgagccaggccaacgagttgagc
gagctgcggcgcagcggggcccatgcgccggtgctggaaacggtgagccgg
gcggcgggcgcgctgctgggggcggcgagcaccgccttgtcacccgacgcg
cacttcaccgatctgggtggagactcgttgtcggcgttgacattcggcaac
ctgctgcgggagatcttcgacgttgacgtaccggtgggtgtgatcgtcagc
ccggccagcgacctggcggcgatcgccgcttacatcgagggcgagcggcag
ggcagcaagcgacccacgttcgccgtgattcacggtcgggacgcgctagag
gtgcatgcgagtgacctcaccctggacaagttcatcgacgcatccaccctg
gcagccgcaccggtactgccgcctccgagcgccgcggtgcgcaccgtcctg
ttgaccggcgcgaccggctttttgggccgctacctggcgctggactggctg

FIG. 8 (Cont.)

gagcgcatggacctggtcgacggcaaggtgatcgccctggtgcgggcgaag
tcggacgacgatgcccgggcacgcctggacaagacgttcgacagcggagac
cccgagctgctgacccactaccggcggctggcgaccgaccacttggaggtc
atcgccggcgacaagggcgaggccaacctcgggctggatcagctgacctgg
cagcggctggccgacaccgtcgacctgatcgtcgacccggccgcgttggtc
aaccacgtgctgccctacagcgagctgttcggccccaacgcgctgggcacc
gccgaattgatccggatagcgctgaccggcaagctcaagccctacacctac
gtctcgaccatcggggtgggcgaccagatcgagccgggcaagttcaccgag
gacgccgacatccggcacatcagcgcgacccggaagatcaacgacagttac
gccaacggctacggcaacagcaagtgggccggcgaggtgctactgcgcgag
gcccacgacctgtgcgggctgccggtcgcggtgttccgctgcgacatgatc
ctggccgacaccacgtgggcgggtcagctcaacgtgccggacatgttcacc
cggatgatgctgagcctggtggccaccggcattgcacccggttcgttctac
gagctggacgccgacggcaaccggcagcgtgcccactacgacggcttgccg
gtcgagttcatcgccgaggcgatcgcgacgctgggcgcccgggacgggaag
ggtttccagacctaccacgtgatgaacccctacgacgacggcatcgggatg
gaccggttcgtcgactggctcgtcgacgccggatgcgccatccaccgcatc
gacgactacggcgactggctgcgacgattcgagaccgcgctgcgcggcctg
cccgaaaagcagcgtcacgcgtcactactgccgttgctgcacaactaccag
aagccggcgccgccgctgcgcgggtcgatggctccgaccgaccggttccgg
gcggccgtgcaggacgcgaaagtgggcccggacaaggacatcccgcacatc
tcgccgcagatcatcgcgaagtacctcagtgatctgcgcttgctcgggctc
ctctga

Amino acid sequence (SEQ ID NO:38)

>gi|240173202|ref|ZP_04751860.1| FadD9
[Mycobacterium kansasii ATCC 12478]
MSTTTRDERLERRIDTLIHDDAQFAAAKPDPAIAAALEKPGLSLPEIIQTA
LQGYADRPALGQRAVEFVTDTQTGRTSVRLLTRFETITYRQLGDRVGALAR
ALTHDSVHAGDRVCVLGFNSLDYTTIDMALAKVGAVSVPLQTSAAVTQLQP
IVAETEPTMMAASVNQLSDAVDVLLSGHLPAKLVVFDYHPEVDDQREALDT
ARERLADTAVVVQTLKDVLDHGATLAAGSVAEPLAASGDNDSLALLIYTSG
STGAPKGAMYRQSNVGKMWRRSSKNWFGPTAASITLNFMPMSHIMGRGVLY
GTLGNGGTAYFAARSDLSTLLEDLRLVRPTELNFVPRIWETLYGEYQRAVD
QRSVDPGEPAAREAVEAQVMAEQRQDLLGGRYIFAMTGSAPMSPELRNWVE
ALLEIPLLDGYGSTEAGMVMFDGEIQRPPVIDYKLVDVPDLGYFSTDQPYP
RGELLLKTENMFPGYYKRPEVTASVFDADGYYRTGDVVAEVAPDRLVYVDR
RNNVLKLAQGEFVTVAKLEAVFGNSPLVRQIYVYGNSAHPYLLAVVVPTEE
ASAGTDIAALKPLIADSLQTVAKEAGLQSYEVPRDFLIETTPFTLENGLLT

FIG. 8 (Cont.)

GIRKLAWPKLRQHYGERLEQLYTELAASQANELSELRRSGAHAPVLETVSR
AAGALLGAASTALSPDAHFTDLGGDSLSALTFGNLLREIFDVDVPVGVIVS
PASDLAAIAAYIEGERQGSKRPTFAVIHGRDALEVHASDLTLDKFIDASTL
AAAPVLPPPSAAVRTVLLTGATGFLGRYLALDWLERMDLVDGKVIALVRAK
SDDDARARLDKTFDSGDPELLTHYRRLATDHLEVIAGDKGEANLGLDQLTW
QRLADTVDLIVDPAALVNHVLPYSELFGPNALGTAELIRIALTGKLKPYTY
VSTIGVGDQIEPGKFTEDADIRHISATRKINDSYANGYGNSKWAGEVLLRE
AHDLCGLPVAVFRCDMILADTTWAGQLNVPDMFTRMMLSLVATGIAPGSFY
ELDADGNRQRAHYDGLPVEFIAEAIATLGARDGKGFQTYHVMNPYDDGIGM
DRFVDWLVDAGCAIHRIDDYGDWLRRFETALRGLPEKQRHASLLPLLHNYQ
KPAPPLRGSMAPTDRFRAAVQDAKVGPDKDIPHISPQIIAKYLSDLRLLGL
L

YP 978699.1

Nucleotide sequence (SEQ ID NO:39)

>uniprot|A1KLT8|A1KLT8_MYCBP Probable fatty-acid-CoA ligase fadD9

ATGTCGATCAACGATCAGCGACTGACACGCCGCGTCGAGGACCTATACGCC
AGCGACGCCCAGTTCGCCGCCGCCAGTCCCAACGAGGCGATCACCCAGGCG
ATCGACCAGCCCGGGGTCGCGCTTCCACAGCTCATCCGTATGGTCATGGAG
GGCTACGCCGATCGGCCGGCACTCGGCCAGCGTGCGCTCCGCTTCGTCACC
GACCCCGACAGCGGCCGCACCATGGTCGAGCTACTGCCGCGGTTCGAGACC
ATCACCTACCGCGAACTGTGGGCCCGCGCCGGCACATTGGCCACCGCGTTG
AGCGCTGAGCCCGCGATCCGGCCGGGCGACCGGGTTTGCGTGCTGGGCTTC
AACAGCGTCGACTACACAACCATCGACATCGCGCTGATCCGGTTGGGCGCC
GTGTCGGTTCCACTGCAGACCAGTGCGCCGGTCACCGGGTTGCGCCCGATC
GTCACCGAGACCGAGCCGACGATGATCGCCACCAGCATCGACAATCTTGGC
GACGCCGTCGAAGTGCTGGCCGGTCACGCCCCGGCCCGGCTGGTCGTATTC
GATTACCACGGCAAGGTTGACACCCACCGCGAGGCCGTCGAAGCCGCCCGA
GCTCGGTTGGCCGGCTCGGTGACCATCGACACACTTGCCGAACTGATCGAA
CGCGGCAGGGCGCTGCCGGCCACACCCATTGCCGACAGCGCCGACGACGCG
CTGGCGCTGCTGATTTACACCTCGGGTAGTACCGGCGCACCCAAAGGCGCC
ATGTATCGCGAGAGCCAGGTGATGAGCTTCTGGCGCAAGTCGAGTGGCTGG
TTCGAGCCGAGCGGTTACCCCTCGATCACGCTGAACTTCATGCCGATGAGC
CACGTCGGGGGCCGTCAGGTGCTCTACGGGACGCTTTCCAACGGCGGTACC
GCCTACTACGTCGCCAAGAGCGACCTGTCGACGCTGTTCGAGGACCTCGCC

FIG. 8 (Cont.)

CTGGTGCGGCCCACAGAATTGTGCTTCGTGCCGCGCATCTGGGACATGGTG
TTCGCAGAGTTCCACAGCGAGGTCGACCGCCGCTTGGTGGACGGCGCCGAT
CGAGCGGCGCTGGAAGCGCAGGTGAAGGCCGAGCTGCGGGAGAACGTGCTC
GGCGGACGGTTTGTCATGGCGCTGACCGGTTCCGCGCCGATCTCCGCTGAG
ATGACGGCGTGGGTCGAGTCCCTGCTGGCCGACGTGCATTTGGTGGAGGGT
TACGGCTCCACCGAGGCCGGGATGGTCCTGAACGACGGCATGGTGCGGCGC
CCCGCGGTGATCGACTACAAGCTGGTCGACGTGCCCGAGCTGGGCTACTTC
GGCACCGATCAGCCCTACCCCCGGGGCGAGCTGCTGGTCAAGACGCAAACC
ATGTTCCCCGGCTACTACCAGCGCCCGGATGTCACCGCCGAGGTGTTCGAC
CCCGACGGCTTCTACCGGACCGGGGACATCATGGCCAAAGTAGGCCCCGAC
CAGTTCGTCTACCTCGACCGCCGCAACAACGTGCTAAAGCTCTCCCAGGGC
GAGTTCATCGCCGTGTCGAAGCTCGAGGCGGTGTTCGGCGACAGCCCGCTG
GTCCGACAGATCTTCATCTACGGCAACAGTGCCCGGGCCTACCCGCTGGCG
GTGGTTGTCCCGTCCGGGGACGCGCTTTCTCGCCATGGCATCGAGAATCTC
AAGCCCGTGATCAGCGAGTCCCTGCAGGAGGTAGCGAGGGCGGCCGGCCTG
CAATCCTACGAGATTCCACGCGACTTCATCATCGAAACCACGCCGTTCACC
CTGGAGAACGGCCTGCTCACCGGCATCCGCAAGCTGGCACGCCCGCAGTTG
AAGAAGTTCTATGGCGAACGTCTCGAGCGGCTCTATACCGAGCTGGCCGAT
AGCCAATCCAACGAGCTGCGCGAGCTGCGGCAAAGCGGTCCCGATGCGCCG
GTGCTTCCGACGCTGTGCCGTGCCGCGGCTGCGTTGCTGGGCTCTACCGCT
GCGGATGTGCGGCCGGACGCGCACTTCGCCGACCTGGGTGGTGACTCGCTC
TCGGCGCTGTCGTTGGCCAACCTGCTGCACGAGATCTTCGGCGTCGACGTG
CCGGTGGGTGTCATTGTCAGCCCGGCAAGCGACCTGCGGGCCCTGGCCGAC
CACATCGAAGCAGCGCGCACCGGCGTCAGGCGACCCAGCTTCGCCTCGATA
CACGGTCGCTCCGCGACGGAAGTGCACGCCAGCGACCTCACGCTGGACAAG
TTCATCGACGCTGCCACCCTGGCCGCAGCCCCGAACCTGCCGGCACCGAGC
GCCCAAGTGCGCACCGTACTGCTGACCGGCGCCACCGGCTTTTTGGGTCGC
TACCTGGCGCTGGAATGGCTCGACCGCATGGACCTGGTCAACGGCAAGCTG
ATCTGCCTGGTCCGCGCCAGATCCGACGAGGAAGCACAAGCCCGGCTGGAC
GCGACGTTCGATAGCGGCGACCCGTATTTGGTGCGGCACTACCGCGAATTG
GGCGCCGGCCGCCTCGAGGTGCTCGCCGGCGACAAGGGCGAGGCCGACCTG
GGCCTGGACCGGGTCACCTGGCAGCGGCTAGCCGACACGGTGGACCTGATC
GTGGACCCCGCGGCCCTGGTCAACCACGTGCTGCCGTATAGCCAGCTGTTC
GGCCCAAACGCGGCGGGCACCGCCGAGTTGCTTCGGCTGGCGCTGACCGGC
AAGCGCAAGCCATACATCTACACCTCGACGATCGCCGTGGGCGAGCAGATC
CCGCCGGAGGCGTTCACCGAGGACGCCGACATCCGGGCCATCAGCCCGACC
CGCAGGATCGACGACAGCTACGCCAACGGCTACGCGAACAGCAAGTGGGCC
GGCGAGGTGCTGCTGCGCGAAGCTCACGAGCAGTGCGGCCTGCCGGTGACG
GTCTTCCGCTGCGACATGATCCTGGCCGACACCAGCTATACCGGTCAGCTC

FIG. 8 (Cont.)

AACCTGCCGGACATGTTCACCCGGCTGATGCTGAGCCTGGCCGCTACCGGC
ATCGCACCCGGTTCGTTCTATGAGCTGGATGCGCACGGCAATCGGCAACGC
GCCCACTATGACGGCTTGCCGGTCGAATTCGTCGCAGAAGCCATTTGCACC
CTTGGGACACATAGCCCGGACCGTTTTGTCACCTACCACGTGATGAACCCC
TACGACGACGGCATCGGGCTGGACGAGTTCGTCGACTGGCTCAACTCCCCA
ACTAGCGGGTCCGGTTGCACGATCCAGCGGATCGCCGACTACGGCGAGTGG
CTGCAGCGGTTCGAGACTTCGCTGCGTGCCTTGCCGGATCGCCAGCGCCAC
GCCTCGCTGCTGCCCTTGCTGCACAACTACCGAGAGCCTGCAAAGCCGATA
TGCGGGTCAATCGCGCCCACCGACCAGTTCCGCGCTGCCGTCCAAGAAGCG
AAAATCGGTCCGGACAAAGACATTCCGCACCTCACGGCGGCGATCATCGCG
AAGTACATCAGCAACCTGCGACTGCTCGGGCTGCTGTGA

Amino acid sequence (SEQ ID NO:40)

>uniprot|A1KLT8|A1KLT8_MYCBP Probable fatty-acid-CoA ligase fadD9

MSINDQRLTRRVEDLYASDAQFAAASPNEAITQAIDQPGVALPQLIRMVM
EGYADRPALGQRALRFVTDPDSGRTMVELLPRFETITYRELWARAGTLAT
ALSAEPAIRPGDRVCVLGFNSVDYTTIDIALIRLGAVSVPLQTSAPVTGL
RPIVTETEPTMIATSIDNLGDAVEVLAGHAPARLVVFDYHGKVDTHREAV
EAARARLAGSVTIDTLAELIERGRALPATPIADSADDALALLIYTSGSTG
APKGAMYRESQVMSFWRKSSGWFEPSGYPSITLNFMPMSHVGGRQVLYGT
LSNGGTAYYVAKSDLSTLFEDLALVRPTELCFVPRIWDMVFAEFHSEVDR
RLVDGADRAALEAQVKAELRENVLGGRFVMALTGSAPISAEMTAWVESLL
ADVHLVEGYGSTEAGMVLNDGMVRRPAVIDYKLVDVPELGYFGTDQPYPR
GELLVKTQTMFPGYYQRPDVTAEVFDPDGFYRTGDIMAKVGPDQFVYLDR
RNNVLKLSQGEFIAVSKLEAVFGDSPLVRQIFIYGNSARAYPLAVVVPSG
DALSRHGIENLKPVISESLQEVARAAGLQSYEIPRDFIIETTPFTLENGL
LTGIRKLARPQLKKFYGERLERLYTELADSQSNELRELRQSGPDAPVLPT
LCRAAAALLGSTAADVRPDAHFADLGGDSLSALSLANLLHEIFGVDVPVG
VIVSPASDLRALADHIEAARTGVRRPSFASIHGRSATEVHASDLTLDKFI
DAATLAAAPNLPAPSAQVRTVLLTGATGFLGRYLALEWLDRMDLVNGKLI
CLVRARSDEEAQARLDATFDSGDPYLVRHYRELGAGRLEVLAGDKGEADL
GLDRVTWQRLADTVDLIVDPAALVNHVLPYSQLFGPNAAGTAELLRLALT
GKRKPYIYTSTIAVGEQIPPEAFTEDADIRAISPTRRIDDSYANGYANSK
WAGEVLLREAHEQCGLPVTVFRCDMILADTSYTGQLNLPDMFTRLMLSLA
ATGIAPGSFYELDAHGNRQRAHYDGLPVEFVAEAICTLGTHSPDRFVTYH
VMNPYDDGIGLDEFVDWLNSPTSGSGCTIQRIADYGEWLQRFETSLRALP
DRQRHASLLPLLHNYREPAKPICGSIAPTDQFRAAVQEAKIGPDKDIPHL

FIG. 8 (Cont.)

TAAIIAKYISNLRLLGLL

ZP_05227804

Nucleotide sequence (SEQ ID NO:41)

>gi|163719878:4480-7968 Mycobacterium intracellulare ATCC 13950 NZ_ABIN01000296, whole genome shotgun sequence

TTGGCCGCGACCGACGAACAGTTCCGCAACGCCCAACCCGATCTGTCGCTG
CAGCAGGCGGCGCGCCAACCGGGCCTGCGCCTGCCCCAGATTCTGGAGCTG
TTCGTCGAGGGCTACGCCGATCGGCCCGCGGTGGGCTGGCGGGCCAGGACA
CTGAGCACCGACCCCGCGACCGGGCGTACCACCACGCGACTGCTCCCCCGC
TTCGACACCATGACCTACCGCGAGCTGTGGGCCGATGTGCGCGCGATCGCC
GCCGCGTGGCGGCACGACGCCGCGAACCCCGTGTCGCCCGGGGACTTCGTC
GCGACGGTCGGTTTCGCCAGCGCCGAATATCTGACCCTCGACCTGGTCTGC
GGCTACCTCGGGCTGGTTGCCGTTCCGCTGCAACACAATACGACCCCCTCA
CGGCTGCGACCGATCGTCGACGAGGTCGAGCCGTCGATACTGGCCGCCGGC
GTCGGTTATCTCGACCTCGCGGTCGAGGCGGCGTCGGGCAGCTCGTCGTTG
CGGCGGCTCGTGGTCTTCGATTACCAGCCCGAGGTCGACGAGCAACGCGAG
GCGCTGCAGCGCGCCCAGGCGACGCTGGCGGCCGCCGGCGCGGCGGTGACC
ATCGAGACCCTCGACGAAATCATCGAACGCGGACGCGCCCTGCCGCCCGAG
CCCATGTACACCGGTGATACCGATCAGCGGCTGGCGATGATCATGTACACC
TCGGGCAGCACCGGGTTACCCAAGGGCGCCATGTACACCGAGCAGATGCTG
GCCAAGGTGTGGACCAACGAGCTGATGCCCGACTTCGCGGACACACCCGTG
TTCAACGTCAACTTCATGCCGCTGAATCACCTCGGTGGCCGGATACCGCTG
TCGACCGCGTTCCAGGCCGGCGGCACCAGCTATTTCGTGCCGGAAAGCGAC
CTGTCCACGTTGTTCGACGACTGGAACCTGGTGCGCCCCACCGAGATGGGC
CTGGTACCCCGGGTGGCGGAAATGCTCTACCAGCGCTACCAGAGCGCCGTC
GACCGACTCGTGGCTTCGGGCGCCGACGCCGGCTCCGCCGAGGCCCGGGCG
CGGGCCGAGCTGCGTGAGCATGTCCTCGGCGGGCGCATCGTGACCGCCTTC
TGCGGGACGGCGCCGCTGGCCGCGGAGATGCGGGCCTTCGTCGAAACCTGT
TTGGACGTCCACGTTCTCGACGGCTACGGGCTGACCGAGGTCGGCATGGTG
ACCAAGGACGGGCGCATGACCCGTCCCCCGGTGCTCGACTACAAGCTCATC
GACGTTCCCGAACTCGGCTATTTCCACACCGACAAGCCTTATCCGCGTGGC
GAATTGCTGGTGAAGTCGCTGACCGCGACGCCGGGCTACTTCAAACGACCG
GACGTCACCGCCAACGCGTTCGATCCCGACGGCTACTACCGGACCGGCGAT
GTGATGGCCGAGCTCGAGCCGGACCGGCTGGCCTACGTCGACCGCCGCAAC
AACGTGTTGAAGTTGGCGCAGGGCGAGTTCGTCGCCGTCGCCCGCCTGGAG

FIG. 8 (Cont.)

GCCGTCTTCGCCAGCGCGCCGCTGATCCGCCAGATCTTCGTATACGGCAAC
AGCGAACGCCCCTATCTGCTGGCCGTCGTCGTGCCGACGGCCGACGCCGCG
GAGCGATTCACCGGAGATCCCGAGGGCCTCAAGGCCGCCGTCGCCGAATCC
CTGCGCCAGTCGGCGCAACTCGCCGAACTGCAGTCCTACGAGGTGCCCGTC
GACTTCGTCGTCGAGACCGAGCCGTTCAGCGAGGACAACGGCCTGCTCTCG
GGCGTGGGCAAGCTGCTGCGGCCGAAGCTCAAGGAGCGCTACGCCGACCGG
CTCGAACAGCTCTACGCCGAGCTGGCCGAAAACCGCGTGACCGAGTTGCGT
GCGCTGCGCGAGGGGGCGGACAAACACCCCGTCGTCTTCACCCTCACCCGG
GCCGCCGAGGCGCTACTGGGTGTGGCCGGCGGCCCGCCCGCCCCCGACGCA
CTGTTCATCGAACTCGGCGGCGATTCCCTGTCGGCGCTGACCTTCTCCAAC
CTGCTGCGCGACATCTTCGACGTCGACGTGCCGGTGGGAATGATCACCGGG
CCCGCGACCGACCTGGGCCAGCTCGCGGAATACGTTGAATCCGAACGCAAA
TCGGGATCACGCCGGCCCACATTCGCGACGGTGCACGGACGCGGCGCCGCC
GAGGTCCGCGCCGCGGAGCTCACCCTCGACAAGTTCATCGACGCGACGACC
TTGGCCGCCGCACCGAACCTGCCGCGCGCGACCGGCACACCCCACACGGTC
CTGCTGACCGGCGCCAACGGCTACCTCGGCCGCTTCCTGGCCCTCGAATGG
CTCGAGCGCCTCGCCGAGACCGGCGGGAAGCTCGTCTCCATCGTCCGCGCG
ACGGACACCGCGGCGGCCGTCAAACGGCTGGAGGCCGTTTTCGACAGCGGG
GATCCGCAGTTGCTGGAGCGGTTCCGGACGCTGGCCGCCGAGCACCTGGAA
GTCATCGTCGGCGACATCGGTGAGCCCAATCTCGGCCTGGACCAAGCGACT
TGGCAGCGCCTGGCCCAGAGCGTGGATCTGATCGTCCACCCGGCCGCGTTG
GTCAACCACGTGCTGCCGTACGACCAACTGTTCGGTCCGAACGTCGTCGGC
ACCGCCGAGTTGATCCGCCTGGCGATCACGACGCGCATCAAGCCCGTCACC
TATCTGTCGACCGTCGCCGTGGCGATGACGGTCGATCCCGGCGAGTTCGCC
GAAGACGGCGACATCCGCGCGGTCAGCGCGGTACGCCCGATCGACGACAGC
TACGCGAACGGGTACGCGAACAGCAAGTGGGCCGGTGAGGTGTTGCTGCGT
GAGGCGCACGACCTGTGCGGGTTGCCGGTCGCCGTCTTCCGCTCCGACATG
ATCCTCGCGCACAGCCGGTATGCCGGGCAGCTGAACGTGCCAGATGCCTTC
ACCCGCTTGATGTTCAGCCTGCTGACCACCGGCATCGCGCCGACCACGTTC
TACCGGACCGACGAACACGGAAACCGAGCCGTGGCCCACTACGACGGGCTG
CCCGCCGACTTCGTGGCCGAAGCGGTCACCACGCTCGGCGAACAGATGGCG
GCCGAGGAATCCGGCGGGTACCGCTCCTATGACGTGATGAACCCACACGAC
GACGGCGTCTCCCTGGACGTGTTCGTCGACTGGCTGATCGCCGCCGGACAC
GACATCCGGCGCATCGAGGACTATGACGAATGGCTGGGCCGCTTCACCACG
GCGCTTCGCGCGTTACCGGACAAGCAGCGCCAGCATTCGGTGTTGCCGCTG
CTGGACGCCTACCGGGAACCCGCGACGCCACTGCGGGGAGCGCCGGCCCCC
ACCGACGTCTTCCGCCACGCGGTGCGGACGGCCAAAATCGGTGCGGACGAG
GACATTCCGCACCTGTCGGCGGCGTTGATCGACAAGTACGTCGCCGACCTA
CGCCTGCTGGGCTTGGTGTAG

FIG. 8 (Cont.)

Amino acid sequence (SEQ ID NO:42)

>gi|254822803|ref|ZP_05227804.1| putative long-chain fatty-acid--CoA ligase [Mycobacterium intracellulare ATCC 13950]
MAATDEQFRNAQPDLSLQQAARQPGLRLPQILELFVEGYADRPAVGWRART
LSTDPATGRTTTRLLPRFDTMTYRELWADVRAIAAAWRHDAANPVSPGDFV
ATVGFASAEYLTLDLVCGYLGLVAVPLQHNTTPSRLRPIVDEVEPSILAAG
VGYLDLAVEAASGSSSLRRLVVFDYQPEVDEQREALQRAQATLAAAGAAVT
IETLDEIIERGRALPPEPMYTGDTDQRLAMIMYTSGSTGLPKGAMYTEQML
AKVWTNELMPDFADTPVFNVNFMPLNHLGGRIPLSTAFQAGGTSYFVPESD
LSTLFDDWNLVRPTEMGLVPRVAEMLYQRYQSAVDRLVASGADAGSAEARA
RAELREHVLGGRIVTAFCGTAPLAAEMRAFVETCLDVHVLDGYGLTEVGMV
TKDGRMTRPPVLDYKLIDVPELGYFHTDKPYPRGELLVKSLTATPGYFKRP
DVTANAFDPDGYYRTGDVMAELEPDRLAYVDRRNNVLKLAQGEFVAVARLE
AVFASAPLIRQIFVYGNSERPYLLAVVVPTADAAERFTGDPEGLKAAVAES
LRQSAQLAELQSYEVPVDFVVETEPFSEDNGLLSGVGKLLRPKLKERYADR
LEQLYAELAENRVTELRALREGADKHPVVFTLTRAAEALLGVAGGPPAPDA
LFIELGGDSLSALTFSNLLRDIFDVDVPVGMITGPATDLGQLAEYVESERK
SGSRRPTFATVHGRGAAEVRAAELTLDKFIDATTLAAAPNLPRATGTPHTV
LLTGANGYLGRFLALEWLERLAETGGKLVSIVRATDTAAAVKRLEAVFDSG
DPQLLERFRTLAAEHLEVIVGDIGEPNLGLDQATWQRLAQSVDLIVHPAAL
VNHVLPYDQLFGPNVVGTAELIRLAITTRIKPVTYLSTVAVAMTVDPGEFA
EDGDIRAVSAVRPIDDSYANGYANSKWAGEVLLREAHDLCGLPVAVFRSDM
ILAHSRYAGQLNVPDAFTRLMFSLLTTGIAPTTFYRTDEHGNRAVAHYDGL
PADFVAEAVTTLGEQMAAEESGGYRSYDVMNPHDDGVSLDVFVDWLIAAGH
DIRRIEDYDEWLGRFTTALRALPDKQRQHSVLPLLDAYREPATPLRGAPAP
TDVFRHAVRTAKIGADEDIPHLSAALIDKYVADLRLLGLV

YP 001288539.1

Nucleotide sequence (SEQ ID NO:43)

>uniprot|A1QUM2|A1QUM2_MYCTF Fatty-acid-CoA ligase fadD9

ATGTCGATCAACGATCAGCGACTGACACGCCGCGTCGAGGACCTATACGCC
AGCGACGCCCAGTTCGCCGCCGCCAGTCCCAACGAGGCGATCACCCAGGCG
ATCGACCAGCCCGGGGTCGCGCTTCCACAGCTCATCCGTATGGTCATGGAG

FIG. 8 (Cont.)

GGCTACGCCGATCGGCCGGCACTCGGCCAGCGTGCGCTCCGCTTCGTCACC
GACCCCGACAGCGGCCGCACCATGGTCGAGCTACTGCCGCGGTTCGAGACC
ATCACCTACCGCGAACTGTGGGCCCGCGCCGGCACATTGGCCACCGCGTTG
AGCGCTGAGCCCGCGATCCGGCCGGGCGACCGGGTTTGCGTGCTGGGCTTC
AACAGCGTCGACTACACAACCATCGACATCGCGCTGATCCGGTTGGGCGCC
GTGTCGGTTCCACTGCAGACCAGTGCGCCGGTCACCGGGTTGCGCCCGATC
GTCACCGAGACCGAGCCGACGATGATCGCCACCAGCATCGACAATCTTGGC
GACGCCGTCGAAGTGCTGGCCGGTCACGCCCCGGCCCGGCTGGTCGTATTC
GATTACCACGGCAAGGTTGACACCCACCGCGAGGCCGTCGAAGCCGCCCGA
GCTCGGTTGGCCGGCTCGGTGACCATCGACACACTTGCCGAACTGATCGAA
CGCGGCAGGGCGCTGCCGGCCACACCCATTGCCGACAGCGCCGACGACGCG
CTGGCGCTGCTGATTTACACCTCGGGTAGTACCGGCGCACCCAAAGGCGCC
ATGTATCGCGAGAGCCAGGTGATGAGCTTCTGGCGCAAGTCGAGTGGCTGG
TTCGAGCCGAGCGGTTACCCCTCGATCACGCTGAACTTCATGCCGATGAGC
CACGTCGGGGGCCGTCAGGTGCTCTACGGGACGCTTTCCAACGGCGGTACC
GCCTACTTCGTCGCCAAGAGCGACCTGTCGACGCTGTTCGAGGACCTCGCC
CTGGTGCGGCCCACAGAATTGTGCTTCGTGCCGCGCATCTGGGACATGGTG
TTCGCAGAGTTCCACAGCGAGGTCGACCGCCGCTTGGTGGACGGCGCCGAT
CGAGCGGCGCTGGAAGCGCAGGTGAAGGCCGAGCTGCGGGAGAACGTGCTC
GGCGGACGGTTTGTCATGGCGCTGACCGGTTCCGCGCCGATCTCCGCTGAG
ATGACGGCGTGGGTCGAGTCCCTGCTGGCCGACGTGCATTTGGTGGAGGGT
TACGGCTCCACCGAGGCCGGGATGGTCCTGAACGACGGCATGGTGCGGCGC
CCCGCGGTGATCGACTACAAGCTGGTCGACGTGCCCGAGCTGGGCTACTTC
GGCACCGATCAGCCCTACCCCCGGGGCGAGCTGCTGGTCAAGACGCAAACC
ATGTTCCCCGGCTACTACCAGCGCCCGGATGTCACCGCCGAGGTGTTCGAC
CCCGACGGCTTCTACCGGACCGGGGACATCATGGCCAAAGTAGGCCCCGAC
CAGTTCGTCTACCTCGACCGCCGCAACAACGTGCTAAAGCTCTCCCAGGGC
GAGTTCATCGCCGTGTCGAAGCTCGAGGCGGTGTTCGGCGACAGCCCGCTG
GTCCGACAGATCTTCATCTACGGCAACAGTGCCCGGGCCTACCCGCTGGCG
GTGGTTGTCCCGTCCGGGGACGCGCTTTCTCGCCATGGCATCGAGAATCTC
AAGCCCGTGATCAGCGAGTCCCTGCAGGAGGTAGCGAGGGCGGCCGGCCTG
CAATCCTACGAGATTCCACGCGACTTCATCATCGAAACCACGCCGTTCACC
CTGGAGAACGGCCTGCTCACCGGCATCCGCAAGCTGGCACGCCCGCAGTTG
AAGAAGTTCTATGGCGAACGTCTCGAGCGGCTCTATACCGAGCTGGCCGAT
AGCCAATCCAACGAGCTGCGCGAGCTGCGGCAAAGCGGTCCCGATGCGCCG
GTGCTTCCGACGCTGTGCCGTGCCGCGGCTGCGTTGCTGGGCTCTACCGCT
GCGGATGTGCGGCCGGACGCGCACTTCGCCGACCTGGGTGGTGACTCGCTC
TCGGCGCTGTCGTTGGCCAACCTGCTGCACGAGATCTTCGGCGTCGACGTG
CCGGTGGGTGTCATTGTCAGCCCGGCAAGCGACCTGCGGGCCCTGGCCGAC

FIG. 8 (Cont.)

CACATCGAAGCAGCGCGCACCGGCGTCAGGCGACCCAGCTTCGCCTCGATA
CACGGTCGCTCCGCGACGGAAGTGCACGCCAGCGACCTCACGCTGGACAAG
TTCATCGACGCTGCCACCCTGGCCGCAGCCCCGAACCTGCCGGCACCGAGC
GCCCAAGTGCGCACCGTACTGCTGACCGGCGCCACCGGCTTTTTGGGTCGC
TACCTGGCGCTGGAATGGCTCGACCGCATGGACCTGGTCAACGGCAAGCTG
ATCTGCCTGGTCCGCGCCAGATCCGACGAGGAAGCACAAGCCCGGCTGGAC
GCGACGTTCGATAGCGGCGACCCGTATTTGGTGCGGCACTACCGCGAATTG
GGCGCCGGCCGCCTCGAGGTGCTCGCCGGCGACAAGGGCGAGGCCGACCTG
GGCCTGGACCGGGTCACCTGGCAGCGGCTAGCCGACACGGTGGACCTGATC
GTGGACCCCGCGGCCCTGGTCAACCACGTGCTGCCGTATAGCCAGCTGTTC
GGCCCAAACGCGGCGGGCACCGCCGAGTTGCTTCGGCTGGCGCTGACCGGC
AAGCGCAAGCCATACATCTACACCTCGACGATCGCCGTGGGCGAGCAGATC
CCGCCGGAGGCGTTCACCGAGGACGCCGACATCCGGGCCATCAGCCCGACC
CGCAGGATCGACGACAGCTACGCCAACGGCTACGCGAACAGCAAGTGGGCC
GGCGAGGTGCTGCTGCGCGAAGCTCACGAGCAGTGCGGCCTGCCGGTGACG
GTCTTCCGCTGCGACATGATCCTGGCCGACACCAGCTATACCGGTCAGCTC
AACCTGCCGGACATGTTCACCCGGCTGATGCTGAGCCTGGCCGCTACCGGC
ATCGCACCCGGTTCGTTCTATGAGCTGGATGCGCACGGCAATCGGCAACGC
GCCCACTATGACGGCTTGCCGGTCGAATTCGTCGCAGAAGCCATTTGCACC
CTTGGGACACATAGCCCGGACCGTTTTGTCACCTACCACGTGATGAACCCC
TACGACGACGGCATCGGGCTGGACGAGTTCGTCGACTGGCTCAACTCCCCA
ACTAGCGGGTCCGGTTGCACGATCCAGCGGATCGCCGACTACGGCGAGTGG
CTGCAGCGGTTCGAGACTTCGCTGCGTGCCTTGCCGGATCGCCAGCGCCAC
GCCTCGCTGCTGCCCTTGCTGCACAACTACCGAGAGCCTGCAAAGCCGATA
TGCGGGTCAATCGCGCCCACCGACCAGTTCCGCGCTGCCGTCCAAGAAGCG
AAAATCGGTCCGGACAAAGACATTCCGCACCTCACGGCGGCGATCATCGCG
AAGTACATCAGCAACCTGCGACTGCTCGGGCTGCTGTGA

Amino acid sequence (SEQ ID NO:44)

>uniprot|A1QUM2|A1QUM2_MYCTF Fatty-acid-CoA ligase fadD9

MSINDQRLTRRVEDLYASDAQFAAASPNEAITQAIDQPGVALPQLIRMVM
EGYADRPALGQRALRFVTDPDSGRTMVELLPRFETITYRELWARAGTLAT
ALSAEPAIRPGDRVCVLGFNSVDYTTIDIALIRLGAVSVPLQTSAPVTGL
RPIVTETEPTMIATSIDNLGDAVEVLAGHAPARLVVFDYHGKVDTHREAV
EAARARLAGSVTIDTLAELIERGRALPATPIADSADDALALLIYTSGSTG
APKGAMYRESQVMSFWRKSSGWFEPSGYPSITLNFMPMSHVGGRQVLYGT
LSNGGTAYFVAKSDLSTLFEDLALVRPTELCFVPRIWDMVFAEFHSEVDR

FIG. 8 (Cont.)

RLVDGADRAALEAQVKAELRENVLGGRFVMALTGSAPISAEMTAWVESLL
ADVHLVEGYGSTEAGMVLNDGMVRRPAVIDYKLVDVPELGYFGTDQPYPR
GELLVKTQTMFPGYYQRPDVTAEVFDPDGFYRTGDIMAKVGPDQFVYLDR
RNNVLKLSQGEFIAVSKLEAVFGDSPLVRQIFIYGNSARAYPLAVVVPSG
DALSRHGIENLKPVISESLQEVARAAGLQSYEIPRDFIIETTPFTLENGL
LTGIRKLARPQLKKFYGERLERLYTELADSQSNELRELRQSGPDAPVLPT
LCRAAAALLGSTAADVRPDAHFADLGGDSLSALSLANLLHEIFGVDVPVG
VIVSPASDLRALADHIEAARTGVRRPSFASIHGRSATEVHASDLTLDKFI
DAATLAAAPNLPAPSAQVRTVLLTGATGFLGRYLALEWLDRMDLVNGKLI
CLVRARSDEEAQARLDATFDSGDPYLVRHYRELGAGRLEVLAGDKGEADL
GLDRVTWQRLADTVDLIVDPAALVNHVLPYSQLFGPNAAGTAELLRLALT
GKRKPYIYTSTIAVGEQIPPEAFTEDADIRAISPTRRIDDSYANGYANSK
WAGEVLLREAHEQCGLPVTVFRCDMILADTSYTGQLNLPDMFTRLMLSLA
ATGIAPGSFYELDAHGNRQRAHYDGLPVEFVAEAICTLGTHSPDRFVTYH
VMNPYDDGIGLDEFVDWLNSPTSGSGCTIQRIADYGEWLQRFETSLRALP
DRQRHASLLPLLHNYREPAKPICGSIAPTDQFRAAVQEAKIGPDKDIPHL
TAAIIAKYISNLRLLGLL

YP 953393.1

Nucleotide sequence (SEQ ID NO:45)

>uniprot|A1T887|A1T887_MYCVP Thioester reductase domain

ATGTCTACTGATACCCGCGAAGACCGCCTTGCCCGCCGCATCGCCGATCTG
TACGCCACCGACCCGCAGTTTGCCGCCGCAGCACCCGACGACGCCATCTCC
CACGCCATCGATCAACCAGGGACGCACCTGCCCGTCATCGTGCAGACCGTC
CTCGACGGATACGCCGAGCGGCCCGCGCTCGGACAGCGCGCGGTCCGTTTC
GTCACCGACCCCGCCACGGGCAGAACCACCACCGAGCTGCTCCCCCGCTTC
GAGACCATCACCTACGCCGAGTTGTCACGCCGCATCCACGCGGTCACCGCC
GCGCTGACCGATGTGCATCCGGGCGACCGCGTGGCGGTGCTGGGCTTCACC
AGCATCGACTACACCACCGTCGACATGGCTTTGGCGATGCTGGGCGCGGTG
GCGGTACCACTTCAGACCAGCGCGCCGGCCACCACCGTCCGGCCGATCGTC
GCCGAGACCGAACCGGTGGTCATCGCATCGTCGGTCGATGCCCTCACCGAC
GCCGTCGGCCTGGCTCTCGACGCTCCCACCGTGACCCGCCTTGTGGTCTTC
GATCATCGCGCCGGGGTCGACGATCATCGCGACGCCCTCATCTCCGCGAGC
GACCGGTTGCGCGCCGCCAACTCGCCGATCGAGGTCGAGACCATTACCGAC
ATCGTCGCTCGGGGTTCGAAACTGCCTGTACGCGCGCAATTCTCGGCCGAC

FIG. 8 (Cont.)

GGTGACGCGCTGAGCCTGCTGATCTACACCTCCGGCAGCACCGGCGCACCC
AAGGGTGCGATGTACCCCCAACACCTGGTCGCCAACTCATGGCGGCGGTTG
GCCCGGTCCTTCTGGGGCGACCTGGGGGTCTTCCCGGCAATCACGCTGAAC
TTCATGCCGATGAGCCACGTGATGGGCCGCGGACTGCTCTACGGCACGCTG
GACGCCGGCGGCACCGCGTATTTCGCGGCCAGGAGCGATCTGTCGACGTTC
CTGGAGGATCTCGCCCTGGTGCGCCCGACGCAGCTGAGCTTCGTGCCGCGG
ATCTGGGACACCATCCACGCCGAGGTGTCCCAGGAACTCGAGCGCCGGCCG
TCGGATGCGACCGAGGTGATCGCCGATCTGCGGCGGAGCCTGCTGGGCGGC
CGCTACGTCACGGCGATGACGGGCTCCGCGCCGCTGTCACCGGAGATGCGG
GCCTTCGTCGAGAACCTGCTCGACGTGCACCTGATCGACGGGTACGGCTCG
ACCGAGGCCGGCGCGGTGTTCGTCGACGGCCGGGTCCAACGCCCGCCGGTC
ATCGATTACAAGCTCGTCGACGTCGCCGACCTCGGCTACTTCTCCACCGAC
CGCCCCCATCCGCGCGGCGAGCTTCTCGTCAAATCCGAGACGCTGTTTCCC
GGCTACTACAAACGCCCCGACGTCACCGCCGAGATGTTCGACGAAGACGGC
TACTACCGCACCGGCGACATCGTCGCCGAGACCGGGGCCGACCAGCTGACC
TATCTGGACCGCCGCAACAACGTCCTCAAACTGTCGCAGGGCGAATTCGTC
ACCGTCTCCCGACTGGAGGCAGTGTTCGGCAACAGCCCGCTGGTCCGCCAG
ATCTACGTCTACGGCAACAGCGCCCGCCCCTACCTGCTGGCTGTGGTCGTG
CCCACCGAAGCCGCGCTGGCCGGTGCTGACGCCAAAGCCGCTGTGGCCGAG
TCACTTCAGGATGTCGCCAAGGCGACCGGGCTGCAGTCCTACGAGATCCCC
CGCGATTTCCTTCTCGAGACGACGCCGTTCACGCTGGAGAACGGCTTGCTG
ACCGGCATCCGAAAACTGGCCCGCCCCAGACTCAGAGAGCGTTACGGCGAA
CAGCTCGAGGCCCTCTACACCATGCTGTCCGAAGAGCAGGCCGACGAGCTG
CGGGAGCTGCGCCGCTCCGGCGGAGAGCGTCCGGCGCTGGAAACCGTCGGA
CGCGCCGCCGGGGCGCTGCTCGGCACCACCGCAGGCGAGCTGGAGCCGAGC
GCCCACTTCACAGATCTGGGCGGGGATTCGCTGTCGGCGTTGACCTTCGCC
AACCTGCTGCGCGACATCTTCGACGTCGACGTCCCCGTCGGTGTGATCGTC
AGCCCGGCCACCGATCTGCAGGCCCTTGCCGACTACGTCGAGTCCGCCCGC
CGGCACGGGTCGGTGCGGCCCACTTTCGAATCGGTGCACGGGCATTCGGGA
CGACCCGGGACCGAGGTGCATGCCCGCGATCTGACGTTGGACGAATTCGTC
GACGCCGCGACCCTGGCGCACGCGCCGACGTTGCCCGGACCGCGCGCCGAG
GTCCGCACCGTCCTGCTGACCGGGGCGACCGGCTTCCTCGGCCGGTATCTC
GCTCTCGAATGGCTTGAGCGGATGGCGCTGGTCGGCGGCAAGCTGATCTGC
CTGGTCCGCGCCAAAGACGATGCGGCAGCGCGGGTTCGGCTGGACAGCACG
TTCGACAGCGGGGACCCGGAGCTGCTGCGGCACTACCGACGGCTGGCAGCC
GACCATCTCGAAGTGATCGCCGGCGACAAGGCCGACGCCGATCTCGGACTC
GACGCGCGGACGTGGCAGCGGCTCGCGGACACCGTCGATCTGATCGTCGAT
CCCGCCGCCCTGGTCAACCACGTCCTGCCGTACCGCCAACTGTTCGCCCCG
AACGTGCTCGGCACCGCCGAACTGCTCCGCATCGCGCTGACAACGAGGATG

FIG. 8 (Cont.)

AAGCCGTTCGTGTACGTGTCGACGATCGGCGTCGGCGCCGGTATCGAACCC
GCGAGGTTCACCGAGGACGCCGACATCAGGCAGATCAGCGCGACGCGCAGA
ATCGACGACAGCTACGCCAACGGCTACGGCAACAGCAAGTGGGCCGGCGAG
GTGCTGCTGCGCGAGGCGCACGATCTGTGCGGCCTGCCCGTTTCGGTGTTC
CGCTGCGACATGATCCTGGCCGACACCACCTACGCCGGTCAGCTGAACCTG
CCGGACATGTTCACCCGCCTGATCTTCAGCCTGGTCGCCACCGGCGTCGCG
CCCGAGTCGTTCTACCACCTCGCCACCGACGGCACCCGGCAACGGGCCCAC
TATGACGGGCTACCGGTGGAGTTCATCGCCGAAGCGATCTCGACCCTCGGC
TCTGACGTCGCCTCCGGGTTCCGGACGTATCACGTGATGAATCCCCACGAC
GACGGGATAGGCCTCGACGAGTACGTGGACTGGCTCATCGACGCAGGCCAT
CCGATTCGACGCGTCGGCGACTACCCGACGTGGCTGCAACGGTTCACGGTG
GCGATCACCGCGCTGCCGGAACGGCAGCGGCAGGCCTCACTGTTGCCGCTG
CTGCACAACTACCAGCACCCGGAGACGCCGATCCGCGGCTCCATCGCACCG
ACAGACAGATTCCGTGAAGCGGTCCAGGACGCCAAAATCGGTCCGGACAAA
GACATTCCACACGTGACACCGCAGATCGTCATCAAGTACGTCACCGACCTG
CAGCGACTCGGACTGCTCTAA

Amino acid sequence (SEQ ID NO:46)

>uniprot|A1T887|A1T887_MYCVP Thioester reductase domain

MSTDTREDRLARRIADLYATDPQFAAAAPDDAISHAIDQPGTHLPVIVQT
VLDGYAERPALGQRAVRFVTDPATGRTTTELLPRFETITYAELSRRIHAV
TAALTDVHPGDRVAVLGFTSIDYTTVDMALAMLGAVAVPLQTSAPATTVR
PIVAETEPVVIASSVDALTDAVGLALDAPTVTRLVVFDHRAGVDDHRDAL
ISASDRLRAANSPIEVETITDIVARGSKLPVRAQFSADGDALSLLIYTSG
STGAPKGAMYPQHLVANSWRRLARSFWGDLGVFPAITLNFMPMSHVMGRG
LLYGTLDAGGTAYFAARSDLSTFLEDLALVRPTQLSFVPRIWDTIHAEVS
QELERRPSDATEVIADLRRSLLGGRYVTAMTGSAPLSPEMRAFVENLLDV
HLIDGYGSTEAGAVFVDGRVQRPPVIDYKLVDVADLGYFSTDRPHPRGEL
LVKSETLFPGYYKRPDVTAEMFDEDGYYRTGDIVAETGADQLTYLDRRNN
VLKLSQGEFVTVSRLEAVFGNSPLVRQIYVYGNSARPYLLAVVVPTEAAL
AGADAKAAVAESLQDVAKATGLQSYEIPRDFLLETTPFTLENGLLTGIRK
LARPRLRERYGEQLEALYTMLSEEQADELRELRRSGGERPALETVGRAAG
ALLGTTAGELEPSAHFTDLGGDSLSALTFANLLRDIFDVDVPVGVIVSPA
TDLQALADYVESARRHGSVRPTFESVHGHSGRPGTEVHARDLTLDEFVDA
ATLAHAPTLPGPRAEVRTVLLTGATGFLGRYLALEWLERMALVGGKLICL
VRAKDDAAARVRLDSTFDSGDPELLRHYRRLAADHLEVIAGDKADADLGL
DARTWQRLADTVDLIVDPAALVNHVLPYRQLFAPNVLGTAELLRIALTTR

FIG. 8 (Cont.)

MKPFVYVSTIGVGAGIEPARFTEDADIRQISATRRIDDSYANGYGNSKWA
GEVLLREAHDLCGLPVSVFRCDMILADTTYAGQLNLPDMFTRLIFSLVAT
GVAPESFYHLATDGTRQRAHYDGLPVEFIAEAISTLGSDVASGFRTYHVM
NPHDDGIGLDEYVDWLIDAGHPIRRVGDYPTWLQRFTVAITALPERQRQA
SLLPLLHNYQHPETPIRGSIAPTDRFREAVQDAKIGPDKDIPHVTPQIVI
KYVTDLQRLGLL

YP 938306.1

Nucleotide sequence (SEQ ID NO:47)

>uniprot|A1UFA8|A1UFA8_MYCSK Thioester reductase domain

ATGTCCACCGAGACCCGTGAAGCGCGCCTGCAGCAGCGCATCGCGCATCTG
TTCACCACCGACCCGCAGTTCGCCGCCGCCCGGCCCGACCCCCGGATCAGC
GATGCCGTCGACCGCGATGACACGCGGCTGACCGCCATCGTCTCCGCGGTC
ATGTCGGGGTACGCCGACCGGCCGGCACTCGGGCAGCGCGCCGCCGAATTC
GTCACCGACCCGCAGACCGGCCGCACCACGATGGAGCTGCTCCCCCGCTTC
GACACCATCACCTACCGGGAGCTGCTCGACCGCGTCCGGGCGCTCACCAAC
GCGTGGCACGCCGACGGGGTGCGCCCCGGCGACCGTGTCGCAATCCTCGGA
TTCACCGGCATCGACTACACCGTGGTCGACCTCGCCCTGATCCAGCTCGGC
GCGGTCGCGGTGCCGCTGCAGACCAGCGCCGCCGTGGAGGCGTTGCGCCCG
ATCGTGGCCGAGACCGAACCGATGCTCATCGCCACCGGCGTCGATCATGTC
GACGCCGCCGCGGAACTCGCACTCACCGGCCACCGTCCGTCCCGGGTGGTG
GTCTTCGACCACCGCGAGCAGGTCGACGACGAACGCGACGCGGTGCGGGCC
GCGACGGCCAGGCTGGGAGACGCCGTCCCCGTCGAGACACTCGCCGAGGTG
TTGCGGCGGGGCGCCCATCTGCCCGCCGTGGCGCCGCACGTGTTCGACGAG
GCCGATCCACTGCGGCTGCTGATCTACACTTCCGGCAGCGCCGGCGCCCCC
AAGGGCGCGATGTATCCCGAGAGCAAGGTCGCCGGCATGTGGCGCGCGTCG
GCCAAGGCCGCCTGGAACAACGATCAGACAGCGATTCCGTCGATCACCCTG
AACTTCCTGCCGATGAGCCACGTCATGGGTCGCGGCCTGCTGTGCGGCACG
CTCAGCACCGGTGGCACCGCGTATTTCGCCGCCCGCAGCGATCTGTCGACG
CTGCTCGAGGACCTGCGCCTGGTACGGCCCACCCAACTCAGCTTCGTGCCG
CGGATCTGGGACATGCTCTTCCAGGAGTTCGTCGGCGAGGTCGACCGGCGG
GTGAACGACGGTGCGGACCGCCCCACCGCGGAGGCCGACGTGCTGGCCGTA
CAGCGTCACGAGCTGCTCGGTGGCCGGTTCGTCACCGCGATGACCGGTTCG
GCGCCCATCTCCCTCGAGATGAAGACATGGGTGGAGACCCTGCTCGACATG
CACCTGGTCGAGGGTTACGGCTCGACGGAGGCCGGCGCGGTGTTCGTCGAC

FIG. 8 (Cont.)

GGCCACATCCAGCGCCCACCGGTGCTCGACTACAAACTCGTCGACGTCCCC
GACCTCGGCTACTTCAGCACCGACCGGCCGCACCCGCGCGGTGAGCTGCTG
GTCCGCTCCACGCAGCTATTCCCCGGCTACTACAAACGTCCCGACGTCACC
GCCGAGGTGTTCGACGACGACGGCTTCTACCGCACGGGCGACATCGTCGCC
GAGGTCGGCCCCGATCAGGTGCAGTACCTCGACCGCCGCAACAACGTGCTC
AAACTCGCCCAGGGTGAGTTCGTCACCATCTCCAAACTCGAGGCGGTCTTC
GCCGGCAGCGCCCTGGTCCGCCAGATCTACGTGTACGGCAACAGTGCGCGC
TCCTACCTGCTGGCCGTCGTCGTGCCGACCGACGATGCGGTGGCCCGGCAC
GACCCGGCATCGCTCAAGACCGCGATCAGCGCCTCGCTGCAGCAGGCCGCG
AAGACCGCCGGTCTGCAGTCCTACGAGCTGCCGCGTGACTTCCTCGTCGAG
ACTCAACCGTTCACGCTGGAGAACGGACTACTGACCGGCATCCGCAAGCTG
GCGCGCCCGAAACTCAAGGCGCGCTACGGCGATCGGCTCGAGGCGCTCTAC
GTCGAACTGGTCGAAGGACAGGCAGGCGAATTGCGCACCCTGCGCCGGGAC
GGCGCGAAGCGTCCGGTGGCCGAGACGGTCGGCCGCGCCGCGGCCGCGCTG
CTCGGCGCCGCCGCCGCCGACGTGCGCCCCGACGCGCACTTCACCGACCTC
GGCGGAGACTCGTTGTCGGCGTTGACCTTCGGCAATCTGCTGCAGGAGATC
TTCGGCGTCGACGTCCCGGTCGGGGTGATCGTCAGCCCGGCGGCCGACCTG
GCGTCGATCGCGGCGTACATCGAGGCCGAACAGGCCTCGACCGGTAAGCGG
CCGACCTACGCGTCGGTGCACGGGCGCGACGCCGAACAGGTACACGCGCGC
GACCTCACCCTGGACAAGTTCATCGACGCCGAAACACTCTCCGCTGCAACA
GAACTGCCCGGCCCGAGCGGTGAGGTGCGCACCGTGCTGCTGACCGGGGCC
ACCGGATTCCTCGGCCGCTACCTGGCGCTGGACTGGCTCGAACGGATGGCC
CTGGTCGACGGCAAGGTCATCTGCCTGGTCCGCGCGAAGGACGATGCGGCC
GCCCGCAAACGCCTCGACGACACCTTCGACAGCGGCGATCCGAAGCTGCTG
GCGCACTACCGCAAGCTGGCCGCCGACCACCTCGAAGTGCTGGCCGGCGAC
AAGGGTGAGGCGGATCTCGGGCTGCCGCATCCTGTCTGGCAGCGCCTGGCC
GACACCGTCGACCTCATCGTCGACCCGGCCGCCCTGGTCAACCACGTACTG
CCGTACAGCCAGCTGTTCGGGCCCAACGCGCTGGGCACTGCCGAGCTGATC
CGGCTTGCGCTCACCACCCGCATCAAACCGTTCACCTACGTGTCGACGATC
GGCGTCGGCGCCGGTATCGAACCGGGCCGTTTCACCGAGGACGACGACATC
CGGGTGATCAGCCCGACGCGGGCCGTCGACACCGGGTACGCCAACGGCTAC
GGCAACAGCAAGTGGGCCGGTGAGGTGTTGTTGCGCGAGGCGCACGATCTG
TGCGGGCTCCCCGTGGCGGTGTTCCGGTGCGACATGATCCTGGCCGACACC
ACCTACGCCGGCCAGCTCAACCTGCCGGACATGTTCACCCGGATGATGCTG
AGCCTGGTGACCACGGGTATCGCGCCGAAATCGTTCCACCCACTCGACGCA
AAGGGCCACCGGCAGAGCGCCCACTACGACGGGCTGCCGGTCGAGTTCGTC
GCCGAATCGATCTCCGCGCTGGGAGCGCAGGCGGTCGACGAGGCGGGAACG
GGTTTCGCCACCTACCACGTGATGAACCCCCACGACGACGGGATCGGCCTC
GACGAATTCGTCGACTGGCTCGTCGAGGCGGGGTATCGCATCGACCGCATC

FIG. 8 (Cont.)

GACTACTACGCGGCCTGGCTGCAGCGGTTCGAAACCGCCCTGCGGGCGCTG
CCCGAGCGCACTCGGCAGTACTCACTGCTCCCGCTGCTGCACAACTACCAG
CGGCCTGCGCACCCGATCAACGGGGCGATGGCCCCGACCGACCGGTTCCGC
GCTGCGGTGCAGGAGGCAAAGCTCGGCCCGGACAAGGACATTCCCCACGTC
ACTCCAGCGGTGATCGTCAAGTACGCCACCGACCTGGAGCTGCTGGGCCTG
ATCTAG

Amino acid sequence (SEQ ID NO:48)

>uniprot|A1UFA8|A1UFA8_MYCSK Thioester reductase domain

MSTETREARLQQRIAHLFTTDPQFAAARPDPRISDAVDRDDTRLTAIVSA
VMSGYADRPALGQRAAEFVTDPQTGRTTMELLPRFDTITYRELLDRVRAL
TNAWHADGVRPGDRVAILGFTGIDYTVVDLALIQLGAVAVPLQTSAAVEA
LRPIVAETEPMLIATGVDHVDAAAELALTGHRPSRVVVFDHREQVDDERD
AVRAATARLGDAVPVETLAEVLRRGAHLPAVAPHVFDEADPLRLLIYTSG
SAGAPKGAMYPESKVAGMWRASAKAAWNNDQTAIPSITLNFLPMSHVMGR
GLLCGTLSTGGTAYFAARSDLSTLLEDLRLVRPTQLSFVPRIWDMLFQEF
VGEVDRRVNDGADRPTAEADVLAVQRHELLGGRFVTAMTGSAPISLEMKT
WVETLLDMHLVEGYGSTEAGAVFVDGHIQRPPVLDYKLVDVPDLGYFSTD
RPHPRGELLVRSTQLFPGYYKRPDVTAEVFDDDGFYRTGDIVAEVGPDQV
QYLDRRNNVLKLAQGEFVTISKLEAVFAGSALVRQIYVYGNSARSYLLAV
VVPTDDAVARHDPASLKTAISASLQQAAKTAGLQSYELPRDFLVETQPFT
LENGLLTGIRKLARPKLKARYGDRLEALYVELVEGQAGELRTLRRDGAKR
PVAETVGRAAAALLGAAAADVRPDAHFTDLGGDSLSALTFGNLLQEIFGV
DVPVGVIVSPAADLASIAAYIEAEQASTGKRPTYASVHGRDAEQVHARDL
TLDKFIDAETLSAATELPGPSGEVRTVLLTGATGFLGRYLALDWLERMAL
VDGKVICLVRAKDDAAARKRLDDTFDSGDPKLLAHYRKLAADHLEVLAGD
KGEADLGLPHPVWQRLADTVDLIVDPAALVNHVLPYSQLFGPNALGTAEL
IRLALTTRIKPFTYVSTIGVGAGIEPGRFTEDDDIRVISPTRAVDTGYAN
GYGNSKWAGEVLLREAHDLCGLPVAVFRCDMILADTTYAGQLNLPDMFTR
MMLSLVTTGIAPKSFHPLDAKGHRQSAHYDGLPVEFVAESISALGAQAVD
EAGTGFATYHVMNPHDDGIGLDEFVDWLVEAGYRIDRIDYYAAWLQRFET
ALRALPERTRQYSLLPLLHNYQRPAHPINGAMAPTDRFRAAVQEAKLGPD
KDIPHVTPAVIVKYATDLELLGLI

FIG. 8 (Cont.)

ZP_05217435

Nucleotide sequence (SEQ ID NO:49)

>gi|222089526:2534-6055 Mycobacterium avium subsp. avium ATCC 25291 NZ_ACFI01000138, whole genome shotgun sequence

ATGTCGACTGCCACCCATGACGAACGACTCGACCGTCGCGTCCACGAACTC
ATCGCCACCGACCCGCAATTCGCCGCCGCCCAACCCGACCCGGCGATCACC
GCCGCCCTCGAACAGCCCGGGCTGCGGCTGCCGCAGATCATCCGCACCGTG
CTCGACGGCTACGCCGACCGGCCGGCGCTGGGACAGCGCGTGGTGGAGTTC
GTCACGGACGCCAAGACCGGGCGCACGTCGGCGCAGCTGCTCCCCCGCTTC
GAGACCATCACGTACGGCGAAGTGGCGCAGCGTGTTTCGGCGCTGGGCCGC
GCCCTGTCCGACGACGCGGTGCACCCCGGCGACCGGGTGTGCGTGCTGGGC
TTCAACAGCGTCGACTACGCCACCATCGACATGGCGCTGGGCGCCATCGGC
GCCGTCTCGGTGCCGCTGCAGACCAGCGCGGCAATCAGCTCGCTGCAGCCG
ATCGTGGCCGAGACCGAGCCCACCCTGATCGCGTCCAGCGTGAACCAGCTG
TCCGACGCGGTGCAGCTGATCACCGGCGCCGAGCAGGCGCCCACCCGGCTG
GTGGTGTTCGACTACCACCCGCAGGTCGACGACCAGCGCGAGGCCGTCCAG
GACGCCGCGGCGCGGCTGTCCAGCACCGGCGTGGCCGTCCAGACGCTGGCC
GAGCTGCTGGAGCGCGGCAAGGACCTGCCCGCCGTCGGGGAGCCGCCCGCC
GACGAGGACTCGCTGGCCCTGCTGATCTACACCTCCGGGTCCACCGGCGCC
CCCAAGGGCGCGATGTACCCGCAGAGCAACGTCGGCAAGATGTGGCGCCGC
GGCAGCAAGAACTGGTTCGGCGAGAGCGCCGCGTCGATCACCCTGAACTTC
ATGCCGATGAGCCACGTGATGGGCCGAAGCATCCTCTACGGCACGCTGGGC
AACGGCGGCACCGCCTACTTCGCCGCCCGCAGCGACCTGTCCACCCTGCTC
GAGGACCTCGAGCTGGTGCGGCCCACCGAGCTCAACTTCGTCCCGCGGATC
TGGGAGACGCTGTACGGCGAATTCCAGCGTCAGGTCGAGCGGCGGCTCTCC
GAGTCCGGGGACGCCGGCGAACGTCGCGCCGTCGAGGCCGAGGTGCTGGCC
GAGCAGCGCCAGTACCTGCTGGGCGGGCGGTTCACCTTCGCGATGACGGGC
TCGGCGCCCATCTCGCCGGAGCTGCGCAACTGGGTCGAGTCGCTGCTCGAA
ATGCACCTGATGGACGGCTACGGCTCCACCGAGGCCGGAATGGTGTTGTTC
GACGGGGAGATTCAGCGCCCGCCGGTGATCGACTACAAGCTGGTCGACGTG
CCGGACCTGGGCTACTTCAGCACCGACCGGCCGCATCCGCGCGGCGAGCTG
CTGCTGCGCACCGAGAACATGTTCCCGGGCTACTACAAGCGGGCCGAAACC
ACCGCGGGCGTCTTCGACGAGGACGGCTACTACCGCACCGGCGACGTGTTC
GCCGAGATCGCCCCGGACCGGCTGGTCTACGTCGACCGCCGCAACAACGTG
CTCAAGCTGGCGCAGGGCGAATTCGTCACGCTGGCCAAGCTGGAGGCGGTG

FIG. 8 (Cont.)

TTCGGCAACAGCCCGCTGATCCGCCAGATCTACGTCTACGGCAACAGCGCC
CAGCCCTACCTGCTGGCGGTCGTGGTGCCCACCGAGGAGGCGCTGGCCTCG
GGTGACCCCGAGACGCTCAAGCCCAAGATCGCCGACTCGCTGCAGCAGGTC
GCCAAGGAGGCCGGCCTGCAGTCCTACGAGGTGCCGCGCGACTTCATCATC
GAGACCACCCCGTTCAGCCTGGAAAACGGTCTGCTGACCGGGATCCGGAAG
CTGGCGTGGCCGAAACTGAAGCAGCACTACGGGGAACGGCTGGAGCAGATG
TACGCCGACCTGGCCGCCGGACAGGCCAACGAGCTGGCCGAGCTGCGCCGC
AACGGTGCCCAGGCGCCGGTGCTGCAGACCGTGAGCCGCGCCGCGGGCGCC
ATGCTGGGTTCGGCCGCCTCCGACCTGTCCCCCGACGCCCACTTCACCGAT
CTGGGCGGAGACTCGTTGTCGGCGTTGACATTCGGCAACCTGCTGCGCGAG
ATCTTCGACGTCGACGTGCCGGTGGGCGTCATCGTCAGCCCGGCCAACGAC
CTGGCGGCCATCGCGAGCTACATCGAGGCCGAGCGGCAGGGCAGCAAGCGC
CCGACGTTCGCCTCGGTGCACGGCCGGGACGCGACCGTGGTGCGCGCCGCC
GACCTGACGCTGGACAAGTTCCTCGACGCCGAGACGCTGGCCGCCGCGCCG
AACCTGCCCAAGCCGGCCACCGAGGTGCGCACCGTGCTGCTGACCGGCGCC
ACCGGCTTCCTGGGCCGCTACCTGGCCCTGGAATGGCTGGAGCGGATGGAC
ATGGTGGACGGCAAGGTCATCGCCCTGGTCCGGGCCCGCTCCGACGAGGAG
GCACGCGCCCGGCTGGACAAGACCTTCGACAGCGGCGACCCGAAGCTGCTC
GCGCACTACCAGCAGCTGGCCGCCGATCACCTGGAGGTCATCGCCGGCGAC
AAGGGCGAGGCCAATCTGGGCCTGGGCCAAGACGTTTGGCAACGACTGGCC
GACACGGTCGACGTGATCGTCGACCCCGCCGCGCTGGTCAACCACGTGTTG
CCGTACAGCGAGCTGTTCGGGCCCAACGCCCTGGGCACCGCGGAGCTGATC
CGGCTGGCGCTGACGTCCAAGCAGAAGCCGTACACCTACGTGTCCACCATC
GGCGTGGGCGACCAGATCGAGCCGGGCAAGTTCGTCGAGAACGCCGACATC
CGGCAGATGAGCGCCACCCGGGCGATCAACGACAGCTACGCCAACGGCTAC
GGCAACAGCAAGTGGGCCGGCGAGGTGCTGCTGCGCGAGGCGCACGACCTG
TGCGGGCTGCCCGTCGCGGTGTTCCGCTGCGACATGATCCTGGCCGACACC
ACGTATGCCGGGCAGCTCAACCTGCCGGACATGTTCACCCGGCTGATGCTG
AGCCTGGTGGCCACCGGGATCGCGCCCGGCTCGTTCTACGAGCTCGACGCC
GACGGCAACCGGCAGCGGGCGCACTACGACGGCCTGCCGGTCGAGTTCATC
GCCGCGGCGATCTCGACGCTGGGTTCGCAGATCACCGACAGCGACACCGGC
TTCCAGACCTACCACGTGATGAACCCCTACGATGACGGCATCGGTCTGGAC
GAGTACGTCGATTGGCTGGTGGACGCCGGCTATTCGATCGAGCGGATTGCC
GACTACTCCGAATGGCTGCGGCGGTTCGAGACCTCGCTGCGGGCCCTGCCG
GACCGGCAGCGCCAGTACTCGCTGCTGCCGCTGCTGCACAACTACCGCACG
CCGGAGAAGCCGATCAACGGGTCGATAGCTCCCACCGACGTGTTCCGGGCA
GCGGTGCAGGAGGCGAAAATCGGCCCCGACAAAGACATTCCGCACGTGTCG
CCGCCGGTCATCGTCAAGTACATCACCGACCTGCAGCTGCTCGGGCTGCTC
TGA

FIG. 8 (Cont.)

Amino acid sequence (SEQ ID NO:50)

>gi|254775919|ref|ZP_05217435.1| FadD9 [Mycobacterium avium subsp. avium ATCC 25291]

MSTATHDERLDRRVHELIATDPQFAAAQPDPAITAALEQPGLRLPQIIRTV
LDGYADRPALGQRVVEFVTDAKTGRTSAQLLPRFETITYGEVAQRVSALGR
ALSDDAVHPGDRVCVLGFNSVDYATIDMALGAIGAVSVPLQTSAAISSLQP
IVAETEPTLIASSVNQLSDAVQLITGAEQAPTRLVVFDYHPQVDDQREAVQ
DAAARLSSTGVAVQTLAELLERGKDLPAVGEPPADEDSLALLIYTSGSTGA
PKGAMYPQSNVGKMWRRGSKNWFGESAASITLNFMPMSHVMGRSILYGTLG
NGGTAYFAARSDLSTLLEDLELVRPTELNFVPRIWETLYGEFQRQVERRLS
ESGDAGERRAVEAEVLAEQRQYLLGGRFTFAMTGSAPISPELRNWVESLLE
MHLMDGYGSTEAGMVLFDGEIQRPPVIDYKLVDVPDLGYFSTDRPHPRGEL
LLRTENMFPGYYKRAETTAGVFDEDGYYRTGDVFAEIAPDRLVYVDRRNNV
LKLAQGEFVTLAKLEAVFGNSPLIRQIYVYGNSAQPYLLAVVVPTEEALAS
GDPETLKPKIADSLQQVAKEAGLQSYEVPRDFIIETTPFSLENGLLTGIRK
LAWPKLKQHYGERLEQMYADLAAGQANELAELRRNGAQAPVLQTVSRAAGA
MLGSAASDLSPDAHFTDLGGDSLSALTFGNLLREIFDVDVPVGVIVSPAND
LAAIASYIEAERQGSKRPTFASVHGRDATVVRAADLTLDKFLDAETLAAAP
NLPKPATEVRTVLLTGATGFLGRYLALEWLERMDMVDGKVIALVRARSDEE
ARARLDKTFDSGDPKLLAHYQQLAADHLEVIAGDKGEANLGLGQDVWQRLA
DTVDVIVDPAALVNHVLPYSELFGPNALGTAELIRLALTSKQKPYTYVSTI
GVGDQIEPGKFVENADIRQMSATRAINDSYANGYGNSKWAGEVLLREAHDL
CGLPVAVFRCDMILADTTYAGQLNLPDMFTRLMLSLVATGIAPGSFYELDA
DGNRQRAHYDGLPVEFIAAAISTLGSQITDSDTGFQTYHVMNPYDDGIGLD
EYVDWLVDAGYSIERIADYSEWLRRFETSLRALPDRQRQYSLLPLLHNYRT
PEKPINGSIAPTDVFRAAVQEAKIGPDKDIPHVSPPVIVKYITDLQLLGLL

YP 001070587.1

Nucleotide sequence (SEQ ID NO:51)

>uniprot|A3PYW9|A3PYW9_MYCSJ Thioester reductase domain

ATGTCCACCGAGACCCGTGAAGCGCGCCTGCAGCAGCGCATCGCGCATCTG
TTCGCCACCGACCCGCAGTTCGCCGCCGCCCGGCCCGACCCCCGGATCAGC
GATGCCGTCGACCGCGATGACGCGCGGCTGACCGCCATCGTCTCCGCCGTC
ATGTCGGGGTACGCCGACCGGCCGGCACTCGGGCAGCGCGCCGCCGAATTC

FIG. 8 (Cont.)

GCCACCGACCCGCAGACCGGCCGCACCACGATGGAGCTGCTCCCCCGCTTC
GACACCATCACCTACCGGGAGCTGCTCGACCGCGTCCGGGCGCTCACCAAC
GCGTGGCACGCCGACGGGGTGCGCCCCGGCGACCGCGTCGCAATCCTCGGA
TTCACCGGCATCGACTACACCGTCGTCGACCTCGCCCTGATCCAGCTCGGC
GCGGTCGCGGTGCCGCTGCAGACCAGCGCCGCCGTGGAGGCGCTGCGCCCG
ATCGTGGCCGAGACCGAACCGATGCTCATCGCCACCGGCGTCGATCATGTC
GACGCCGCCGCGGAACTCGCACTCACCGGCCACCGTCCGTCCCAGGTGGTG
GTCTTCGACCACCGCGAGCAGGTCGACGACGAACGCGACGCGGTGCGGGCC
GCGACGGCCCGGCTGGGAGACGCGGTGCCCGTCGAGACACTCGCCGAGGTG
TTGCGGCGTGGCGCCCATCTGCCCGCCGTGGCACCGCACGTGTTCGACGAG
GCCGATCCACTGCGGCTGCTGATCTACACCTCCGGCAGCACCGGCGCCCCC
AAGGGCGCGATGTATCCCGAGAGCAAGGTCGCCGGCATGTGGCGCGCGTCG
GCCAAGGCCGCCTGGAACAACGATCAGACGGCGATTCCGTCGATCACCCTG
AACTTCCTGCCGATGAGCCACGTCATGGGTCGCGGCCTGCTGTGCGGCACG
CTCAGCACCGGTGGCACCGCGTATTTCGCCGCCCGCAGCGATCTGTCGACG
CTGCTCGAGGACCTGCGCCTGGTACGGCCCACCCAACTCAGCTTCGTGCCA
CGGATCTGGGACATGCTCTTCCAGGAGTTCGTCGGCGAGGTCGACCGGCGG
GTGAACGACGGTGCAGACCGCCCCACCGCGGAGGCCGACGTGCTGGCCGAA
CTGCGCCAGGAGCTGCTCGGTGGCCGGTTCGTCACCGCGATGACCGGTTCG
GCGCCCATCTCCCCCGAGATGAAGACATGGGTGGAGACCCTGCTCGACATG
CACCTGGTCGAGGGTTACGGCTCGACGGAGGCCGGCGCGGTGTTCGTCGAC
GGCCACATCCAGCGCCCGCCGGTGCTCGACTACAAACTCGTCGACGTCCCC
GACCTCGGCTACTTCAGCACCGACCGGCCGCACCCGCGCGGTGAGCTGCTG
GTCCGCTCCACGCAGCTGTTCCCCGGCTACTACAAGCGTCCCGACGTCACC
GCCGAGGTGTTCGACGACGACGGCTTCTACCGCACGGGCGACATCGTCGCC
GAGCTCGGCCCCGACCAGCTGCAGTACCTCGACCGCCGCAACAACGTGCTC
AAACTCGCCCAGGGTGAGTTCGTCACCATCTCCAAACTCGAGGCGGTCTTC
GCCGGCAGCGCCCTCGTCCGCCAGATCTTCGTGTACGGCAACAGTGCGCGC
TCCTACCTGCTGGCCGTCGTCGTGCCGACCGACGATGCGGTGGCCCGGCAC
GACCCGGCATCGCTCAAGACCGCGATCAGCGCCTCGCTGCAGCAGGCCGCG
AAGACCGCCGGTCTGCAGTCCTACGAGCTGCCGCGTGACTTCCTCGTCGAG
ACTCAACCGTTCACGCTGGAGAACGGACTACTGACCGGCATCCGCAAGCTG
GCCCGCCCGAAACTCAAGGCGCGCTACGGCGATCGGCTCGAGGCGCTCTAC
GTCGAACTGGCCGAAGGACAGGCAGGCGAACTGCGCACCCTGCGCCGGGAC
GGCGCGAAGCGTCCGGTGGCCGAGACGGTCGGCCGCGCCGCGGCCGCGCTG
CTCGGCGCCGCCGCCGCCGACGTGCGCCCCGACGCGCACTTCACCGACCTC
GGCGGAGACTCGTTGTCGGCGTTGACCTTCGGCAATCTGCTGCAGGAGATC
TTCGGCGTCGACGTCCCGGTCGGGGTGATCGTCAGCCCGGCGGCCGACCTG
GCGTCGATCGCGGCGTACATCGAGACCGAACAGGCCTCGACCGGCAAGCGG

FIG. 8 (Cont.)

CCGACCTACGCGTCGGTGCACGGGCGCGACGCCGAACAGGTACGCGCGCGC
GACCTCACCCTGGACAAGTTCATCGACGCCGAAACACTCTCCGCTGCAACA
GAACTGCCCGTCCCGATCGGTGAGGTGCGCACCGTGCTGCTGACCGGGGCC
ACCGGATTCCTCGGCCGCTACCTGGCGCTGGACTGGCTCGAAAGGATGGCC
CTGGTCGACGGCAAGGTCATCTGCCTGGTCCGCGCGAAGGACGATGCGGCC
GCCCGCAAACGCCTCGACGACACCTTCGACAGCGGCGATCCGAAGCTGCTG
GCGCACTACCGCAAGCTGGCCGCCGACCACCTCGAAGTGCTGGCCGGCGAC
AAGGGTGAGGCGGATCTCGGGCTGCCGCATCAGGTCTGGCAGCGACTGGCC
GACACCGTCGACCTCATCGTCGACCCGGCCGCCCTGGTCAACCACGTACTG
CCGTACAGCCAGCTGTTCGGGCCCAACGCGCTGGGCACTGCCGAGCTGATC
CGGCTTGCGCTCACCACCCGCATCAAACCGTTCACCTACGTGTCGACGATC
GGCGTCGGCGCCGGTATCGAACCGGGCCGTTTCACCGAGGACGACGACATC
CGGGTGATCAGCCCGACGCGGGCCGTCGACACCGGGTACGCCAACGGCTAC
GGCAACAGCAAGTGGGCCGGTGAGGTGTTGTTGCGCGAGGCGCACGATCTG
TGCGGGCTCCCCGTGGCGGTGTTCCGGTGCGACATGATCCTGGCCGACACC
ACCTACGCCGGCCAGCTCAACCTGCCGGACATGTTCACCCGGATGATGGTG
AGCCTGGTGACCACGGGTATCGCGCCGAAGTCGTTCCACCCACTCGACGCA
AAGGGCCACCGGCAGCGCGCCCACTACGACGGGCTGCCGGTCGAGTTCGTC
GCCGAATCGATCTCCGCGCTGGGAGCGCAGGCGGTCGACGAGGCGGGAACG
GGTTTCGCCACCTACCACGTGATGAACCCCCACGACGACGGGATCGGCCTC
GACGAATTCGTCGACTGGCTCGTCGAGGCGGGGTATCGCATCGACCGCATC
GACGACTACGCGGCCTGGCTGCAACGGTTCGAAACCGCGCTGCGGGCGCTG
CCCGAGCGCACTCGGCAGTACTCACTGCTCCCGCTGCTGCACAACTACCAG
CGGCCTGCGCACCCGATCAACGGGGCGATGGCCCCGACCGACCGGTTCCGC
GCCGCGGTGCAGGAGGCAAAGCTCGGCCCGGACAAGGACATTCCCCACGTC
ACCCCCGGGGTGATCGTCAAGTACGCCACCGACCTGGAACTGCTGGGGCTG
ATCTAG

Amino acid sequence (SEQ ID NO:52)

>uniprot|A3PYW9|A3PYW9_MYCSJ Thioester reductase domain

MSTETREARLQQRIAHLFATDPQFAAARPDPRISDAVDRDDARLTAIVSA
VMSGYADRPALGQRAAEFATDPQTGRTTMELLPRFDTITYRELLDRVRAL
TNAWHADGVRPGDRVAILGFTGIDYTVVDLALIQLGAVAVPLQTSAAVEA
LRPIVAETEPMLIATGVDHVDAAAELALTGHRPSQVVVFDHREQVDDERD
AVRAATARLGDAVPVETLAEVLRRGAHLPAVAPHVFDEADPLRLLIYTSG
STGAPKGAMYPESKVAGMWRASAKAAWNNDQTAIPSITLNFLPMSHVMGR
GLLCGTLSTGGTAYFAARSDLSTLLEDLRLVRPTQLSFVPRIWDMLFQEF

FIG. 8 (Cont.)

