



(86) Date de dépôt PCT/PCT Filing Date: 2012/07/05
(87) Date publication PCT/PCT Publication Date: 2013/01/10
(45) Date de délivrance/Issue Date: 2021/06/29
(85) Entrée phase nationale/National Entry: 2014/01/03
(86) N° demande PCT/PCT Application No.: US 2012/045622
(87) N° publication PCT/PCT Publication No.: 2013/006733
(30) Priorités/Priorities: 2011/07/06 (US61/505,092);
2011/08/12 (US61/523,216)

(51) Cl.Int./Int.Cl. *C12N 9/02* (2006.01),
C12N 15/81 (2006.01), *C12P 7/44* (2006.01)
(72) Inventeurs/Inventors:
BEARDSLEE, TOM, US;
PICATAGGIO, STEPHEN, US;
HUTAGALUNG, ALEX, US;
FAHLAND, TOM, US
(73) Propriétaire/Owner:
RADICI CHIMICA S.P.A., IT
(74) Agent: SMART & BIGGAR LLP

(54) Titre : PROCEDES BIOLOGIQUES POUR LA PREPARATION D'ACIDE GRAS DICARBOXYLIQUE
(54) Title: BIOLOGICAL METHODS FOR PREPARING A FATTY DICARBOXYLIC ACID

(57) **Abrégé/Abstract:**

The technology relates in part to biological methods for producing a fatty dicarboxylic acid and engineered microorganisms capable of such production.

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

(19) World Intellectual Property
Organization
International Bureau



(43) International Publication Date
10 January 2013 (10.01.2013)

WIPO | PCT

(10) International Publication Number
WO 2013/006733 A3

- (51) **International Patent Classification:**
C12N 9/02 (2006.01) *C12P 7/44* (2006.01)
C12N 15/81 (2006.01)
- (21) **International Application Number:**
PCT/US2012/045622
- (22) **International Filing Date:**
5 July 2012 (05.07.2012)
- (25) **Filing Language:** English
- (26) **Publication Language:** English
- (30) **Priority Data:**
61/505,092 6 July 2011 (06.07.2011) US
61/523,216 12 August 2011 (12.08.2011) US
- (71) **Applicant** (for all designated States except US): **VER-DEZYNE, INC.** [US/US]; 2715 Loker Avenue West, Carlsbad, CA 92010 (US).
- (72) **Inventors; and**
- (75) **Inventors/Applicants** (for US only): **BEARDSLEE, Tom** [US/US]; 2715 Loker Avenue West, Carlsbad, CA 92010 (US). **PICATAGGIO, Stephen** [US/US]; 2715 Loker Avenue West, Carlsbad, CA 92010 (US). **HUTAGALUNG, Alex** [US/US]; 2715 Loker Avenue West, Carlsbad, CA 92010 (US). **FAHLAND, Tom** [US/US]; 2715 Loker Avenue West, Carlsbad, CA 92010 (US).
- (74) **Agents:** **FORCE, Walker R.** et al.; Grant Anderson LLP, c/o PortfolioIP, P.O. Box 52050, Minneapolis, MN 55402 (US).
- (81) **Designated States** (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.
- (84) **Designated States** (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).
- Published:**
— with international search report (Art. 21(3))
— with sequence listing part of description (Rule 5.2(a))
- (88) **Date of publication of the international search report:**
25 April 2013

(54) **Title:** BIOLOGICAL METHODS FOR PREPARING A FATTY DICARBOXYLIC ACID

(57) **Abstract:** The technology relates in part to biological methods for producing a fatty dicarboxylic acid and engineered microorganisms capable of such production.

WO 2013/006733 A3

BIOLOGICAL METHODS FOR PREPARING A FATTY DICARBOXYLIC ACID

Related Patent Applications

- 5 This patent application claims the benefit of U.S. Provisional Patent Application No. 61/505,092 filed on July 6, 2011, entitled "BIOLOGICAL METHODS FOR PREPARING SEBACIC ACID" naming Stephen Picataggio and Tom Beardslee as inventors, and designated by Attorney Docket No. VRD-1005-PV. This patent application also claims the benefit of U.S. provisional patent application no. 61/523,216 filed August 12, 2011, entitled "BIOLOGICAL METHODS FOR
- 10 PREPARING DODECANEDIOIC ACID" naming Stephen Picataggio and Tom Beardslee as inventors, and designated by Attorney Docket No. VRD-1006-PV.

Sequence Listing

- 15 This description contains a sequence listing in electronic form in ASCII text format. A copy of the sequence listing in electronic form is available from the Canadian Intellectual Property Office.

Field

- 20 The technology relates in part to biological methods for producing a fatty dicarboxylic acid and engineered microorganisms capable of such production.

Background

- 25 Microorganisms employ various enzyme-driven biological pathways to support their own metabolism and growth. A cell synthesizes native proteins, including enzymes, in vivo from deoxyribonucleic acid (DNA). DNA first is transcribed into a complementary ribonucleic acid (RNA) that comprises a ribonucleotide sequence encoding the protein. RNA then directs translation of the encoded protein by interaction with various cellular components, such as ribosomes. The

resulting enzymes participate as biological catalysts in pathways involved in production of molecules by the organism.

These pathways can be exploited for the harvesting of the naturally produced products. The pathways also can be altered to increase production or to produce different products that may be commercially valuable. Advances in recombinant molecular biology methodology allow researchers to isolate DNA from one organism and insert it into another organism, thus altering the cellular synthesis of enzymes or other proteins. Advances in recombinant molecular biology methodology also allow endogenous genes, carried in the genomic DNA of a microorganism, to be increased in copy number, thus altering the cellular synthesis of enzymes or other proteins. Such genetic engineering can change the biological pathways within the host organism, causing it to produce a desired product. Microorganic industrial production can minimize the use of caustic chemicals and the production of toxic byproducts, thus providing a "clean" source for certain compounds. The use of appropriate plant derived feedstocks allows production of "green" compounds while further minimizing the need for and use of petroleum derived compounds.

Summary

Provided in certain aspects is a genetically modified yeast that includes an active, modified endogenous acyl-coA oxidase polypeptide or an active, modified endogenous acyl-coA dehydrogenase polypeptide, which yeast is capable of producing a diacid from a feedstock comprising one or more components from a vegetable oil.

In certain instances a modified endogenous acyl-coA oxidase polypeptide comprises one or more amino acid modifications in one or more structures chosen from the N-terminal loop, D alpha helix, loop between the D alpha helix and the E' alpha helix, an amino acid in effective contact with carbons 6 to 9 in a feedstock component, an amino acid in effective contact with carbons 10 to 12 in a feedstock component, L alpha helix, loop C-terminal to the L alpha helix, and loop between the L alpha helix and the M alpha helix.

In some instances a modified endogenous acyl-coA oxidase polypeptide is a modified POX4 or POX5 polypeptide from a *Candida* spp. yeast (e.g., strain ATCC20336 or ATCC20962). In some cases a modified POX4 polypeptide comprises a modified amino acid sequence of SEQ ID NO: 30. Sometimes the POX4 polypeptide comprises an amino acid modification at one or more amino

acid positions chosen from 88, 90, 96, 98, 99, 100, 102, 103, 302, 309, 310, 473, 474, 475, 476, 477, 478, 479, 480, 481, 482, 483, 484, 485, 486, 487, 488, 489, 490, 491, 492, 493, 494, 495, 496, 497, 498, 499, 500, 501, 502, 503, 504 and 505. A modified endogenous acyl-coA oxidase polypeptide that is not a modified POX4 polypeptide can include an amino acid modification at one or more positions corresponding to one or more of the foregoing positions in the POX4 polypeptide.

In some instances a modified POX5 polypeptide comprises a modified amino acid sequence of SEQ ID NO: 32. Sometimes the POX5 polypeptide comprises an amino acid modification at one or more amino acid positions chosen from 81, 82, 83, 84, 85, 86, 88, 93, 94, 95, 96, 98, 102, 284, 287, 290, 291, 292, 294, 295, 436, 453, 454, 455, 456, 457, 458, 459, 460, 461, 462 and 463. A modified endogenous acyl-coA oxidase polypeptide that is not a modified POX5 polypeptide can include an amino acid modification at one or more positions corresponding to one or more of the foregoing positions in the POX5 polypeptide.

In certain instances a modified endogenous acyl-coA dehydrogenase polypeptide is chosen from a modified ACAD, VLCAD, LCAD, MCAD and SCAD polypeptide. In some cases the acyl-coA dehydrogenase polypeptide comprises an amino acid modification corresponding to position 461 of a VLCAD polypeptide.

Provided in certain aspects is a genetically modified yeast that includes a heterologous acyl-coA oxidase polypeptide or a heterologous acyl-coA dehydrogenase polypeptide, which yeast is capable of producing a diacid from a feedstock comprising one or more components from a vegetable oil. The heterologous acyl-coA oxidase polypeptide sometimes is a native polypeptide and sometimes is an active, modified polypeptide. In some embodiments, the heterologous acyl-coA dehydrogenase polypeptide sometimes is a native polypeptide and sometimes is an active, modified polypeptide. In certain instances, a heterologous acyl-coA oxidase polypeptide is chosen from a polypeptide having an amino acid sequence set forth in SEQ ID NO: 51 to SEQ ID NO: 3673. In some cases a heterologous acyl-coA dehydrogenase polypeptide is chosen from SEQ ID NOs: 3679 to 3683, 3686, 3689, 3691, 3693, 3695, 3697, 3699, 3701 and 3703.

A genetically modified yeast sometimes is chosen from a *Candida* spp. yeast, *Yarrowia* spp. yeast, *Pichia* spp. yeast, *Saccharomyces* spp. yeast and *Kluyveromyces* spp. yeast. In certain cases a genetically modified yeast includes one or more genetic modifications that reduce the activity of

one or more native endogenous acyl-coA oxidase polypeptides. In some instances a genetically modified yeast includes a genetic modification that reduces the activity of an enoyl coA isomerase polypeptide (e.g., ECI polypeptide (e.g., ECI1, ECI2)). In certain cases, a genetically modified yeast includes a genetic modification that reduces the activity of an acyl-CoA synthetase (ACS) polypeptide (e.g., a genetic modification reduces the activity of a cytoplasmic and/or peroxisomal ACS polypeptide). An ACS polypeptide sometimes is an ACS1 and/or FAT1 polypeptide. In certain instances, a genetically modified yeast includes a genetic modification that reduces the activity of a polypeptide that transports long-chain fatty acyl-CoA molecules from the cytoplasm into the peroxisome (e.g., the peroxisomal matrix). Such a polypeptide sometimes is a PXA polypeptide. In some cases, a genetic modification that reduces the activity of a certain polypeptide disrupts a polynucleotide that encodes the polypeptide.

In some aspects provided is a method for producing a diacid, which includes contacting a genetically modified yeast described herein with a feedstock comprising one or more components from a vegetable oil capable of being converted by the yeast to a diacid; and culturing the yeast under conditions in which the diacid is produced from the feedstock. The diacid sometimes is a C4 to C24 diacid, and sometimes is chosen from one or more of a C10, C12, C14, C16, C18 or C20 diacid. A diacid sometimes contains no unsaturation (e.g., double bond) and sometimes contains one or more unsaturations. A particular diacid sometimes is the predominant diacid in a mixture of diacids. In some instances the feedstock includes a substantially pure oil. Sometimes the feedstock includes a plurality of fatty acids, and sometimes the feedstock includes a soapstock and/or fatty acid distillate. In certain cases the vegetable oil is from a plant chosen from palm, palm kernel, coconut, soy, safflower, canola, palm, palm kernel or combination thereof.

In certain aspects, provided herein are isolated nucleic acids described herein.

Various embodiments of the present disclosure relate to a genetically modified yeast, comprising an active, modified endogenous acyl-coA oxidase polypeptide, wherein the endogenous acyl-coA oxidase polypeptide is a POX5 polypeptide, which yeast is capable of producing a diacid from a feedstock comprising one or more components from a vegetable oil, wherein the POX5 polypeptide comprises a modified amino acid sequence of SEQ ID NO: 32.

Various embodiments of the present disclosure relate to a method for producing a diacid, comprising: contacting a genetically modified yeast as described herein with a feedstock comprising one or more components from a vegetable oil capable of being converted by the yeast to a diacid; and culturing the yeast under conditions in which the diacid is produced from the feedstock.

Various embodiments of the present disclosure relate to an isolated nucleic acid as claimed, comprising a polynucleotide that encodes a modified acyl-coA oxidase polypeptide from a yeast, wherein the modified acyl-coA oxidase polypeptide is a POX5 polypeptide, and wherein the modified acyl-coA oxidase polypeptide comprises an amino acid modification in the N-terminal loop.

Various embodiments of the present disclosure relate to a cell comprising a nucleic acid as described herein.

Various embodiments of the present disclosure relate to a genetically modified yeast, comprising: an active, modified endogenous POX5 polypeptide, wherein the POX5 polypeptide comprises an amino acid substitution as compared to the sequence shown in SEQ ID NO: 32 at one or more amino acid positions chosen from 81, 82, 83, 84, 85, 86, 88, 93, 94, 95, 96, 98, 102, 284, 287, 290, 291, 292, 294, 295, 436, 453, 454, 455, 456, 457, 458, 459, 460, 461, 462 and 463; which yeast produces a diacid from a feedstock comprising one or more components from a vegetable oil capable of being converted by the yeast to a diacid.

Certain embodiments are described further in the following description, examples, claims and drawings.

Brief Description of the Drawings

The drawings illustrate embodiments of the technology and are not limiting. For clarity and ease of illustration, the drawings are not made to scale and, in some instances, various aspects may be

FIG. 1 is a schematic representation of the conversion of decane to sebacic acid in a beta-oxidation blocked microorganism. Capric acid is formed as an intermediate during omega oxidation.

- 5 FIG. 2 is a schematic representation of the conversion of dodecane to dodecanedioic acid in a beta-oxidation blocked microorganism. Lauric acid is formed as an intermediate during omega oxidation.