VGEVDRRVNDGADRPTAEADVLAELRQELLGGRFVTAMTGSAPISPEMKT
WVETLLDMHLVEGYGSTEAGAVFVDGHIQRPPVLDYKLVDVPDLGYFSTD
RPHPRGELLVRSTQLFPGYYKRPDVTAEVFDDDGFYRTGDIVAELGPDQL
QYLDRRNNVLKLAQGEFVTISKLEAVFAGSALVRQIFVYGNSARSYLLAV
VVPTDDAVARHDPASLKTAISASLQQAAKTAGLQSYELPRDFLVETQPFT
LENGLLTGIRKLARPKLKARYGDRLEALYVELAEGQAGELRTLRRDGAKR
PVAETVGRAAAALLGAAAADVRPDAHFTDLGGDSLSALTFGNLLQEIFGV
DVPVGVIVSPAADLASIAAYIETEQASTGKRPTYASVHGRDAEQVRARDL
TLDKFIDAETLSAATELPVPIGEVRTVLLTGATGFLGRYLALDWLERMAL
VDGKVICLVRAKDDAAARKRLDDTFDSGDPKLLAHYRKLAADHLEVLAGD
KGEADLGLPHQVWQRLADTVDLIVDPAALVNHVLPYSQLFGPNALGTAEL
IRLALTTRIKPFTYVSTIGVGAGIEPGRFTEDDDIRVISPTRAVDTGYAN
GYGNSKWAGEVLLREAHDLCGLPVAVFRCDMILADTTYAGQLNLPDMFTR
MMVSLVTTGIAPKSFHPLDAKGHRQRAHYDGLPVEFVAESISALGAQAVD
EAGTGFATYHVMNPHDDGIGLDEFVDWLVEAGYRIDRIDDYAAWLQRFET
ALRALPERTRQYSLLPLLHNYQRPAHPINGAMAPTDRFRAAVQEAKLGPD
KDIPHVTPGVIVKYATDLELLGLI

CAR70557

Nucleotide sequence (SEQ ID NO:53)

>gi|219932450:561817-562104 Mycobacterium leprae Br4923, complete genome sequence
ATGAGCACCGTGGATGGTGTGTGTGATGAAGACATATATTACTATTATGAC
TGCATATATTGCGCTGCCGATGTTGCCAGCCCGACTGGGTATTTACCCGGA
GGACTCTTTAGCTTGGCGAAGCTGCTTGTTATACGTGGGGAGAAGCAGTAT
GTCTTCGGCCTGAATGACTCGACGAGTGTCGGCCCCGGGCCGGGCAACATG
CCCTGGCTGGACGACCCAGGCCTGTCTGCTGTCATTGCGAGTTCCGTCGCA
GCGGCTGAATTGGCTGCTGCGCGATCCTGGTAA

Amino acid sequence (SEQ ID NO:54)

>gi|219932734|emb|CAR70557.1| hypothetical protein [Mycobacterium leprae Br4923]
MSTVDGVCDEDIYYYYDCIYCAADVASPTGYLPGGLFSLAKLLVIRGEKQY
VFGLNDSTSVGPGPGNMPWLDDPGLSAVIASSVAAAELAAARSW

FIG. 8 (Cont.)

YP 001220863.1

Nucleotide sequence (SEQ ID NO:55)

>uniprot|A5CM59|A5CM59_CLAM3 Putative acyl-CoA synthetase

ATGAGCACTGAGCAGATGGGCACCGAGCAGATGGGGAGCCAGCACGAGGAC
ACGTCGATCGAGGCGATCTTCGCGCAGCACGCCGACCGGACGGCGCTGCGG
CAGCGATCCGGGCCGGACATCACGGACATGGGCTTCCGGGAGCTGTGGGAC
CGGGCGGGCGCGCTGGCCGCGGCGCTCGGCGAGACCGTGTCCGCCGGAGAC
CGCATCGCGGTGCTCGGCACGGCGACCGCGGACGCCGTCACGCTCGACCTC
GCCGCGTGGATCCTCGGCGCGGTGAGCGTGCCCCTCCAGGCGAGCGCGCCC
GTCGCGGCCCTGCGCGCGATCGTCGAGGAGACGACGCCGGTGTGGATCGCG
GCCACCGCCGACCAGGCCGCGACGGCCCGGGCGGTCGCCGAGGCATCGGGC
GACGGCATCCGGACGATGCGGCTCGACACCGACACCGACGCGGACACCGAC
ACCGACGCCGCCCTGACGCTCGGCGCGCTCGTCGCCCGCGGGGCCGGCCTG
CGTCGCCGGAGCCCGTGGCACCCCGCACCCGGCGACGACCCGCTCGCGCTC
CTGCTCTACACGTCGGGCAGCACGGGCACGCCGAAGGGCGCGATGTACACG
CGCTCCATGGTCGAGCGGATGTGGCACGCCCTCCGGCCCGACCCCGCCGCG
CCCGCCGACGCATCCACGACCGCGGACGACGGAGACGCCGCAGCCATCGTC
GGGTACGCGTACCTCCCCATGAGCCACCTCACCGGCCGCTCCTCCCTGCTC
GCCACGCTCGGCCGCGGCGGCACGGTCGCGCTCGCGACCTCGACCGACCTG
TCGACGCTCTTCGACGACCTGCGGACCTTCGCCCCGACCGAGTTCGTCTTC
GTGCCGCGCGTCGCGGAGCTGGTGCGGCAGGAGGGCGACCGCGAGGAGCAG
CGCCGACTCACCGCCGGCAGCACGGACCGGGACGCGGTCCGCGCCGAGGTC
CAGGCCGACCTGCGCGCCCGCGCGTTCGGCGGACGGATCCACCGGGCGATC
TGCACCAGCGCACCGCTGACGCCGGAGCTCCGCACGTACATCGAGGGATGC
CTCGGGCTGACGCTGCACGACCTGTACGGGTCGACCGAGGCCGGCGGCATC
CTCCACGACGGGGTGATCCAGCAGCCTCCCGTCACCGAGCACAAGCTCGTC
GACGTCCCCGAGCTCGGGTACCGCACCACCGACCGCCCGCACCCGCGGGGC
GAGCTGCTCGTCAAGAGCGCCTCCGTGATCGCCGGGTACTTCCGCCGGCCG
GACGTCACCGCCGCGGTGTTCGACGAGGACGGCTTCTACCGGACCGGCGAC
GTCATGGCGCAGACCGGACCCGGCACCTACGAGTACGTCGACCGTCGGAAC
AACGTGATCAAGCTGTCGCAGGGCGAGTTCGTCGCGGTGGCATCGCTCGAG
GCGACGTACGGCGGGACGCCCGAGGTGCACCAGATCGCGCTGCACGGCGAC
AGCCGGCACGCGTTCCTCGTCGCGGTCGTCGTCCCGGCGGACCCCGCGGCG
TCGGAGCGCGACATCCTCGCGGCCCTCCAGCGCACCGCCCGCGAGCACGGC
CTCGCCCCCTACGAGGTGCCCCGCGGCGTGATCGTCGAACCGGATCCGTTC
ACGGTCGACGGCGGCATGCTCTCCGACGCCGGCAAGCTCCTGCGCCTGCGC

FIG. 8 (Cont.)

CTCACGCAGCGGTACGGCGAGCGCCTCGCCGCCCTCTACGACGCGCTCGAG
GAGCAGCAGAGCGGCACCCTCGTCGCCGCGCTCCGCGAACGCGCCGACGAC
GAGCCGACGGTCGACACGGTGGTGCGCGCCGCCCTCCTGCTCCTCGGGGCC
GAGGTGTCCCCCGCGACCGCCGCCGCTGCCCGGTTCTCCGACCTCGGCGGC
GACTCGCTGTCCGCGCTGACGTTCTCCGGGATCCTCGAGGACGTCTTCGGC
ACCGAGGTGCCCGTCGGCGTCCTCACCGACCCGACCAACGACCTCGCCGCC
GTCGCCGCCTACGTGGAGCGCTCCGCCTCCGACGACCGGCCGACCGTGACC
CGCGTGCACGGCGCCGGCGCATCCACCCTCCGCGTGGGCGACCTCCGGCTC
GACCGGATGCTCGGCGGCATCCCGACGCCGGTCCCCCGGGCCTCGGCCGCC
CGGCCGGGATCCCGCACGGTCCTGCTCACCGGCGCGAACGGCTACCTCGGC
CGGTTCCTGGCCATCGACTGGCTCGAGCGCCTCGCTGCGACGGGCGGCACG
CTGGTGTGCATCGTGCGCGGCGCCGACGACGCCGACGCCCGGCGCCGCCTC
GAGGCCGCGTTCGCCGCGGATCCCGCGTTCGCCCGGCGCTTCGCCGAGCTG
TCGGGCTCGCTCGAGGTGCTGGCCGGAGACGTCAGCGAGCACCGCCTCGGC
CTCGACGACGAGCGGTGGATCGACCTGGCCGCGCGGGTCGACCTCGTCGCG
CACGCGGCTGCCCTCGTGAACCACGTCCTCCCGTACTCGGCGCTGTTCGGT
CCGAACGTCGTCGGCACCGCCGAGGCGATCCGCCTGGCGATCGCCGCCGGC
AGCGTGCCCGTCACCTTCGTCTCGAGCGTCGCGGTCGCGGGCGGCGCGCGG
CCGGGCGCGACCGCCGACGCCGAGCCGTCGGCGCCCGGCGCGCTCGACGAG
CACGCCGACATCCGCGCCACGATCCCCGAGTGGGCCGTCGGCGACGAGTAC
GCCAACGGGTACGGCGCGAGCAAGTGGGCGAGCGAGGTGCTGCTCCGCGAG
GCGCACGAGCACCACGGCGTCCCCGTGGCCGTGTTCCGCTCCGACATGATC
CTGGCGCACCCCCGCTGGCGCGGTCAGGTGAACCTCCCCGACGTCTTCACC
CGGCTGATCTGGAGCGTGCTCACCACCGGCCTCGCCCCCGCATCGTTCGTG
AGGCGCGGCCCCGACGGCGAGCGGCAGCGGTCGCACTACGACGGGCTGCCG
GCCGACTTCACGGCGGCGGCGATCGACGGGATCGGCGCGGCGCTCACCGAG
GGGCACCGCACCTTCAACGTCGTGAACCCCCACGACGACGGCGTCTCGCTC
GACACCTTCGTCGACTGGATCCGCGAGGACGGCCACGACATCGCGCGCGTG
GACGACCACGCGGAGTGGGTCGACCGGTTCCGCGCGGCGCTGGGAGCGCTC
CCGGACGCGGACCGCGCCCGGTCCGTCCTGCCGCTGATGCACGCGTTCGCC
TCGCCCGAGGAGCCGCACGCCGGCTCGGCGATCCCGGCGGATGCGTTCGCC
GAGGCCGTCCGCGCGGTGCGCCCGCTCGGGTCGCCGGACATCCCGTCTCTC
GACCACGCGCTCATCGCCAAGGTCGCCGACGACCTCGCGTTCCTGGGGCTG
CTCGCGCCGGCGCGGGCGGCGGCTGCCTGA

Amino acid sequence (SEQ ID NO:56)

>uniprot|A5CM59|A5CM59_CLAM3 Putative acyl-CoA synthetase

MSTEQMGTEQMGSQHEDTSIEAIFAQHADRTALRQRSGPDITDMGFRELW

FIG. 8 (Cont.)

DRAGALAAALGETVSAGDRIAVLGTATADAVTLDLAAWILGAVSVPLQAS
APVAALRAIVEETTPVWIAATADQAATARAVAEASGDGIRTMRLDTDTDA
DTDTDAALTLGALVARGAGLRRRSPWHPAPGDDPLALLLYTSGSTGTPKG
AMYTRSMVERMWHALRPDPAAPADASTTADDGDAAAIVGYAYLPMSHLTG
RSSLLATLGRGGTVALATSTDLSTLFDDLRTFAPTEFVFVPRVAELVRQE
GDREEQRRLTAGSTDRDAVRAEVQADLRARAFGGRIHRAICTSAPLTPEL
RTYIEGCLGLTLHDLYGSTEAGGILHDGVIQQPPVTEHKLVDVPELGYRT
TDRPHPRGELLVKSASVIAGYFRRPDVTAAVFDEDGFYRTGDVMAQTGPG
TYEYVDRRNNVIKLSQGEFVAVASLEATYGGTPEVHQIALHGDSRHAFLV
AVVVPADPAASERDILAALQRTAREHGLAPYEVPRGVIVEPDPFTVDGGM
LSDAGKLLRLRLTQRYGERLAALYDALEEQQSGTLVAALRERADDEPTVD
TVVRAALLLLGAEVSPATAAAARFSDLGGDSLSALTFSGILEDVFGTEVP
VGVLTDPTNDLAAVAAYVERSASDDRPTVTRVHGAGASTLRVGDLRLDRM
LGGIPTPVPRASAARPGSRTVLLTGANGYLGRFLAIDWLERLAATGGTLV
CIVRGADDADARRRLEAAFAADPAFARRFAELSGSLEVLAGDVSEHRLGL
DDERWIDLAARVDLVAHAAALVNHVLPYSALFGPNVVGTAEAIRLAIAAG
SVPVTFVSSVAVAGGARPGATADAEPSAPGALDEHADIRATIPEWAVGDE
YANGYGASKWASEVLLREAHEHHGVPVAVFRSDMILAHPRWRGQVNLPDV
FTRLIWSVLTTGLAPASFVRRGPDGERQRSHYDGLPADFTAAAIDGIGAA
LTEGHRTFNVVNPHDDGVSLDTFVDWIREDGHDIARVDDHAEWVDRFRAA
LGALPDADRARSVLPLMHAFASPEEPHAGSAIPADAFAEAVRAVRPLGSP
DIPSLDHALIAKVADDLAFLGLLAPARAAAA

YP 001537947.1

Nucleotide sequence (SEQ ID NO:57)

>uniprot|A8M8D3|A8M8D3_SALAI Thioester reductase domain

GTGACCACCACGGAGCAGACCCTCACCGAGCGGCTCATCGCCGAGGACGAG
CAGATCCGGCGGGCCCAGGTCAGCGCCGAAGTCTCCGCCGCGATGCGGGTG
CCGGGCATGTCGCAGGCCCAGATCGTGGCCGCCGGATTCACCGGTTACGCC
GACCGCGCCGCCCTGGGTGAGCGCGCCCGCGAAGCCGTCACCGACCCGGTC
ACCGGCCGCACCACCCACCGGCTTCTGCCATGGTTCGACACCATCACCTAC
GGCGAGGTCCGGTCGCGGGTGCTGGCGATCTCCGCCGCCTGGTGGCACGAC
GTGGACGCTCCACTCCGTCCCGGCGCCTTCGTCGTTTCGGTCGGCGTCCCC
AGCGCCGACCTCGTGACGGTCGAGCTCGCGGTGCTACACACCGGCGCGGTC
AGTGTGCCACTGCAGGTCAGCTCCACCGCCGAGCAACTGCGCCCGATCCTC

FIG. 8 (Cont.)

GACGAGGCCGCCCCGCTCATCGTGGCCACGAGCGTGGACCGGCTCGCTGTG
GTGACCGCGGCGATGTCGGGCAACGCGTCGGTGCGCCGGATCATGGTCCTG
AACCACGACGCAGCGATCACCGCCCACCGGGATGCCGTGGACGCCGCGCGA
TCGGCGCTCGCCGGCACCGCAGTCGTCGTGCACACATTGACCGAGGTGTTG
GACCGTGGACGGGGCCTGCCCGCCCCTGAGCCCTACGCGGCGCCCACGGGG
GAGGATCCCCTGTCGCTGTTGATCTACACCTCGGGCAGTACCGGTACGCCC
AAGGGCGCAATGTTTCCGGAGAGCATGACCCGCGCCAACTGGGTGCGTTTC
GACCCCAAGCCGACCGACATGGCGGTCATCCGGCTCAACTACCTGCCGCTG
AGCCACAACGTCGGCCGCATCGTGCTGTTCGAGGCGCTCGCGGTGGGCGGC
ATCGCCTTCTTCACCGCACACAGTGACCTGTCCACGCTCCTGGAGGACATG
GCCCTCGCCCGGCCCACCGACCTGTTCCTTATCCCGCGGCTGTGCGACATG
CTCGCCCAGCGCCACGACAGCGAACTGGCCCGCCGCCGCATCACCACCGCG
GATCACGAGGGGGTCCGACAGGTCCACACCCATCTGCGCGAGGCTGTCCTC
GGCGGCAGGGTGACCCGCGCGATGTCGCTGTCCGCGCCGCTGAGCCCACAG
CTGCGTCGGTTCGTGGAGTCGTGTCTCGGCTTCGCGGTGCACGATGTCTTC
GGGTCGACCGAGGCCGGCGGGCTGCTCGTCAACGGCCGGGTGCTGCGCCCG
CCGGTGCTCGACTATCGCTTGGTCGACGTCCCCGACCTCGGCTACTTCACC
ACCGACCGTCCGTACCCTCGCGGGGAGCTGCTGGTGCGGACCGCGACGATC
ATCCCCGGCTACTACCAGCGGCCCGAGCTCAACGCCGAGCTGTTCACCGAG
GACGGCTACTACCGCACCGGCGACATCATGGCCGAGTACGGCCCCGACCAC
CTCGGCTATGTCGACCGCACCACGAGCGTGCTGAAGCTGTCACAGGGCGAG
TTCGTGGCCGTGTCACGGTTGGAGGAACTGTTCGCCGCCTCCCCGCTGATC
CGGCAGATCTACCTGTACGGCAACAGCGAGCGGCCGTACCTGCTCGCCGTG
GTCGTGCCCACGGAGGAGGCGCACGCCGCCACCCGGGAACCCGCGGCGCTC
AAGGCGGTGCTCGGCGAGTCGCTGCAACGCATCGCTCAGCAGCACGGCCTG
CACCCGTACGAGGTGCCGCGCGACCTCCTCATCGAGACCACCCCGTTCAGC
ACCGCCAACGGTCTGCTCTCCGACATCCGTAAGCCCCTGCGTCCGAAGCTC
AAGACCCGGTACGCTCCTCGACTCGAAGCGCTCTACACCGAGCTCGCCGAG
CGCGAGGCCGACCGGATCCGCACGCTGCGCGACGCCGGTTCCGCGCAACCC
GTGCTGCCCGCGTTGCGCGAGGCTGCCCGGGCGTTCCTCGGCCGCCCAGGC
GCAGCGCTCGACGTGAACGACCGCTTTGTGGACCTCGGCGGCGACTCCCTG
TCGGCCCTGGCCCTGTCGAACCTGCTGAGCGACATCTTCGAGGTCCGCGTC
CCGGTCGGCATCATGATCAGCGCGACCGGCACGCTCGGTTCCGTGGCGGCC
TGGATCGAGGCCGAGCGTGCCACCGCCGGAGCGGGTATCGGCCGCGCGACG
CCCACCTCCGTGCACGGTGCGAACCTCACCCAGGTACACGCCGATGACCTG
ACCCTCGGCACGTTCCTCGACGTGACGACCCTCGCCGCCGCTGCCTGCCTG
CCCCGGGCGCCGCTGTCCGACCCGCGCGTGGTGCTGCTGACGGGTGCGACC
GGCTATCTGGGCCGGTTCCTGGCCCTCGAGTGGCTGGACCGCCTTTCCCGT
AGCGGCGGGACGCTCGTGTGCGTGGTGCGCGCCGCCGACGATGCGGAAGCC

FIG. 8 (Cont.)

GCGCGCCGCCTGGAAAGTGTCTATGGCTCCAGCGACCCCGAGTTGCTGGAG
CGCTTCCGTTCACTCGCCGGCCACGTGCGCGTGTTGGCCGGCGATGTTGCC
GAAGCCAGGTTCGGCCTGCCGGCCGGGGTGTGGCAGGAACTGGCCGAAACG
GTGGACCTGATCGTGCACTCCGCGGCACTGGTCAACCACGTTCTCCCGTAC
GAACAGCTGTTCGGGCCCAACGTGGCGGGAACGGCGGAACTGGTGCGCCTC
GCCGTCAGCGTACGGGTGAAGGGAATTGCCTTTCTCTCCACCGTTGCCGTG
ATCACCTCGCAGACCACGACACCCGACGAGGACGCGGACATCCGGCAGGCG
AGCCCGCACCGGGTGCTCGACGACAGCTACGCGAACGGCTACGCGGCCAGC
AAGTGGGCAGGTGAGGTGCTCCTGCGACGCGCCCACGAGGAGTACGGCGTG
CCGGTCAGCGTGTTCCGCTCGGACGTCATCCTGGCCCACTCCCGCTACGCC
GGGCAGCTCAACGTCCCGGATATGTTCACGCGCCTGCTCCTGAGCATCCTG
GCGACCGGTATTGCCCCAGCGTCGTTTTATCGCACCGGCCCGGACGGCGAA
CGCCAGCCGGCGCACTACGACGGTCTCCCGGTCGACTTCACCGCGGCGGCC
GTAGCCGCGGTGGGTGTCACCGAGGGACACCGCACCTTCAACGTACTGAAT
CCACACGAGGACGGCATCGGGCTGGATACCTTTGTGGACTGGCTCGTGGCA
GCCGGACACCCGGTGCAGCGCATCGCGGACCACGACGAGTGGGTGACCCGC
TTCGCCACGGCCATGCGTGGGCTGCCTGAACGCCAGCGCCGCAGCTCGATC
CTGCCGCTCCTACACGCCTTTGCCGAGCCCGCTCCGCCGACCTTCGGATCC
AGACTGCCGACGGACCGGTTTCGCGCCGCCGTGAAAGCCGCCAACGTGGTC
CCCGGCAACGAGATCCCGCACCTCGATGCGGCCCTCGTCACCAAGTACGCC
GACGACCTCAGGCTGCTCGACCTTCTCTGA

Amino acid sequence (SEQ ID NO:58)

>uniprot|A8M8D3|A8M8D3_SALAI Thioester reductase domain

VTTTEQTLTERLIAEDEQIRRAQVSAEVSAAMRVPGMSQAQIVAAGFTGY
ADRAALGERAREAVTDPVTGRTTHRLLPWFDTITYGEVRSRVLAISAAWW
HDVDAPLRPGAFVVSVGVPSADLVTVELAVLHTGAVSVPLQVSSTAEQLR
PILDEAAPLIVATSVDRLAVVTAAMSGNASVRRIMVLNHDAAITAHRDAV
DAARSALAGTAVVVHTLTEVLDRGRGLPAPEPYAAPTGEDPLSLLIYTSG
STGTPKGAMFPESMTRANWVRFDPKPTDMAVIRLNYLPLSHNVGRIVLFE
ALAVGGIAFFTAHSDLSTLLEDMALARPTDLFLIPRLCDMLAQRHDSELA
RRRITTADHEGVRQVHTHLREAVLGGRVTRAMSLSAPLSPQLRRFVESCL
GFAVHDVFGSTEAGGLLVNGRVLRPPVLDYRLVDVPDLGYFTTDRPYPRG
ELLVRTATIIPGYYQRPELNAELFTEDGYYRTGDIMAEYGPDHLGYVDRT
TSVLKLSQGEFVAVSRLEELFAASPLIRQIYLYGNSERPYLLAVVVPTEE
AHAATREPAALKAVLGESLQRIAQQHGLHPYEVPRDLLIETTPFSTANGL
LSDIRKPLRPKLKTRYAPRLEALYTELAEREADRIRTLRDAGSAQPVLPA

FIG. 8 (Cont.)

LREAARAFLGRPGAALDVNDRFVDLGGDSLSALALSNLLSDIFEVRVPVG
IMISATGTLGSVAAWIEAERATAGAGIGRATPTSVHGANLTQVHADDLTL
GTFLDVTTLAAAACLPRAPLSDPRVVLLTGATGYLGRFLALEWLDRLSRS
GGTLVCVVRAADDAEAARRLESVYGSSDPELLERFRSLAGHVRVLAGDVA
EARFGLPAGVWQELAETVDLIVHSAALVNHVLPYEQLFGPNVAGTAELVR
LAVSVRVKGIAFLSTVAVITSQTTTPDEDADIRQASPHRVLDDSYANGYA
ASKWAGEVLLRRAHEEYGVPVSVFRSDVILAHSRYAGQLNVPDMFTRLLL
SILATGIAPASFYRTGPDGERQPAHYDGLPVDFTAAAVAAVGVTEGHRTF
NVLNPHEDGIGLDTFVDWLVAAGHPVQRIADHDEWVTRFATAMRGLPERQ
RRSSILPLLHAFAEPAPPTFGSRLPTDRFRAAVKAANVVPGNEIPHLDAA
LVTKYADDLRLLDLL

YP 001703694.1

Nucleotide sequence (SEQ ID NO:59)

>uniprot|B1MCR9|B1MCR9_MYCAB Probable fatty-acid-CoA ligase FadD

ATGACCGTGACCAACGAAACCAACCCACAGCAGGAGCAGCTATCCCGCCGT
ATTGAAAGTCTGCGCGAAAGCGATCCGCAGTTCCGGGCGGCCCAGCCCGAC
CCGGCGGTCGCCGAACAGGTGCTGCGCCCGGGCCTGCATCTTTCTGAAGCC
ATTGCGGCGTTGATGACTGGATACGCTGAGCGCCCGGCGCTCGGTGAGCGC
GCACGCGAGTTGGTCACCGACCAGGATGGCCGCACCACGCTGCGCCTGTTG
CCACGCTTCGACACCACCACATACGGCGAATTATGGTCCCGCACAACATCA
GTCGCCGCTGCATGGCACCACGACGCCGCCCACCCGGTTAAGGCCGGCGAT
CTGGTGGCCACCCTGGGATTCACCAGCATCGACTACACCGTGCTGGATCTG
GCGATCATGATCCTCGGTGGCGTGGCGGTTCCGCTACAGACCAGCGCCCCG
GCTTCGCAGTGGACGACCATTCTGGCCGAAGCGGAACCCAACACTCTTGCG
GTAAGCATCGAATTGATCGGCGCTGCAATGGAATCTGTGCGGGCCACGCCT
TCCATCAAGCAGGTCGTCGTGTTCGACTACACCCCCGAGGTCGATGATCAA
CGGGAGGCATTCGAGGCAGCAAGCACACAACTCGCCGGCACCGGCATCGCC
ATTGAGACCCTCGATGCCGTCATCGCCCGCGGCGCCGCACTTCCGGCCGCA
CCGCTCTACGCACCATCGGCCGGCGACGATCCGCTGGCGCTGCTCATCTAC
ACCTCCGGCAGCACCGGGGCTCCAAAGGGCGCCATGCACAGCGAAAACATC
GTGCGCCGCTGGTGGATTCGTGAGGACGTCATGGCCGGCACCGAGAACCTG
CCCATGATCGGGCTGAACTTCATGCCGATGAGTCACATCATGGGACGCGGC
ACCCTCACCTCCACCCTGTCTACCGGTGGAACCGGATACTTCGCGGCGTCC
AGTGACATGTCAACGCTCTTCGAGGACATGGAGCTGATCCGCCCGACGGCC

FIG. 8 (Cont.)

CTGGCCTTGGTTCCACGCGTGTGCGACATGGTGTTCCAGCGATTCCAGACC
GAGGTGGACCGGCGTCTGGCGAGCAGCGACACCGCCAGTGCCGAGGCCGTT
GCGGCCGAGGTCAAGGCCGATATCCGTGACAACCTCTTCGGTGGCCGCGTA
TCGGCGGTCATGGTCGGTTCTGCTCCGTTGTCCGAGGAGCTGGGTGAGTTC
ATCGAATCCTGCTTCGAGCTGAATCTGACCGATGGCTACGGCTCCACCGAA
GCCGGCATGGTGTTCCGCGACGGCATCGTGCAACGCCCGCCGGTCATTGAC
TACAAGCTGGTTGACGTGCCCGAACTGGGCTACTTCTCCACCGACAAGCCG
CACCCGCGCGGTGAGCTGCTGCTGAAGACCGACGGCATGTTCCTCGGGTAC
TACAAACGCCCCGAGGTGACTGCCGGCGTCTTCGACGCGGACGGTTTTTAC
ATGACCGGCGACATCGTCGCCGAGCTGGCCCACGACAACATCGAGATCATC
GATCGCCGCAACAACGTGCTCAAACTCTCACAGGGAGAGTTTGTCGCGGTC
GCCACCTTGGAGGCCGAGTACGCCAATAGCCCTGTGGTGCACCAGATCTAC
GTCTACGGCAGCAGCGAACGGTCCTACCTGCTAGCAGTCGTGGTGCCGACG
CCGGAGGCCGTGGCCGCCGCCAAGGGCGACGCGGCGGCACTCAAGACGACC
ATCGCGGACTCGCTGCAGGACATTGCCAAGGAGATCCAGCTGCAGTCCTAC
GAAGTCCCCCGTGACTTCATCATCGAACCGCAGCCATTCACCCAGGGCAAC
GGCCTGCTGACGGGTATCGCCAAGCTGGCGCGTCCGAACCTGAAGGCGCAC
TATGGACCGCGGCTGGAGCAGATGTACGCCGAAATCGCCGAGCAGCAGGCT
GCCGAGCTTCGGGCGTTGCACGGAGTGGACCCAGACAAGCCCGCGCTGGAA
ACGGTCCTCAAGGCGGCGCAGGCCCTGCTCGGCGTCTCGTCGGCCGAACTG
GCCGCGGACGCGCATTTCACCGATCTAGGTGGCGATTCGCTGTCCGCACTG
TCCTTCTCGGATCTGCTGCGCGATATCTTCGCGGTCGAAGTACCGGTCGGA
GTCATCGTCAGTGCCGCAAACGATCTCAGCGGTGTTGCGAAATTTGTTGAT
GAACAACGCTATTCGGGCGGGACGCGGCCGACCGCGGAGACGGTGCACGGC
GCCGGGCATACGGAGATCCGGGCCGCGGACCTGACCCTGGATAAGTTCATC
GACGAGGCCACCCTGCATGCGGCACCGTCGCTTCCGAAGGCCGTCGGGATC
CCACACACCGTCCTGCTCACCGGGTCCAACGGCTACCTGGGCCACTACCTG
GCACTGGAATGGCTTGAGCGCCTGGACAAGACAGAAGGCAAGCTGATCGCC
ATCGTCCGCGGTAAGAATGCCGAGGCCGCCTACCGCCGCCTCGAGGAAGCC
TTCGACACCGGCGACACGCAGCTGTTGGCGCACTTCCGGTCGCTGGCCGAC
AAGCACCTCGAAGTACTGGCCGGCGATATCGGCGACCCCAACCTTGGCCTG
GATGCCGACACCTGGCAGCGCCTGGCCGACACCGTCGACGTCATCGTGCAC
CCCGCCGCCCTGGTCAACCACGTACTGCCCTACAGCCAGCTGTTCGGACCG
AATGTCGTCGGCACCGCCGAGATCATCAAGCTGGCCATCACTACCAAGATC
AAGCCGGTCACCTACCTGTCCACGGTCGCGGTCGCGGCATATGTCGATCCG
ACGACATTCGACGAAGAGTCCGATATCCGGCTCATCAGCGCGGTGCGTCCC
GTGGACGAGCTGTACGCGAACGGCTACGGCAACAGCAAGTGGGCCGGCGAG
GTACTGCTGCGCGAAGCCCACGATCTGTGCGGACTACCCGTCGCGGTCTTC
CGCTCCGACATGATCTTGGCCCACAGCCGCTACACCGGACAGCTCAACGTG

FIG. 8 (Cont.)

CCCGACCAGTTCACCCGACTAATCCTCAGCCTCATCGCCACCGGAATCGCA
CCCGGCTCCTTCTACCAAGCACACGCCACCGGCGAACGCCCACTCGCCCAC
TACGACGGGCTACCCGGTGACTTCACCGCCGAGGCGATCACCACGTTGGGC
ACCCAGGTGGTCGACAGCTACGAGACCTACGACTGCGTGAACCCGCATGCA
GACGGAGTCTCGCTGGACAACTTCGTCGACTGGCTCATCGAAGCCGGCTAC
CCCATCGCACGCATCGACAACTACACCGAATGGTTCACCCGCTTCGACACC
GCCATCCGAAGCCTCCCCGAAAAACAGAAACAACACTCCCTACTACCACTG
CTCCACGCATTCGAACAGCCGTCCGCCGCCGAGAACCACGGCGTCGTCCCG
GCAAAGCGTTTCCAGCACGCTGTGCAGGCCGCCGGAATCGGTCCGGCCGGG
CAAGACGGCACTACCGACATTCCCCACCTGTCGCGGCGGCTGATCGTGAAA
TACGCCAAGGACCTCGAACAGCTCGGACTCCTATGA

Amino acid sequence (SEQ ID NO:60)

>uniprot|B1MCR9|B1MCR9_MYCAB Probable fatty-acid-CoA ligase FadD

MTVTNETNPQQEQLSRRIESLRESDPQFRAAQPDPAVAEQVLRPGLHLSE
AIAALMTGYAERPALGERARELVTDQDGRTTLRLLPRFDTTTYGELWSRT
TSVAAAWHHDAAHPVKAGDLVATLGFTSIDYTVLDLAIMILGGVAVPLQT
SAPASQWTTILAEAEPNTLAVSIELIGAAMESVRATPSIKQVVVFDYTPE
VDDQREAFEAASTQLAGTGIAIETLDAVIARGAALPAAPLYAPSAGDDPL
ALLIYTSGSTGAPKGAMHSENIVRRWWIREDVMAGTENLPMIGLNFMPMS
HIMGRGTLTSTLSTGGTGYFAASSDMSTLFEDMELIRPTALALVPRVCDM
VFQRFQTEVDRRLASSDTASAEAVAAEVKADIRDNLFGGRVSAVMVGSAP
LSEELGEFIESCFELNLTDGYGSTEAGMVFRDGIVQRPPVIDYKLVDVPE
LGYFSTDKPHPRGELLLKTDGMFLGYYKRPEVTAGVFDADGFYMTGDIVA
ELAHDNIEIIDRRNNVLKLSQGEFVAVATLEAEYANSPVVHQIYVYGSSE
RSYLLAVVVPTPEAVAAAKGDAAALKTTIADSLQDIAKEIQLQSYEVPRD
FIIEPQPFTQGNGLLTGIAKLARPNLKAHYGPRLEQMYAEIAEQQAAELR
ALHGVDPDKPALETVLKAAQALLGVSSAELAADAHFTDLGGDSLSALSFS
DLLRDIFAVEVPVGVIVSAANDLSGVAKFVDEQRYSGGTRPTAETVHGAG
HTEIRAADLTLDKFIDEATLHAAPSLPKAVGIPHTVLLTGSNGYLGHYLA
LEWLERLDKTEGKLIAIVRGKNAEAAYRRLEEAFDTGDTQLLAHFRSLAD
KHLEVLAGDIGDPNLGLDADTWQRLADTVDVIVHPAALVNHVLPYSQLFG
PNVVGTAEIIKLAITTKIKPVTYLSTVAVAAYVDPTTFDEESDIRLISAV
RPVDELYANGYGNSKWAGEVLLREAHDLCGLPVAVFRSDMILAHSRYTGQ
LNVPDQFTRLILSLIATGIAPGSFYQAHATGERPLAHYDGLPGDFTAEAI
TTLGTQVVDSYETYDCVNPHADGVSLDNFVDWLIEAGYPIARIDNYTEWF
TRFDTAIRSLPEKQKQHSLLPLLHAFEQPSAAENHGVVPAKRFQHAVQAA

FIG. 8 (Cont.)

GIGPAGQDGTTDIPHLSRRLIVKYAKDLEQLGLL

YP 001703695.1

Nucleotide sequence (SEQ ID NO:61)

>uniprot|B1MCS0|B1MCS0_MYCAB Probable fatty-acid-CoA ligase FadD

ATGACGATCGACGCCACCGCGGACAACACCAAGGAAGCACGTCGTCAGCGG
TTAGGCGACCGCATCAGGCGCCTATTCACCGACGATGAGCAGTTCCGTGCC
GCCAAGCCCGATACCGCGGTTGATACCGCCGTCGCCCAGCCTGGTCTGCGT
CTCGCCCAGGTGGTCGCCACGATCATGAACGGGTACGCGGACCGTCCGGCG
CTCGGGCACCGAGTCCAGGAGCTCGTCGCCGACGCCGCCGGCCGTTCGACG
CTGCGCCCGTTGCCCGAGTTCGAGACGGTCACCTACGGAGAGCTGTGGGGC
ATGGCCCGCGCGTTGGCCTCCACCTGGTACCACGATCCCGCCGCTCCGGTA
CGGGCCGGGGACTTCGTCGCGATGCTCGGCTTCACCAGCGTGGACTACACC
GCTGTCGACTTGGCATGCATCCACCTTGGCGCGGTGGCGGTTCCATTGCAG
ACCAGCGCATCGGCATCCAACTGGACCGCGATCCTGGCCGAATCGGAACCT
GCCGTCCTGGCGGTAAGCGCCGAGTTGCTCGATACGGCAATGGAATCGGTG
CTCGCCACGCCGTCGCTGCGGCACATCACCGTCTTTGACTATCATCCCGGT
GTCGACGTGCAGCGCGAAAGTCTCGAATCCGCACAGCACCGGATCGCCGAG
GCCGGCCTGCCGATTTCGGTAGACCCGATACCCCTGGCGATCGGGCACGGG
CGCGCCTTGCCGGATGCGCCGTTGTTCACCGCAGAGGAGGGTACCGACCCG
CTGGCCCTGGTGATCTACACCTCGGGAAGCACCGGAACTCCCAAGGGCGCC
ACCTATAGCGAAAAGATGGTCGCCAAGCCCTGGCTGCGGGCCGACACGTTG
AGCTCTAAGGCCGAAATTCCTTTGATCAACCTGAATTTCATGCCAATGAGC
CATGTGATGGGACGCGGTAGTCTGGTCACTGCCCTGGCCTGCGGCGGCCTG
GCTTACTTTGCCGCGTCCAGCGACATGTCCACGCTGTTCGAAGACATCACG
CTTACGCGCCCCACCGTGGTGACACTCGTGCCCCGTGTGTGCGACATGCTT
TTCCAGCGCTACCGCAACGAGGTTGAACGCCGTACCGGGCTTGATCCGGCG
GCCGACCTGGCCACCCTTGATGCCGATGTCAAGACCGATATCCGCGAAAAC
CTGTTCGGCGGGCGTGTTCTGACAATCGTGTGCGGCTCTGCCCCACTGTCC
GAGGAACTGGCCGCCTTCATCGAATCCTGCCTCGATGCCCGTATCACCGAT
GGCTACGGCTCCACCGAGGCGGGCGTCATCGTGCGCAACGGCCGCATTCAG
CGCCCGCCCGTCATCGACTACAAGCTGGTCGACGTGCCTGAGCTCGGTTAC
TTCTCCACCGACAAGCCGCACCCGCGCGGCGAGCTGCTCGTGAAAGCCGAA
TCGGTGTTCGGCGGCTACTTCAAACGCCCCGACGTCACCGCCGACGTATTC
GATCCCGACGGGTACTACAAGACCGGGGACATCGTCGCCGAGCTCGAGCCC
GACAAGATCCAGATCGTGGACCGGCGCAACAACGTGATCAAGCTGTCCCAG

FIG. 8 (Cont.)

GGTGAGTTCGTGGCGATCGCCAACCTGGAAGCCGAGTTCGCCAATAGTCCA
CTGGTGCATCAGATCTGCGTCTACGGCAGCAGCGAGCGGTCGTATCTGCTA
GCGGTGGTCGTGCCGACCGCTGAGGCATATGAACAAAGCGGTGGAGATGAG
GATCTACTCAAACGCCTGATCGCGGACTCTCTCGCGCAGGTTGCCCGCGAG
GCCCAACTGCAGTCCTACGAGGTACCGCGCGACTTCCTGCTGGAGACCGAA
CCGTTCACCGCTGCCAACGGCCTACTGACCGGCATCGCGAAGCTGGCCCGA
CCGAAGCTCCATGAGAAGTACGGCGCCCGCCTGGAGCAGCTGTACTCCGAT
ATCGCCGCCGCCCAGGCGCTTGAGCTGCAAGCACTGCACTCTGCCGGACAT
GAGGACAAGCCTGTCCTGGATACCGTGCAACGCGCGGTCACGGCGTTGTTG
GGGCTGTCGGCGGCCGAGGTGAGCCCAGACGCGCATTTCATCGACCTTGGT
GGCGATTCACTATCCGCCCTTGCCTTCTCGGACCTGCTGCGCGATATCTTC
ACTGTGGAGGTTCCGGTTGGCGATATCGTCAGCGCCGCCAACGATCTGACC
GCTATCGCACGCATCGTGGAAAGACACCGGGAAGCAGACGGTCATTCGGTA
ACTCCCACCGCCGAATCCGTGCACGGTGCCGGGCACCGCGAGATCCGGGCC
GCGGACCTGACGCTGGACAAGTTCATCGACGCGGACACCCTGCGCGCGGCC
CCGGCACTGTCCACATTCACCGGCACCCCGCACACGGTGCTGCTCACCGGC
GCCAACGGCTACCTGGGGCGGTTCCTGGCCCTGGAATGGCTTGAGCGCCTG
GACAAGACAGACGGCAAGCTGATCGCCATCGTCCGCGGTAAGAATGCCGAG
GCCGCCTACCGCCGCCTGGAGGAAGCCTTCGACACCGGCGACACGCAGCTG
TTGGCGCACTTCCGGTCGCTGGCCGACAAGCACCTCGAAGTACTGGCCGGC
GATATCGGCGACCCCAACCTTGGCCTGGATGCCGACACCTGGCAGCGCCTG
GCCGAGACCGTCGACGTCATCGTGCACCCCGCCGCCCTGGTCAACCACGTA
CTGCCCTACAGCCAGCTGTTCGGACCCAATGTCGTTGGCACCGCCGAAATC
ATCAAGCTGGCACTCACCACCAAGATCAAGCCCATCACCTACCTCTCCACA
GTGGCCGTGGCAATCTCGGTGGACCCCAAGGTATTCGATGAAGACTCCGAC
ATCCGCACGATCAGCGCGGTACGACCAATCAACGACGGCTACGCCAACGGA
TACGGCAACGCGAAATGGGCTGGCGAGGTACTGCTGCGCGAAGCCCACGAC
CTGTGCGGACTACCCGTCGCGGTCTTCCGCTCCGACATGATCTTGGCCCAC
AGCCGCTACACCGGACAGCTCAACGTGCCCGACCAGTTCACCCGACTAATC
CTCAGCCTCATCGCCACCGGAGTCGCACCCGGCTCCTTCTACCAAGCACAC
GCCACCGGCGAACGCCCACTCGCCCACTACGACGGCCTGCCTGCGGATTTC
ACGGCATCGGCCATCACCGCCCTCGGGCCCATCGAGGAGTTCCACACCTAC
GATTCGGTGAACCCGCATGCCGATGGGATCTCGCTGGACAACTTCGTCGAC
TGGCTCATCGAAGCCGGCTACCCCATCGCACGCATCGACAACTACACCGAA
TGGTTCACCCGCTTCGACACCGCCATCCGAAGCCTCCCCGAAAAACAGAAA
CAACACTCCCTACTACCACTACTACACGCGTACAGGCATCCACAACACCCA
CACAACGGCGCATTCCTGCCCGCGATCAGGTTCAGTGAAGGCGTCCAGGCC
CATCTGAACGCCGACATCCCGCACCTCACGCGGGAACTCATCGCGAAATAC
GCGGCCGACCTGAAGCAGCTCGGGTTACTCTAG

FIG. 8 (Cont.)

Amino acid sequence (SEQ ID NO:62)

>uniprot|B1MCS0|B1MCS0_MYCAB Probable fatty-acid-CoA ligase FadD

MTIDATADNTKEARRQRLGDRIRRLFTDDEQFRAAKPDTAVDTAVAQPGL
RLAQVVATIMNGYADRPALGHRVQELVADAAGRSTLRPLPEFETVTYGEL
WGMARALASTWYHDPAAPVRAGDFVAMLGFTSVDYTAVDLACIHLGAVAV
PLQTSASASNWTAILAESEPAVLAVSAELLDTAMESVLATPSLRHITVFD
YHPGVDVQRESLESAQHRIAEAGLPISVDPIPLAIGHGRALPDAPLFTAE
EGTDPLALVIYTSGSTGTPKGATYSEKMVAKPWLRADTLSSKAEIPLINL
NFMPMSHVMGRGSLVTALACGGLAYFAASSDMSTLFEDITLTRPTVVTLV
PRVCDMLFQRYRNEVERRTGLDPAADLATLDADVKTDIRENLFGGRVLTI
VCGSAPLSEELAAFIESCLDARITDGYGSTEAGVIVRNGRIQRPPVIDYK
LVDVPELGYFSTDKPHPRGELLVKAESVFGGYFKRPDVTADVFDPDGYYK
TGDIVAELEPDKIQIVDRRNNVIKLSQGEFVAIANLEAEFANSPLVHQIC
VYGSSERSYLLAVVVPTAEAYEQSGGDEDLLKRLIADSLAQVAREAQLQS
YEVPRDFLLETEPFTAANGLLTGIAKLARPKLHEKYGARLEQLYSDIAAA
QALELQALHSAGHEDKPVLDTVQRAVTALLGLSAAEVSPDAHFIDLGGDS
LSALAFSDLLRDIFTVEVPVGDIVSAANDLTAIARIVERHREADGHSVTP
TAESVHGAGHREIRAADLTLDKFIDADTLRAAPALSTFTGTPHTVLLTGA
NGYLGRFLALEWLERLDKTDGKLIAIVRGKNAEAAYRRLEEAFDTGDTQL
LAHFRSLADKHLEVLAGDIGDPNLGLDADTWQRLAETVDVIVHPAALVNH
VLPYSQLFGPNVVGTAEIIKLALTTKIKPITYLSTVAVAISVDPKVFDED
SDIRTISAVRPINDGYANGYGNAKWAGEVLLREAHDLCGLPVAVFRSDMI
LAHSRYTGQLNVPDQFTRLILSLIATGVAPGSFYQAHATGERPLAHYDGL
PADFTASAITALGPIEEFHTYDSVNPHADGISLDNFVDWLIEAGYPIARI
DNYTEWFTRFDTAIRSLPEKQKQHSLLPLLHAYRHPQHPHNGAFLPAIRF
SEGVQAHLNADIPHLTRELIAKYAADLKQLGLL

YP 001704097.1

Nucleotide sequence (SEQ ID NO:63)

>uniprot|B1MDX4|B1MDX4_MYCAB Putative fatty-acid-CoA ligase

ATGACGGCTGGTGCGGCGGCTCGCGTTGCCAAACTGTTCGAGTCCGATCCC
CAATTCCGGGCAGCCATGCCGGATCCAGCGGTGATGGACTCGCTGCTGGCG

FIG. 8 (Cont.)

CCCGGCCTGCGTTTATCCCAGGTACTCCACGCGTTGCTCAGCGGTTACGCG
GAGCGCCCGGTGATGGGTTTCCGGTCCCGCGAGTCGGTGGTCGACACCGCC
ACCGGCCGCACGGTCGACCGGCTGCTCCCTGCCTTTGAAACCATCACCTAT
GGGCAACTCCTGGAAGACATCTCGGCCATCCTCGCGGAGTGGCAGCATGGC
GACATTCCCATGGGCGCCGGCGACTTCATCGCCACCATCGGCTTCTCCAGT
CCCGACTACGTCACCCTGGATCTGGCCACCCTCATGAATGGTTCGGTCTCG
ATCCCACTGCAGCACAACACATCTGTGGCGCAGCTGCGGATGATGCTGGAG
GAGACCAGCCCACGCCTGGTGGCGGCGAGCGCGGACTGCCTGGATCTCGCG
GTCGAGGCAGCTGTCGGGCTTACCGATCTGCGACGGGTTGTGGTGTTCGAT
TACCGCGCCGAGACCGACGATCATCGCGAAAAACTGGCCACGGCAAGAGAA
CGCTTGCACGCGGCCGGTATGGACGTTGTAGTCGAACCGCTCGCAGAGGTG
ATCGGGAGAGGACGAGACCTACCCGAACCCGTGCTGTACACGGCCGGGGAC
GATCAGCGCACGGCCCTGATCATGTACACCTCCGGTAGCACCGGCGCGCCC
AAGGGGGCGATGTTCACCGAGTGGACGGTGACCCGCTTCTGGTCCTCGGGC
GCCGCCCCCAACCGGGACACCCCGATCATCAACGTGAACTTCCTGCCGCTC
AACCACCTTGCGGGCCGGGTAGGACTGCTGACGGCCTTCATTCCCGGCGGC
ACATGCTACTTCGTCCCCGAGAGCGATCTGTCCACCCTGTTCGAGGACTGG
CAGCTGGCACGGCCCACCCATATGGGTGTGGTTCCCCGTGTCGTCGACATG
CTCTTCCAGCACTACCAAACGCGAGTGGACGCACTGATGGCCGGGGGAACC
GACGTCGACACCGCCGATCGGCTAGCCAAAACCGAACTGCGCGAAGATGTC
CTGGGCGGGCGTGTGGTCGCCGGCATGCTCGCCACCGCGCCGTTGTCCCCC
GAGATGAAGGCTTTCCTGGAGTCCTCATTGGACTTTCATCTGCTTGATCTG
TACGGCCTGACCGAGGTCGGCGGCGTGTTCCGAGACGGCAAGATTTCCCGG
CCGCCGGTGCTCGACTACAAGCTCGTCGATGTTCCTGAGCTCGGGTACTAC
ACCACCGACAAGCCCCATCCGCGTGGCGAATTGCTGGTCAAGAGTGCCACC
GCAACGCCCGGCTACTACAAGCGTCCCGACGTCACCGCCGAGGTGTTCGAC
GCCGATGGCTACTACCGCACGGGCGATGTCATGGCGGAGGTCGCGCCGGAC
CAATTGGTGTACGTGGACAGGCGCAATAACGTCATCAAGCTCGCCCAGGGC
GAGTTCGTCGCGGTCGCCAATTTGGAAACGGTCTATGTGGGTGCGCCGCTG
GTGCGCCAGATCTTCGTCTACGGCAACAGCGAACGCGCATACCTCCTCGCC
GTTGTGGTGCCCACCGAGGAAGCCCTGCGGGCACACCCGGACCCCGTCGAA
CTGAAGAATTCGATCCGGGAGTCACTGCAGCGGACCGCCCGCTCCAACCAC
CTGCATTCCTACGAGCTGCCCGCCGACTTCATTATCGAAACCACTCCATTC
ACGATCGAGAGTGGGATGCTTGCGGCTGTCGGTAAGCCGATACGTCCCAAG
ATGATCGAGCACTACGGCGACCGGCTCGAGCAGCTCTACGTCGACCTCGCC
GAGGCACGCGTCCAGGAACTGCGGCAGCTCCGCGATACGGCGCAACAACGC
CCGGTCCTCGATACCGTCACCGAGGCCGCCCAGGCCCTCCTCGGCATGTCT
GCGGACGCCGTCCGTCCCGACCACCACTTCATCGACCTCGGCGGAGATTCG
CTGTCCGCGTTGACATTCTCCAATCTTCTTCGAGACCTCTTCGACGTCGAG

FIG. 8 (Cont.)

GTTCCGGTCGGTGTGATCACCGGCCCGGCGGCCGATCTGCGCAAGCTCGCC
GCTTACATCCAGCACGAACGGGAGCACAGCACCGCGACCGCTGCCAGCGTG
CACGGGCTCGACACCACCGTCATCAGCGCCACCGAACTGACACTCGACAAG
TTCATCGACGCCGAGACACTCCACAACGCTTCGCAACTCGACGTGCCGGCG
GGCGCGGTAGCTACCGTTCTGCTCACCGGCGCCAACGGATATCTCGGAAGA
TTCCTCTGCCTGGAGTGGCTGCAACGGCTGTCCCAGACAGGTGGACAACTG
ATCTGCCTGGTCCGCGGCGACAACGCCGATCAAGCCCTCGCGCGCCTCGTT
GCCGCCTACGGCGACACCGATCGCACACTGCTCGAGGAGTTCCACACCCTG
GCTCGACGGCACCTGCGCGTGATCGCCGCCGATATCGCTCAGCCGCGCTTC
GGCGTGGATGACGCCACCTGGGAGCAGCTGGCCCGCGATGTCGACAAGATC
GTGCATCCGGCCGCGCTGGTCAACCACGTGCTGCCCTACAACCAGCTGTTC
GGCCCCAATGTGTTTGGCACGGCGGAGGTTATCCGGCTGGCCCTGACCACC
CGGATCAAGCCGGTGACCTATCTGTCGACGATGGCCGTGGCCATGACCGTG
CCCGATTTCGACGAGGACGGGGACATCCGCACGGTGAGTCCCACCCGGCAT
ATCGACCCCGGCTACGCCAACGGGTACGCCAACAGCAAATGGGCCGGCGAG
GTGCTGCTGCGGGAGGCACACGACATATGCGGCCTGCCGGTCAGCGTGTTC
CGGTCCGACATGATCCTGACGCACCGCCGTTACAGCGGACAACTCAACGTC
ACCGACGCCTTCACCCGCATGCTGCTGAGCCTGGTGCTCACCGGCATCGCG
CCGCGAAGCTTTTACCAAGGCGATGGCAGCGGTGCCCGCCCACGCGCTCAC
TACGAGGGGCTGCCGGTCGATTTCGTCACCGAAGCCATTACCAGCCTCGGC
CTGTCCTCGTCCGAGGGATTTCGCTCGTACGACGTCATGAATCCTCACGAT
GACGGCATTTCTGTGGACACCTTTGTCGACTGGCTCATGGAAGATGGGCAT
TCCATCGACATCATCGACAACTACGACGAATGGCTGTCCCGTTTCGAGACG
GCATTGCGAGGTCTGCCCGACGAGCAGCGGCGCGCCTCAGTACTTCCGCTC
CTCGATGCGTATCGGATACCGGGCAACCCGCGCCGTGCTGCCGCCACGCCC
AATCATGTATTCCGGAAAGCCGTACAGGAGAACAACATCGGAGGTGACGGC
GCCGATATTCCGCAAATCGATCGTGCGCTGATCGCCAAATACATCGCCGAT
CTACGAGCACACAGGCTGCTGTGA

Amino acid sequence (SEQ ID NO:64)

>uniprot|B1MDX4|B1MDX4_MYCAB Putative fatty-acid-CoA ligase

MTAGAAARVAKLFESDPQFRAAMPDPAVMDSLLAPGLRLSQVLHALLSGY
AERPVMGFRSRESVVDTATGRTVDRLLPAFETITYGQLLEDISAILAEWQ
HGDIPMGAGDFIATIGFSSPDYVTLDLATLMNGSVSIPLQHNTSVAQLRM
MLEETSPRLVAASADCLDLAVEAAVGLTDLRRVVVFDYRAETDDHREKLA
TARERLHAAGMDVVVEPLAEVIGRGRDLPEPVLYTAGDDQRTALIMYTSG
STGAPKGAMFTEWTVTRFWSSGAAPNRDTPIINVNFLPLNHLAGRVGLLT

FIG. 8 (Cont.)

AFIPGGTCYFVPESDLSTLFEDWQLARPTHMGVVPRVVDMLFQHYQTRVD
ALMAGGTDVDTADRLAKTELREDVLGGRVVAGMLATAPLSPEMKAFLESS
LDFHLLDLYGLTEVGGVFRDGKISRPPVLDYKLVDVPELGYYTTDKPHPR
GELLVKSATATPGYYKRPDVTAEVFDADGYYRTGDVMAEVAPDQLVYVDR
RNNVIKLAQGEFVAVANLETVYVGAPLVRQIFVYGNSERAYLLAVVVPTE
EALRAHPDPVELKNSIRESLQRTARSNHLHSYELPADFIIETTPFTIESG
MLAAVGKPIRPKMIEHYGDRLEQLYVDLAEARVQELRQLRDTAQQRPVLD
TVTEAAQALLGMSADAVRPDHHFIDLGGDSLSALTFSNLLRDLFDVEVPV
GVITGPAADLRKLAAYIQHEREHSTATAASVHGLDTTVISATELTLDKFI
DAETLHNASQLDVPAGAVATVLLTGANGYLGRFLCLEWLQRLSQTGGQLI
CLVRGDNADQALARLVAAYGDTDRTLLEEFHTLARRHLRVIAADIAQPRF
GVDDATWEQLARDVDKIVHPAALVNHVLPYNQLFGPNVFGTAEVIRLALT
TRIKPVTYLSTMAVAMTVPDFDEDGDIRTVSPTRHIDPGYANGYANSKWA
GEVLLREAHDICGLPVSVFRSDMILTHRRYSGQLNVTDAFTRMLLSLVLT
GIAPRSFYQGDGSGARPRAHYEGLPVDFVTEAITSLGLSSSEGFRSYDVM
NPHDDGISVDTFVDWLMEDGHSIDIIDNYDEWLSRFETALRGLPDEQRRA
SVLPLLDAYRIPGNPRRAAATPNHVFRKAVQENNIGGDGADIPQIDRALI
AKYIADLRAHRLL

YP 001705436.1

Nucleotide sequence (SEQ ID NO:65)

>uniprot|B1MLD7|B1MLD7_MYCAB Probable fatty-acid-coa ligase FadD

ATGACTGAAACGATCTCCACAGCGGCTGTCCCCACTACGGATCTCGAAGAG
CAGGTGAAGCGACGCATCGAGCAGGTCGTGTCCAACGATCCGCAGCTGGCG
GCGCTTCTCCCGGAAGATTCGGTCACCGAGGCGGTCAACGAGCCCGATCTA
CCGCTGGTCGAGGTGATCAGGCGACTGCTGGAGGGCTACGGTGACCGCCCG
GCACTCGGCCAGCGCGCCTTCGAGTTCGTCACCGGGGACGACGGTGCGACC
GTGATCGCGCTGAAGCCCGAATACACCACCGTCTCCTACCGCGAGTTGTGG
GAACGTGCCGAGGCTATCGCTGCCGCGTGGCACGAGCAGGGCATCCGTGAC
GGCGACTTCGTCGCTCAGTTGGGTTTCACCAGCACGGACTTCGCGTCGCTC
GACGTCGCGGGATTGCGTCTGGGCACCGTCTCGGTGCCCCTGCAGACGGGC
GCGTCGCTGCAGCAGCGCAACGCGATTCTCGAAGAGACCCGGCCCGCAGTC
TTTGCCGCGAGTATCGAATACCTTGATGCCGCCGTCGATTCGGTGCTTGCG
ACCCCCTCGGTGCGACTCCTCTCGGTTTTCGACTATCACGCGGAGGTCGAC
AGCCAGCGCGAAGCGCTGGAGGCTGTGCGGGCCCGGCTTGAGAGTGCCGGC

FIG. 8 (Cont.)

CGGACGATCGTCGTCGAGGCCCTGGCGGAGGCTCTCGCGCGGGGGCGGGAC
CTGCCCGCCGCGCCGCTGCCCAGTGCAGATCCCGATGCCTTGCGTCTGCTC
ATCTACACCTCCGGCAGCACCGGTACCCCCAAGGGCGCCATGTATCCGCAA
TGGCTGGTCGCCAACTTGTGGCAGAAGAAGTGGCTCACCGACGATGTGATT
CCGTCCATAGGCGTGAACTTCATGCCCATGAGCCACCTGGCGGGTCGCCTC
ACTCTCATGGGCACCCTTTCCGGTGGCGGAACCGCCTACTACATCGCTTCG
AGCGATCTTTCGACTTTCTTCGAGGACATCGCGCTCATCCGCCCCTCCGAA
GTGCTCTTCGTGCCGCGTGTGGTGGAGATGGTGTTCCAGCGTTTTCAGGCA
GAATTGGACCGGTCCCTTGCCCCGGGTGAGAGCAACTCCGAGATCGCGGAG
CGAATCAAGGTCCGCATCCGGGAACAGGACTTCGGCGGGCGTGTGCTCAGT
GCTGGCTCCGGGTCGGCCCCGTTGTCTCCTGAGATGACGGAGTTCATGGAG
TCGCTGCTGCAGGTGCCGTTGCGCGACGGGTATGGGTCCACCGAGGCCGGT
GGTGTGTGGCGTGACGGAGTCCTGCAGCGTCCGCCCGTCACCGACTACAAG
CTGGTTGACGTTCCGGAACTCGGATACTTCACCACAGATTCGCCGCATCCC
CGTGGCGAGCTGCGGTTGAAGTCGGAGACGATGTTCCCCGGCTACTACAAG
CGCCCGGAGACCACTGCCGATGTCTTCGATGACGAGGGGTACTACAAGACC
GGTGACGTGGTCGCCGAGCTCGGGCCGGATCACCTCAAGTACCTCGACCGC
GTCAAGAACGTCCTCAAGCTCGCGCAGGGAGAGTTTGTCGCGGTGTCAAAG
CTGGAGGCCGCTTACACCGGCAGCCCGCTGGTCCGGCAGATCTTTGTGTAC
GGGAACAGTGAACGCTCGTTCCTGCTGGCTGTCGTGGTCCCGACACCCGAA
GTCCTTGAGCGGTACGCAGATTCGCCAGATGCGCTCAAGCCCTTGATCCAG
GATTCGCTGCAGCAGGTCGCCAAGGACGCGGAGCTGCAATCCTATGAGATA
CCGCGCGACTTCATCGTTGAGACGGTGCCGTTCACCGTCGAGTCCGGATTG
CTATCGGACGCGCGAAAGCTGCTGCGCCCCAAGCTGAAGGATCACTACGGA
GAGAGGCTGGAGGCGCTGTACGCCGAACTGGCGGAAAGCCAGAATGAGCGG
CTGCGCCAGTTGGCCAGGGAGGCAGCCACGCGCCCGGTCCTGGAGACGGTG
ACCGATGCGGCCGCCGCGCTGCTGGGCGCATCGTCCTCGGATCTGGCTCCT
GATGTGCGATTCATCGACCTCGGTGGCGACTCACTGTCGGCGCTGTCGTAC
TCCGAGCTGCTGCGCGACATCTTTGAGGTGGACGTTCCGGTGGGCGTCATC
AACAGCGTCGCCAACGACCTTGCCGCGATCGCCCGGCACATCGAGGCGCAG
CGGACCGGCGCCGCTACGCAGCCGACCTTTGCGTCGGTCCACGGCAAGGAC
GCGACGGTCATCACCGCCGGTGAACTCACCCTCGACAAGTTCTTGGACGAG
TCACTGTTGAAAGCGGCCAAGGACGTTCAGCCGGCAACGGCCGATGTCAAG
ACCGTTCTAGTGACCGGCGGCAACGGCTGGTTGGGTCGTTGGCTGGTGCTC
GATTGGCTGGAGCGGTTGGCACCCAATGGTGGCAAGGTCTACGCCCTCATT
CGTGGCGCCGATGCCGAAGCAGCCCGGGCACGGTTGGACGCCGTGTACGAA
TCGGGTGATCCCAAGCTGTCCGCGCATTATCGTCAGCTGGCGCAACAGAGT
CTGGAAGTTATCGCCGGCGATTTCGGCGACCAGGATCTCGGTCTATCCCAG
GAAGTTTGGCAGAAGCTGGCCAAGGACGTGGACCTGATCGTGCACTCCGGT

FIG. 8 (Cont.)

GCCTTGGTGAACCACGTGCTGCCGTACAGCCAGTTGTTCGGTCCGAATGTG
GCGGGTACCGCCGAGATCATCAAGCTGGCAATTTCGGAGCGGCTCAAGCCG
GTCACCTACCTGTCGACGGTGGGCATCGCCGACCAGATTCCGGTGACGGAG
TTCGAGGAAGACTCCGATGTTCGTGTGATGTCGGCCGAGCGCCAGATCAAT
GACGGCTACGCGAACGGATACGGCAACTCAAAATGGGCCGGCGAGGTGCTG
TTGCGGGAGGCTCATGACCTAGCGGGGCTGCCGGTGCGTGTGTTCCGCTCC
GACATGATCCTGGCGCACAGTGACTACCACGGACAGCTCAACGTCACCGAC
GTGTTCACCCGGAGCATCCAGAGTCTGCTGCTCACCGGTGTTGCACCGGCC
AGCTTCTATGAATTGGATGCCGACGGCAATCGGCAGCGCGCTCACTATGAC
GGTGTGCCCGGCGATTTCACCGCCGCATCGATCACCGCCATCGGCGGTGTG
AACGTGGTAGACGGTTACCGCAGCTTCGACGTGTTCAACCCGCACCATGAC
GGTGTCTCGATGGATACCTTCGTCGACTGGCTGATCGACGCAGGCTACAAG
ATCGCGCGGATCGACGATTACGACCAGTGGCTCGCCCGGTTCGAGCTGGCC
CTCAAGGGATTGCCCGAGCAGCAGCGGCAACAGTCGGTGTTGCCACTTCTC
AAGATGTACGAGAAGCCGCAACCGGCGATCGACGGAAGTGCACTTCCGACC
GCAGAATTCAGTCGCGCCGTGCACGAGGCGAAGGTCGGAGACAGCGGTGAG
ATACCGCACGTCACCAAGGAGCTGATCCTCAAGTACGCCAGCGATATTCAG
CTGTTGGGCCTGGTGTAG

Amino acid sequence (SEQ ID NO:66)

>uniprot|B1MLD7|B1MLD7_MYCAB Probable fatty-acid-coa ligase FadD

MTETISTAAVPTTDLEEQVKRRIEQVVSNDPQLAALLPEDSVTEAVNEPD
LPLVEVIRRLLEGYGDRPALGQRAFEFVTGDDGATVIALKPEYTTVSYRE
LWERAEAIAAAWHEQGIRDGDFVAQLGFTSTDFASLDVAGLRLGTVSVPL
QTGASLQQRNAILEETRPAVFAASIEYLDAAVDSVLATPSVRLLSVFDYH
AEVDSQREALEAVRARLESAGRTIVVEALAEALARGRDLPAAPLPSADPD
ALRLLIYTSGSTGTPKGAMYPQWLVANLWQKKWLTDDVIPSIGVNFMPMS
HLAGRLTLMGTLSGGGTAYYIASSDLSTFFEDIALIRPSEVLFVPRVVEM
VFQRFQAELDRSLAPGESNSEIAERIKVRIREQDFGGRVLSAGSGSAPLS
PEMTEFMESLLQVPLRDGYGSTEAGGVWRDGVLQRPPVTDYKLVDVPELG
YFTTDSPHPRGELRLKSETMFPGYYKRPETTADVFDDEGYYKTGDVVAEL
GPDHLKYLDRVKNVLKLAQGEFVAVSKLEAAYTGSPLVRQIFVYGNSERS
FLLAVVVPTPEVLERYADSPDALKPLIQDSLQQVAKDAELQSYEIPRDFI
VETVPFTVESGLLSDARKLLRPKLKDHYGERLEALYAELAESQNERLRQL
AREAATRPVLETVTDAAAALLGASSSDLAPDVRFIDLGGDSLSALSYSEL
LRDIFEVDVPVGVINSVANDLAAIARHIEAQRTGAATQPTFASVHGKDAT

FIG. 8 (Cont.)

VITAGELTLDKFLDESLLKAAKDVQPATADVKTVLVTGGNGWLGRWLVLD
WLERLAPNGGKVYALIRGADAEAARARLDAVYESGDPKLSAHYRQLAQQS
LEVIAGDFGDQDLGLSQEVWQKLAKDVDLIVHSGALVNHVLPYSQLFGPN
VAGTAEIIKLAISERLKPVTYLSTVGIADQIPVTEFEEDSDVRVMSAERQ
INDGYANGYGNSKWAGEVLLREAHDLAGLPVRVFRSDMILAHSDYHGQLN
VTDVFTRSIQSLLLTGVAPASFYELDADGNRQRAHYDGVPGDFTAASITA
IGGVNVVDGYRSFDVFNPHHDGVSMDTFVDWLIDAGYKIARIDDYDQWLA
RFELALKGLPEQQRQQSVLPLLKMYEKPQPAIDGSALPTAEFSRAVHEAK
VGDSGEIPHVTKELILKYASDIQLLGLV

YP 001828302.1

Nucleotide sequence (SEQ ID NO:67)

>uniprot|B1VMZ4|B1VMZ4_STRGG Putative carboxylic acid reductase

ATGGCCGAACCGCTGGACGCCGCCACCGCGTCGGCGCACGATCCGGGCCAG
GGGCTCGCCGAGGCCCTGGCCGCCGTCGAACCGGGCCGGGCGCTCGCCGAG
GTCATGGCGTCCGTCCTGGAGGGCCACGGGGACCGGCCCGCCCTCGGCGAA
CGGGCCCGGGAGCCGGAGACCGGGCGTCTCCTCCCGCACTTCGACACCATC
AGCTACCGCGAACTGTGGTCCCGCGTCCGTGCGCTGGCCGGCCGGTGGCAC
CACGACCCGGAATACCCCCTGGGCCCCGGCGACCGGATCTGCACCCTCGGC
TTCACCAGCACCGACTACGCGACGCTCGACCTGGCGTGCATCCACCTGGGG
GCCGTGCCCGTCCCCCTCCCGTCCAACGCCCCGCTGCCCCGACTGGCGCCG
GTCGTCGAGGAGTCCGGGCCGACGGTACTGGCCGCGAGCGTCGACCGGCTC
GACACCGCCATCGACGTCGTCCTCGCGTCGAGCACGATCCGCCGCCTCCTC
GTCTTCGACGACGGCCCGGGGGCCACCCGCCCGGGCGGGGCACTGGCGGCC
GCCCGCCAACGCCTGTCCGGCAGCCCGGTCACCGTCGACACCCTGGCCGGA
CTCATCGACCGGGGCAGGGACCTGCCGCCCCCGCCCCTGTACATCCCGGAC
CCGGGGGAGGACCCGCTCGCCCTGCTCATCTACACCTCCGGCAGCACCGGC
GCGCCCAAGGGCGCCATGTACACCCAACGGCTCCTGGGCACCGCGTGGTAC
GGGTTCAGCTACGGGGCGGCCGACACCCCCGCGATCAGCGTCCTCTACCTG
CCGCAGAGCCACCTCGCGGGCCGCTACGCGGTGATGGGATCGCTCGTGAAG
GGGGGCACCGGCTACTTCACCGCTGCCGACGACCTGTCCACCCTGTTCGAG
GACATCGCCCTGGTCCGCCCCACGGAGCTGACCATGGTCCCGCGCCTGTGC
GACATGCTCCTCCAGCACTACCGGAGCGAACGGGACCGCCGGGCCGACGAA
CCGGGCGACATCGAGGCGGCGGTCACGAAGGCGGTGCGGGAGGACTTCCTG
GGCGGGCGCGTCGCCAAGGCGTTCGTCGGCACCGCGCCGCTCTCCGCCGAA

FIG. 8 (Cont.)

CTCACGGCGTTCGTCGAGTCCGTCCTCGGCTTCCACCTCTACACCGGCTAC
GGCTCCACCGAAGCCGGCGGAGTGCTGCTGGACACGGTGGTGCAGCGCCCT
CCGGTCACCGACTACAAACTGGTCGACGTCCCCGAACTGGGCTACTACGCG
ACCGACCTGCCCCATCCGCGCGGCGAACTGCTGCTGAAGTCCCACACGCTC
ATCCCCGGCTACTACCGGCGCCCCGACCTCACCGCCGCGATCTTCGACGCG
GACGGCTACTACCGCACCGGTGACGTCTTCGCCGAGACCGGACCCGACCGG
CTGGTCTACGTCGACCGCACGAAGGACACCCTGAAGCTGTCCCAGGGCGAG
TTCGTGGCCGTGTCCCGCCTGGAGACCGTCCTCCTCGACAGCCCTCTCGTC
CAGCACCTCTACCTGTACGGCAACAGCGAGCGCGCCTACCTCCTCGCGGTG
GTGGTGCCCACCCCGGACGCGCTGGCCGGGTGCGGCGGGGACACCGAGGCG
CTCAGGCCGCTGCTCATGGAATCCCTCCGCAGCGTCGCCAGGAGGGCCGGG
CTCAACGCGTACGAGATCCCGCGCGGCATCCTCGTCGAGCCCGAGCCCTTC
AGCCCGGAGAACGGCCTCTTCACCGAGAGCCACAAACTGCTGCGCCCCCGC
CTCAAGGAGCGCTACGGGCCCGCTCTGGAGCTGCTGTACGACCGACTGGCC
GACGGGCAGGACCGCCGGCTGCGCGAGCTGCGGCGCACCGGTGCGGACCGG
CCGGTGCAGGAGACGGTCCTCCGGGCCGCCCAGGCCCTGCTGGGATCCCCG
GGCTCCGACCTCCGGCCCGGCGCGCACTTCACCGACCTCGGCGGGGACTCC
CTCTCCGCCGTCTCGTTCTCCGAGCTGATGAAGGAGATCTTCCACGTCGAC
GTCCCGGTCGGTGCGATCATCGGCCCGGCCGCCGACCTGGCGGAGGTGGCG
CGGTACATCACGGCGGCCCGTCGGCCGGCCGGGGCACCGCGGCCCACGCCC
GCCTCCGTGCACGGGGAGCACCGCACCGAAGTCCGCGCCGGGGACCTCGCC
CCGGAGAAGTTCCTCGACGCGCCCACGCTCGCCGCCGCTCCGGCGCTGCCC
CGCCCCGACGGCGACGTCCGGACGGTCCTGCTGACCGGCGCCACCGGCTAC
CTCGGCCGGTTCCTCTGCCTGGAGTGGCTGGAGCGGCTGGCGCCCTCGGGC
GGACGGCTGGTCTGCCTCGTCCGCGGCAGCGACGCGACCGTCGCGGCGAGG
CGGCTGGAGGCAGCCTTCGACAGCGGCGACACCGCCCTGCTCCGGCGCTAC
CGGAAGGCGGCCGGGAAGACCCTGGACGTGGTCGCGGGGGACATCGGCGAA
CCACTGCTGGGCCTGGCGGAGGAGACCTGGCGGGAACTGGCCGGCGCCGTG
GACCTGATCGTGCACCCGGCCGCGCTGGTCAACCACCTGCTGCCGTACGGC
GAACTGTTCGGCCCCAACGTCGTCGGCACCGCCGAAGCGATCCGGCTGGCG
CTCACCACCCGGCTGAAGCCCGTCAACCACGTCTCGACCGTCGCGGTCTGC
CTCGGCACCCCCGCCGAGACGGCCGACGAGAACGCCGACATCCGGGCCGCC
GTCCCGGTACGGACCACCGGCCAGGGGTACGCCGACGGATACGCGACCAGC
AAATGGGCCGGCGAGGTCCTCCTGCGTGAGGCGCACGAGCGCTATGGCCTC
CCCGTCGCCGTCTTCCGGTCCGACATGGTCCTGGCGCACCGCACCTACACC
GGACAGGTCAACGTCCCCGACGTCCTCACCCGGCTGCTGCTCAGCCTGGTC
GCCACCGGCATCGCCCCCGGCTCGTTCTACCGCACGGACACCCGTGCCCAC
TACGACGGCCTGCCGGTCGACTTCACCGCGGAGGCCGTCGTCGCGCTGGGC
GCACCGATCACCGAGGGCCACCGGACCTTCAACGTCCTCAACCCGCACGAC

FIG. 8 (Cont.)

GACGGCGTTTCCCTGGACACCTTCGTCGACTGGCTCATCGAGGCCGGCCAC
CCGATCCGGCGGATCGACGACCACGGTGCCTGGCTCACCCGCTTCACCGCG
GCGCTCCGCGCCCTGCCGGAGAAGCAGCGGCAGCACTCCCTGCTCCCGCTG
ATCGGCGCCTGGGCGGAACCCGGCGAAGGAGCCCCCGGGCCGCTGCTCCCC
GCCCGGCGCTTCCACGCCGCCGTCCGGGCGGCGGGGGTCGGCCCCGAGCGG
GACATTCCCCGGGTGTCGCCGGACCTCATCCGCAAGTACGTCACCGACCTG
CGCGCACTCGGGCTCCTCGCCGGCCCCTGA

Amino acid sequence (SEQ ID NO:68)

>uniprot|B1VMZ4|B1VMZ4_STRGG Putative carboxylic acid reductase

MAEPLDAATASAHDPGQGLAEALAAVEPGRALAEVMASVLEGHGDRPALG
ERAREPETGRLLPHFDTISYRELWSRVRALAGRWHHDPEYPLGPGDRICT
LGFTSTDYATLDLACIHLGAVPVPLPSNAPLPRLAPVVEESGPTVLAASV
DRLDTAIDVVLASSTIRRLLVFDDGPGATRPGGALAAARQRLSGSPVTVD
TLAGLIDRGRDLPPPPLYIPDPGEDPLALLIYTSGSTGAPKGAMYTQRLL
GTAWYGFSYGAADTPAISVLYLPQSHLAGRYAVMGSLVKGGTGYFTAADD
LSTLFEDIALVRPTELTMVPRLCDMLLQHYRSERDRRADEPGDIEAAVTK
AVREDFLGGRVAKAFVGTAPLSAELTAFVESVLGFHLYTGYGSTEAGGVL
LDTVVQRPPVTDYKLVDVPELGYYATDLPHPRGELLLKSHTLIPGYYRRP
DLTAAIFDADGYYRTGDVFAETGPDRLVYVDRTKDTLKLSQGEFVAVSRL
ETVLLDSPLVQHLYLYGNSERAYLLAVVVPTPDALAGCGGDTEALRPLLM
ESLRSVARRAGLNAYEIPRGILVEPEPFSPENGLFTESHKLLRPRLKERY
GPALELLYDRLADGQDRRLRELRRTGADRPVQETVLRAAQALLGSPGSDL
RPGAHFTDLGGDSLSAVSFSELMKEIFHVDVPVGAIIGPAADLAEVARYI
TAARRPAGAPRPTPASVHGEHRTEVRAGDLAPEKFLDAPTLAAAPALPRP
DGDVRTVLLTGATGYLGRFLCLEWLERLAPSGGRLVCLVRGSDATVAARR
LEAAFDSGDTALLRRYRKAAGKTLDVVAGDIGEPLLGLAEETWRELAGAV
DLIVHPAALVNHLLPYGELFGPNVVGTAEAIRLALTTRLKPVNHVSTVAV
CLGTPAETADENADIRAAVPVRTTGQGYADGYATSKWAGEVLLREAHERY
GLPVAVFRSDMVLAHRTYTGQVNVPDVLTRLLLSLVATGIAPGSFYRTDT
RAHYDGLPVDFTAEAVVALGAPITEGHRTFNVLNPHDDGVSLDTFVDWLI
EAGHPIRRIDDHGAWLTRFTAALRALPEKQRQHSLLPLIGAWAEPGEGAP
GPLLPARRFHAAVRAAGVGPERDIPRVSPDLIRKYVTDLRALGLLAGP

YP 001851230.1

FIG. 8 (Cont.)