- 10 FIG. 3 is a schematic representation of the conversion of a feedstock containing mixed chain-length alkanes to mixed diacids products, including sebacic acid in a beta-oxidation blocked microorganism. Mixed chain-length fatty acids are formed as intermediates during omega oxidation. Sebacic acid can be separated from other diacid products by the use of appropriate separation techniques.

- 15 FIG. 4 is a schematic representation of the conversion of a feedstock containing mixed chain-length alkanes to mixed diacids products, including dodecanedioic acid in a beta-oxidation blocked microorganism. Mixed chain-length fatty acids are formed as intermediates during omega oxidation. Dodecanedioic acid can be separated from other diacid products by the use of appropriate separation techniques.

- 20 FIG. 5 is a schematic representation of the conversion of a long-chain alkane into sebacic acid in a partially beta-oxidation blocked microorganism. The long-chain alkane is first converted into a long-chain fatty acid and then into a long-chain diacid by activities in the omega-oxidation pathway. The long-chain diacid can be converted to sebacic acid by activities in the beta-oxidation pathway, with the simultaneous generation of acetyl-CoA.
- 25

- FIG. 6 is a schematic representation of the conversion of a long-chain alkane into dodecanedioic acid in a partially beta-oxidation blocked microorganism. The long-chain alkane is first converted into a long-chain fatty acid and then into a long-chain diacid by activities in the omega-oxidation pathway. The long-chain diacid can be converted to dodecanedioic acid by activities in the beta-oxidation pathway, with the simultaneous generation of acetyl-CoA.
- 30

FIG. 7 is a schematic representation of the conversion of a feedstock containing mixed chain-length alkanes into sebacic acid in a partially beta-oxidation blocked microorganism. The mixed

chain-length alkanes are first converted into mixed chain-length fatty acids and then mixed diacids by activities in the omega-oxidation pathway. Mixed diacids can be converted to sebacic acid by activities in the beta-oxidation pathway, with the simultaneous generation of acetyl-CoA.

5 FIG. 8 is a schematic representation of the conversion of a feedstock containing mixed chain-length alkanes into dodecanedioic acid in a partially beta-oxidation blocked microorganism. The mixed chain-length alkanes are first converted into mixed chain-length fatty acids and then mixed diacids by activities in the omega-oxidation pathway. Mixed diacids can be converted to dodecanedioic acid by activities in the beta-oxidation pathway, with the simultaneous generation of
10 acetyl-CoA.

FIG. 9 graphically illustrates the conversion of decane to sebacic acid in a fully beta-oxidation blocked *Candida* yeast strain. After incubation for the times shown in the graph, the media was subjected to gas chromatography. The results indicate that greater than 99% of the decane was
15 converted into sebacic acid, with a minimal amount of capric acid also detected by gas chromatography. No significant accumulation of any other monoacid or diacid was detected by gas chromatography. Experimental details and results are given in Example 1.

FIG. 10 graphically illustrates the conversion of capric acid to sebacic acid in a *Candida* yeast
20 strain. GC analysis was performed after a predetermined period of growth. Nearly all the capric acid added was converted to sebacic acid using a starting concentration of capric acid. Experimental details and results are given in Example 2.

FIG. 11 graphically illustrates the distribution of diacids produced during the conversion of long-
25 chain fatty acids to mixed diacids under fermentation conditions using a partially beta-oxidation blocked *Candida tropicalis* strain (e.g., sAA106). Experimental details and results are given in Example 5.

FIG. 12 graphically illustrates the conversion of decane to sebacic acid in a fully beta-oxidation
30 blocked *Candida* yeast strain having additional genetic modifications. Strain sAA003 is the fully beta-oxidation blocked control strain. +CPR indicates the fully beta-oxidation blocked strain also includes an increased number of copies of cytochrome P450 reductase. +CPR+A12 indicates starting strain sAA003 includes the addition genetic modifications of an increased number of copies of cytochrome P450 reductase and also includes an increased number of copies of

cytochrome P450 A12 (e.g., CYP52A12). +CPR+A18 indicates starting strain sAA003 includes the addition genetic modifications of an increased number of copies of cytochrome P450 reductase and also includes an increased number of copies of cytochrome P450 A18 (e.g., CYP52A18). +CPR+A19 indicates starting strain sAA003 includes the addition genetic modifications of an increased number of copies of cytochrome P450 reductase and also includes an increased number of copies of cytochrome P450 A19 (e.g., CYP52A19). +CPR+A20 indicates starting strain sAA003 includes the addition genetic modifications of an increased number of copies of cytochrome P450 reductase and also includes an increased number of copies of cytochrome P450 A20 (e.g., CYP52A20). The y-axis of FIG. 12 is percent of theoretical maximum yield. Experimental details and results are given in Example 7.

FIG. 13 graphically illustrates the results of conversion of methyl laurate to dodecanedioic acid in a fully beta-oxidation blocked Candida yeast strain also contain genetic alterations to a monooxygenase reductase activity, a monooxygenase activity, or a monooxygenase reductase activity and a monooxygenase activity. After 48 hours of incubation the media was subjected to gas chromatography. The results indicate that Candida strains containing an increased number of copies of a CYP52A18 monooxygenase activity and an increased number of copies of a monooxygenase reductase activity (e.g., CPR750) gave the highest yield of dodecanedioic acid (e.g., DDDA), in shake flask fermentation experiments. Experimental details and results are given in Example 8.

FIG. 14 and FIG. 15 schematically illustrate a screening and/or selection method for identifying acyl-CoA oxidase activities with specific substrate specificities. The method can be utilized in conjunction with generating and/or identifying acyl-CoA oxidase activities with altered chain-length substrate specificities. Screening/selection method details are given in Example 9.

FIG. 16 graphically illustrates the results of engineered microorganisms described herein converting decane to sebacic acid under fermentation conditions using different amounts of decane as the feedstock. Experimental details and results are given in Example 3.

FIG. 17 graphically illustrates the results of engineered microorganisms described herein converting a mixed fatty acid feedstock (e.g., mixed chain-length fatty acids) to sebacic acid under fermentation conditions. Experimental details and results are given in Example 4.

FIG. 18 shows a diagram of a plasmid designated pAA073 containing a POX4 promoter and a POX4 terminator.

FIG. 19 illustrates the generation of a full-length deletion cassette for EC11 using PCR overlap
5 extension.

FIG. 20 illustrates the generation of a full-length deletion cassette for the second allele of EC11 using PCR overlap extension.

10 FIG. 21 shows an acyl CoA oxidase activity profile for Pox5 isolated from a *Candida* strain.

FIG. 22 illustrates a PCR overlap extension method for introducing site-directed point mutations into Acyl-CoA Oxidase genes.

15 FIG. 23 – shows a sequence alignment of the N-terminal 180 amino acids of Acol and Acoll. The amino acids highlighted in grey are located within alpha helices and those in bold are located within beta sheets. The center sequence represents the consensus sequence showing conserved residues.

20 FIG. 24 – illustrates a HotSpot Wizard analysis of Pox4 (FIG. 24A) and Pox5 (FIG. 24B) from *Candida* strain ATCC20336. Residues highlighted in dark grey or light grey are mutagenic “hot spots”. Dark grey residues show greater variability at that position than light grey residues. Residues in bold are found within or close to the substrate binding pocket.

25 FIG. 25 - illustrates a HotSpot Wizard analysis of RnAcoll. Residues highlighted in dark grey or light grey are mutagenic “hot spots”. Dark grey residues show greater variability at that position than light grey residues. Residues in bold are found within or close to the substrate binding pocket.

30 FIG. 26 shows a multiple sequence alignment of all three proteins. The underlined portion of RnAcoll (Acoll from *R. norvegicus*) represents the alternatively spliced exon 3.

FIG. 27 shows the acyl CoA activity profile associated with Pox5 mutants. The substrate referred to as “C18” is shortened and pertains to a C18:1 substrate.

FIG. 28 shows the acyl CoA activity profile associated with Pox4 mutants.

FIG. 29 shows a diagram of a plasmid designated pAA298.

5 Detailed Description

Certain fatty dicarboxylic acids (i.e., diacids, e.g., dodecanedioic acid or sebacic acid) are chemical intermediates in manufacturing processes used to make certain polyamides, polyurethanes and plasticizers, all of which have wide applications in producing items such as antiseptics, top-grade
 10 coatings, hot-melt coating and adhesives, painting materials, corrosion inhibitor, surfactant, engineering plastics and can also be used as a starting material in the manufacture of fragrances, for example. For example dodecanedioic acid, also known as 1,12 dodecanedioic acid, and DDDA, is a 12 carbon organic molecule that is a fatty dicarboxylic acid. In another example, sebacic acid, also known as 1,10 decanedioic acid, and 1,8 octanedicarboxylic acid, is a 10 carbon
 15 organic molecule that is a fatty dicarboxylic acid.

Provided herein are methods for producing a fatty dicarboxylic acid (also referred to herein as a diacid). Any suitable diacid can be produced, and a diacid produced often includes acid moieties at each terminus of the molecule (e.g., alpha omega diacids). A diacid sometimes is a C4 to a C24
 20 diacid (i.e., a diacid containing 4 carbons to 24 carbons) and sometimes is a C8, C10, C12, C14, C16, C18, or C20 diacid. Yeast and processes herein are capable of producing a diacid containing an odd number of carbons, and sometimes a product contains one or more diacids chosen from a C5, C7, C9, C11, C13, C15, C17, C19, C21 and C23 diacid. A hydrocarbon portion of a diacid sometimes is fully saturated and sometimes a diacid includes one or more unsaturations (e.g.,
 25 double bonds).

Non-limiting examples of diacids include octadecanedioic acid, decanedioic acid, dodecanedioic acid, tetradecanedioic acid, hexadecanedioic acid, octadecanedioic acid, eicosanedioic acid) and other organic intermediates using biological systems. Non-limiting examples of fatty dicarboxylic
 30 acids include suberic acid (i.e., octanedioic acid, 1,8-octanedioic acid, octanedioic acid, octane-1,8-dioic acid, 1,6-hexanedicarboxylic acid, caprylic diacids), sebacic acid (i.e., 1,10-decanedioic acid, decanedioic acid, decane-1,10-dioic acid, 1,8-octanedicarboxylic acid, capric diacid), dodecanedioic acid (i.e., DDDA, 1,12-dodecanedioic acid, dodecanedioic acid, dodecane-1,12-dioic acid, 1,10-decanedicarboxylic acid, decamethylenedicarboxylic acid, 1,10-dicarboxydecane,

lauric diacid), tetradecanedioic acid (i.e., TDDA, 1,14-tetradecanedioic acid, tetradecanedioic acid, tetradecane-1,14-dioic acid, 1,12-dodecanedicarboxylic acid, myristic diacid), thapsic acid (i.e., hexadecanedioic acid, 1,16-hexadecanedioic acid, hexadecanedioic acid, hexadecane-1,16-dioic acid, 1,14-tetradecanedicarboxylic acid, palmitic diacid), cis-9-hexadecenedioic acid (i.e.,
 5 palmitoleic diacids), octanedioic acid (i.e., 1,18-octadecanedioic acid, octadecanedioic acid, octadecane-1,18-dioic acid, 1,16-hexadecanedicarboxylic acid, stearic diacid), cis-9-octadecenedioic acid (i.e., oleic diacids), cis-9,12-octadecenedioic acid (i.e., linoleic diacids), cis-9,12,15-octadecenedioic acid (i.e., linolenic diacids), arachidic diacid (i.e., eicosanoic diacid, icosanoic diacid), 11-eicosenoic diacid (i.e., cis-11-eicosenedioic acid), 13-eicosenoic diacids (i.e.,
 10 cis-13-eicosenedioic acid), arachidonic diacid (i.e., cis-5,8,11,14-eicosatetraenedioic acid).

A genetically modified yeast can be provided with a feedstock to produce a diacid, and the feedstock sometimes includes a substantially pure aliphatic molecule from which the diacid is produced. In certain embodiments, the feedstock contains a mixed set of aliphatic molecules from
 15 which diacids may be produced. In some embodiments, an aliphatic molecule in the feedstock is the predominant aliphatic species and sometimes a particular diacid produced from that aliphatic molecule is the predominant diacid species produced. A predominant species generally is 51% or more by weight of aliphatic molecule species in a feedstock or 51% or more by weight of diacid species in a product (e.g., about 55% or more, 60% or more, 65% or more, 70% or more, 75% or
 20 more, 80% or more, 85% or more, 90% or more or 95% or more).

Such production systems may have significantly less environmental impact and could be economically competitive with current manufacturing systems. Thus, provided in part herein are methods for manufacturing a fatty dicarboxylic acid (e.g., octanedioic acid, decanedioic acid,
 25 dodecanedioic acid, tetradecanedioic acid, hexadecanedioic acid, octadecanedioic acid, eicosanedioic acid) by engineered microorganisms. In some embodiments microorganisms are engineered to contain at least one heterologous gene encoding an enzyme, where the enzyme is a member of a novel and/or altered pathway engineered into the microorganism. In certain embodiments, an organism may be selected for elevated activity of a native enzyme.

Microorganisms

A microorganism selected often is suitable for genetic manipulation and often can be cultured at cell densities useful for industrial production of a target fatty dicarboxylic acid product. A
 5 microorganism selected often can be maintained in a fermentation device.

The term "engineered microorganism" as used herein refers to a modified microorganism that includes one or more activities distinct from an activity present in a microorganism utilized as a starting point (hereafter a "host microorganism"). An engineered microorganism includes a
 10 heterologous polynucleotide in some embodiments, and in certain embodiments, an engineered organism has been subjected to selective conditions that alter an activity, or introduce an activity, relative to the host microorganism. Thus, an engineered microorganism has been altered directly or indirectly by a human being. A host microorganism sometimes is a native microorganism, and at times is a microorganism that has been engineered to a certain point.