Nucleotide sequence (SEQ ID NO:69)

>uniprot|B2HE95|B2HE95_MYCMM Fatty-acid-CoA ligase FadD9 1

TTGTCAATTACCTGTGTGGATACCCGTGCACAGCGGAGCGCCCGTCGCATC
GAGCAGCTTTACTCCACCGATGCGCAATTCGCCGCCGCCCGGCCCAGTACG
GCGGTCGGTATCGCAATCAGCAAGTCCGGGTTGGGATTACCACAGATCATT
CAAACGGTGATGGACGGATACCCGCAACGTCCGGCACTTGGGCAGCGGGCG
ACGCGCGTTGTTACCGATCCGAATACCGGGCGTAGCTCGGCGCAGCTGTTG
GCGGAGTTCGAGACCATCACCTACCGGGAGTTGTGGAACCGCACCAATGCA
TTGACCAACGCATTCGCCGCCGAGGCACTTGCGGATCGCGGTCAGCGGGTC
TGTGTGCTGGGATTCGCGAGCATCGACTACGCCACCATCGACTTGGCGCTG
ATGTTGCTCGGCGCGGTATCGGTTCCGTTGCCGACGAATGCGGCTCGCGCC
CAGCTGTGCCATATCGTCTCCGAGACCCAGCCCAGCCTGATCGCCTCGAGT
ACCGAAAACCTGCCCGATGCAATCTCTTTGGTGCTGTCGCACCGCGCACCA
CACCGGGTGGTGGTGTTCGACTACCGCCCCGAACTCGACGCACACCGCGAA
GCCCTCGAAGCCGCTCGCGCGCGCCTGGCCGCCATCCCGGTGACCGTCGAA
ACGCTCACCGCCATCATCGCGCGCGGTCGAACGGTGCGGCCGGCCGAGGCC
GATTGCGGCGCCCAGTCCGCTGATGCACCGGCGCTTTTGATCTATACCTCC
GGAAGCACCGGGGCACCCAAGGGCGTCGTCTACACCCGCAACCGGGTGGCG
GACTTCTGGCGCACCTCGAAAGCCGAGGTCGAAGCGACCGAACAACGAACC
GCTCCTTCGATCACCCTCAACTTCATGCCGATGAGCCACGCGAACGGCCGC
CAGGTGCTCTACGGGACGCTGTCCAACGGCGGCACCGCGTATTTCACGGCC
CGCAGCGACCTCTCGACGCTCTTCGATGATCTCGCGTTGGTCCGGCCCACC
GAATTGGGCTTTCCACCGCGCATTTGGGACATGCTGTTGGAGAGGTTTGGG
CGCGAAGTCGACCGTCGGCTCCGGGACGGCACAGCCGAGGGCGCCGACCCG
GGCGCGCTGAAGGCTCGCGTGGCGGCCGACCTACGCCAGGTGCTGCTCGGC
GGACGGTATGCGCTGGCGATGATGGGCTCCGCGCCAATCTCCGAGCAGATG
AAAGCATCCGTCGAATCCCTGCTCGATCTGGACGTCATGGAGGGCTATGGC
TCCACGGAAGCCGGAACGGTCATCATCAACAACGAGGTTCAGCGTCCCCAG
GTGATCGACTACAAGCTGGTCGACGTTGCGGAACTGGGCTATTTCCTTACC
GACCGGCCATATCCGCGGGGCGAACTGCTGGTCAAAACGCGGACACTGTTT
TCCGGCTACTACCGGGACCCCGAAGACGGCGCCCAGGTCTTCGACCCGGAC
GGCTTCTACCGGACCGGCGACATCATGGCCCAAGTCGGCCCCGATCGGCTC
GCCTACCTCGACCGGCGCAACAACGTGCTGAAGCTGTCGCAGGGGGAGTTC
GTCGCGGTCTCGCGACTAGAAGCAATATTTGCCAATAGCCCGTTGGTCCGG
CAGATCTTCGTCTATGCCAACGGTGCTCGCGCCTACCCACTGGCGGTAGTC
GTGCCCACCCAGGACGCACAGTCGCGCCACGGTCGCGCCGAACTCAAGGCC

FIG. 8 (Cont.)

GAACTCCATACATCGCTGCACCGCGTTGCCATGTCGGCCGGTCTGGCACCC
TACGAGATCCCACGCGACTTCATTGTCGAGACAACCCCCTTCACGCCGCAG
AACGGCCTGCTCACCGCAATCCACAAGCTGGCCCGGCCGCACCTCACGCAG
CGCTATGGCGCACGTCTGGAGCTGCTGTACACCGAGCTGGCCGACAGCCAG
ACCCGCCGGCTGCACCGATTGCGCCAAACCGGTGGGCGGCTGCCGGCGCTC
GAGACCATCAGGCGTGCCGCCGGGGCACTGTTGGGCACGGAGACCACCGAG
CCGCGGCCCGAGGCCCACTTCAAAGATCTGGGCGGGGATTCGGTATCGGCG
GTGACGTTCTCCAACCTGCTACACGACATCTACGGTTTCGATGTTCCGGTC
GGTGTGATCCTCGGCCCGGCAACCGATTTGCGGGCGCTGGCCAGCCACGTC
GAGAGCCGGCGCGGTGCCGGATGGTCGGGGCCCAGCTTCGCGTCGGTGCAC
GTGCCCCGGGCGACCTCGGTACACGCCGGCGACCTGAAACTGGCCAAGTTC
CTGGACACCAAGACACTCGCAGCTGCCACGAGCCTGCCCGCTGCCGATGCC
CGGGCACGGACGGTGCTACTCACCGGCGCAACCGGATTCCTGGGACGCTAC
CTGGTGCTGGAATGGCTGCGCCGGTTGCGGGCCGTCGGCGGCAAGCTGATC
TGTCTGGTGCGCGCCGCGTCCGACGAACAAGCCCGGGTTCGGCTGGATACG
GCCTTCGATAGCGGCGATCCGCAGCTGCCCGAGCACTTTCGGCAGCTCGCT
GTCGACCGCCTGGAGGTCCTCGCCGGCGATAAGAGCGAACCAGGTCTCGGT
CTGGACGGCCCAACCTGGCAGCGACTGGCCGACACGGTCGACCTGATCGTC
GACCCCGCCACGCTGGTCAACCACGTGCTGTCGTACCGGCAGCTGTTCGCT
CCCAACGTGGCGGGCACCGCCGAGTTGCTCCGCCTCGCACTCACCACCAAA
CGCAAGCCCTATGCCTACGTCTCGACCGTCAGCGTGGCCAACCAGATCGAA
CCGTCCGCATTCACCGAAGACGCCGACATCCGGGAGATCAGCCGCACCCGA
ACCATCGATGACAGCTTTGCCAACGGCTACACCACCAGCAAGTGGGCCAGC
GAGGTGCTGTTGCGTGAGGCTCACGATCTGTGCGGACTGCCGGTCACGGTC
TTTCGTTGCGACATGATCCTGGCGGACACCAGCTACGCCGGCCAGCTCAAC
CTCGCCGATACCTTCACCCGGCTGATGCTCAGTGTGGCGGCCACCGGGATC
GCGCCCGCCTCGTTCTACCGGCTGGGCCCCGACGGCAAACGCCAGCCCGCC
CACTTCGACGGATTGCCCGTCGAATTCATCGCCGAGGCGGTGGCCACCCTG
GGGGCGCGGCGCCACGACGGGTTCCAGGTCCACCATGTGGCGAATCCGCAC
CACGACGGCGTTGGGTTGGACGAGTACGTCGACTGGCTAGTCGATGCCGGT
TGCCCCATCCGGCGCATTCCCGACTATGACGAGTGGCTGAGTCGATTCGAG
ACGGCGCTGCACGCGCTGCCGGATCGCAAGCGTCGTCATTCACTGCTTCCG
CTGCTGCAGAACTATCGAGAACCCGCCGAGCCGATCCGGGGCGGCATCGCG
CCCGCACCACGGTTTCGCGGTGCGGTACGGCAGGCGAAAATCGGCCGCGAC
AACGACATTCCCCATGTCGGCCCGGCGATCATCGCCAAGTACGCCAGCGAC
CTGCAGCTTCTCGGCCTGGCTTGA

Amino acid sequence (SEQ ID NO:70)

>uniprot|B2HE95|B2HE95_MYCMM Fatty-acid-CoA ligase

FIG. 8 (Cont.)

FadD9_1

MSITCVDTRAQRSARRIEQLYSTDAQFAAARPSTAVGIAISKSGLGLPQI
IQTVMDGYPQRPALGQRATRVVTDPNTGRSSAQLLAEFETITYRELWNRT
NALTNAFAAEALADRGQRVCVLGFASIDYATIDLALMLLGAVSVPLPTNA
ARAQLCHIVSETQPSLIASSTENLPDAISLVLSHRAPHRVVVFDYRPELD
AHREALEAARARLAAIPVTVETLTAIIARGRTVRPAEADCGAQSADAPAL
LIYTSGSTGAPKGVVYTRNRVADFWRTSKAEVEATEQRTAPSITLNFMPM
SHANGRQVLYGTLSNGGTAYFTARSDLSTLFDDLALVRPTELGFPPRIWD
MLLERFGREVDRRLRDGTAEGADPGALKARVAADLRQVLLGGRYALAMMG
SAPISEQMKASVESLLDLDVMEGYGSTEAGTVIINNEVQRPQVIDYKLVD
VAELGYFLTDRPYPRGELLVKTRTLFSGYYRDPEDGAQVFDPDGFYRTGD
IMAQVGPDRLAYLDRRNNVLKLSQGEFVAVSRLEAIFANSPLVRQIFVYA
NGARAYPLAVVVPTQDAQSRHGRAELKAELHTSLHRVAMSAGLAPYEIPR
DFIVETTPFTPQNGLLTAIHKLARPHLTQRYGARLELLYTELADSQTRRL
HRLRQTGGRLPALETIRRAAGALLGTETTEPRPEAHFKDLGGDSVSAVTF
SNLLHDIYGFDVPVGVILGPATDLRALASHVESRRGAGWSGPSFASVHVP
RATSVHAGDLKLAKFLDTKTLAAATSLPAADARARTVLLTGATGFLGRYL
VLEWLRRLRAVGGKLICLVRAASDEQARVRLDTAFDSGDPQLPEHFRQLA
VDRLEVLAGDKSEPGLGLDGPTWQRLADTVDLIVDPATLVNHVLSYRQLF
APNVAGTAELLRLALTTKRKPYAYVSTVSVANQIEPSAFTEDADIREISR
TRTIDDSFANGYTTSKWASEVLLREAHDLCGLPVTVFRCDMILADTSYAG
QLNLADTFTRLMLSVAATGIAPASFYRLGPDGKRQPAHFDGLPVEFIAEA
VATLGARRHDGFQVHHVANPHHDGVGLDEYVDWLVDAGCPIRRIPDYDEW
LSRFETALHALPDRKRRHSLLPLLQNYREPAEPIRGGIAPAPRFRGAVRQ
AKIGRDNDIPHVGPAIIAKYASDLQLLGLA

YP 001850422.1

Nucleotide sequence (SEQ ID NO:71)

>uniprot|B2HN69|B2HN69_MYCMM Fatty-acid-CoA ligase FadD9

ATGTCGCCAATCACGCGTGAAGAGCGGCTCGAGCGCCGCATCCAGGACCTC
TACGCCAACGACCCGCAGTTCGCCGCCGCCAAACCCGCCACGGCGATCACC
GCAGCAATCGAGCGGCCGGGTCTACCGCTACCCCAGATCATCGAGACCGTC
ATGACCGGATACGCCGATCGGCCGGCTCTCGCTCAGCGCTCGGTCGAATTC
GTGACCGACGCCGGCACCGGCCACACCACGCTGCGACTGCTCCCCCACTTC

FIG. 8 (Cont.)

GAAACCATCAGCTACGGCGAGCTTTGGGACCGCATCAGCGCACTGGCCGAC
GTGCTCAGCACCGAACAGACGGTGAAACCGGGCGACCGGGTCTGCTTGTTG
GGCTTCAACAGCGTCGACTACGCCACGATCGACATGACTTTGGCGCGGCTG
GGCGCGGTGGCCGTACCACTGCAGACCAGCGCGGCGATAACCCAGCTGCAG
CCGATCGTCGCCGAGACCCAGCCCACCATGATCGCGGCCAGCGTCGACGCA
CTCGCTGACGCCACCGAATTGGCTCTGTCCGGTCAGACCGCTACCCGAGTC
CTGGTGTTCGACCACCACCGGCAGGTTGACGCACACCGCGCAGCGGTCGAA
TCCGCCCGGGAGCGCCTGGCCGGCTCGGCGGTCGTCGAAACCCTGGCCGAG
GCCATCGCGCGCGGCGACGTGCCCCGCGGTGCGTCCGCCGGCTCGGCGCCC
GGCACCGATGTGTCCGACGACTCGCTCGCGCTACTGATCTACACCTCGGGC
AGCACGGGTGCGCCCAAGGGCGCGATGTACCCCCGACGCAACGTTGCGACC
TTCTGGCGCAAGCGCACCTGGTTCGAAGGCGGCTACGAGCCGTCGATCACG
CTGAACTTCATGCCAATGAGCCACGTCATGGGCCGCCAAATCCTGTACGGC
ACGCTGTGCAATGGCGGCACCGCCTACTTCGTGGCGAAAAGCGATCTCTCC
ACCTTGTTCGAAGACCTGGCGCTGGTGCGGCCCACCGAGCTGACCTTCGTG
CCGCGCGTGTGGGACATGGTGTTCGACGAGTTTCAGAGTGAGGTCGACCGC
CGCCTGGTCGACGGCGCCGACCGGGTCGCGCTCGAAGCCCAGGTCAAGGCC
GAGATACGCAACGACGTGCTCGGTGGACGGTATACCAGCGCACTGACCGGC
TCCGCCCCTATCTCCGACGAGATGAAGGCGTGGGTCGAGGAGCTGCTCGAC
ATGCATCTGGTCGAGGGCTACGGCTCCACCGAGGCCGGGATGATCCTGATC
GACGGAGCCATTCGGCGCCCGGCGGTACTCGACTACAAGCTGGTCGATGTT
CCCGACCTGGGTTACTTCCTGACCGACCGGCCACATCCGCGGGGCGAGTTG
CTGGTCAAGACCGATAGTTTGTTCCCGGGCTACTACCAGCGAGCCGAAGTC
ACCGCCGACGTGTTCGATGCTGACGGCTTCTACCGGACCGGCGACATCATG
GCCGAGGTCGGCCCCGAACAGTTCGTGTACCTCGACCGCCGCAACAACGTG
TTGAAGCTGTCGCAGGGCGAGTTCGTCACCGTCTCCAAACTCGAAGCGGTG
TTTGGCGACAGCCCACTGGTACGGCAGATCTACATCTACGGCAACAGCGCC
CGTGCCTACCTGTTGGCGGTGATCGTCCCCACCCAGGAGGCGCTGGACGCC
GTGCCTGTCGAGGAGCTCAAGGCGCGGCTGGGCGACTCGCTGCAAGAGGTC
GCAAAGGCCGCCGGCCTGCAGTCCTACGAGATCCCGCGCGACTTCATCATC
GAAACAACACCATGGACGCTGGAGAACGGCCTGCTCACCGGCATCCGCAAG
TTGGCCAGGCCGCAGCTGAAAAAGCATTACGGCGAGCTTCTCGAGCAGATC
TACACGGACCTGGCACACGGCCAGGCCGACGAACTGCGCTCGCTGCGCCAA
AGCGGTGCCGATGCGCCGGTGCTGGTGACGGTGTGCCGTGCGGCGGCCGCG
CTGTTGGGCGGCAGCGCCTCTGACGTCCAGCCCGATGCGCACTTCACCGAT
TTGGGCGGCGACTCGCTGTCGGCGCTGTCGTTCACCAACCTGCTGCACGAG
ATCTTCGACATCGAAGTGCCGGTGGGCGTCATCGTCAGCCCCGCCAACGAC
TTGCAGGCCCTGGCCGACTACGTCGAGGCGGCTCGCAAACCCGGCTCGTCA
CGGCCGACCTTCGCCTCGGTCCACGGCGCCTCGAATGGGCAGGTCACCGAG

FIG. 8 (Cont.)

GTGCATGCCGGTGACCTGTCCCTGGACAAATTCATCGATGCCGCAACCCTG
GCCGAAGCTCCCCGGCTGCCCGCCGCAAACACCCAAGTGCGCACCGTGCTG
CTGACCGGCGCCACCGGCTTCCTCGGGCGCTACCTGGCCCTGGAATGGCTG
GAGCGGATGGACCTGGTCGACGGCAAACTGATCTGCCTGGTCCGGGCCAAG
TCCGACACCGAAGCACGGGCGCGGCTGGACAAGACGTTCGACAGCGGCGAC
CCCGAACTGCTGGCCCACTACCGCGCACTGGCCGGCGACCACCTCGAGGTG
CTCGCCGGTGACAAGGGCGAAGCCGACCTCGGACTGGACCGGCAGACCTGG
CAACGCCTGGCCGACACGGTCGACCTGATCGTCGACCCCGCGGCCCTGGTC
AACCACGTACTGCCATACAGCCAGCTGTTCGGGCCCAACGCGCTGGGCACC
GCCGAGCTGCTGCGGCTGGCGCTCACCTCCAAGATCAAGCCCTACAGCTAC
ACCTCGACAATCGGTGTCGCCGACCAGATCCCGCCGTCGGCGTTCACCGAG
GACGCCGACATCCGGGTCATCAGCGCCACCCGCGCGGTCGACGACAGCTAC
GCCAATGGCTACTCGAACAGCAAGTGGGCCGGCGAGGTGCTGTTGCGCGAG
GCGCATGACCTGTGTGGCCTGCCGGTTGCGGTGTTCCGCTGCGACATGATC
CTGGCCGACACCACATGGGCGGGACAGCTCAATGTGCCGGACATGTTCACC
CGGATGATCCTGAGCCTGGCGGCCACCGGTATCGCGCCGGGTTCGTTCTAT
GAGCTTGCGGCCGACGGCGCCCGGCAACGCGCCCACTATGACGGTCTGCCC
GTCGAGTTCATCGCCGAGGCGATTTCGACTTTGGGTGCGCAGAGCCAGGAT
GGTTTCCACACGTATCACGTGATGAACCCCTACGACGACGGCATCGGACTC
GACGAGTTCGTCGACTGGCTCAACGAGTCCGGTTGCCCCATCCAGCGCATC
GCTGACTATGGCGACTGGCTGCAGCGCTTCGAAACCGCACTGCGCGCACTG
CCCGATCGGCAGCGGCACAGCTCACTGCTGCCGCTGTTGCACAACTATCGG
CAGCCGGAGCGGCCCGTCCGCGGGTCGATCGCCCCTACCGATCGCTTCCGG
GCAGCGGTGCAAGAGGCCAAGATCGGCCCCGACAAAGACATTCCGCACGTC
GGCGCGCCGATCATCGTGAAGTACGTCAGCGACCTGCGCCTACTCGGCCTG
CTCTGA

Amino acid sequence (SEQ ID NO:72)

>uniprot|B2HN69|B2HN69_MYCMM Fatty-acid-CoA ligase FadD9

MSPITREERLERRIQDLYANDPQFAAAKPATAITAAIERPGLPLPQIIET
VMTGYADRPALAQRSVEFVTDAGTGHTTLRLLPHFETISYGELWDRISAL
ADVLSTEQTVKPGDRVCLLGFNSVDYATIDMTLARLGAVAVPLQTSAAIT
QLQPIVAETQPTMIAASVDALADATELALSGQTATRVLVFDHHRQVDAHR
AAVESARERLAGSAVVETLAEAIARGDVPRGASAGSAPGTDVSDDSLALL
IYTSGSTGAPKGAMYPRRNVATFWRKRTWFEGGYEPSITLNFMPMSHVMG
RQILYGTLCNGGTAYFVAKSDLSTLFEDLALVRPTELTFVPRVWDMVFDE
FQSEVDRRLVDGADRVALEAQVKAEIRNDVLGGRYTSALTGSAPISDEMK

FIG. 8 (Cont.)

AWVEELLDMHLVEGYGSTEAGMILIDGAIRRPAVLDYKLVDVPDLGYFLT
DRPHPRGELLVKTDSLFPGYYQRAEVTADVFDADGFYRTGDIMAEVGPEQ
FVYLDRRNNVLKLSQGEFVTVSKLEAVFGDSPLVRQIYIYGNSARAYLLA
VIVPTQEALDAVPVEELKARLGDSLQEVAKAAGLQSYEIPRDFIIETTPW
TLENGLLTGIRKLARPQLKKHYGELLEQIYTDLAHGQADELRSLRQSGAD
APVLVTVCRAAAALLGGSASDVQPDAHFTDLGGDSLSALSFTNLLHEIFD
IEVPVGVIVSPANDLQALADYVEAARKPGSSRPTFASVHGASNGQVTEVH
AGDLSLDKFIDAATLAEAPRLPAANTQVRTVLLTGATGFLGRYLALEWLE
RMDLVDGKLICLVRAKSDTEARARLDKTFDSGDPELLAHYRALAGDHLEV
LAGDKGEADLGLDRQTWQRLADTVDLIVDPAALVNHVLPYSQLFGPNALG
TAELLRLALTSKIKPYSYTSTIGVADQIPPSAFTEDADIRVISATRAVDD
SYANGYSNSKWAGEVLLREAHDLCGLPVAVFRCDMILADTTWAGQLNVPD
MFTRMILSLAATGIAPGSFYELAADGARQRAHYDGLPVEFIAEAISTLGA
QSQDGFHTYHVMNPYDDGIGLDEFVDWLNESGCPIQRIADYGDWLQRFET
ALRALPDRQRHSSLLPLLHNYRQPERPVRGSIAPTDRFRAAVQEAKIGPD
KDIPHVGAPIIVKYVSDLRLLGLL

O69484

Nucleotide sequence (SEQ ID NO:73)

>uniprot|O69484|O69484_MYCLE Putative Acyl-CoA synthetase

ATGTCGACTATCACTAAGCAGGAAAAGCAGCTCGCACGCCGCGTTGACGAC
CTCACCGCCAACGACCCGCAGTTCGCCGCCGCCAAACCCGACCCGGCGGTA
GCCGCCGCCCTTGCCCAGCCCGGGCTTCGACTGCCCCAAATCATCCAGACC
GCGCTGGACGGTTACGCGGAGCGGCCGGCACTGGGCCAGCGCGTCGCCGAG
TTCACCAAAGACCCTAAGACCGGACGCACCTCGATGGAGCTGCTCCCCAGC
TTTGAGACCATCACCTACCGCCAGTTGGGCGACCGTGTCGGAGCGCTGGCG
CGCGCCTGGAGGCACGACCTACTGCACGCCGGCTACCGGGTCTGCGTGCTA
GGTTTCAACAGTGTCGATTACGCCATCATCGACATGGCGCTCGGCGTGATT
GGTGCTGTGGCGGTTCCACTGCAGACCAGTGCGGCGATCACCCAGCTGCAG
TCGATCGTGACCGAGACCGAACCCAGTATGATCGCGACGAGCGTAAACCAG
CTGCCCGATACTGTCGAGCTGATCCTGTCTGGCCAGGCGCCAGCGAAGCTC
GTTGTGTTTGACTACCACCCCGAGGTCGACGAGCAGCATGACGCAGTGGCA
ACCGCCCGGGCGCGGTTGGCGGACAGTAGCGTGGTGGTCGAGAGCCTGACC
GAGGTCCTCGGTCGCGGCAAGACGCTGCCAGCTACGCCGATCCCCGTGGCC
GATGACTCTGCTGACCCGTTGGCGTTGCTGATCTACACATCTGGCAGCACC

FIG. 8 (Cont.)

GGCGCACCCAAGGGCGCGATGTATCTGCAAAGCAATGTCGGCAAGATGTGG
CGCCGGTCAGACGGAAACTGGTTCGGGCCAACCGCCGCGTCAATCACTCTT
AACTTCATGCCGATGAGCCACGTCATGGGCCGCGGAATCCTCTACGGCACG
CTCGGTAACGGCGGCACGGCTTACTTCGCCGCCCGCAGCGACCTCTCGACG
CTGCTGGAGGATCTCAAGCTGGTGCGGCCGACCGAGTTGAACTTTGTACCG
CGCATCTGGGAAACCCTCTACGATGAATCCAAACGCGCAGTTGACCGTCGG
TTAGCCAACAGCGGCTCCGCCGACCGTGCAGCCATCAAAGCCGAAGTTATG
GATGAACAGCGGCAATCCCTGCTGGGAGGACGGTACATCGCGGCTATGACG
GGCTCGGCGCCAACCTCCCCGGAGTTGAAACACGGGGTCGAGTCCCTACTC
GAAATGCATCTGTTGGAAGGCTACGGCTCCACCGAAGCCGGCATGGTCTTG
TTTGACGGCGAAGTGCAACGTCCGCCGGTTATCGATTACAAGCTGGTCGAC
GTTCCGGATTTGGGCTACTTCAGCACCGACCAGCCTTATCCGAGAGGTGAA
TTGCTGCTCAAGACCCAGAACATGTTCCCCGGCTACTACAAGCGTCCTGAG
GTTACCGCCACCGTGTTCGACAGCGACGGTTACTACCAGACCGGAGACATT
GTCGCCGAAGTCGGTCCCGACCGGCTCGTGTACGTCGATCGCCGCAACAAC
GTGCTGAAACTCGCGCAGGGCCAGTTCGTCACCGTCGCGAAACTCGAGGCA
GCGTTCAGCAATAGCCCACTGGTCCGGCAGATCTACATCTATGGCAACAGC
GCACACCCCTACCTGTTGGCTGTTGTGGTGCCGACCGAGGATGCGTTGGCT
ACCAATGACATTGAGGTGCTCAAACCGCTGATTATCGATTCTTTACAGAAA
GTAGCGAAAGAAGCCGACCTGCAGTCCTACGAGGTGCCGCGCGACTTAATC
GTCGAGACTACACCGTTCAGCCTGGAGAATGGCCTGCTCACCGGTATTCGC
AAGCTGGCGTGGCCGAAGCTCAAGCAGCACTACGGCGCGCGACTCGAACAG
CTCTACGCCGATCTGGTTGAAGGTCAGGCAAATGCACTGCACGTGCTAAAA
CAAAGCGTGGCGAACGCTCCGGTACTGCAGACGGTGAGCCGAGCCGTGGGC
ACCATTCTGGGAGTGGCGACCACCGATTTGCCGTCGAATGCGCACTTCACC
GACTTAGGAGGAGACTCGTTGTCCGCGCTGACATTCGGTAGCCTGCTACGC
GAACTCTTCGACATCGATGTGCCGGTGGGCGTCATTGTCAGCCCTGTCAAC
AACTTGGTGGCGATCGCCGACTACATCGAGCGCGAGCGGCAGGGCACGAAG
CGGCCCACTTTCATTGCCATACACGGTCGTGACGCTGGCAAAGTGCATGCC
AGTGACCTCACTCTAGACAAATTCATCGATGTATCAACGCTGACTGCCGCG
CCCGTATTGGCGCAACCCGGCACCGAGGTGCGCACCGTCCTGTTGACCGGC
GCTACCGGCTTCCTGGGGCGCTACTTGGCCCTGAAATGGCTCGAACGGATG
GACCTGGTCGAAGGGAAGGTAATCGCTCTGGTAAGAGCCAAGTCCAACGAG
GACGCTCGGGCCCGGCTCGACAAGACCTTCGATAGCGGAGACCCCAAACTG
CTGGCGCACTACCAGGAACTGGCAACCGACCACCTGGAGGTCATCGCCGGC
GACAAAGGCGAAGTAGATCTGGAATTGGACCGGCAAACGTGGCGACGACTG
GCCGACACGGTCGATCTGATCGTCGACCCCGCCGCCCTGGTCAACCACGTG
CTGCCGTACAGCGAGCTATTCGGCCCCAATACGTTAGGCACCGCCGAGCTG
ATTCGGATCGCGCTGACCAGTAAGCAAAAGCCGTACATCTATGTGTCGACA

FIG. 8 (Cont.)

ATCGGCGTCGGTAATCAGATTGAGCCAGCAAAATTCACCGAAGACTCCGAC
ATCCGAGTCATTAGCCCGACGCGCAACATCAACAACAACTATGCCAACGGC
TACGGCAACAGCAAGTGGGCCGGCGAAGTGCTGCTGCGCGAAGCTCACGAC
CTATGCGGTCTGCCGGTCACGGTCTTCCGCTGCGACATGATCTTGGCCGAC
ACCAGCTATGCCGGTCAGCTCAACGTCCCCGACATGTTTACTCGAATGATG
CTGAGTCTAGCCGCCACCGGCATCGCACCCGGCTCGTTCTACGAGCTAGAC
GCCGAGAGCAATCGGCAACGCGCCCACTACGACGGTCTGCCCGTCGAGTTC
ATCGCCGAAGCGATCTCCACCCTGGGAGACCAAAGCCTGCACGATCGAGAC
GGGTTCACGACCTATCATGTAATGAACCCGCACGACGACGGCATCGGTATG
GACGAGTTTGTGGACTGGTTAATTGATGCCGGCTGCCCTATACAACGCATC
AACGACTACGACGAATGGCTGCGACGGTTTGAGATTTCGCTGCGCGCCCTG
CCCGAAAGGCAGCGTCACAGCTCACTGTTGCCGTTGTTGCACAACTACCAG
AAGCCGGAGAAGCCATTGCACGGGTCGCTGGCACCCACAATCCGGTTCCGT
ACGGCCGTTCAAAACGCGAACATTGGTCAGGACAAAGATATTCCGCATATC
TCGCCGGCAATCATCGCCAAATATGTCAGCGATCTGCAGCTGCTCGGGCTG
GTTTGA

Amino acid sequence (SEQ ID NO:74)

>uniprot|O69484|O69484_MYCLE Putative Acyl-CoA synthetase

MSTITKQEKQLARRVDDLTANDPQFAAAKPDPAVAAALAQPGLRLPQIIQ
TALDGYAERPALGQRVAEFTKDPKTGRTSMELLPSFETITYRQLGDRVGA
LARAWRHDLLHAGYRVCVLGFNSVDYAIIDMALGVIGAVAVPLQTSAAIT
QLQSIVTETEPSMIATSVNQLPDTVELILSGQAPAKLVVFDYHPEVDEQH
DAVATARARLADSSVVVESLTEVLGRGKTLPATPIPVADDSADPLALLIY
TSGSTGAPKGAMYLQSNVGKMWRRSDGNWFGPTAASITLNFMPMSHVMGR
GILYGTLGNGGTAYFAARSDLSTLLEDLKLVRPTELNFVPRIWETLYDES
KRAVDRRLANSGSADRAAIKAEVMDEQRQSLLGGRYIAAMTGSAPTSPEL
KHGVESLLEMHLLEGYGSTEAGMVLFDGEVQRPPVIDYKLVDVPDLGYFS
TDQPYPRGELLLKTQNMFPGYYKRPEVTATVFDSDGYYQTGDIVAEVGPD
RLVYVDRRNNVLKLAQGQFVTVAKLEAAFSNSPLVRQIYIYGNSAHPYLL
AVVVPTEDALATNDIEVLKPLIIDSLQKVAKEADLQSYEVPRDLIVETTP
FSLENGLLTGIRKLAWPKLKQHYGARLEQLYADLVEGQANALHVLKQSVA
NAPVLQTVSRAVGTILGVATTDLPSNAHFTDLGGDSLSALTFGSLLRELF
DIDVPVGVIVSPVNNLVAIADYIERERQGTKRPTFIAIHGRDAGKVHASD
LTLDKFIDVSTLTAAPVLAQPGTEVRTVLLTGATGFLGRYLALKWLERMD
LVEGKVIALVRAKSNEDARARLDKTFDSGDPKLLAHYQELATDHLEVIAG
DKGEVDLELDRQTWRRLADTVDLIVDPAALVNHVLPYSELFGPNTLGTAE

FIG. 8 (Cont.)

LIRIALTSKQKPYIYVSTIGVGNQIEPAKFTEDSDIRVISPTRNINNNYA
NGYGNSKWAGEVLLREAHDLCGLPVTVFRCDMILADTSYAGQLNVPDMFT
RMMLSLAATGIAPGSFYELDAESNRQRAHYDGLPVEFIAEAISTLGDQSL
HDRDGFTTYHVMNPHDDGIGMDEFVDWLIDAGCPIQRINDYDEWLRRFEI
SLRALPERQRHSSLLPLLHNYQKPEKPLHGSLAPTIRFRTAVQNANIGQD
KDIPHISPAIIAKYVSDLQLLGLV

Q10896

Nucleotide sequence (SEQ ID NO:75)

>uniprot|Q10896|Q10896_MYCTU PROBABLE PEPTIDE
SYNTHETASE NRP (PEPTIDE SYNTHASE)

ATGTCGATCAACGATCAGCGACTGACACGCCGCGTCGAGGACCTATACGCC
AGCGACGCCCAGTTCGCCGCCGCCAGTCCCAACGAGGCGATCACCCAGGCG
ATCGACCAGCCCGGGGTCGCGCTTCCACAGCTCATCCGTATGGTCATGGAG
GGCTACGCCGATCGGCCGGCACTCGGCCAGCGTGCGCTCCGCTTCGTCACC
GACCCCGACAGCGGCCGCACCATGGTCGAGCTACTGCCGCGGTTCGAGACC
ATCACCTACCGCGAACTGTGGGCCCGCGCCGGCACATTGGCCACCGCGTTG
AGCGCTGAGCCCGCGATCCGGCCGGGCGACCGGGTTTGCGTGCTGGGCTTC
AACAGCGTCGACTACACAACCATCGACATCGCGCTGATCCGGTTGGGCGCC
GTGTCGGTTCCACTGCAGACCAGTGCGCCGGTCACCGGGTTGCGCCCGATC
GTCACCGAGACCGAGCCGACGATGATCGCCACCAGCATCGACAATCTTGGC
GACGCCGTCGAAGTGCTGGCCGGTCACGCCCCGGCCCGGCTGGTCGTATTC
GATTACCACGGCAAGGTTGACACCCACCGCGAGGCCGTCGAAGCCGCCCGA
GCTCGGTTGGCCGGCTCGGTGACCATCGACACACTTGCCGAACTGATCGAA
CGCGGCAGGGCGCTGCCGGCCACACCCATTGCCGACAGCGCCGACGACGCG
CTGGCGCTGCTGATTTACACCTCGGGTAGTACCGGCGCACCCAAAGGCGCC
ATGTATCGCGAGAGCCAGGTGATGAGCTTCTGGCGCAAGTCGAGTGGCTGG
TTCGAGCCGAGCGGTTACCCCTCGATCACGCTGAACTTCATGCCGATGAGC
CACGTCGGGGGCCGTCAGGTGCTCTACGGGACGCTTTCCAACGGCGGTACC
GCCTACTTCGTCGCCAAGAGCGACCTGTCGACGCTGTTCGAGGACCTCGCC
CTGGTGCGGCCCACAGAATTGTGCTTCGTGCCGCGCATCTGGGACATGGTG
TTCGCAGAGTTCCACAGCGAGGTCGACCGCCGCTTGGTGGACGGCGCCGAT
CGAGCGGCGCTGGAAGCGCAGGTGAAGGCCGAGCTGCGGGAGAACGTGCTC
GGCGGACGGTTTGTCATGGCGCTGACCGGTTCCGCGCCGATCTCCGCTGAG
ATGACGGCGTGGGTCGAGTCCCTGCTGGCCGACGTGCATTTGGTGGAGGGT
TACGGCTCCACCGAGGCCGGGATGGTCCTGAACGACGGCATGGTGCGGCGC

FIG. 8 (Cont.)

CCCGCGGTGATCGACTACAAGCTGGTCGACGTGCCCGAGCTGGGCTACTTC
GGCACCGATCAGCCCTACCCCCGGGGCGAGCTGCTGGTCAAGACGCAAACC
ATGTTCCCCGGCTACTACCAGCGCCCGGATGTCACCGCCGAGGTGTTCGAC
CCCGACGGCTTCTACCGGACCGGGGACATCATGGCCAAAGTAGGCCCCGAC
CAGTTCGTCTACCTCGACCGCCGCAACAACGTGCTAAAGCTCTCCCAGGGC
GAGTTCATCGCCGTGTCGAAGCTCGAGGCGGTGTTCGGCGACAGCCCGCTG
GTCCGACAGATCTTCATCTACGGCAACAGTGCCCGGGCCTACCCGCTGGCG
GTGGTTGTCCCGTCCGGGGACGCGCTTTCTCGCCATGGCATCGAGAATCTC
AAGCCCGTGATCAGCGAGTCCCTGCAGGAGGTAGCGAGGGCGGCCGGCCTG
CAATCCTACGAGATTCCACGCGACTTCATCATCGAAACCACGCCGTTCACC
CTGGAGAACGGCCTGCTCACCGGCATCCGCAAGCTGGCACGCCCGCAGTTG
AAGAAGTTCTATGGCGAACGTCTCGAGCGGCTCTATACCGAGCTGGCCGAT
AGCCAATCCAACGAGCTGCGCGAGCTGCGGCAAAGCGGTCCCGATGCGCCG
GTGCTTCCGACGCTGTGCCGTGCCGCGGCTGCGTTGCTGGGCTCTACCGCT
GCGGATGTGCGGCCGGACGCGCACTTCGCCGACCTGGGTGGTGACTCGCTC
TCGGCGCTGTCGTTGGCCAACCTGCTGCACGAGATCTTCGGCGTCGACGTG
CCGGTGGGTGTCATTGTCAGCCCGGCAAGCGACCTGCGGGCCCTGGCCGAC
CACATCGAAGCAGCGCGCACCGGCGTCAGGCGACCCAGCTTCGCCTCGATA
CACGGTCGCTCCGCGACGGAAGTGCACGCCAGCGACCTCACGCTGGACAAG
TTCATCGACGCTGCCACCCTGGCCGCAGCCCCGAACCTGCCGGCACCGAGC
GCCCAAGTGCGCACCGTACTGCTGACCGGCGCCACCGGCTTTTTGGGTCGC
TACCTGGCGCTGGAATGGCTCGACCGCATGGACCTGGTCAACGGCAAGCTG
ATCTGCCTGGTCCGCGCCAGATCCGACGAGGAAGCACAAGCCCGGCTGGAC
GCGACGTTCGATAGCGGCGACCCGTATTTGGTGCGGCACTACCGCGAATTG
GGCGCCGGCCGCCTCGAGGTGCTCGCCGGCGACAAGGGCGAGGCCGACCTG
GGCCTGGACCGGGTCACCTGGCAGCGGCTAGCCGACACGGTGGACCTGATC
GTGGACCCCGCGGCCCTGGTCAACCACGTGCTGCCGTATAGCCAGCTGTTC
GGCCCAAACGCGGCGGGCACCGCCGAGTTGCTTCGGCTGGCGCTGACCGGC
AAGCGCAAGCCATACATCTACACCTCGACGATCGCCGTGGGCGAGCAGATC
CCGCCGGAGGCGTTCACCGAGGACGCCGACATCCGGGCCATCAGCCCGACC
CGCAGGATCGACGACAGCTACGCCAACGGCTACGCGAACAGCAAGTGGGCC
GGCGAGGTGCTGCTGCGCGAAGCTCACGAGCAGTGCGGCCTGCCGGTGACG
GTCTTCCGCTGCGACATGATCCTGGCCGACACCAGCTATACCGGTCAGCTC
AACCTGCCGGACATGTTCACCCGGCTGATGCTGAGCCTGGCCGCTACCGGC
ATCGCACCCGGTTCGTTCTATGAGCTGGATGCGCACGGCAATCGGCAACGC
GCCCACTATGACGGCTTGCCGGTCGAATTCGTCGCAGAAGCCATTTGCACC
CTTGGGACACATAGCCCGGACCGTTTTGTCACCTACCACGTGATGAACCCC
TACGACGACGGCATCGGGCTGGACGAGTTCGTCGACTGGCTCAACTCCCCA
ACTAGCGGGTCCGGTTGCACGATCCAGCGGATCGCCGACTACGGCGAGTGG

FIG. 8 (Cont.)

CTGCAGCGGTTCGAGACTTCGCTGCGTGCCTTGCCGGATCGCCAGCGCCAC
GCCTCGCTGCTGCCCTTGCTGCACAACTACCGAGAGCCTGCAAAGCCGATA
TGCGGGTCAATCGCGCCCACCGACCAGTTCCGCGCTGCCGTCCAAGAAGCG
AAAATCGGTCCGGACAAAGACATTCCGCACCTCACGGCGGCGATCATCGCG
AAGTACATCAGCAACCTGCGACTGCTCGGGCTGCTGTGA

Amino acid sequence (SEQ ID NO:76)

>uniprot|Q10896|Q10896_MYCTU PROBABLE PEPTIDE SYNTHETASE NRP (PEPTIDE SYNTHASE)

MSINDQRLTRRVEDLYASDAQFAAASPNEAITQAIDQPGVALPQLIRMVM
EGYADRPALGQRALRFVTDPDSGRTMVELLPRFETITYRELWARAGTLAT
ALSAEPAIRPGDRVCVLGFNSVDYTTIDIALIRLGAVSVPLQTSAPVTGL
RPIVTETEPTMIATSIDNLGDAVEVLAGHAPARLVVFDYHGKVDTHREAV
EAARARLAGSVTIDTLAELIERGRALPATPIADSADDALALLIYTSGSTG
APKGAMYRESQVMSFWRKSSGWFEPSGYPSITLNFMPMSHVGGRQVLYGT
LSNGGTAYFVAKSDLSTLFEDLALVRPTELCFVPRIWDMVFAEFHSEVDR
RLVDGADRAALEAQVKAELRENVLGGRFVMALTGSAPISAEMTAWVESLL
ADVHLVEGYGSTEAGMVLNDGMVRRPAVIDYKLVDVPELGYFGTDQPYPR
GELLVKTQTMFPGYYQRPDVTAEVFDPDGFYRTGDIMAKVGPDQFVYLDR
RNNVLKLSQGEFIAVSKLEAVFGDSPLVRQIFIYGNSARAYPLAVVVPSG
DALSRHGIENLKPVISESLQEVARAAGLQSYEIPRDFIIETTPFTLENGL
LTGIRKLARPQLKKFYGERLERLYTELADSQSNELRELRQSGPDAPVLPT
LCRAAAALLGSTAADVRPDAHFADLGGDSLSALSLANLLHEIFGVDVPVG
VIVSPASDLRALADHIEAARTGVRRPSFASIHGRSATEVHASDLTLDKFI
DAATLAAAPNLPAPSAQVRTVLLTGATGFLGRYLALEWLDRMDLVNGKLI
CLVRARSDEEAQARLDATFDSGDPYLVRHYRELGAGRLEVLAGDKGEADL
GLDRVTWQRLADTVDLIVDPAALVNHVLPYSQLFGPNAAGTAELLRLALT
GKRKPYIYTSTIAVGEQIPPEAFTEDADIRAISPTRRIDDSYANGYANSK
WAGEVLLREAHEQCGLPVTVFRCDMILADTSYTGQLNLPDMFTRLMLSLA
ATGIAPGSFYELDAHGNRQRAHYDGLPVEFVAEAICTLGTHSPDRFVTYH
VMNPYDDGIGLDEFVDWLNSPTSGSGCTIQRIADYGEWLQRFETSLRALP
DRQRHASLLPLLHNYREPAKPICGSIAPTDQFRAAVQEAKIGPDKDIPHL
TAAIIAKYISNLRLLGLL

FIG. 8 (Cont.)

Q5YY80

Nucleotide sequence (SEQ ID NO:77)

>uniprot|Q5YY80|Q5YY80_NOCFA Putative carboxylic acid reductase

GTGGAATCCACACGAGCGACGCGTCTGCGGCAGCGGATCGCCGCCCTGTAC
GCCGACGACGCGCAGGTGCGCGACGCGCGGCCGGACGAGGCGATCAGCACG
GCGCTGCGGGAACCGGGCCTGCGGCTGCGCGAGCTCGTCGCCACCGTGGTC
GACGGCTACCGCGACCGGCCCGCGCTGGCCGCGCGCTCGGTGCAGCCGGCC
GTCGACGCCGCGACCGGTGCCTGCGTGGCACGGTTGCTGCCCGAGTACACG
ACGATGAGTTACGGCGAGCTCGGCCTGCGGTTGCGCGCGGTGGCCGCGGCC
TGGCAGCACGACGACGAAACCCGCCTTCGCCCCGGCGAATTCGTCGCGACC
CTGGGTTTCACCAGCCCCGACTACGCCGTCGTCGACCTGGCCTGCGTGTGG
GCGGGCGCGGTGGCGGTGCCGCTGCAGGCGAGCGCGTCGGTGACGCAGCTG
ACCGCCATCCTGGCCGAGACCGCGCCCGCAATCCTGGCCACCGGCCTGGAC
ACGCTGCCGCACGCGGTGGACTGTGTGCTCGCCGGGGCCACGCCACGGGCA
CTGCACGTGTTCGACTTCGACCCCGCCATCGACGCGCAGCGCACGGTGTAC
GAGGCGGCGTGTGCGCGGCTGGCCGGTACCGGTGTGCGCGTGCGCACGCTC
GCCGAGGTCGAGGACCGCGGCCGGGCGCTGCCGCCTGCCGTGATCGACGAC
GGCCCCGGCGACGACCGGCTCGCCCTGTTGATCTACACCTCCGGCAGTACC
GGCACGCCCAAGGGGGCGATGTACACCGAGCGGCTGGTCGCGCTGATGTGG
CTGGGCCAGCCGCAGGTCGCCGCGCTCACCGTCAACTACCTGCCGCTCAGC
CACGTCGCCGGGCGGCTGGCGCTGTTCGGGCTGCTCGCGCGCGGCGGCACC
GCCTACTTCACCGCGCGCGCCGACATGTCCACGCTGTTCGAGGATCTGGCG
CTGGCCAGGCCGACCGAGCTGTTCGTGGTGCCGCGCGTGTGCGAGATGGTG
CTGCAACGATTCCAGACCGAGCGGCTGCGGCGCCAGGCCGACGACGACCGG
GTCAAGGCCGACCTGCGCCTCGAACTGTTCGGCGACCGGCTGCTCTCGGTG
GTGTGCGGCAGCGCGCCGCTGGCCCCGGAGCTGAAGGCGTTCATGGAATCG
GTGCTCGACCTGACCCTGCACGACGGCTACGGCTCCACCGAGGCGGGCGGC
AGCGTGGTCATCGACACCACCGTGCGCAGGCCGCCGGTGCTGGACTACCGG
CTCGCCGACGTGCCCGAACTGGGCTATTTCCGTACCGACAAGCCGCATCCG
CGCGGCGAGCTGCTGCTCAAGACCACCACCATGATCCCCGGCTACTACCGG
CGGCCGGAGCTCAACGCCCAGATCTTCGACGAGGACGGCTTCTACCGCACC
GGCGACGTGGTCGCGGAACTGGCGCCGGACCGGCTCGTGTACGTCGATCGC
CGCAACAATGTGCTCAAGCTGGCGCAGGGCGAGTTCGTCACCATCGCCCGG
CTGGAGGCGATCTTCGCCAACAGTCCGCTGGTGCGCCAGATCTTCGTCTAC
GGCAACAGCGAACGCGCCTATCTGCTGGCGGTGATCGTGCCGAGCCGACAG
GCGATGGCGGGCGATCCGGCCACGCTGAAGACGCGGATCGCGGAGTCGTTG
CAGCTCATCGGCCGGGACGCCGAGCTGGAGGCCTACGAGATCCCGCGCGAC

FIG. 8 (Cont.)

TTCCTGATCGAGACCGAGCCGTTCACCACCGAATCCGGGCTGCTCTCGGGC
ATCGGCAAGATCCTGCGTCCCGCCGTCGAGGCGCGCTATCGCGACCGGCTC
GAACAGCTCTACGCCGACCTGGCCGCGGCCCAGCAGGACGAGCTGGCGGCG
CTGCGCCGCGAGGCCGGGCAGCGTCCGGTGCTCGAGACCGTCACCCGCGCG
GCCGCCGCGATCCTCGGCGGCACGGCGAGCGACCTGAGCCCGGCCGCGCAC
TTCACCGATCTCGGCGGCGATTCGCTGGCGGCGCTGGCGCTGTCGAACCTG
CTGCGTGAGATCTTCGCCGTCGAGGTGCCGGTCGGCGTCATCACCGGCCCC
GCGACCGACCTCCGTGGCCTGGCCGCCCACATCGCCGCGGAACGCGAAAAC
CGCACCGAGACACCGCTGTTCGACCGGGTGCATCCCGACCAGATCCTGATC
CGGGCCACCGACCTCGCCCTGGAGAAGTTCTTCGACGCCGAGGAGTTGGCC
GCCGCGGCCACCGCCGCGCCGCCGGTCGCCGAGCCCCGGGTGGTGCTGCTG
ACCGGTGCCAACGGCTATCTCGGCCGGTTTCTGTGCCTGGAATGGCTGGAA
CGGCTCGACCGCGTCGACGGACGGCTGATCTGCCTGGTGCGCGGCGCGGAC
GAGGCCGCCGCGCTGGCCCGCCTGGAAGCCGCCTTCGACAGCGGCGATCCC
GAATTGGTGCGCCGCTTCAAGGAATTGGCCCAGCGCAGGCTCACCGTGGTG
GCCGGCGACATCGGCGAGCCCGGCCTGGGCCTGGCCACCGCCACGTGGCGA
CGGCTCGCCGCCGAGGTCGAGCACATCGTGCACCCGGCCGCGCTGGTCAAC
CACGTGCTGCCCTACCGGCAGCTGTTCGGGCCCAACGTGGCGGGCACCGCG
GAGATCCTGCGGCTCGCGCTCACCGAGCGGCGCAAGCCGATCGACTTCCTG
TCCACGGTCGCCGTTGCCGCGCAGATACCCGCCGACCGGTTCGCCGAGGAC
GGCGACATCCGCGTGATCAGCCCGACCCGCACGGTGGACCGCGGCTACGCC
AACGGCTACGGCAACAGCAAATGGGCCGCCGAGGTGTTGCTGCGTGCGGCG
CACGACCGCTTCGATCTCCCGGTGGCGGTGTTCCGCTCGGACATGATCCTG
GCCCACGGCAGCTTCGCCGGACAGCTCAACATCCCCGACGTGTTCACCCGG
CTGCTGCTCAGCCTGCTGGTCACCGGTATCGCGCCCGCCTCGTTCCACGCC
GCGACGGTCACCGGCGAGCGCCCGCGCGCCCACTACGACGGGCTGCCCGCG
GACTTCACCGCTGCCGCGATCACCGCGCTCGGGGCGCGCACCGCGGGATTC
CACACCTACGACGTGCTCAACCCGCACGACGACGGCATCAGCCTGGACACC
TTCGTGGACTGGCTGATCGAGGCCGGACATCCCATCGAACGCATCCCCGAG
CACAGCGAGTGGGTCACCCGTTTCGAGACGGCGTTGCACGCCCTGCCCGAA
CGTCAGCGCAAACACTCGCTGCTCCCGCTGTTGCACGCCTACCGCAGGCCG
GTGCCCGCGCTGCGCGGCTCGGCGCTGCCCGCCGCGGAGTTCCGGGCGGCG
GTGCGGGCCGCAGGCATCACCGCCGACGGTGACATCCCGCACCTGACGCGC
GCGCTGATCGAGAAGTACGTCGCCGATCTCCGCCTGCACGGACTGTTGTAG

Amino acid sequence (SEQ ID NO:78)

>uniprot|Q5YY80|Q5YY80_NOCFA Putative carboxylic acid reductase

FIG. 8 (Cont.)

VESTRATRLRQRIAALYADDAQVRDARPDEAISTALREPGLRLRELVATV
VDGYRDRPALAARSVQPAVDAATGACVARLLPEYTTMSYGELGLRLRAVA
AAWQHDDETRLRPGEFVATLGFTSPDYAVVDLACVWAGAVAVPLQASASV
TQLTAILAETAPAILATGLDTLPHAVDCVLAGATPRALHVFDFDPAIDAQ
RTVYEAACARLAGTGVRVRTLAEVEDRGRALPPAVIDDGPGDDRLALLIY
TSGSTGTPKGAMYTERLVALMWLGQPQVAALTVNYLPLSHVAGRLALFGL
LARGGTAYFTARADMSTLFEDLALARPTELFVVPRVCEMVLQRFQTERLR
RQADDDRVKADLRLELFGDRLLSVVCGSAPLAPELKAFMESVLDLTLHDG
YGSTEAGGSVVIDTTVRRPPVLDYRLADVPELGYFRTDKPHPRGELLLKT
TTMIPGYYRRPELNAQIFDEDGFYRTGDVVAELAPDRLVYVDRRNNVLKL
AQGEFVTIARLEAIFANSPLVRQIFVYGNSERAYLLAVIVPSRQAMAGDP
ATLKTRIAESLQLIGRDAELEAYEIPRDFLIETEPFTTESGLLSGIGKIL
RPAVEARYRDRLEQLYADLAAAQQDELAALRREAGQRPVLETVTRAAAAI
LGGTASDLSPAAHFTDLGGDSLAALALSNLLREIFAVEVPVGVITGPATD
LRGLAAHIAAERENRTETPLFDRVHPDQILIRATDLALEKFFDAEELAAA
ATAAPPVAEPRVVLLTGANGYLGRFLCLEWLERLDRVDGRLICLVRGADE
AAALARLEAAFDSGDPELVRRFKELAQRRLTVVAGDIGEPGLGLATATWR
RLAAEVEHIVHPAALVNHVLPYRQLFGPNVAGTAEILRLALTERRKPIDF
LSTVAVAAQIPADRFAEDGDIRVISPTRTVDRGYANGYGNSKWAAEVLLR
AAHDRFDLPVAVFRSDMILAHGSFAGQLNIPDVFTRLLLSLLVTGIAPAS
FHAATVTGERPRAHYDGLPADFTAAAITALGARTAGFHTYDVLNPHDDGI
SLDTFVDWLIEAGHPIERIPEHSEWVTRFETALHALPERQRKHSLLPLLH
AYRRPVPALRGSALPAAEFRAAVRAAGITADGDIPHLTRALIEKYVADLR
LHGLL

Q6RKB1

Nucleotide sequence (SEQ ID NO:79)

>uniprot|Q6RKB1|Q6RKB1_9NOCA ATP/NADPH-dependent carboxylic acid reductase

ATGGCAGTGGATTCACCGGATGAGCGGCTACAGCGCCGCATTGCACAGTTG
TTTGCAGAAGATGAGCAGGTCAAGGCCGCACGTCCGCTCGAAGCGGTGAGC
GCGGCGGTGAGCGCGCCCGGTATGCGGCTGGCGCAGATCGCCGCCACTGTT
ATGGCGGGTTACGCCGACCGCCCGGCCGCCGGGCAGCGTGCGTTCGAACTG
AACACCGACGACGCGACGGGCCGCACCTCGCTGCGGTTACTTCCCCGATTC
GAGACCATCACCTATCGCGAACTGTGGCAGCGAGTCGGCGAGGTTGCCGCG

FIG. 8 (Cont.)

GCCTGGCATCATGATCCCGAGAACCCCTTGCGCGCAGGTGATTTCGTCGCC
CTGCTCGGCTTCACCAGCATCGACTACGCCACCCTCGACCTGGCCGATATC
CACCTCGGCGCGGTTACCGTGCCGTTGCAGGCCAGCGCGGCGGTGTCCCAG
CTGATCGCTATCCTCACCGAGACTTCGCCGCGGCTGCTCGCCTCGACCCCG
GAGCACCTCGATGCGGCGGTCGAGTGCCTACTCGCGGGCACCACACCGGAA
CGACTGGTGGTCTTCGACTACCACCCCGAGGACGACGACCAGCGTGCGGCC
TTCGAATCCGCCCGCCGCCGCCTTGCCGACGCGGGCAGCTTGGTGATCGTC
GAAACGCTCGATGCCGTGCGTGCCCGGGGCCGCGACTTACCGGCCGCGCCA
CTGTTCGTTCCCGACACCGACGACGACCCGCTGGCCCTGCTGATCTACACC
TCCGGCAGCACCGGAACGCCGAAGGGCGCGATGTACACCAATCGGTTGGCC
GCCACGATGTGGCAGGGGAACTCGATGCTGCAGGGGAACTCGCAACGGGTC
GGGATCAATCTCAACTACATGCCGATGAGCCACATCGCCGGTCGCATATCG
CTGTTCGGCGTGCTCGCTCGCGGTGGCACCGCATACTTCGCGGCCAAGAGC
GACATGTCGACACTGTTCGAAGACATCGGCTTGGTACGTCCCACCGAGATC
TTCTTCGTCCCGCGCGTGTGCGACATGGTCTTCCAGCGCTATCAGAGCGAG
CTGGACCGGCGCTCGGTGGCGGGCGCCGACCTGGACACGCTCGATCGGGAA
GTGAAAGCCGACCTCCGGCAGAACTACCTCGGTGGGCGCTTCCTGGTGGCG
GTCGTCGGCAGCGCGCCGCTGGCCGCGGAGATGAAGACGTTCATGGAGTCC
GTCCTCGATCTGCCACTGCACGACGGGTACGGGTCGACCGAGGCGGGCGCA
AGCGTGCTGCTCGACAACCAGATCCAGCGGCCGCCGGTGCTCGATTACAAG
CTCGTCGACGTGCCCGAACTGGGTTACTTCCGCACCGACCGGCCGCATCCG
CGCGGTGAGCTGTTGTTGAAGGCGGAGACCACGATTCCGGGCTACTACAAG
CGGCCCGAGGTCACCGCGGAGATCTTCGACGAGGACGGCTTCTACAAGACC
GGCGATATCGTGGCCGAGCTCGAGCACGATCGGCTGGTCTATGTCGACCGT
CGCAACAATGTGCTCAAACTGTCGCAGGGCGAGTTCGTGACCGTCGCCCAT
CTCGAGGCCGTGTTCGCCAGCAGCCCGCTGATCCGGCAGATCTTCATCTAC
GGCAGCAGCGAACGTTCCTATCTGCTCGCGGTGATCGTCCCCACCGACGAC
GCGCTGCGCGGCCGCGACACCGCCACCTTGAAATCGGCACTGGCCGAATCG
ATTCAGCGCATCGCCAAGGACGCGAACCTGCAGCCCTACGAGATTCCGCGC
GATTTCCTGATCGAGACCGAGCCGTTCACCATCGCCAACGGACTGCTCTCC
GGCATCGCGAAGCTGCTGCGCCCCAATCTGAAGGAACGCTACGGCGCTCAG
CTGGAGCAGATGTACACCGATCTCGCGACAGGCCAGGCCGATGAGCTGCTC
GCCCTGCGCCGCGAAGCCGCCGACCTGCCGGTGCTCGAAACCGTCAGCCGG
GCAGCGAAAGCGATGCTCGGCGTCGCCTCCGCCGATATGCGTCCCGACGCG
CACTTCACCGACCTGGGCGGCGATTCCCTTTCCGCGCTGTCGTTCTCGAAC
CTGCTGCACGAGATCTTCGGGGTCGAGGTGCCGGTGGGTGTCGTCGTCAGC
CCGGCGAACGAGCTGCGCGATCTGGCGAATTACATTGAGGCGGAACGCAAC
TCGGGCGCGAAGCGTCCCACCTTCACCTCGGTGCACGGCGGCGGTTCCGAG
ATCCGCGCCGCCGATCTGACCCTCGACAAGTTCATCGATGCCCGCACCCTG

FIG. 8 (Cont.)

GCCGCCGCCGACAGCATTCCGCACGCGCCGGTGCCAGCGCAGACGGTGCTG
CTGACCGGCGCGAACGGCTACCTCGGCCGGTTCCTGTGCCTGGAATGGCTG
GAGCGGCTGGACAAGACGGGTGGCACGCTGATCTGCGTCGTGCGCGGTAGT
GACGCGGCCGCGGCCCGTAAACGGCTGGACTCGGCGTTCGACAGCGGCGAT
CCCGGCCTGCTCGAGCACTACCAGCAACTGGCCGCACGGACCCTGGAAGTC
CTCGCCGGTGATATCGGCGACCCGAATCTCGGTCTGGACGACGCGACTTGG
CAGCGGTTGGCCGAAACCGTCGACCTGATCGTCCATCCCGCCGCGTTGGTC
AACCACGTCCTTCCCTACACCCAGCTGTTCGGCCCCAATGTCGTCGGCACC
GCCGAAATCGTCCGGTTGGCGATCACGGCGCGGCGCAAGCCGGTCACCTAC
CTGTCGACCGTCGGAGTGGCCGACCAGGTCGACCCGGCGGAGTATCAGGAG
GACAGCGACGTCCGCGAGATGAGCGCGGTGCGCGTCGTGCGCGAGAGTTAC
GCCAACGGCTACGGCAACAGCAAGTGGGCGGGGGAGGTCCTGCTGCGCGAA
GCACACGATCTGTGTGGCTTGCCGGTCGCGGTGTTCCGTTCGGACATGATC
CTGGCGCACAGCCGGTACGCGGGTCAGCTCAACGTCCAGGACGTGTTCACC
CGGCTGATCCTCAGCCTGGTCGCCACCGGCATCGCGCCGTACTCGTTCTAC
CGAACCGACGCGGACGGCAACCGGCAGCGGGCCCACTATGACGGCTTGCCG
GCGGACTTCACGGCGGCGGCGATCACCGCGCTCGGCATCCAAGCCACCGAA
GGCTTCCGGACCTACGACGTGCTCAATCCGTACGACGATGGCATCTCCCTC
GATGAATTCGTCGACTGGCTCGTCGAATCCGGCCACCCGATCCAGCGCATC
ACCGACTACAGCGACTGGTTCCACCGTTTCGAGACGGCGATCCGCGCGCTG
CCGGAAAAGCAACGCCAGGCCTCGGTGCTGCCGTTGCTGGACGCCTACCGC
AACCCCTGCCCGGCGGTCCGCGGCGCGATACTCCCGGCCAAGGAGTTCCAA
GCGGCGGTGCAAACAGCCAAAATCGGTCCGGAACAGGACATCCCGCATTTG
TCCGCGCCACTGATCGATAAGTACGTCAGCGATCTGGAACTGCTTCAGCTG
CTCTGA

Amino acid sequence (SEQ ID NO:80)

>uniprot|Q6RKB1|Q6RKB1_9NOCA ATP/NADPH-dependent carboxylic acid reductase

MAVDSPDERLQRRIAQLFAEDEQVKAARPLEAVSAAVSAPGMRLAQIAAT
VMAGYADRPAAGQRAFELNTDDATGRTSLRLLPRFETITYRELWQRVGEV
AAAWHHDPENPLRAGDFVALLGFTSIDYATLDLADIHLGAVTVPLQASAA
VSQLIAILTETSPRLLASTPEHLDAAVECLLAGTTPERLVVFDYHPEDDD
QRAAFESARRRLADAGSLVIVETLDAVRARGRDLPAAPLFVPDTDDDPLA
LLIYTSGSTGTPKGAMYTNRLAATMWQGNSMLQGNSQRVGINLNYMPMSH
IAGRISLFGVLARGGTAYFAAKSDMSTLFEDIGLVRPTEIFFVPRVCDMV
FQRYQSELDRRSVAGADLDTLDREVKADLRQNYLGGRFLVAVVGSAPLAA
EMKTFMESVLDLPLHDGYGSTEAGASVLLDNQIQRPPVLDYKLVDVPELG

FIG. 8 (Cont.)

YFRTDRPHPRGELLLKAETTIPGYYKRPEVTAEIFDEDGFYKTGDIVAEL
EHDRLVYVDRRNNVLKLSQGEFVTVAHLEAVFASSPLIRQIFIYGSSERS
YLLAVIVPTDDALRGRDTATLKSALAESIQRIAKDANLQPYEIPRDFLIE
TEPFTIANGLLSGIAKLLRPNLKERYGAQLEQMYTDLATGQADELLALRR
EAADLPVLETVSRAAKAMLGVASADMRPDAHFTDLGGDSLSALSFSNLLH
EIFGVEVPVGVVVSPANELRDLANYIEAERNSGAKRPTFTSVHGGGSEIR
AADLTLDKFIDARTLAAADSIPHAPVPAQTVLLTGANGYLGRFLCLEWLE
RLDKTGGTLICVVRGSDAAAARKRLDSAFDSGDPGLLEHYQQLAARTLEV
LAGDIGDPNLGLDDATWQRLAETVDLIVHPAALVNHVLPYTQLFGPNVVG
TAEIVRLAITARRKPVTYLSTVGVADQVDPAEYQEDSDVREMSAVRVVRE
SYANGYGNSKWAGEVLLREAHDLCGLPVAVFRSDMILAHSRYAGQLNVQD
VFTRLILSLVATGIAPYSFYRTDADGNRQRAHYDGLPADFTAAAITALGI
QATEGFRTYDVLNPYDDGISLDEFVDWLVESGHPIQRITDYSDWFHRFET
AIRALPEKQRQASVLPLLDAYRNPCPAVRGAILPAKEFQAAVQTAKIGPE
QDIPHLSAPLIDKYVSDLELLQLL

Q741P9

Nucleotide sequence (SEQ ID NO:81)

>uniprot|Q741P9|Q741P9_MYCPA FadD9

ATGTCGACTGCCACCCATGACGAACGACTCGACCGTCGCGTCCACGAACTC
ATCGCCACCGACCCGCAATTCGCCGCCGCCCAACCCGACCCGGCGATCACC
GCCGCCCTCGAACAGCCCGGGCTGCGGCTGCCGCAGATCATCCGCACCGTG
CTCGACGGCTACGCCGACCGGCCGGCGCTGGGACAGCGCGTGGTGGAGTTC
GTCACGGACGCCAAGACCGGGCGCACGTCGGCGCAGCTGCTCCCCCGCTTC
GAGACCATCACGTACAGCGAAGTAGCGCAGCGTGTTTCGGCGCTGGGCCGC
GCCCTGTCCGACGACGCGGTGCACCCCGGCGACCGGGTGTGCGTGCTGGGC
TTCAACAGCGTCGACTACGCCACCATCGACATGGCGCTGGGCGCCATCGGC
GCCGTCTCGGTGCCGCTGCAGACCAGCGCGGCAATCAGCTCGCTGCAGCCG
ATCGTGGCCGAGACCGAGCCCACCCTGATCGCGTCCAGCGTGAACCAGCTG
TCCGACGCGGTGCAGCTGATCACCGGCGCCGAGCAGGCGCCCACCCGGCTG
GTGGTGTTCGACTACCACCCGCAGGTCGACGACCAGCGCGAGGCCGTCCAG
GACGCCGCGGCGCGGCTGTCCAGCACCGGCGTGGCCGTCCAGACGCTGGCC
GAGCTGCTGGAGCGCGGCAAGGACCTGCCCGCCGTCGCGGAGCCGCCCGCC
GACGAGGACTCGCTGGCCCTGCTGATCTACACCTCCGGGTCCACCGGCGCC
CCCAAGGGCGCGATGTACCCACAGAGCAACGTCGGCAAGATGTGGCGCCGC
GGCAGCAAGAACTGGTTCGGCGAGAGCGCCGCGTCGATCACCCTGAACTTC

FIG. 8 (Cont.)

ATGCCGATGAGCCACGTGATGGGCCGAAGCATCCTCTACGGCACGCTGGGC
AACGGCGGCACCGCCTACTTCGCCGCCCGCAGCGACCTGTCCACCCTGCTT
GAGGACCTCGAGCTGGTGCGGCCCACCGAGCTCAACTTCGTCCCGCGGATC
TGGGAGACGCTGTACGGCGAATTCCAGCGTCAGGTCGAGCGGCGGCTCTCC
GAGGCCGGGGACGCCGGCGAACGTCGCGCCGTCGAGGCCGAGGTGCTGGCC
GAGCAGCGCCAGTACCTGCTGGGCGGGCGGTTCACCTTCGCGATGACGGGC
TCGGCGCCCATCTCGCCGGAGCTGCGCAACTGGGTCGAGTCGCTGCTCGAA
ATGCACCTGATGGACGGCTACGGCTCCACCGAGGCCGGAATGGTGTTGTTC
GACGGGGAGATTCAGCGCCCGCCGGTGATCGACTACAAGCTGGTCGACGTG
CCGGACCTGGGCTACTTCAGCACCGACCGGCCGCATCCGCGCGGCGAGCTG
CTGCTGCGCACCGAGAACATGTTCCCGGGCTACTACAAGCGGGCCGAAACC
ACCGCGGGCGTCTTCGACGAGGACGGCTACTACCGCACCGGCGACGTGTTC
GCCGAGATCGCCCCGGACCGGCTGGTCTACGTCGACCGCCGCAACAACGTG
CTCAAGCTGGCGCAGGGCGAATTCGTCACGCTGGCCAAGCTGGAGGCGGTG
TTCGGCAACAGCCCGCTGATCCGCCAGATCTACGTCTACGGCAACAGCGCC
CAGCCCTACCTGCTGGCGGTCGTGGTGCCCACCGAGGAGGCGCTGGCCTCG
GGTGACCCCGAGACGCTCAAGCCCAAGATCGCCGACTCGCTGCAGCAGGTC
GCCAAGGAGGCCGGCCTGCAGTCCTACGAGGTGCCGCGCGACTTCATCATC
GAGACCACCCCGTTCAGCCTGGAAAACGGTCTGCTGACCGGGATCCGGAAG
CTGGCGTGGCCGAAACTGAAGCAGCACTACGGGGAACGGCTGGAGCAGATG
TACGCCGACCTGGCCGCCGGACAGGCCAACGAGCTGGCCGAGCTGCGCCGC
AACGGTGCCCAGGCGCCGGTGTTGCAGACCGTGAGCCGCGCCGCGGGCGCC
ATGCTGGGTTCGGCCGCCTCCGACCTGTCCCCCGACGCCCACTTCACCGAT
CTGGGCGGAGACTCGTTGTCGGCGTTGACATTCGGCAACCTGCTGCGCGAG
ATCTTCGACGTCGACGTGCCGGTAGGCGTGATCGTCAGCCCGGCCAACGAC
CTGGCGGCCATCGCGAGCTACATCGAGGCCGAGCGGCAGGGCAGCAAGCGC
CCGACGTTCGCCTCGGTGCACGGCCGGGACGCGACCGTGGTGCGCGCCGCC
GACCTGACGCTGGACAAGTTCCTCGACGCCGAGACGCTGGCCGCCGCGCCG
AACCTGCCCAAGCCGGCCACCGAGGTGCGCACCGTGCTGCTGACCGGCGCC
ACCGGCTTCCTGGGCCGCTACCTGGCCCTGGAATGGCTGGAGCGGATGGAC
ATGGTGGACGGCAAGGTCATCGCCCTGGTCCGGGCCCGCTCCGACGAGGAG
GCACGCGCCCGGCTGGACAAGACCTTCGACAGCGGCGACCCGAAACTGCTC
GCGCACTACCAGCAGCTGGCCGCCGATCACCTGGAGGTCATCGCCGGCGAC
AAGGGCGAGGCCAATCTGGGCCTGGGCCAAGACGTTTGGCAACGACTGGCC
GACACGGTCGACGTGATCGTCGACCCCGCCGCGCTGGTCAACCACGTGTTG
CCGTACAGCGAGCTGTTCGGGCCCAACGCCCTGGGCACCGCGGAGCTGATC
CGGCTGGCGCTGACGTCCAAGCAGAAGCCGTACACCTACGTGTCCACCATC
GGCGTGGGCGACCAGATCGAGCCGGGCAAGTTCGTCGAGAACGCCGACATC
CGGCAGATGAGCGCCACCCGGGCGATCAACGACAGCTACGCCAACGGCTAT

FIG. 8 (Cont.)

GGCAACAGCAAGTGGGCCGGCGAGGTGCTGCTGCGCGAGGCGCACGACCTG
TGCGGGCTGCCCGTCGCGGTGTTCCGCTGCGACATGATCCTGGCCGACACC
ACGTATGCCGGGCAGCTCAACCTGCCGGACATGTTCACCCGGCTGATGCTG
AGCCTGGTGGCCACCGGGATCGCGCCCGGCTCGTTCTACGAGCTCGACGCC
GACGGCAACCGGCAGCGGGCGCACTACGACGGCCTGCCGGTCGAGTTCATC
GCCGCGGCGATCTCGACGCTGGGTTCGCAGATCACCGACAGCGACACCGGC
TTCCAGACCTACCACGTGATGAACCCCTACGATGACGGCGTCGGTCTGGAC
GAGTACGTCGATTGGCTGGTGGACGCCGGCTATTCGATCGAGCGGATCGCC
GACTACTCCGAATGGCTGCGGCGGTTCGAGACCTCGCTGCGGGCCCTGCCG
GACCGGCAGCGCCAGTACTCGCTGCTGCCGCTGCTGCACAACTACCGCACG
CCGGAGAAGCCGATCAACGGGTCGATAGCTCCCACCGACGTGTTCCGGGCA
GCGGTGCAGGAGGCGAAAATCGGCCCCGACAAAGACATTCCGCACGTGTCG
CCGCCGGTCATCGTCAAGTACATCACCGACCTGCAGCTGCTCGGGCTGCTC
TGA

Amino acid sequence (SEQ ID NO:82)

>uniprot|Q741P9|Q741P9_MYCPA FadD9

MSTATHDERLDRRVHELIATDPQFAAAQPDPAITAALEQPGLRLPQIIRT
VLDGYADRPALGQRVVEFVTDAKTGRTSAQLLPRFETITYSEVAQRVSAL
GRALSDDAVHPGDRVCVLGFNSVDYATIDMALGAIGAVSVPLQTSAAISS
LQPIVAETEPTLIASSVNQLSDAVQLITGAEQAPTRLVVFDYHPQVDDQR
EAVQDAAARLSSTGVAVQTLAELLERGKDLPAVAEPPADEDSLALLIYTS
GSTGAPKGAMYPQSNVGKMWRRGSKNWFGESAASITLNFMPMSHVMGRSI
LYGTLGNGGTAYFAARSDLSTLLEDLELVRPTELNFVPRIWETLYGEFQR
QVERRLSEAGDAGERRAVEAEVLAEQRQYLLGGRFTFAMTGSAPISPELR
NWVESLLEMHLMDGYGSTEAGMVLFDGEIQRPPVIDYKLVDVPDLGYFST
DRPHPRGELLLRTENMFPGYYKRAETTAGVFDEDGYYRTGDVFAEIAPDR
LVYVDRRNNVLKLAQGEFVTLAKLEAVFGNSPLIRQIYVYGNSAQPYLLA
VVVPTEEALASGDPETLKPKIADSLQQVAKEAGLQSYEVPRDFIIETTPF
SLENGLLTGIRKLAWPKLKQHYGERLEQMYADLAAGQANELAELRRNGAQ
APVLQTVSRAAGAMLGSAASDLSPDAHFTDLGGDSLSALTFGNLLREIFD
VDVPVGVIVSPANDLAAIASYIEAERQGSKRPTFASVHGRDATVVRAADL
TLDKFLDAETLAAAPNLPKPATEVRTVLLTGATGFLGRYLALEWLERMDM
VDGKVIALVRARSDEEARARLDKTFDSGDPKLLAHYQQLAADHLEVIAGD
KGEANLGLGQDVWQRLADTVDVIVDPAALVNHVLPYSELFGPNALGTAEL
IRLALTSKQKPYTYVSTIGVGDQIEPGKFVENADIRQMSATRAINDSYAN
GYGNSKWAGEVLLREAHDLCGLPVAVFRCDMILADTTYAGQLNLPDMFTR

FIG. 8 (Cont.)