15 In some embodiments an engineered microorganism is a single cell organism, often capable of dividing and proliferating. A microorganism can include one or more of the following features: aerobe, anaerobe, filamentous, non-filamentous, monoploid, diploid, auxotrophic and/or non-auxotrophic. In certain embodiments, an engineered microorganism is a prokaryotic
 20 microorganism (e.g., bacterium), and in certain embodiments, an engineered microorganism is a non-prokaryotic microorganism. In some embodiments, an engineered microorganism is a eukaryotic microorganism (e.g., yeast, fungi, amoeba). In some embodiments, an engineered microorganism is a fungus. In some embodiments, an engineered organism is a yeast.

25 Any suitable yeast may be selected as a host microorganism, engineered microorganism, genetically modified organism or source for a heterologous or modified polynucleotide. Yeast include, but are not limited to, *Yarrowia* yeast (e.g., *Y. lipolytica* (formerly classified as *Candida lipolytica*)), *Candida* yeast (e.g., *C. revkaufi*, *C. viswanathii*, *C. pulcherrima*, *C. tropicalis*, *C. utilis*), *Rhodotorula* yeast (e.g., *R. glutinus*, *R. graminis*), *Rhodospiridium* yeast (e.g., *R. toruloides*),
 30 *Saccharomyces* yeast (e.g., *S. cerevisiae*, *S. bayanus*, *S. pastorianus*, *S. carlsbergensis*), *Cryptococcus* yeast, *Trichosporon* yeast (e.g., *T. pullans*, *T. cutaneum*), *Pichia* yeast (e.g., *P. pastoris*) and *Lipomyces* yeast (e.g., *L. starkeyii*, *L. lipoferus*). In some embodiments, a suitable yeast is of the genus *Arachniotus*, *Aspergillus*, *Aureobasidium*, *Auxarthron*, *Blastomyces*, *Candida*, *Chrysosporium*, *Chrysosporium* *Debaryomyces*, *Coccidiodes*, *Cryptococcus*, *Gymnoascus*,

Hansenula, Histoplasma, Issatchenkia, Kluyveromyces, Lipomyces, Lssatchenkia, Microsporum, Myxotrichum, Myxozyma, Oidiodendron, Pachysolen, Penicillium, Pichia, Rhodosporidium, Rhodotorula, Rhodotorula, Saccharomyces, Schizosaccharomyces, Scopulariopsis, Sepedonium, Trichosporon, or Yarrowia. In some embodiments, a suitable yeast is of the species Arachniotus

5 flavoluteus, Aspergillus flavus, Aspergillus fumigatus, Aspergillus niger, Aureobasidium pullulans, Auxarthron thaxteri, Blastomyces dermatitidis, Candida albicans, Candida dubliniensis, Candida famata, Candida glabrata, Candida guilliermondii, Candida kefir, Candida krusei, Candida lambica, Candida lipolytica, Candida lusitanae, Candida parapsilosis, Candida pulcherrima, Candida revkaufi, Candida rugosa, Candida tropicalis, Candida utilis, Candida viswanathii,

10 Candida xestobii, Chrysosporium keratinophilum, Coccidioides immitis, Cryptococcus albidus var. diffluens, Cryptococcus laurentii, Cryptococcus neoformans, Debaryomyces hansenii, Gymnoascus dugwayensis, Hansenula anomala, Histoplasma capsulatum, Issatchenkia occidentalis, Issatchenkia orientalis, Kluyveromyces lactis, Kluyveromyces marxianus, Kluyveromyces thermotolerans, Kluyveromyces waltii, Lipomyces lipoferus, Lipomyces starkeyii, Microsporum

15 gypseum, Myxotrichum deflexum, Oidiodendron echinulatum, Pachysolen tannophilis, Penicillium notatum, Pichia anomala, Pichia pastoris, Pichia stipitis, Rhodosporidium toruloides, Rhodotorula glutinus, Rhodotorula graminis, Saccharomyces cerevisiae, Saccharomyces kluyveri, Schizosaccharomyces pombe, Scopulariopsis acremonium, Sepedonium chrysospermum, Trichosporon cutaneum, Trichosporon pullans, Yarrowia lipolytica, or Yarrowia lipolytica (formerly

20 classified as Candida lipolytica). In some embodiments, a yeast is a Y. lipolytica strain that includes, but is not limited to, ATCC20362, ATCC8862, ATCC18944, ATCC20228, ATCC76982 and LGAM S(7)1 strains (Papanikolaou S., and Aggelis G., Bioresour. Technol. 82(1):43-9 (2002)). In certain embodiments, a yeast is a Candida species (i.e., Candida spp.) yeast. Any suitable Candida species can be used and/or genetically modified for production of a fatty dicarboxylic acid

25 (e.g., octanedioic acid, decanedioic acid, dodecanedioic acid, tetradecanedioic acid, hexadecanedioic acid, octadecanedioic acid, eicosanedioic acid). In some embodiments, suitable Candida species include, but are not limited to Candida albicans, Candida dubliniensis, Candida famata, Candida glabrata, Candida guilliermondii, Candida kefir, Candida krusei, Candida lambica, Candida lipolytica, Candida lusitanae, Candida parapsilosis, Candida pulcherrima,

30 Candida revkaufi, Candida rugosa, Candida tropicalis, Candida utilis, Candida viswanathii, Candida xestobii and any other Candida spp. yeast described herein. Non-limiting examples of Candida spp. strains include, but are not limited to, sAA001 (ATCC20336), sAA002 (ATCC20913), sAA003 (ATCC20962), sAA496 (US2012/0077252), sAA106 (US2012/0077252), SU-2 (ura3-

/ura3-), H5343 (beta oxidation blocked; US Patent No. 5648247) strains. Any suitable strains from *Candida* spp. yeast may be utilized as parental strains for genetic modification.

Yeast genera, species and strains are often so closely related in genetic content that they can be difficult to distinguish, classify and/or name. In some cases strains of *C. lipolytica* and *Y. lipolytica* can be difficult to distinguish, classify and/or name and can be, in some cases, considered the same organism. In some cases, various strains of *C. tropicalis* and *C. viswanathii* can be difficult to distinguish, classify and/or name (for example see Arie et.al., J. Gen. Appl. Microbiol., 46, 257-262 (2000). Some *C. tropicalis* and *C. viswanathii* strains obtained from ATCC as well as from other commercial or academic sources can be considered equivalent and equally suitable for the embodiments described herein. In some embodiments, some parental strains of *C. tropicalis* and *C. viswanathii* are considered to differ in name only.

Any suitable fungus may be selected as a host microorganism, engineered microorganism or source for a heterologous polynucleotide. Non-limiting examples of fungi include, but are not limited to, *Aspergillus* fungi (e.g., *A. parasiticus*, *A. nidulans*), *Thraustochytrium* fungi, *Schizochytrium* fungi and *Rhizopus* fungi (e.g., *R. arrhizus*, *R. oryzae*, *R. nigricans*). In some embodiments, a fungus is an *A. parasiticus* strain that includes, but is not limited to, strain ATCC24690, and in certain embodiments, a fungus is an *A. nidulans* strain that includes, but is not limited to, strain ATCC38163.

Any suitable prokaryote may be selected as a host microorganism, engineered microorganism or source for a heterologous polynucleotide. A Gram negative or Gram positive bacteria may be selected. Examples of bacteria include, but are not limited to, *Bacillus* bacteria (e.g., *B. subtilis*, *B. megaterium*), *Acinetobacter* bacteria, *Nocardia* bacteria, *Xanthobacter* bacteria, *Escherichia* bacteria (e.g., *E. coli* (e.g., strains DH10B, Stbl2, DH5-alpha, DB3, DB3.1), DB4, DB5, JDP682 and ccdA-over (e.g., U.S. Application No. 09/518,188))), *Streptomyces* bacteria, *Erwinia* bacteria, *Klebsiella* bacteria, *Serratia* bacteria (e.g., *S. marcescens*), *Pseudomonas* bacteria (e.g., *P. aeruginosa*), *Salmonella* bacteria (e.g., *S. typhimurium*, *S. typhi*), *Megasphaera* bacteria (e.g., *Megasphaera elsdenii*). Bacteria also include, but are not limited to, photosynthetic bacteria (e.g., green non-sulfur bacteria (e.g., *Chloroflexus* bacteria (e.g., *C. aurantiacus*), *Chloronema* bacteria (e.g., *C. gigateum*)), green sulfur bacteria (e.g., *Chlorobium* bacteria (e.g., *C. limicola*), *Pelodictyon* bacteria (e.g., *P. luteolum*), purple sulfur bacteria (e.g., *Chromatium* bacteria (e.g., *C. okenii*)), and

purple non-sulfur bacteria (e.g., *Rhodospirillum* bacteria (e.g., *R. rubrum*), *Rhodobacter* bacteria (e.g., *R. sphaeroides*, *R. capsulatus*), and *Rhodomicrobium* bacteria (e.g., *R. vanellii*)).

- Cells from non-microbial organisms can be utilized as a host microorganism, engineered
 5 microorganism or source for a heterologous polynucleotide. Examples of such cells, include, but are not limited to, insect cells (e.g., *Drosophila* (e.g., *D. melanogaster*), *Spodoptera* (e.g., *S. frugiperda* Sf9 or Sf21 cells) and *Trichoplusia* (e.g., High-Five cells); nematode cells (e.g., *C. elegans* cells); avian cells; amphibian cells (e.g., *Xenopus laevis* cells); reptilian cells; mammalian cells (e.g., NIH3T3, 293, CHO, COS, VERO, C127, BHK, Per-C6, Bowes melanoma and HeLa
 10 cells); and plant cells (e.g., *Arabidopsis thaliana*, *Nicotiana tabacum*, *Cuphea acinifolia*, *Cuphea aequipetala*, *Cuphea angustifolia*, *Cuphea appendiculata*, *Cuphea avigera*, *Cuphea avigera* var. *pulcherrima*, *Cuphea axilliflora*, *Cuphea bahiensis*, *Cuphea baillonis*, *Cuphea brachypoda*, *Cuphea bustamanta*, *Cuphea calcarata*, *Cuphea calophylla*, *Cuphea calophylla* subsp. *mesostemon*, *Cuphea carthagenensis*, *Cuphea circaeoides*, *Cuphea confertiflora*, *Cuphea cordata*, *Cuphea*
 15 *crassiflora*, *Cuphea cyanea*, *Cuphea decandra*, *Cuphea denticulata*, *Cuphea disperma*, *Cuphea epilobiifolia*, *Cuphea ericoides*, *Cuphea flava*, *Cuphea flavisetula*, *Cuphea fuchsiifolia*, *Cuphea gaumeri*, *Cuphea glutinosa*, *Cuphea heterophylla*, *Cuphea hookeriana*, *Cuphea hyssopifolia* (Mexican-heather), *Cuphea hyssopoides*, *Cuphea ignea*, *Cuphea ingrata*, *Cuphea jorullensis*, *Cuphea lanceolata*, *Cuphea linarioides*, *Cuphea llavea*, *Cuphea lophostoma*, *Cuphea lutea*,
 20 *Cuphea lutescens*, *Cuphea melanium*, *Cuphea melvilla*, *Cuphea micrantha*, *Cuphea micropetala*, *Cuphea mimuloides*, *Cuphea nitidula*, *Cuphea palustris*, *Cuphea parsonsia*, *Cuphea pascuorum*, *Cuphea paucipetala*, *Cuphea procumbens*, *Cuphea pseudosilene*, *Cuphea pseudovaccinium*, *Cuphea pulchra*, *Cuphea racemosa*, *Cuphea repens*, *Cuphea salicifolia*, *Cuphea salvadorensis*, *Cuphea schumannii*, *Cuphea sessiliflora*, *Cuphea sessilifolia*, *Cuphea setosa*, *Cuphea spectabilis*,
 25 *Cuphea spermacoce*, *Cuphea splendida*, *Cuphea splendida* var. *viridiflava*, *Cuphea strigulosa*, *Cuphea subuligera*, *Cuphea teleandra*, *Cuphea thymoides*, *Cuphea toluhana*, *Cuphea urens*, *Cuphea utriculosa*, *Cuphea viscosissima*, *Cuphea watsoniana*, *Cuphea wrightii*, *Cuphea lanceolata*)
- 30 Microorganisms or cells used as host organisms or source for a heterologous polynucleotide are commercially available. Microorganisms and cells described herein, and other suitable microorganisms and cells are available, for example, from Invitrogen Corporation, (Carlsbad, CA), American Type Culture Collection (Manassas, Virginia), and Agricultural Research Culture Collection (NRRL; Peoria, Illinois).

Host microorganisms and engineered microorganisms may be provided in any suitable form. For example, such microorganisms may be provided in liquid culture or solid culture (e.g., agar-based medium), which may be a primary culture or may have been passaged (e.g., diluted and cultured) one or more times. Microorganisms also may be provided in frozen form or dry form (e.g., lyophilized). Microorganisms may be provided at any suitable concentration.

Carbon Processing Pathways and Activities

Figures 1-8 schematically illustrate non-limiting embodiments of engineered pathways that can be used to produce a fatty dicarboxylic acid (e.g., octanedioic acid, decanedioic acid, dodecanedioic acid, tetradecanedioic acid, hexadecanedioic acid, octadecanedioic acid, eicosanedioic acid) from various starting carbon sources or feedstocks. FIG. 1 depicts an embodiment of a non-limiting engineered biological pathway for producing sebacic acid in microorganisms having a fully blocked beta-oxidation pathway, using decane as the carbon source starting material. FIG. 2 depicts an embodiment of a non-limiting engineered biological pathway for producing dodecanedioic acid in microorganisms having a fully blocked beta-oxidation pathway, using dodecane as the carbon source starting material. Figures 3 and 4 depict an embodiment of a non-limiting engineered biological pathway for producing mixed chain-length diacids in a microorganism having a fully blocked beta-oxidation pathway, using mixed chain-length alkanes as the carbon source starting material. Sebacic acid (FIG. 3) and dodecanedioic acid (FIG. 4) can be separated and/or purified away from other diacid products using a suitable combination of centrifugation, organic solvent extraction, chromatography, and/or other purification/separation techniques. Figures 5 and 6 depict an embodiment of a non-limiting engineered biological pathway for producing sebacic acid (FIG. 5) and dodecanedioic acid (FIG. 6) in microorganisms having a partially blocked beta oxidation pathway, using long-chain alkanes as the carbon source starting material. Figures 7 and 8 depict an embodiment of a non-limiting engineered biological pathway for producing sebacic acid (FIG. 7) and dodecanedioic acid (FIG. 8) in microorganisms having a partially blocked beta oxidation pathway, using mixed-chain length alkanes as the carbon source starting material. The alkane carbon source starting materials are initially metabolized using naturally occurring and/or engineered activities in naturally occurring and/or engineered pathways to yield an intermediate alcohol which can then be converted to a carboxylic acid (e.g., fatty acid) by the action of other naturally occurring and/or engineered activities in the omega-oxidation pathway depicted in Figures 1-8.