LMLSLVATGIAPGSFYELDADGNRQRAHYDGLPVEFIAAAISTLGSQITD
SDTGFQTYHVMNPYDDGVGLDEYVDWLVDAGYSIERIADYSEWLRRFETS
LRALPDRQRQYSLLPLLHNYRTPEKPINGSIAPTDVFRAAVQEAKIGPDK
DIPHVSPPVIVKYITDLQLLGLL

Q7D6X4

Nucleotide sequence (SEQ ID NO:83)

>uniprot|Q7D6X4|Q7D6X4_MYCTU Substrate--CoA ligase, putative

ATGTCGATCAACGATCAGCGACTGACACGCCGCGTCGAGGACCTATACGCC
AGCGACGCCCAGTTCGCCGCCGCCAGTCCCAACGAGGCGATCACCCAGGCG
ATCGACCAGCCCGGGGTCGCGCTTCCACAGCTCATCCGTATGGTCATGGAG
GGCTACGCCGATCGGCCGGCACTCGGCCAGCGTGCGCTCCGCTTCGTCACC
GACCCCGACAGCGGCCGCACCATGGTCGAGCTACTGCCGCGGTTCGAGACC
ATCACCTACCGCGAACTGTGGGCCCGCGCCGGCACATTGGCCACCGCGTTG
AGCGCTGAGCCCGCGATCCGGCCGGGCGACCGGGTTTGCGTGCTGGGCTTC
AACAGCGTCGACTACACAACCATCGACATCGCGCTGATCCGGTTGGGCGCC
GTGTCGGTTCCACTGCAGACCAGTGCGCCGGTCACCGGGTTGCGCCCGATC
GTCACCGAGACCGAGCCGACGATGATCGCCACCAGCATCGACAATCTTGGC
GACGCCGTCGAAGTGCTGGCCGGTCACGCCCCGGCCCGGCTGGTCGTATTC
GATTACCACGGCAAGGTTGACACCCACCGCGAGGCCGTCGAAGCCGCCCGA
GCTCGGTTGGCCGGCTCGGTGACCATCGACACACTTGCCGAACTGATCGAA
CGCGGCAGGGCGCTGCCGGCCACACCCATTGCCGACAGCGCCGACGACGCG
CTGGCGCTGCTGATTTACACCTCGGGTAGTACCGGCGCACCCAAAGGCGCC
ATGTATCGCGAGAGCCAGGTGATGAGCTTCTGGCGCAAGTCGAGTGGCTGG
TTCGAGCCGAGCGGTTACCCCTCGATCACGCTGAACTTCATGCCGATGAGC
CACGTCGGGGGCCGTCAGGTGCTCTACGGGACGCTTTCCAACGGCGGTACC
GCCTACTTCGTCGCCAAGAGCGACCTGTCGACGCTGTTCGAGGACCTCGCC
CTGGTGCGGCCCACAGAATTGTGCTTCGTGCCGCGCATCTGGGACATGGTG
TTCGCAGAGTTCCACAGCGAGGTCGACCGCCGCTTGGTGGACGGCGCCGAT
CGAGCGGCGCTGGAAGCGCAGGTGAAGGCCGAGCTGCGGGAGAACGTGCTC
GGCGGACGGTTTGTCATGGCGCTGACCGGTTCCGCGCCGATCTCCGCTGAG
ATGACGGCGTGGGTCGAGTCCCTGCTGGCCGACGTGCATTTGGTGGAGGGT
TACGGCTCCACCGAGGCCGGGATGGTCCTGAACGACGGCATGGTGCGGCGC
CCCGCGGTGATCGACTACAAGCTGGTCGACGTGCCCGAGCTGGGCTACTTC
GGCACCGATCAGCCCTACCCCCGGGGCGAGCTGCTGGTCAAGACGCAAACC

FIG. 8 (Cont.)

ATGTTCCCCGGCTACTACCAGCGCCCGGATGTCACCGCCGAGGTGTTCGAC
CCCGACGGCTTCTACCGGACCGGGGACATCATGGCCAAAGTAGGCCCCGAC
CAGTTCGTCTACCTCGACCGCCGCAACAACGTGCTAAAGCTCTCCCAGGGC
GAGTTCATCGCCGTGTCGAAGCTCGAGGCGGTGTTCGGCGACAGCCCGCTG
GTCCGACAGATCTTCATCTACGGCAACAGTGCCCGGGCCTACCCGCTGGCG
GTGGTTGTCCCGTCCGGGGACGCGCTTTCTCGCCATGGCATCGAGAATCTC
AAGCCCGTGATCAGCGAGTCCCTGCAGGAGGTAGCGAGGGCGGCCGGCCTG
CAATCCTACGAGATTCCACGCGACTTCATCATCGAAACCACGCCGTTCACC
CTGGAGAACGGCCTGCTCACCGGCATCCGCAAGCTGGCACGCCCGCAGTTG
AAGAAGTTCTATGGCGAACGTCTCGAGCGGCTCTATACCGAGCTGGCCGAT
AGCCAATCCAACGAGCTGCGCGAGCTGCGGCAAAGCGGTCCCGATGCGCCG
GTGCTTCCGACGCTGTGCCGTGCCGCGGCTGCGTTGCTGGGCTCTACCGCT
GCGGATGTGCGGCCGGACGCGCACTTCGCCGACCTGGGTGGTGACTCGCTC
TCGGCGCTGTCGTTGGCCAACCTGCTGCACGAGATCTTCGGCGTCGACGTG
CCGGTGGGTGTCATTGTCAGCCCGGCAAGCGACCTGCGGGCCCTGGCCGAC
CACATCGAAGCAGCGCGCACCGGCGTCAGGCGACCCAGCTTCGCCTCGATA
CACGGTCGCTCCGCGACGGAAGTGCACGCCAGCGACCTCACGCTGGACAAG
TTCATCGACGCTGCCACCTTGGCCGCAGCCCCGAACCTGCCGGCACCGAGC
GCCCAAGTGCGCACCGTACTGCTGACCGGCGCCACCGGCTTTTTGGGTCGC
TACCTGGCGCTGGAATGGCTCGACCGCATGGACCTGGTCAACGGCAAGCTG
ATCTGCCTGGTCCGCGCCAGATCCGACGAGGAAGCACAAGCCCGGCTGGAC
GCGACGTTCGATAGCGGCGACCCGTATTTGGTGCGGCACTACCGCGAATTG
GGCGCCGGCCGCCTCGAGGTGCTCGCCGGCGACAAGGGCGAGGCCGACCTG
GGCCTGGACCGGGTCACCTGGCAGCGGCTAGCCGACACGGTGGACCTGATC
GTGGACCCCGCGGCCCTGGTCAACCACGTGCTGCCGTATAGCCAGCTGTTC
GGCCCAAACGCGGCGGGCACCGCCGAGTTGCTTCGGCTGGCGCTGACCGGC
AAGCGCAAGCCATACATCTACACCTCGACGATCGCCGTGGGCGAGCAGATC
CGCCCGGAGGCGTTCACCGAGGACGCCGACATCCGGGCCATCAGCCCGACC
CGCAGGATCGACGACAGCTACGCCAACGGCTACGCGAACAGCAAGTGGGCC
GGCGAGGTGCTGCTGCGCGAAGCTCACGAGCAGTGCGGCCTGCCGGTGACG
GTCTTCCGCTGCGACATGATCCTGGCCGACACCAGCTATACCGGTCAGCTC
AACCTGCCGGACATGTTCACCCGGCTGATGCTGAGCCTGGCCGCTACCGGC
ATCGCACCCGGTTCGTTCTATGAGCTGGATGCGCACGGCAATCGGCAACGC
GCCCACTATGACGGCTTGCCGGTCGAATTCGTCGCAGAAGCCATTTGCACC
CTTGGGACACATAGCCCGGACCGTTTTGTCACCTACCACGTGATGAACCCC
TACGACGACGGCATCGGGCTGGACGAGTTCGTCGACTGGCTCAACTCCCCA
ACTAGCGGGTCCGGTTGCACGATCCAGCGGATCGCCGACTACGGCGAGTGG
CTGCAGCGGTTCGAGACTTCGCTGCGTGCCTTGCCGGATCGCCAGCGCCAC
GCCTCGCTGCTGCCCTTGCTGCACAACTACCGAGAGCCTGCAAAGCCGATA

FIG. 8 (Cont.)

TGCGGGTCAATCGCGCCCACCGACCAGTTCCGCGCTGCCGTCCAAGAAGCG
AAAATCGGTCCGGACAAAGACATTCCGCACCTCACGGCGGCGATCATCGCG
AAGTACATCAGCAACCTGCGACTGCTCGGGCTGCTGTGA

Amino acid sequence (SEQ ID NO:84)

>uniprot|Q7D6X4|Q7D6X4_MYCTU Substrate--CoA ligase, putative

MSINDQRLTRRVEDLYASDAQFAAASPNEAITQAIDQPGVALPQLIRMVM
EGYADRPALGQRALRFVTDPDSGRTMVELLPRFETITYRELWARAGTLAT
ALSAEPAIRPGDRVCVLGFNSVDYTTIDIALIRLGAVSVPLQTSAPVTGL
RPIVTETEPTMIATSIDNLGDAVEVLAGHAPARLVVFDYHGKVDTHREAV
EAARARLAGSVTIDTLAELIERGRALPATPIADSADDALALLIYTSGSTG
APKGAMYRESQVMSFWRKSSGWFEPSGYPSITLNFMPMSHVGGRQVLYGT
LSNGGTAYFVAKSDLSTLFEDLALVRPTELCFVPRIWDMVFAEFHSEVDR
RLVDGADRAALEAQVKAELRENVLGGRFVMALTGSAPISAEMTAWVESLL
ADVHLVEGYGSTEAGMVLNDGMVRRPAVIDYKLVDVPELGYFGTDQPYPR
GELLVKTQTMFPGYYQRPDVTAEVFDPDGFYRTGDIMAKVGPDQFVYLDR
RNNVLKLSQGEFIAVSKLEAVFGDSPLVRQIFIYGNSARAYPLAVVVPSG
DALSRHGIENLKPVISESLQEVARAAGLQSYEIPRDFIIETTPFTLENGL
LTGIRKLARPQLKKFYGERLERLYTELADSQSNELRELRQSGPDAPVLPT
LCRAAAALLGSTAADVRPDAHFADLGGDSLSALSLANLLHEIFGVDVPVG
VIVSPASDLRALADHIEAARTGVRRPSFASIHGRSATEVHASDLTLDKFI
DAATLAAAPNLPAPSAQVRTVLLTGATGFLGRYLALEWLDRMDLVNGKLI
CLVRARSDEEAQARLDATFDSGDPYLVRHYRELGAGRLEVLAGDKGEADL
GLDRVTWQRLADTVDLIVDPAALVNHVLPYSQLFGPNAAGTAELLRLALT
GKRKPYIYTSTIAVGEQIRPEAFTEDADIRAISPTRRIDDSYANGYANSK
WAGEVLLREAHEQCGLPVTVFRCDMILADTSYTGQLNLPDMFTRLMLSLA
ATGIAPGSFYELDAHGNRQRAHYDGLPVEFVAEAICTLGTHSPDRFVTYH
VMNPYDDGIGLDEFVDWLNSPTSGSGCTIQRIADYGEWLQRFETSLRALP
DRQRHASLLPLLHNYREPAKPICGSIAPTDQFRAAVQEAKIGPDKDIPHL
TAAIIAKYISNLRLLGLL

FIG. 8 (Cont.)

Q7TY99

Nucleotide sequence (SEQ ID NO:85)

>uniprot|Q7TY99|Q7TY99_MYCBO PROBABLE FATTY-ACID-CoA LIGASE FADD9 (FATTY-ACID-COA SYNTHETASE) (FATTY-ACID-COA SYNTHASE)

ATGTCGATCAACGATCAGCGACTGACACGCCGCGTCGAGGACCTATACGCC
AGCGACGCCCAGTTCGCCGCCGCCAGTCCCAACGAGGCGATCACCCAGGCG
ATCGACCAGCCCGGGGTCGCGCTTCCACAGCTCATCCGTATGGTCATGGAG
GGCTACGCCGATCGGCCGGCACTCGGCCAGCGTGCGCTCCGCTTCGTCACC
GACCCCGACAGCGGCCGCACCATGGTCGAGCTACTGCCGCGGTTCGAGACC
ATCACCTACCGCGAACTGTGGGCCCGCGCCGGCACATTGGCCACCGCGTTG
AGCGCTGAGCCCGCGATCCGGCCGGGCGACCGGGTTTGCGTGCTGGGCTTC
AACAGCGTCGACTACACAACCATCGACATCGCGCTGATCCGGTTGGGCGCC
GTGTCGGTTCCACTGCAGACCAGTGCGCCGGTCACCGGGTTGCGCCCGATC
GTCACCGAGACCGAGCCGACGATGATCGCCACCAGCATCGACAATCTTGGC
GACGCCGTCGAAGTGCTGGCCGGTCACGCCCCGGCCCGGCTGGTCGTATTC
GATTACCACGGCAAGGTTGACACCCACCGCGAGGCCGTCGAAGCCGCCCGA
GCTCGGTTGGCCGGCTCGGTGACCATCGACACACTTGCCGAACTGATCGAA
CGCGGCAGGGCGCTGCCGGCCACACCCATTGCCGACAGCGCCGACGACGCG
CTGGCGCTGCTGATTTACACCTCGGGTAGTACCGGCGCACCCAAAGGCGCC
ATGTATCGCGAGAGCCAGGTGATGAGCTTCTGGCGCAAGTCGAGTGGCTGG
TTCGAGCCGAGCGGTTACCCCTCGATCACGCTGAACTTCATGCCGATGAGC
CACGTCGGGGGCCGTCAGGTGCTCTACGGGACGCTTTCCAACGGCGGTACC
GCCTACTACGTCGCCAAGAGCGACCTGTCGACGCTGTTCGAGGACCTCGCC
CTGGTGCGGCCCACAGAATTGTGCTTCGTGCCGCGCATCTGGGACATGGTG
TTCGCAGAGTTCCACAGCGAGGTCGACCGCCGCTTGGTGGACGGCGCCGAT
CGAGCGGCGCTGGAAGCGCAGGTGAAGGCCGAGCTGCGGGAGAACGTGCTC
GGCGGACGGTTTGTCATGGCGCTGACCGGTTCCGCGCCGATCTCCGCTGAG
ATGACGGCGTGGGTCGAGTCCCTGCTGGCCGACGTGCATTTGGTGGAGGGT
TACGGCTCCACCGAGGCCGGGATGGTCCTGAACGACGGCATGGTGCGGCGC
CCCGCGGTGATCGACTACAAGCTGGTCGACGTGCCCGAGCTGGGCTACTTC
GGCACCGATCAGCCCTACCCCCGGGGCGAGCTGCTGGTCAAGACGCAAACC
ATGTTCCCCGGCTACTACCAGCGCCCGGATGTCACCGCCGAGGTGTTCGAC
CCCGACGGCTTCTACCGGACCGGGGACATCATGGCCAAAGTAGGCCCCGAC
CAGTTCGTCTACCTCGACCGCCGCAACAACGTGCTAAAGCTCTCCCAGGGC
GAGTTCATCGCCGTGTCGAAGCTCGAGGCGGTGTTCGGCGACAGCCCGCTG

FIG. 8 (Cont.)

GTCCGACAGATCTTCATCTACGGCAACAGTGCCCGGGCCTACCCGCTGGCG
GTGGTTGTCCCGTCCGGGGACGCGCTTTCTCGCCATGGCATCGAGAATCTC
AAGCCCGTGATCAGCGAGTCCCTGCAGGAGGTAGCGAGGGCGGCCGGCCTG
CAATCCTACGAGATTCCACGCGACTTCATCATCGAAACCACGCCGTTCACC
CTGGAGAACGGCCTACTCACCGGCATCCGCAAGCTGGCACGCCCGCAGTTG
AAGAAGTTCTATGGCGAACGTCTCGAGCGGCTCTATACCGAGCTGGCCGAT
AGCCAATCCAACGAGCTGCGCGAGCTGCGGCAAAGCGGTCCCGATGCGCCG
GTGCTTCCGACGCTGTGCCGTGCCGCGGCTGCGTTGCTGGGCTCTACCGCT
GCGGATGTGCGGCCGGACGCGCACTTCGCCGACCTGGGTGGTGACTCGCTC
TCGGCGCTGTCGTTGGCCAACCTGCTGCACGAGATCTTCGGCGTCGACGTG
CCGGTGGGTGTCATTGTCAGCCCGGCAAGCGACCTGCGGGCCCTGGCCGAC
CACATCGAAGCAGCGCGCACCGGCGTCAGGCGACCCAGCTTCGCCTCGATA
CACGGTCGCTCCGCGACGGAAGTGCACGCCAGCGACCTCACGCTGGACAAG
TTCATCGACGCTGCCACCCTGGCCGCAGCCCCGAACCTGCCGGCACCGAGC
GCCCAAGTGCGCACCGTACTGCTGACCGGCGCCACCGGCTTTTTGGGTCGC
TACCTGGCGCTGGAATGGCTCGACCGCATGGACCTGGTCAACGGCAAGCTG
ATCTGCCTGGTCCGCGCCAGATCCGACGAGGAAGCACAAGCCCGGCTGGAC
GCGACGTTCGATAGCGGCGACCCGTATTTGGTGCGGCACTACCGCGAATTG
GGCGCCGGCCGCCTCGAGGTGCTCGCCGGCGACAAGGGCGAGGCCGACCTG
GGCCTGGACCGGGTCACCTGGCAGCGGCTAGCCGACACGGTGGACCTGATC
GTGGACCCCGCGGCCCTGGTCAACCACGTGCTGCCGTATAGCCAGCTGTTC
GGCCCAAACGCGGCGGGCACCGCCGAGTTGCTTCGGCTGGCGCTGACCGGC
AAGCGCAAGCCATACATCTACACCTCGACGATCGCCGTGGGCGAGCAGATC
CCGCCGGAGGCGTTCACCGAGGACGCCGACATCCGGGCCATCAGCCCGACC
CGCAGGATCGACGACAGCTACGCCAACGGCTACGCGAACAGCAAGTGGGCC
GGCGAGGTGCTGCTGCGCGAAGCTCACGAGCAGTGCGGCCTGCCGGTGACG
GTCTTCCGCTGCGACATGATCCTGGCCGACACCAGCTATACCGGTCAGCTC
AACCTGCCGGACATGTTCACCCGGCTGATGCTGAGCCTGGCCGCTACCGGC
ATCGCACCCGGTTCGTTCTATGAGCTGGATGCGCACGGCAATCGGCAACGC
GCCCACTATGACGGCTTGCCGGTCGAATTCGTCGCAGAAGCCATTTGCACC
CTTGGGACACATAGCCCGGACCGTTTTGTCACCTACCACGTGATGAACCCC
TACGACGACGGCATCGGGCTGGACGAGTTCGTCGACTGGCTCAACTCCCCA
ACTAGCGGGTCCGGTTGCACGATCCAGCGGATCGCCGACTACGGCGAGTGG
CTGCAGCGGTTCGAGACTTCGCTGCGTGCCTTGCCGGATCGCCAGCGCCAC
ACCTCGCTGCTGCCCTTGCTGCACAACTACCGAGAGCCTGCAAAGCCGATA
TGCGGGTCAATCGCGCCCACCGACCAGTTCCGCGCTGCCGTCCAAGAAGCG
AAAATCGGTCCGGACAAAGACATTCCGCACCTCACGGCGGCGATCATCGCG
AAGTACATCAGCAACCTGCGACTGCTCGGGCTGCTGTGA

FIG. 8 (Cont.)

Amino acid sequence (SEQ ID NO:.86)

>uniprot|Q7TY99|Q7TY99_MYCBO PROBABLE FATTY-ACID-CoA LIGASE FADD9 (FATTY-ACID-COA SYNTHETASE) (FATTY-ACID-COA SYNTHASE)

MSINDQRLTRRVEDLYASDAQFAAASPNEAITQAIDQPGVALPQLIRMVM
EGYADRPALGQRALRFVTDPDSGRTMVELLPRFETITYRELWARAGTLAT
ALSAEPAIRPGDRVCVLGFNSVDYTTIDIALIRLGAVSVPLQTSAPVTGL
RPIVTETEPTMIATSIDNLGDAVEVLAGHAPARLVVFDYHGKVDTHREAV
EAARARLAGSVTIDTLAELIERGRALPATPIADSADDALALLIYTSGSTG
APKGAMYRESQVMSFWRKSSGWFEPSGYPSITLNFMPMSHVGGRQVLYGT
LSNGGTAYYVAKSDLSTLFEDLALVRPTELCFVPRIWDMVFAEFHSEVDR
RLVDGADRAALEAQVKAELRENVLGGRFVMALTGSAPISAEMTAWVESLL
ADVHLVEGYGSTEAGMVLNDGMVRRPAVIDYKLVDVPELGYFGTDQPYPR
GELLVKTQTMFPGYYQRPDVTAEVFDPDGFYRTGDIMAKVGPDQFVYLDR
RNNVLKLSQGEFIAVSKLEAVFGDSPLVRQIFIYGNSARAYPLAVVVPSG
DALSRHGIENLKPVISESLQEVARAAGLQSYEIPRDFIIETTPFTLENGL
LTGIRKLARPQLKKFYGERLERLYTELADSQSNELRELRQSGPDAPVLPT
LCRAAAALLGSTAADVRPDAHFADLGGDSLSALSLANLLHEIFGVDVPVG
VIVSPASDLRALADHIEAARTGVRRPSFASIHGRSATEVHASDLTLDKFI
DAATLAAAPNLPAPSAQVRTVLLTGATGFLGRYLALEWLDRMDLVNGKLI
CLVRARSDEEAQARLDATFDSGDPYLVRHYRELGAGRLEVLAGDKGEADL
GLDRVTWQRLADTVDLIVDPAALVNHVLPYSQLFGPNAAGTAELLRLALT
GKRKPYIYTSTIAVGEQIPPEAFTEDADIRAISPTRRIDDSYANGYANSK
WAGEVLLREAHEQCGLPVTVFRCDMILADTSYTGQLNLPDMFTRLMLSLA
ATGIAPGSFYELDAHGNRQRAHYDGLPVEFVAEAICTLGTHSPDRFVTYH
VMNPYDDGIGLDEFVDWLNSPTSGSGCTIQRIADYGEWLQRFETSLRALP
DRQRHTSLLPLLHNYREPAKPICGSIAPTDQFRAAVQEAKIGPDKDIPHL
TAAIIAKYISNLRLLGLL

FIG. 8 (Cont.)

Q9CCT4

Nucleotide sequence (SEQ ID NO:87)

>uniprot|Q9CCT4|Q9CCT4_MYCLE Putative acyl-CoA synthetase

GTGTGGCGCCAACAGTCTATTTCTCATCGCAAGGAATCTGTTATGTCGACT
ATCACTAAGCAGGAAAAGCAGCTCGCACGCCGCGTTGACGACCTCACCGCC
AACGACCCGCAGTTCGCCGCCGCCAAACCCGACCCGGCGGTAGCCGCCGCC
CTTGCCCAGCCCGGGCTTCGACTGCCCCAAATCATCCAGACCGCGCTGGAC
GGTTACGCGGAGCGGCCGGCACTGGGCCAGCGCGTCGCCGAGTTCACCAAA
GACCCTAAGACCGGACGCACCTCGATGGAGCTGCTCCCCAGCTTTGAGACC
ATCACCTACCGCCAGTTGGGCGACCGTGTCGGAGCGCTGGCGCGCGCCTGG
AGGCACGACCTACTGCACGCCGGCTACCGGGTCTGCGTGCTAGGTTTCAAC
AGTGTCGATTACGCCATCATCGACATGGCGCTCGGCGTGATTGGTGCTGTG
GCGGTTCCACTGCAGACCAGTGCGGCGATCACCCAGCTGCAGTCGATCGTG
ACCGAGACCGAACCCAGTATGATCGCGACGAGCGTAAACCAGCTGCCCGAT
ACTGTCGAGCTGATCCTGTCTGGCCAGGCGCCAGCGAAGCTCGTTGTGTTT
GACTACCACCCCGAGGTCGACGAGCAGCATGACGCAGTGGCAACCGCCCGG
GCGCGGTTGGCGGACAGTAGCGTGGTGGTCGAGAGCCTGACCGAGGTCCTC
GGTCGCGGCAAGACGCTGCCAGCTACGCCGATCCCCGTGGCCGATGACTCT
GCTGACCCGTTGGCGTTGCTGATCTACACATCTGGCAGCACCGGCGCACCC
AAGGGCGCGATGTATCTGCAAAGCAATGTCGGCAAGATGTGGCGCCGGTCA
GACGGAAACTGGTTCGGGCCAACCGCCGCGTCAATCACTCTTAACTTCATG
CCGATGAGCCACGTCATGGGCCGCGGAATCCTCTACGGCACGCTCGGTAAC
GGCGGCACGGCTTACTTCGCCGCCCGCAGCGACCTCTCGACGCTGCTGGAG
GATCTCAAGCTGGTGCGGCCGACCGAGTTGAACTTTGTACCGCGCATCTGG
GAAACCCTCTACGATGAATCCAAACGCGCAGTTGACCGTCGGTTAGCCAAC
AGCGGCTCCGCCGACCGTGCAGCCATCAAAGCCGAAGTTATGGATGAACAG
CGGCAATCCCTGCTGGGAGGACGGTACATCGCGGCTATGACGGGCTCGGCG
CCAACCTCCCCGGAGTTGAAACACGGGGTCGAGTCCCTACTCGAAATGCAT
CTGTTGGAAGGCTACGGCTCCACCGAAGCCGGCATGGTCTTGTTTGACGGC
GAAGTGCAACGTCCGCCGGTTATCGATTACAAGCTGGTCGACGTTCCGGAT
TTGGGCTACTTCAGCACCGACCAGCCTTATCCGAGAGGTGAATTGCTGCTC
AAGACCCAGAACATGTTCCCCGGCTACTACAAGCGTCCTGAGGTTACCGCC
ACCGTGTTCGACAGCGACGGTTACTACCAGACCGGAGACATTGTCGCCGAA
GTCGGTCCCGACCGGCTCGTGTACGTCGATCGCCGCAACAACGTGCTGAAA
CTCGCGCAGGGCCAGTTCGTCACCGTCGCGAAACTCGAGGCAGCGTTCAGC

FIG. 8 (Cont.)

AATAGCCCACTGGTCCGGCAGATCTACATCTATGGCAACAGCGCACACCCC
TACCTGTTGGCTGTTGTGGTGCCGACCGAGGATGCGTTGGCTACCAATGAC
ATTGAGGTGCTCAAACCGCTGATTATCGATTCTTTACAGAAAGTAGCGAAA
GAAGCCGACCTGCAGTCCTACGAGGTGCCGCGCGACTTAATCGTCGAGACT
ACACCGTTCAGCCTGGAGAATGGCCTGCTCACCGGTATTCGCAAGCTGGCG
TGGCCGAAGCTCAAGCAGCACTACGGCGCGCGACTCGAACAGCTCTACGCC
GATCTGGTTGAAGGTCAGGCAAATGCACTGCACGTGCTAAAACAAAGCGTG
GCGAACGCTCCGGTACTGCAGACGGTGAGCCGAGCCGTGGGCACCATTCTG
GGAGTGGCGACCACCGATTTGCCGTCGAATGCGCACTTCACCGACTTAGGA
GGAGACTCGTTGTCCGCGCTGACATTCGGTAGCCTGCTACGCGAACTCTTC
GACATCGATGTGCCGGTGGGCGTCATTGTCAGCCCTGTCAACAACTTGGTG
GCGATCGCCGACTACATCGAGCGCGAGCGGCAGGGCACGAAGCGGCCCACT
TTCATTGCCATACACGGTCGTGACGCTGGCAAAGTGCATGCCAGTGACCTC
ACTCTAGACAAATTCATCGATGTATCAACGCTGACTGCCGCGCCCGTATTG
GCGCAACCCGGCACCGAGGTGCGCACCGTCCTGTTGACCGGCGCTACCGGC
TTCCTGGGGCGCTACTTGGCCCTGAAATGGCTCGAACGGATGGACCTGGTC
GAAGGGAAGGTAATCGCTCTGGTAAGAGCCAAGTCCAACGAGGACGCTCGG
GCCCGGCTCGACAAGACCTTCGATAGCGGAGACCCCAAACTGCTGGCGCAC
TACCAGGAACTGGCAACCGACCACCTGGAGGTCATCGCCGGCGACAAAGGC
GAAGTAGATCTGGAATTGGACCGGCAAACGTGGCGACGACTGGCCGACACG
GTCGATCTGATCGTCGACCCCGCCGCCCTGGTCAACCACGTGCTGCCGTAC
AGCGAGCTATTCGGCCCCAATACGTTAGGCACCGCCGAGCTGATTCGGATC
GCGCTGACCAGTAAGCAAAAGCCGTACATCTATGTGTCGACAATCGGCGTC
GGTAATCAGATTGAGCCAGCAAAATTCACCGAAGACTCCGACATCCGAGTC
ATTAGCCCGACGCGCAACATCAACAACAACTATGCCAACGGCTACGGCAAC
AGCAAGTGGGCCGGCGAAGTGCTGCTGCGCGAAGCTCACGACCTATGCGGT
CTGCCGGTCACGGTCTTCCGCTGCGACATGATCTTGGCCGACACCAGCTAT
GCCGGTCAGCTCAACGTCCCCGACATGTTTACTCGAATGATGCTGAGTCTA
GCCGCCACCGGCATCGCACCCGGCTCGTTCTACGAGCTAGACGCCGAGAGC
AATCGGCAACGCGCCCACTACGACGGTCTGCCCGTCGAGTTCATCGCCGAA
GCGATCTCCACCCTGGGAGACCAAAGCCTGCACGATCGAGACGGGTTCACG
ACCTATCATGTAATGAACCCGCACGACGACGGCATCGGTATGGACGAGTTT
GTGGACTGGTTAATTGATGCCGGCTGCCCTATACAACGCATCAACGACTAC
GACGAATGGCTGCGACGGTTTGAGATTTCGCTGCGCGCCCTGCCCGAAAGG
CAGCGTCACAGCTCACTGTTGCCGTTGTTGCACAACTACCAGAAGCCGGAG
AAGCCATTGCACGGGTCGCTGGCACCCACAATCCGGTTCCGTACGGCCGTT
CAAAACGCGAACATTGGTCAGGACAAAGATATTCCGCATATCTCGCCGGCA
ATCATCGCCAAATATGTCAGCGATCTGCAGCTGCTCGGGCTGGTTTGA

FIG. 8 (Cont.)

Amino acid sequence (SEQ ID NO:88)

>uniprot|Q9CCT4|Q9CCT4_MYCLE Putative acyl-CoA synthetase

VWRQQSISHRKESVMSTITKQEKQLARRVDDLTANDPQFAAAKPDPAVAA
ALAQPGLRLPQIIQTALDGYAERPALGQRVAEFTKDPKTGRTSMELLPSF
ETITYRQLGDRVGALARAWRHDLLHAGYRVCVLGFNSVDYAIIDMALGVI
GAVAVPLQTSAAITQLQSIVTETEPSMIATSVNQLPDTVELILSGQAPAK
LVVFDYHPEVDEQHDAVATARARLADSSVVVESLTEVLGRGKTLPATPIP
VADDSADPLALLIYTSGSTGAPKGAMYLQSNVGKMWRRSDGNWFGPTAAS
ITLNFMPMSHVMGRGILYGTLGNGGTAYFAARSDLSTLLEDLKLVRPTEL
NFVPRIWETLYDESKRAVDRRLANSGSADRAAIKAEVMDEQRQSLLGGRY
IAAMTGSAPTSPELKHGVESLLEMHLLEGYGSTEAGMVLFDGEVQRPPVI
DYKLVDVPDLGYFSTDQPYPRGELLLKTQNMFPGYYKRPEVTATVFDSDG
YYQTGDIVAEVGPDRLVYVDRRNNVLKLAQGQFVTVAKLEAAFSNSPLVR
QIYIYGNSAHPYLLAVVVPTEDALATNDIEVLKPLIIDSLQKVAKEADLQ
SYEVPRDLIVETTPFSLENGLLTGIRKLAWPKLKQHYGARLEQLYADLVE
GQANALHVLKQSVANAPVLQTVSRAVGTILGVATTDLPSNAHFTDLGGDS
LSALTFGSLLRELFDIDVPVGVIVSPVNNLVAIADYIERERQGTKRPTFI
AIHGRDAGKVHASDLTLDKFIDVSTLTAAPVLAQPGTEVRTVLLTGATGF
LGRYLALKWLERMDLVEGKVIALVRAKSNEDARARLDKTFDSGDPKLLAH
YQELATDHLEVIAGDKGEVDLELDRQTWRRLADTVDLIVDPAALVNHVLP
YSELFGPNTLGTAELIRIALTSKQKPYIYVSTIGVGNQIEPAKFTEDSDI
RVISPTRNINNNYANGYGNSKWAGEVLLREAHDLCGLPVTVFRCDMILAD
TSYAGQLNVPDMFTRMMLSLAATGIAPGSFYELDAESNRQRAHYDGLPVE
FIAEAISTLGDQSLHDRDGFTTYHVMNPHDDGIGMDEFVDWLIDAGCPIQ
RINDYDEWLRRFEISLRALPERQRHSSLLPLLHNYQKPEKPLHGSLAPTI
RFRTAVQNANIGQDKDIPHISPAIIAKYVSDLQLLGLV

Q54JK0

Nucleotide sequence (SEQ ID NO:89)

>uniprot|Q54JK0|Q54JK0_DICDI Putative uncharacterized protein

ATGTTAAAACATATTAAAAATTTTTTAACTAGAAAAGAAGAAAGAAAAGAA
AAAGAAGTAGAGAAATTAAAAGATGGAGTATCAATAACTGAGGTTAAACAA

FIG. 8 (Cont.)

TCGAACCTAGTAGTTTATTCATGCAATGGTTGTGGATCAGAGATATGGCCA
CCAAAACAAGAGAGATATGCATGTAATGAATGCTCAAATTTCGATTTATGT
AGTGAGTGTTATAGAAAAGAAATGATATTAATAAATGGTACACAAGAAGAG
AAAGATAAATTAGTTAGTGGGGAGAGTAATAATGGAATAAAGTACGAGCCA
GTGAGACATTATGATCCATCACCATTACCTCATCAATTAACATTAGAGAAT
GAGACTCAATTTCAATTAGTTTATAGTTTACGTGGTAATTCAACATTTGAA
ACAATGGAGAAATCATTTAAATACTTTAAAAACAGACCATGTCTTGGTATT
AGAGAGAGATTAGGAGAGGATAATGTTTTATCAGAAAGATATAAATGGTTA
ACATATGGTGAAGTGTATGAGAAATCTTTAACCTTGGCAAAGGCATTAACT
AATTTCATCGAAAGAAGAGATTTCATTTCAATCTATATGGATAATTGCTTA
GAATGGTATTTCACAGATTTTGCATCATTATGGGCCGGATTAATAGTGGTA
CCATTACATCATGCTTCAAATAGTTTTAATCTTTTAGAGATTCTTTGGAAT
TCCGAATCAAAATGTATAGTTTGTTCTGGTGAATCATTTAAAAATTTAATA
GAACTTTATGATCAATTGACTGAGCAAGATAAATTAGAGAAACCAATAGTG
TTGAAATTGATAGTTCATAAGGAGGATCTATTCGATCAGTCATTAGTCGAT
AGATTGCCAAGTGGCGTAGAATTTAAAACTTTCAATGAGATGATTAAAATT
GGGGAATCATTAAGTCAGGCTAAATATGAATTTGTCCCAGTTGGTCCAAAT
GATCTTTCCTCGGTGACTTATACAAGTGGTAGTACTGGTGTACCAAAAGGT
GTAATGAAGTTAGATTCAATTTTCAATTTACTAATTGTCAATTCCTATGTT
CAATTCCCAAATGCAGTTTATAGTTATAATACCCTATCACATAGTCAACGT
TTAAGTGATTGGAGATATATTTATATGGGTGGTAGAGTAGCTATCTATTCA
GGTGATATGAATCTATTATTTGAAGATTTAGCCTTGGTTAGACCTCATTCA
TTTTGGGCTGTACCAAGATTTTGGAATTTATTATTCACCCAGTTCAAGAGT
GATCTAAAGCAATACATGTTTGAAAATCCACAATTGGATGAAAGAACTGCC
ACACTCTATTGCTATAAAGGTATTAGAAAGTTATTAGGTGATAGAATTAAT
AATCTAGTGACTGGTGGTGCTCCGACTGCAAATGAAGTACTCAAATTTATG
AGTGATTGTTGGAAAGATATAAACATTTCAAATTCTTATGGTTTAACTGAA
GTATCAGGTGTTTGTATAGATGGTTATATCTCTGACGAAGTAGAATTCAAA
ATTGAACCAGTACCTTCTTTTGGATATTACCCAACTGATTTACCACATCCT
CGTGGTGAATTGGTTGTAAAATCATCAACAATGTCTGCAGGTTATTATAAA
AATACTCAATTAACTTCAGAATCATTTGAAGGTGGTTGGTTTAAAACTGGT
GATGTCGTAGAATTAATTGGAGTTAGAAAAGTTAAAATCATTGATAGAATT
AAACATGCCTTCAAATTGGCAAATGGAGAATTCGTTACACCAGAACCATTG
GAAAATAATTTCGTTTCACTTTGTATTAATCAAATTTTTATTTATGGTAAT
TCACTTAAAACATTTTTAGTTGCAATTGTTAAACCATCACAAGATTGTTTA
AAACAATTAGGACTTCAAGATATACCAATCGATCAATTAATTGAAAATCCA
ACTTTAAAATCAAAACTTTTATCAGAGATTAATAAAATTTCAAAAGAAAAA
AAACTAGCAAATTATGAAATTCCAAAAATTATTACAATAGATTTCACTGAA
TGGACAATTGATAATAAATTAATCACTGGTTCTGGTAAATTTAATAGAGGT

FIG. 8 (Cont.)

GAATTATATAAATTTTATAAAATTAAAATTAATAATATGTTTGATATAATT
GATAAAATTCAACAAGGTTTAAGAAATAATAATAATAATAATAATGAT
AATATTAATAATAATGATAATAATAATAATAATGAATCAAATAAAGATAAT
TTTGAAAATTATATAAAATCAATTTTAAATTTAGACGGTAGAATTGAAGAT
AATAATTTTAATTTAGAGAATTTATCATTTATTCAAATTGGTGGTGATTCG
TTAGGTGCTGTTAAATTATCATCACTTTTAAAAGAAAAAGAAAATATTGAT
ATTTCACCTTCAACAATTTTAAATCAAAATTTTAATTTATCTTCATTATCG
AAATTAATAAATGAAAAAGAATCAAATCAATCAATTGTTGAAGATTTTAAA
GAAAATTTTAAAATCAATTGGAATGAAGAGATGATTTTAGATGAAGATATT
AAAAAATCAATTGACCAAATTAAAAATGCACAACCATCATCAACTCCCTCT
TCATCAAAATCAACACCATCACAATCATCATCATCACCACCACCATCATTA
AATTCAAATAATATTGGTCAAAATGCTTTTCATATGAAATCAATATTTATT
ACAGGTGTTACAGGTTATTTAGGTACATTTTTATTATTTAATTTATTAGAG
GATAAATCAATTGGTATTGAGAGAATTTATTGTTTAGTTAGAAATGTAAAG
AATGAAGAAGAAGGTTTTAAATTAATTGAAAGAATATTTGAAAAATCTTGT
ATCAATGGTATGAATGAAAAGATTAGAGAAAAGGTAATTCCAGTTTGTGGT
GACTTATCAAAACCATTTTTCGGTGTTTCTACTGAAACCTTCAAAATGTTA
TCTTTAGCGGTCGATATGGTAATTCACAATGGTGCCATTGTTAATATGGCC
TATCCATATGCGAATATGAAATCAACAAATGTTACATCAACTCGTGATATC
CTAAGATTATGCACTACCGGAAGAGCCTCTTTTAAAAAGTTGGTCTACGTT
TCAACAGTTGGTGTATTCTTTGGAAATGGTGATGAAAAGATAGATGAATCA
ACAGCACCATCAACTTTCTTTTTAGATCATGGTAATGGTTATAACCAAACA
AAACTAATATCAGATATACTTGTTAGAGAAGCGGCTTCATATGGTTTACCA
ACAATGATTTTCAGACCAGGTACAATCTTTAGTCATACCCAATCTGGTTTC
AACAATCAAAATGATTCAATCGGTTTAATAATTAAAGGTATCCTAGGTTCG
AGTTCCTATCCAACTAAAAAAGATTACTCTAGTGGCGATTTAAATCTTTCA
CCAGTTGATTGGGTATCATCTTCAATGGTTTCTTTAATTAAGCATCTTCCA
TTTTGGTGTAATAATACAAAAATTTATCATATGGTAAATGATAATCGTTTA
TCTTTAGATTTGCTATGTCAATATATAAATAAAGAAAAACAATTAGAAGAA
ATTAATTATTTCGATTGGATCGATGCTCAACTTAATTCTTCAAATAATCCA
TTGTATTCTATTAAACATTTATTTAAAAAGAATGATCGTTTCCCAATTGGT
TCTCAGTCAATTAAAAATCCAAAAACTATTAAAGATTTAGAATCAATTGGT
GAACTTCAATGTCAACCAATATCTGATTCCACAGTAATCAATTATGTAAAA
TATTTAATCTCAAATAATTTAATTCAAACAATTAATAAATAA

FIG. 8 (Cont.)

Amino acid sequence (SEQ ID NO:90)

>uniprot|Q54JK0|Q54JK0_DICDI Putative uncharacterized protein

MLKHIKNFLTRKEERKEKEVEKLKDGVSITEVKQSNLVVYSCNGCGSEIWP
PKQERYACNECSNFDLCSECYRKEMILINGTQEEKDKLVSGESNNGIKYEP
VRHYDPSPLPHQLTLENETQFQLVYSLRGNSTFETMEKSFKYFKNRPCLGI
RERLGEDNVLSERYKWLTYGEVYEKSLTLAKALTNFIERRDFISIYMDNCL
EWYFTDFASLWAGLIVVPLHHASNSFNLLEILWNSESKCIVCSGESFKNLI
ELYDQLTEQDKLEKPIVLKLIVHKEDLFDQSLVDRLPSGVEFKTFNEMIKI
GESLSQAKYEFVPVGPNDLSSVTYTSGSTGVPKGVMKLDSIFNLLIVNSYV
QFPNAVYSYNTLSHSQRLSDWRYIYMGGRVAIYSGDMNLLFEDLALVRPHS
FWAVPRFWNLLFTQFKSDLKQYMFENPQLDERTATLYCYKGIRKLLGDRIN
NLVTGGAPTANEVLKFMSDCWKDINISNSYGLTEVSGVCIDGYISDEVEFK
IEPVPSFGYYPTDLPHPRGELVVKSSTMSAGYYKNTQLTSESFEGGWFKTG
DVVELIGVRKVKIIDRIKHAFKLANGEFVTPEPLENNFVSLCINQIFIYGN
SLKTFLVAIVKPSQDCLKQLGLQDIPIDQLIENPTLKSKLLSEINKISKEK
KLANYEIPKIITIDFTEWTIDNKLITGSGKFNRGELYKFYKIKINNMFDII
DKIQQGLRNNNNNNNNDNINNNDNNNNNESNKDNFENYIKSILNLDGRIED
NNFNLENLSFIQIGGDSLGAVKLSSLLKEKENIDISPSTILNQNFNLSSLS
KLINEKESNQSIVEDFKENFKINWNEEMILDEDIKKSIDQIKNAQPSSTPS
SSKSTPSQSSSSPPPSLNSNNIGQNAFHMKSIFITGVTGYLGTFLLFNLLE
DKSIGIERIYCLVRNVKNEEEGFKLIERIFEKSCINGMNEKIREKVIPVCG
DLSKPFFGVSTETFKMLSLAVDMVIHNGAIVNMAYPYANMKSTNVTSTRDI
LRLCTTGRASFKKLVYVSTVGVFFGNGDEKIDESTAPSTFFLDHGNGYNQT
KLISDILVREAASYGLPTMIFRPGTIFSHTQSGFNNQNDSIGLIIKGILGS
SSYPTKKDYSSGDLNLSPVDWVSSSMVSLIKHLPFWCNNTKIYHMVNDNRL
SLDLLCQYINKEKQLEEINYFDWIDAQLNSSNNPLYSIKHLFKKNDRFPIG
SQSIKNPKTIKDLESIGELQCQPISDSTVINYVKYLI
SNNLIQTINK

Q2MFQ3

Nucleotide sequence (SEQ ID NO:91)

>uniprot|Q2MFQ3|Q2MFQ3_STRRY Putative non-ribosomal peptide synthetase

FIG. 8 (Cont.)

ATGACCGACACGTACGTTTCTTCACGCCCGTTATCCAAGCGGCCCCAGGTC
CCGGGTGCCCGCACGCCGGCGCCCGGCTACCCACGGGACAGCCGCATCCCG
GAGCTGTTCGAGGCACAGGCCGCGGCGCTGCCGCAGGCCCCGGCGGCCCGG
CACGGGGACCGCACCCTGACCTACGGCCAACTCGACGCCCACGCCGACGCG
TTGGCGGACCGGCTGGCTGCCGGCGGGGTCCGGCCGGGTGACCTGATCGGC
GTGTGCGGCAGCCGTTCCCTGGAGGCGCTGGTGGCGCTGCTGGGCATCCTC
AAGGCCGGCTGCGCGTACGTACCGCTCGACGAGGAACTGCCGCCGGCCCGG
CTGCGGGCCATGGCCGAGGACGCGGGCATCAGCGCCGCGGTCACCCTGCCG
GGCAGCACGCGCCGGGTGCGGGGTCTACGCGTGTCAGTCGAGGTCGGCTCC
CTCGGCCGGCCCGCCCCCGAGCGCGCGAGCGGCCCCGCCCCCGACCGGGCC
ACCGGCTCCGCCGCCGACTGCGCCTACGTCGCCTTCACTTCCGGCACGACC
GGCCGGCCCAAGCCCGTAGCGCTGTCCCACCGGGGCGTGGTCCGCCTCGTG
CTGTCCGACCCCGGCCTCACGCCACCCGGACCGGGCGACGGAGTGCTGCAC
GCCTACAGCCTGTCCTCCGACGCCTCGACCATCGAGATCTGGGGCGCGCTG
CTGACCGGCGCCTGCCTGGTCGTCGCCGACCGCGAGGAACTGCTCTCGCCC
ACCGCCCTGGAACGGCTGCTCCGCGCGGGCGGCGTCACCGTGGCGTACCTG
ACGACGAGCGTCTTCCACCTCGTCGCCCGGACCCGCCCCGAGGCGCTGGCC
GGCCTGCGGTTCGTCTCCGCGGGCGGGGAGGCGATGGACCCGCGCCTGGCG
AACGCCGTCCTCGCGGCCTGCCCCCGCACCACGGTGGTCAACTTCTACGGC
CCGACCGAGAACGCCGTGGTCTCCACCGCCCATGTGCTCACCCCCCTCCCC
GAGGACGCCGCACACGTCCCCCTGGGACGCCCCTTCGGCGCTTCCACCTGC
CACGTCCTGCGGGCCGACGGCTCGCCCGCGCGGCCGGGCGAGGAAGGGGAG
CTGTACGTCGGCGGGGACGGGCTGGCGCTGGGCTACCTCGGCGACCCGCAG
CTGACCGCCGAGCGGTTCGTGACGCTGCCCGCGGTCGAGCCGGACGGACCG
CTGTACCGGACCGGCGACCGGGCCGTACGGCACGCCGACGGGCTGCTGGAG
TACCGCGGACGGCTCGACCGCCAGGTCAAGCTGCGCGGCGCCCGCATCGAG
CTGGACGAGGTGGAGACCCGCTTACGGGCCCACCCCGAGGTCGGCGAAGCG
GCCGTCGAGGTCGACGGGCACTCCCTGACCGCCTACGTCACGGCCACCGTC
CCCGGCCGCCCGCTGCCGCTGGCCGACCTGCGCGCGTACTGCGCCAAGTGG
CTGCCCCCGCAGGCCGTCCCGGCCCTGATACCCCTGGACCGCTTCCCGGTC
ACCAGCGGCGGCAAGATCGACCGCAGCCGTCTGAAGCCGACCGCGCCACCG
CCCGGCCCCGAAGACACCGCGGAGGCCGCGCGGCGCCCGGACGAGCCGGAG
GCCACCGACGGCCTGTCCGGTCTCCTCTCGGAAGTGTGGCACCAGGTGCTG
CGTGTCCGGCCCACGCCCCGGGACGACTTCTTCCTCCTCGGCGGCGACTCC
CTGCTCGCCTCGGAGACCGTCACCCGCACCCTCGCCGTACTCGGCCTCGAC
GCGGCCCTGGGCTCCACCCTCATCAGGGCGCTGCTGGCCGCGCCCACCCTC
GAAAGCTTCACCGCCGCCGTACGCGGAGTCCGCGGCGGCACCGGCGGACCG
GCCGGCGGCCAGGAACCGGCCGTCGACTTCGCCGCGGAGACCGGACTCGGC
TTCGCCCTCCCGCCCGCCGAAGGCCCGGCGCCGAACCCGCACGACCCCGAG

FIG. 8 (Cont.)

GACGTCCTGCTCACGGGCGCTTCCGGCTTCGTCGGCGGATTCCTGCTGCAC
CGTCTGCTGCACGCGACGGCCGCCCGCGTCCACTGCCCCGTACGGGCGACG
AGCCCCGCCCACGCCCGGCAGCGGGTCCGCACCGCCCTCACCCGCTACGGG
CTGCACCTCGACGAGGCCGACTGGCAGCGCGTGGAGTGCTTCCCCGGGGAC
CTGACCCAGCCGCGCCTGGGGCTCGACCACGAGCGCGCCGACGCACTGGCC
CAGCGCCTGGACCTGATCGTGCACAACGGTGCCCGGGTCAACTTCCTCTAC
CCCTACCAGCAGTTGCGCCCGGCGAACGTCGACGGAACCCGCGAGGTCGTC
CGGATCGCCGCGCGCCGCCGGGTGCCGGTGCACTTCGTGTCCACCGTCGCA
GTCGTCGCGGGCTTCGGCACCGCCGGGGTGCGCGAGGTGGACGAAGATCTG
CCGCCGGCCCACGCCGACGGGCTGACCATGGGGTACGCGGAGAGCAAGTGG
GTCGCCGAAGGGGTGCTGCGGCAGGCGGCCGCGCAGGGCCTGCCGGTGGCC
GTGTACCGGCCGTACGAGGTCACGGGCGACCGGACGCACGGCGCGTGCAAC
ACCGAGACGGCCATCTGCTCGCTGTTCAAGATGATCGCCGACACGGGAGTG
GCCCCCGACATCAAGCTGCCGATGGACTTCGTACCCGTCGACCACCTCGCC
GAGTCCCTGGTGCACATCGCCACGCACCGGCCGGCCGACGGCCGGGTCTAC
CACCTGACCAACCCGCGCCCGGCGATGCTGTCGGACGTCCTCGACCGGATG
CGCGCGGCGGGCTTCACCCTGCGCACCCTGCCGTACGACGCGTGGGTCGGC
GAGCTCGTCCGGCACGTCGCCGAGAACCCGACGAGCGCCACGGCTCCGTTC
GTGTCCCTGTGCGTGGACCGCAGCCGCACCGCCGACATGTCCGTCAAGGAG
ATGTACCTCAAGGGCACCTTCCCGGTCCTGGGGCGGCGCAACGCCGAGGAG
GCGCTGGCCGGCAGCGGGCTGCACTGCCCGCCGGTCGACTCCGCTCTGCTG
GACCGCTACCTGGAGTACTTCTTCACCTCCGGCTACCTCACGCGCCCGGCG
GCCGGCCCCGGGCCCGAATCCGAGGCCGAGCGAATACCGGAGGACGAGCCG
GTGTCCGGGACCGAACCGATATCCGGGACCGAACCGATTTCTGCCGCCGAG
CCGATATCCGGGACCGAACCGATTTCTGCCGCCGGGCCGATATCCGGGACC
GAGCCGATACCCGCCGCCGAGCCGATATCCGGGACCGCAGCCGCAGCCCGC
ACGGAGCGCAGCCGATGA

Amino acid sequence (SEQ ID NO:92)

>uniprot|Q2MFQ3|Q2MFQ3_STRRY Putative non-ribosomal peptide synthetase

MTDTYVSSRPLSKRPQVPGARTPAPGYPRDSRIPELFEAQAAALPQAPAAR
HGDRTLTYGQLDAHADALADRLAAGGVRPGDLIGVCGSRSLEALVALLGIL
KAGCAYVPLDEELPPARLRAMAEDAGISAAVTLPGSTRRVRGLRVSVEVGS
LGRPAPERASGPAPDRATGSAADCAYVAFTSGTTGRPKPVALSHRGVVRLV
LSDPGLTPPGPGDGVLHAYSLSSDASTIEIWGALLTGACLVVADREELLSP
TALERLLRAGGVTVAYLTTSVFHLVARTRPEALAGLRFVSAGGEAMDPRLA

FIG. 8 (Cont.)

NAVLAACPRTTVVNFYGPTENAVVSTAHVLTPLPEDAAHVPLGRPFGASTC
HVLRADGSPARPGEEGELYVGGDGLALGYLGDPQLTAERFVTLPAVEPDGP
LYRTGDRAVRHADGLLEYRGRLDRQVKLRGARIELDEVETRLRAHPEVGEA
AVEVDGHSLTAYVTATVPGRPLPLADLRAYCAKWLPPQAVPALIPLDRFPV
TSGGKIDRSRLKPTAPPPGPEDTAEAARRPDEPEATDGLSGLLSEVWHQVL
RVRPTPRDDFFLLGGDSLLASETVTRTLAVLGLDAALGSTLIRALLAAPTL
ESFTAAVRGVRGGTGGPAGGQEPAVDFAAETGLGFALPPAEGPAPNPHDPE
DVLLTGASGFVGGFLLHRLLHATAARVHCPVRATSPAHARQRVRTALTRYG
LHLDEADWQRVECFPGDLTQPRLGLDHERADALAQRLDLIVHNGARVNFLY
PYQQLRPANVDGTREVVRIAARRRVPVHFVSTVAVVAGFGTAGVREVDEDL
PPAHADGLTMGYAESKWVAEGVLRQAAAQGLPVAVYRPYEVTGDRTHGACN
TETAICSLFKMIADTGVAPDIKLPMDFVPVDHLAESLVHIATHRPADGRVY
HLTNPRPAMLSDVLDRMRAAGFTLRTLPYDAWVGELVRHVAENPTSATAPF
VSLCVDRSRTADMSVKEMYLKGTFPVLGRRNAEEALAGSGLHCPPVDSALL
DRYLEYFFTSGYLTRPAAGPGPESEAERIPEDEPVSGTEPISGTEPISAAE
PISGTEPISAAGPISGTEPIPAAEPISGTAAAARTERSR

ZP_03429464

Nucleotide sequence (SEQ ID NO:263)

>gi|189214978:54695-58201 Mycobacterium tuberculosis EAS054 NZ_ABOV01000087, whole genome shotgun sequence

ATGTCGATCAACGATCAGCGACTGACACGCCGCGTCGAGGACCTATACGCC
AGCGACGCCCAGTTCGCCGCCGCCAGTCCCAACGAGGCGATCACCCAGGCG
ATCGACCAGCCCGGGGTCGCGCTTCCACAGCTCATCCGTATGGTCATGGAG
GGCTACGCCGATCGGCCGGCACTCGGCCAGCGTGCGCTCCGCTTCGTCACC
GACCCCGACAGCGGCCGCACCATGGTCGAGCTACTGCCGCGGTTCGAGACC
ATCACCTACCGCGAACTGTGGGCCCGCGCCGGCACATTGGCCACCGCGTTG
AGCGCTGAGCCCGCGATCCGGCCGGGCGACCGGGTTTGCGTGCTGGGCTTC
AACAGCGTCGACTACACAACCATCGACATCGCGCTGATCCGGTTGGGCGCC
GTGTCGGTTCCACTGCAGACCAGTGCGCCGGTCACCGGGTTGCGCCCGATC
GTCACCGAGACCGAGCCGACGATGATCGCCACCAGCATCGACAATCTTGGC
GACGCCGTCGAAGTGCTGGCCGGTCACGCCCCGGCCCGGCTGGTCGTATTC
GATTACCACGGCAAGGTTGACACCCACCGCGAGGCCGTCGAAGCCGCCCGA
GCTCGGTTGGCCGGCTCGGTGACCATCGACACACTTGCCGAACTGATCGAA
CGCGGCAGGGCGCTGCCGGCCACACCCATTGCCGACAGCGCCGACGACGCG

FIG. 8 (Cont.)

CTGGCGCTGCTGACTTACACCTCGGGTAGTACCGGCGCACCCAAAGGCGCC
ATGTATCGCGAGAGCCAGGTGATGAGCTTCTGGCGCAAGTCGAGTGGCTGG
TTCGAGCCGAGCGGTTACCCCTCGATCACGCTGAACTTCATGCCGATGAGC
CACGTCGGGGGCCGTCAGGTGCTCTACGGGACGCTTTCCAACGGCGGTACC
GCCTACTTCGTCGCCAAGAGCGACCTGTCGACGCTGTTCGAGGACCTCGCC
CTGGTGCGGCCCACAGAATTGTGCTTCGTGCCGCGCATCTGGGACATGGTG
TTCGCAGAGTTCCACAGCGAGGTCGACCGCCGCTTGGTGGACGGCGCCGAT
CGAGCGGCGCTGGAAGCGCAGGTGAAGGCCGAGCTGCGGGAGAACGTGCTC
GGCGGACGGTTTGTCATGGCGCTGACCGGTTCCGCGCCGATCTCCGCTGAG
ATGACGGCGTGGGTCGAGTCCCTGCTGGCCGACGTGCATTTGGTGGAGGGT
TACGGCTCCACCGAGGCCGGGATGGTCCTGAACGACGGCATGGTGCGGCGC
CCCGCGGTGATCGACTACAAGCTGGTCGACGTGCCCGAGCTGGGCTACTTC
GGCACCGATCAGCCCTACCCCCGGGGCGAGCTGCTGGTCAAGACGCAAACC
ATGTTCCCCGGCTACTACCAGCGCCCGGATGTCACCGCCGAGGTGTTCGAC
CCCGACGGCTTCTACCGGACCGGGGACATCATGGCCAAAGTAGGCCCCGAC
CAGTTCGTCTACCTCGACCGCCGCAACAACGTGCTAAAGCTCTCCCAGGGC
GAGTTCATCGCCGTGTCGAAGCTCGAGGCGGTGTTCGGCGACAGCCCGCTG
GTCCGACAGATCTTCATCTACGGCAACAGTGCCCGGGCCTACCCGCTGGCG
GTGGTTGTCCCGTCCGGGGACGCGCTTTCTCGCCATGGCATCGAGAATCTC
AAGCCCGTGATCAGCGAGTCCCTGCAGGAGGTAGCGAGGGCGGCCGGCCTG
CAATCCTACGAGATTCCACGCGACTTCATCATCGAAACCACGCCGTTCACC
CTGGAGAACGGCCTGCTCACCGGCATCCGCAAGCTGGCACGCCCGCAGTTG
AAGAAGTTCTATGGCGAACGTCTCGAGCGGCTCTATACCGAGCTGGCCGAT
AGCCAATCCAACGAGCTGCGCGAGCTGCGGCAAAGCGGTCCCGATGCGCCG
GTGCTTCCGACGCTGTGCCGTGCCGCGGCTGCGTTGCTGGGCTCTACCGCT
GCGGATGTGCGGCCGGACGCGCACTTCGCCGACCTGGGTGGTGACTCGCTC
TCGGCGCTGTCGTTGGCCAACCTGCTGCACGAGATCTTCGGCGTCGACGTG
CCGGTGGGTGTCATTGTCAGCCCGGCAAGCGACCTGCGGGCCCTGGCCGAC
CACATCGAAGCAGCGCGCACCGGCGTCAGGCGACCCAGCTTCGCCTCGATA
CACGGTCGCTCCGCGACGGAAGTGCACGCCAGCGACCTCACGCTGGACAAG
TTCATCGACGCTGCCACCCTGGCCGCAGCCCCGAACCTGCCGGCACCGAGC
GCCCAAGTGCGCACCGTACTGCTGACCGGCGCCACCGGCTTTTTGGGTCGC
TACCTGGCGCTGGAATGGCTCGACCGCATGGACCTGGTCAACGGCAAGCTG
ATCTGCCTGGTCCGCGCCAGATCCGACGAGGAAGCACAAGCCCGGCTGGAC
GCGACGTTCGATAGCGGCGACCCGTATTTGGTGCGGCACTACCGCGAATTG
GGCGCCGGCCGCCTCGAGGTGCTCGCCGGCGACAAGGGCGAGGCCGACCTG
GGCCTGGACCGGGTCACCTGGCAGCGGCTAGCCGACACGGTGGACCTGATC
GTGGACCCCGCGGCCCTGGTCAACCACGTGCTGCCGTATAGCCAGCTGTTC
GGCCCAAACGCGGCGGGCACCGCCGAGTTGCTTCGGCTGGCGCTGACCGGC

FIG. 8 (Cont.)

AAGCGCAAGCCATACATCTACACCTCGACGATCGCCGTGGGCGAGCAGATC
CCGCCGGAGGCGTTCACCGAGGACGCCGACATCCGGGCCATCAGCCCGACC
CGCAGGATCGACGACAGCTACGCCAACGGCTACGCGAACAGCAAGTGGGCC
GGCGAGGTGCTGCTGCGCGAAGCTCACGAGCAGTGCGGCCTGCCGGTGACG
GTCTTCCGCTGCGACATGATCCTGGCCGACACCAGCTATACCGGTCAGCTC
AACCTGCCGGACATGTTCACCCGGCTGATGCTGAGCCTGGCCGCTACCGGC
ATCGCACCCGGTTCGTTCTATGAGCTGGATGCGCACGGCAATCGGCAACGC
GCCCACTATGACGGCTTGCCGGTCGAATTCGTCGCAGAAGCCATTTGCACC
CTTGGGACACATAGCCCGGACCGTTTTGTCACCTACCACGTGATGAACCCC
TACGACGACGGCATCGGGCTGGACGAGTTCGTCGACTGGCTCAACTCCCCA
ACTAGCGGGTCCGGTTGCACGATCCAGCGGATCGCCGACTACGGCGAGTGG
CTGCAGCGGTTCGAGACTTCGCTGCGTGCCTTGCCGGATCGCCAGCGCCAC
GCCTCGCTGCTGCCCTTGCTGCACAACTACCGAGAGCCTGCAAAGCCGATA
TGCGGGTCAATCGCGCCCACCGACCAGTTCCGCGCTGCCGTCCAAGAAGCG
AAAATCGGTCCGGACAAAGACATTCCGCACCTCACGGCGGCGATCATCGCG
AAGTACATCAGCAACCTGCGACTGCTCGGGCTGCTGTGA

Amino acid sequence (SEQ ID NO:264)

>gi|215431545|ref|ZP_03429464.1| fatty-acid-CoA ligase fadD9 [Mycobacterium tuberculosis EAS054]

MSINDQRLTRRVEDLYASDAQFAAASPNEAITQAIDQPGVALPQLIRMVME
GYADRPALGQRALRFVTDPDSGRTMVELLPRFETITYRELWARAGTLATAL
SAEPAIRPGDRVCVLGFNSVDYTTIDIALIRLGAVSVPLQTSAPVTGLRPI
VTETEPTMIATSIDNLGDAVEVLAGHAPARLVVFDYHGKVDTHREAVEAAR
ARLAGSVTIDTLAELIERGRALPATPIADSADDALALLTYTSGSTGAPKGA
MYRESQVMSFWRKSSGWFEPSGYPSITLNFMPMSHVGGRQVLYGTLSNGGT
AYFVAKSDLSTLFEDLALVRPTELCFVPRIWDMVFAEFHSEVDRRLVDGAD
RAALEAQVKAELRENVLGGRFVMALTGSAPISAEMTAWVESLLADVHLVEG
YGSTEAGMVLNDGMVRRPAVIDYKLVDVPELGYFGTDQPYPRGELLVKTQT
MFPGYYQRPDVTAEVFDPDGFYRTGDIMAKVGPDQFVYLDRRNNVLKLSQG
EFIAVSKLEAVFGDSPLVRQIFIYGNSARAYPLAVVVPSGDALSRHGIENL
KPVISESLQEVARAAGLQSYEIPRDFIIETTPFTLENGLLTGIRKLARPQL
KKFYGERLERLYTELADSQSNELRELRQSGPDAPVLPTLCRAAAALLGSTA
ADVRPDAHFADLGGDSLSALSLANLLHEIFGVDVPVGVIVSPASDLRALAD
HIEAARTGVRRPSFASIHGRSATEVHASDLTLDKFIDAATLAAAPNLPAPS
AQVRTVLLTGATGFLGRYLALEWLDRMDLVNGKLICLVRARSDEEAQARLD
ATFDSGDPYLVRHYRELGAGRLEVLAGDKGEADLGLDRVTWQRLADTVDLI

FIG. 8 (Cont.)

VDPAALVNHVLPYSQLFGPNAAGTAELLRLALTGKRKPYIYTSTIAVGEQI
PPEAFTEDADIRAISPTRRIDDSYANGYANSKWAGEVLLREAHEQCGLPVT
VFRCDMILADTSYTGQLNLPDMFTRLMLSLAATGIAPGSFYELDAHGNRQR
AHYDGLPVEFVAEAICTLGTHSPDRFVTYHVMNPYDDGIGLDEFVDWLNSP
TSGSGCTIQRIADYGEWLQRFETSLRALPDRQRHASLLPLLHNYREPAK
PICGSIAPTDQFRAAVQEAKIGPDKDIPHLTAAIIAKYISNLRLLGLL

ZP_03533123

Nucleotide sequence (SEQ ID NO:265)

>gi|192384451:15291-18572 Mycobacterium tuberculosis GM 1503 NZ_ABQG01000169, whole genome shotgun sequence

ATGGTCGAGCTACTGCCGCGGTTCGAGACCATCACCTACCGCGAACTGTGG
GCCCGCGCCGGCACATTGGCCACCGCGTTGAGCGCTGAGCCCGCGATCCGG
CCGGGCGACCGGGTTTGCGTGCTGGGCTTCAACAGCGTCGACTACACAACC
ATCGACATCGCGCTGATCCGGTTGGGCGCCGTGTCGGTTCCACTGCAGACC
AGTGCGCCGGTCACCGGGTTGCGCCCGATCGTCACCGAGACCGAGCCGACG
ATGATCGCCACCAGCATCGACAATCTTGGCGACGCCGTCGAAGTGCTGGCC
GGTCACGCCCCGGCCCGGCTGGTCGTATTCGATTACCACGGCAAGGTTGAC
ACCCACCGCGAGGCCGTCGAAGCCGCCCGAGCTCGGTTGGCCGGCTCGGTG
ACCATCGACACACTTGCCGAACTGATCGAACGCGGCAGGGCGCTGCCGGCC
ACACCCATTGCCGACAGCGCCGACGACGCGCTGGCGCTGCTGATTTACACC
TCGGGTAGTACCGGCGCACCCAAAGGCGCCATGTATCGCGAGAGCCAGGTG
ATGAGCTTCTGGCGCAAGTCGAGTGGCTGGTTCGAGCCGAGCGGTTACCCC
TCGATCACGCTGAACTTCATGCCGATGAGCCACGTCGGGGGCCGTCAGGTG
CTCTACGGGACGCTTTCCAACGGCGGTACCGCCTACTTCGTCGCCAAGAGC
GACCTGTCGACGCTGTTCGAGGACCTCGCCCTGGTGCGGCCCACAGAATTG
TGCTTCGTGCCGCGCATCTGGGACATGGTGTTCGCAGAGTTCCACAGCGAG
GTCGACCGCCGCTTGGTGGACGGCGCCGATCGAGCGGCGCTGGAAGCGCAG
GTGAAGGCCGAGCTGCGGGAGAACGTGCTCGGCGGACGGTTTGTCATGGCG
CTGACCGGTTCCGCGCCGATCTCCGCTGAGATGACGGCGTGGGTCGAGTCC
CTGCTGGCCGACGTGCATTTGGTGGAGGGTTACGGCTCCACCGAGGCCGGG
ATGGTCCTGAACGACGGCATGGTGCGGCGCCCCGCGGTGATCGACTACAAG
CTGGTCGACGTGCCCGAGCTGGGCTACTTCGGCACCGATCAGCCCTACCCC
CGGGGCGAGCTGCTGGTCAAGACGCAAACCATGTTCCCCGGCTACTACCAG
CGCCCGGATGTCACCGCCGAGGTGTTCGACCCCGACGGCTTCTACCGGACC

FIG. 8 (Cont.)

GGGGACATCATGGCCAAAGTAGGCCCCGACCAGTTCGTCTACCTCGACCGC
CGCAACAACGTGCTAAAGCTCTCCCAGGGCGAGTTCATCGCCGTGTCGAAG
CTCGAGGCGGTGTTCGGCGACAGCCCGCTGGTCCGACAGATCTTCATCTAC
GGCAACAGTGCCCGGGCCTACCCGCTGGCGGTGGTTGTCCCGTCCGGGGAC
GCGCTTTCTCGCCATGGCATCGAGAATCTCAAGCCCGTGATCAGCGAGTCC
CTGCAGGAGGTAGCGAGGGCGGCCGGCCTGCAATCCTACGAGATTCCACGC
GACTTCATCATCGAAACCACGCCGTTCACCCTGGAGAACGGCCTGCTCACC
GGCATCCGCAAGCTGGCACGCCCGCAGTTGAAGAAGTTCTATGGCGAACGT
CTCGAGCGGCTCTATACCGAGCTGGCCGATAGCCAATCCAACGAGCTGCGC
GAGCTGCGGCAAAGCGGTCCCGATGCGCCGGTGCTTCCGACGCTGTGCCGT
GCCGCGGCTGCGTTGCTGGGCTCTACCGCTGCGGATGTGCGGCCGGACGCG
CACTTCGCCGACCTGGGTGGTGACTCGCTCTCGGCGCTGTCGTTGGCCAAC
CTGCTGCACGAGATCTTCGGCGTCGACGTGCCGGTGGGTGTCATTGTCAGC
CCGGCAAGCGACCTGCGGGCCCTGGCCGACCACATCGAAGCAGCGCGCACC
GGCGTCAGGCGACCCAGCTTCGCCTCGATACACGGTCGCTCCGCGACGGAA
GTGCACGCCAGCGACCTCACGCTGGACAAGTTCATCGACGCTGCCACCCTG
GCCGCAGCCCCGAACCTGCCGGCACCGAGCGCCCAAGTGCGCACCGTACTG
CTGACCGGCGCCACCGGCTTTTTGGGTCGCTACCTGGCGCTGGAATGGCTC
GACCGCATGGACCTGGTCAACGGCAAGCTGATCTGCCTGGTCCGCGCCAGA
TCCGACGAGGAAGCACAAGCCCGGCTGGACGCGACGTTCGATAGCGGCGAC
CCGTATTTGGTGCGGCACTACCGCGAATTGGGCGCCGGCCGCCTCGAGGTG
CTCGCCGGCGACAAGGGCGAGGCCGACCTGGGCCTGGACCGGGTCACCTGG
CAGCGGCTAGCCGACACGGTGGACCTGATCGTGGACCCCGCGGCCCTGGTC
AACCACGTGCTGCCGTATAGCCAGCTGTTCGGCCCAAACGCGGCGGGCACC
GCCGAGTTGCTTCGGCTGGCGCTGACCGGCAAGCGCAAGCCATACATCTAC
ACCTCGACGATCGCCGTGGGCGAGCAGATCCCGCCGGAGGCGTTCACCGAG
GACGCCGACATCCGGGCCATCAGCCCGACCCGCAGGATCGACGACAGCTAC
GCCAACGGCTACGCGAACAGCAAGTGGGCCGGCGAGGTGCTGCTGCGCGAA
GCTCACGAGCAGTGCGGCCTGCCGGTGACGGTCTTCCGCTGCGACATGATC
CTGGCCGACACCAGCTATACCGGTCAGCTCAACCTGCCGGACATGTTCACC
CGGCTGATGCTGAGCCTGGCCGCTACCGGCATCGCACCCGGTTCGTTCTAT
GAGCTGGATGCGCACGGCAATCGGCAACGCGCCCACTATGACGGCTTGCCG
GTCGAATTCGTCGCAGAAGCCATTTGCACCCTTGGGACACATAGCCCGGAC
CGTTTTGTCACCTACCACGTGATGAACCCCTACGACGACGGCATCGGGCTG
GACGAGTTCGTCGACTGGCTCAACTCCCCAACTAGCGGGTCCGGTTGCACG
ATCCAGCGGATCGCCGACTACGGCGAGTGGCTGCAGCGGTTCGAGACTTCG
CTGCGTGCCTTGCCGGATCGCCAGCGCCACGCCTCGCTGCTGCCCTTGCTG
CACAACTACCGAGAGCCTGCAAAGCCGATATGCGGGTCAATCGCGCCCACC
GACCAGTTCCGCGCTGCCGTCCAAGAAGCGAAAATCGGTCCGGACAAAGAC

FIG. 8 (Cont.)

ATTCCGCACCTCACGGCGGCGATCATCGCGAAGTACATCAGCAACCTGCGA
CTGCTCGGGCTGCTGTGA

Amino acid sequence (SEQ ID NO:266)

>gi|218754327|ref|ZP_03533123.1| fatty-acid-CoA ligase fadD9 [Mycobacterium tuberculosis GM 1503]

MVELLPRFETITYRELWARAGTLATALSAEPAIRPGDRVCVLGFNSVDYTT
IDIALIRLGAVSVPLQTSAPVTGLRPIVTETEPTMIATSIDNLGDAVEVLA
GHAPARLVVFDYHGKVDTHREAVEAARARLAGSVTIDTLAELIERGRALPA
TPIADSADDALALLIYTSGSTGAPKGAMYRESQVMSFWRKSSGWFEPSGYP
SITLNFMPMSHVGGRQVLYGTLSNGGTAYFVAKSDLSTLFEDLALVRPTEL
CFVPRIWDMVFAEFHSEVDRRLVDGADRAALEAQVKAELRENVLGGRFVMA
LTGSAPISAEMTAWVESLLADVHLVEGYGSTEAGMVLNDGMVRRPAVIDYK
LVDVPELGYFGTDQPYPRGELLVKTQTMFPGYYQRPDVTAEVFDPDGFYRT
GDIMAKVGPDQFVYLDRRNNVLKLSQGEFIAVSKLEAVFGDSPLVRQIFIY
GNSARAYPLAVVVPSGDALSRHGIENLKPVISESLQEVARAAGLQSYEIPR
DFIIETTPFTLENGLLTGIRKLARPQLKKFYGERLERLYTELADSQSNELR
ELRQSGPDAPVLPTLCRAAAALLGSTAADVRPDAHFADLGGDSLSALSLAN
LLHEIFGVDVPVGVIVSPASDLRALADHIEAARTGVRRPSFASIHGRSATE
VHASDLTLDKFIDAATLAAAPNLPAPSAQVRTVLLTGATGFLGRYLALEWL
DRMDLVNGKLICLVRARSDEEAQARLDATFDSGDPYLVRHYRELGAGRLEV
LAGDKGEADLGLDRVTWQRLADTVDLIVDPAALVNHVLPYSQLFGPNAAGT
AELLRLALTGKRKPYIYTSTIAVGEQIPPEAFTEDADIRAISPTRRIDDSY
ANGYANSKWAGEVLLREAHEQCGLPVTVFRCDMILADTSYTGQLNLPDMFT
RLMLSLAATGIAPGSFYELDAHGNRQRAHYDGLPVEFVAEAICTLGTHSPD
RFVTYHVMNPYDDGIGLDEFVDWLNSPTSGSGCTIQRIADYGEWLQRFETS
LRALPDRQRHASLLPLLHNYREPAKPICGSIAPTDQFRAAVQEAKIGPDKD
IPHLTAAIIAKYISNLRLLGLL

ZP_03433592

Nucleotide sequence (SEQ ID NO:267)

>gi|189214744:1-3126 Mycobacterium tuberculosis T85 NZ_ABOW01000178, whole genome shotgun sequence

FIG. 8 (Cont.)

GACATCGCGCTGATCCGGTTGGGCGCCGTGTCGGTTCCACTGCAGACCAGT
GCGCCGGTCACCGGGTTGCGCCCGATCGTCACCGAGACCGAGCCGACGATG
ATCGCCACCAGCATCGACAATCTTGGCGACGCCGTCGAAGTGCTGGCCGGT
CACGCCCCGGCCCGGCTGGTCGTATTCGATTACCACGGCAAGGTTGACACC
CACCGCGAGGCCGTCGAAGCCGCCCGAGCTCGGTTGGCCGGCTCGGTGACC
ATCGACACACTTGCCGAACTGATCGAACGCGGCAGGGCGCTGCCGGCCACA
CCCATTGCCGACAGCGCCGACGACGCGCTGGCGCTGCTGATTTACACCTCG
GGTAGTACCGGCGCACCCAAAGGCGCCATGTATCGCGAGAGCCAGGTGATG
AGCTTCTGGCGCAAGTCGAGTGGCTGGTTCGAGCCGAGCGGTTACCCCTCG
ATCACGCTGAACTTCATGCCGATGAGCCACGTCGGGGGCCGTCAGGTGCTC
TACGGGACGCTTTCCAACGGCGGTACCGCCTACTTCGTCGCCAAGAGCGAC
CTGTCGACGCTGTTCGAGGACCTCGCCCTGGTGCGGCCCACAGAATTGTGC
TTCGTGCCGCGCATCTGGGACATGGTGTTCGCAGAGTTCCACAGCGAGGTC
GACCGCCGCTTGGTGGACGGCGCCGATCGAGCGGCGCTGGAAGCGCAGGTG
AAGGCCGAGCTGCGGGAGAACGTGCTCGGCGGACGGTTTGTCATGGCGCTG
ACCGGTTCCGCGCCGATCTCCGCTGAGATGACGGCGTGGGTCGAGTCCCTG
CTGGCCGACGTGCATTTGGTGGAGGGTTACGGCTCCACCGAGGCCGGGATG
GTCCTGAACGACGGCATGGTGCGGCGCCCCGCGGTGATCGACTACAAGCTG
GTCGACGTGCCCGAGCTGGGCTACTTCGGCACCGATCAGCCCTACCCCCGG
GGCGAGCTGCTGGTCAAGACGCAAACCATGTTCCCCGGCTACTACCAGCGC
CCGGATGTCACCGCCGAGGTGTTCGACCCCGACGGCTTCTACCGGACCGGG
GACATCATGGCCAAAGTAGGCCCCGACCAGTTCGTCTACCTCGACCGCCGC
AACAACGTGCTAAAGCTCTCCCAGGGCGAGTTCATCGCCGTGTCGAAGCTC
GAGGCGGTGTTCGGCGACAGCCCGCTGGTCCGACAGATCTTCATCTACGGC
AACAGTGCCCGGGCCTACCCGCTGGCGGTGGTTGTCCCGTCCGGGGACGCG
CTTTCTCGCCATGGCATCGAGAATCTCAAGCCCGTGATCAGCGAGTCCCTG
CAGGAGGTAGCGAGGGCGGCCGGCCTGCAATCCTACGAGATTCCACGCGAC
TTCATCATCGAAACCACGCCGTTCACCCTGGAGAACGGCCTGCTCACCGGC
ATCCGCAAGCTGGCACGCCCGCAGTTGAAGAAGTTCTATGGCGAACGTCTC
GAGCGGCTCTATACCGAGCTGGCCGATAGCCAATCCAACGAGCTGCGCGAG
CTGCGGCAAAGCGGTCCCGATGCGCCGGTGCTTCCGACGCTGTGCCGTGCC
GCGGCTGCGTTGCTGGGCTCTACCGCTGCGGATGTGCGGCCGGACGCGCAC
TTCGCCGACCTGGGTGGTGACTCGCTCTCGGCGCTGTCGTTGGCCAACCTG
CTGCACGAGATCTTCGGCGTCGACGTGCCGGTGGGTGTCATTGTCAGCCCG
GCAAGCGACCTGCGGGCCCTGGCCGACCACATCGAAGCAGCGCGCACCGGC
GTCAGGCGACCCAGCTTCGCCTCGATACACGGTCGCTCCGCGACGGAAGTG
CACGCCAGCGACCTCACGCTGGACAAGTTCATCGACGCTGCCACCCTGGCC
GCAGCCCCGAACCTGCCGGCACCGAGCGCCCAAGTGCGCACCGTACTGCTG
ACCGGCGCCACCGGCTTTTTGGGTCGCTACCTGGCGCTGGAATGGCTCGAC

FIG. 8 (Cont.)

CGCATGGACCTGGTCAACGGCAAGCTGATCTGCCTGGTCCGCGCCAGATCC
GACGAGGAAGCACAAGCCCGGCTGGACGCGACGTTCGATAGCGGCGACCCG
TATTTGGTGCGGCACTACCGCGAATTGGGCGCCGGCCGCCTCGAGGTGCTC
GCCGGCGACAAGGGCGAGGCCGACCTGGGCCTGGACCGGGTCACCTGGCAG
CGGCTAGCCGACACGGTGGACCTGATCGTGGACCCCGCGGCCCTGGTCAAC
CACGTGCTGCCGTATAGCCAGCTGTTCGGCCCAAACGCGGCGGGCACCGCC
GAGTTGCTTCGGCTGGCGCTGACCGGCAAGCGCAAGCCATACATCTACACC
TCGACGATCGCCGTGGGCGAGCAGATCCCGCCGGAGGCGTTCACCGAGGAC
GCCGACATCCGGGCCATCAGCCCGACCCGCAGGATCGACGACAGCTACGCC
AACGGCTACGCGAACAGCAAGTGGGCCGGCGAGGTGCTGCTGCGCGAAGCT
CACGAGCAGTGCGGCCTGCCGGTGACGGTCTTCCGCTGCGACATGATCCTG
GCCGACACCAGCTATACCGGTCAGCTCAACCTGCCGGACATGTTCACCCGG
CTGATGCTGAGCCTGGCCGCTACCGGCATCGCACCCGGTTCGTTCTATGAG
CTGGATGCGCACGGCAATCGGCAACGCGCCCACTATGACGGCTTGCCGGTC
GAATTCGTCGCAGAAGCCATTTGCACCCTTGGGACACATAGCCCGGACCGT
TTTGTCACCTACCACGTGATGAACCCCTACGACGACGGCATCGGGCTGGAC
GAGTTCGTCGACTGGCTCAACTCCCCAACTAGCGGGTCCGGTTGCACGATC
CAGCGGATCGCCGACTACGGCGAGTGGCTGCAGCGGTTCGAGACTTCGCTG
CGTGCCTTGCCGGATCGCCAGCGCCACGCCTCGCTGCTGCCCTTGCTGCAC
AACTACCGAGAGCCTGCAAAGCCGATATGCGGGTCAATCGCGCCCACCGAC
CAGTTCCGCGCTGCCGTCCAAGAAGCGAAAATCGGTCCGGACAAAGACATT
CCGCACCTCACGGCGGCGATCATCGCGAAGTACATCAGCAACCTGCGACTG
CTCGGGCTGCTGTGA

Amino acid sequence (SEQ ID NO:268)

>gi|215446840|ref|ZP_03433592.1| fatty-acid-CoA ligase fadD9 [Mycobacterium tuberculosis T85]

DIALIRLGAVSVPLQTSAPVTGLRPIVTETEPTMIATSIDNLGDAVEVLAG
HAPARLVVFDYHGKVDTHREAVEAARARLAGSVTIDTLAELIERGRALPAT
PIADSADDALALLIYTSGSTGAPKGAMYRESQVMSFWRKSSGWFEPSGYPS
ITLNFMPMSHVGGRQVLYGTLSNGGTAYFVAKSDLSTLFEDLALVRPTELC
FVPRIWDMVFAEFHSEVDRRLVDGADRAALEAQVKAELRENVLGGRFVMAL
TGSAPISAEMTAWVESLLADVHLVEGYGSTEAGMVLNDGMVRRPAVIDYKL
VDVPELGYFGTDQPYPRGELLVKTQTMFPGYYQRPDVTAEVFDPDGFYRTG
DIMAKVGPDQFVYLDRRNNVLKLSQGEFIAVSKLEAVFGDSPLVRQIFIYG
NSARAYPLAVVVPSGDALSRHGIENLKPVISESLQEVARAAGLQSYEIPRD
FIIETTPFTLENGLLTGIRKLARPQLKKFYGERLERLYTELADSQSNELRE
LRQSGPDAPVLPTLCRAAAALLGSTAADVRPDAHFADLGGDSLSALSLANL

FIG. 8 (Cont.)

LHEIFGVDVPVGVIVSPASDLRALADHIEAARTGVRRPSFASIHGRSATEV
HASDLTLDKFIDAATLAAAPNLPAPSAQVRTVLLTGATGFLGRYLALEWLD
RMDLVNGKLICLVRARSDEEAQARLDATFDSGDPYLVRHYRELGAGRLEVL
AGDKGEADLGLDRVTWQRLADTVDLIVDPAALVNHVLPYSQLFGPNAAGTA
ELLRLALTGKRKPYIYTSTIAVGEQIPPEAFTEDADIRAISPTRRIDDSYA
NGYANSKWAGEVLLREAHEQCGLPVTVFRCDMILADTSYTGQLNLPDMFTR
LMLSLAATGIAPGSFYELDAHGNRQRAHYDGLPVEFVAEAICTLGTHSPDR
FVTYHVMNPYDDGIGLDEFVDWLNSPTSGSGCTIQRIADYGEWLQRFETSL
RALPDRQRHASLLPLLHNYREPAKPICGSIAPTDQFRAAVQEAKIGPDKDI
PHLTAAIIAKYISNLRLLGLL

ZP_03537669

Nucleotide sequence (SEQ ID NO:269)

>gi|192384126:19547-22546 Mycobacterium tuberculosis T17 NZ_ABQH01000288, whole genome shotgun sequence

ATGTCGATCAACGATCAGCGACTGACACGCCGCGTCGAGGACCTATACGCC
AGCGACGCCCAGTTCGCCGCCGCCAGTCCCAACGAGGCGATCACCCAGGCG
ATCGACCAGCCCGGGGTCGCGCTTCCACAGCTCATCCGTATGGTCATGGAG
GGCTACGCCGATCGGCCGGCACTCGGCCAGCGTGCGCTCCGCTTCGTCACC
GACCCCGACAGCGGCCGCACCATGGTCGAGCTACTGCCGCGGTTCGAGACC
ATCACCTACCGCGAACTGTGGGCCCGCGCCGGCACATTGGCCACCGCGTTG
AGCGCTGAGCCCGCGATCCGGCCGGGCGACCGGGTTTGCGTGCTGGGCTTC
AACAGCGTCGACTACACAACCATCGACATCGCGCTGATCCGGTTGGGCGCC
GTGTCGGTTCCACTGCAGACCAGTGCGCCGGTCACCGGGTTGCGCCCGATC
GTCACCGAGACCGAGCCGACGATGATCGCCACCAGCATCGACAATCTTGGC
GACGCCGTCGAAGTGCTGGCCGGTCACGCCCCGGCCCGGCTGGTCGTATTC
GATTACCACGGCAAGGTTGACACCCACCGCGAGGCCGTCGAAGCCGCCCGA
GCTCGGTTGGCCGGCTCGGTGACCATCGACACACTTGCCGAACTGATCGAA
CGCGGCAGGGCGCTGCCGGCCACACCCATTGCCGACAGCGCCGACGACGCG
CTGGCGCTGCTGATTTACACCTCGGGTAGTACCGGCGCACCCAAAGGCGCC
ATGTATCGCGAGAGCCAGGTGATGAGCTTCTGGCGCAAGTCGAGTGGCTGG
TTCGAGCCGAGCGGTTACCCCTCGATCACGCTGAACTTCATGCCGATGAGC
CACGTCGGGGGCCGTCAGGTGCTCTACGGGACGCTTTCCAACGGCGGTACC
GCCTACTTCGTCGCCAAGAGCGACCTGTCGACGCTGTTCGAGGACCTCGCC
CTGGTGCGGCCCACAGAATTGTGCTTCGTGCCGCGCATCTGGGACATGGTG
TTCGCAGAGTTCCACAGCGAGGTCGACCGCCGCTTGGTGGACGGCGCCGAT

FIG. 8 (Cont.)

CGAGCGGCGCTGGAAGCGCAGGTGAAGGCCGAGCTGCGGGAGAACGTGCTC
GGCGGACGGTTTGTCATGGCGCTGACCGGTTCCGCGCCGATCTCCGCTGAG
ATGACGGCGTGGGTCGAGTCCCTGCTGGCCGACGTGCATTTGGTGGAGGGT
TACGGCTCCACCGAGGCCGGGATGGTCCTGAACGACGGCATGGTGCGGCGC
CCCGCGGTGATCGACTACAAGCTGGTCGACGTGCCCGAGCTGGGCTACTTC
GGCACCGATCAGCCCTACCCCCGGGGCGAGCTGCTGGTCAAGACGCAAACC
ATGTTCCCCGGCTACTACCAGCGCCCGGATGTCACCGCCGAGGTGTTCGAC
CCCGACGGCTTCTACCGGACCGGGGACATCATGGCCAAAGTAGGCCCCGAC
CAGTTCGTCTACCTCGACCGCCGCAACAACGTGCTAAAGCTCTCCCAGGGC
GAGTTCATCGCCGTGTCGAAGCTCGAGGCGGTGTTCGGCGACAGCCCGCTG
GTCCGACAGATCTTCATCTACGGCAACAGTGCCCGGGCCTACCCGCTGGCG
GTGGTTGTCCCGTCCGGGGACGCGCTTTCTCGCCATGGCATCGAGAATCTC
AAGCCCGTGATCAGCGAGTCCCTGCAGGAGGTAGCGAGGGCGGCCGGCCTG
CAATCCTACGAGATTCCACGCGACTTCATCATCGAAACCACGCCGTTCACC
CTGGAGAACGGCCTGCTCACCGGCATCCGCAAGCTGGCACGCCCGCAGTTG
AAGAAGTTCTATGGCGAACGTCTCGAGCGGCTCTATACCGAGCTGGCCGAT
AGCCAATCCAACGAGCTGCGCGAGCTGCGGCAAAGCGGTCCCGATGCGCCG
GTGCTTCCGACGCTGTGCCGTGCCGCGGCTGCGTTGCTGGGCTCTACCGCT
GCGGATGTGCGGCCGGACGCGCACTTCGCCGACCTGGGTGGTGACTCGCTC
TCGGCGCTGTCGTTGGCCAACCTGCTGCACGAGATCTTCGGCGTCGACGTG
CCGGTGGGTGTCATTGTCAGCCCGGCAAGCGACCTGCGGGCCCTGGCCGAC
CACATCGAAGCAGCGCGCACCGGCGTCAGGCGACCCAGCTTCGCCTCGATA
CACGGTCGCTCCGCGACGGAAGTGCACGCCAGCGACCTCACGCTGGACAAG
TTCATCGACGCTGCCACCCTGGCCGCAGCCCCGAACCTGCCGGCACCGAGC
GCCCAAGTGCGCACCGTACTGCTGACCGGCGCCACCGGCTTTTTGGGTCGC
TACCTGGCGCTGGAATGGCTCGACCGCATGGACCTGGTCAACGGCAAGCTG
ATCTGCCTGGTCCGCGCCAGATCCGACGAGGAAGCACAAGCCCGGCTGGAC
GCGACGTTCGATAGCGGCGACCCGTATTTGGTGCGGCACTACCGCGAATTG
GGCGCCGGCCGCCTCGAGGTGCTCGCCGGCGACAAGGGCGAGGCCGACCTG
GGCCTGGACCGGGTCACCTGGCAGCGGCTAGCCGACACGGTGGACCTGATC
GTGGACCCCGCGGCCCTGGTCAACCACGTGCTGCCGTATAGCCAGCTGTTC
GGCCCAAACGCGGCGGGCACCGCCGAGTTGCTTCGGCTGGCGCTGACCGGC
AAGCGCAAGCCATACATCTACACCTCGACGATCGCCGTGGGCGAGCAGATC
CCGCCGGAGGCGTTCACCGAGGACGCCGACATCCGGGCCATCAGCCCGACC
CGCAGGATCGACGACAGCTACGCCAACGGCTACGCGAACAGCAAGTGGGCC
GGCGAGGTGCTGCTGCGCGAAGCTCACGAGCAGTGCGGCCTGCCGGTGACG
GTCTTCCGCTGCGACATGATCCTGGGCCGACACCAGCTATACCGGTCAGCT
CAACCTGCCGGACATGTCACCCGGGCTGATGCTGAGCCTGGC

FIG. 8 (Cont.)

Amino acid sequence (SEQ ID NO:270)

>gi|219558593|ref|ZP_03537669.1| fatty-acid-CoA ligase fadD9 [Mycobacterium tuberculosis T17]

MSINDQRLTRRVEDLYASDAQFAAASPNEAITQAIDQPGVALPQLIRMVME
GYADRPALGQRALRFVTDPDSGRTMVELLPRFETITYRELWARAGTLATAL
SAEPAIRPGDRVCVLGFNSVDYTTIDIALIRLGAVSVPLQTSAPVTGLRPI
VTETEPTMIATSIDNLGDAVEVLAGHAPARLVVFDYHGKVDTHREAVEAAR
ARLAGSVTIDTLAELIERGRALPATPIADSADDALALLIYTSGSTGAPKGA
MYRESQVMSFWRKSSGWFEPSGYPSITLNFMPMSHVGGRQVLYGTLSNGGT
AYFVAKSDLSTLFEDLALVRPTELCFVPRIWDMVFAEFHSEVDRRLVDGAD
RAALEAQVKAELRENVLGGRFVMALTGSAPISAEMTAWVESLLADVHLVEG
YGSTEAGMVLNDGMVRRPAVIDYKLVDVPELGYFGTDQPYPRGELLVKTQT
MFPGYYQRPDVTAEVFDPDGFYRTGDIMAKVGPDQFVYLDRRNNVLKLSQG
EFIAVSKLEAVFGDSPLVRQIFIYGNSARAYPLAVVVPSGDALSRHGIENL
KPVISESLQEVARAAGLQSYEIPRDFIIETTPFTLENGLLTGIRKLARPQL
KKFYGERLERLYTELADSQSNELRELRQSGPDAPVLPTLCRAAAAALLGSTA
ADVRPDAHFADLGGDSLSALSLANLLHEIFGVDVPVGVIVSPASDLRALAD
HIEAARTGVRRPSFASIHGRSATEVHASDLTLDKFIDAATLAAAPNLPAPS
AQVRTVLLTGATGFLGRYLALEWLDRMDLVNGKLICLVRARSDEEAQARLD
ATFDSGDPYLVRHYRELGAGRLEVLAGDKGEADLGLDRVTWQRLADTVDLI
VDPAALVNHVLPYSQLFGPNAAGTAELLRLALTGKRKPYIYTSTIAVGEQI
PPEAFTEDADIRAISPTRRIDDSYANGYANSKWAGEVLLREAHEQCGLPVT
VFRCDMILGRHQLYRSAQPAGHVTRADAEPG

ZP_05224908

Nucleotide sequence (SEQ ID NO:271)

>gi|163719654:2489-6013 Mycobacterium intracellulare ATCC 13950 NZ_ABIN01000072, whole genome shotgun sequence

ATGTCGACTGCCATTCATGACGAACACCTCGACCGTCGCATCGAGGAACTC
ATCGCCAACGACCCCCAATTCGCCGCCGCCCGACCGGACCCGGCCATCACC
GCCGCCACCGAAGCGCCCGGGCTGCGGCTGCCGCAGATCATCCGGACCGTG
CTCGACGGCTACGCCGACCGGCCTGCCCTGGCGCAGCGCGTCGTGGAGTTC
GTCACCGACGCCAAGACCGGGCGGACGACGGCCGAGCTGCTCCCCCGTTTC

FIG. 8 (Cont.)

GAGACCATCACGTATGGCGAACTCGGCGAACGGGTTTCGGCCCTCGGCCGT
GCCTGGGCCGGCGACGCGGTGCGCCCCGGCGACCGCGTCTGCGTGCTCGGC
TTCAACAGCGTTGACTACGCCACCATCGACATCGCGCTGGGCACCATCGGG
GCCGTGTCGGTGCCGCTGCAGACCAGCGCGGCGATCTCCTCGTTGCAGCCG
ATCGTCGCCGAGACCGAGCCCAGCCTGATCGCCTCGAGCGTCAACCAGCTG
CCCGACGCGGTGGAGCTGATCCTGGCCGGCGACCACGTGCCCGGCAAGCTC
GTCGTGTTCGACTACCAGCCCCAGGTCGACGACCAGCGCGAGGCCGTGGAG
GCCGCCGCCGCGCGGTTGGCCGACTCCGGCGTCGCGGTCGAGGCTCTCGCC
GACGTGCTGCGGCGCGGCAAGGACCTGCCGGCCGTCGAGCCGCCGGCGAGC
GACGAGGACTCGCTGGCCCTGCTGATCTACACCTCCGGCAGCACCGGCGCG
CCCAAGGGCGCGATGTACCCGCAGAGCAACGTCGGCAAGATGTGGCGGCGC
GGGAGCAAGAACTGGTTCGGGGAAAGCGCCGCGTCGATCACCCTCAACTTC
ATGCCGATGAGCCACGTCATGGGGCGCGGAATCCTCTACGGCACGCTGGGC
AACGGCGGCACCGCGTACTTCGCCGCCCGCAGCGACCTGTCCACCCTGCTC
GAGGACCTCGAGTTGGTGCGGCCCACCGAGATGAACTTCGTCCCCCGCATC
TGGGAGACGCTGTACGGCGAATTCCAGCGCCAGGTCGAGCGGCGGCTGGCC
GACGGCGATGCGGGCCCGGAGGCCCGCGAGACTGTGGCGGCTGCGGTGTTG
GAAGAACAGCGCCAGTACCTGCTGGGCGGGCGGTTCATCTTCGCGATGACG
GGCTCGGCACCCACCTCGCCGGAGCTCAAGGCGTGGGCCGAGTCGCTCCTG
CAGATGCACCTGATGGACGGCTACGGCTCCACCGAGGCCGGAATGGTGTTG
TTCGACGGGGAGATTCAGCGTCCGCCGGTTATTGATTACAAGCTGGTCGAC
GTTCCGGATCTGGGCTATTTCAGCACCGACCGTCCGCATCCGCGCGGTGAG
TTGTTGCTGCGGACCGAGAACATGTTCCCGGGTTATTACAAGCGGGCCGAG
ACCACCGCGAACGTGTTCGACGAGGACGGTTATTACCGCACCGGTGACGTG
TTCGCCGAGATCGCGCCGGACCGGCTGGTGTATGTCGATCGCCGCAACAAC
GTGCTCAAGTTGGCCCAGGGCGAGTTCGTGACGCTGGCCAAGCTGGAGGCG
GTGTTCGGCAACAGCCCGCTGATCCGCCAGATCTACGTTTACGGCAACAGC
TCCCAGCCCTACCTGCTGGCCGTGGTGGTGCCGACCGAGGAAGCGTTGGCG
GACAACGATCTTGAGTCGCTCAAGCCGAAGATCGCCGACTCGCTGCAGAAG
GTCGCCAAGGAGACCGGCCTGCAGTCCTACGAGGTGCCGCGCGACTTCATC
ATCGAGACCACGCCGTTCACCCTGGAAAACGGCCTGCTGACCGGGATCCGC
AAGCTGGCGTGGCCCAAGCTCAAGGCGCACTACGGGGATCGGCTCGAGCAG
ATGTATGCCGAGCTGGCCGCGGGACAGGCCAACGAGTTGGCCGAACTGCGC
CGCAGCGGCGCGGCGGCGCCGGTGGCCCAGACCGTGAGCCGGGCCGCGGCC
GCCCTGCTGGGTGCGACGGCCGGGGATCTGTCCGCAGATGCCCACTTCACC
GATCTTGGTGGAGACTCGTTGTCGGCGTTGACCTTCGGCAACCTGCTGCGC
GAGATCTTCGATGTCGACGTGCCGGTGGGGGTGATCGTCAGCCCGGCCAAC
GACCTGGCGGGGATCGCCGCCTACATCGAGGCCGAGCGGCAGGGCTCCAAG
CGCCCGACGTTCGCCGCCGTGCACGGTCGCGGTGCGACCATGGTGCACGCC

FIG. 8 (Cont.)

AGTGACCTCACGCTGGACAAGTTCCTCGACGAGGCGACCCTGGCCGCCGCG
CCCAGCCTGCCCAAGCCGGCCACCGAGGTGCGCACCGTGCTGTTGACCGGC
GCGACCGGCTTTTTGGGCCGCTACCTGGCGCTGGACTGGCTCGAGCGGATG
GACATGGTCGACGGCAAGGTCATCGCCCTGGTGCGGGCCCGCACCGATGAG
GAGGCGCGCGCCCGGCTGGACAAGACCTTCGACAGCGGCGACCCCAAACTG
CTGGCGCACTACCAGCGGCTGGCCGCCGACCACCTCGAGGTCATCGCCGGC
GACAAGGGTGAGGCCAACCTCGGCCTGGACCCCCAGACCTGGCAGCGACTG
GCCGAGGAGGTCGACGTCATCGTCGACCCCGCCGCGCTGGTCAACCACGTG
CTGCCCTACAGCGAGCTGTTCGGCCCCAACGCCCTGGGCACCGCGGAGCTG
ATCCGGATCGCGCTGACCTCCAGGCAAAAGCCCTACACCTACGTGTCGACG
ATCGGGGTGGGCGATCAGATCCAGCCAGGTGAGTTCGTCGAGAACGCCGAC
ATCCGCCAGATCAGCGCCACCCGCGAGATCAACGACGGCTACGCCAACGGC
TACGGCAACAGCAAGTGGGCCGGCGAGGTGTTGCTGCGCGAGGCCCACGAC
CTGTGCGGCCTGCCCGTCACGGTGTTCCGCTGCGACATGATCCTGGCCGAC
ACCACCTATGCCGGGCAGCTCAACCTGCCCGACATGTTCACCCGGCTGATG
CTGAGCCTGGTCGCCACCGGTATCGCGCCCGGGTCGTTCTACGAACTGGAC
GCCGACGGCAACCGCCAGCGGGCACACTACGACGGTTTGCCGGTCGAGTTC
ATCGCCGCGGCGATCTCGACGCTGGGGACCCAAATCACCGACAGCGACACG
GGCTTTCAGACCTACCACGTGATGAACCCCTACGACGACGGCATCGGGCTG
GATGAGTACATCGATTGGCTGATCGAGGCCGGGTATTCGATCGAGCGGATC
GCCGATTACTCCGAGTGGCTGCGGCGCTTCGAGACCTCGCTGCGGGCCCTG
CCCGATCGGCAGCGTCAGTACTCGCTGCTGCCGCTGCTGCACAACTACCAG
AAGCCGGAAAAGCCGATCAACGGCTCGATGGCGCCCACCGACGTGTTCCGT
GCCGCGGTGCAGGAAGCGAAAATCGGCCCCGACAAAGACATCCCGCACGTC
TCGGCGCCGGTGATCGTCAAGTACATCACCGACCTGGAGTTGCTCGGACTC
CTCTGA

Amino acid sequence (SEQ ID NO:272)

>gi|254819907|ref|ZP_05224908.1| FadD9
[Mycobacterium intracellulare ATCC 13950]

MSTAIHDEHLDRRIEELIANDPQFAAARPDPAITAATEAPGLRLPQIIRTV
LDGYADRPALAQRVVEFVTDAKTGRTTAELLPRFETITYGELGERVSALGR
AWAGDAVRPGDRVCVLGFNSVDYATIDIALGTIGAVSVPLQTSAAISSLQP
IVAETEPSLIASSVNQLPDAVELILAGDHVPGKLVVFDYQPQVDDQREAVE
AAAARLADSGVAVEALADVLRRGKDLPAVEPPASDEDSLALLIYTSGSTGA
PKGAMYPQSNVGKMWRRGSKNWFGESAASITLNFMPMSHVMGRGILYGTLG
NGGTAYFAARSDLSTLLEDLELVRPTEMNFVPRIWETLYGEFQRQVERRLA

FIG. 8 (Cont.)

DGDAGPEARETVAAAVLEEQRQYLLGGRFIFAMTGSAPTSPELKAWAESLL
QMHLMDGYGSTEAGMVLFDGEIQRPPVIDYKLVDVPDLGYFSTDRPHPRGE
LLLRTENMFPGYYKRAETTANVFDEDGYYRTGDVFAEIAPDRLVYVDRRNN
VLKLAQGEFVTLAKLEAVFGNSPLIRQIYVYGNSSQPYLLAVVVPTEEALA
DNDLESLKPKIADSLQKVAKETGLQSYEVPRDFIIETTPFTLENGLLTGIR
KLAWPKLKAHYGDRLEQMYAELAAGQANELAELRRSGAAAPVAQTVSRAAA
ALLGATAGDLSADAHFTDLGGDSLSALTFGNLLREIFDVDVPVGVIVSPAN
DLAGIAAYIEAERQGSKRPTFAAVHGRGATMVHASDLTLDKFLDEATLAAA
PSLPKPATEVRTVLLTGATGFLGRYLALDWLERMDMVDGKVIALVRARTDE
EARARLDKTFDSGDPKLLAHYQRLAADHLEVIAGDKGEANLGLDPQTWQRL
AEEVDVIVDPAALVNHVLPYSELFGPNALGTAELIRIALTSRQKPYTYVST
IGVGDQIQPGEFVENADIRQISATREINDGYANGYGNSKWAGEVLLREAHD
LCGLPVTVFRCDMILADTTYAGQLNLPDMFTRLMLSLVATGIAPGSFYELD
ADGNRQRAHYDGLPVEFIAAAISTLGTQITDSDTGFQTYHVMNPYDDGIGL
DEYIDWLIEAGYSIERIADYSEWLRRFETSLRALPDRQRQYSLLPLLHNYQ
KPEKPINGSMAPTDVFRAAVQEAKIGPDKDIPHVSAPVIVKYITDLELLGL
L

FIG. 9

Accession Numbers are from NCBI, GenBank, Release 159.0 as of March 2008
EC Numbers are from KEGG, Release 42.0 as of April 2007 (plus daily updates up to March 2008)

CATEGORY	GENE	NAME	ACCESSION	EC NUMBER	MODIFICATION	USE	ORGANISM
1. Fatty Acid Production Increase / Product Production Increase							
increase acyl-CoA							
reduce catabolism of derivatives and intermediates							
reduce feedback inhibition							
attenuate other pathways that consume fatty acids							
	accA	Acetyl-CoA carboxylase, subunit A (carboxyltransferase alpha)	AAC73296, NP_414727	6.4.1.2	Over-express	increase Malonyl-CoA production	*Escherichia coli, Lactococci*
	accB	Acetyl-CoA carboxylase, subunit B (BCCP: biotin carboxyl carrier protein)	NP_417721	6.4.1.2	Over-express	increase Malonyl-CoA production	*Escherichia coli, Lactococci*
	accC	Acetyl-CoA carboxylase, subunit C (biotin carboxylase)	NP_417722	6.4.1.2, 6.3.4.14	Over-express	increase Malonyl-CoA production	*Escherichia coli, Lactococci*
	accD	Acetyl-CoA carboxylase, subunit D (carboxyltransferase beta)	NP_416819	6.4.1.2	Over-express	increase Malonyl-CoA production	*Escherichia coli, Lactococci*

SUBSTITUTE SHEET (RULE 26)

FIG. 9 (Cont.)

	aceE	pyruvate dehydrogenase, subunit E1	NP_414656, AAC73226	1.2.4.1	Over-express	increase Acetyl-CoA production	*Escherichia coli*
	aceF	pyruvate dehydrogenase, subunit E2	NP_414657	2.3.1.12	Over-express	increase Acetyl-CoA production	*Escherichia coli*
	ackA	acetate kinase	AAC75356, NP_416799	2.7.2.1	Delete or reduce	increase Acetyl-CoA production	*Escherichia coli*
	ackB	acetate kinase AckB	BAB81430	2.7.2.1	Delete or reduce	increase Acetyl-CoA production	*Escherichia coli*
	acpP	acyl carrier protein	AAC74178	NONE	Over-express	increase Acetyl-CoA production	*Escherichia coli*
	fadD	acyl-CoA synthase	AP_002424	2.3.1.86, 6.2.1.3	Over-express	increase Fatty acid production	*Escherichia coli W3110*
	adhE	alcohol dehydrogenase	CAA47743	1.1.1.1, 1.2.1.10	Delete or reduce	increase Acetyl-CoA production	*Escherichia coli W3111*
	cer1	Aldehyde decarbonylase	BAA11024	4.1.99.5	Over-express	increase Acetyl-CoA production	*Arabidopsis thaliana*
	fabA	beta-hydroxydecanoyl thioester dehydrase	NP_415474	4.2.1.60	express	fatty acyl-CoA production	*E. coli K12*
	fabD	[acyl-carrier-protein] S-malonyltransferase	AAC74176	2.3.1.39	Over-express	increase Acetyl-CoA production	*E. coli K12*
	fabF	3-oxoacyl-[acyl-carrier-protein] synthase II	AAC74179	2.3.1.179	Delete or OverExpress	increase Acetyl-CoA production	*E. coli K12*
	fabG	3-oxoacyl-[acyl-carrier protein] reductase	AAC74177	1.1.1.100	Over-express	increase Acetyl-CoA production	*E. coli K12*
	fabH	3-oxoacyl-[acyl-carrier-protein] synthase III	AAC74175	2.3.1.180	Over-express	increase Acetyl-CoA production	*E. coli K12, lactococci*

FIG. 9 (Cont.)

	fabI	enoyl-[acyl-carrier-protein] reductase, NADH-dependent	NP_415804	1.3.1.9	express	fatty acyl-CoA production	*E. coli K12, lactococci*
	fabR	Transcriptional Repressor	NP_418398	NONE	Delete or reduce	modulate unsaturated fatty acid production	*E. coli K12*
	fabZ	(3R)-hydroxymyristol acyl carrier protein dehydratase	NP_414722	4.2.1.-			*E. coli K12*
	fadE	acyl-CoA dehydrogenase	AAC73325	1.3.99.3, 1.3.99.-	Delete or reduce	increase Acetyl-CoA production	
	acr1	Fatty Acyl-CoA reductase	YP_047869, AAC45217	1.2.1.42	Over-express	for fatty alcohol production	*Acinetobacter sp., i.e. calcoaceticus*
	GST, gshB	Glutathione synthase	P04425	6.3.2.3	Delete or reduce	increase Acyl-CoA	*E. coli K12*
	gpsA	biosynthetic sn-glycerol 3-phosphate dehydrogenase	AAC76632, NP_418065	EC: 1.1.1.94	Delete or reduce	increase Acetyl-CoA production	*E. coli K12*
	ldhA	lactate dehydrogenase	AAC74462, NP_415898	EC: 1.1.1.27, 1.1.1.28	Delete or reduce	increase Acetyl-CoA production	*E. coli K12*
	Lipase	Triglyceride Lipase	CAA89087, CAA98876	3.1.1.3	express	increase Fatty acid production	*Saccharomyces cerevisiae*
		Malonyl-CoA decarboxylase	AAA26500	4.1.1.9, 4.1.1.41	Over-express		*Saccharopolyspora erythraea*
	panD	aspartate 1-decarboxylase	BAB96708	4.1.1.11	Over-express	increase Acyl-CoA	*Escherichia coli W3110*

SUBSTITUTE SHEET (RULE 26)

FIG. 9 (Cont.)

	panK a.k.a. coaA	pantothenate kinase	AAC76952	2.7.1.33	Over-express	increase Acetyl-CoA production	*E. coli*
	panK a.k.a. coaA, R106K	pantothenate kinase	AAC76952	2.7.1.33	Express, Over-express, R106K mutation	increase Acetyl-CoA production	*E. coli*
	pdh	Pyruvate dehydrogenase	BAB34380, AAC73226, NP_415392	1.2.4.1	Over-express	increase Acetyl-CoA production	
	pflB	formate acetyltransferase (pyruvate formate lyase)	AAC73989, P09373	EC: 2.3.1.54	Delete or reduce	increase Acetyl-CoA production	
	plsB	acyltransferase	AAC77011	2.3.1.15	D311E mutation	reduce limits on Acyl-CoA pool	*E. coli K12*
	poxB	pyruvate oxidase	AAC73958, NP_415392	1.2.2.2	Delete or reduce	increase Acetyl-CoA production	
	pta	phosphotransacetylase	AAC75357, NP_416800	2.3.1.8	Delete or reduce	increase Acetyl-CoA production	
	udhA	pyridine nucleotide transhydrogenase	CAA46822	1.6.1.1	Over-express	conversion NADH to NADPH or vice versa	
	fadB	fused 3-hydroxybutyryl-CoA epimerase/delta(3)-cis-delta(2)-trans-enoyl-CoA isomerase/enoyl-CoA hydratase and 3-hydroxyacyl-CoA dehydrogenase	AP_003956	4.2.1.17, 5.1.2.3, 5.3.3.8, 1.1.1.35	Delete or reduce	Block fatty acid degradation	*E. coli*

SUBSTITUTE SHEET (RULE 26)

FIG. 9 (Cont.)

	fadJ	3-hydroxyacyl-CoA dehydrogenase; K01692 enoyl-CoA hydratase; K01782 3-hydroxybutyryl-CoA epimerase	AAC75401	1.1.1.35, 4.2.1.17, 5.1.2.3	Delete or reduce	Block fatty acid degradation	*E. coli*
	fadA	3-ketoacyl-CoA thiolase	BAE77458	2.3.1.16	Delete or reduce	Block fatty acid degradation	*E. coli*
	fadI	beta-ketoacyl-CoA thiolase	AAC75402	2.3.1.16	Delete or reduce	Block fatty acid degradation	*E. coli*
	YdiO	acyl-coA dehydrogenase	YP_852786	1.3.99.-	Delete or reduce	Block fatty acid degradation	*E. coli*
2. Structure Control							
2A. Chain Length Control							
2	tesA	thioesterase	P0ADA1	3.1.2.-, 3.1.1.5	Delete and/or express	C18 Chain Length	
	tesA without leader sequence	thioesterase	AAC73596, NP_415027	3.1.2.-, 3.1.1.5	express or overexpress	C18:1	*E.coli*
	tesA without leader sequence:L109P	thioesterase	P0ADA1	3.1.2.-, 3.1.1.5	Express and/or overexpress mutation L109P	<C18 Chain Length	*E. coli*
	fatB1 (umbellularia)	thioesterase	Q41635	3.1.2.14	express or overexpress	C12:0	*Umbellularia californica*
	fatB2 (umbellularia)DELETE umbelluria)	thioesterase	AAC49269	3.1.2.14	express or overexpress	C8:0 - C10:0	*Cuphea hookeriana*

SUBSTITUTE SHEET (RULE 26)

FIG. 9 (Cont.)

	fatB3	thioesterase	AAC72881	3.1.2.14	express or overexpress	C14:0 - C16:0	*Cuphea hookeriana*
	fatB (cinnamonum)	thioesterase	Q39473	3.1.2.14	express or overexpress	C14:0	*Cinnamomum camphora*
	fatB[M141T]*	thioesterase	CAA85388	3.1.2.14	express or overexpress	C16:1	*Arabidopsis thaliana*
	fatA1 (Helianthus)	thioesterase	AAL79361	3.1.2.14	express or overexpress	*C18:1*	*Helianthus annuus*
	atfata (ARABIDOPSIS FATA ACYL-ACP THIOESTERASE)	thioesterase	NP_189147, NP_193041	3.1.2.14	express or overexpress	C18:1	*Arabidopsis thaliana*
	fatA	thioesterase	CAC39106	3.1.2.14	express or overexpress	C18:1	*Brassica juncea*
	fatA (cuphea)	thioesterase	AAC72883	3.1.2.14	express or overexpress	C18:1	*Cuphea hookeriana*
2B. Branching Control							
	attenuate FabH						
	express FabH from S. glaucescens or S. coelicolor and knock out endogenouse FabH					increase branched chain fatty acid derivatives	

SUBSTITUTE SHEET (RULE 26)

FIG. 9 (Cont.)

	express FabH from B. subtilis and knock out endogenouse FabH						
	bdk - E3 - dihydroplipoyl dehyrodgenase subunit			EC 1.2.4.4			
	bkd - E1-alpha/beta subunit	decarboxylase subunits of branched-chain α-keto acid dehydrogenase complex		EC 1.2.4.4			
	bkd - E2 - dihydrolipoyl transacylase subunit			EC 1.2.4.4			
	bkdA1	branched-chain α-keto acid dehydrogenase a-subunit (E1a)	NP_628006	EC 1.2.4.4	express or Over-Express	make branched-chain acyl-CoA precursors	*Streptomyces coelicolor*
	bkdB1	branched-chain α-keto acid dehydrogenase a-subunit (E1b)	NP_628005	EC 1.2.4.4	express or Over-Express	make branched-chain acyl-CoA precursors	*Streptomyces coelicolor*
	bkdC1	dihydrolipoyl transacetylase (E2)	NP_628004	EC 2.3.1.168	express or Over-Express	make branched-chain acyl-CoA precursors	*Streptomyces coelicolor*

SUBSTITUTE SHEET (RULE 26)

CA 02740037 2011-04-08

FIG. 9 (Cont.)

	bkdA2	branched-chain a-ketoacid dehydrogenase a-subunit (E1a)	NP_733618	EC 1.2.4.4	express or Over-Express	make branched-chain acyl-CoA precursors	*Streptomyces coelicolor*
	bkdB2	branched-chain a-ketoacid dehydrogenase b-subunit (E1b)	NP_628019	EC 1.2.4.4	express or Over-Express	make branched-chain acyl-CoA precursors	*Streptomyces coelicolor*
	bkdC2	dihydrolipoyl transacetylase (E2)	NP_628018	EC 2.3.1.168	express or Over-Express	make branched-chain acyl-CoA precursors	*Streptomyces coelicolor*
	bkdA	branched-chain a-ketoacid dehydrogenase a-subunit (E1a)	BAC72074	EC 1.2.4.4	express or Over-Express	make branched-chain acyl-CoA precursors	*Streptomyces avermitilis*
	bkdB	branched-chain a-ketoacid dehydrogenase b-subunit (E1b)	BAC72075	EC 1.2.4.4	express or Over-Express	make branched-chain acyl-CoA precursors	*Streptomyces avermitilis*
	bkdC	dihydrolipoyl transacetylase (E2)	BAC72076	EC 2.3.1.168	express or Over-Express	make branched-chain acyl-CoA precursors	*Streptomyces avermitilis*
	bkdF	branched-chain a-ketoacid dehydrogenase a-subunit (E1a)	BAC72088	EC 1.2.4.4	express or Over-Express	make branched-chain acyl-CoA precursors	*Streptomyces avermitilis*

SUBSTITUTE SHEET (RULE 26)

FIG. 9 (Cont.)

	bkdG	branched-chain a-ketoacid dehydrogenase b-subunit (E1b)	BAC72089	EC 1.2.4.4	express or Over-Express	make branched-chain acyl-CoA precursors	*Streptomyces avermitilis*
	bkdH	dihydrolipoyl transacetylase (E2)	BAC72090	EC 2.3.1.168	express or Over-Express	make branched-chain acyl-CoA precursors	*Streptomyces avermitilis*
	bkdAA	branched-chain a-ketoacid dehydrogenase a-subunit (E1a)	NP_390285	EC 1.2.4.4	express or Over-Express	make branched-chain acyl-CoA precursors	*Bacillus subtilis*
	bkdAB	branched-chain a-ketoacid dehydrogenase b-subunit (E1b)	NP_390284	EC 1.2.4.4	express or Over-Express	make branched-chain acyl-CoA precursors	*Bacillus subtilis*
	bkdB	dihydrolipoyl transacetylase (E2)	NP_390283	EC 2.3.1.168	express or Over-Express	make branched-chain acyl-CoA precursors	*Bacillus subtilis*
	bkdA1	branched-chain a-ketoacid dehydrogenase a-subunit (E1a)	AAA65614	EC 1.2.4.4	express or Over-Express	make branched-chain acyl-CoA precursors	*Pseudomonas putida*
	bkdA2	branched-chain a-ketoacid dehydrogenase b-subunit (E1b)	AAA65615	EC 1.2.4.4	express or Over-Express	make branched-chain acyl-CoA precursors	*Pseudomonas putida*
	bkdC	dihydrolipoyl transacetylase (E2)	AAA65617	EC 2.3.1.168	express or Over-Express	make branched-chain acyl-CoA precursors	*Pseudomonas putida*

SUBSTITUTE SHEET (RULE 26)

FIG. 9 (Cont.)

	lpd	dihydrolipoamide dehydrogenase (E3)	NP_414658	1.8.1.4	express or Over-Express	make branched-chain acyl-CoA precursors	*Escherichia coli*
	IlvE	branched-chain amino acid aminotransferase	YP_026247	2.6.1.42	express or Over-Express	make branched a-ketoacids	*Escherichia coli*
	IlvE	branched-chain amino acid aminotransferase	AAF34406	2.6.1.42	express or Over-Express	make branched a-ketoacids	*Lactococcus lactis*
	IlvE	branched-chain amino acid aminotransferase	NP_745648	2.6.1.42	express or Over-Express	make branched a-ketoacids	*Pseudomonas putida*
	IlvE	branched-chain amino acid aminotransferase	NP_629657	2.6.1.42	express or Over-Express	make branched a-ketoacids	*Streptomyces coelicolor*
	ccr	crotonyl-CoA reductase	NP_630556	1.6.5.5,1.1.1.1	express or Over-Express	Converting crotonyl-CoA to butyryl-CoA	*Streptomyces coelicolor*
	ccr	crotonyl-CoA reductase	AAD53915	1.6.5.5,1.1.1.1	express or Over-Express	Converting crotonyl-CoA to butyryl-CoA	*Streptomyces cinnamonensis*
	IcmA, isobutyryl-CoA mutase	isobutyryl-CoA mutase, subunit A	NP_629554	5.4.99.2	express or Over-Express	converting butyryl-CoA to isobutyryl-CoA	*Streptomyces coelicolor*
	IcmA, isobutyryl-CoA mutase	isobutyryl-CoA mutase, subunit A	AAC08713	5.4.99.2	express or Over-Express	converting butyryl-CoA to isobutyryl-CoA	*Streptomyces cinnamonensis*
	IcmB, isobutyryl-CoA mutase	isobutyryl-CoA mutase, subunit B	NP_630904	5.4.99.2	express or Over-Express	converting butyryl-CoA to isobutyryl-CoA	*Streptomyces coelicolor*
	IcmB, isobutyryl-CoA mutase	isobutyryl-CoA mutase, subunit B	CAB59633	5.4.99.2	express or Over-Express	converting butyryl-CoA to isobutyryl-CoA	*Streptomyces cinnamonensis*

SUBSTITUTE SHEET (RULE 26)

FIG. 9 (Cont.)

	FabH, ACPs and fabF genes with specificity for branched chain acyl-CoAs						
	IlvE	branched-chain amino acid aminotransferase	CAC12788	EC2.6.1.42	over express	branched chain amino acid amino transferase	Staphylococcus carnosus
	FabH1	beta-ketoacyl-ACP synthase III	NP_626634	2.3.1.180	express or Over-Express	initiation of branched-chain fatty acid biosynthesis	Streptomyces coelicolor
	ACP	acyl-carrier protein	NP_626635	NONE	express or Over-Express	initiation and elongation of branched-chain fatty acid biosynthesis	*Streptomyces coelicolor*
	FabF	beta-ketoacyl-ACP synthase II	NP_626636	2.3.1.179	express or Over-Express	elongation of branched-chain fatty acid biosynthesis	*Streptomyces coelicolor*
	FabH3	beta-ketoacyl-ACP synthase III	NP_823466	2.3.1.180	express or Over-Express	initiation of branched-chain fatty acid biosynthesis	*Streptomyces avermitilis*

FIG. 9 (Cont.)

	FabC3 (ACP)	acyl-carrier protein	NP_823467	NONE	express or Over-Express	initiation and elongation of branched-chain fatty acid biosynthesis	*Streptomyces avermitilis*
	FabF	beta-ketoacyl-ACP synthase II	NP_823468	2.3.1.179	express or Over-Express	elongation of branched-chain fatty acid biosynthesis	*Streptomyces avermitilis*
	FabH_A	beta-ketoacyl-ACP synthase III	NP_389015	2.3.1.180	express or Over-Express	initiation of branched-chain fatty acid biosynthesis	*Bacillus subtillis*
	FabH_B	beta-ketoacyl-ACP synthase III	NP_388898	2.3.1.180	express or Over-Express	initiation of branched-chain fatty acid biosynthesis	*Bacillus subtillis*
	ACP	acyl-carrier protein	NP_389474	NONE	express or Over-Express	initiation and elongation of branched-chain fatty acid biosynthesis	*Bacillus subtillis*
	FabF	beta-ketoacyl-ACP synthase II	NP_389016	2.3.1.179	express or Over-Express	elongation of branched-chain fatty acid biosynthesis	*Bacillus subtillis*

SUBSTITUTE SHEET (RULE 26)

FIG. 9 (Cont.)

	SmalDRAFT_0818	beta-ketoacyl-ACP synthase III	ZP_01643059	2.3.1.180	express or Over-Express	initiation of branched-chain fatty acid biosynthesis	*Stenotrophomonas maltophilia*
	SmalDRAFT_0821	acyl-carrier protein	ZP_01643063	NONE	express or Over-Express	initiation and elongation of branched-chain fatty acid biosynthesis	*Stenotrophomonas maltophilia*
	SmalDRAFT_0822	beta-ketoacyl-ACP synthase II	ZP_01643064	2.3.1.179	express or Over-Express	elongation of branched-chain fatty acid biosynthesis	*Stenotrophomonas maltophilia*
	FabH	beta-ketoacyl-ACP synthase III	YP_123672	2.3.1.180	express or Over-Express	initiation of branched-chain fatty acid biosynthesis	*Legionella pneumophila*
	ACP	acyl-carrier protein	YP_123675	NONE	express or Over-Express	initiation and elongation of branched-chain fatty acid biosynthesis	*Legionella pneumophila*
	FabF	beta-ketoacyl-ACP synthase II	YP_123676	2.3.1.179	express or Over-Express	elongation of branched-chain fatty acid biosynthesis	*Legionella pneumophila*

FIG. 9 (Cont.)

	FabH	beta-ketoacyl-ACP synthase III	NP_415609	2.3.1.180	delete or reduce	initiation of branched-chain fatty acid biosynthesis	*Escherichia coli*
	FabF	beta-ketoacyl-ACP synthase II	NP_415613	2.3.1.179	delete or reduce	elongation of branched-chain fatty acid biosynthesis	*Escherichia coli*
To Produce Cyclic Fatty Acids							
	AnsJ	dehydratase (putative)	not available	not available	express or Over-Express	cyclohexylcarbonyl-CoA boiosynthesis	*Streptomyces collinus*
	AnsK	CoA ligase (putative)	not available	not available	express or Over-Express	cyclohexylcarbonyl-CoA boiosynthesis	*Streptomyces collinus*
	AnsL	dehydrogenase (putative)	not available	not available	express or Over-Express	cyclohexylcarbonyl-CoA boiosynthesis	*Streptomyces collinus*
	ChcA	enoyl-CoA reductase	U72144	EC 1.3.1.34	express or Over-Express	cyclohexylcarbonyl-CoA boiosynthesis	*Streptomyces collinus*
	AnsM	oxidorecutase (putative)	not available	not available	express or Over-Express	cyclohexylcarbonyl-CoA boiosynthesis	*Streptomyces collinus*
	PlmJ	dehydratase (putative)	AAQ84158	not available	express or Over-Express	cyclohexylcarbonyl-CoA boiosynthesis	*Streptomyces sp. HK803*

SUBSTITUTE SHEET (RULE 26)

FIG. 9 (Cont.)

	PlmK	CoA ligase (putative)	AAQ84158	not available	express or Over-Express	cyclohexylcarbonyl-CoA boiosynthesis	*Streptomyces sp. HK803*
	PlmL	dehydrogenase (putative)	AAQ84159	not available	express or Over-Express	cyclohexylcarbonyl-CoA boiosynthesis	*Streptomyces sp. HK803*
	ChcA	enoyl-CoA reductase	AAQ84160	EC 1.3.1.34	express or Over-Express	cyclohexylcarbonyl-CoA boiosynthesis	*Streptomyces sp. HK803*
	PlmM	oxidorecutase (putative)	AAQ84161	not available	express or Over-Express	cyclohexylcarbonyl-CoA boiosynthesis	*Streptomyces sp. HK803*
	ChcB	enoyl-CoA isomerase	AF268489	not available	express or Over-Express	cyclohexylcarbonyl-CoA boiosynthesis	*Streptomyces collinus*
	ChcB/CaiD	enoyl-CoA isomerase	NP_629292	4.2.1.-	express or Over-Express	cyclohexylcarbonyl-CoA boiosynthesis	*Streptomyces coelicolor*
	ChcB/CaiD	enoyl-CoA isomerase	NP_824296	4.2.1.-	express or Over-Express	cyclohexylcarbonyl-CoA boiosynthesis	*Streptomyces avermitilis*
2C. Saturation Level Control							
	Sfa	Suppressor of FabA	AAN79592, AAC44390	NONE	Over-express	increase monounsaturated fatty acids	*E.coli*
	also *see* FabA in sec. 1				express	produce unsaturated fatty acids	

SUBSTITUTE SHEET (RULE 26)

FIG. 9 (Cont.)

	GnsA	suppressors of the secG null mutation	ABD18647.1	NONE	Over-express	increase unsaturated fatty acid esters	*E.coli*
	GnsB	suppressors of the secG null mutation	AAC74076.1	NONE	Over-express	increase unsaturated fatty acid esters	*E.coli*
	also *see* section 2A - items with :0 are unsaturated (no double bonds) and with :1 are saturated (1 double bond)						
	fabB	3-oxoacyl-[acyl-carrier-protein] synthase I	BAA16180	EC:2.3.1.41	overexpress	modulate unsaturated fatty acid production	*Escherichia coli*
	fabK	trans-2-enoyl-ACP reductase II	AAF98273	1.3.1.9	express	modulate unsaturated fatty acid production	*Streptococcus pneumoniae*
	fabL	enoyl-(acyl carrier protein) reductase	AAU39821	1.3.1.9	express	modulate unsaturated fatty acid production	*Bacillus licheniformis DSM 13*

SUBSTITUTE SHEET (RULE 26)

FIG. 9 (Cont.)

	fabM	trans-2, cis-3-decenoyl-ACP isomerase	DAA05501	4.2.1.17	Over-express	modulate unsaturated fatty acid production	*Streptococcus mutans*
Fatty Aldehyde Output							
	thioesterase	*see* chain length control section			express	produce	
Export							
	Wax ester exporter (FATP family, Fatty Acid (long chain) Transport Protein)		NP_524723	NONE	express	export wax	*Drosophila melanogaster*
	ABC transport protein	putative alkane transporter	AAN73268	NONE	express	export products	*Rhodococcus erythropolis*
	CER5	wax transporter	At1g51500, AY734542, At3g21090, At1g51460	NONE	express	export products	*Arabidopsis thaliana*
	AtMRP5	Arabidopsis thaliana multidrug resistance-associated	NP_171908	NONE	express	export products	*Arabidopsis thaliana*
	AmiS2	ABC transporter AmiS2	JC5491	NONE	express	export products	*Rhodococcus sp.*
	AtPGP1	ARABIDOPSIS THALIANA P GLYCOPROTEIN1	NP_181228	NONE	express	export products	*Arabidopsis thaliana*

SUBSTITUTE SHEET (RULE 26)

FIG. 9 (Cont.)

	AcrA	putative multidrug-efflux transport protein acrA	CAF23274	NONE	express	export products	*Candidatus Protochlamydia amoebophila UWE25*
	AcrB	probable multidrug-efflux transport protein, acrB	CAF23275	NONE	express	export products	*Candidatus Protochlamydia amoebophila UWE25*
	TolC	Outer membrane protein [Cell envelope biogenesis,	ABD59001	NONE	express	export products	*Francisella tularensis subsp. novicida*
	AcrE	transmembrane protein affects septum formation and cell membrane permeability	YP_312213	NONE	express	export products	*Shigella sonnei Ss046*
	AcrF	Acriflavine resistance protein F	P24181	NONE	express	export products	*Escherichia coli*
	tll1618	multidrug efflux transporter	NP_682408.1	NONE	express	export products	*Thermosynechococcus elongatus BP-1]*
	tll1619	multidrug efflux transporter	NP_682409.1	NONE	express	export products	*Thermosynechococcus elongatus BP-1]*
	tll0139	multidrug efflux transporter	NP_680930.1	NONE	express	export products	*Thermosynechococcus elongatus BP-1]*

SUBSTITUTE SHEET (RULE 26)

CA 02740037 2011-04-08

FIG. 9 (Cont.)

5. Fermentation							
	replication checkpoint genes					increase output efficiency	
	umuD	DNA polymerase V, subunit	YP_310132	3.4.21.-	Over-express	increase output efficiency	*Shigella sonnei Ss046*
	umuC	DNA polymerase V, subunit	ABC42261	2.7.7.7	Over-express	increase output efficiency	*Escherichia coli*
	NADH:NADPH transhydrogenase (alpha and beta subunits) (pntA, pntB)		P07001, P0AB70	1.6.1.2	express	increase output efficiency	*Shigella flexneri*

SUBSTITUTE SHEET (RULE 26)

	Annotations	Pfam	Accession #
Zn-dep. Alcohol Dehydrogenases			
YjgB	best homolog of *Acinetobacter* alrA (50%)	00107, 08240	NP 418690
YahK	best homolog of yjgB (31%)	00107, 08240	NP 414859
AdhP	active on ethanol, more efficient as aldehyde reductase	00107, 08240	NP 415995
YdjL	predicted oxidoreductase, Zn-dependent and NAD(P)-binding	00107, 08240	NP 416290
YdjJ	predicted oxidoreductase, Zn-dependent and NAD(P)-binding	00107, 08240	NP 416288
YjgV (idnD)	oxidizes L-idonate using NAD and reduces of 5-ketogluconate using NADH or NADPH	00107, 08240	NP 418688
Tdh	converts threonine and NAD to 1,2-amino-3-oxobutanoate and NADH	00107, 08240	NP 418073
YjjN	possible L-galactonate oxidoreductase	00707, 08240	NP 418778
RspB	predicted oxidoreductase, Zn-dependent and NAD(P)-binding unknown substrate	00707, 08240	NP 416097
GatD	galactitol-1-phosphate dehydrogenase, Zn-dependent and NAD(P)-binding	00107, 08240	NP 416594
YphC	predicted oxidoreductase, Zn-dependent and NAD(P)-binding	00107, 08240	NP 417040
YhdH	predicted oxidoreductase, Zn-dependent and NAD(P)-binding(3D)	00107, 08240	NP 417719
YcjQ	predicted oxidoreductase, Zn-dependent and NAD(P)-binding large operon	00107	NP 415829
YncB	predicted oxidoreductase, Zn-dependent and NAD(P)-binding	00107	NP 415966
Qor	NADPH:quinone reductase and related Zn-dependent oxidoreductases	00107, 08240	NP 418475
Adh3(fmnA)	alcohol dehydrogenase class III/glutathione-dependent formaldehyde dehydrogenase	00107, 08240	NP 414890
YbdR	predicted oxidoreductase, Zn-dependent and NAD(P)-binding	00107, 08240	NP 415141
YggP	predicted oxidoreductase, Zn-dependent and NAD(P)-binding	00107, 08240	YP 026187
Fe-dep. Alcohol Dehydrogenases			
YiaY	unknown funtion, operon with aldB (aldehyde dehydrogenase)	00465	YP 026233
FucO	glyceroaldehyde to ethylene glycole or lactaldehyde to 1,3-propanediol	00465	NP 417279
EutG	ethanolamine utilization?	00465	NP 416948
YqhD	reduces butanal, response mechanism to lipid peroxygenation? (3D)	00465	NP 417484
AdhE	three Fe++dependent catalytic functions: alcohol dh, acetaldehyde dh and pyruvate formate-lyase deactivase, poorly active with pentanol	00465	NP 415757

FIG. 10

SUBSTITUTE SHEET (RULE 26)

CA 02740037 2011-04-08

SUBSTITUTE SHEET (RULE 26)

	Annotations	Pfam	Accession #
Aldo-Keto Reductases			
yafB (dkgB)	reduces methylglyoxal, 2,5-diketogluconate and 4-nitrobenzaldehyde	00248	NP 414743
YdjG	reduces methy glyoxal, NADH specific	00248	NP 416285
YeaE	reduces methy glyoxal	00248	NP 416295
YqhE(dkgA)	reduces methy glyoxal and 2,5-diketogluconate	00248	NP 417485
YajO	putative NAD(P)H-dependent xylose reductase	00248	NP 414953
YghZ	reduces methy glyoxal	00248	NP 417474
Tas	Multicopy expression suppresses prephenate dehydrogenase defect	00248	NP 417311
YdhF	predicted oxidoreductase	00248	YP 025305
YdbC	predicted oxidoreductase	00248	NP 415924
Short-Chain Dehydrogenases			
ybbO	short chain dehydrogenase	pfam00106	NP 415026
yohF	predicted oxidoreductase with NAD(P)-binding Rossmann-fold domain	pfam00106	NP 416641
YciK	short chain dehydrogenase	pfam00106	NP 415787
YgfF	predicted NAD(P)-binding oxidoreductase withNAD(P)-binding Rossmann-fold domain	pfam00106	NP 417378
YghA	oxidoreductase	pfam00106	NP 417476
Yjgl	predicted oxidoreductase with NAD(P)-binding Rossmann-fold domain	pfam00106	NP 418670
YdfG	putative oxidoreductase, L-allo-threonine dehydrogenase, NAD(P)-binding	pfam00106	NP 416057
YgcW	putative oxidoreductase, predicted deoxygluconate dehydrogenase	pfam00106	NP 417254
UcpA	predicted acetoin dehydrogenase (diacetyl reductase)	pfam00106	NP 416921
EntA	2,3-Dihydro-2,3-dihydroxybenzoate dehydrogenase	pfam00106	NP 415128
FolM	dihydrofolate reductase, THF biosynthesis	pfam00106	NP 416123
HdhA	7-alpha-hydroxysteroid dehydrogenase	pfam00106	NP 416136
HcaB	2,3-dihydroxy-2,3-dihydrophenylpropionate dehydrogenase, (aromatization)	pfam00106	NP 417036
SrlD	sorbitol-6-phosphate 2-dehydrogenase, glucitol-6-phosphate dehydrogenase	pfam00106	NP 417185
KduD	2-deoxy-D-gluconate 3-dehydrogenase	pfam00106	NP 417319
IdnO	gluconate 5-dehydrogenase, 5-keto-D-gluconate 5-reductase	pfam00106	NP 418687
FabG	3-oxoacyl-[acyl-carrier-protein] reductase	pfam00106	NP 415611
FabI	enoyl-(acyl carrier protein) reductase	pfam00106	NP 415804
YdjA	predicted oxidoreductase (FAD dep.)	pfam00881	NP 416279

FIG. 10 Cont.

FIG. 10 (Cont)

ygjB (NP_418690)

Nucleotide sequence (SEQ ID NO:93)

```
   1 atgtcgatga taaaaagcta tgccgcaaaa gaagcgggcg gcgaactgga agtttatgag
  61 tacgatcccg gtgagctgag gccacaagat gttgaagtgc aggtggatta ctgcgggatc
 121 tgccattccg atctgtcgat gatcgataac gaatggggat tttcacaata tccgctggtt
 181 gccgggcatg aggtgattgg gcgcgtggtg gcactcggga gcgccgcgca ggataaaggt
 241 ttgcaggtcg gtcagcgtgt cgggattggc tggacggcgc gtagctgtgg tcactgcgac
 301 gcctgtatta gcggtaatca gatcaactgc gagcaaggtg cggtgccgac gattatgaat
 361 cgcggtggct ttgccgagaa gttgcgtgcg gactggcaat gggtgattcc actgccagaa
 421 aatattgata tcgagtccgc cgggccgctg ttgtgcggcg gtatcacggt ctttaaacca
 481 ctgttgatgc accatatcac tgctaccagc cgcgttgggg taattggtat tggcgggctg
 541 gggcatatcg ctataaaact tctgcacgca atgggatgcg aggtgacagc ctttagttct
 601 aatccggcga aagagcagga agtgctggcg atgggtgccg ataaagtggt gaatagccgc
 661 gatccgcagg cactgaaagc actggcgggg cagtttgatc tcattatcaa caccgtcaac
 721 gtcagcctcg actggcagcc ctattttgag gcgctgacct atggcggtaa tttccatacg
 781 gtcggtgcgg ttctcacgcc gctgtctgtt ccggccttta cgttaattgc gggcgatcgc
 841 agcgtctctg gttctgctac cggcacgcct tatgagctgc gtaagctgat gcgttttgcc
 901 gcccgcagca aggttgcgcc gaccaccgaa ctgttcccga tgtcgaaaat taacgacgcc
 961 atccagcatg tgcgcgacgg taaggcgcgt taccgcgtgg tgttgaaagc cgatttttga
```

Amino acid sequence (SEQ ID NO:94)

```
   1 MSMIKSYAAK EAGGELEVYE YDPGELRPQD VEVQVDYCGI CHSDLSMIDN EWGFSQYPLV
  61 AGHEVIGRVV ALGSAAQDKG LQVGQRVGIG WTARSCGHCD ACISGNQINC EQGAVPTIMN
 121 RGGFAEKLRA DWQWVIPLPE NIDIESAGPL LCGGITVFKP LLMHHITATS RVGVIGIGGL
 181 GHIAIKLLHA MGCEVTAFSS NPAKEQEVLA MGADKVVNSR DPQALKALAG QFDLIINTVN
 241 VSLDWQPYFE ALTYGGNFHT VGAVLTPLSV PAFTLIAGDR SVSGSATGTP YELRKLMRFA
 301 ARSKVAPTTE LFPMSKINDA IQHVRDGKAR YRVVLKADF
```

yahK (NP_414859)

Nucleotide sequence (SEQ ID NO:95)

```
   1 atgaagatca aagctgttgg tgcatattcc gctaaacaac cacttgaacc gatggatatc
  61 acccggcgtg aaccgggacc gaatgatgtc aaaatcgaaa tcgcttactg tggcgtttgc
 121 cattccgatc tccaccaggt ccgttccgag tgggcgggga cggtttaccc ctgcgtgccg
 181 ggtcatgaaa ttgtggggcg tgtggtagcc gttggtgatc aggtagaaaa atatgcgccg
 241 ggcgatctgg tcggtgtcgg ctgcattgtc gacagttgta aacattgcga agagtgtgaa
 301 gacgggttgg aaaactactg tgatcacatg accggcacct ataactcgcc gacgccggac
 361 gaaccgggcc atactctggg cggctactca caacagatcg tcgttcatga gcgatatgtt
 421 ctgcgtattc gtcacccgca agagcagctg gcggcggtgg ctcctttgtt gtgtgcaggg
 481 atcaccacgt attcgccgct acgtcactgg caggccgggc cgggtaaaaa agtgggcgtg
```

FIG. 10 (Cont)

```
541 gtcggcatcg gcggtctggg acatatgggg attaagctgg cccacgcgat gggggcacat
601 gtggtggcat ttaccacttc tgaggcaaaa cgcgaagcgg caaaagccct gggggccgat
661 gaagttgtta actcacgcaa tgccgatgag atggcggctc atctgaagag tttcgatttc
721 attttgaata cagtagctgc gccacataat ctcgacgatt ttaccacctt gctgaagcgt
781 gatggcacca tgacgctggt tggtgcgcct gcgacaccgc ataaatcgcc ggaagttttc
841 aacctgatca tgaaacgccg tgcgatagcc ggttctatga ttggcggcat tccagaaact
901 caggagatgc tcgatttttg cgccgaacat ggcatcgtgg ctgatataga gatgattcgg
961 gccgatcaaa ttaatgaagc ctatgagcga atgctgcgcg gtgatgtgaa atatcgtttt
1021 gttatcgata atcgcacact aacagactga
```

Amino acid sequence (SEQ ID NO:96)

```
  1 MKIKAVGAYS AKQPLEPMDI TRREPGPNDV KIEIAYCGVC HSDLHQVRSE WAGTVYPCVP
 61 GHEIVGRVVA VGDQVEKYAP GDLVGVGCIV DSCKHCEECE DGLENYCDHM TGTYNSPTPD
121 EPGHTLGGYS QQIVVHERYV LRIRHPQEQL AAVAPLLCAG ITTYSPLRHW QAGPGKKVGV
181 VGIGGLGHMG IKLAHAMGAH VVAFTTSEAK REAAKALGAD EVVNSRNADE MAAHLKSFDF
241 ILNTVAAPHN LDDFTTLLKR DGTMTLVGAP ATPHKSPEVF NLIMKRRAIA GSMIGGIPET
301 QEMLDFCAEH GIVADIEMIR ADQINEAYER MLRGDVKYRF VIDNRTLTD
```

adhP (NP_415995)

Nucleotide sequence (SEQ ID NO:97)

```
  1 atgaaggctg cagttgttac gaaggatcat catgttgacg ttacgtataa aacactgcgc
 61 tcactgaaac atggcgaagc cctgctgaaa atggagtgtt gtggtgtatg tcataccgat
121 cttcatgtta agaatggcga ttttggtgac aaaaccggcg taattctggg ccatgaaggc
181 atcggtgtgg tggcagaagt gggtccaggt gtcacctcat taaaaccagg cgatcgtgcc
241 agcgtggcgt ggttctacga aggatgcggt cattgcgaat actgtaacag tggtaacgaa
301 acgctctgcc gttcagttaa aaatgccgga tacagcgttg atggcgggat ggcggaagag
361 tgcatcgtgg tcgccgatta cgcggtaaaa gtgccagatg gtctggactc ggcggcggcc
421 agcagcatta cctgtgcggg agtcaccacc tacaaagccg ttaagctgtc aaaaattcgt
481 ccagggcagt ggattgctat ctacggtctt ggcggtctgg gtaacctcgc cctgcaatac
541 gcgaagaatg tctttaacgc caaagtgatc gccattgatg tcaatgatga gcagttaaaa
601 ctggcaaccg aaatgggcgc agatttagcg attaactcac acaccgaaga cgccgccaaa
661 attgtgcagg agaaaactgg tggcgctcac gctgcggtgg taacagcggt agctaaagct
721 gcgtttaact cggcagttga tgctgtccgt gcaggcggtc gtgttgtggc tgtcggtcta
781 ccgccggagt ctatgagcct ggatatccca cgtcttgtgc tggatggtat tgaagtggtc
841 ggttcgctgg tcggcacgcg ccaggattta actgaagcct tccagtttgc cgccgaaggt
901 aaagtggtgc cgaaagtcgc cctgcgtccg ttagcggaca tcaacaccat ctttactgag
961 atggaagaag gcaaaatccg tggccgcatg gtgattgatt tccgtcacta a
```

FIG. 10 (Cont)

Amino acid sequence (SEQ ID NO:98)

```
  1 MKAAVVTKDH HVDVTYKTLR SLKHGEALLK MECCGVCHTD LHVKNGDFGD KTGVILGHEG
 61 IGVVAEVGPG VTSLKPGDRA SVAWFYEGCG HCEYCNSGNE TLCRSVKNAG YSVDGGMAEE
121 CIVVADYAVK VPDGLDSAAA SSITCAGVTT YKAVKLSKIR PGQWIAIYGL GGLGNLALQY
181 AKNVFNAKVI AIDVNDEQLK LATEMGADLA INSHTEDAAK IVQEKTGGAH AAVVTAVAKA
241 AFNSAVDAVR AGGRVVAVGL PPESMSLDIP RLVLDGIEVV GSLVGTRQDL TEAFQFAAEG
301 KVVPKVALRP LADINTIFTE MEEGKIRGRM VIDFRH
```

ydjL (NP_416290)

Nucleotide sequence (SEQ ID NO:99)

```
   1 atgaaagcac tggctcggtt tggcaaggcc tttggcggct acaagatgat tgatgtccca
  61 caacccatgt gtggcccgga agatgtagtg attgaaatta aagccgcggc aatctgcggc
 121 gcagacatga agcactacaa tgtcgatagc ggttctgatg agtttaactc tatccgcggc
 181 catgagttcg caggttgtat tgcgcaggtt ggtgaaaaag tcaaagactg gaaagtgggg
 241 caacgcgtcg tatcggataa cagcggtcac gtttgcggtg tttgtccggc ctgtgaacaa
 301 ggtgattttc tgtgttgtac agaaaaagta aaccttggtc tggataataa tacctggggc
 361 ggtggttttt ccaaatattg tctggttcct ggtgaaattc tcaaaattca tcgtcatgcg
 421 ttgtgggaaa tccctgatgg tgttgattat gaggacgcag ccgtacttga ccctatctgt
 481 aatgcctaca aatccatcgc gcagcaatcg aaattccttc ctggtcagga tgtggtcgtc
 541 atcggcactg gcccactcgg gctgttctcc gtacaaatgg cgcgaattat gggggcggta
 601 aatatcgtcg tcgttggtct gcaagaagat gtggcggtcc gcttcccggt tgcaaaagaa
 661 ctgggtgcga cggcagtagt aaatggttct accgaagatg tggtggcgcg ctgccagcaa
 721 atttgtggca aagacaatct gggactggtg attgaatgct ccggtgccaa tatcgcactg
 781 aaacaagcca tcgaaatgct ccgcccgaac ggggaagtgg tacgcgttgg aatgggcttc
 841 aaacctcttg atttctcgat taatgacatt accgcctgga acaaaaagcat cattgggcat
 901 atggcctatg actccacctc atggcgtaac gctatcaggc tattagccag cggcgctatc
 961 aaagtcaaac cgatgatcac gcatcgtatc ggcctgtcgc aatggcgcga agggtttgat
1021 gcgatggtcg ataaaaccgc aatcaaagtg atcatgactt acgactttga tgaataa
```

Amino acid sequence (SEQ ID NO:100)

```
  1 MKALARFGKA FGGYKMIDVP QPMCGPEDVV IEIKAAAICG ADMKHYNVDS GSDEFNSIRG
 61 HEFAGCIAQV GEKVKDWKVG QRVVSDNSGH VCGVCPACEQ GDFLCCTEKV NLGLDNNTWG
121 GGFSKYCLVP GEILKIHRHA LWEIPDGVDY EDAAVLDPIC NAYKSIAQQS KFLPGQDVVV
181 IGTGPLGLFS VQMARIMGAV NIVVVGLQED VAVRFPVAKE LGATAVVNGS TEDVVARCQQ
241 ICGKDNLGLV IECSGANIAL KQAIEMLRPN GEVVRVGMGF KPLDFSINDI TAWNKSIIGH
301 MAYDSTSWRN AIRLLASGAI KVKPMITHRI GLSQWREGFD AMVDKTAIKV IMTYDFDE
```

FIG. 10 (Cont)

ydjJ (NP_416288)

Nucleotide sequence (SEQ ID NO:101)

```
   1 atgaaaaatt caaaagcaat attgcaggtg ccgggcacaa tgaaaattat ttcagcagaa
  61 ataccagtgc ctaaagaaga tgaagttttg attaaagtag aatatgtcgg tatttgtggt
 121 tcagatgtac atggttttga atcaggcccg tttattccgc ctaaagaccc aaatcaagaa
 181 attggcctgg gtcatgaatg cgccgggacg gttgtggctg tgggaagccg cgtgcgcaaa
 241 tttaaaccgg gggatcgggt aaatatcgaa cctggcgttc cttgcggtca ctgtcgttac
 301 tgtctggaag gcaaatataa catctgcccg gacgttgatt ttatggcgac acaacccaac
 361 taccgcggcg cattaacgca ctatctgtgt catccggaga gctttactta caaactgccc
 421 gacaatatgg acacgatgga aggggcgctg gtggagcctg ccgcagtcgg gatgcatgcc
 481 gcgatgctgg cagatgttaa accgggtaag aagataatta ttctgggagc aggttgtatt
 541 ggtttgatga cgttgcaagc gtgcaaatgc ctgggagcaa cggaaattgc cgtcgttgat
 601 gtgctggaaa aacgtctggc aatggcggaa cagcttggtg cgacagtggt tattaacggc
 661 gcaaaagaag acactattgc acgctgtcag caatttaccg aagacatggg cgcagatatt
 721 gttttcgaaa cagcgggttc tgcggtcacc gttaaacagg caccttatct ggtaatgcgc
 781 ggcggtaaaa ttatgattgt tggtactgta cccggcgatt cggcaatcaa tttcctcaaa
 841 atcaatcgcg aagtcactat ccagacggta ttccgctatg ccaatcgtta tccggtcacg
 901 attgaagcta tttcttcagg gcgattcgat gtgaaatcga tggtgacgca tatttacgat
 961 tatcgggatg tacaacaggc atttgaagag tcagttaaca acaaacgcga cattattaaa
1021 ggcgttatta aaattagcga ttaa
```

Amino acid sequence (SEQ ID NO:102)

```
   1 MKNSKAILQV PGTMKIISAE IPVPKEDEVL IKVEYVGICG SDVHGFESGP FIPPKDPNQE
  61 IGLGHECAGT VVAVGSRVRK FKPGDRVNIE PGVPCGHCRY CLEGKYNICP DVDFMATQPN
 121 YRGALTHYLC HPESFTYKLP DNMDTMEGAL VEPAAVGMHA AMLADVKPGK KIIILGAGCI
 181 GLMTLQACKC LGATEIAVVD VLEKRLAMAE QLGATVVING AKEDTIARCQ QFTEDMGADI
 241 VFETAGSAVT VKQAPYLVMR GGKIMIVGTV PGDSAINFLK INREVTIQTV FRYANRYPVT
 301 IEAISSGRFD VKSMVTHIYD YRDVQQAFEE SVNNKRDIIK GVIKISD
```

idnD (NP_418688)

Nucleotide sequence (SEQ ID NO:103)

```
   1 atgcaagtga aaacacagtc ctgcgttgtt gcgggcaaga aaactgttgc cgttaccgag
  61 cagacgatag attggaataa taatggaaca ttagtacaaa taacccgagg tggaatttgc
 121 ggttccgatt tacattatta tcaggaagga aaagtaggta atttcatgat aaaggcaccg
 181 atggtgttag gtcatgaagt tatcggtaaa gttattcata gcgactcatc agaattacat
 241 gaagggcaaa cggtagccat taatccgtct aaaccgtgcg gtcactgcaa atactgcatt
 301 gaacataacg agaatcagtg tacagatatg cgttttttg gcagtgccat gtatttccct
 361 catgttgatg gtggttttac ccgttataaa atggtcgaaa cgtcgcaatg tgtcccttat
 421 ccggccaaag ctgacgaaaa ggttatggct tttgccgaac ctttagccgt cgcgattcat
```

FIG. 10 (Cont)

```
481 gccgcacatc aggccggcga gttacagggc aagcgagtat ttatttccgg tgttggaccc
541 attggctgcc tgattgtcag tgcagtgaaa acactggggg ccgcggaaat tgtctgtgct
601 gatgtgagtc cccgttccct ttcgctgggc aaagagatgg gggcggatgt gctcgtaaac
661 ccacaaaacg acgacatgga tcactggaaa gcggaaaaag gctatttcga tgtcagcttt
721 gaagtgtccg gtcatccttc atcagtgaat acctgtctgg aggtcactcg tgcacgcggc
781 gtaatggtgc aggtaggtat gggaggcgcg atggcagaat tcccaatgat gacgttgatt
841 ggtaaggaga tttcactcag aggctctttc cgttttacca gcgaatttaa taccgcagtg
901 tcatggctgg cgaatggcgt tatcaatcca ctgcctttac tgagtgctga atatcccttc
961 actgacctgg aagaggcgct acgtttcgcc ggtgataaaa cccaggcagc aaaagtccag
1021 cttgttttct aa
```

Amino acid sequence (SEQ ID NO:104)

```
   1 MQVKTQSCVV AGKKTVAVTE QTIDWNNNGT LVQITRGGIC GSDLHYYQEG KVGNFMIKAP
  61 MVLGHEVIGK VIHSDSSELH EGQTVAINPS KPCGHCKYCI EHNENQCTDM RFFGSAMYFP
 121 HVDGGFTRYK MVETSQCVPY PAKADEKVMA FAEPLAVAIH AAHQAGELQG KRVFISGVGP
 181 IGCLIVSAVK TLGAAEIVCA DVSPRSLSLG KEMGADVLVN PQNDDMDHWK AEKGYFDVSF
 241 EVSGHPSSVN TCLEVTRARG VMVQVGMGGA MAEFPMMTLI GKEISLRGSF RFTSEFNTAV
 301 SWLANGVINP LPLLSAEYPF TDLEEALRFA GDKTQAAKVQ LVF
```

tdh (NP_418073)

Nucleotide sequence (SEQ ID NO:105)

```
   1 atgaaagcgt tatccaaact gaaagcggaa gagggcatct ggatgaccga cgttcctgta
  61 ccggaactcg ggcataacga tctgctgatt aaaatccgta aaacagccat ctgcgggact
 121 gacgttcaca tctataactg ggatgagtgg tcgcaaaaaa ccatcccggt gccgatggtc
 181 gtgggccatg aatatgtcgg tgaagtggta ggtattggtc aggaagtgaa aggcttcaag
 241 atcggcgatc gcgtttctgg cgaaggccat atcacctgtg gtcattgccg caactgtcgt
 301 ggtggtcgta cccatttgtg ccgcaacacg ataggcgttg gtgttaatcg cccgggctgc
 361 tttgccgaat atctggtgat cccggcattc aacgccttca aaatccccga caatatttcc
 421 gatgacttag ccgcaatttt tgatcccttc ggtaacgccg tgcataccgc gctgtcgttt
 481 gatctggtgg gcgaagatgt gctggtttct ggtgcaggcc cgattggtat tatggcagcg
 541 gcggtggcga aacacgttgg tgcacgcaat gtggtgatca ctgatgttaa cgaataccgc
 601 cttgagctgg cgcgtaaaat gggtatcacc cgtgcggtta acgtcgccaa agaaaatctc
 661 aatgacgtga tggcggagtt aggcatgacc gaaggttttg atgtcggtct ggaaatgtcc
 721 ggtgcgccgc cagcgtttcg taccatgctt gacaccatga atcacggcgg ccgtattgcg
 781 atgctgggta ttccgccgtc tgatatgtct atcgactgga ccaaagtgat ctttaaaggc
 841 ttgttcatta aaggtattta cggtcgtgag atgtttgaaa cctggtacaa gatggcggcg
 901 ctgattcagt ctggcctcga tctttcgccg atcattaccc atcgtttctc tatcgatgat
 961 ttccagaagg gctttgacgc tatgcgttcg ggccagtccg ggaaagttat tctgagctgg
1021 gattaa
```