Alkanes are omega-hydroxylated by the activity of cytochrome P450 enzymes, thereby generating the equivalent chain-length alcohol derivative of the starting alkane carbon source material. In certain embodiments, a cytochrome P450 activity can be increased by increasing the number of copies of a cytochrome P450 gene (e.g., 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more copies of the gene), by increasing the activity of a promoter that regulates transcription of a cytochrome P450 gene, or by increasing the number of copies of a cytochrome P450 gene and increasing the activity of a promoter that regulates transcription of a cytochrome P450 gene, thereby increasing the production of target product (e.g., sebacic or dodecanedioic acid) via increased activity of one or more cytochrome P450 enzymes. In some embodiments, a cytochrome P450 enzyme is endogenous to the host microorganism. One or more cytochrome P450 activities can be added and/or increased dependent on the carbon source starting material, in certain embodiments. Cytochrome P450's sometimes exhibit increased activities in response to stimulation by certain feedstocks or carbon source starting materials. In some embodiments, an engineered microorganism includes an increased number of copies of one or more cytochrome P450s that are stimulated by a chosen carbon source starting material or feedstock. Cytochrome P450 responsiveness to a chosen starting carbon source or feedstock can be determined using any suitable assay. Non-limiting examples of assays suitable for identification of cytochrome P450 responsiveness to a starting carbon source or feedstock include RT-PCR or qRT-PCR after the host microorganism has been exposed to the chosen carbon source or feedstock for varying amounts of time.

Cytochrome P450 is reduced by the activity of cytochrome P450 reductase (CPR), thereby recycling cytochrome P450 to allow further enzymatic activity. In certain embodiments, the CPR enzyme is endogenous to the host microorganism. In some embodiments, host CPR activity can be increased by increasing the number of copies of a CPR gene (e.g., 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more copies of the gene), by increasing the activity of a promoter that regulates transcription of a CPR gene, or by increasing the number of copies of a CPR gene and increasing the activity of a promoter that regulates transcription of a CPR gene, thereby increasing the production of target product (e.g., sebacic or dodecanedioic acid) via increased recycling of cytochrome P450. In certain embodiments, the promoter can be a heterologous promoter (e.g., endogenous or exogenous promoter). In some embodiments, the CPR gene is heterologous and exogenous and can be isolated from any suitable organism. Non-limiting examples of organisms from which a CPR gene can be isolated include *C. tropicalis*, *S. cerevisiae* and *Bacillus megaterium*.

Oxidation of the alcohol to an aldehyde may be performed by an enzyme in the fatty alcohol oxidase family (e.g., long-chain fatty alcohol oxidase EC 1.1.3.20), or an enzyme in the alcohol dehydrogenase family (e.g., fatty alcohol dehydrogenase; EC 1.1.1.1). The aldehyde may be oxidized to a carboxylic acid (e.g., sebacic or dodecanedioic acid) by the activity of the enzyme
5 aldehyde dehydrogenase (e.g., long-chain-aldehyde dehydrogenase or fatty aldehyde dehydrogenase; EC 1.2.1.48). In some embodiments, the long chain fatty alcohol oxidase, fatty alcohol dehydrogenase and/or the long-chain-aldehyde dehydrogenase exist in a host organism. Flux through these two steps may sometimes be augmented by increasing the copy number of the enzymes, or by increasing the activity of the promoter transcribing the genes. In some
10 embodiments alcohol and aldehyde dehydrogenases specific for 10, 12 or 14 carbon substrates may be isolated from another organism, and inserted into the host organism.

FIG. 1 depicts a non-limiting embodiment of an engineered biological pathway for making sebacic acid using decane (e.g., a C10 alkane) as the carbon source starting material. Due to the carbon
15 chain length of decane, no chain shortening is necessary to arrive at the 10 carbon diacid, sebacic acid. Thus a fully beta oxidation blocked microorganism can be utilized to minimize conversion of the desired 10 carbon diacid into diacids having shorter chain lengths.

FIG. 2 depicts a non-limiting embodiment of an engineered biological pathway for making
20 dodecanedioic acid using dodecane (e.g., a C12 alkane) as the carbon source starting material. Due to the carbon chain length of dodecane, no chain shortening is necessary to arrive at the 12 carbon diacid, dodecanedioic acid. Thus a fully beta oxidation blocked microorganism can be utilized to minimize conversion of the desired 12 carbon diacid into diacids having shorter chain lengths.

25
Figures 3 and 4 depict a non-limiting embodiment of an engineered biological pathway for generating a mixed population of diacid (fatty dicarboxylic acid) products, including sebacic acid (FIG. 3) and dodecanedioic acid (FIG. 4), using a carbon source or feedstock that contains mixed-chain-length alkanes as the carbon source starting material. Any suitable mixed-chain-length
30 alkane, fatty alcohol, mixed chain length fatty alcohol feedstock, fatty acid, mixed fatty acid feedstock, paraffin, fat or oil can be used. In some embodiments, the distribution of carbon chain lengths in the starting material is substantially similar to the desired carbon chain length distribution in the mixed diacid product. In certain embodiments, the feedstock is enriched for a desired chain length. In some embodiments, the enriched fraction is enriched for carbon chain lengths of about

10 carbons. In some embodiments, the enriched fraction is enriched for carbon chain lengths of about 12 carbons. Because, in some embodiments, the diacids generated have substantially the same chain lengths as the chain lengths found in the carbon source starting material, a fully beta-oxidation blocked microorganism can be utilized to minimize conversion of the diacids of desired
 5 chain length into diacids of shorter chain lengths. The lower part of the pathways in FIG. 3 AND FIG. 4 show the separation of sebacic acid and dodecanedioic acid, respectively, away from the mixed diacid products by the use of separation techniques described herein, or those known in the art.

10 In certain embodiments involving genetically modified organisms having partially blocked beta-oxidation pathways (see Figures 5 - 8), feedstocks suitable for use include, but are not limited to, fatty acid distillates or soapstocks of renewable oils (palm oil fatty acid distillate, soybean oil soapstock, coconut oil soapstock), renewable oils (coconut oil, palm oil, palm kernel oil, soybean oil, corn oil, etc.), fatty acids of chain length equal to or greater than C10 (in substantially single
 15 form (e.g., in substantially pure form) or in mixture form, alkanes of chain length equal to or greater than C10 in substantially single form (e.g., substantially pure form) or in mixture form.