FIG. 10 (Cont)

Amino acid sequence (SEQ ID NO:106)

```
  1 MKALSKLKAE EGIWMTDVPV PELGHNDLLI KIRKTAICGT DVHIYNWDEW SQKTIPVPMV
 61 VGHEYVGEVV GIGQEVKGFK IGDRVSGEGH ITCGHCRNCR GGRTHLCRNT IGVGVNRPGC
121 FAEYLVIPAF NAFKIPDNIS DDLAAIFDPF GNAVHTALSF DLVGEDVLVS GAGPIGIMAA
181 AVAKHVGARN VVITDVNEYR LELARKMGIT RAVNVAKENL NDVMAELGMT EGFDVGLEMS
241 GAPPAFRTML DTMNHGGRIA MLGIPPSDMS IDWTKVIFKG LFIKGIYGRE MFETWYKMAA
301 LIQSGLDLSP IITHRFSIDD FQKGFDAMRS GQSGKVILSW D
```

yjjN (NP_418778)

Nucleotide sequence (SEQ ID NO:107)

```
   1 atgtctacga tgaatgtttt aatttgccag cagccgaaag aattagtctg gaaacaacgc
  61 gagataccta ttccgggtga caatgaagca ttaataaaaa ttaagtctgt cgggatttgc
 121 ggtaccgata ttcatgcctg gggtggaaat caaccatttt ttagttatcc acgtgtttta
 181 ggccatgaaa tatgtgggga gattgttggg ctgggtaaaa atattgctga tcttaaaaat
 241 ggtcagcaag ttgctgtgat cccttatgtt gcctgtcagc aatgcccggc gtgtaaaagc
 301 gggcgtacca attgctgtga aaaaatttca gtcattggcg tgcatcagga tggcggtttt
 361 agtgagtatt tgagcgtgcc ggtggcgaac attttgcccg cagacggtat tgacccgcag
 421 gcggcagcat tgattgaacc tttcgctatt agcgctcatg cggtgcgtcg cgcagccatt
 481 gctcccggcg agcaggtgct ggtggtcggg gcggggccaa tcggtctggg cgcggcggca
 541 atcgctaaag ccgatggcgc acaggtggtg gtggcggata ccagtccggc gcgccgtgaa
 601 catgtggcaa cgcgtctgga attaccttta ctggacccgt cagccgaaga ttttgacgcg
 661 cagctacggg cgcagtttgg tggttcgctg gcgcagaaag tgatcgacgc gacaggtaat
 721 caacatgcga tgaataacac cgtgaatttg attcgtcacg gcggcacggt ggtatttgtc
 781 ggcctgttta aaggtgagtt gcagttctcc gatccggaat tccataaaaa agaaacgacg
 841 atgatgggca gccgcaacgc cacgccggaa gattttgcta aagtcggtcg actgatggcg
 901 gaaggaaaaa tcactgctga catgatgtta acccatcgct atccgttcgc cacgctggca
 961 gaaacctacg agcgcgatgt gattaacaat cgtgagttaa ttaaaggcgt aattactttc
1021 tga
```

Amino acid sequence (SEQ ID NO:108)

```
  1 MSTMNVLICQ QPKELVWKQR EIPIPGDNEA LIKIKSVGIC GTDIHAWGGN QPFFSYPRVL
 61 GHEICGEIVG LGKNIADLKN GQQVAVIPYV ACQQCPACKS GRTNCCEKIS VIGVHQDGGF
121 SEYLSVPVAN ILPADGIDPQ AAALIEPFAI SAHAVRRAAI APGEQVLVVG AGPIGLGAAA
181 IAKADGAQVV VADTSPARRE HVATRLELPL LDPSAEDFDA QLRAQFGGSL AQKVIDATGN
241 QHAMNNTVNL IRHGGTVVFV GLFKGELQFS DPEFHKKETT MMGSRNATPE DFAKVGRLMA
301 EGKITADMML THRYPFATLA ETYERDVINN RELIKGVITF
```

FIG. 10 (Cont)

rspB (NP_416097)

Nucleotide sequence (SEQ ID NO:109)

```
  1 atgaaaagca tattaattga aaaaccgaat caactggcga ttgtcgaacg tgaaataccc
 61 accccgtcag cgggtgaagt acgagtaaaa gtgaaacttg ccggaatttg tggttcagat
121 agccatattt atcgtgggca taatcctttt gcgaaatatc cgcgcgtcat tggtcatgaa
181 ttctttggcg tcattgatgc agtgggtgaa ggcgtggaaa gcgccagagt cggtgaacgt
241 gttgctgtcg atccggtggt cagctgtggg cattgctatc cgtgctctat aggtaaaccg
301 aacgtttgta cgacactggc tgtattaggt gtgcacgctg acggtggttt cagtgaatat
361 gccgtggttc cggcaaaaaa tgcgtggaaa attcctgaag cagtggccga tcaatatgcg
421 gtaatgatcg aacctttac cattgcggct aacgtaaccg gacatggtca accgactgaa
481 aatgataccg ttctggttta tggtgccggt ccaatcggcc tgacgatcgt tcaggtatta
541 aaaggcgtct ataacgttaa aaatgtgatt gttgccgatc gcattgatga acgactggaa
601 aaagcgaaag agagcggggc tgactgggcg attaataaca gccagacacc gcttggcgag
661 attttcactg aaaaaggcat caagccgaca ttaattatcg atgcggcttg tcatccttct
721 atcctgaaag aggccgtaac gctggcttct ccagcggcac gtattgtatt gatggggttc
781 tccagtgaac cgtctgaagt gattcagcaa ggaattaccg gaaaagaact ctctattttc
841 tcttcacgct taaatgcaaa taaattcccg atcgttatcg actggttaag taaagggtta
901 attaaaccag aaaaattaat tacccatacg tttgatttcc agcatgttgc tgatgccatt
961 agtttatttg aacaggatca aaaacattgc tgcaaagtct tactcacttt ttctgaataa
```

Amino acid sequence (SEQ ID NO:110)

```
  1 MKSILIEKPN QLAIVEREIP TPSAGEVRVK VKLAGICGSD SHIYRGHNPF AKYPRVIGHE
 61 FFGVIDAVGE GVESARVGER VAVDPVVSCG HCYPCSIGKP NVCTTLAVLG VHADGGFSEY
121 AVVPAKNAWK IPEAVADQYA VMIEPFTIAA NVTGHGQPTE NDTVLVYGAG PIGLTIVQVL
181 KGVYNVKNVI VADRIDERLE KAKESGADWA INNSQTPLGE IFTEKGIKPT LIIDAACHPS
241 ILKEAVTLAS PAARIVLMGF SSEPSEVIQQ GITGKELSIF SSRLNANKFP IVIDWLSKGL
301 IKPEKLITHT FDFQHVADAI SLFEQDQKHC CKVLLTFSE
```

gatD (NP_416594)

Nucleotide sequence (SEQ ID NO:111)

```
  1 atgaaatcag tggtgaatga tactgatggt atcgtgcgcg ttgcagaaag cgtcattcct
 61 gaaattaaac atcaggatga ggtgcgggta aaaattgcca gctcgggctt atgtggttcc
121 gatttaccca ggatatttaa aaatggtgca cattattatc caataacgtt aggccatgaa
181 tttagcggct atattgatgc ggtgggatcc ggtgttgatg atttacaccc tggcgatgcg
241 gttgcctgtg tgccgttatt accctgtttt acttgtccag agtgtctgaa agggttttat
301 tcccagtgcg caaaatatga ttttattggc tcgcggcgtg atggtggatt tgctgaatat
361 attgtcgtta agcgaaaaaa tgtctttgct ctacccacgg atatgcctat tgaggatggg
421 gcttttattg agccgattac cgttggtctg catgcttttc atttagcgca aggttgtgag
481 aataaaaacg ttattattat tggtgccgga accattggcc tgctggccat tcagtgcgct
541 gtcgcgctgg gagcaaagag tgtgacggcg atcgacatta gttcagaaaa actggcactg
```

FIG. 10 (Cont)

```
601 gcaaaatctt tcggtgcgat gcaaacattt aacagtagcg aaatgagcgc gccgcaaatg
661 cagagcgttt tacgcgaact gcgctttaat cagcttatcc tcgagacggc tggcgtaccg
721 caaactgtcg aactggcggt agagattgcc ggtcctcatg cccaactggc gctggtgggc
781 acgttgcatc aggatctgca tttaacatcg gcaacgtttg gcaaaatatt gcgtaaagag
841 ctgacggtta tcggcagttg gatgaactac tccagccctt ggccggggca ggagtgggaa
901 acggcgagcc ggttgctgac agaacgtaag ttaagcctgg agccattaat cgctcaccgt
961 ggaagctttg aaagcttcgc ccaggcggtg cgtgacatcg ctcgtaatgc tatgccgggc
1021 aaagtgttgc tcattccctg a
```

Amino acid sequence (SEQ ID NO:112)

```
   1 MKSVVNDTDG IVRVAESVIP EIKHQDEVRV KIASSGLCGS DLPRIFKNGA HYYPITLGHE
  61 FSGYIDAVGS GVDDLHPGDA VACVPLLPCF TCPECLKGFY SQCAKYDFIG SRRDGGFAEY
 121 IVVKRKNVFA LPTDMPIEDG AFIEPITVGL HAFHLAQGCE NKNVIIIGAG TIGLLAIQCA
 181 VALGAKSVTA IDISSEKLAL AKSFGAMQTF NSSEMSAPQM QSVLRELRFN QLILETAGVP
 241 QTVELAVEIA GPHAQLALVG TLHQDLHLTS ATFGKILRKE LTVIGSWMNY SSPWPGQEWE
 301 TASRLLTERK LSLEPLIAHR GSFESFAQAV RDIARNAMPG KVLLIP
```

<u>yphC (NP_417040)</u>

Nucleotide sequence (SEQ ID NO:113)

```
   1 atgaaaacga tgctggcagc ttatttacca ggaaattcga ccgtcgatct gcgggaagtt
  61 gcggtgccga cgccggggat taaccaggta ctgatcaaaa tgaaatcctc cgggatttgc
 121 ggaagcgatg tccactatat ctatcatcaa caccgtgcca cagcggcggc acccgataaa
 181 ccgttatacc agggctttat caacggtcat gaaccgtgcg ggcagattgt ggcgatgggg
 241 caaggctgcc gccattttaa agagggcgac cgcgtgctgg tgtatcacat ttctggctgt
 301 ggtttttgcc cgaactgccg tcgcggtttt cctatttctt gtactggcga aggaaaagcg
 361 gcttacggct ggcagcgtga cggcggtcat gccgaatact tgctggcgga agaaaaagat
 421 ctgatcctcc tgccggatgc gctgagctac gaagatggtg cgtttatcag ttgcggcgtt
 481 ggtacagcgt atgaaggaat tttgcgcggc gaagtttccg gcagtgataa cgtgctggtg
 541 gtcggtctgg ggccagtcgg catgatggcg atgatgctgg cgaaaggtcg cggtgcaaaa
 601 cggatcatcg gcgttgatat gctgccggaa cgtctggcga tggcaaaaca gttaggggtg
 661 atggatcacg gctatttagc aaccaccgaa ggtctgccgc agattatcgc cgaactcacc
 721 cacggtggcg cggatgttgc gctcgattgt tccggtaatg ccgcaggtcg cttgctggca
 781 ctgcaatcca ccgctgactg gggacgggtg gtttacattg gtgaaaccgg aaaagtggaa
 841 ttcgaggtca gcgccgatct gatgcaccat caacggcgga ttatcggctc ctgggtgacc
 901 agtctgttcc atatggaaaa atgcgcccat gatctgacgg actggaagct gtggccgcgt
 961 aacgccatta cccatcgctt ctcgctggaa caggcaggtg atgcctatgc gctgatggcg
1021 agcggcaaat gcgggaaagt tgtgattaac ttcccggatt aa
```

FIG. 10 (Cont)

Amino acid sequence (SEQ ID NO:114)

```
  1 MKTMLAAYLP GNSTVDLREV AVPTPGINQV LIKMKSSGIC GSDVHYIYHQ HRATAAAPDK
 61 PLYQGFINGH EPCGQIVAMG QGCRHFKEGD RVLVYHISGC GFCPNCRRGF PISCTGEGKA
121 AYGWQRDGGH AEYLLAEEKD LILLPDALSY EDGAFISCGV GTAYEGILRG EVSGSDNVLV
181 VGLGPVGMMA MMLAKGRGAK RIIGVDMLPE RLAMAKQLGV MDHGYLATTE GLPQIIAELT
241 HGGADVALDC SGNAAGRLLA LQSTADWGRV VYIGETGKVE FEVSADLMHH QRRIIGSWVT
301 SLFHMEKCAH DLTDWKLWPR NAITHRFSLE QAGDAYALMA SGKCGKVVIN FPD
```

yhdH (NP_417719)

Nucleotide sequence (SEQ ID NO:115)

```
  1 atgcaggcgt tacttttaga acagcaggac ggcaaaactc tcgcatcagt acagactctg
 61 gacgaaagtc gcctgccgga gggcgatgtc acggtcgatg ttcactggtc gagcctgaac
121 tataaagatg cgctggcgat taccggtaag ggaaaaatca tccgtaattt tccgatgatt
181 cctgggatcg attttgccgg aactgtacgc accagcgaag atccgcgttt tcatgccggt
241 caggaggtgt tactcactgg ctggggcgtt ggtgaaaacc actggggtgg gctggcggag
301 caggcgcgag tgaaaggtga ctggctggtt gccatgccgc aagggctgga cgcgcgtaaa
361 gcaatgatta tcggtactgc cggttttacc gccatgctgt gtgtgatggc gctggaagat
421 gccggtgttc gcccgcagga cggggagatt gtcgtgacgg gtgccagtgg tggcgtcggc
481 agtaccgccg tggcgctgct gcataagttg ggttatcagg tcgttgccgt ttccggtcgc
541 gaaagtaccc atgaatatct gaaaagttta ggtgctagcc gtgttctccc tcgtgatgag
601 tttgccgaat cccgtcctct ggaaaaacaa gtctgggctg gggcaattga caccgttggc
661 gacaaagtgc tggcaaaagt gctggcgcaa atgaattacg gcggctgcgt ggcggcctgt
721 ggtctggcgg gtggttttac tctgccaacc acggtcatgc catttattct gcgtaatgtc
781 cgtttgcaag gggtggattc agtaatgacg ccaccagaac gccgcgcaca agcctggcag
841 cgactggtcg ccgatttacc ggaatcattc tatacccagg cggcaaaaga gatatctctg
901 tcagaggcac cgaactttgc cgaggccatc attaataacc agatccaggg tcgcacgctg
961 gtgaaggtta actaa
```

Amino acid sequence (SEQ ID NO:116)

```
  1 MQALLLEQQD GKTLASVQTL DESRLPEGDV TVDVHWSSLN YKDALAITGK GKIIRNFPMI
 61 PGIDFAGTVR TSEDPRFHAG QEVLLTGWGV GENHWGGLAE QARVKGDWLV AMPQGLDARK
121 AMIIGTAGFT AMLCVMALED AGVRPQDGEI VVTGASGGVG STAVALLHKL GYQVVAVSGR
181 ESTHEYLKSL GASRVLPRDE FAESRPLEKQ VWAGAIDTVG DKVLAKVLAQ MNYGGCVAAC
241 GLAGGFTLPT TVMPFILRNV RLQGVDSVMT PPERRAQAWQ RLVADLPESF YTQAAKEISL
301 SEAPNFAEAI INNQIQGRTL VKVN
```

ycjQ (NP_415829)

Nucleotide sequence (SEQ ID NO:117)

```
  1 atgaaaaagt tagtagccac agcaccgcgt gttgctgcgc tggttgagta tgaagatcgg
```

FIG. 10 (Cont)

```
  61 gcgattttag ctaatgaagt gaagatccgc gtgcgtttcg gcgcaccgaa acacggaacg
 121 gaagtggtcg acttccgcgc cgccagcccg tttattgatg aagactttaa cggcgaatgg
 181 cagatgttca ctccgcgtcc ggcagatgcg ccgcgcggca ttgagtttgg caaattccag
 241 cttggcaaca tggtggttgg cgacattatc gagtgcggca gcgacgttac cgactacgcg
 301 gtgggcgaca gcgtatgcgg ctacggcccg ctctccgaga cggtcatcat taacgcagtg
 361 aataactaca agctgcgcaa aatgccgcaa ggcagctcct ggaaaaacgc cgtctgctac
 421 gacccggcgc agtttgccat gagcggcgtg cgcgatgcca acgtgcgcgt aggggatttt
 481 gtggtggtgg tagggcttgg cgcaatcggt caaattgcca tccaactggc taaacgcgct
 541 ggcgcttctg tggtgattgg cgtcgatcct atcgcccatc gctgtgatat tgcccgtcgc
 601 cacggcgcgg atttctgcct taaccccatc ggcactgatg taggtaaaga gatcaaaacg
 661 ctgaccggca agcagggtgc cgatgtgatt atcgaaacca gcggctacgc cgacgcgctg
 721 caatcggcgc tccgcggtct ggcttatggc ggcaccatct cctatgtcgc gtttgccaag
 781 ccgtttgccg aaggttttaa cctcggacgc gaagcgcatt tcaataacgc caaaattgtc
 841 ttctctcgcg cgtgcagcga accgaacccg gattatccgc gctggagccg caagcgtatt
 901 gaagaaacct gttgggaact gctgatgaac ggttatctca attgcgaaga tttaatcgac
 961 ccggtggtga cctttgccaa tagcccggaa agctatatgc agtatgtcga ccagcatccg
1021 gaacagagca tcaaaatggg cgttacgttt taa
```

Amino acid sequence (SEQ ID NO:118)

```
   1 MKKLVATAPR VAALVEYEDR AILANEVKIR VRFGAPKHGT EVVDFRAASP FIDEDFNGEW
  61 QMFTPRPADA PRGIEFGKFQ LGNMVVGDII ECGSDVTDYA VGDSVCGYGP LSETVIINAV
 121 NNYKLRKMPQ GSSWKNAVCY DPAQFAMSGV RDANVRVGDF VVVVGLGAIG QIAIQLAKRA
 181 GASVVIGVDP IAHRCDIARR HGADFCLNPI GTDVGKEIKT LTGKQGADVI IETSGYADAL
 241 QSALRGLAYG GTISYVAFAK PFAEGFNLGR EAHFNNAKIV FSRACSEPNP DYPRWSRKRI
 301 EETCWELLMN GYLNCEDLID PVVTFANSPE SYMQYVDQHP EQSIKMGVTF
```

yncB (NP_415966)

Nucleotide sequence (SEQ ID NO:119)

```
   1 atggggcaac aaaagcagcg taatcgacgt tgggttctgg cctcgcgtcc acatggcgca
  61 cctgttccgg agaatttccg tcttgaagaa gatgatgtcg ccacaccggg tgaaggacag
 121 gtgttactgc gcacagttta tttgtccctg gacccgtata tgcgtggacg tatgagcgat
 181 gagccatctt attcaccgcc tgttgatatt ggcggcgtga tggtcggcgg tacggtgagc
 241 cgtgtcgtgg agtcgaatca tcctgattat cagtctggcg actgggtgct gggctacagt
 301 ggatggcaag actatgacat atccagtggt gatgatctgg tgaaacttgg cgatcatccg
 361 caaaatccat cgtggtcgct gggtgtgcta gggatgccag gctttaccgc ttatatgggc
 421 ctactggata tcggtcagcc taaagagggc gaaacgttgg tggtagctgc ggcgacagga
 481 ccagtggggg cgacggtggg gcaaatcggc aaacttaaag gttgcagagt ggtgggggta
 541 gccggtggcg cggaaaaatg ccgccatgct accgaggtgt taggcttcga tgtttgtctt
 601 gatcaccacg cggatgattt tgccgaacaa ctggcgaaag cgtgcccaaa aggtattgat
 661 atctattatg aaaacgtggg cggtaaggta ttcgatgcgg tgctaccgtt acttaataca
 721 tctgcgcgca ttcccgtctg cggattagtg agcagctata acgctacaga gctaccaccc
```

FIG. 10 (Cont)

```
 781 ggtccggatc gtttacctct gttgatggct acagtgctga aaaaacgtat tcgcttgcaa
 841 ggttttatta tcgctcagga ttatggtcac cgcatccatg agtttcagag ggagatgggg
 901 caatgggtga aagaggataa aatccactac cgcgaagaaa ttactgacgg tttagagaat
 961 gcgccacaga cgtttatcgg cctgctgaag ggtaaaaact tcggcaaagt ggtgatccgc
1021 gtggcgggtg atgattaa
```

Amino acid sequence (SEQ ID NO:120)

```
  1 MGQQKQRNRR WVLASRPHGA PVPENFRLEE DDVATPGEGQ VLLRTVYLSL DPYMRGRMSD
 61 EPSYSPPVDI GGVMVGGTVS RVVESNHPDY QSGDWVLGYS GWQDYDISSG DDLVKLGDHP
121 QNPSWSLGVL GMPGFTAYMG LLDIGQPKEG ETLVVAAATG PVGATVGQIG KLKGCRVVGV
181 AGGAEKCRHA TEVLGFDVCL DHHADDFAEQ LAKACPKGID IYYENVGGKV FDAVLPLLNT
241 SARIPVCGLV SSYNATELPP GPDRLPLLMA TVLKKRIRLQ GFIIAQDYGH RIHEFQREMG
301 QWVKEDKIHY REEITDGLEN APQTFIGLLK GKNFGKVVIR VAGDD
```

qor (NP_418475)

Nucleotide sequence (SEQ ID NO:121)

```
   1 atggcaacac gaattgaatt tcacaagcac ggtggcccgg aagtacttca agccgtagag
  61 ttcactcctg ccgatccggc ggagaatgaa atccaggtcg aaaataaagc catcggcatc
 121 aattttatcg acacatatat ccgcagcggc ctttacccgc cgccatcgct acccagcgga
 181 ttaggcaccg aagcagcagg catcgtgagt aaagtcggca gtggtgtaaa gcatattaag
 241 gcaggcgatc gtgtagtcta tgcgcagtcg gcgttaggcg cttacagctc tgtgcataac
 301 attattgcgg ataaagcggc gattctgcct gcggcaattt cttttgagca agctgcggca
 361 tccttcctga aaggcttaac ggtttattat ctgctgcgca aaacctatga aattaaaccc
 421 gatgaacagt tcctgttcca cgcagcggct ggcggcgttg gcttaattgc ctgccagtgg
 481 gcaaaagccc tgggcgcgaa acttatcggc accgtaggaa ccgcgcaaaa agcgcagagc
 541 gcgctaaaag cgggcgcgtg gcaggttatt aactatcgtg aagaggatct ggtcgagcgg
 601 ttaaaagaga tcaccggcgg taagaaagtg cgcgtggtgt acgattccgt gggcagagac
 661 acctgggaac ggtcgctgga ttgcctgcaa cgccgcggct taatggtcag ttttggcaac
 721 tcatcaggtg cggttaccgg tgtgaactta ggcattctca atcaaaaagg ctcgttgtat
 781 gtgacacgcc cttccctgca aggctatatc accacgcggg aggaattaac cgaggccagt
 841 aatgaactgt tctctttgat tgccagcggt gtgattaagg tcgatgtcgc cgagcagcag
 901 aaatatccgc tgaaggatgc gcagcgtgcg catgagattc tggaaagccg ggcgacgcaa
 961 ggttccagcc tgctgattcc ataa
```

Amino acid sequence (SEQ ID NO:122)

```
  1 MATRIEFHKH GGPEVLQAVE FTPADPAENE IQVENKAIGI NFIDTYIRSG LYPPPSLPSG
 61 LGTEAAGIVS KVGSGVKHIK AGDRVVYAQS ALGAYSSVHN IIADKAAILP AAISFEQAAA
121 SFLKGLTVYY LLRKTYEIKP DEQFLFHAAA GGVGLIACQW AKALGAKLIG TVGTAQKAQS
181 ALKAGAWQVI NYREEDLVER LKEITGGKKV RVVYDSVGRD TWERSLDCLQ RRGLMVSFGN
241 SSGAVTGVNL GILNQKGSLY VTRPSLQGYI TTREELTEAS NELFSLIASG VIKVDVAEQQ
301 KYPLKDAQRA HEILESRATQ GSSLLIP
```

FIG. 10 (Cont)

frmA (NP_414890)

Nucleotide sequence (SEQ ID NO:123)

```
   1 tcagtaacga attacggttc gaatggattt gccttcatgc atcaggtcga aggcgtcatt
  61 aatttcatcc aggctcatgg tatgcgtgac aaacggttcc agatcgatat caccttcat
 121 cgcatcttca accatgcccg gtaactggga acgacctttc acgccgccaa acgcggaacc
 181 tttccatacg cgaccagtga ccaactggaa tggacgggtg gagatttcct gaccggcaac
 241 cgcgaccccg atgatcaccg actgacccca gccgcggtgc gcactttcca gcgccgcacg
 301 catcacgttg acgttaccga tgcattcaaa ggtatggtcg ataccccatt tgttgatatc
 361 caacaggaca tcttttatcg gtttgtcgta gtcattcggg ttaatgcagt cggtagcacc
 421 gaagcgacgc gccagatcga atttcttcgg gttggtatcg atagcgataa tccgacccgc
 481 tttcgcctga cgcgcgccct gaaccactgc cagaccaatc gcgccaagac caaacacggc
 541 aacagaatca cctggctgga ctttagctgt gttgtgtacc gcgccaatac cggtggtcac
 601 gccacagccc agcaggcaga cgtgttcatg gtttgcttct ggattaattt ttgccagaga
 661 cacttccgcg actacggtgt attcactgaa tgtagagcac cccatgtagt gataaagcgg
 721 ctgcccgttg taagaaaaac gggtggtgcc gtctggcatc aggcctttac cctgggtttc
 781 gcgaaccgca acacagaggt tagttttgcc agaacgacag aactcacact cgccgcactc
 841 cgcggtgtaa agcgggatca catggtcgcc aggtttgacg ctggttacgc cttcaccgac
 901 ttcaaccaca acgccggccc cttcgtgacc gagaaccacc gggaatacac cttccgggtc
 961 atcgccggag agggtaaatg cgtcggtatg gcaaacgccg gtatgggtga ctttaattag
1021 cacttcacct tttttcggtg gtgcaacgtc aatttcaacg atttccagcg gtttaccggg
1081 agcaaatgca acggcagcac gtgatttcat
```

Amino acid sequence (SEQ ID NO:124)

```
  1 MKSRAAVAFA PGKPLEIVEI DVAPPKKGEV LIKVTHTGVC HTDAFTLSGD DPEGVFPVVL
 61 GHEGAGVVVE VGEGVTSVKP GDHVIPLYTA ECGECEFCRS GKTNLCVAVR ETQGKGLMPD
121 GTTRFSYNGQ PLYHYMGCST FSEYTVVAEV SLAKINPEAN HEHVCLLGCG VTTGIGAVHN
181 TAKVQPGDSV AVFGLGAIGL AVVQGARQAK AGRIIAIDTN PKKFDLARRF GATDCINPND
241 YDKPIKDVLL DINKWGIDHT FECIGNVNVM RAALESAHRG WGQSVIIGVA VAGQEISTRP
301 FQLVTGRVWK GSAFGGVKGR SQLPGMVEDA MKGDIDLEPF VTHTMSLDEI NDAFDLMHEG
361 KSIRTVIRY
```

ybdR (NP_415141)

Nucleotide sequence (SEQ ID NO:125)

```
   1 atgaaagcat tgacttatca cggcccacat cacgttcagg tagaaaatgt tcccgatccg
  61 ggcgttgaac aggcagatga tattattctg cgtattacgg caacggcgat ctgtggctct
 121 gacctccatc tttatcgagg caaaatacct caggttaaac atggcgatat ttttggtcat
 181 gaatttatgg gggaagtagt tgaaaccgga aaggacgtaa aaaatttgca aaaaggcgac
 241 cgagtggtaa ttccgttcgt cattgcttgt ggcgactgtt ttttctgtcg attgcaacaa
 301 tatgccgcct gcgaaaatac caatgcgggt aaaggcgctg cgctcaataa aaaacagata
```

FIG. 10 (Cont)

```
361 ccagctccag cggcattgtt tggttatagt cacctgtatg gcggcgttcc tggtgggcag
421 gcggaatatg tccgcgtccc taaagggaat gtggggccgt ttaaagtacc gcctttgctt
481 tcagatgata aagcgctttt cctttctgat attctgccaa cggcatggca ggcagcaaaa
541 aatgcgcaga tccaacaagg ttcaagcgtt gcagtctatg gtgctggtcc tgtgggattg
601 ttgacaatcg cctgtgcacg gttgctcggt gcggaacaga tttttgttgt tgatcatcat
661 ccctaccgct tgcatttcgc cgccgaccgc tacggcgcga tcccgattaa ttttgatgaa
721 gacagcgatc cggcacagtc aattattgaa caaacggcag gtcaccgggg cgtggatgca
781 gtaatagacg ccgtcggttt tgaagcgaaa ggcagcacca cggaaacggt gctgactaac
841 ctgaaactgg agggcagcag cggtaaagcg ttgcgtcagt gtattgcggc ggtcaggcgt
901 ggcggcattg ttagcgtacc gggcgtctac gctggattta ttcacggttt cctgtttggc
961 gacgcctttg ataaagggtt gtcgtttaaa atgggacaga cccacgttca cgcatggctg
1021 ggagaattat taccgttaat tgagaaagga ttactgaaac cagaagaaat tgttacccac
1081 tatatgccgt ttgaagaggc cgcccgggga tatgagattt tcgaaaaacg tgaagaggag
1141 tgccgtaagg tgattctggt acccggtgca caaagcgcag aggcggcgca gaaggcggtt
1201 tcaggtctgg tgaatgcgat gccgggggga acaatatga
```

Amino acid sequence (SEQ ID NO:126)

```
   1 MKALTYHGPH HVQVENVPDP GVEQADDIIL RITATAICGS DLHLYRGKIP QVKHGDIFGH
  61 EFMGEVVETG KDVKNLQKGD RVVIPFVIAC GDCFFCRLQQ YAACENTNAG KGAALNKKQI
 121 PAPAALFGYS HLYGGVPGGQ AEYVRVPKGN VGPFKVPPLL SDDKALFLSD ILPTAWQAAK
 181 NAQIQQGSSV AVYGAGPVGL LTIACARLLG AEQIFVVDHH PYRLHFAADR YGAIPINFDE
 241 DSDPAQSIIE QTAGHRGVDA VIDAVGFEAK GSTTETVLTN LKLEGSSGKA LRQCIAAVRR
 301 GGIVSVPGVY AGFIHGFLFG DAFDKGLSFK MGQTHVHAWL GELLPLIEKG LLKPEEIVTH
 361 YMPFEEAARG YEIFEKREEE CRKVILVPGA QSAEAAQKAV SGLVNAMPGG TI
```

YggP (YP_026187)

Nucleotide sequence (SEQ ID NO:127)

```
   1 atgaaaacca aagttgctgc tatttatggc aagcgggatg tccgtctgcg cgtatttgaa
  61 ctgccagaaa ttaccgataa tgaattactg gtgagtgtaa tttctgacag cgtctgttta
 121 tcgacctgga aagcggcgtt actcggtagt gaacataaac gcgtacccga cgatttagaa
 181 aatcatccgg tcattaccgg gcatgaatgt gccggggtta ttgtcgaagt gggtaaaaat
 241 ctcactggca aatataaaaa aggccagcgt tttgtattgc aaccggcgat ggggttacca
 301 agcggatatt cagcgggcta cagctacgaa tattttggcg gcaatgccac ttatatgatt
 361 attcccgaaa tcgccattaa tttgggctgc gtattaccgt atcacggctc ttattttgct
 421 gcggcgtcgc tggcagagcc tatgtgctgc attattggtg cttatcatgc caattatcac
 481 accacgcaat atgtttatga gcatcgcatg ggcgtcaaac ctggcggcaa tattgcactg
 541 ctggcgtgtg caggtccgat gggcattggc gctatcgatt acgccattaa cggcggcata
 601 caaccgtcgc gggtggtggt ggtcgatatc gacgacaaac gtctggcgca ggtacagaag
 661 ctgctgccgg tggaactggc ggccagtaaa ggcattgagc tggtgtatgt gaataccaaa
 721 gggatgagcg atcctgtcca gatgctgcgg gcgctgacag gagatgccgg gttcgatgac
 781 atttttgttt atgcggcggt gcctgctgtc gttgagatgg ctgatgaatt actggcggaa
```

FIG. 10 (Cont)

```
 841 gatggctgtc tgaacttctt tgccgggccg acggataaaa acttcaaagt gccgtttaat
 901 ttctacaacg tccattacaa cagcacgcac gtcgtcggta catctggcgg ttcaacggac
 961 gacatgaaag aggcgattgc ccttagcgcc actgggcagt tacagccgtc gtttatggtg
1021 acccatatcg gtggcctgga tgcggtgcca gaaaccgtgc tcaatctgcc ggatatccct
1081 ggcggtaaaa aactcattta taacggcgtg accatgccgc tcactgccat tgccgatttt
1141 gccgaaaaag gcaaaaccga tccgctgttt aaagagttgg cgcggctggt tgaggaaacg
1201 cacggcatct ggaatgaaca ggccgagaaa tatctgctgg cacaatttgg cgttgatatc
1261 ggggaggccg cgcaatga
```

Amino acid sequence (SEQ ID NO:128)

```
   1 MKTKVAAIYG KRDVRLRVFE LPEITDNELL VSVISDSVCL STWKAALLGS EHKRVPDDLE
  61 NHPVITGHEC AGVIVEVGKN LTGKYKKGQR FVLQPAMGLP SGYSAGYSYE YFGGNATYMI
 121 IPEIAINLGC VLPYHGSYFA AASLAEPMCC IIGAYHANYH TTQYVYEHRM GVKPGGNIAL
 181 LACAGPMGIG AIDYAINGGI QPSRVVVVDI DDKRLAQVQK LLPVELAASK GIELVYVNTK
 241 GMSDPVQMLR ALTGDAGFDD IFVYAAVPAV VEMADELLAE DGCLNFFAGP TDKNFKVPFN
 301 FYNVHYNSTH VVGTSGGSTD DMKEAIALSA TGQLQPSFMV THIGGLDAVP ETVLNLPDIP
 361 GGKKLIYNGV TMPLTAIADF AEKGKTDPLF KELARLVEET HGIWNEQAEK YLLAQFGVDI
 421 GEAAQ
```

YiaY (YP_026233)

Nucleotide sequence (SEQ ID NO:129)

```
   1 atggcagctt caacgttctt tattccttct gtgaatgtca tcggcgctga ttcattgact
  61 gatgcaatga atatgatggc agattatgga tttacccgta ccttaattgt cactgacaat
 121 atgttaacga aattaggtat ggcgggcgat gtgcaaaaag cactggaaga acgcaatatt
 181 tttagcgtta tttatgatgg cacccaacct aaccccacca cggaaaacgt cgccgcaggt
 241 ttgaaattac ttaaagagaa taattgcgat agcgtgatct ccttaggcgg tggttctcca
 301 cacgactgcg caaaaggtat tgcgctggtg gcagccaatg gcggcgatat tcgcgattac
 361 gaaggcgttg accgctctgc aaaaccgcag ctgccgatga tcgccatcaa taccacggcg
 421 ggtacggcct ctgaaatgac ccgtttctgc atcatcactg acgaagcgcg tcatatcaaa
 481 atggcgattg ttgataaaca tgtcactccg ctgctttctg tcaatgactc ctctctgatg
 541 attggtatgc cgaagtcact gaccgccgca acgggtatgg atgccttaac gcacgctatc
 601 gaagcatatg tttctattgc cgccacgccg atcactgacg cttgtgcact gaaagccgtg
 661 accatgattg ccgaaaacct gccgttagcc gttgaagatg gcagtaatgc gaaagcgcgt
 721 gaagcaatgg cttatgccca gttcctcgcc ggtatggcgt tcaataatgc ttctctgggt
 781 tatgttcatg cgatggcgca ccagctgggc ggtttctaca acctgccaca cggtgtatgt
 841 aacgccgttt tgctgccgca cgttcaggta ttcaacagca aagtcgccgc tgcacgtctg
 901 cgtgactgtg ccgctgcaat gggcgtgaac gtgacaggta aaaacgacgc ggaaggtgct
 961 gaagcctgca ttaacgccat ccgtgaactg gcgaagaaag tggatatccc ggcaggccta
1021 cgcgacctga acgtgaaaga agaagatttc gcggtattgg cgactaatgc cctgaaagat
1081 gcctgtggct ttactaaccc gatccaggca actcacgaag aaattgtggc gatttatcgc
1141 gcagcgatgt aa
```

FIG. 10 (Cont)

Amino acid sequence (SEQ ID NO:130)

```
  1 MAASTFFIPS VNVIGADSLT DAMNMMADYG FTRTLIVTDN MLTKLGMAGD VQKALEERNI
 61 FSVIYDGTQP NPTTENVAAG LKLLKENNCD SVISLGGGSP HDCAKGIALV AANGGDIRDY
121 EGVDRSAKPQ LPMIAINTTA GTASEMTRFC IITDEARHIK MAIVDKHVTP LLSVNDSSLM
181 IGMPKSLTAA TGMDALTHAI EAYVSIAATP ITDACALKAV TMIAENLPLA VEDGSNAKAR
241 EAMAYAQFLA GMAFNNASLG YVHAMAHQLG GFYNLPHGVC NAVLLPHVQV FNSKVAAARL
301 RDCAAAMGVN VTGKNDAEGA EACINAIREL AKKVDIPAGL RDLNVKEEDF AVLATNALKD
361 ACGFTNPIQA THEEIVAIYR AAM
```

FucO (NP_417279)

Nucleotide sequence (SEQ ID NO:131)

```
   1 atgatggcta acagaatgat tctgaacgaa acggcatggt ttggtcgggg tgctgttggg
  61 gctttaaccg atgaggtgaa acgccgtggt tatcagaagg cgctgatcgt caccgataaa
 121 acgctggtgc aatgcggcgt ggtggcgaaa gtgaccgata agatggatgc tgcagggctg
 181 gcatgggcga tttacgacgg cgtagtgccc aacccaacaa ttactgtcgt caaagaaggg
 241 ctcggtgtat tccagaatag cggcgcggat tacctgatcg ctattggtgg tggttctcca
 301 caggatactt gtaaagcgat tggcattatc agcaacaacc cggagtttgc cgatgtgcgt
 361 agcctggaag ggctttcccc gaccaataaa cccagtgtac cgattctggc aattcctacc
 421 acagcaggta ctgcggcaga agtgaccatt aactacgtga tcactgacga agagaaacgg
 481 cgcaagtttg tttgcgttga tccgcatgat atcccgcagg tggcgtttat tgacgctgac
 541 atgatggatg gtatgcctcc agcgctgaaa gctgcgacgg gtgtcgatgc gctcactcat
 601 gctattgagg ggtatattac ccgtggcgcg tgggcgctaa ccgatgcact gcacattaaa
 661 gcgattgaaa tcattgctgg ggcgctgcga ggatcggttg ctggtgataa ggatgccgga
 721 gaagaaatgg cgctcgggca gtatgttgcg ggtatgggct tctcgaatgt tgggttaggg
 781 ttggtgcatg gtatggcgca tccactgggc gcgttttata acactccaca cggtgttgcg
 841 aacgccatcc tgttaccgca tgtcatgcgt tataacgctg actttaccgg tgagaagtac
 901 cgcgatatcg cgcgcgttat gggcgtgaaa gtggaaggta tgagcctgga agaggcgcgt
 961 aatgccgctg ttgaagcggt gtttgctctc aaccgtgatg tcggtattcc gccacatttg
1021 cgtgatgttg gtgtacgcaa ggaagacatt ccggcactgg cgcaggcggc actggatgat
1081 gtttgtaccg gtggcaaccc gcgtgaagca acgcttgagg atattgtaga gctttaccat
1141 accgcctggt aa
```

Amino acid sequence (SEQ ID NO:132)

```
  1 MMANRMILNE TAWFGRGAVG ALTDEVKRRG YQKALIVTDK TLVQCGVVAK VTDKMDAAGL
 61 AWAIYDGVVP NPTITVVKEG LGVFQNSGAD YLIAIGGGSP QDTCKAIGII SNNPEFADVR
121 SLEGLSPTNK PSVPILAIPT TAGTAAEVTI NYVITDEEKR RKFVCVDPHD IPQVAFIDAD
181 MMDGMPPALK AATGVDALTH AIEGYITRGA WALTDALHIK AIEIIAGALR GSVAGDKDAG
241 EEMALGQYVA GMGFSNVGLG LVHGMAHPLG AFYNTPHGVA NAILLPHVMR YNADFTGEKY
301 RDIARVMGVK VEGMSLEEAR NAAVEAVFAL NRDVGIPPHL RDVGVRKEDI PALAQAALDD
361 VCTGGNPREA TLEDIVELYH TAW
```

FIG. 10 (Cont)

EutG (NP_416948)

Nucleotide sequence (SEQ ID NO:133)

```
   1 atgcaaaatg aattgcagac cgcgctcttt caggcgttcg ataccctgaa tctgcaacgg
  61 gtaaaaacat ttagcgttcc accggtgacg ctttgcggtc cgggctcggt gagcagttgc
 121 ggacagcaag cgcaaacgcg tgggctgaaa catctgttcg tgatggcaga cagcttttg
 181 catcaggcag ggatgaccgc cgggctgacg cgtagcctga ccgttaaagg tatcgccatg
 241 acgctctggc catgtccggt gggcgaaccg tgcattaccg acgtgtgtgc agccgtggcg
 301 cagttgcgtg agtcaggctg tgatggggtg atcgcgtttg gcggcggctc ggtgctggat
 361 gcggcgaaag ccgtgacgtt gctggtgacg aacccggata gcacgctggc agagatgtca
 421 gaaaccagcg ttctgcaacc gcgcttgccg ctgattgcca ttccaactac cgccggaacc
 481 ggctctgaaa ccaccaatgt aacggtgatt atcgacgcgg tgagcgggcg caagcaggtg
 541 ttagcccatg cctcgctgat gccggatgtg gcgatcctcg acgccgcatt gaccgaaggt
 601 gtgccgtcgc atgtcacggc gatgaccggc attgatgcgt taacccatgc cattgaagca
 661 tacagcgccc tgaacgctac accgtttacc gacagtctgg cgattggtgc cattgcgatg
 721 attggcaaat cgctgccgaa agcggtgggc tacggtcacg accttgccgc gcgcgagagc
 781 atgttgctgg cttcatgtat ggcgggaatg gcgttttcca gtgcgggtct tgggttgtgc
 841 cacgcgatgg cgcatcagcc gggcgcggcg ctgcatattc cgcacggtct cgcgaacgcc
 901 atgttgctgc caacggtgat ggaatttaac cggatggttt gtcgtgaacg ctttagtcag
 961 attggtcggg cactgcgaac taaaaaatcc gacgatcgtg acgctattaa cgcggtaagt
1021 gagctgattg cggaagttgg gattggtaaa cgactgggcg atgttggtgc gacatctgcg
1081 cattacggcg catgggcgca ggccgcgctg gaagatattt gtctgcgcag taacccgcgt
1141 accgccagcc tggagcagat tgtcggcctg tacgcagcgg cgcaataa
```

Amino acid sequence (SEQ ID NO:134)

```
   1 MQNELQTALF QAFDTLNLQR VKTFSVPPVT LCGPGSVSSC GQQAQTRGLK HLFVMADSFL
  61 HQAGMTAGLT RSLTVKGIAM TLWPCPVGEP CITDVCAAVA QLRESGCDGV IAFGGGSVLD
 121 AAKAVTLLVT NPDSTLAEMS ETSVLQPRLP LIAIPTTAGT GSETTNVTVI IDAVSGRKQV
 181 LAHASLMPDV AILDAALTEG VPSHVTAMTG IDALTHAIEA YSALNATPFT DSLAIGAIAM
 241 IGKSLPKAVG YGHDLAARES MLLASCMAGM AFSSAGLGLC HAMAHQPGAA LHIPHGLANA
 301 MLLPTVMEFN RMVCRERFSQ IGRALRTKKS DDRDAINAVS ELIAEVGIGK RLGDVGATSA
 361 HYGAWAQAAL EDICLRSNPR TASLEQIVGL YAAAQ
```

YqhD (NP_417484)

Nucleotide sequence (SEQ ID NO:135)

```
   1 atgaacaact ttaatctgca caccccaacc cgcattctgt ttggtaaagg cgcaatcgct
  61 ggtttacgcg aacaaattcc tcacgatgct cgcgtattga ttacctacgg cggcggcagc
 121 gtgaaaaaaa ccggcgttct cgatcaagtt ctggatgccc tgaaaggcat ggacgtgctg
 181 gaatttggcg gtattgagcc aaacccggct tatgaaacgc tgatgaacgc cgtgaaactg
 241 gttcgcgaac agaaagtgac tttcctgctg gcggttggcg gcggttctgt actggacggc
 301 accaaattta tcgccgcagc ggctaactat ccggaaaata tcgatccgtg gcacattctg
```

FIG. 10 (Cont)

```
361 caaacgggcg gtaaagagat taaaagcgcc atcccgatgg gctgtgtgct gacgctgcca
421 gcaaccggtt cagaatccaa cgcaggcgcg gtgatctccc gtaaaaccac aggcgacaag
481 caggcgttcc attctgccca tgttcagccg gtatttgccg tgctcgatcc ggtttatacc
541 tacaccctgc cgccgcgtca ggtggctaac ggcgtagtgg acgcctttgt acacaccgtg
601 gaacagtatg ttaccaaacc ggttgatgcc aaaattcagg accgtttcgc agaaggcatt
661 ttgctgacgc taatcgaaga tggtccgaaa gccctgaaag agccagaaaa ctacgatgtg
721 cgcgccaacg tcatgtgggc ggcgactcag gcgctgaacg gtttgattgg cgctggcgta
781 ccgcaggact gggcaacgca tatgctgggc cacgaactga ctgcgatgca cggtctggat
841 cacgcgcaaa cactggctat cgtcctgcct gcactgtgga atgaaaaacg cgataccaag
901 cgcgctaagc tgctgcaata tgctgaacgc gtctggaaca tcactgaagg ttccgatgat
961 gagcgtattg acgccgcgat tgccgcaacc cgcaatttct ttgagcaatt aggcgtgccg
1021 acccacctct ccgactacgg tctggacggc agctccatcc cggctttgct gaaaaaactg
1081 gaagagcacg gcatgaccca actgggcgaa aatcatgaca ttacgttgga tgtcagccgc
1141 cgtatatacg aagccgcccg ctaa
```

Amino acid sequence (SEQ ID NO:136)

```
  1 MNNFNLHTPT RILFGKGAIA GLREQIPHDA RVLITYGGGS VKKTGVLDQV LDALKGMDVL
 61 EFGGIEPNPA YETLMNAVKL VREQKVTFLL AVGGGSVLDG TKFIAAAANY PENIDPWHIL
121 QTGGKEIKSA IPMGCVLTLP ATGSESNAGA VISRKTTGDK QAFHSAHVQP VFAVLDPVYT
181 YTLPPRQVAN GVVDAFVHTV EQYVTKPVDA KIQDRFAEGI LLTLIEDGPK ALKEPENYDV
241 RANVMWAATQ ALNGLIGAGV PQDWATHMLG HELTAMHGLD HAQTLAIVLP ALWNEKRDTK
301 RAKLLQYAER VWNITEGSDD ERIDAAIAAT RNFFEQLGVP THLSDYGLDG SSIPALLKKL
361 EEHGMTQLGE NHDITLDVSR RIYEAAR
```

AdhE (NP_415757)

Nucleotide sequence (SEQ ID NO:137)

```
  1 atggctgtta ctaatgtcgc tgaacttaac gcactcgtag agcgtgtaaa aaaagcccag
 61 cgtgaatatg ccagtttcac tcaagagcaa gtagacaaaa tcttccgcgc cgccgctctg
121 gctgctgcag atgctcgaat cccactcgcg aaaatggccg ttgccgaatc cggcatgggt
181 atcgtcgaag ataaagtgat caaaaaccac tttgcttctg aatatatcta caacgcctat
241 aaagatgaaa aaacctgtgg tgttctgtct gaagacgaca cttttggtac catcactatc
301 gctgaaccaa tcggtattat ttgcggtatc gttccgacca ctaacccgac ttcaactgct
361 atcttcaaat cgctgatcag tctgaagacc cgtaacgcca ttatcttctc cccgcacccg
421 cgtgcaaaag atgccaccaa caaagcggct gatatcgttc tgcaggctgc tatcgctgcc
481 ggtgctccga aagatctgat cggctggatc gatcaacctt ctgttgaact gtctaacgca
541 ctgatgcacc acccagacat caacctgatc ctcgcgactg gtggtccggg catggttaaa
601 gccgcataca gctccggtaa accagctatc ggtgtaggcg cgggcaacac tccagttgtt
661 atcgatgaaa ctgctgatat caaacgtgca gttgcatctg tactgatgtc caaaaccttc
721 gacaacggcg taatctgtgc ttctgaacag tctgttgttg ttgttgactc tgtttatgac
781 gctgtacgtg aacgttttgc aacccacggc ggctatctgt tgcagggtaa agagctgaaa
841 gctgttcagg atgttatcct gaaaaacggt gcgctgaacg cggctatcgt tggtcagcca
```

FIG. 10 (Cont)

```
 901 gcctataaaa ttgctgaact ggcaggcttc tctgtaccag aaaacaccaa gattctgatc
 961 ggtgaagtga ccgttgttga tgaaagcgaa ccgttcgcac atgaaaaact gtccccgact
1021 ctggcaatgt accgcgctaa agatttcgaa gacgcggtag aaaaagcaga gaaactggtt
1081 gctatgggcg gtatcggtca tacctcttgc ctgtacactg accaggataa ccaaccggct
1141 cgcgtttctt acttcggtca gaaaatgaaa acggcgcgta tcctgattaa caccccagcg
1201 tctcagggtg gtatcggtga cctgtataac ttcaaactcg caccttccct gactctgggt
1261 tgtggttctt ggggtggtaa ctccatctct gaaaacgttg gtccgaaaca cctgatcaac
1321 aagaaaaccg ttgctaagcg agctgaaaac atgttgtggc acaaacttcc gaaatctatc
1381 tacttccgcc gtggctccct gccaatcgcg ctggatgaag tgattactga tggccacaaa
1441 cgtgcgctca tcgtgactga ccgcttcctg ttcaacaatg gttatgctga tcagatcact
1501 tccgtactga aagcagcagg cgttgaaact gaagtcttct tcgaagtaga agcggacccg
1561 accctgagca tcgttcgtaa aggtgcagaa ctggcaaact ccttcaaacc agacgtgatt
1621 atcgcgctgg gtggtggttc cccgatggac gccgcgaaga tcatgtgggt tatgtacgaa
1681 catccggaaa ctcacttcga agagctggcg ctgcgcttta tggatatccg taaacgtatc
1741 tacaagttcc cgaaaatggg cgtgaaagcg aaaatgatcg ctgtcaccac cacttctggt
1801 acaggttctg aagtcactcc gtttgcggtt gtaactgacg acgctactgg tcagaaatat
1861 ccgctggcag actatgcgct gactccggat atggcgattg tcgacgccaa cctggttatg
1921 gacatgccga agtccctgtg tgctttcggt ggtctggacg cagtaactca cgccatggaa
1981 gcttatgttt ctgtactggc atctgagttc tctgatggtc aggctctgca ggcactgaaa
2041 ctgctgaaag aatatctgcc agcgtcctac cacgaagggt ctaaaaatcc ggtagcgcgt
2101 gaacgtgttc acagtgcagc gactatcgcg ggtatcgcgt ttgcgaacgc cttcctgggt
2161 gtatgtcact caatggcgca caaactgggt tcccagttcc atattccgca cggtctggca
2221 aacgccctgc tgatttgtaa cgttattcgc tacaatgcga acgacaaccc gaccaagcag
2281 actgcattca gccagtatga ccgtccgcag gctcgccgtc gttatgctga aattgccgac
2341 cacttgggtc tgagcgcacc gggcgaccgt actgctgcta agatcgagaa actgctggca
2401 tggctggaaa cgctgaaagc tgaactgggt attccgaaat ctatccgtga agctggcgtt
2461 caggaagcag acttcctggc gaacgtggat aaactgtctg aagatgcatt cgatgaccag
2521 tgcaccggcg ctaacccgcg ttacccgctg atctccgagc tgaaacagat tctgctggat
2581 acctactacg gtcgtgatta tgtagaaggt gaaactgcag cgaagaaaga agctgctccg
2641 gctaaagctg agaaaaaagc gaaaaaatcc gcttaa
```

Amino acid sequence (SEQ ID NO:138)

```
  1 MAVTNVAELN ALVERVKKAQ REYASFTQEQ VDKIFRAAAL AAADARIPLA KMAVAESGMG
 61 IVEDKVIKNH FASEYIYNAY KDEKTCGVLS EDDTFGTITI AEPIGIICGI VPTTNPTSTA
121 IFKSLISLKT RNAIIFSPHP RAKDATNKAA DIVLQAAIAA GAPKDLIGWI DQPSVELSNA
181 LMHHPDINLI LATGGPGMVK AAYSSGKPAI GVGAGNTPVV IDETADIKRA VASVLMSKTF
241 DNGVICASEQ SVVVVDSVYD AVRERFATHG GYLLQGKELK AVQDVILKNG ALNAAIVGQP
301 AYKIAELAGF SVPENTKILI GEVTVVDESE PFAHEKLSPT LAMYRAKDFE DAVEKAEKLV
361 AMGGIGHTSC LYTDQDNQPA RVSYFGQKMK TARILINTPA SQGGIGDLYN FKLAPSLTLG
421 CGSWGGNSIS ENVGPKHLIN KKTVAKRAEN MLWHKLPKSI YFRRGSLPIA LDEVITDGHK
481 RALIVTDRFL FNNGYADQIT SVLKAAGVET EVFFEVEADP TLSIVRKGAE LANSFKPDVI
541 IALGGGSPMD AAKIMWVMYE HPETHFEELA LRFMDIRKRI YKFPKMGVKA KMIAVTTTSG
601 TGSEVTPFAV VTDDATGQKY PLADYALTPD MAIVDANLVM DMPKSLCAFG GLDAVTHAME
```

FIG. 10 (Cont)

661 AYVSVLASEF SDGQALQALK LLKEYLPASY HEGSKNPVAR ERVHSAATIA GIAFANAFLG
721 VCHSMAHKLG SQFHIPHGLA NALLICNVIR YNANDNPTKQ TAFSQYDRPQ ARRRYAEIAD
781 HLGLSAPGDR TAAKIEKLLA WLETLKAELG IPKSIREAGV QEADFLANVD KLSEDAFDDQ
841 CTGANPRYPL ISELKQILLD TYYGRDYVEG ETAAKKEAAP AKAEKKAKKS A

dkgB (NP_414743)

Nucleotide sequence (SEQ ID NO:139)

```
   1 atggctatcc ctgcatttgg tttaggtact ttccgtctga aagacgacgt tgttatttca
  61 tctgtgataa cggcgcttga acttggttat cgcgcaattg ataccgcaca aatctatgat
 121 aacgaagccg cagtaggtca ggcgattgca gaaagtggcg tgccacgtca tgaactctac
 181 atcaccacta aaatctggat tgaaaatctc agcaaagaca aattgatccc aagtctgaaa
 241 gagagcctgc aaaaattgcg taccgattat gttgatctga cgctaatcca ctggccgtca
 301 ccaaacgatg aagtctctgt tgaagagttt atgcaggcgc tgctggaagc caaaaaacaa
 361 gggctgacgc gtgagatcgg tatttccaac ttcacgatcc cgttgatgga aaaagcgatt
 421 gctgctgttg gtgctgaaaa catcgctact aaccagattg aactctctcc ttatctgcaa
 481 aaccgtaaag tggttgcctg ggctaaacag cacggcatcc atattacttc ctatatgacg
 541 ctggcgtatg gtaaggccct gaaagatgag gttattgctc gtatcgcagc taaacacaat
 601 gcgactccgg cacaagtgat tctggcgtgg gctatggggg aaggttactc agtaattcct
 661 tcttctacta aacgtaaaaa cctggaaagt aatcttaagg cacaaaattt acagcttgat
 721 gccgaagata aaaaagcgat cgccgcactg gattgcaacg accgcctggt tagcccggaa
 781 ggtctggctc ctgaatggga ttaa
```

Amino acid sequence (SEQ ID NO:140)

```
   1 MAIPAFGLGT FRLKDDVVIS SVITALELGY RAIDTAQIYD NEAAVGQAIA ESGVPRHELY
  61 ITTKIWIENL SKDKLIPSLK ESLQKLRTDY VDLTLIHWPS PNDEVSVEEF MQALLEAKKQ
 121 GLTREIGISN FTIPLMEKAI AAVGAENIAT NQIELSPYLQ NRKVVAWAKQ HGIHITSYMT
 181 LAYGKALKDE VIARIAAKHN ATPAQVILAW AMGEGYSVIP SSTKRKNLES NLKAQNLQLD
 241 AEDKKAIAAL DCNDRLVSPE GLAPEWD
```

YdjG (NP_416285)

Nucleotide sequence (SEQ ID NO:141)

```
   1 atgaaaaaga tacctttagg cacaacggat attacgcttt cgcgaatggg gttggggaca
  61 tgggccattg gcggcggtcc tgcatggaat ggcgatctcg atcggcaaat atgtattgat
 121 acgattcttg aagcccatcg ttgtggcatt aatctgattg atactgcgcc aggatataac
 181 tttggcaata gtgaagttat cgtcggtcag gcgttaaaaa aactgccccg tgaacaggtt
 241 gtagtagaaa ccaaatgcgg cattgtctgg gaacgaaaag gaagtttatt caacaaagtt
 301 ggcgatcggc agttgtataa aaacctttcc ccggaatcta tccgcgaaga ggtagcagcg
 361 agcttgcaac gtctgggtat tgattacatc gatatctaca tgacgcactg gcagtcggtg
 421 ccgccatttt ttacgccgat cgctgaaact gtcgcagtgc ttaatgagtt aaagtctgaa
```

FIG. 10 (Cont)

```
481 gggaaaattc gcgctatagg cgctgctaac gtcgatgctg accatatccg cgagtatctg
541 caatatggtg aactggatat tattcaggcg aaatacagta tcctcgaccg ggcaatggaa
601 aacgaactgc tgccactatg tcgtgataat ggcattgtgg ttcaggttta ttccccgcta
661 gagcagggat tgttgaccgg caccatcact cgtgattacg ttccgggcgg cgctcgggca
721 ataaaagtct ggttccagcg tgaaaacatg ctgaaagtga ttgatatgct tgaacagtgg
781 cagccacttt gtgctcgtta tcagtgcaca attcccactc tggcactggc gtggatatta
841 aaacagagtg atttaatctc cattcttagt ggggctactg caccggaaca ggtacgcgaa
901 aatgtcgcgg cactgaatat caacttatcg gatgcagacg caacattgat gagggaaatg
961 gcagaggccc tggagcgtta a
```

Amino acid sequence (SEQ ID NO:142)

```
  1 MKKIPLGTTD ITLSRMGLGT WAIGGGPAWN GDLDRQICID TILEAHRCGI NLIDTAPGYN
 61 FGNSEVIVGQ ALKKLPREQV VVETKCGIVW ERKGSLFNKV GDRQLYKNLS PESIREEVAA
121 SLQRLGIDYI DIYMTHWQSV PPFFTPIAET VAVLNELKSE GKIRAIGAAN VDADHIREYL
181 QYGELDIIQA KYSILDRAME NELLPLCRDN GIVVQVYSPL EQGLLTGTIT RDYVPGGARA
241 NKVWFQRENM LKVIDMLEQW QPLCARYQCT IPTLALAWIL KQSDLISILS GATAPEQVRE
301 NVAALNINLS DADATLMREM AEALER
```

YeaE (NP_416295)

Nucleotide sequence (SEQ ID NO:143)

```
  1 atgcaacaaa aaatgattca atttagtggc gatgtctcac tgccagccgt agggcaggga
 61 acatggtata tgggcgaaga tgccagtcag cgcaaaacag aagttgctgc actacgcgcg
121 ggcattgaac tcggtttaac cctcattgat accgccgaaa tgtatgccga tggcggtgcc
181 gaaaaggtgg ttggggaagc attaaccggt ctgcgagaga aggtctttct cgtctctaaa
241 gtctatccgt ggaatgctgg cgggcaaaaa gcgataaatg catgcgaagc cagtttacgc
301 cgtctcaata ctgattatct cgatctttac ttattacact ggtctggcag tttcgctttt
361 gaagagactg tcgcagcgat ggaaaaattg atcgcccagg gaaaaatccg ccgctggggc
421 gtttctaacc ttgattatgc tgatatgcag gaactctggc agctgccggg gggaaatcag
481 tgtgccacta atcaggtgct ttaccatctc ggttcacgag gaattgagta cgatctactc
541 ccctggtgcc agcaacagca gatgccggtg atggcttaca gtccgttagc ccaggccggg
601 cggttgcgca atggactgtt aaaaaacgcg gtagtcaacg aaattgcaca tgctcacaat
661 atcagcgcgg cacaagtatt gttggcgtgg gtgatcagtc atcagggtgt gatggcgatt
721 ccaaaagcgg ccacgattgc ccatgtccaa caaaatgcgg ctgtgcttga ggtcgaactt
781 tcttcagcgg aattagctat gctggataag gcatatccgg caccaaaagg aaaaactgcg
841 ctggatatgg tgtga
```

Amino acid sequence (SEQ ID NO:144)

```
  1 MQQKMIQFSG DVSLPAVGQG TWYMGEDASQ RKTEVAALRA GIELGLTLID TAEMYADGGA
 61 EKVVGEALTG LREKVFLVSK VYPWNAGGQK AINACEASLR RLNTDYLDLY LLHWSGSFAF
121 EETVAAMEKL IAQGKIRRWG VSNLDYADMQ ELWQLPGGNQ CATNQVLYHL GSRGIEYDLL
```

FIG. 10 (Cont)

181 PWCQQQQMPV MAYSPLAQAG RLRNGLLKNA VVNEIAHAHN ISAAQVLLAW VISHQGVMAI
241 PKAATIAHVQ QNAAVLEVEL SSAELAMLDK AYPAPKGKTA LDMV

dkgA(NP_417485)

Nucleotide sequence (SEQ ID NO:145)

```
  1 atggctaatc caaccgttat taagctacag gatggcaatg tcatgcccca gctgggactg
 61 ggcgtctggc aagcaagtaa tgaggaagta atcaccgcca ttcaaaaagc gttagaagtg
121 ggttatcgct cgattgatac cgccgcggcc tacaagaacg aagaaggtgt cggcaaagcc
181 ctgaaaaatg cctcagtcaa cagagaagaa ctgttcatca ccactaagct gtggaacgac
241 gaccacaagc gcccccgcga agccctgctc gacagcctga aaaaactcca gcttgattat
301 atcgacctct acttaatgca ctggcccgtt cccgctatcg accattatgt cgaagcatgg
361 aaaggcatga tcgaattgca aaaagaggga ttaatcaaaa gcatcggcgt gtgcaacttc
421 cagatccatc acctgcaacg cctgattgat gaaactggcg tgacgcctgt gataaaccag
481 atcgaacttc atccgctgat gcaacaacgc cagctacacg cctggaacgc gacacacaaa
541 atccagaccg aatcctggag cccattagcg caaggaggga aaggcgtttt cgatcagaaa
601 gtcattcgcg atctggcaga taaatacggc aaaaccccgg cgcagattgt tatccgctgg
661 catctggata gcggcctggt ggtgatcccg aaatcggtca caccttcacg tattgccgaa
721 aactttgatg tctgggattt ccgtctcgac aaagacgaac tcggcgaaat tgcaaaactc
781 gatcagggca agcgtctcgg tcccgatcct gaccagttcg gcggctaa
```

Amino acid sequence (SEQ ID NO:146)

```
  1 MANPTVIKLQ DGNVMPQLGL GVWQASNEEV ITAIQKALEV GYRSIDTAAA YKNEEGVGKA
 61 LKNASVNREE LFITTKLWND DHKRPREALL DSLKKLQLDY IDLYLMHWPV PAIDHYVEAW
121 KGMIELQKEG LIKSIGVCNF QIHHLQRLID ETGVTPVINQ IELHPLMQQR QLHAWNATHK
181 IQTESWSPLA QGGKGVFDQK VIRDLADKYG KTPAQIVIRW HLDSGLVVIP KSVTPSRIAE
241 NFDVWDFRLD KDELGEIAKL DQGKRLGPDP DQFGG
```

YajO (NP_414953)

Nucleotide sequence (SEQ ID NO:147)

```
  1 atgcaataca acccttagg aaaaaccgac cttcgcgttt cccgactttg cctcggctgt
 61 atgacctttg gcgagccaga tcgcggtaat cacgcatgga cactgccgga agaaagcagc
121 cgtcccataa ttaaacgtgc actggaaggc ggcataaatt tctttgatac cgccaacagt
181 tattctgacg gcagcagcga agagatcgtc ggtcgcgcac tgcgggattt cgcccgtcgt
241 gaagacgtgg tcgttgcgac caaagtgttc catcgcgttg gtgatttacc ggaaggatta
301 tcccgtgcgc aaattttgcg ctctatcgac gacagcctgc gacgtctcgg catggattat
361 gtcgatatcc tgcaaattca tcgctgggat tacaacacgc cgatcgaaga gacgctggaa
421 gccctcaacg acgtggtaaa agccgggaaa gcgcgttata tcggcgcgtc atcaatgcac
481 gcttcgcagt ttgctcaggc actggaactc caaaaacagc acggctgggc gcagtttgtc
541 agtatgcagg atcactacaa tctgatttat cgtgaagaag agcgcgagat gctaccactg
```

FIG. 10 (Cont)

601 tgttatcagg agggcgtggc ggtaattcca tggagcccgc tggcaagggg ccgtctgacg
661 cgtccgtggg gagaaactac cgcacgactg gtgtctgatg aggtggggaa aaatctctat
721 aaagaaagcg atgaaaatga cgcgcagatc gcagagcggt taacaggcgt cagtgaagaa
781 ctgggggcga cacgagcaca agttgcgctg gcctggttgt tgagtaaacc gggcattgcc
841 gcaccgatta tcggaacttc gcgcgaagaa cagcttgatg agctattgaa cgcggtggat
901 atcactttga agccggaaca gattgccgaa ctggaaacgc cgtataaacc gcatcctgtc
961 gtaggattta aataa

Amino acid sequence (SEQ ID NO:148)

1 MQYNPLGKTD LRVSRLCLGC MTFGEPDRGN HAWTLPEESS RPIIKRALEG GINFFDTANS
61 YSDGSSEEIV GRALRDFARR EDVVVATKVF HRVGDLPEGL SRAQILRSID DSLRRLGMDY
121 VDILQIHRWD YNTPIEETLE ALNDVVKAGK ARYIGASSMH ASQFAQALEL QKQHGWAQFV
181 SMQDHYNLIY REEEREMLPL CYQEGVAVIP WSPLARGRLT RPWGETTARL VSDEVGKNLY
241 KESDENDAQI AERLTGVSEE LGATRAQVAL AWLLSKPGIA APIIGTSREE QLDELLNAVD
301 ITLKPEQIAE LETPYKPHPV VGFK

YghZ (NP_417474)

Nucleotide sequence (SEQ ID NO:149)

1 atggtctggt tagcgaatcc cgaacgttac gggcagatgc aataccgcta ttgcggaaaa
61 agtggtttac gcctgcccgc gttatcgctc ggtttatggc acaatttcgg tcacgttaac
121 gcgctggaat cacagcgtgc gatcctgcgt aaagcgtttg atttgggcat tacgcacttt
181 gatttagcca acaattacgg gccgcctcca ggaagcgcag aagagaactt tggtcgcctg
241 ctgcgggagg attttgccgc ttatcgcgat gaactgatta tctctaccaa ggctggctac
301 gatatgtggc ccggccctta cggctctggc ggttcacgta aatacctgct cgccagcctc
361 gaccaaagcc tgaagcgtat ggggcttgag tatgtcgata tcttttactc tcatcgcgtc
421 gatgaaaata cgccgatgga agaaaccgcc tctgcgctgg ctcatgcggt acaaagcggt
481 aaggcgctgt atgtcgggat ctcctcttac tcgccagagc ggacgcaaaa aatggtcgag
541 ttgctgcgcg agtggaaaat tccgctgtta attcatcaac cttcgtacaa tttactgaac
601 cgctgggtgg ataaaagcgg cctgctggat accctgcaaa ataacggcgt gggctgtatt
661 gcctttactc ctctggctca gggattgctg accggaaaat atctcaacgg cattccgcaa
721 gattcacgga tgcatcgtga agggaataaa gttcgtggtc tgacaccgaa aatgcttacc
781 gaagccaacc tcaacagcct gcgcttattg aatgaaatgg cacagcagcg tggacaatca
841 atggcgcaaa tggcgttaag ctggttgctg aaagatgatc gcgtgacgtc ggtattgatt
901 ggtgccagcc gcgcggagca actagaggag aacgtgcagg cgctgaataa tctgacattt
961 agcaccaagg agctggcgca gattgatcag catatcgccg atggcgagct gaatctgtgg
1021 caggcgtctt ccgataaatg a

Amino acid sequence (SEQ ID NO:150)

1 MVWLANPERY GQMQYRYCGK SGLRLPALSL GLWHNFGHVN ALESQRAILR KAFDLGITHF
61 DLANNYGPPP GSAEENFGRL LREDFAAYRD ELIISTKAGY DMWPGPYGSG GSRKYLLASL

FIG. 10 (Cont)

121 DQSLKRMGLE YVDIFYSHRV DENTPMEETA SALAHAVQSG KALYVGISSY SPERTQKMVE
181 LLREWKIPLL IHQPSYNLLN RWVDKSGLLD TLQNNGVGCI AFTPLAQGLL TGKYLNGIPQ
241 DSRMHREGNK VRGLTPKMLT EANLNSLRLL NEMAQQRGQS MAQMALSWLL KDDRVTSVLI
301 GASRAEQLEE NVQALNNLTF STKELAQIDQ HIADGELNLW QASSDK

Tas (NP_417311)

Nucleotide sequence (SEQ ID NO:151)

```
   1 atgcaatatc accgtatacc ccacagttcg ctggaagtca gcacgctggg gcttggcacg
  61 atgacgtttg gtgaacagaa cagcgaagcc gacgcccacg cacaactcga ctatgccgtc
 121 gctcagggca ttaaccttat cgacgttgcc gaaatgtacc cagtacctcc gcgccccgaa
 181 acgcaagggt taaccgaaac ctacgtcggc aactggctgg cgaaacatgg cagccgcgaa
 241 aagttaatta tcgcctccaa agtgagcgga ccgtcgcgca ataatgacaa gggcatccgc
 301 ccggatcagg cgctggatcg gaagaatatc cgcgaagcgc tgcatgacag cctcaagcgc
 361 ctacagactg attacctcga tctttatcag gtgcactggc cgcagcgccc gaccaactgc
 421 ttcggcaaac tcggttatag ctggacggat tctgcgcctg cggtttcgct gctggatacg
 481 ctggacgcac tggcagagta ccaacgcgcg ggaaaaattc gttatatcgg cgtgtcgaac
 541 gaaactgcat ttggcgtaat gcgctacctg catctggcgg acaaacacga tctgccgcgt
 601 attgtcacca ttcagaaccc ttacagtctg ttaaaccgca gttttgaagt aggtctggca
 661 gaagtcagcc agtatgaagg ggtcgaactg ctggcctatt cgtgcctggg tttcggcacg
 721 ctgaccggga aatatctcaa tggtgcaaaa cccgctggcg cacgtaatac gctctttagt
 781 cggttcaccc gctatagcgg tgagcaaacg caaaaagccg tcgcggcgta tgttgatatc
 841 gccagacgtc atggcctgga ccctgctcag atggcgctcg cgtttgtacg ccgtcaaccg
 901 tttgttgcca gcactctgct gggcgcaacc acgatggatc agctgaaaac taacatcgaa
 961 agtttgcatc tggagttaag cgaagacgta ttagctgaaa ttgaagcggt gcatcaggtt
1021 tatacttatc cggcaccata a
```

Amino acid sequence (SEQ ID NO:152)

```
   1 MQYHRIPHSS LEVSTLGLGT MTFGEQNSEA DAHAQLDYAV AQGINLIDVA EMYPVPPRPE
  61 TQGLTETYVG NWLAKHGSRE KLIIASKVSG PSRNNDKGIR PDQALDRKNI REALHDSLKR
 121 LQTDYLDLYQ VHWPQRPTNC FGKLGYSWTD SAPAVSLLDT LDALAEYQRA GKIRYIGVSN
 181 ETAFGVMRYL HLADKHDLPR IVTIQNPYSL LNRSFEVGLA EVSQYEGVEL LAYSCLGFGT
 241 LTGKYLNGAK PAGARNTLFS RFTRYSGEQT QKAVAAYVDI ARRHGLDPAQ MALAFVRRQP
 301 FVASTLLGAT TMDQLKTNIE SLHLELSEDV LAEIEAVHQV YTYPAP
```

YdhF (YP_025305)

Nucleotide sequence (SEQ ID NO:153)

```
   1 atggttcagc gtattactat tgcgccgcaa ggcccggagt tttcccgttt tgtgatgggc
  61 tactggcgat tgatggactg gaatatgtcc gcccgccagc tggtcagttt tattgaagag
 121 catctggatc tcggcgtgac caccgtggac catgctgata tttatggtgg ctatcagtgc
 181 gaagcggcgt ttggcgaggc actgaaactg gcacctcacc tgcgtgaacg gatggaaatc
```

FIG. 10 (Cont)

241 gtcagtaaat gcggtatcgc gacgaccgcg cgtgaagaaa acgtcattgg tcattacatc
301 actgaccgcg atcacatcat taagagcgcc gaacagtcgc taattaatct cgcgaccgat
361 catctggatt tgctgttaat ccaccgacca gacccgttaa tggatgccga tgaagtggcg
421 gacgcgttca aacatctgca tcagagcggc aaagtgcgtc attttggcgt atcgaacttt
481 acgcctgcgc aatttgccct gttgcaatca cgtctgccgt ttacccttgc cactaatcag
541 gtggaaatat ccccggtgca tcagccgtta ctgctggatg gcacgctcga ccaactacaa
601 caactgcgtg ttcgtccgat ggcgtggtcc tgccttggtg gtggtcgtct gtttaatgat
661 gattatttcc agccgctgcg tgatgaactg gctgtggtgg cagaggagtt aaacgcgggc
721 tcgattgaac aggtggttta cgcctgggta ttacgtttac catcgcagcc gctgccaatt
781 atcggttcag gtaaaattga gcgcgtacgg gcagctgtcg aagcagaaac actgaaaatg
841 acccgtcaac aatggtttcg tatccgtaaa gcggcactgg ggtacgacgt accgtaa

Amino acid sequence (SEQ ID NO:154)

1 MVQRITIAPQ GPEFSRFVMG YWRLMDWNMS ARQLVSFIEE HLDLGVTTVD HADIYGGYQC
61 EAAFGEALKL APHLRERMEI VSKCGIATTA REENVIGHYI TDRDHIIKSA EQSLINLATD
121 HLDLLLIHRP DPLMDADEVA DAFKHLHQSG KVRHFGVSNF TPAQFALLQS RLPFTLATNQ
181 VEISPVHQPL LLDGTLDQLQ QLRVRPMAWS CLGGGRLFND DYFQPLRDEL AVVAEELNAG
241 SIEQVVYAWV LRLPSQPLPI IGSGKIERVR AAVEAETLKM TRQQWFRIRK AALGYDVP

YdbC (NP_415924)

Nucleotide sequence (SEQ ID NO:155)

1 atgagcagca atacatttac tctcggtaca aaatccgtta accgtcttgg ttatggcgcg
61 atgcaactgg caggtcctgg agtttttggc cccccacgag atcgccacgt cgctataacc
121 gtgctgcgtg aggcgctggc attgggcgtc aatcatattg ataccagcga cttttatggt
181 ccgcacgtca ccaatcagat tatccgcgaa gcgctttatc cttactctga cgacctgaca
241 attgtcacta aaattggtgc gcggcgtgga gaggacgcat cctggttgcc cgcattttct
301 ccggcagagc tgcaaaaagc ggtgcacgat aatctacgta atctcgggct ggacgtgctg
361 gatgtggtta acctgcgcgt tatgatgggg gatggtcatg gcccagcgga aggatcgatt
421 gaggccagcc tgaccgtgct ggcagagatg caacaacaag gcctggtaaa acatattggc
481 ctgagcaacg tcacaccgac gcaggttgca gaggcgcgca agattgccga aattgtctgt
541 gtgcaaaacg aatacaacat cgcgcaccgt gctgatgatg caatgattga tgctttggcc
601 cacgatggca ttgcctacgt gccgttcttc ccgctcgggg gctttacacc gctgcaatcg
661 tccacacttt ccgatgttgc tgcgagcctg ggtgcaacac caatgcaggt ggcgctggcg
721 tggctgttac agcgttcacc gaatattttg ctgatcccag ggacgtcttc ggttgcgcat
781 ttacgggaga atatggctgc tgaaaaattg catctttctg aggaagtgtt gtctacgttg
841 gatggtattt cgcgagaata a

Amino acid sequence (SEQ ID NO:156)

1 MSSNTFTLGT KSVNRLGYGA MQLAGPGVFG PPRDRHVAIT VLREALALGV NHIDTSDFYG
61 PHVTNQIIRE ALYPYSDDLT IVTKIGARRG EDASWLPAFS PAELQKAVHD NLRNLGLDVL

FIG. 10 (Cont)

121 DVVNLRVMMG DGHGPAEGSI EASLTVLAEM QQQGLVKHIG LSNVTPTQVA EARKIAEIVC
181 VQNEYNIAHR ADDAMIDALA HDGIAYVPFF PLGGFTPLQS STLSDVAASL GATPMQVALA
241 WLLQRSPNIL LIPGTSSVAH LRENMAAEKL HLSEEVLSTL DGISRE

ybbO (NP_415026)

Nucleotide sequence (SEQ ID NO:157)

```
   1 atgactcata aagcaacgga gatcctgaca ggtaaagtta tgcaaaaatc ggtcttaatt
  61 accggatgtt ccagtggaat tggcctggaa agcgcgctcg aattaaaacg ccagggtttt
 121 catgtgctgg caggttgccg gaaaccggat gatgttgagc gcatgaacag catgggattt
 181 accggcgtgt tgatcgatct ggattcacca gaaagtgttg atcgcgcagc cgacgaggtg
 241 atcgccctga ccgataattg tctgtatggg atctttaaca atgccggatt cggcatgtat
 301 ggcccccttt ccaccatcag ccgtgcgcag atggaacagc agttttccgc caacttttc
 361 ggcgcacacc agctcaccat gcgcctgtta cccgcgatgt taccgcacgg tgaagggcgt
 421 attgtgatga catcatcggt gatgggatta atctccacgc cgggtcgtgg cgcttacgcg
 481 gccagtaaat atgcgctgga ggcgtggtca gatgcactgc gcatggagct gcgccacagc
 541 ggaattaaag tcagcctgat cgaacccggt cccattcgta ctcgcttcac cgacaacgtc
 601 aaccagacgc aaagtgataa accagtcgaa aatcccggca tcgccgcccg ctttacgttg
 661 ggaccggaag cggtggtgga caaagtacgc catgctttta ttagcgagaa gccgaagatg
 721 cgctatccgg tgacgctggt gacctgggcg gtaatggtgc ttaagcgcct gctgccgggg
 781 cgcgtgatgg acaaaatatt gcaggggtga
```

Amino acid sequence (SEQ ID NO:158)

```
  1 MTHKATEILT GKVMQKSVLI TGCSSGIGLE SALELKRQGF HVLAGCRKPD DVERMNSMGF
 61 TGVLIDLDSP ESVDRAADEV IALTDNCLYG IFNNAGFGMY GPLSTISRAQ MEQQFSANFF
121 GAHQLTMRLL PAMLPHGEGR IVMTSSVMGL ISTPGRGAYA ASKYALEAWS DALRMELRHS
181 GIKVSLIEPG PIRTRFTDNV NQTQSDKPVE NPGIAARFTL GPEAVVDKVR HAFISEKPKM
241 RYPVTLVTWA VMVLKRLLPG RVMDKILQG
```

yohF (NP_416641)

Nucleotide sequence (SEQ ID NO:159)

```
   1 atggcacagg ttgcgattat taccgcctcc gattcgggga tcggcaaaga gtgcgcgtta
  61 ttactggcgc agcaggggtt tgatattggt attacctggc actcagatga agaaggggca
 121 aaagataccg cgcgtgaggt agttagccac ggcgtacgtg cggagatcgt gcagctggat
 181 ctcggcaatc taccagaagg ggcactggcg ctggagaaac tcattcaacg gctggggcgc
 241 attgatgtgc tggtgaataa tgcgggtgca atgaccaaag cgccgtttct tgatatggct
 301 tttgatgagt ggcgcaagat ttttaccgtt gatgtcgatg gtgcattctt atgctcgcaa
 361 attgcggctc gtcagatggt gaaacaaggg cagggcggtc gcatcatcaa cattacgtcg
 421 gtacatgaac atacgccgct gccggatgcc agcgcctaca cagccgctaa acatgcgctc
 481 ggtgggttaa ccaaagcgat ggcgctggag ctggtcaggc ataagatttt ggtgaacgca
```

FIG. 10 (Cont)

```
541 gtcgcgcctg gggcgatcgc cacgccaatg aatggcatgg atgacagcga cgtgaagccc
601 gacgcggagc cttcgattcc cttgcggcgt tttggcgcaa cgcatgagat tgccagcctg
661 gtggtgtggc tttgttcgga gggcgcaaat tacaccaccg ggcagtcgtt gatagtggat
721 ggcggcttta tgttggcgaa tccacagttc aacccagaat ag
```

Amino acid sequence (SEQ ID NO:160)

```
  1 MAQVAIITAS DSGIGKECAL LLAQQGFDIG ITWHSDEEGA KDTAREVVSH GVRAEIVQLD
 61 LGNLPEGALA LEKLIQRLGR IDVLVNNAGA MTKAPFLDMA FDEWRKIFTV DVDGAFLCSQ
121 IAARQMVKQG QGGRIINITS VHEHTPLPDA SAYTAAKHAL GGLTKAMALE LVRHKILVNA
181 VAPGAIATPM NGMDDSDVKP DAEPSIPLRR FGATHEIASL VVWLCSEGAN YTTGQSLIVD
241 GGFMLANPQF NPE
```

YciK (NP_415787)

Nucleotide sequence (SEQ ID NO:161)

```
   1 atgcattacc agccaaaaca agatttactc aatgatcgca ttatcctggt gacgggagcc
  61 agcgatggta ttggtcgtga agccgcgatg acgtatgcac gctatggtgc gacagtgatt
 121 ctgttgggcc gtaatgaaga aaaattacgt caggtagcca gccacataaa cgaagaaact
 181 gggcgtcagc cacagtggtt tattctcgat ttgctgacct gcacgtccga aaattgccaa
 241 caactggcac agcgcattgc cgttaattat ccgcgtctgg atggtgtttt gcataatgcc
 301 ggattgctcg gcgatgtttg cccaatgagc gaacaaaatc cgcaggtctg gcaggacgtc
 361 atgcaggtca acgttaatgc cacctttatg ctcacccagg cactgcttcc tttattactc
 421 aaatcggacg ccggttcact ggtctttact tcatcaagcg ttggacgtca gggacgagcc
 481 aactggggtg catatgcagc gtcgaaattt gccaccgaag ggatgatgca ggtactggcc
 541 gatgaatatc agcagcgcct gcgtgtcaac tgcattaacc caggcggtac gcgcaccgca
 601 atgcgtgcca gcgccttccc gaccgaagat ccacagaaac ttaaaacacc cgctgatatc
 661 atgccgctct acctctggct gatgggcgat gacagccgcc gtaaaaccgg catgaccttt
 721 gacgcccaac cgggccgtaa accaggaatt tcccaatga
```

Amino acid sequence (SEQ ID NO:162)

```
  1 MHYQPKQDLL NDRIILVTGA SDGIGREAAM TYARYGATVI LLGRNEEKLR QVASHINEET
 61 GRQPQWFILD LLTCTSENCQ QLAQRIAVNY PRLDGVLHNA GLLGDVCPMS EQNPQVWQDV
121 MQVNVNATFM LTQALLPLLL KSDAGSLVFT SSSVGRQGRA NWGAYAASKF ATEGMMQVLA
181 DEYQQRLRVN CINPGGTRTA MRASAFPTED PQKLKTPADI MPLYLWLMGD DSRRKTGMTF
241 DAQPGRKPGI SQ
```

YgfF (NP_417378)

Nucleotide sequence (SEQ ID NO:163)

```
   1 atggctatag cacttgtgac tggtggcagt cgcggcatcg ggcgggcaac tgcattactg
```

FIG. 10 (Cont)

```
 61 ttggcgcaag aagggtatac ggtggcggtt aattatcagc aaaacctcca cgcggcgcag
121 gaagtgatga acttaataac gcaagccggt ggcaaagcat tcgtgctcca ggcggatatc
181 agcgacgaaa accaggtcgt tgcgatgttt acagcaatcg atcagcacga tgaaccgcta
241 gcagcgctgg tcaataacgc cgggatcttg tttacccagt gcaccgttga aaaccttacc
301 gcagagcgaa tcaaccgagt actttccacc aacgtgacgg gatattttct ctgctgccgc
361 gaggcggtaa aacgcatggc gcttaaaaat ggtggcagtg gcggcgctat cgtcaatgtc
421 tcttcggtgg cctcacggtt gggttcgcca ggggaatatg ttgattacgc ggcatcgaaa
481 ggggcgattg atacgttaac caccggacta tcgctggaag tcgccgcgca ggggatccgc
541 gttaactgcg tgcggccagg gtttatttat accgaaatgc acgccagcgg cggcgagcct
601 ggacgcgtcg atcgcgttaa gtcgaacatc cccatgcagc gtggtggaca ggcagaagag
661 gtcgcgcagg ccattgtctg gctactaagt gataaagcct cttacgtcac gggaagtttt
721 atcgatttgg cgggcgggaa ataa
```

Amino acid sequence (SEQ ID NO:164)

```
  1 MAIALVTGGS RGIGRATALL LAQEGYTVAV NYQQNLHAAQ EVMNLITQAG GKAFVLQADI
 61 SDENQVVAMF TAIDQHDEPL AALVNNAGIL FTQCTVENLT AERINRVLST NVTGYFLCCR
121 EAVKRMALKN GGSGGAIVNV SSVASRLGSP GEYVDYAASK GAIDTLTTGL SLEVAAQGIR
181 VNCVRPGFIY TEMHASGGEP GRVDRVKSNI PMQRGGQAEE VAQAIVWLLS DKASYVTGSF
241 IDLAGGK
```

YghA (NP_417476)

Nucleotide sequence (SEQ ID NO:165)

```
  1 atgtctcatt taaaagaccc gaccacgcag tattacactg gtgaatatcc caaacagaaa
 61 caaccgacgc caggcatcca ggcgaagatg acaccggtac cggattgcgg cgagaaaacc
121 tatgttggta gcggtcgcct gaaagatcgt aaagcactgg tgacaggggg cgattccgga
181 ataggtcgcg ctgccgccat cgcttacgcg cgtgaagggg ctgacgtggc gatcagttat
241 cttcccgtgg aagaagaaga cgctcaggat gtgaaaaaga tcattgaaga atgcggacgc
301 aaagccgttc tgctgccagg cgatttaagc gatgagaaat ttgcccgttc gctggttcac
361 gaagcgcaca aggcgttagg cgggctggat attatggcgc tggtcgccgg gaaacaggtt
421 gccattccgg atattgcaga cctcaccagc gaacagtttc aaaagacctt tgccattaac
481 gttttcgcgc tgttctggct aacccaggaa gcgatccccc tgctaccgaa aggtgcaagt
541 atcatcacca cttcgtcaat ccaggcatac cagccaagtc cgcatttact ggactatgcg
601 gctacgaagg cggcgattct gaactacagc cgtggcttgg caaaacaggt cgcggagaaa
661 ggtattcggg tgaatattgt cgcgccaggc ccgatctgga cagcactgca aatttccggc
721 ggacaaacgc aggataagat cccgcagttt ggtcagcaaa cgccgatgaa acgtgcgggg
781 caaccggcgg aactggcccc tgtatatgtt tatctggcaa gtcaggagtc gagctacgtc
841 accgcagaag tgcacggcgt gtgcggcggc gagcatttag gttaa
```

Amino acid sequence (SEQ ID NO:166)

```
  1 MSHLKDPTTQ YYTGEYPKQK QPTPGIQAKM TPVPDCGEKT YVGSGRLKDR KALVTGGDSG
```

FIG. 10 (Cont)

```
 61 IGRAAAIAYA REGADVAISY LPVEEEDAQD VKKIIEECGR KAVLLPGDLS DEKFARSLVH
121 EAHKALGGLD IMALVAGKQV AIPDIADLTS EQFQKTFAIN VFALFWLTQE AIPLLPKGAS
181 IITTSSIQAY QPSPHLLDYA ATKAAILNYS RGLAKQVAEK GIRVNIVAPG PIWTALQISG
241 GQTQDKIPQF GQQTPMKRAG QPAELAPVYV YLASQESSYV TAEVHGVCGG EHLG
```

Yjgl (NP_418670)

Nucleotide sequence (SEQ ID NO:167)

```
   1 atgggcgctt ttacaggtaa gacagttctc atcctcggtg gcagtcgtgg tatcggtgcc
  61 gctatcgtac gtcgtttcgt caccgatggg gccaatgtac gattcaccta tgcggggtcg
 121 aaagatgccg ctaaacgcct ggcacaagag actggagcga cagcagtatt cacagatagt
 181 gctgacagag acgctgtcat tgatgtcgtt cgtaagagcg gcgcattgga tatcctggtg
 241 gtaaatgcag gtattggcgt ctttggcgag gccctggaat taaatgccga cgatattgat
 301 cgccttttca aaatcaatat tcatgctcct tatcatgcct ctgttgaagc cgcccggcag
 361 atgcccgaag gcgggcgcat cttaatcatc ggctccgtga atggcgatcg tatgcctgtt
 421 gcaggcatgg ctgcttatgc cgccagcaaa tctgccctgc aaggcatggc gcgcgggctg
 481 gcccgtgatt ttggaccgcg tgggatcacc attaacgtcg tccagccagg gccaattgat
 541 accgacgcta atcccgccaa cgggccaatg cgcgatatgt tgcatagttt gatggctatc
 601 aaaagacatg ggcaaccgga agaggtcgct ggtatggtcg catggttagc agggccagaa
 661 gccagttttg ttaccggcgc gatgcatacc attgatggcg cgtttggcgc ataa
```

Amino acid sequence (SEQ ID NO:168)

```
  1 MGAFTGKTVL ILGGSRGIGA AIVRRFVTDG ANVRFTYAGS KDAAKRLAQE TGATAVFTDS
 61 ADRDAVIDVV RKSGALDILV VNAGIGVFGE ALELNADDID RLFKINIHAP YHASVEAARQ
121 MPEGGRILII GSVNGDRMPV AGMAAYAASK SALQGMARGL ARDFGPRGIT INVVQPGPID
181 TDANPANGPM RDMLHSLMAI KRHGQPEEVA GMVAWLAGPE ASFVTGAMHT IDGAFGA
```

YdfG (NP_416057)

Nucleotide sequence (SEQ ID NO:169)

```
   1 atgatcgttt tagtaactgg agcaacggca ggttttggtg aatgcattac tcgtcgtttt
  61 attcaacaag ggcataaagt tatcgccact ggccgtcgcc aggaacggtt gcaggagtta
 121 aaagacgaac tgggagataa tctgtatatc gcccaactgg acgttcgcaa ccgcgccgct
 181 attgaagaga tgctggcatc gcttcctgcc gagtggtgca atattgatat cctggtaaat
 241 aatgccggcc tggcgttggg catggagcct gcgcataaag ccagcgttga agactgggaa
 301 acgatgattg ataccaacaa caaaggcctg gtatatatga cgcgcgccgt cttaccgggt
 361 atggttgaac gtaatcatgg tcatattatt aacattggct caacggcagg tagctggccg
 421 tatgccggtg gtaacgttta cggtgcgacg aaagcgtttg ttcgtcagtt tagcctgaat
 481 ctgcgtacgg atctgcatgg tacggcggtg cgcgtcaccg acatcgaacc gggtctggtg
 541 ggtggtaccg agttttccaa tgtccgcttt aaaggcgatg acggtaaagc agaaaaaacc
```

FIG. 10 (Cont)

601 tatcaaaata ccgttgcatt gacgccagaa gatgtcagcg aagccgtctg gtgggtgtca
661 acgctgcctg ctcacgtcaa tatcaatacc ctggaaatga tgccggttac ccaaagctat
721 gccggactga atgtccaccg tcagtaa

Amino acid sequence (SEQ ID NO:170)

1 MIVLVTGATA GFGECITRRF IQQGHKVIAT GRRQERLQEL KDELGDNLYI AQLDVRNRAA
61 IEEMLASLPA EWCNIDILVN NAGLALGMEP AHKASVEDWE TMIDTNNKGL VYMTRAVLPG
121 MVERNHGHII NIGSTAGSWP YAGGNVYGAT KAFVRQFSLN LRTDLHGTAV RVTDIEPGLV
181 GGTEFSNVRF KGDDGKAEKT YQNTVALTPE DVSEAVWWVS TLPAHVNINT LEMMPVTQSY
241 AGLNVHRQ

YgcW (NP_417254)

Nucleotide sequence (SEQ ID NO:171)

1 atgtcaatcg aatctctcaa tgcgttctca atggattttt tctccctgaa aggtaaaacc
61 gcaattgtta ccggtgggaa tagcggttta ggccaggcat ttgccatggc gttggccaaa
121 gctggcgcaa atatctttat tcctagtttc gtcaaagata acggcgaaac aaaggaaatg
181 attgaaaaac agggtgttga ggtggacttc atgcaggtgg gtatcaccgc agaaggcgcg
241 ccgcagaaga ttatcgctgc ttgctgtgag cgtttcggta cagttgatat tctggttaac
301 aatgccggta tttgtaagct gaataaggtg ctggacttcg gtcgtgccga ctgggatccg
361 atgattgatg tgaacctgac cgccgcattc gagttaagct atgaagctgc aaaaattatg
421 atcccgcaga aaagcggcaa aattattaat atctgttcat tgttctctta cttaggtgga
481 caatggtcac ctgcatattc tgccactaaa catgctcttg ccgggttcac caaagcttat
541 tgtgatgaac taggtcaata taatattcag gtaaatggta tcgcccctgg ctattatgca
601 acagatatta cgctggcgac acgcagtaat ccagaaacca atcagcgcgt tcttgatcat
661 attccggcaa accgttgggg cgatactcag gatttaatgg gcgcagccgt attcctcgca
721 agtccggcat cgaattatgt caacgggcat ttattagtgg ttgatggcgg ttatttagtg
781 cgctaa

Amino acid sequence (SEQ ID NO:172)

1 MSIESLNAFS MDFFSLKGKT AIVTGGNSGL GQAFAMALAK AGANIFIPSF VKDNGETKEM
61 IEKQGVEVDF MQVGITAEGA PQKIIAACCE RFGTVDILVN NAGICKLNKV LDFGRADWDP
121 MIDVNLTAAF ELSYEAAKIM IPQKSGKIIN ICSLFSYLGG QWSPAYSATK HALAGFTKAY
181 CDELGQYNIQ VNGIAPGYYA TDITLATRSN PETNQRVLDH IPANRWGDTQ DLMGAAVFLA
241 SPASNYVNGH LLVVDGGYLV R

UcpA(NP_416921)

Nucleotide sequence (SEQ ID NO:173)

1 tcagataccg acgctaaccg tctccggcag tgtgctgccg ccatcaatca cattctgtgt

FIG. 10 (Cont)

```
 61 accggttaaa tagctggatt catccgatgc gaggaaggcc gccagttcgc cgacttccag
121 cggatcggcg aggcgacgca tcgggattgc tttcgccatt tcagtcagca ccgactctgg
181 atcttccggg ttcgactggc gggcaatgct ttccgccatt ggtgtgcgca cgtatcccgg
241 gcaaatggcg ttaacgcgaa taccagactg cgcgtactcc accgccagcg attttgtcag
301 gccaacaatc gccgctttcg ttaaggcgta cgccgtttcg ccaggatcgg ccaccatatc
361 accagtgact gaagacatca tcacaatgcg accatctttg cgggcaatca tctccggcag
421 caccgccttc gtgacgttcc atacgccttt aatattgatg tcaatatgga aatcgcgatc
481 gtcatcgctc atatcgagga aactgcccag acgacaaacg cctgcgttat tcaccaggat
541 atcaatgcgc ccttcttttt ccttcgcgcg tttgatagct gcggctaccg acgccgggtc
601 acgcacatcg gcgacaaccg ccgtacagcg atgaccacga ccacacagtt cgtccgccag
661 cttttcgatc tcaggggaga tatccagcaa gattaggttc gcgccatgac gtgcaaaagt
721 tctggcaatt ccttcgccaa ttccctgcaa tgcgcccgta atcagtgctg tcttgcccgt
781 gagtttaccc at
```

Amino acid sequence (SEQ ID NO:174)

```
  1 MGKLTGKTAL ITGALQGIGE GIARTFARHG ANLILLDISP EIEKLADELC GRGHRCTAVV
 61 ADVRDPASVA AAIKRAKEKE GRIDILVNNA GVCRLGSFLD MSDDDRDFHI DINIKGVWNV
121 TKAVLPEMIA RKDGRIVMMS SVTGDMVADP GETAYALTKA AIVGLTKSLA VEYAQSGIRV
181 NAICPGYVRT PMAESIARQS NPEDPESVLT EMAKAIPMRR LADPLEVGEL AAFLASDESS
241 YLTGTQNVID GGSTLPETVS VGI
```

EntA (NP_415128)

Nucleotide sequence (SEQ ID NO:175)

```
  1 atggatttca gcggtaaaaa tgtctgggta accggcgcag gtaaaggtat cggctacgcc
 61 acggcgctgg cgtttgttga ggcgggagcg aaagttacag gttttgatca agcgttcact
121 caggagcaat atccctttgc gaccgaagtg atggatgttg ccgacgctgc gcaggtcgcg
181 caagtgtgtc agcgactgtt agctgaaacg gagcgactgg acgcgctggt caatgcggcg
241 ggaattttac gcatgggcgc gaccgatcag ctcagtaaag aggactggca gcagactttt
301 gcggttaacg tcggcggtgc gtttaacctg ttccagcaaa ccatgaacca gtttcgccgt
361 cagcggggcg gggcgattgt cactgtggcg tccgacgccg cgcacacgcc gcgtattggc
421 atgagtgctt atggcgcatc gaaagcggcg ctgaaaagcc tggcgttgag cgtcgggctg
481 gaactggcgg gtagcggcgt gcgctgtaat gtggtttcgc ctggctccac cgacaccgat
541 atgcaacgca cgctgtgggt gagcgatgac gccgaagaac agcgtattcg cggctttggc
601 gagcagttta aactcggcat tccgctgggg aaaatcgccc gtccacaaga gatcgccaac
661 acgattttgt tcctcgcctc tgacctcgcc agccatatta ccctacagga tattgtggtc
721 gatggcggct caacgctggg ggcataa
```

Amino acid sequence (SEQ ID NO:176)

```
  1 MDFSGKNVWV TGAGKGIGYA TALAFVEAGA KVTGFDQAFT QEQYPFATEV MDVADAAQVA
```

FIG. 10 (Cont)

```
 61 QVCQRLLAET ERLDALVNAA GILRMGATDQ LSKEDWQQTF AVNVGGAFNL FQQTMNQFRR
121 QRGGAIVTVA SDAAHTPRIG MSAYGASKAA LKSLALSVGL ELAGSGVRCN VVSPGSTDTD
181 MQRTLWVSDD AEEQRIRGFG EQFKLGIPLG KIARPQEIAN TILFLASDLA SHITLQDIVV
241 DGGSTLGA
```

FolM (NP_416123)

Nucleotide sequence (SEQ ID NO:177)

```
  1 atgggtaaaa cccagccctt gccaatatta attactggcg gaggtcgtcg catcggcctc
 61 gccctcgcat ggcatttcat taatcaaaag caaccggtga ttgtcagcta tcggacacac
121 tatccagcca ttgatggact gattaatgca ggtgcgcagt gtattcaggc tgatttttcg
181 accaacgacg gtgtgatggc gtttgccgat gaagtactaa aaagcaccca tggtctgcgt
241 gctattttgc ataacgccag tgcgtggatg gcggaaaaac cgggtgcgcc actggccgac
301 gtactggctt gcatgatgca gatccacgtt aataccccat acctgctcaa ccatgcgctg
361 gaaagattac tgcgtgggca cggacacgcc gccagcgata tcattcactt taccgattat
421 gtggtggagc gcggtagcga caaacatatt gcgtatgctg caagcaaagc ggcactggat
481 aatatgaccc gctcgtttgc ccgcaagctg gcaccggaag tgaaagtgaa ttctattgcg
541 ccatcgctga tcctgtttaa tgaacatgat gatgccgaat atcgacaaca ggcgctgaat
601 aaatcactga tgaaaaccgc gcctggcgag aaagaagtga tcgacctggt cgattactta
661 cttaccagtt gctttgtcac cggacgcagt ttcccacttg atggcggtcg tcatctgcgt
721 taa
```

Amino acid sequence (SEQ ID NO:178)

```
  1 MGKTQPLPIL ITGGGRRIGL ALAWHFINQK QPVIVSYRTH YPAIDGLINA GAQCIQADFS
 61 TNDGVMAFAD EVLKSTHGLR AILHNASAWM AEKPGAPLAD VLACMMQIHV NTPYLLNHAL
121 ERLLRGHGHA ASDIIHFTDY VVERGSDKHI AYAASKAALD NMTRSFARKL APEVKVNSIA
181 PSLILFNEHD DAEYRQQALN KSLMKTAPGE KEVIDLVDYL LTSCFVTGRS FPLDGGRHLR
```

HdhA (NP_416136)

Nucleotide sequence (SEQ ID NO:179)

```
  1 gtgtttaatt ctgacaacct gagactcgac ggaaaatgcg ccatcatcac aggtgcgggt
 61 gcaggtattg gtaaagaaat cgccattaca ttcgcgacag ctggcgcatc tgtggtggtc
121 agtgatatta acgccgacgc agctaaccat gttgtagacg aaattcaaca actgggtggt
181 caggcatttg cctgccgttg tgatattact tccgaacagg aactctctgc actggcagac
241 tttgctatca gtaagctggg taaagttgat attctggtta acaacgccgg tggcggtgga
301 cctaaaccgt ttgatatgcc aatggcggat tttcgccgtg cttatgaact gaatgtgttt
361 tctttttcc atctgtcaca acttgttgcg ccagaaatgg aaaaaaatgg cggtggcgtt
421 attctgacca tcacttctat ggcggcagaa ataaaaata taaacatgac ttcctatgca
481 tcatctaaag ctgcggccag tcatctggtc agaaatatgg cgtttgacct gggtgaaaaa
541 aatattcggg taaatggcat tgcgccgggg gcaatattaa ccgatgccct gaaatccgtt
```

FIG. 10 (Cont)

601 attacaccag aaattgaaca aaaaatgtta cagcacacgc cgatcagacg tctgggccaa
661 ccgcaagata ttgctaacgc agcgctgttc ctttgctcgc ctgctgcgag ctgggtaagc
721 ggacaaattc tcaccgtctc cggtggtggg gtacaggagc tcaattaa

Amino acid sequence (SEQ ID NO:180)

1 MFNSDNLRLD GKCAIITGAG AGIGKEIAIT FATAGASVVV SDINADAANH VVDEIQQLGG
61 QAFACRCDIT SEQELSALAD FAISKLGKVD ILVNNAGGGG PKPFDMPMAD FRRAYELNVF
121 SFFHLSQLVA PEMEKNGGGV ILTITSMAAE NKNINMTSYA SSKAAASHLV RNMAFDLGEK
181 NIRVNGIAPG AILTDALKSV ITPEIEQKML QHTPIRRLGQ PQDIANAALF LCSPAASWVS
241 GQILTVSGGG VQELN

HcaB (NP_417036)

Nucleotide sequence (SEQ ID NO:181)

1 atgagcgatc tgcataacga gtccattttt attaccggcg gcggatcggg attagggctg
61 gcgctggtcg agcgatttat cgaagaaggc gcgcaggttg ccacgctgga actgtcggcg
121 gcaaaagtcg ccagtctgcg tcagcgattt ggcgaacata ttctggcggt ggaaggtaac
181 gtgacctgtt atgccgatta tcaacgcgcg gtcgatcaga tcctgactcg ttccggcaag
241 ctggattgtt ttatcggcaa tgcaggcatc tgggatcaca atgcctcact ggttaatact
301 cccgcagaga cgctcgaaac cggcttccac gagctgttta acgtcaatgt tctcggttac
361 ctgctgggcg caaaagcctg cgctccggcg ttaatcgcca gtgaaggcag catgattttc
421 acactgtcaa atgccgcctg gtatcctggc ggcggtggcc cgctgtacac cgccagtaaa
481 catgccgcaa ccggacttat tcgccaactg gcttatgaac tggcaccgaa agtgcgggtg
541 aatggcgtcg gcccgtgtgg tatggccagc gacctgcgcg gcccacaggc gctcgggcaa
601 agtgaaacct cgataatgca gtctctgacg ccggagaaaa ttgccgccat tttaccgctg
661 caatttttcc cgcaaccggc ggattttacg gggccgtatg tgatgttgac atcgcggcgc
721 aataatcgcg cattaagcgg tgtgatgatc aacgctgatg cgggtttagc gattcgcggc
781 attcgccacg tagcggctgg gctggatctt taa

Amino acid sequence (SEQ ID NO:182)

1 MSDLHNESIF ITGGGSGLGL ALVERFIEEG AQVATLELSA AKVASLRQRF GEHILAVEGN
61 VTCYADYQRA VDQILTRSGK LDCFIGNAGI WDHNASLVNT PAETLETGFH ELFNVNVLGY
121 LLGAKACAPA LIASEGSMIF TLSNAAWYPG GGGPLYTASK HAATGLIRQL AYELAPKVRV
181 NGVGPCGMAS DLRGPQALGQ SETSIMQSLT PEKIAAILPL QFFPQPADFT GPYVMLTSRR
241 NNRALSGVMI NADAGLAIRG IRHVAAGLDL

SrlD (NP_417185)

Nucleotide sequence (SEQ ID NO:183)

1 atgaatcagg ttgccgttgt catcggtggt gggcaaacct taggcgcgtt cctgtgccac

FIG. 10 (Cont)

```
 61 ggtctggctg ccgaggggta tcgcgtcgcg gttgtcgata ttcagagcga caaagccgca
121 aatgtggcac aagaaattaa cgccgaatat ggtgaaagta tggcgtacgg ttttggtgct
181 gacgccacta gcgagcaaag cgttctggcg ctctctcgtg gggtagatga aatctttggt
241 cgcgtggatt tgctggtcta cagcgccgga atagcaaaag cagcctttat cagcgacttc
301 cagctcggcg attttgaccg ttcgctacag gtgaatctgg tgggttattt cctgtgtgcg
361 cgtgaatttt cgcgtttgat gatccgcgac gggattcagg ggcgcattat tcagatcaac
421 tcgaaatccg gcaaagtggg cagcaaacac aactctggct acagcgcagc gaaatttggt
481 ggcgtcgggc tgactcaatc actggcgctg gatctggcgg agtacggcat tacggtgcat
541 tcactgatgc tcggtaacct gctgaaatcg ccgatgttcc agtcactgtt gccacaatac
601 gcgaccaagc tgggtatcaa accggatcaa gtcgagcagt attacatcga caaagtaccg
661 ctcaaacgcg gctgcgatta tcaagatgtg ctgaatatgc tgctgttcta cgccagtcct
721 aaggcgtcgt actgcaccgg acagtcgatc aatgtcaccg gcggtcaggt gatgttctga
```

Amino acid sequence (SEQ ID NO:184)

```
  1 MNQVAVVIGG GQTLGAFLCH GLAAEGYRVA VVDIQSDKAA NVAQEINAEY GESMAYGFGA
 61 DATSEQSVLA LSRGVDEIFG RVDLLVYSAG IAKAAFISDF QLGDFDRSLQ VNLVGYFLCA
121 REFSRLMIRD GIQGRIIQIN SKSGKVGSKH NSGYSAAKFG GVGLTQSLAL DLAEYGITVH
181 SLMLGNLLKS PMFQSLLPQY ATKLGIKPDQ VEQYYIDKVP LKRGCDYQDV LNMLLFYASP
241 KASYCTGQSI NVTGGQVMF
```

KduD (NP_417319)

Nucleotide sequence (SEQ ID NO:185)

```
  1 atgattttaa gtgcattttc tctcgaaggt aaagttgcgg tcgtcactgg ttgtgatact
 61 ggactgggtc aggggatggc gttggggctg gcgcaagcgg gctgtgacat tgttggcatt
121 aacatcgttg aaccgactga aaccatcgag caggtcacag cgctggggcg tcgttttta
181 agcctgaccg ccgatctgcg aaagattgat ggtattccag cactgctgga tcgcgcggta
241 gcggagtttg gtcatattga tatcctggtg aataacgccg gattgattcg ccgcgaagat
301 gctctcgagt tcagcgaaaa ggactgggac gatgtcatga acctgaatat caagagcgta
361 ttcttcatgt ctcaggcagc ggcgaaacac tttatcgcgc aaggcaatgg cggcaagatt
421 atcaatatcg cgtcaatgct ctccttccag ggcgggatcc gtgtgccttc ttataccgca
481 tcaaaaagcg gcgtgatggg tgtgacgcga ttgatggcga acgaatgggc taaacacaac
541 attaatgtta atgcgatagc cccgggttac atggcgacca acaatactca acaactacgg
601 gcagatgaac aacgtagcgc ggaaattctc gaccgcattc cagctggtcg ttggggactg
661 ccgagtgacc tgatggggcc gatagtgttc cttgcctcca gcgcttcaga ttatgtgaat
721 ggttatacca ttgccgtgga tggcggttgg ctggcgcgtt aa
```

Amino acid sequence (SEQ ID NO:186)

```
  1 MILSAFSLEG KVAVVTGCDT GLGQGMALGL AQAGCDIVGI NIVEPTETIE QVTALGRRFL
 61 SLTADLRKID GIPALLDRAV AEFGHIDILV NNAGLIRRED ALEFSEKDWD DVMNLNIKSV
121 FFMSQAAAKH FIAQGNGGKI INIASMLSFQ GGIRVPSYTA SKSGVMGVTR LMANEWAKHN
181 INVNAIAPGY MATNNTQQLR ADEQRSAEIL DRIPAGRWGL PSDLMGPIVF LASSASDYVN
241 GYTIAVDGGW LAR
```

FIG. 10 (Cont)

IdnO (NP_418687)

Nucleotide sequence (SEQ ID NO:187)

```
   1 atgaacgatc tattttcact ggcaggaaaa aatatcttga ttaccggttc agcacagggc
  61 attggctttt tactggcaac cggcctgggt aaatatggcg cacaaataat tattaatgat
 121 attactgccg aacgcgcaga acttgctgta gaaaaactcc accaggaggg tattcaggcc
 181 gttgccgcac cttttaatgt tactcataaa catgaaattg atgccgccgt tgaacatatc
 241 gaaaaggaca tcggccccat tgatgtgctg gtgaataacg ccggtatcca gcgccgtcat
 301 ccttttactg agttccctga acaagagtgg aatgatgtga tcgcagtaaa ccagaccgcc
 361 gtgttcctgg tatcgcaagc ggtaactcgt cacatggttg aacgcaaggc aggtaaagtt
 421 attaatattt gctcgatgca aagcgaactg ggacgtgaca ccatcacccc ttatgccgca
 481 tcgaaagggg cggtaaaaat gctcacccgc ggcatgtgcg tcgagctggc gcgccacaat
 541 attcaggtca acggtattgc gccgggctat ttcaaaacag aaatgactaa agcactggtt
 601 gaggacgaag ccttcaccgc ctggttgtgc aaacggaccc ccgccgcacg ctggggagat
 661 ccgcaggaac tgattggtgc tgcggtgttc ctttcttcaa aagcctctga tttcgtaaac
 721 ggccacctgt tgtttgttga tggcggcatg ttagtggctg tttaa
```

Amino acid sequence (SEQ ID NO:188)

```
   1 MNDLFSLAGK NILITGSAQG IGFLLATGLG KYGAQIIIND ITAERAELAV EKLHQEGIQA
  61 VAAPFNVTHK HEIDAAVEHI EKDIGPIDVL VNNAGIQRRH PFTEFPEQEW NDVIAVNQTA
 121 VFLVSQAVTR HMVERKAGKV INICSMQSEL GRDTITPYAA SKGAVKMLTR GMCVELARHN
 181 IQVNGIAPGY FKTEMTKALV EDEAFTAWLC KRTPAARWGD PQELIGAAVF LSSKASDFVN
 241 GHLLFVDGGM LVAV
```

FabG (NP_415611)

Nucleotide sequence (SEQ ID NO:189)

```
   1 atgaattttg aaggaaaaat cgcactggta accggtgcaa gccgcggaat tggccgcgca
  61 attgctgaaa cgctcgcagc ccgtggcgcg aaagttattg gcactgcgac cagtgaaaat
 121 ggcgctcagg cgatcagtga ttatttaggt gccaacggca aaggtctgat gttgaatgtg
 181 accgacccgg catctatcga atctgttctg gaaaaaattc gcgcagaatt tggtgaagtg
 241 gatatcctgg tcaataatgc cggtatcact cgtgataacc tgttaatgcg aatgaaagat
 301 gaagagtgga acgatattat cgaaaccaac ctttcatctg ttttccgtct gtcaaaagcg
 361 gtaatgcgcg ctatgatgaa aaagcgtcat ggtcgtatta tcactatcgg ttctgtggtt
 421 ggtaccatgg gaaatggcgg tcaggccaac tacgctgcgg cgaaagcggg cttgatcggc
 481 ttcagtaaat cactggcgcg cgaagttgcg tcacgcggta ttactgtaaa cgttgttgct
 541 ccgggcttta ttgaaacgga catgacacgt gcgctgagcg atgaccagcg tgcgggtatc
 601 ctggcgcagg ttcctgcggg tcgcctcggc ggcgcacagg aaatcgccaa cgcggttgca
 661 ttcctggcat ccgacgaagc agcttacatc acgggtgaaa ctttgcatgt gaacggcggg
 721 atgtacatgg tctga
```

FIG. 10 (Cont)

Amino acid sequence (SEQ ID NO:190)

```
  1 MNFEGKIALV TGASRGIGRA IAETLAARGA KVIGTATSEN GAQAISDYLG ANGKGLMLNV
 61 TDPASIESVL EKIRAEFGEV DILVNNAGIT RDNLLMRMKD EEWNDIIETN LSSVFRLSKA
121 VMRAMMKKRH GRIITIGSVV GTMGNGGQAN YAAAKAGLIG FSKSLAREVA SRGITVNVVA
181 PGFIETDMTR ALSDDQRAGI LAQVPAGRLG GAQEIANAVA FLASDEAAYI TGETLHVNGG
241 MYMV
```

FabI (NP_415804)

Nucleotide sequence (SEQ ID NO:191)

```
  1 atgggttttc tttccggtaa gcgcattctg gtaaccggtg ttgccagcaa actatccatc
 61 gcctacggta tcgctcaggc gatgcaccgc gaaggagctg aactggcatt cacctaccag
121 aacgacaaac tgaaaggccg cgtagaagaa tttgccgctc aattgggttc tgacatcgtt
181 ctgcagtgcg atgttgcaga agatgccagc atcgacacca tgttcgctga actggggaaa
241 gtttggccga aatttgacgg tttcgtacac tctattggtt ttgcacctgg cgatcagctg
301 gatggtgact atgttaacgc cgttacccgt gaaggcttca aaattgccca cgacatcagc
361 tcctacagct tcgttgcaat ggcaaaagct tgccgctcca tgctgaatcc gggttctgcc
421 ctgctgaccc tttcctacct tggcgctgag cgcgctatcc cgaactacaa cgttatgggt
481 ctggcaaaag cgtctctgga agcgaacgtg cgctatatgg cgaacgcgat gggtccggaa
541 ggtgtgcgtg ttaacgccat ctctgctggt ccgatccgta ctctggcggc ctccggtatc
601 aaagacttcc gcaaaatgct ggctcattgc gaagccgtta ccccgattcg ccgtaccgtt
661 actattgaag atgtgggtaa ctctgcggca ttcctgtgct ccgatctctc tgccggtatc
721 tccggtgaag tggtccacgt tgacggcggt ttcagcattg ctgcaatgaa cgaactcgaa
781 ctgaaataa
```

Amino acid sequence (SEQ ID NO:192)

```
  1 MGFLSGKRIL VTGVASKLSI AYGIAQAMHR EGAELAFTYQ NDKLKGRVEE FAAQLGSDIV
 61 LQCDVAEDAS IDTMFAELGK VWPKFDGFVH SIGFAPGDQL DGDYVNAVTR EGFKIAHDIS
121 SYSFVAMAKA CRSMLNPGSA LLTLSYLGAE RAIPNYNVMG LAKASLEANV RYMANAMGPE
181 GVRVNAISAG PIRTLAASGI KDFRKMLAHC EAVTPIRRTV TIEDVGNSAA FLCSDLSAGI
241 SGEVVHVDGG FSIAAMNELE LK
```

YdjA (NP_416279)

Nucleotide sequence (SEQ ID NO:193)

```
  1 atggatgcac tcgaactatt gatcaatcgc cgtagcgcct cccgcttggc tgaacccgcg
 61 ccaacgggtg aacaactgca aaacatcctg cgtgcgggta tgcgtgcgcc ggaccataag
121 tccatgcaac cgtggcattt ttttgtgatt gaaggggaag ggcgcgagcg tttcagcgcc
181 gtactggaac agggggcgat tgctgccggt agtgatgaca aagctatcga caaagcccgt
241 aatgcgccgt tccgcgcacc gctcatcatc acggtggtgg cgaaatgcga agagaatcat
```

FIG. 10 (Cont)

```
301 aaagtcccgc gctgggaaca ggaaatgtct gccggatgcg cggtcatggc gatgcaaatg
361 gcagcagttg cccaggggtt tggcggcatc tggcgcagtg gcgcattaac tgaaagtccg
421 gtagtgcgtg aagcattcgg ttgccgtgag caggataaaa ttgtcggttt tctctacctc
481 ggtacgccgc agctgaaagc atctacgtcg attaacgtcc cggacccgac gccgtttgta
541 acttatttct ga
```

Amino acid sequence (SEQ ID NO:194)

```
  1 MDALELLINR RSASRLAEPA PTGEQLQNIL RAGMRAPDHK SMQPWHFFVI EGEGRERFSA
 61 VLEQGAIAAG SDDKAIDKAR NAPFRAPLII TVVAKCEENH KVPRWEQEMS AGCAVMAMQM
121 AAVAQGFGGI WRSGALTESP VVREAFGCRE QDKIVGFLYL GTPQLKASTS INVPDPTPFV
181 TYF
```

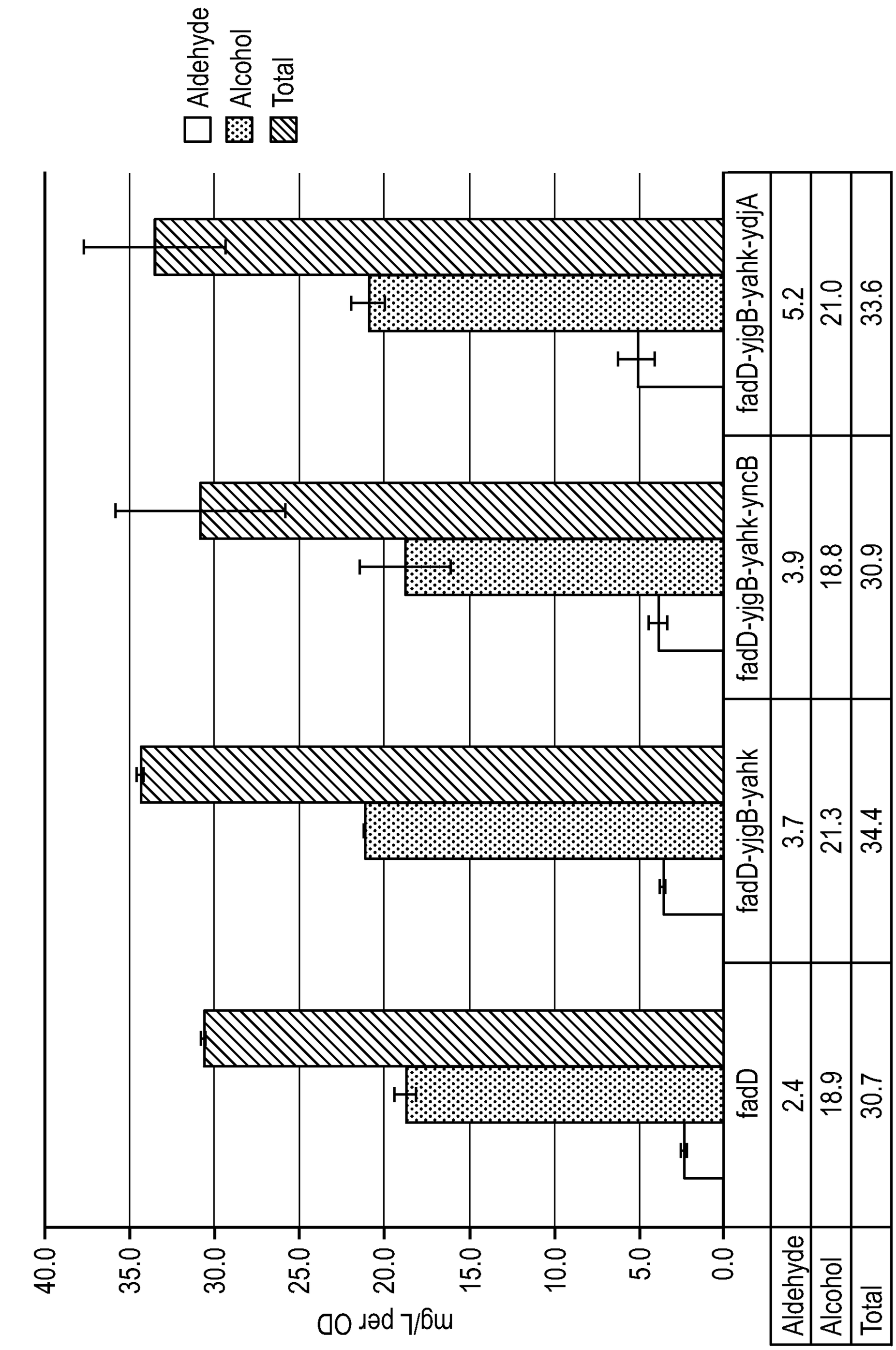
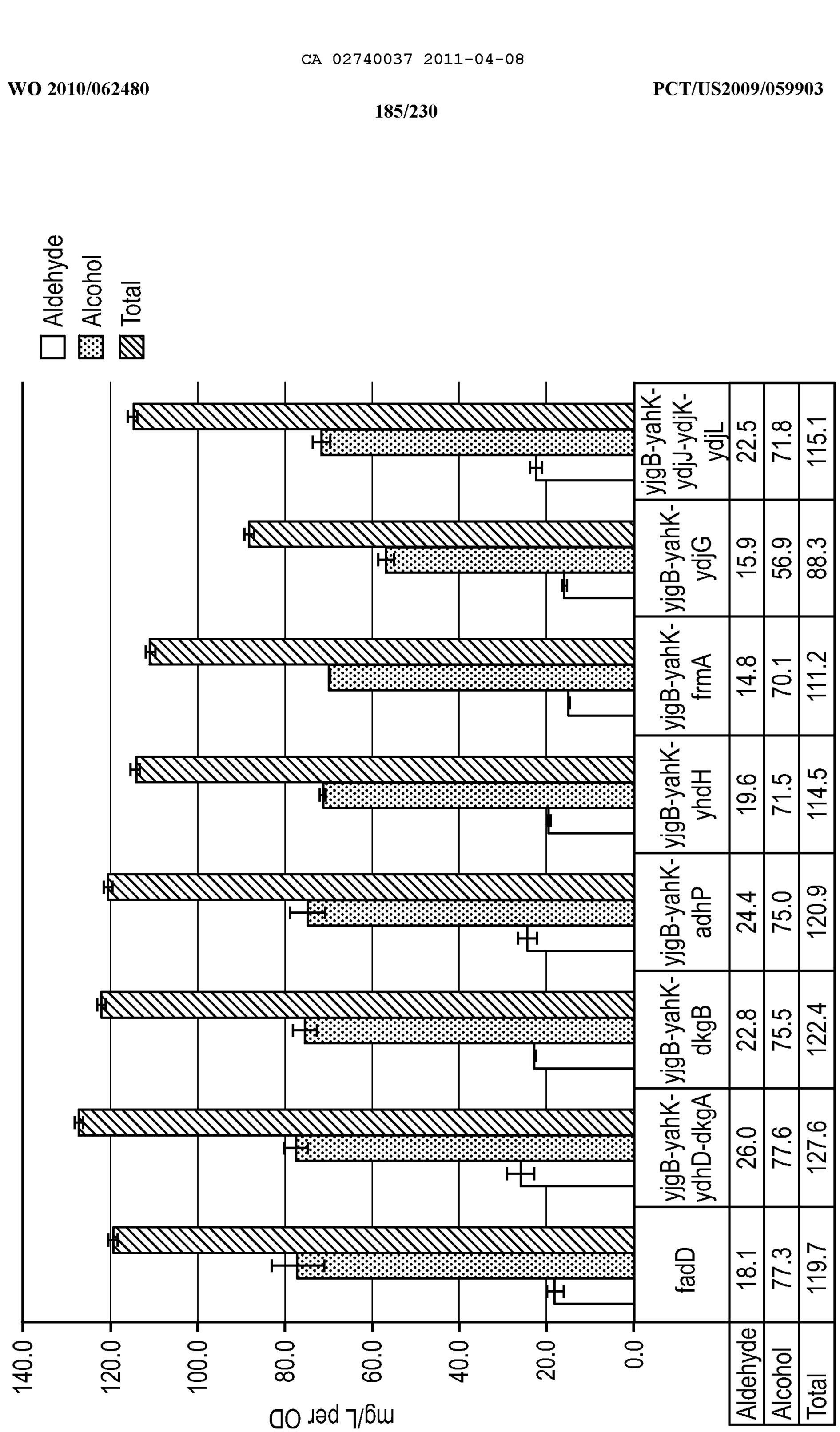

Aldehyde
Alcohol
Total
40.0
35.0
30.0
25.0
20.0
15.0
10.0
5.0
0.0
mg/L per OD
Aldehyde
Alcohol
Total
fadD
2.4
18.9
30.7
fadD-yjgB-yahk
3.7
21.3
34.4
fadD-yjgB-yahk-yncB
3.9
18.8
30.9
fadD-yjgB-yahk-ydjA
5.2
21.0
33.6

FIG. 11



	yjgB-yahK-ydjJ-ydjK-ydjL	yjgB-yahK-ydjG	yjgB-yahK-frmA	yjgB-yahK-yhdH	yjgB-yahK-adhP	yjgB-yahK-dkgB	yjgB-yahK-ydhD-dkgA	fadD
Aldehyde	22.5	15.9	14.8	19.6	24.4	22.8	26.0	18.1
Alcohol	71.8	56.9	70.1	71.5	75.0	75.5	77.6	77.3
Total	115.1	88.3	111.2	114.5	120.9	122.4	127.6	119.7

FIG. 12

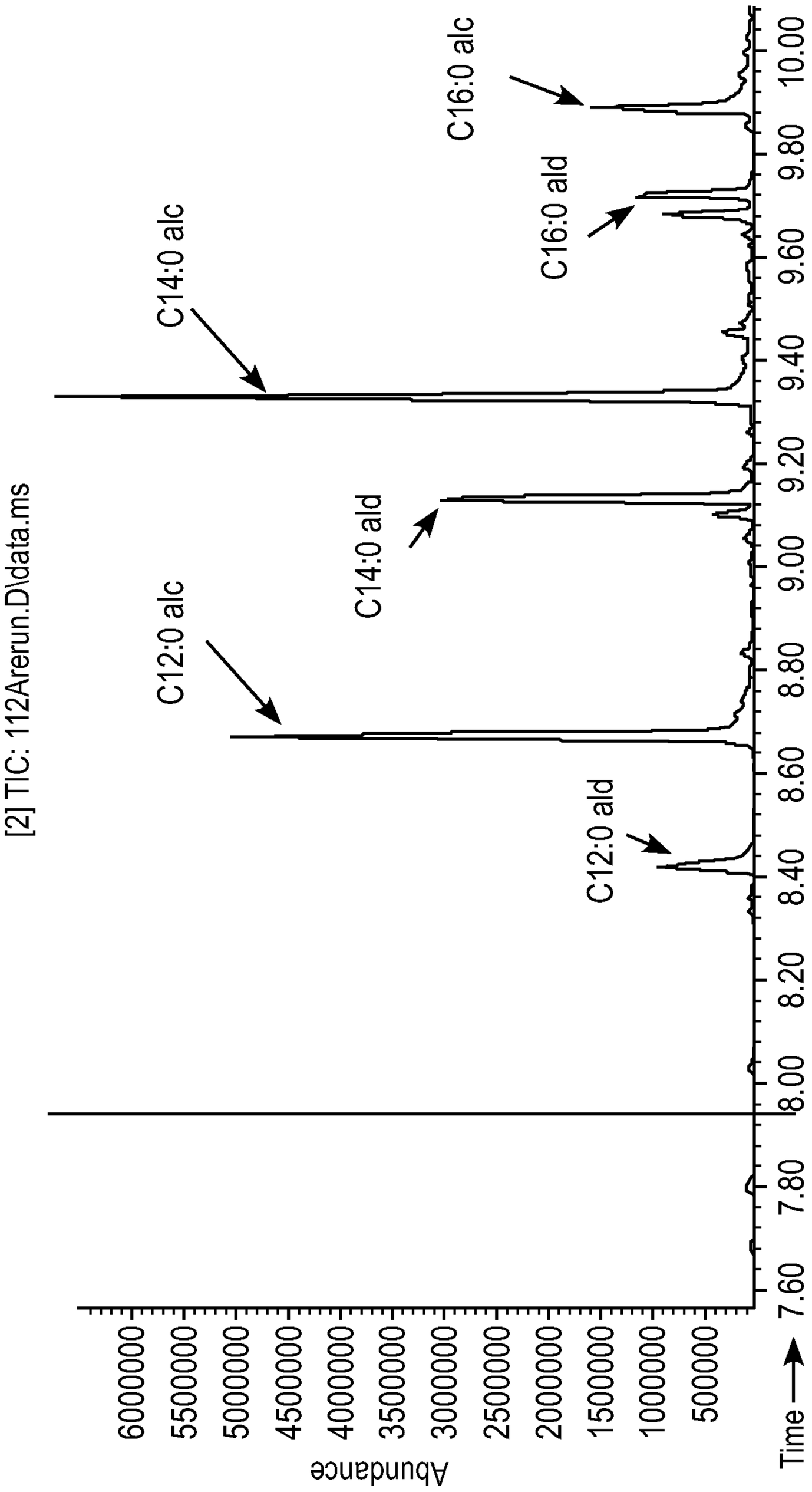

[2] TIC: 112Arerun.D\data.ms
C12:0 alc
C12:0 ald
C14:0 alc
C14:0 ald
C16:0 alc
C16:0 ald
Abundance
Time
6000000
5500000
5000000
4500000
4000000
3500000
3000000
2500000
2000000
1500000
1000000
500000
7.60
7.80
8.00
8.20
8.40
8.60
8.80
9.00
9.20
9.40
9.60
9.80
10.00

FIG. 13

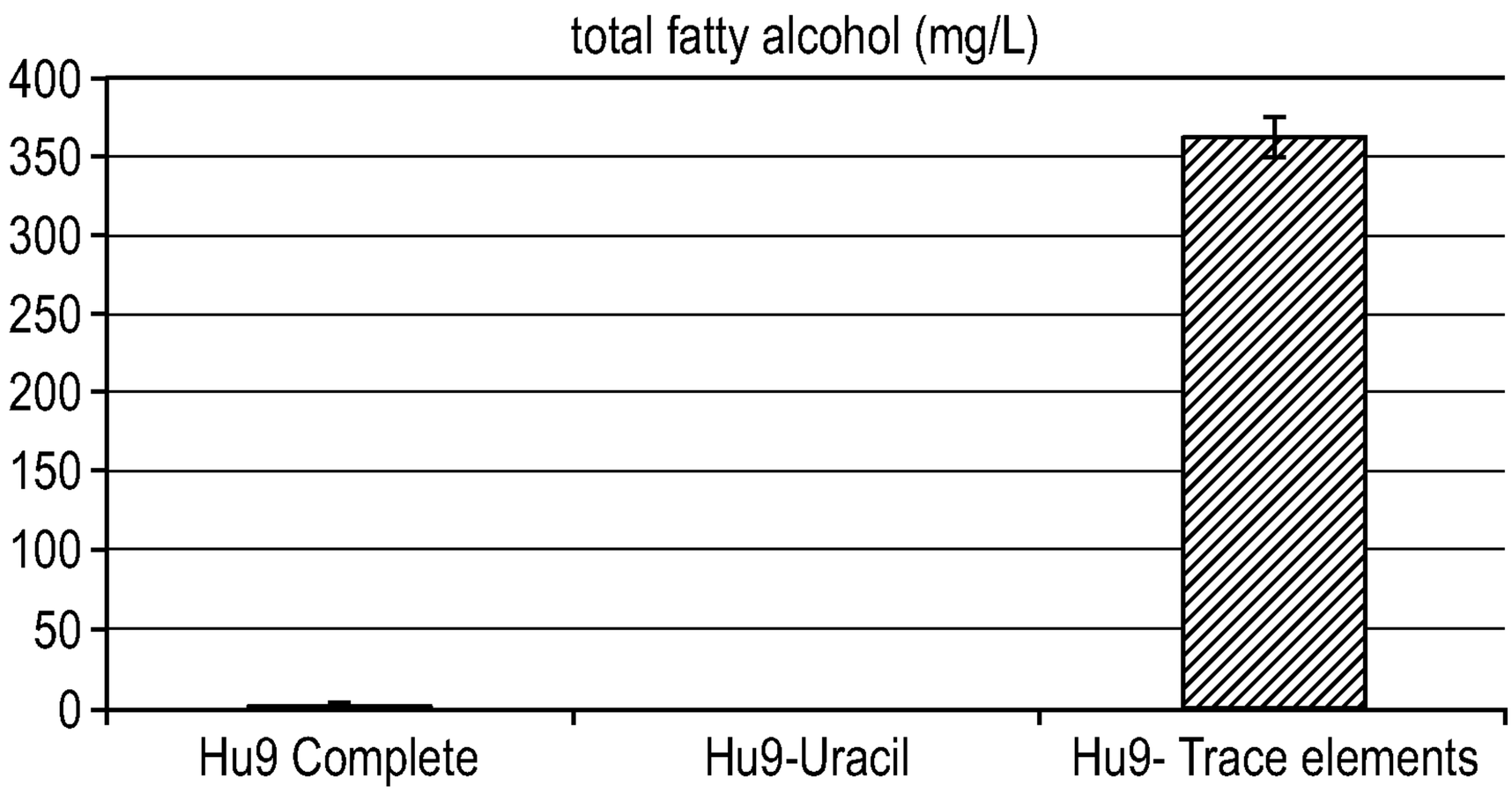


FIG. 14A

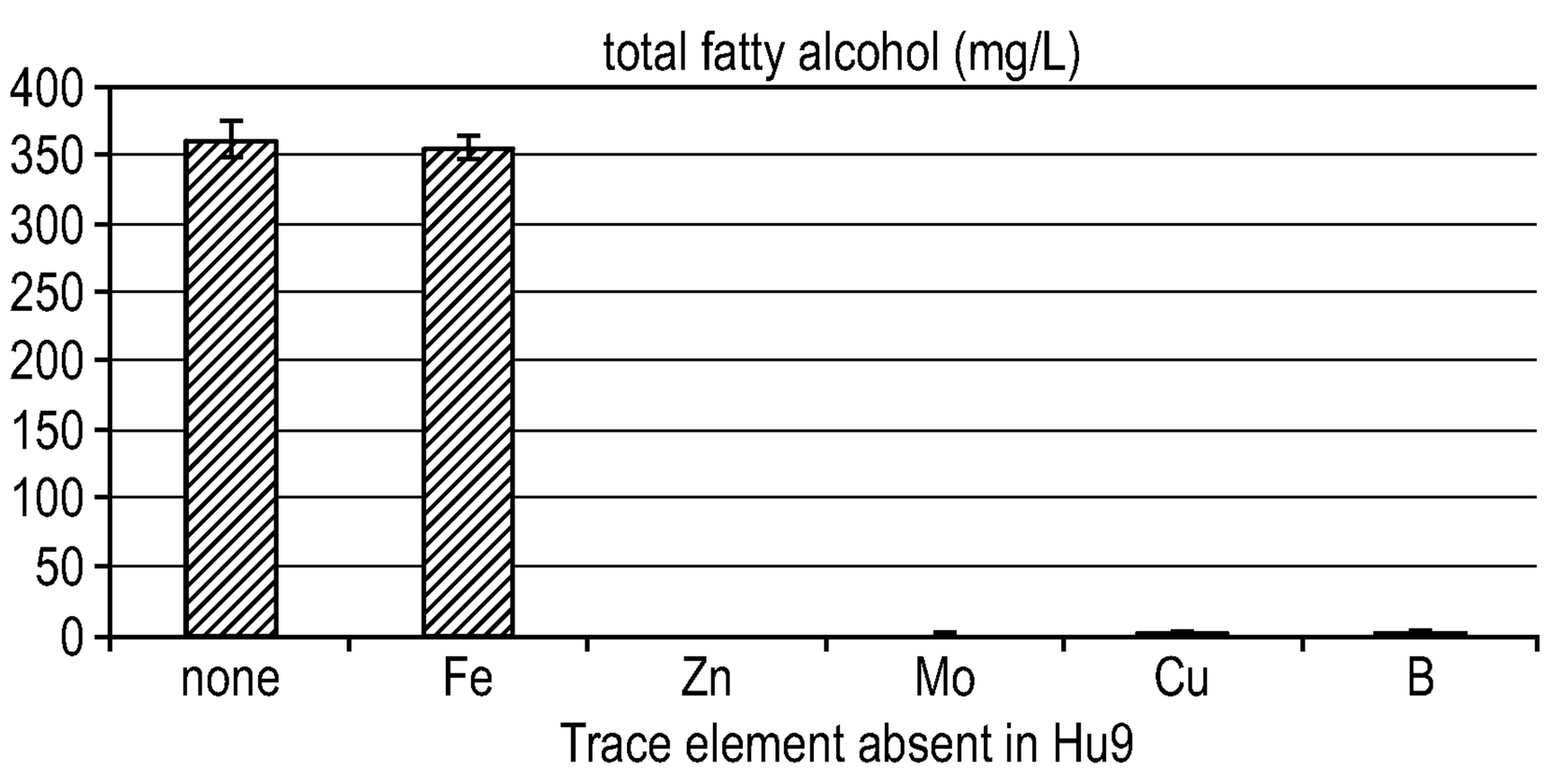


FIG. 14B

FIG. 15

Gene	Name	Nucleotide Sequence	Protein Sequence
fabA	beta-hydroxydecanoyl thioester dehydrase	atgGTAGATA AACGCGAATC CTATACAAAA GAAGACCTTC TTGCCTCTGG TCGCGGTGAA CTGTTTGGCG CTAAAGGCCC GCAATTGCCA GCACCGAACA TGCTGATGAT GGACCGTGTG GTCAAAATGA CCGAAACGGG TGGTAACTTC GACAAAGGGT ATGTTGAAGC AGAACTGGAT ATCAATCCGG ATCTGTGGTT CTTCGGATGC CACTTTATTG GCGATCCGGT TATGCCGGGA TGCCTGGGCC TGGACGCAAT GTGGCAGCTG GTAGGGTTCT ACCTCGGCTG GCTGGGCGGC GAAGGTAAAG GCCGCGCGCT GGGCGTTGGC GAAGTGAAAT TCACTGGTCA GGTACTGCCG ACAGCGAAAA AAGTGACCTA CCGTATTCAC	MVDKRESYTK EDLLASGRGE LFGAKGPQLP APNMLMMDRV VKMTETGGNF DKGYVEAELD INPDLWFFGC HFIGDPVMPG CLGLDAMWQL VGFYLGWLGG EGKGRALGVG EVKFTGQVLP TAKKVTYRIH FKRIVNRRLI MGLADGEVLV DGRLIYTASD LKVGLFQDTS AF (SEQ ID NO:244)

		TTTAAACGCA TTGTTAACCG TCGTCTGATT ATGGGCCTGG CGGATGGCGA AGTGCTGGTT GATGGTCGTC TGATCTATAC CGCCAGCGAC CTGAAAGTCG GTCTGTTCCA GGATACGTCT GCCTTCTGA (SEQ ID NO:243)	
fabZ	(3R)-hydroxymyristol acyl carrier protein dehydratase	ttgACTACTA ACACTCATAC TCTGCAGATT GAAGAGATTT TAGAACTTCT GCCGCACCGT TTCCCGTTCT TACTGGTGGA TCGCGTGCTG GATTTTGAAG AAGGTCGTTT TCTGCGCGCA GTAAAAAATG TCTCTGTCAA TGAGCCATTC TTCCAGGGCC ATTTCCCTGG AAAACCGATT TTCCCGGGTG TGCTGATTCT GGAAGCAATG GCACAGGCAA CAGGTATTCT GGCGTTTAAA AGCGTAGGAA AACTGGAACC GGGTGAGCTG	MTTNTHTLQI EEILELLPHR FPFLLVDRVL DFEEGRFLRA VKNVSVNEPF FQGHFPGKPI FPGVLILEAM AQATGILAFK SVGKLEPGEL YYFAGIDEAR FKRPVVPGDQ MIMEVTFEKT RRGLTRFKGV ALVDGKVVCE ATMMCARSRE A (SEQ ID NO:246)

		TACTACTTCG CTGGTATTGA CGAAGCGCGC TTCAAGCGCC CGGTCGTGCC TGGCGATCAA ATGATCATGG AAGTCACTTT CGAAAAAACG CGCCGCGGCC TGACCCGTTT TAAAGGGGTT GCTCTGGTCG ATGGTAAAGT AGTTTGCGAA GCAACGATGA TGTGTGCTCG TAGCCGGGAG GCCTGA (SEQ ID NO:245)	
cysM	cysteine synthase B (O-acetylserine sulfhydrolase B)	gtgAGTACAT TAGAACAAAC AATAGGCAAT ACGCCTCTGG TGAAGTTGCA GCGAATGGGG CCGGATAACG GCAGTGAAGT GTGGTTAAAA CTGGAAGGCA ATAACCCGGC AGGTTCGGTG AAAGATCGTG CGGCACTTTC GATGATCGTC GAGGCGGAAA AGCGCGGGGA AATTAAACCG GGTGATGTCT TAATCGAAGC CACCAGTGGT	MSTLEQTIGN TPLVKLQRMG PDNGSEVWLK LEGNNPAGSV KDRAALSMIV EAEKRGEIKP GDVLIEATSG NTGIALAMIA ALKGYRMKLL MPDNMSQERR AAMRAYGAEL ILVTKEQGME GARDLALEMA NRGEGKLLDQ FNNPDNPYAH YTTTGPEIWQ QTGGRITHFV SSMGTTGTIT GVSRFMREQS KPVTIVGLQP EEGSSIPGIR

		AACACCGGCA	RWPTEYLPGI
		TTGCGCTGGC	FNASLVDEVL
		AATGATTGCC	DIHQRDAENT
		GCGCTGAAAG	MRELAVREGI
		GCTATCGCAT	FCGVSSGGAV
		GAAATTGCTG	AGALRVAKAN
		ATGCCCGACA	PDAVVVAIIC
		ACATGAGCCA	DRGDRYLSTG
		GGAACGCCGT	VFGEEHFSQG
		GCGGCGATGC	AGI
		GTGCTTATGG	(SEQ ID
		TGCGGAACTG	NO:248)
		ATTCTTGTCA	
		CCAAAGAGCA	
		GGGCATGGAA	
		GGTGCGCGCG	
		ATCTGGCGCT	
		GGAGATGGCG	
		AATCGTGGCG	
		AAGGAAAGCT	
		GCTCGATCAG	
		TTCAATAATC	
		CCGATAACCC	
		TTATGCGCAT	
		TACACCACCA	
		CTGGGCCGGA	
		AATCTGGCAG	
		CAAACCGGCG	
		GGCGCATCAC	
		TCATTTTGTC	
		TCCAGCATGG	
		GGACGACCGG	
		CACTATCACC	
		GGCGTCTCAC	
		GCTTTATGCG	
		CGAACAATCC	
		AAACCGGTGA	
		CCATTGTCGG	
		CCTGCAACCG	
		GAAGAGGGCA	
		GCAGCATTCC	
		CGGCATTCGC	

		CGCTGGCCTA CGGAATATCT GCCGGGGATT TTCAACGCTT CTCTGGTGGA TGAGGTGCTG GATATTCATC AGCGCGATGC GGAAAACACC ATGCGCGAAC TGGCGGTGCG GGAAGGAATA TTCTGTGGCG TCAGCTCCGG CGGCGCGGTT GCCGGAGCAC TGCGGGTGGC AAAAGCTAAC CCTGACGCGG TGGTGGTGGC GATCATCTGC GATCGTGGCG ATCGCTACCT TTCTACCGGG GTGTTTGGGG AAGAGCATTT TAGCCAGGGG GCGGGGATTT AA (SEQ ID NO:247)	
maoC	fused aldehyde dehydrogenase/enoyl-CoA hydratase	atgCAGCAGT TAGCCAGTTT CTTATCCGGT ACCTGGCAGT CTGGCCGGGG CCGTAGCCGT TTGATTCACC ACGCTATTAG CGGCGAGGCG TTATGGGAAG TGACCAGTGA	MQQLASFLSG TWQSGRGRSR LIHHAISGEA LWEVTSEGLD MAAARQFAIE KGAPALRAMT FIERAAMLKA VAKHLLSEKE RFYALSAQTG ATRADSWVDI EGGIGTLFTY

		AGGTCTTGAT	ASLGSRELPD
		ATGGCGGCTG	DTLWPEDELI
		CCCGCCAGTT	PLSKEGGFAA
		TGCCATTGAA	RHLLTSKSGV
		AAAGGTGCCC	AVHINAFNFP
		CCGCCCTTCG	CWGMLEKLAP
		CGCTATGACC	TWLGGMPAII
		TTTATCGAAC	KPATATAQLT
		GTGCGGCGAT	QAMVKSIVDS
		GCTTAAAGCG	GLVPEGAISL
		GTCGCTAAAC	ICGSAGDLLD
		ATCTGCTGAG	HLDSQDVVTF
		TGAAAAAGAG	TGSAATGQML
		CGTTTCTATG	RVQPNIVAKS
		CTCTTTCTGC	IPFTMEADSL
		GCAAACAGGC	NCCVLGEDVT
		GCAACGCGGG	PDQPEFALFI
		CAGACAGTTG	REVVREMTTK
		GGTTGATATT	AGQKCTAIRR
		GAAGGTGGCA	IIVPQALVNA
		TTGGGACGTT	VSDALVARLQ
		ATTTACTTAC	KVVVGDPAQE
		GCCAGCCTCG	GVKMGALVNA
		GTAGCCGGGA	EQRADVQEKV
		GCTGCCTGAC	NILLAAGCEI
		GATACGCTGT	RLGGQADLSA
		GGCCGGAAGA	AGAFFPPTLL
		TGAATTGATC	YCPQPDETPA
		CCCTTATCGA	VHATEAFGPV
		AAGAAGGTGG	ATLMPAQNQR
		ATTTGCCGCG	HALQLACAGG
		CGCCATTTAC	GSLAGTLVTA
		TGACCTCAAA	DPQIARQFIA
		GTCAGGCGTG	DAARTHGRIQ
		GCAGTGCATA	ILNEESAKES
		TTAACGCCTT	TGHGSPLPQL
		TAACTTCCCC	VHGGPGRAGG
		TGCTGGGGAA	GEELGGLRAV
		TGCTGGAAAA	KHYMQRTAVQ
		GCTGGCACCA	GSPTMLAAIS
		ACGTGGCTGG	KQWVRGAKVE
		GCGGAATGCC	EDRIHPFRKY

		AGCCATCATC	FEELQPGDSL
		AAACCAGCTA	LTPRRTMTEA
		CCGCGACGGC	DIVNFACLSG
		CCAACTGACT	DHFYAHMDKI
		CAGGCGATGG	AAAESIFGER
		TGAAATCAAT	VVHGYFVLSA
		TGTCGATAGT	AAGLFVDAGV
		GGTCTTGTTC	GPVIANYGLE
		CCGAAGGCGC	SLRFIEPVKP
		AATTAGTCTG	GDTIQVRLTC
		ATCTGCGGTA	KRKTLKKQRS
		GTGCTGGCGA	AEEKPTGVVE
		CTTGTTGGAT	WAVEVFNQHQ
		CATCTGGACA	TPVALYSILT
		GCCAGGATGT	LVARQHGDFV
		GGTGACTTTC	D
		ACGGGGTCAG	(SEQ ID
		CGGCGACCGG	NO:250)
		ACAGATGCTG	
		CGAGTTCAGC	
		CAAATATCGT	
		CGCCAAATCT	
		ATCCCCTTCA	
		CTATGGAAGC	
		TGATTCCCTG	
		AACTGCTGCG	
		TACTGGGCGA	
		AGATGTCACC	
		CCGGATCAAC	
		CGGAGTTTGC	
		GCTGTTTATT	
		CGTGAAGTTG	
		TGCGTGAGAT	
		GACCACAAAA	
		GCCGGGCAAA	
		AATGTACGGC	
		AATCCGGCGG	
		ATTATTGTGC	
		CGCAGGCATT	
		GGTTAATGCT	
		GTCAGTGATG	
		CTCTGGTTGC	

		GCGATTACAG	
		AAAGTCGTGG	
		TCGGTGATCC	
		TGCTCAGGAA	
		GGCGTGAAAA	
		TGGGCGCACT	
		GGTAAATGCT	
		GAGCAGCGTG	
		CCGATGTGCA	
		GGAAAAAGTG	
		AACATATTGC	
		TGGCTGCAGG	
		ATGCGAGATT	
		CGCCTCGGTG	
		GTCAGGCGGA	
		TTTATCTGCT	
		GCGGGTGCCT	
		TCTTCCCGCC	
		AACCTTATTG	
		TACTGTCCGC	
		AGCCGGATGA	
		AACACCGGCG	
		GTACATGCAA	
		CAGAAGCCTT	
		TGGCCCTGTC	
		GCAACGCTGA	
		TGCCAGCACA	
		AAACCAGCGA	
		CATGCTCTGC	
		AACTGGCTTG	
		TGCAGGCGGC	
		GGTAGCCTTG	
		CGGGAACGCT	
		GGTGACGGCT	
		GATCCGCAAA	
		TTGCGCGTCA	
		GTTTATTGCC	
		GACGCGGCAC	
		GTACGCATGG	
		GCGAATTCAG	
		ATCCTCAATG	
		AAGAGTCGGC	

		AAAAGAATCC	
		ACCGGGCATG	
		GCTCCCCACT	
		GCCACAACTG	
		GTACATGGTG	
		GGCCTGGTCG	
		CGCAGGAGGC	
		GGTGAAGAAT	
		TAGGCGGTTT	
		ACGAGCGGTG	
		AAACATTACA	
		TGCAGCGAAC	
		CGCTGTTCAG	
		GGTAGTCCGA	
		CGATGCTTGC	
		CGCTATCAGT	
		AAACAGTGGG	
		TGCGCGGTGC	
		GAAAGTCGAA	
		GAAGATCGTA	
		TTCATCCGTT	
		CCGCAAATAT	
		TTTGAGGAGC	
		TACAACCAGG	
		CGACAGCCTG	
		TTGACTCCCC	
		GCCGCACAAT	
		GACAGAGGCC	
		GATATTGTTA	
		ACTTTGCTTG	
		CCTCAGCGGC	
		GATCATTTCT	
		ATGCACATAT	
		GGATAAGATT	
		GCTGCTGCCG	
		AATCTATTTT	
		CGGTGAGCGG	
		GTGGTGCATG	
		GGTATTTTGT	
		GCTTTCTGCG	
		GCTGCGGGTC	
		TGTTTGTCGA	

		TGCCGGTGTC GGTCCGGTCA TTGCTAACTA CGGGCTGGAA AGCTTGCGTT TTATCGAACC CGTAAAGCCA GGCGATACCA TCCAGGTGCG TCTCACCTGT AAGCGCAAGA CGCTGAAAAA ACAGCGTAGC GCAGAAGAAA AACCAACAGG TGTGGTGGAA TGGGCTGTAG AGGTATTCAA TCAGCATCAA ACCCCGGTGG CGCTGTATTC AATTCTGACG CTGGTGGCCA GGCAGCACGG TGATTTTGTC GATTAA (SEQ ID NO:249)	

Source	Genbank Accession Number
Shigella sp. D9	ZP_05432652
Citrobacter youngae ATCC 29220	ZP_04561391.1
Salmonella enterica	YP_001570967.1
Escherichia fergusonii ATCC 35469	YP_002382254.1
Klebsiella pneumoniae NTUH-K2044	YP_002918743.1
Enterobacter cancerogenus ATCC 35316	ZP_03281954.1
Cronobacter turicensis	CBA29728.1
Erwinia pyrifoliae Ep1/96	YP_002649242.1
Pectobacterium carotovorum subsp. carotovorum PC1	YP_003018119.1
Dickeya dadantii Ech703	YP_002987184.1

Edwardsiella ictaluri 93-146	YP_002932813.1
Providencia alcalifaciens DSM 30120	ZP_03317956.1
Yersinia kristensenii ATCC 33638	ZP_04624337.1
Photorhabdus asymbiotica	YP_003041580.1
Pantoea sp. At-9b	ZP_05728924.1
Actinobacillus succinogenes 130Z	YP_001344737.1
Mannheimia succiniciproducens MBEL55E	YP_088386.1
Pasteurella multocida subsp. multocida str. Pm70	NP_245421.1
Haemophilus somnus 129PT	YP_719117.1
Proteus mirabilis HI4320	YP_002150544.1
Sodalis glossinidius str. 'morsitans'	YP_454706.1
Candidatus Blochmannia pennsylvanicus str. BPEN	YP_277927.1
Aggregatibacter aphrophilus NJ8700	YP_003007342.1
Vibrio cholerae MZO-3	ZP_01958381.1
Baumannia cicadellinicola str. Hc (Homalodisca coagulata)	YP_588853.1
Vibrionales bacterium SWAT-3	ZP_01815187.1
Aliivibrio salmonicida LFI1238	YP_002262988.1
Aeromonas salmonicida subsp. salmonicida A449	YP_001141819.1
Wigglesworthia glossinidia endosymbiont of Glossina brevipalpis	NP_871303.1
Glaciecola sp. HTCC2999	ZP_03560821.1
Alteromonas macleodii ATCC 27126	ZP_04714556.1

FIG. 16

Gene	Name	Nucleotide Sequence	Protein Sequence
fabB	B-ketoacyl synthase/ 3-oxoacyl-[acyl-carrier-protein] synthase I	atgAAACGTG CAGTGATTAC TGGCCTGGGC ATTGTTTCCA GCATCGGTAA TAACCAGCAG GAAGTCCTGG CATCTCTGCG TGAAGGACGT TCAGGGATCA CTTTCTCTCA GGAGCTGAAG GATTCCGGCA TGCGTAGCCA CGTCTGGGGC AACGTAAAAC TGGATACCAC TGGCCTCATT GACCGCAAAG TTGTGCGCTT TATGAGCGAC GCATCCATTT ATGCATTCCT TTCTATGGAG CAGGCAATCG CTGATGCGGG CCTCTCTCCG GAAGCTTACC AGAATAACCC GCGCGTTGGC CTGATTGCAG GTTCCGGCGG CGGCTCCCCG CGTTTCCAGG TGTTCGGCGC TGACGCAATG CGCGGCCCGC GCGGCCTGAA AGCGGTTGGC	MKRAVITGLG IVSSIGNNQQ EVLASLREGR SGITFSQELK DSGMRSHVWG NVKLDTTGLI DRKVVRFMSD ASIYAFLSME QAIADAGLSP EAYQNNPRVG LIAGSGGGSP RFQVFGADAM RGPRGLKAVG PYVVTKAMAS GVSACLATPF KIHGVNYSIS SACATSAHCI GNAVEQIQLG KQDIVFAGGG EELCWEMACE FDAMGALSTK YNDTPEKASR TYDAHRDGFV IAGGGGMVVV EELEHALARG AHIYAEIVGY GATSDGADMV APSGEGAVRC MKMAMHGVDT PIDYLNSHGT STPVGDVKEL AAIREVFGDK SPAISATKAM TGHSLGAAGV QEAIYSLLML EHGFIAPSIN IEELDEQAAG LNIVTETTDR ELTTVMSNSF

		CCGTATGTGG TCACCAAAGC GATGGCATCC GGCGTTTCTG CCTGCCTCGC CACCCCGTTT AAAATTCATG GCGTTAACTA CTCCATCAGC TCCGCGTGTG CGACTTCCGC ACACTGTATC GGTAACGCAG TAGAGCAGAT CCAACTGGGC AAACAGGACA TCGTGTTTGC TGGCGGCGGC GAAGAGCTGT GCTGGGAAAT GGCTTGCGAA TTCGACGCAA TGGGTGCGCT GTCTACTAAA TACAACGACA CCCCGGAAAA AGCCTCCCGT ACTTACGACG CTCACCGTGA CGGTTTCGTT ATCGCTGGCG GCGGCGGTAT GGTAGTGGTT GAAGAGCTGG AACACGCGCT GGCGCGTGGT GCTCACATCT ATGCTGAAAT CGTTGGCTAC GGCGCAACCT CTGATGGTGC AGACATGGTT	GFGGTNATLV MRKLKD (SEQ ID NO:252)

		GCTCCGTCTG GCGAAGGCGC AGTACGCTGC ATGAAGATGG CGATGCATGG CGTTGATACC CCAATCGATT ACCTGAACTC CCACGGTACT TCGACTCCGG TTGGCGACGT GAAAGAGCTG GCAGCTATCC GTGAAGTGTT CGGCGATAAG AGCCCGGCGA TTTCTGCAAC CAAAGCCATG ACCGGTCACT CTCTGGGCGC TGCTGGCGTA CAGGAAGCTA TCTACTCTCT GCTGATGCTG GAACACGGCT TTATCGCCCC GAGCATCAAC ATTGAAGAGC TGGACGAGCA GGCTGCGGGT CTGAACATCG TGACCGAAAC GACCGATCGC GAACTGACCA CCGTTATGTC TAACAGCTTC GGCTTCGGCG GCACCAACGC CACGCTGGTA ATGCGCAAGC TGAAAGATTA A	

		(SEQ ID NO:251)	
fabF	3-oxoacyl-[acyl-carrier-protein] synthase II	gtgTCTAAGC GTCGTGTAGT TGTGACCGGA CTGGGCATGT TGTCTCCTGT CGGCAATACC GTAGAGTCTA CCTGGAAAGC TCTGCTTGCC GGTCAGAGTG GCATCAGCCT AATCGACCAT TTCGATACTA GCGCCTATGC AACGAAATTT GCTGGCTTAG TAAAGGATTT TAACTGTGAG GACATTATCT CGCGCAAAGA ACAGCGCAAG ATGGATGCCT TCATTCAATA TGGAATTGTC GCTGGCGTTC AGGCCATGCA GGATTCTGGC CTTGAAATAA CGGAAGAGAA CGCAACCCGC ATTGGTGCCG CAATTGGCTC CGGGATTGGC GGCCTCGGAC TGATCGAAGA AAACCACACA TCTCTGATGA ACGGTGGTCC ACGTAAGATC AGCCCATTCT TCGTTCCGTC	MSKRRVVVTG LGMLSPVGNT VESTWKALLA GQSGISLIDH FDTSAYATKF AGLVKDFNCE DIISRKEQRK MDAFIQYGIV AGVQAMQDSG LEITEENATR IGAAIGSGIG GLGLIEENHT SLMNGGPRKI SPFFVPSTIV NMVAGHLTIM YGLRGPSISI ATACTSGVHN IGHAARIIAY GDADVMVAGG AEKASTPLGV GGFGAARALS TRNDNPQAAS RPWDKERDGF VLGDGAGMLV LEEYEHAKKR GAKIYAELVG FGMSSDAYHM TSPPENGAGA ALAMANALRD AGIEASQIGY VNAHGTSTPA GDKAEAQAVK TIFGEAASRV LVSSTKSMTG HLLGAAGAVE SIYSILALRD QAVPPTINLD NPDEGCDLDF VPHEARQVSG MEYTLCNSFG FGGTNGSLIF

		AACGATTGTG AACATGGTGG CAGGTCATCT GACTATCATG TATGGCCTGC GTGGCCCGAG CATCTCTATC GCGACTGCCT GTACTTCCGG CGTGCACAAC ATTGGCCATG CTGCGCGTAT TATCGCGTAT GGCGATGCTG ACGTGATGGT TGCAGGTGGC GCAGAGAAAG CCAGTACGCC GCTGGGCGTT GGTGGTTTTG GCGCGGCACG TGCATTATCT ACCCGCAATG ATAACCCGCA AGCGGCGAGC CGCCCGTGGG ATAAAGAGCG TGATGGTTTC GTACTGGGCG ATGGTGCCGG TATGCTGGTA CTTGAAGAGT ACGAACACGC GAAAAAACGC GGTGCGAAAA TTTACGCTGA ACTCGTCGGC TTTGGTATGA GCAGCGATGC TTATCATATG ACGTCACCGC CAGAAAATGG	KKI (SEQ ID NO:254)

		CGCAGGCGCA GCTCTGGCGA TGGCAAATGC TCTGCGTGAT GCAGGCATTG AAGCGAGTCA GATTGGCTAC GTTAACGCGC ACGGTACTTC TACGCCGGCT GGCGATAAAG CTGAAGCGCA GGCGGTGAAA ACCATCTTCG GTGAAGCTGC AAGCCGTGTG TTGGTAAGCT CCACGAAATC TATGACCGGT CACCTGTTAG GTGCGGCGGG TGCAGTAGAA TCTATCTACT CCATCCTGGC GCTGCGCGAT CAGGCTGTTC CGCCAACCAT CAACCTGGAT AACCCGGATG AAGGTTGCGA TCTGGATTTC GTACCGCACG AAGCGCGTCA GGTTAGCGGA ATGGAATACA CTCTGTGTAA CTCCTTCGGC TTCGGTGGCA CTAATGGTTC TTTGATCTTT AAAAAGATCT AA	

		(SEQ ID NO:253)	
fadJ	fused enoyl-CoA hydratase and epimerase and isomerase/3-hydroxyacyl-CoA dehydrogenase	atgGAAATGA CATCAGCGTT TACCCTTAAT GTTCGTCTGG ACAACATTGC CGTTATCACC ATCGACGTAC CGGGTGAGAA AATGAATACC CTGAAGGCGG AGTTTGCCTC GCAGGTGCGC GCCATTATTA AGCAACTCCG TGAAAACAAA GAGTTGCGAG GCGTGGTGTT TGTCTCCGCT AAACCGGACA ACTTCATTGC TGGCGCAGAC ATCAACATGA TCGGCAACTG CAAAACGGCG CAAGAAGCGG AAGCTCTGGC GCGGCAGGGC CAACAGTTGA TGGCGGAGAT TCATGCTTTG CCCATTCAGG TTATCGCGGC TATTCATGGC GCTTGCCTGG GTGGTGGGCT GGAGTTGGCG CTGGCGTGCC ACGGTCGCGT TTGTACTGAC GATCCTAAAA CGGTGCTCGG	MEMTSAFTLN VRLDNIAVIT IDVPGEKMNT LKAEFASQVR AIIKQLRENK ELRGVVFVSA KPDNFIAGAD INMIGNCKTA QEAEALARQG QQLMAEIHAL PIQVIAAIHG ACLGGGLELA LACHGRVCTD DPKTVLGLPE VQLGLLPGSG GTQRLPRLIG VSTALEMILT GKQLRAKQAL KLGLVDDVVP HSILLEAAVE LAKKERPSSR PLPVRERILA GPLGRALLFK MVGKKTEHKT QGNYPATERI LEVVETGLAQ GTSSGYDAEA RAFGELAMTP QSQALRSIFF ASTDVKKDPG SDAPPAPLNS VGILGGGLMG GGIAYVTACK AGIPVRIKDI NPQGINHALK YSWDQLEGKV RRRHLKASER DKQLALISGT TDYRGFAHRD LIIEAVFENL ELKQQMVAEV

		TTTGCCTGAA	EQNCAAHTIF
		GTACAACTTG	ASNTSSLPIG
		GATTGTTACC	DIAAHATRPE
		CGGTTCAGGC	QVIGLHFFSP
		GGCACCCAGC	VEKMPLVEII
		GTTTACCGCG	PHAGTSAQTI
		TCTGATAGGC	ATTVKLAKKQ
		GTCAGCACAG	GKTPIVVRDK
		CATTAGAGAT	AGFYVNRILA
		GATCCTCACC	PYINEAIRML
		GGAAAACAAC	TQGERVEHID
		TTCGGGCGAA	AALVKFGFPV
		ACAGGCATTA	GPIQLLDEVG
		AAGCTGGGGC	IDTGTKIIPV
		TGGTGGATGA	LEAAYGERFS
		CGTTGTTCCG	APANVVSSIL
		CACTCCATTC	NDDRKGRKNG
		TGCTGGAAGC	RGFYLYGQKG
		CGCTGTTGAG	RKSKKQVDPA
		CTGGCAAAGA	IYPLIGTQGQ
		AGGAGCGCCC	GRISAPQVAE
		ATCTTCCCGC	RCVMLMLNEA
		CCTCTACCTG	VRCVDEQVIR
		TACGCGAGCG	SVRDGDIGAV
		TATTCTGGCG	FGIGFPPFLG
		GGGCCGTTAG	GPFRYIDSLG
		GTCGTGCGCT	AGEVVAIMQR
		GCTGTTCAAA	LATQYGSRFT
		ATGGTCGGCA	PCERLVEMGA
		AGAAAACAGA	RGESFWKTTA
		ACACAAAACT	TDLQ
		CAAGGCAATT	(SEQ ID
		ATCCGGCGAC	NO:256)
		AGAACGCATC	
		CTGGAGGTTG	
		TTGAAACGGG	
		ATTAGCGCAG	
		GGCACCAGCA	
		GCGGTTATGA	
		CGCCGAAGCT	
		CGGGCGTTTG	
		GCGAACTGGC	

		GATGACGCCA	
		CAATCGCAGG	
		CGCTGCGTAG	
		TATCTTTTTT	
		GCCAGTACGG	
		ACGTGAAGAA	
		AGATCCCGGC	
		AGTGATGCGC	
		CGCCTGCGCC	
		ATTAAACAGC	
		GTGGGGATTT	
		TAGGTGGTGG	
		CTTGATGGGC	
		GGCGGTATTG	
		CTTATGTCAC	
		TGCTTGTAAA	
		GCGGGGATTC	
		CGGTCAGAAT	
		TAAAGATATC	
		AACCCGCAGG	
		GCATAAATCA	
		TGCGCTGAAG	
		TACAGTTGGG	
		ATCAGCTGGA	
		GGGCAAAGTT	
		CGCCGTCGTC	
		ATCTCAAAGC	
		CAGCGAACGT	
		GACAAACAGC	
		TGGCATTAAT	
		CTCCGGAACG	
		ACGGACTATC	
		GCGGCTTTGC	
		CCATCGCGAT	
		CTGATTATTG	
		AAGCGGTGTT	
		TGAAAATCTC	
		GAATTGAAAC	
		AACAGATGGT	
		GGCGGAAGTT	
		GAGCAAAATT	
		GCGCCGCTCA	

		TACCATCTTT	
		GCTTCGAATA	
		CGTCATCTTT	
		ACCGATTGGT	
		GATATCGCCG	
		CTCACGCCAC	
		GCGACCTGAG	
		CAAGTTATCG	
		GCCTGCATTT	
		CTTCAGTCCG	
		GTGGAAAAAA	
		TGCCGCTGGT	
		GGAGATTATT	
		CCTCATGCGG	
		GGACATCGGC	
		GCAAACCATC	
		GCTACCACAG	
		TAAAACTGGC	
		GAAAAAACAG	
		GGTAAAACGC	
		CAATTGTCGT	
		GCGTGACAAA	
		GCCGGTTTTT	
		ACGTCAATCG	
		CATCTTAGCG	
		CCTTACATTA	
		ATGAAGCTAT	
		CCGCATGTTG	
		ACCCAAGGTG	
		AACGGGTAGA	
		GCACATTGAT	
		GCCGCGCTAG	
		TGAAATTTGG	
		TTTTCCGGTA	
		GGCCCAATCC	
		AACTTTTGGA	
		TGAGGTAGGA	
		ATCGACACCG	
		GGACTAAAAT	
		TATTCCTGTA	
		CTGGAAGCCG	
		CTTATGGAGA	

		ACGTTTTAGC GCGCCTGCAA ATGTTGTTTC TTCAATTTTG AACGACGATC GCAAAGGCAG AAAAAATGGC CGGGGTTTCT ATCTTTATGG TCAGAAAGGG CGTAAAAGCA AAAAACAGGT CGATCCCGCC ATTTACCCGC TGATTGGCAC ACAAGGGCAG GGGCGAATCT CCGCACCGCA GGTTGCTGAA CGGTGTGTGA TGTTGATGCT GAATGAAGCA GTACGTTGTG TTGATGAGCA GGTTATCCGT AGCGTGCGTG ACGGGGATAT TGGCGCGGTA TTTGGCATTG GTTTTCCGCC ATTTCTCGGT GGACCGTTCC GCTATATCGA TTCTCTCGGC GCGGGCGAAG TGGTTGCAAT AATGCAACGA CTTGCCACGC AGTATGGTTC CCGTTTTACC CCTTGCGAGC GTTTGGTCGA	

		GATGGGCGCG CGTGGGGAAA GTTTTTGGAA AACAACTGCA ACTGACCTGC AATAA (SEQ ID NO:255)	
xerC	site-specific tyrosine recombinase	atgACCGATT TACACACCGA TGTAGAACGC TACCTACGTT ATCTGAGCGT GGAGCGCCAG CTTAGCCCGA TAACCCTGCT TAACTACCAG CGTCAGCTTG AGGCGATCAT CAATTTTGCC AGCGAAAACG GCCTGCAAAG CTGGCAGCAA TGTGATGTGA CGATGGTGCG CAATTTTGCT GTACGCAGTC GCCGTAAAGG GCTGGGAGCA GCAAGTCTGG CGTTACGGCT TTCTGCGCTA CGTAGCTTTT TTGACTGGCT GGTCAGCCAG AACGAACTCA AAGCTAACCC GGCGAAAGGT GTTTCGGCAC CGAAAGCGCC GCGTCATCTG CCGAAAAACA TCGACGTCGA	MTDLHTDVER YLRYLSVERQ LSPITLLNYQ RQLEAIINFA SENGLQSWQQ CDVTMVRNFA VRSRRKGLGA ASLALRLSAL RSFFDWLVSQ NELKANPAKG VSAPKAPRHL PKNIDVDDMN RLLDIDINDP LAVRDRAMLE VMYGAGLRLS ELVGLDIKHL DLESGEVWVM GKGSKERRLP IGRNAVAWIE HWLDLRDLFG SEDDALFLSK LGKRISARNV QKRFAEWGIK QGLNNHVHPH KLRHSFATHM LESSGDLRGV QELLGHANLS TTQIYTHLDF QHLASVYDAA HPRAKRGK (SEQ ID NO:258)

		CGATATGAAT	
		CGGCTGCTGG	
		ATATTGATAT	
		CAATGATCCC	
		CTCGCTGTAC	
		GCGACCGTGC	
		AATGCTGGAA	
		GTGATGTACG	
		GCGCGGGTCT	
		GCGTCTTTCT	
		GAGCTGGTGG	
		GGCTGGATAT	
		TAAACACCTC	
		GACCTGGAGT	
		CTGGTGAAGT	
		GTGGGTTATG	
		GGGAAAGGCA	
		GCAAAGAGCG	
		CCGCCTGCCG	
		ATTGGTCGCA	
		ACGCTGTGGC	
		GTGGATTGAG	
		CACTGGCTTG	
		ATTTGCGCGA	
		CCTGTTTGGT	
		AGCGAAGACG	
		ACGCGCTTTT	
		TCTGTCGAAA	
		CTGGGCAAGC	
		GTATCTCCGC	
		GCGTAATGTG	
		CAGAAACGCT	
		TTGCCGAATG	
		GGGCATAAAA	
		CAAGGGCTGA	
		ATAATCACGT	
		TCATCCGCAT	
		AAATTACGTC	
		ACTCGTTCGC	
		CACGCATATG	
		CTGGAGTCGA	
		GCGGCGATCT	

		TCGTGGTGTG CAGGAGCTGC TGGGTCATGC CAACCTCTCC ACCACGCAAA TCTATACTCA TCTTGATTTT CAACACCTTG CCTCGGTGTA CGATGCGGCG CATCCACGCG CCAAACGGGG GAAATAA (SEQ ID NO:257)	
yqeF	predicted acyltransferase	atgAAAGACG TTGTGATTGT CGGGGCGTTA CGGACACCTA TCGGCTGCTT TCGTGGTGCG TTAGCGGGTC ATTCCGCCGT GGAACTTGGT AGTCTGGTCG TGAAAGCGTT AATAGAACGT ACCGGCGTTC CTGCATATGC GGTGGATGAA GTAATTCTTG GTCAGGTGTT GACTGCAGGG GCAGGGCAGA ATCCGGCAAG GCAATCGGCT ATTAAAGGTG GTCTGCCTAA TAGCGTTTCT GCAATCACTA TTAATGACGT TTGCGGTTCC GGGCTTAAAG	MKDVVIVGAL RTPIGCFRGA LAGHSAVELG SLVVKALIER TGVPAYAVDE VILGQVLTAG AGQNPARQSA IKGGLPNSVS AITINDVCGS GLKALHLATQ AIQCGEADIV IAGGQENMSR APHVLTDSRT GAQLGNSQLV DSLVHDGLWD AFNDYHIGVT AENLAREYGI SRQLQDAYAL SSQQKARAAI DAGRFKDEIV PVMTQSNGQT LVVDTDEQPR TDASAEGLAR LNPSFDSLGS VTAGNASSIN DGAAAVMMMS EAKARALNLP VLARIRAFAS

		CACTGCATCT	VGVDPALMGI
		GGCTACTCAG	APVYATRRCL
		GCGATACAGT	ERVGWQLAEV
		GTGGCGAGGC	DLIEANEAFA
		TGATATTGTC	AQALSVGKML
		ATCGCCGGTG	EWDERRVNVN
		GCCAGGAAAA	GGAIALGHPI
		CATGAGCCGC	GASGCRILVS
		GCACCACATG	LVHEMVKRNA
		TTCTGACTGA	RKGLATLCIG
		TAGCCGCACC	GGQGVALTIE
		GGTGCACAGC	RDE
		TTGGCAATAG	(SEQ ID
		CCAGTTGGTT	NO:260)
		GACAGTCTTG	
		TGCATGATGG	
		GTTGTGGGAT	
		GCCTTCAATG	
		ATTATCATAT	
		TGGTGTCACC	
		GCCGAAAATC	
		TGGCTCGCGA	
		ATATGGCATC	
		AGCCGTCAGT	
		TGCAGGATGC	
		TTACGCACTT	
		AGCTCGCAAC	
		AAAAAGCGCG	
		AGCGGCGATT	
		GACGCCGGAC	
		GATTTAAAGA	
		TGAGATCGTC	
		CCGGTAATGA	
		CCCAAAGTAA	
		CGGGCAGACG	
		TTGGTTGTTG	
		ATACCGATGA	
		ACAGCCACGC	
		ACTGACGCCA	
		GCGCAGAAGG	
		CTTAGCCCGT	
		TTAAATCCTT	

		CATTTGATAG	
		TCTCGGTTCT	
		GTGACAGCGG	
		GTAATGCATC	
		ATCCATAAAC	
		GATGGCGCAG	
		CTGCGGTAAT	
		GATGATGAGC	
		GAAGCCAAAG	
		CACGAGCGTT	
		GAATTTACCC	
		GTGCTGGCCC	
		GCATTCGCGC	
		ATTTGCCAGC	
		GTTGGTGTAG	
		ATCCGGCATT	
		GATGGGAATT	
		GCGCCGGTGT	
		ATGCGACCCG	
		CCGTTGCCTG	
		GAGCGTGTAG	
		GCTGGCAGTT	
		GGCTGAAGTC	
		GATCTTATCG	
		AGGCTAATGA	
		AGCGTTTGCT	
		GCACAGGCGC	
		TTTCGGTTGG	
		CAAGATGCTT	
		GAGTGGGATG	
		AGCGTCGGGT	
		CAATGTCAAT	
		GGTGGCGCGA	
		TCGCACTCGG	
		TCACCCGATA	
		GGCGCTTCCG	
		GTTGCCGAAT	
		CCTGGTTTCT	
		CTGGTTCATG	
		AAATGGTGAA	
		ACGTAATGCC	
		CGCAAAGGAC	

		TGGCAACGCT TTGTATCGGC GGGGGCCAGG GTGTGGCATT GACCATTGAA CGTGACGAAT AG (SEQ ID NO:259)	
murQ	predicted PTS component	atgCAATTTG AAAAGATGAT TACTGAAGGC TCGAACACCG CCTCGGCTGA AATTGACCGC GTATCGACGC TGGAAATGTG CCGGATTATC AACGATGAAG ATAAAACCGT ACCGCTTGCC GTTGAGCGCG TACTGCCGGA TATCGCCGCG GCGATCGATG TTATCCACGC CCAGGTAAGC GGCGGCGGGC GTCTGATTTA CCTCGGTGCG GGAACATCCG GTCGTCTGGG GATTCTGGAT GCCAGCGAAT GTCCGCCCAC CTACGGCGTG AAACCGGGTC TGGTGGTTGG TTTGATTGCT GGCGGCGAAT ATGCCATTCA GCACGCGGTG GAAGGCGCGG	MQFEKMITEG SNTASAEIDR VSTLEMCRII NDEDKTVPLA VERVLPDIAA AIDVIHAQVS GGGRLIYLGA GTSGRLGILD ASECPPTYGV KPGLVVGLIA GGEYAIQHAV EGAEDSREGG VNDLKNINLT AQDVVVGIAA SGRTPYVIAG LEYARQLGCR TVGISCNPGS AVSTTAEFAI TPIVGAEVVT GSSRMKAGTA QKLVLNMLST GLMIKSGKVF GNLMVDVVAT NEKLHVRQVN IVKNATGCSA EQAEAALIAC ERNCKTAIVM VLKNLDAAEA KKRLDQHGGF IRQVLDKE (SEQ ID NO:262)

		AAGATAGCCG	
		GGAAGGCGGT	
		GTTAATGATC	
		TGAAAAATAT	
		TAATTTAACG	
		GCACAGGATG	
		TGGTGGTTGG	
		CATTGCTGCC	
		AGCGGTCGCA	
		CGCCGTATGT	
		GATTGCCGGA	
		CTGGAATACG	
		CACGCCAGCT	
		CGGCTGCCGC	
		ACAGTGGGAA	
		TTTCCTGTAA	
		TCCGGGGAGC	
		GCCGTTTCAA	
		CCACCGCTGA	
		GTTTGCCATT	
		ACACCGATTG	
		TAGGTGCCGA	
		AGTTGTTACC	
		GGTTCTTCGC	
		GGATGAAAGC	
		AGGTACAGCG	
		CAGAAACTGG	
		TGCTCAATAT	
		GCTTTCCACC	
		GGGCTGATGA	
		TTAAATCCGG	
		CAAAGTGTTC	
		GGCAACCTGA	
		TGGTCGATGT	
		GGTCGCCACC	
		AACGAAAAAC	
		TGCATGTGCG	
		ACAGGTCAAT	
		ATTGTTAAAA	
		ACGCCACCGG	
		ATGTAGCGCA	
		GAGCAAGCGG	

		AAGCGGCGTT AATTGCTTGC GAGCGCAACT GTAAAACGGC CATTGTGATG GTGCTGAAAA ATCTCGATGC CGCAGAAGCT AAAAAACGCC TGGATCAACA CGGCGGCTTT ATTCGTCAGG TTTTAGACAA GGAATAA (SEQ ID NO:261)	

Source	Genbank Accession Number
Shigella boydii CDC 3083-94	YP_001881145.1
Escherichia fergusonii ATCC 35469	YP_002382013.1
Salmonella enterica subsp. arizonae	YP_001569590.1
Citrobacter sp. 30_2	ZP_04562837.1
Klebsiella pneumoniae subsp. pneumoniae MGH 78578	YP_001336360.1
Pectobacterium carotovorum subsp. carotovorum WPP14	ZP_03831287.1
Enterobacter cancerogenus ATCC 35316	ZP_03283474.1
Pantoea sp. At-9b	ZP_05730617.1
Cronobacter turicensis	CBA32510.1
Dickeya dadantii Ech586	ZP_05723897.1
Erwinia tasmaniensis Et1/99	YP_001907100.1
Serratia proteamaculans 568	YP_001479594.1
Edwardsiella ictaluri 93-146	YP_002934130.1
Sodalis glossinidius str. 'morsitans'	YP_455303.1
Yersinia aldovae ATCC 35236	ZP_04620215.1
Providencia stuartii ATCC 25827	ZP_02961167.1
Photorhabdus asymbiotica	YP_003040275.1
Proteus mirabilis HI4320	YP_002151524.1
Candidatus Blochmannia pennsylvanicus str. BPEN	YP_278005.1
Glaciecola sp. HTCC2999	ZP_03561088.1

Vibrio cholerae V51	ZP_04919940.1
Wigglesworthia glossinidia endosymbiont of Glossina brevipalpis	NP_871411.1
Tolumonas auensis DSM 9187	YP_002892770.1
Actinobacillus pleuropneumoniae serovar 1 str. 4074	ZP_00134992.2
Aggregatibacter aphrophilus NJ8700	YP_003007711.1
Pseudoalteromonas tunicata D2	ZP_01135065.1
Vibrionales bacterium SWAT-3	ZP_01816638.1
Pasteurella multocida subsp. multocida str. Pm70	NP_245276.1
Mannheimia succiniciproducens MBEL55E	YP_088783.1
Haemophilus somnus 129PT	YP_718877.1
Shewanella loihica PV-4	YP_001094535.1
Aliivibrio salmonicida LFI1238	YP_002262558.1

FIG. 17

AAR_7942 sequence

SEQ ID NO:195 - *Synechococcus elongatus* PCC7942 Synpcc7942_1594 DNA

```
   1 atgttcggtc ttatcggtca tctcaccagt ttggagcagg cccgcgacgt ttctcgcagg
  61 atgggctacg acgaatacgc cgatcaagga ttggagtttt ggagtagcgc tcctcctcaa
 121 atcgttgatg aaatcacagt caccagtgcc acaggcaagg tgattcacgg tcgctacatc
 181 gaatcgtgtt tcttgccgga aatgctggcg gcgcgccgct tcaaaacagc cacgcgcaaa
 241 gttctcaatg ccatgtccca tgcccaaaaa cacggcatcg acatctcggc cttgggggcc
 301 tttacctcga ttattttcga gaatttcgat ttggccagtt tgcggcaagt gcgcgacact
 361 accttggagt ttgaacggtt caccaccggc aatactcaca cggcctacgt aatctgtaga
 421 caggtggaag ccgctgctaa aacgctgggc atcgacatta cccaagcgac agtagcggtt
 481 gtcggcgcga ctggcgatat cggtagcgct gtctgccgct ggctcgacct caaactgggt
 541 gtcggtgatt tgatcctgac ggcgcgcaat caggagcgtt tggataacct gcaggctgaa
 601 ctcggccggg gcaagattct gcccttggaa gccgctctgc cggaagctga ctttatcgtg
 661 tgggtcgcca gtatgcctca gggcgtagtg atcgacccag caaccctgaa gcaaccctgc
 721 gtcctaatcg acgggggcta ccccaaaaac ttgggcagca aagtccaagg tgagggcatc
 781 tatgtcctca atggcggggt agttgaacat tgcttcgaca tcgactggca gatcatgtcc
 841 gctgcagaga tggcgcggcc cgagcgccag atgtttgcct gctttgccga ggcgatgctc
 901 ttggaatttg aaggctggca tactaacttc tcctggggcc gcaaccaaat cacgatcgag
 961 aagatggaag cgatcggtga ggcatcggtg cgccacggct tccaaccctt ggcattggca
1021 atttga
```

SEQ ID NO:196 - *Synechococcus elongatus* PCC7942 Synpcc7942_1594 protein (YP_400611)

```
   1 MFGLIGHLTS LEQARDVSRR MGYDEYADQG LEFWSSAPPQ IVDEITVTSA TGKVIHGRYI
  61 ESCFLPEMLA ARRFKTATRK VLNAMSHAQK HGIDISALGG FTSIIFENFD LASLRQVRDT
 121 TLEFERFTTG NTHTAYVICR QVEAAAKTLG IDITQATVAV VGATGDIGSA VCRWLDLKLG
 181 VGDLILTARN QERLDNLQAE LGRGKILPLE AALPEADFIV WVASMPQGVV IDPATLKQPC
 241 VLIDGGYPKN LGSKVQGEGI YVLNGGVVEH CFDIDWQIMS AAEMARPERQ MFACFAEAML
 301 LEFEGWHTNF SWGRNQITIE KMEAIGEASV RHGFQPLALA I
```

SEQ ID NO:213 – Nucleotide sequence of plasmid pCL-Ptrc-carB_'tesA

```
Cactataccaattgagatgggctagtcaatgataattactagtccttttcctttgagttgtgggtatct
gtaaattctgctagacctttgctggaaaacttgtaaattctgctagaccctctgtaaattccgctagac
```

FIG. 17 (Cont)

ctttgtgtgtttttttgtttatattcaagtggttataatttatagaataaagaaagaataaaaaaaga
taaaaagaatagatcccagccctgtgtataactcactactttagtcagttccgcagtattacaaaagga
tgtcgcaaacgctgtttgctcctctacaaaacagaccttaaaaccctaaaggcgtcggcatccgcttac
agacaagctgtgaccgtctccgggagctgcatgtgtcagaggttttcaccgtcatcaccgaaacgcgcg
aggcagcagatcaattcgcgcgcgaaggcgaagcggcatgcatttacgttgacaccatcgaatggtgca
aaacctttcgcggtatggcatgatagcgcccggaagagagtcaattcagggtggtgaatgtgaaaccag
taacgttatacgatgtcgcagagtatgccggtgtctcttatcagaccgtttcccgcgtggtgaaccagg
ccagccacgtttctgcgaaaacgcgggaaaaagtggaagcggcgatggcggagctgaattacattccca
accgcgtggcacaacaactggcgggcaaacagtcgttgctgattggcgttgccacctccagtctggccc
tgcacgcgccgtcgcaaattgtcgcggcgattaaatctcgcgccgatcaactgggtgccagcgtggtgg
tgtcgatggtagaacgaagcggcgtcgaagcctgtaaagcggcggtgcacaatcttctcgcgcaacgcg
tcagtgggctgatcattaactatccgctggatgaccaggatgccattgctgtggaagctgcctgcacta
atgttccggcgttatttcttgatgtctctgaccagacacccatcaacagtattattttctcccatgaag
acggtacgcgactgggcgtggagcatctggtcgcattgggtcaccagcaaatcgcgctgttagcgggcc
cattaagttctgtctcggcgcgtctgcgtctggctggctggcataaatatctcactcgcaatcaaattc
agccgatagcggaacgggaaggcgactggagtgccatgtccggttttcaacaaaccatgcaaatgctga
atgagggcatcgttcccactgcgatgctggttgccaacgatcagatggcgctgggcgcaatgcgcgcca
ttaccgagtccgggctgcgcgttggtgcggatatctcggtagtgggatacgacgataccgaagacagct
catgttatatcccgccgttaaccaccatcaaacaggattttcgcctgctggggcaaaccagcgtggacc
gcttgctgcaactctctcagggccaggcggtgaagggcaatcagctgttgcccgtctcactggtgaaaa
gaaaaaccaccctggcgcccaatacgcaaaccgcctctccccgcgcgttggccgattcattaatgcagc
tggcacgacaggtttcccgactggaaagcgggcagtgagcgcaacgcaattaatgtaagttagcgcgaa
ttgatctggtttgacagcttatcatcgactgcacggtgcaccaatgcttctggcgtcaggcagccatcg
gaagctgtggtatggctgtgcaggtcgtaaatcactgcataattcgtgtcgctcaaggcgcactcccgt
tctggataatgtttttgcgccgacatcataacggttctggcaaatattctgaaatgagctgttgacaa
ttaatcatccggctcgtataatgtgtggaattgtgagcggataacaatttcacacaggaaacagcgccg
ctgagaaaaagcgaagcggcactgctctttaacaatttatcagacaatctgtgtgggcactcgaccgga
attatcgattaactttattattaaaaattaaagaggtatatattaatgtatcgattaaataaggaggaa
taaaccatgaccagcgatgttcacgacgccacagacggcgtcaccgaaaccgcactcgacgacgagcag
tcgacccgccgcatcgccgagctgtacgccaccgatcccgagttcgccgccgccgcaccgttgcccgcc
gtggtcgacgcggcgcacaaacccgggctgcggctggcagagatcctgcagaccctgttcaccggctac
ggtgaccgcccggcgctgggataccgcgcccgtgaactggccaccgacgagggcgggcgcaccgtgacg
cgtctgctgccgcggttcgacaccctcacctacgcccaggtgtggtcgcgcgtgcaagcggtcgccgcg
gccctgcgccacaacttcgcgcagccgatctaccccggcgacgccgtcgcgacgatcggtttcgcgagt
cccgattacctgacgctggatctcgtatgcgcctacctgggcctcgtgagtgttccgctgcagcacaac
gcaccggtcagccggctcgccccgatcctggccgaggtcgaaccgcggatcctcaccgtgagcgccgaa
tacctcgacctcgcagtcgaatccgtgcgggacgtcaactcggtgtcgcagctcgtggtgttcgaccat
caccccgaggtcgacgaccaccgcgacgcactggcccgcgcgcgtgaacaactcgccggcaagggcatc
gccgtcaccaccctggacgcgatcgccgacgagggcgccgggctgccggccgaaccgatctacaccgcc
gaccatgatcagcgcctcgcgatgatcctgtacacctcgggttccaccggcgcacccaagggtgcgatg
tacaccgaggcgatggtggcgcggctgtggaccatgtcgttcatcacgggtgaccccacgccggtcatc
aacgtcaacttcatgccgctcaaccacctgggcgggcgcatccccatttccaccgccgtgcagaacggt

FIG. 17 (Cont)

ggaaccagttacttcgtaccggaatccgacatgtccacgctgttcgaggatctcgcgctggtgcgcccg
accgaactcggcctggttccgcgcgtcgccgacatgctctaccagcaccacctcgccaccgtcgaccgc
ctggtcacgcagggcgccgacgaactgaccgccgagaagcaggccggtgccgaactgcgtgagcaggtg
ctcggcggacgcgtgatcaccggattcgtcagcaccgcaccgctggccgcggagatgagggcgttcctc
gacatcaccctgggcgcacacatcgtcgacggctacgggctcaccgagaccggcgccgtgacacgcgac
ggtgtgatcgtgcggccaccggtgatcgactacaagctgatcgacgttcccgaactcggctacttcagc
accgacaagccctacccgcgtggcgaactgctggtcaggtcgcaaacgctgactcccgggtactacaag
cgccccgaggtcaccgcgagcgtcttcgaccgggacggctactaccacaccggcgacgtcatggccgag
accgcacccgaccacctggtgtacgtggaccgtcgcaacaacgtcctcaaactcgcgcagggcgagttc
gtggcggtcgccaacctggaggcggtgttctccggcgcggcgctggtgcgccagatcttcgtgtacggc
aacagcgagcgcagtttccttctggccgtggtggtcccgacgccggaggcgctcgagcagtacgatccg
gccgcgctcaaggccgcgctggccgactcgctgcagcgcaccgcacgcgacgccgaactgcaatcctac
gaggtgccggccgatttcatcgtcgagaccgagccgttcagcgccgccaacgggctgctgtcgggtgtc
ggaaaactgctgcggcccaacctcaaagaccgctacgggcagcgcctggagcagatgtacgccgatatc
gcggccacgcaggccaaccagttgcgcgaactgcggcgcgcggccgccacacaaccggtgatcgacacc
ctcacccaggccgctgccacgatcctcggcaccgggagcgaggtggcatccgacgcccacttcaccgac
ctgggcggggattccctgtcggcgctgacactttcgaacctgctgagcgatttcttcggtttcgaagtt
cccgtcggcaccatcgtgaacccggccaccaacctcgcccaactcgcccagcacatcgaggcgcagcgc
accgcgggtgaccgcaggccgagtttcaccaccgtgcacggcgcggacgccaccgagatccgggcgagt
gagctgaccctggacaagttcatcgacgccgaaacgctccgggccgcaccgggtctgcccaaggtcacc
accgagccacggacggtgttgctctcgggcgccaacggctggctgggccggttcctcacgttgcagtgg
ctggaacgcctggcacctgtcggcggcaccctcatcacgatcgtgcggggccgcgacgacgccgcggcc
cgcgcacggctgacccaggcctacgacaccgatcccgagttgtcccgccgcttcgccgagctggccgac
cgccacctgcgggtggtcgccggtgacatcggcgacccgaatctgggcctcacacccgagatctggcac
cggctcgccgccgaggtcgacctggtggtgcatccggcagcgctggtcaaccacgtgctccccctaccgg
cagctgttcggccccaacgtcgtgggcacggccgaggtgatcaagctggccctcaccgaacggatcaag
cccgtcacgtacctgtccaccgtgtcggtggccatggggatccccgacttcgaggaggacggcgacatc
cggaccgtgagcccggtgcgcccgctcgacggcggatacgccaacggctacggcaacagcaagtgggcc
ggcgaggtgctgctgcgggaggcccacgatctgtgcgggctgcccgtggcgacgttccgctcggacatg
atcctggcgcatccgcgctaccgcggtcaggtcaacgtgccagacatgttcacgcgactcctgttgagc
ctcttgatcaccggcgtcgcgccgcggtcgttctacatcggagacggtgagcgcccgcgggcgcactac
cccggcctgacggtcgatttcgtggccgaggcggtcacgacgctcggcgcgcagcagcgcgagggatac
gtgtcctacgacgtgatgaacccgcacgacgacgggatctccctggatgtgttcgtggactggctgatc
cgggcgggccatccgatcgaccgggtcgacgactacgacgactgggtgcgtcggttcgagaccgcgttg
accgcgcttcccgagaagcgccgcgcacagaccgtactgccgctgctgcacgcgttccgcgctccgcag
gcaccgttgcgcggcgcacccgaacccacggaggtgttccacgccgcggtgcgcaccgcgaaggtgggc
ccgggagacatcccgcacctcgacgaggcgctgatcgacaagtacatacgcgatctgcgtgagttcggt
ctgatctgagaattctagatctgatcgttgcgggcggggcgagagtctcgccccgcccgcgaccgcggt
gaaaatacgagaatattatttgtattgatctcctaggcggggtaccgtattttggatgataacgaggcg
caaaaaatggcggacacgttattgattctgggtgatagcctgagcgccgggtatcgaatgtctgccagc
gcggcctggcctgccttgttgaatgataagtggcagagtaaaacgtcggtagttaatgccagcatcagc
ggcgacacctcgcaacaaggactggcgcgccttccggctctgctgaaacagcatcagccgcgttgggtg
ctggttgaactgggcggcaatgacggtttgcgtggttttcagccacagcaaaccgagcaaacgctgcgc

FIG. 17 (Cont)

cagattttgcaggatgtcaaagccgccaacgctgaaccattgttaatgcaaatacgtctgcctgcaaac
tatggtcgccgttataatgaagcctttagcgccatttaccccaaactcgccaaagagtttgatgttccg
ctgctgccctttttatggaagaggtctacctcaagccacaatggatgcaggatgacggtattcatccc
aaccgcgacgcccagccgtttattgccgactggatggcgaagcagttgcagcctttagtaaatcatgac
tcataacctaggggtaccgctagcgagctctctagagaagcttgggcccgaacaaaaactcatctcaga
agaggatctgaatagcgccgtcgaccatcatcatcatcatcattgagtttaaacggtctccagcttggc
tgttttggcggatgagagaagattttcagcctgatacagattaaatcagaacgcagaagcggtctgata
aaacagaatttgcctggcggcagtagcgcggtggtcccacctgaccccatgccgaactcagaagtgaaa
cgccgtagcgccgatggtagtgtggggtctccccatgcgagagtagggaactgccaggcatcaaataaa
acgaaaggctcagtcgaaagactgggcctttcgttttatctgttgtttgtcggtgaacgctctcctgac
gcctgatgcggtattttctccttacgcatctgtgcggtatttcacaccgcatatggtgcactctcagta
caatctgctctgatgccgcatagttaagccagccccgacacccgccaacacccgctgacgagcttagta
aagccctcgctagattttaatgcggatgttgcgattacttcgccaactattgcgataacaagaaaaagc
cagcctttcatgatatatctcccaatttgtgtagggcttattatgcacgcttaaaaataataaaagcag
acttgacctgatagtttggctgtgagcaattatgtgcttagtgcatctaacgcttgagttaagccgcgc
cgcgaagcggcgtcggcttgaacgaattgttagacattatttgccgactaccttggtgatctcgccttt
cacgtagtggacaaattcttccaactgatctgcgcgcgaggccaagcgatcttcttcttgtccaagata
agcctgtctagcttcaagtatgacgggctgatactgggccggcaggcgctccattgcccagtcggcagc
gacatccttcggcgcgattttgccggttactgcgctgtaccaaatgcgggacaacgtaagcactacatt
tcgctcatcgccagcccagtcgggcggcgagttccatagcgttaaggtttcatttagcgcctcaaatag
atcctgttcaggaaccggatcaaagagttcctccgccgctggacctaccaaggcaacgctatgttctct
tgcttttgtcagcaagatagccagatcaatgtcgatcgtggctggctcgaagatacctgcaagaatgtc
attgcgctgccattctccaaattgcagttcgcgcttagctggataacgccacggaatgatgtcgtcgtg
cacaacaatggtgacttctacagcgcggagaatctcgctctctccaggggaagccgaagtttccaaaag
gtcgttgatcaaagctcgccgcgttgtttcatcaagccttacggtcaccgtaaccagcaaatcaatatc
actgtgtggcttcaggccgccatccactgcggagccgtacaaatgtacggccagcaacgtcggttcgag
atggcgctcgatgacgccaactacctctgatagttgagtcgatacttcggcgatcaccgcttccctcat
gatgtttaactttgttttagggcgactgccctgctgcgtaacatcgttgctgctccataacatcaaaca
tcgacccacggcgtaacgcgcttgctgcttggatgcccgaggcatagactgtaccccaaaaaaacagtc
ataacaagccatgaaaaccgccactgcgccgttaccaccgctgcgttcggtcaaggttctggaccagtt
gcgtgagcgcatacgctacttgcattacagcttacgaaccgaacaggcttatgtccactgggttcgtgc
cttcatccgtttccacggtgtgcgtcacccggcaaccttgggcagcagcgaagtcgaggcatttctgtc
ctggctggcgaacgagcgcaaggtttcggtctccacgcatcgtcaggcattggcggccttgctgttctt
ctacggcaaggtgctgtgcacggatctgccctggcttcaggagatcggaagacctcggccgtcgcggcg
cttgccggtggtgctgaccccggatgaagtggttcgcatcctcggttttctggaaggcgagcatcgttt
gttcgcccagcttctgtatggaacgggcatgcggatcagtgagggtttgcaactgcgggtcaaggatct
ggatttcgatcacggcacgatcatcgtgcgggagggcaagggctccaaggatcgggccttgatgttacc
cgagagcttggcacccagcctgcgcgagcaggggaattaattcccacgggttttgctgcccgcaaacgg
gctgttctggtgttgctagtttgttatcagaatcgcagatccggcttcagccggtttgccggctgaaag
cgctatttcttccagaattgccatgatttttttccccacgggaggcgtcactggctcccgtgttgtcggc
agctttgattcgataagcagcatcgcctgtttcaggctgtctatgtgtgactgttgagctgtaacaagt
tgtctcaggtgttcaatttcatgttctagttgctttgttttactggtttcacctgttctattaggtgtt
acatgctgttcatctgttacattgtcgatctgttcatggtgaacagctttgaatgcaccaaaaactcgt

FIG. 17 (Cont)

aaaagctctgatgtatctatctttttacaccgttttcatctgtgcatatggacagttttccctttgat
atgtaacggtgaacagttgttctacttttgtttgttagtcttgatgcttcactgatagatacaagagcc
ataagaacctcagatccttccgtatttagccagtatgttctctagtgtggttcgttgtttttgcgtgag
ccatgagaacgaaccattgagatcatacttactttgcatgtcactcaaaaattttgcctcaaaactggt
gagctgaatttttgcagttaaagcatcgtgtagtgtttttcttagtccgttatgtaggtaggaatctga
tgtaatggttgttggtattttgtcaccattcatttttatctggttgttctcaagttcggttacgagatc
catttgtctatctagttcaacttggaaaatcaacgtatcagtcgggcggcctcgcttatcaaccaccaa
tttcatattgctgtaagtgtttaaatctttacttattggtttcaaaacccattggttaagccttttaaa
ctcatggtagttattttcaagcattaacatgaacttaaattcatcaaggctaatctctatatttgcctt
gtgagttttcttttgtgttagttcttttaataaccactcataaatcctcatagagtatttgttttcaaa
agacttaacatgttccagattatattttatgaatttttttaactggaaaagataaggcaatatctcttc
actaaaaactaattctaatttttcgcttgagaacttggcatagtttgtccactggaaaatctcaaagcc
tttaaccaaaggattcctgatttccacagttctcgtcatcagctctctggttgctttagctaatacacc
ataagcattttccctactgatgttcatcatctgagcgtattggttataagtgaacgataccgtccgttc
tttccttgtagggttttcaatcgtggggttgagtagtgccacacagcataaaattagcttggtttcatg
ctccgttaagtcatagcgactaatcgctagttcatttgctttgaaaacaactaattcagacatacatct
caattggtctaggtgattttaat

TER_*Euglena gracilis* (Q5EU90)without N-terminal transit peptide

Amino acid sequence (SEQ ID NO:228)

```
  1 VPQMAEGFSG EATSAWAAAG PQWAAPLVAA ASSALALWWW AARRSVRRPL AALAELPTAV
 61 THLAPPMAMF TTTAKVIQPK IRGFICTTTH PIGCEKRVQE EIAYARAHPP TSPGPKRVLV
121 IGCSTGYGLS TRITAAFGYQ AATLGVFLAG PPTKGRPAAA GWYNTVAFEK AALEAGLYAR
181 SLNGDAFDST TKARTVEAIK RDLGTVDLVV YSIAAPKRTD PATGVLHKAC LKPIGATYTN
241 RTVNTDKAEV TDVSIEPASP EEIADTVKVM GGEDWELWIQ ALSEAGVLAE GAKTVAYSYI
301 GPEMTWPVYW SGTIGEAKKD VEKAAKRITQ QYGCPAYPVV AKALVTQASS AIPVVPLYIC
361 LLYRVMKEKG THEGCIEQMV RLLTTKLYPE NGAPIVDEAG RVRVDDWEMA EDVQQAVKDL
421 WSQVSTANLK DISDFAGYQT EFLRLFGFGI DGVDYDQPVD VEADLPSAAQ Q
```

fadA (YP026272)

Nucleotide sequence (SEQ ID NO:229)

```
  1 atggaacagg ttgtcattgt cgatgcaatt cgcaccccga tgggccgttc gaagggcggt
 61 gcttttcgta acgtgcgtgc agaagatctc tccgctcatt taatgcgtag cctgctggcg
121 cgtaacccgg cgctggaagc ggcggccctc gacgatattt actggggttg tgtgcagcag
181 acgctggagc agggttttaa tatcgcccgt aacgcggcgc tgctggcaga agtaccacac
241 tctgtcccgg cggttaccgt taatcgcttg tgtggttcat ccatgcaggc actgcatgac
```

FIG. 17 (Cont)

```
301 gcagcacgaa tgatcatgac tggcgatgcg caggcatgtc tggttggcgg cgtggagcat
361 atgggccatg tgccgatgag tcacggcgtc gattttcacc ccggcctgag ccgcaatgtc
421 gccaaagcgg cgggcatgat gggcttaacg gcagaaatgc tggcgcgtat gcacggtatc
481 agccgtgaaa tgcaggatgc ctttgccgcg cggtcacacg cccgcgcctg ggccgccacg
541 cagtcggccg catttaaaaa tgaaatcatc ccgaccggtg gtcacgatgc cgacggcgtc
601 ctgaagcagt ttaattacga cgaagtgatt cgcccggaaa ccaccgtgga agccctcgcc
661 acgctgcgtc cggcgtttga tccagtaaac ggtatggtaa cggcgggcac atcttctgca
721 ctttccgatg gcgcagctgc catgctggtg atgagtgaaa gccgcgccca tgaattaggt
781 cttaagccgc gcgctcgtgt gcgttcgatg gcggtcgttg gttgtgaccc atcgattatg
841 ggttacggcc cggttccggc ctcgaaactg gcgctgaaaa aagcggggct ttctgccagc
901 gatatcggcg tgtttgaaat gaacgaagcc tttgccgcgc agatcctgcc atgtattaaa
961 gatctgggac taattgagca gattgacgag aagatcaacc tcaacggtgg cgcgatcgcg
1021 ctgggtcatc cgctgggttg ttccggtgcg cgtatcagca ccacgctgct gaatctgatg
1081 gaacgcaaag acgttcagtt tggtctggcg acgatgtgta tcggtctggg tcagggtatt
1141 gcgacggtgt ttgagcgggt ttaa
```

Amino acid sequence (SEQ ID NO:230)

```
  1 MEQVVIVDAI RTPMGRSKGG AFRNVRAEDL SAHLMRSLLA RNPALEAAAL DDIYWGCVQQ
 61 TLEQGFNIAR NAALLAEVPH SVPAVTVNRL CGSSMQALHD AARMIMTGDA QACLVGGVEH
121 MGHVPMSHGV DFHPGLSRNV AKAAGMMGLT AEMLARMHGI SREMQDAFAA RSHARAWAAT
181 QSAAFKNEII PTGGHDADGV LKQFNYDEVI RPETTVEALA TLRPAFDPVN GMVTAGTSSA
241 LSDGAAAMLV MSESRAHELG LKPRARVRSM AVVGCDPSIM GYGPVPASKL ALKKAGLSAS
301 DIGVFEMNEA FAAQILPCIK DLGLIEQIDE KINLNGGAIA LGHPLGCSGA RISTTLLNLM
361 ERKDVQFGLA TMCIGLGQGI ATVFERV
```

fadB (NP_416843)

Nucleotide sequence (SEQ ID NO:231)

```
   1 atgctttaca aaggcgacac cctgtacctt gactggctgg aagatggcat tgccgaactg
  61 gtatttgatg ccccaggttc agttaataaa ctcgacactg cgaccgtcgc cagcctcggc
 121 gaggccatcg gcgtgctgga acagcaatca gatctaaaag ggctgctgct gcgttcgaac
 181 aaagcagcct ttatcgtcgg tgctgatatc accgaatttt tgtccctgtt cctcgttcct
 241 gaagaacagt taagtcagtg gctgcacttt gccaatagcg tgtttaatcg cctggaagat
 301 ctgccggtgc cgaccattgc tgccgtcaat ggctatgcgc tgggcggtgg ctgcgaatgc
 361 gtgctggcga ccgattatcg tctggcgacg ccggatctgc gcatcggtct gccggaaacc
 421 aaactgggca tcatgcctgg ctttggcggt tctgtacgta tgccacgtat gctgggcgct
```

FIG. 17 (Cont)

```
481 gacagtgcgc tggaaatcat tgccgccggt aaagatgtcg gcgcggatca ggcgctgaaa
541 atcggtctgg tggatggcgt agtcaaagca gaaaaactgg ttgaaggcgc aaaggcggtt
601 ttacgccagg ccattaacgg cgacctcgac tggaaagcaa aacgtcagcc gaagctggaa
661 ccactaaaac tgagcaagat tgaagccacc atgagcttca ccatcgctaa agggatggtc
721 gcacaaacag cggggaaaca ttatccggcc cccatcaccg cagtaaaaac cattgaagct
781 gcggcccgtt ttggtcgtga agaagcctta aacctggaaa acaaaagttt tgtcccgctg
 841 gcgcatacca acgaagcccg cgcactggtc ggcattttcc ttaacgatca atatgtaaaa
 901 ggcaaagcga agaaactcac caaagacgtt gaaaccccga aacaggccgc ggtgctgggt
 961 gcaggcatta tgggcggcgg catcgcttac cagtctgcgt ggaaaggcgt gccggttgtc
1021 atgaaagata tcaacgacaa gtcgttaacc ctcggcatga ccgaagccgc gaaactgctg
1081 aacaagcagc ttgagcgcgg caagatcgat ggtctgaaac tggctggcgt gatctccaca
1141 atccacccaa cgctcgacta cgccggattt gaccgcgtgg atattgtggt agaagcggtt
1201 gttgaaaacc cgaaagtgaa aaaagccgta ctggcagaaa ccgaacaaaa agtacgccag
1261 gataccgtgc tggcgtctaa cacttcaacc attcctatca gcgaactggc caacgcgctg
1321 gaacgcccgg aaaacttctg cgggatgcac ttctttaacc cggtccaccg aatgccgttg
1381 gtagaaatta ttcgcggcga gaaaagctcc gacgaaacca tcgcgaaagt tgtcgcctgg
1441 gcgagcaaga tgggcaagac gccgattgtg gttaacgact gccccggctt ctttgttaac
1501 cgcgtgctgt tcccgtattt cgccggtttc agccagctgc tgcgcgacgg cgcggatttc
1561 cgcaagatcg acaaagtgat ggaaaaacag tttggctggc cgatgggccc ggcatatctg
1621 ctggacgttg tgggcattga taccgcgcat cacgctcagg ctgtcatggc agcaggcttc
1681 ccgcagcgga tgcagaaaga ttaccgcgat gccatcgacg cgctgtttga tgccaaccgc
1741 tttggtcaga agaacggcct cggtttctgg cgttataaag aagacagcaa aggtaagccg
1801 aagaaagaag aagacgccgc cgttgaagac ctgctggcag aagtgagcca gccgaagcgc
1861 gatttcagcg aagaagagat tatcgcccgc atgatgatcc cgatggtcaa cgaagtggtg
1921 cgctgtctgg aggaaggcat tatcgccact ccggcggaag cggatatggc gctggtctac
1981 ggcctgggct tccctccgtt ccacggcggc gcgttccgct ggctggacac cctcggtagc
2041 gcaaaatacc tcgatatggc acagcaatat cagcacctcg gcccgctgta tgaagtgccg
2101 gaaggtctgc gtaataaagc gcgtcataac gaaccgtact atcctccggt tgagccagcc
2161 cgtccggttg gcgacctgaa aacggcttaa
```

Amino acid sequence (SEQ ID NO:232)

```
  1 MLYKGDTLYL DWLEDGIAEL VFDAPGSVNK LDTATVASLG EAIGVLEQQS DLKGLLLRSN
 61 KAAFIVGADI TEFLSLFLVP EEQLSQWLHF ANSVFNRLED LPVPTIAAVN GYALGGGCEC
121 VLATDYRLAT PDLRIGLPET KLGIMPGFGG SVRMPRMLGA DSALEIIAAG KDVGADQALK
181 IGLVDGVVKA EKLVEGAKAV LRQAINGDLD WKAKRQPKLE PLKLSKIEAT MSFTIAKGMV
241 AQTAGKHYPA PITAVKTIEA AARFGREEAL NLENKSFVPL AHTNEARALV GIFLNDQYVK
301 GKAKKLTKDV ETPKQAAVLG AGIMGGGIAY QSAWKGVPVV MKDINDKSLT LGMTEAAKLL
361 NKQLERGKID GLKLAGVIST IHPTLDYAGF DRVDIVVEAV VENPKVKKAV LAETEQKVRQ
421 DTVLASNTST IPISELANAL ERPENFCGMH FFNPVHRMPL VEIIRGEKSS DETIAKVVAW
481 ASKMGKTPIV VNDCPGFFVN RVLFPYFAGF SQLLRDGADF RKIDKVMEKQ FGWPMGPAYL
541 LDVVGIDTAH HAQAVMAAGF PQRMQKDYRD AIDALFDANR FGQKNGLGFW RYKEDSKGKP
601 KKEEDAAVED LLAEVSQPKR DFSEEEIIAR MMIPMVNEVV RCLEEGIIAT PAEADMALVY
```

FIG. 17 (Cont)

661 GLGFPPFHGG AFRWLDTLGS AKYLDMAQQY QHLGPLYEVP EGLRNKARHN EPYYPPVEPA
721 RPVGDLKTA

fadl (NP_416844)

Nucleotide sequence (SEQ ID NO:233)

```
   1 atgggtcagg ttttaccgct ggttacccgc cagggcgatc gtatcgccat tgttagcggt
  61 ttacgtacgc cttttgcccg tcaggcgacg gcttttcatg gcattcccgc ggttgattta
 121 gggaagatgg tggtaggcga actgctggca cgcagcgaga tccccgccga agtgattgaa
 181 caactggtct ttggtcaggt cgtacaaatg cctgaagccc ccaacattgc gcgtgaaatt
 241 gttctcggta cgggaatgaa tgtacatacc gatgcttaca gcgtcagccg cgcttgcgct
 301 accagtttcc aggcagttgc aaacgtcgca gaaagcctga tggcgggaac tattcgagcg
 361 gggattgccg gtggggcaga ttcctcttcg gtattgccaa ttggcgtcag taaaaaactg
 421 gcgcgcgtgc tggttgatgt caacaaagct cgtaccatga gccagcgact gaaactcttc
 481 tctcgcctgc gtttgcgcga cttaatgccc gtaccacctg cggtagcaga atattctacc
 541 ggcttgcgga tgggcgacac cgcagagcaa atggcgaaaa cctacggcat cacccgagaa
 601 cagcaagatg cattagcgca ccgttcgcat cagcgtgccg ctcaggcatg gtcagacgga
 661 aaactcaaag aagaggtgat gactgccttt atccctcctt ataaacaacc gcttgtcgaa
 721 gacaacaata ttcgcggtaa ttcctcgctt gccgattacg caaagctgcg cccggcgttt
 781 gatcgcaaac acggaacggt aacggcggca aacagtacgc cgctgaccga tggcgcggca
 841 gcggtgatcc tgatgactga atcccgggcg aaagaattag ggctggtgcc gctggggtat
 901 ctgcgcagct acgcatttac tgcgattgat gtctggcagg acatgttgct cggtccagcc
 961 tggtcaacac cgctggcgct ggagcgtgcc ggtttgacga tgagcgatct gacattgatc
1021 gatatgcacg aagcctttgc agctcagacg ctggcgaata ttcagttgct gggtagtgaa
1081 cgttttgctc gtgaagcact ggggcgtgca catgccactg gcgaagtgga cgatagcaaa
1141 tttaacgtgc ttggcggttc gattgcttac gggcatccct tcgcggcgac cggcgcgcgg
1201 atgattaccc agacattgca tgaacttcgc cgtcgcggcg gtggatttgg tttagttacc
1261 gcctgtgctg ccggtgggct tggcgcggca atggttctgg aggcggaata a
```

Amino acid sequence (SEQ ID NO:234)

```
   1 MGQVLPLVTR QGDRIAIVSG LRTPFARQAT AFHGIPAVDL GKMVVGELLA RSEIPAEVIE
  61 QLVFGQVVQM PEAPNIAREI VLGTGMNVHT DAYSVSRACA TSFQAVANVA ESLMAGTIRA
 121 GIAGGADSSS VLPIGVSKKL ARVLVDVNKA RTMSQRLKLF SRLRLRDLMP VPPAVAEYST
 181 GLRMGDTAEQ MAKTYGITRE QQDALAHRSH QRAAQAWSDG KLKEEVMTAF IPPYKQPLVE
 241 DNNIRGNSSL ADYAKLRPAF DRKHGTVTAA NSTPLTDGAA AVILMTESRA KELGLVPLGY
 301 LRSYAFTAID VWQDMLLGPA WSTPLALERA GLTMSDLTLI DMHEAFAAQT LANIQLLGSE
 361 RFAREALGRA HATGEVDDSK FNVLGGSIAY GHPFAATGAR MITQTLHELR RRGGGFGLVT
 421 ACAAGGLGAA MVLEAE
```

FIG. 17 (Cont)

fadJ (NP_416843)

Nucleotide sequence (SEQ ID NO:235)

```
   1 atggaaatga catcagcgtt tacccttaat gttcgtctgg acaacattgc cgttatcacc
  61 atcgacgtac cgggtgagaa aatgaatacc ctgaaggcgg agtttgcctc gcaggtgcgc
 121 gccattatta agcaactccg tgaaaacaaa gagttgcgag gcgtggtgtt tgtctccgct
 181 aaaccggaca acttcattgc tggcgcagac atcaacatga tcggcaactg caaaacggcg
 241 caagaagcgg aagctctggc gcggcagggc caacagttga tggcggagat tcatgctttg
 301 cccattcagg ttatcgcggc tattcatggc gcttgcctgg gtggtgggct ggagttggcg
 361 ctggcgtgcc acggtcgcgt ttgtactgac gatcctaaaa cggtgctcgg tttgcctgaa
 421 gtacaacttg gattgttacc cggttcaggc ggcacccagc gtttaccgcg tctgataggc
 481 gtcagcacag cattagagat gatcctcacc ggaaaacaac ttcgggcgaa acaggcatta
 541 aagctggggc tggtggatga cgttgttccg cactccattc tgctggaagc cgctgttgag
 601 ctggcaaaga aggagcgccc atcttcccgc cctctacctg tacgcgagcg tattctggcg
 661 gggccgttag gtcgtgcgct gctgttcaaa atggtcggca agaaaacaga acacaaaact
 721 caaggcaatt atccggcgac agaacgcatc ctggaggttg ttgaaacggg attagcgcag
 781 ggcaccagca gcggttatga cgccgaagct cgggcgtttg gcgaactggc gatgacgcca
 841 caatcgcagg cgctgcgtag tatctttttt gccagtacgg acgtgaagaa agatcccggc
 901 agtgatgcgc cgcctgcgcc attaaacagc gtggggattt taggtggtgg cttgatgggc
 961 ggcggtattg cttatgtcac tgcttgtaaa gcggggattc cggtcagaat taaagatatc
1021 aacccgcagg gcataaatca tgcgctgaag tacagttggg atcagctgga gggcaaagtt
1081 cgccgtcgtc atctcaaagc cagcgaacgt gacaaacagc tggcattaat ctccggaacg
1141 acggactatc gcggctttgc ccatcgcgat ctgattattg aagcggtgtt tgaaaatctc
1201 gaattgaaac aacagatggt ggcggaagtt gagcaaaatt gcgccgctca taccatcttt
1261 gcttcgaata cgtcatcttt accgattggt gatatcgccg ctcacgccac gcgacctgag
1321 caagttatcg gcctgcattt cttcagtccg gtggaaaaaa tgccgctggt ggagattatt
1381 cctcatgcgg ggacatcggc gcaaaccatc gctaccacag taaaactggc gaaaaaacag
1441 ggtaaaacgc caattgtcgt gcgtgacaaa gccggttttt acgtcaatcg catcttagcg
1501 ccttacatta atgaagctat ccgcatgttg acccaaggtg aacgggtaga gcacattgat
1561 gccgcgctag tgaaatttgg ttttccggta ggcccaatcc aacttttgga tgaggtagga
1621 atcgacaccg ggactaaaat tattcctgta ctggaagccg cttatggaga acgttttagc
1681 gcgcctgcaa atgttgtttc ttcaattttg aacgacgatc gcaaaggcag aaaaaatggc
1741 cggggtttct atctttatgg tcagaaaggg cgtaaaagca aaaaacaggt cgatcccgcc
1801 atttacccgc tgattggcac acaagggcag gggcgaatct ccgcaccgca ggttgctgaa
1861 cggtgtgtga tgttgatgct gaatgaagca gtacgttgtg ttgatgagca ggttatccgt
1921 agcgtgcgtg acggggatat tggcgcggta tttggcattg gttttccgcc atttctcggt
1981 ggaccgttcc gctatatcga ttctctcggc gcgggcgaag tggttgcaat aatgcaacga
2041 cttgccacgc agtatggttc ccgttttacc ccttgcgagc gtttggtcga gatgggcgcg
2101 cgtggggaaa gtttttggaa aacaactgca actgacctgc aataa
```

FIG. 17 (Cont)

Amino Acid sequence (SEQ ID NO:236)

```
  1 MEMTSAFTLN VRLDNIAVIT IDVPGEKMNT LKAEFASQVR AIIKQLRENK ELRGVVFVSA
 61 KPDNFIAGAD INMIGNCKTA QEAEALARQG QQLMAEIHAL PIQVIAAIHG ACLGGGLELA
121 LACHGRVCTD DPKTVLGLPE VQLGLLPGSG GTQRLPRLIG VSTALEMILT GKQLRAKQAL
181 KLGLVDDVVP HSILLEAAVE LAKKERPSSR PLPVRERILA GPLGRALLFK MVGKKTEHKT
241 QGNYPATERI LEVVETGLAQ GTSSGYDAEA RAFGELAMTP QSQALRSIFF ASTDVKKDPG
301 SDAPPAPLNS VGILGGGLMG GGIAYVTACK AGIPVRIKDI NPQGINHALK YSWDQLEGKV
361 RRRHLKASER DKQLALISGT TDYRGFAHRD LIIEAVFENL ELKQQMVAEV EQNCAAHTIF
421 ASNTSSLPIG DIAAHATRPE QVIGLHFFSP VEKMPLVEII PHAGTSAQTI ATTVKLAKKQ
481 GKTPIVVRDK AGFYVNRILA PYINEAIRML TQGERVEHID AALVKFGFPV GPIQLLDEVG
541 IDTGTKIIPV LEAAYGERFS APANVVSSIL NDDRKGRKNG RGFYLYGQKG RKSKKQVDPA
601 IYPLIGTQGQ GRISAPQVAE RCVMLMLNEA VRCVDEQVIR SVRDGDIGAV FGIGFPPFLG
661 GPFRYIDSLG AGEVVAIMQR LATQYGSRFT PCERLVEMGA RGESFWKTTA TDLQ
```

fabI (NP_415804)

Nucleotide sequence (SEQ ID NO:237)

```
  1 atgggttttc tttccggtaa gcgcattctg gtaaccggtg ttgccagcaa actatccatc
 61 gcctacggta tcgctcaggc gatgcaccgc gaaggagctg aactggcatt cacctaccag
121 aacgacaaac tgaaaggccg cgtagaagaa tttgccgctc aattgggttc tgacatcgtt
181 ctgcagtgcg atgttgcaga agatgccagc atcgacacca tgttcgctga actggggaaa
241 gtttggccga aatttgacgg tttcgtacac tctattggtt ttgcacctgg cgatcagctg
301 gatggtgact atgttaacgc cgttacccgt gaaggcttca aaattgccca cgacatcagc
361 tcctacagct tcgttgcaat ggcaaaagct tgccgctcca tgctgaatcc gggttctgcc
421 ctgctgaccc tttcctacct tggcgctgag cgcgctatcc cgaactacaa cgttatgggt
481 ctggcaaaag cgtctctgga agcgaacgtg cgctatatgg cgaacgcgat gggtccggaa
541 ggtgtgcgtg ttaacgccat ctctgctggt ccgatccgta ctctggcggc ctccggtatc
601 aaagacttcc gcaaaatgct ggctcattgc gaagccgtta ccccgattcg ccgtaccgtt
661 actattgaag atgtgggtaa ctctgcggca ttcctgtgct ccgatctctc tgccggtatc
721 tccggtgaag tggtccacgt tgacggcggt ttcagcattg ctgcaatgaa cgaactcgaa
781 ctgaaataa
```

Amino acid sequence (SEQ ID NO:238)

```
  1 MGFLSGKRIL VTGVASKLSI AYGIAQAMHR EGAELAFTYQ NDKLKGRVEE FAAQLGSDIV
 61 LQCDVAEDAS IDTMFAELGK VWPKFDGFVH SIGFAPGDQL DGDYVNAVTR EGFKIAHDIS
121 SYSFVAMAKA CRSMLNPGSA LLTLSYLGAE RAIPNYNVMG LAKASLEANV RYMANAMGPE
181 GVRVNAISAG PIRTLAASGI KDFRKMLAHC EAVTPIRRTV TIEDVGNSAA FLCSDLSAGI
241 SGEVVHVDGG FSIAAMNELE LK
```

tesB (NP_414986)

FIG. 17 (Cont)

Nucleotide sequence (SEQ ID NO:239)

```
  1 atgagtcagg cgctaaaaaa tttactgaca ttgttaaatc tggaaaaaat tgaggaagga
 61 ctctttcgcg gccagagtga agatttaggt ttacgccagg tgtttggcgg ccaggtcgtg
121 ggtcaggcct tgtatgctgc aaaagagacc gtccctgaag agcggctggt acattcgttt
181 cacagctact ttcttcgccc tggcgatagt aagaagccga ttatttatga tgtcgaaacg
241 ctgcgtgacg gtaacagctt cagcgcccgc cgggttgctg ctattcaaaa cggcaaaccg
301 atttttttata tgactgcctc tttccaggca ccagaagcgg gtttcgaaca tcaaaaaaca
361 atgccgtccg cgccagcgcc tgatggcctc ccttcggaaa cgcaaatcgc ccaatcgctg
421 gcgcacctgc tgccgccagt gctgaaagat aaattcatct gcgatcgtcc gctggaagtc
481 cgtccggtgg agtttcataa cccactgaaa ggtcacgtcg cagaaccaca tcgtcaggtg
541 tggatccgcg caaatggtag cgtgccggat gacctgcgcg ttcatcagta tctgctcggt
601 tacgcttctg atcttaactt cctgccggta gctctacagc cgcacggcat cggttttctc
661 gaaccgggga ttcagattgc caccattgac cattccatgt ggttccatcg cccgtttaat
721 ttgaatgaat ggctgctgta tagcgtggag agcacctcgg cgtccagcgc acgtggcttt
781 gtgcgcggtg agttttatac ccaagacggc gtactggttg cctcgaccgt tcaggaaggg
841 gtgatgcgta atcacaatta a
```

Amino acid sequence (SEQ ID NO:240)

```
  1 MSQALKNLLT LLNLEKIEEG LFRGQSEDLG LRQVFGGQVV GQALYAAKET VPEERLVHSF
 61 HSYFLRPGDS KKPIIYDVET LRDGNSFSAR RVAAIQNGKP IFYMTASFQA PEAGFEHQKT
121 MPSAPAPDGL PSETQIAQSL AHLLPPVLKD KFICDRPLEV RPVEFHNPLK GHVAEPHRQV
181 WIRANGSVPD DLRVHQYLLG YASDLNFLPV ALQPHGIGFL EPGIQIATID HSMWFHRPFN
241 LNEWLLYSVE STSASSARGF VRGEFYTQDG VLVASTVQEG VMRNHN
```

Acr1_Acinetobacter (YP_047869)

Nucleotide sequence (SEQ ID NO:241)

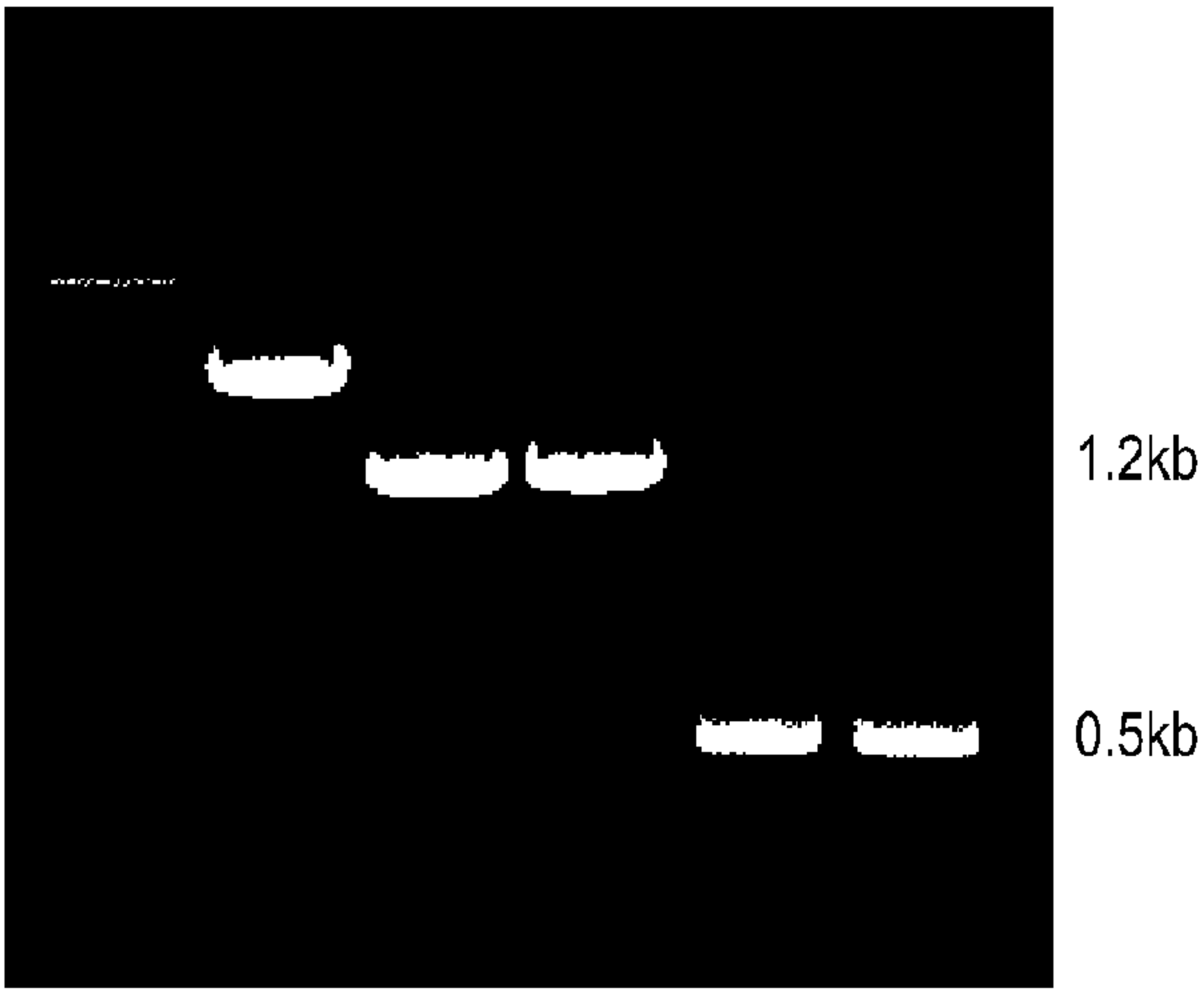
```
  1 ttgatatcaa tcagggaaaa acgcgtgaac aaaaaacttg aagctctctt ccgagagaat
 61 gtaaaaggta aagtggcttt gatcactggt gcatctagtg gaatcggttt gacgattgca
121 aaaagaattg ctgcggcagg tgctcatgta ttattggttg cccgaaccca agaaacactg
181 gaagaagtga aagctgcaat tgaacagcaa gggggacagg cctctatttt tccttgtgac
241 ctgactgaca tgaatgcgat tgaccagtta tcacaacaaa ttatggccag tgtcgatcat
301 gtcgatttcc tgatcaataa tgcagggcgt tcgattcgcc gtgccgtaca cgagtcgttt
361 gatcgcttcc atgattttga acgcaccatg cagctgaatt actttggtgc ggtacgttta
421 gtgttaaatt tactgccaca tatgattaag cgtaaaaatg gccagatcat caatatcagc
481 tctattggtg tattggccaa tgcgacccgt ttttctgctt atgtcgcgtc taaagctgcg
541 ctggatgcct tcagtcgctg tctttcagcc gaggtactca agcataaaat ctcaattacc
601 tcgatttata tgccattggt gcgtacccca atgatcgcac ccaccaaaat ttataaatac
661 gtgcccacgc tttccccaga agaagccgca gatctcattg tctacgccat tgtgaaacgt
721 ccaaaacgta ttgcgacgca cttgggtcgt ctggcgtcaa ttacctatgc catcgcacca
```

FIG. 17 (Cont)

```
781 gacatcaata atattctgat gtcgattgga tttaacctat tcccaagctc aacggctgca
841 ctgggtgaac aggaaaaatt gaatctgcta caacgtgcct atgcccgctt gttcccaggc
901 gaacactggt aa
```

Amino acid sequence (SEQ ID NO:242)

```
  1 MISIREKRVN KKLEALFREN VKGKVALITG ASSGIGLTIA KRIAAAGAHV LLVARTQETL
 61 EEVKAAIEQQ GGQASIFPCD LTDMNAIDQL SQQIMASVDH VDFLINNAGR SIRRAVHESF
121 DRFHDFERTM QLNYFGAVRL VLNLLPHMIK RKNGQIINIS SIGVLANATR FSAYVASKAA
181 LDAFSRCLSA EVLKHKISIT SIYMPLVRTP MIAPTKIYKY VPTLSPEEAA DLIVYAIVKR
241 PKRIATHLGR LASITYAIAP DINNILMSIG FNLFPSSTAA LGEQEKLNLL QRAYARLFPG
301 EHW
```


wt
ΔfadD::Cm
ΔfadD
1.2kb
0.5kb