Carbon sources with longer chain lengths (e.g., 12 carbons or greater in length) can be metabolized using naturally occurring and/or engineered pathways to yield molecules that can be
 20 further metabolized using the beta oxidation pathway shown in the lower portion of Figures 5 through 8. In some embodiments, beta-oxidation activities in the pathways shown in Figures 5 through 8 also can be engineered (e.g., as described herein) to enhance metabolism and target product formation. In some embodiments, one acyl-CoA oxidase activity of the beta-oxidation pathway is engineered to be enhanced, and in certain embodiments, the other acyl-CoA oxidase
 25 activity in the beta-oxidation pathway is altered to reduce or eliminate the activity, thereby optimizing the production of a diacid of a desired chain-length or diacids with a distribution of desired chain lengths. In some embodiments, an acyl-CoA oxidase is selected and/or engineered to alter the substrate specificity of the enzyme. In certain embodiments, the substrate specificity of a heterologous and/or engineered acyl-CoA oxidase is for carbon chain lengths of between about
 30 12 carbons and about 18 carbons, and in some embodiments a heterologous and/or engineered acyl-CoA oxidase exhibits no activity on substrates below 12 carbons in length. In certain embodiments, a heterologous acyl-CoA oxidase with a desired chain length specificity can be isolated from any suitable organism. Non-limiting examples of organisms that include, or can be used as donors for acyl-CoA oxidase enzymes include yeast (e.g., Candida, Saccharomyces,

Debaryomyces, Meyerozyma, Lodderomyces, Scheffersomyces, Clavispora, Yarrowia, Pichia, Kluyveromyces, Eremothecium, Zygosaccharomyces, Lachancea, Nakaseomyces), animals (e.g., Homo, Rattus), bacteria (e.g., Escherichia, Pseudomonas, Bacillus), or plants (e.g., Arabidopsis, Nicotiana, Cuphea).

5

In certain embodiments, a carbon source starting material (e.g., alkane, fatty acid, fatty alcohol, dicarboxylic acid) of intermediate or long chain length (e.g., between about 10 carbons and 22 carbons) is converted into an acyl-CoA derivative for entry into the beta-oxidation pathway. The acyl-CoA derivative can be generated by the activity of an acyl-CoA ligase enzyme, in some
 10 embodiments. The acyl-CoA derivative is subsequently oxidized by the activity of an acyl-CoA oxidase enzyme (e.g., also known as acyl-CoA oxidoreductase and fatty acyl-coenzyme A oxidase) of natural or altered substrate specificity, in certain embodiments. The trans-2,3-dehydroacyl-CoA derivative long chain fatty alcohol, fatty acid or dicarboxylic acid may be further converted to 3-hydroxyacyl-CoA by the activity of enoyl-CoA hydratase. 3-hydroxyacyl-CoA can be
 15 converted to 3-oxoacyl-CoA by the activity of 3-hydroxyacyl-CoA dehydrogenase. 3-oxoacyl-CoA may be converted to an acyl-CoA molecule, shortened by 2 carbons and an acetyl-CoA, by the activity of Acetyl-CoA C-acyltransferase (e.g., also known as beta-ketothiolase and beta-ketothiolase). In some embodiments, acyl-CoA molecules may be repeatedly shortened by beta oxidation until a desired carbon chain length is generated (e.g., 10 or 12 carbons, sebacic acid or
 20 dodecanedioic acid, respectively). A shortened fatty acid can be further processed using omega oxidation to yield a dicarboxylic acid (e.g., dodecanedioic acid).

Beta-oxidation activities

25 The term "beta oxidation pathway" as used herein, refers to a series of enzymatic activities utilized to metabolize fatty alcohols, fatty acids, or dicarboxylic acids. The activities utilized to metabolize fatty alcohols, fatty acids, or dicarboxylic acids include, but are not limited to, acyl-CoA ligase activity, acyl-CoA oxidase activity, acyl-CoA hydrolase activity, acyl-CoA thioesterase activity, enoyl-CoA hydratase activity, 3-hydroxyacyl-CoA dehydrogenase activity and acetyl-CoA C-
 30 acyltransferase activity. The term "beta oxidation activity" refers to any of the activities in the beta oxidation pathway utilized to metabolize fatty alcohols, fatty acids or dicarboxylic acids.

Beta-oxidation – Acyl-CoA Ligase

An acyl-CoA ligase enzyme sometimes is encoded by the host organism and can be added to generate an engineered organism. In some embodiments, host acyl-CoA ligase activity can be increased by increasing the number of copies of an acyl-CoA ligase gene, by increasing the activity of a promoter that regulates transcription of an acyl-CoA ligase gene, or by increasing the number of copies of the gene and by increasing the activity of a promoter that regulates transcription of the gene, thereby increasing production of target product (e.g., sebacic or dodecanedioic acid) due to increased carbon flux through the pathway. In certain embodiments, the acyl-CoA ligase gene can be isolated from any suitable organism. Non-limiting examples of organisms that include, or can be used as donors for, acyl-CoA ligase enzymes include *Candida*, *Saccharomyces*, or *Yarrowia*.

Beta-oxidation – Enoyl-CoA Hydratase

An enoyl-CoA hydratase enzyme catalyzes the addition of a hydroxyl group and a proton to the unsaturated β -carbon on a fatty-acyl CoA and sometimes is encoded by the host organism and sometimes can be added to generate an engineered organism. In certain embodiments, the enoyl-CoA hydratase activity is unchanged in a host or engineered organism. In some embodiments, the host enoyl-CoA hydratase activity can be increased by increasing the number of copies of an enoyl-CoA hydratase gene, by increasing the activity of a promoter that regulates transcription of an enoyl-CoA hydratase gene, or by increasing the number copies of the gene and by increasing the activity of a promoter that regulates transcription of the gene, thereby increasing the production of target product (e.g., sebacic or dodecanedioic acid) due to increased carbon flux through the pathway. In certain embodiments, the enoyl-CoA hydratase gene can be isolated from any suitable organism. Non-limiting examples of organisms that include, or can be used as donors for, enoyl-CoA hydratase enzymes include *Candida*, *Saccharomyces*, or *Yarrowia*.

Beta-oxidation – 3-Hydroxyacyl-CoA Dehydrogenase

3-hydroxyacyl-CoA dehydrogenase enzyme catalyzes the formation of a 3-ketoacyl-CoA by removal of a hydrogen from the newly formed hydroxyl group created by the activity of enoyl-CoA hydratase. In some embodiments, the activity is encoded by the host organism and sometimes can be added or increased to generate an engineered organism. In certain embodiments, the 3-hydroxyacyl-CoA activity is unchanged in a host or engineered organism. In some embodiments,

the host 3-hydroxyacyl-CoA dehydrogenase activity can be increased by increasing the number of copies of a 3-hydroxyacyl-CoA dehydrogenase gene, by increasing the activity of a promoter that regulates transcription of a 3-hydroxyacyl-CoA dehydrogenase gene, or by increasing the number of copies of the gene and by increasing the activity of a promoter that regulates transcription of the gene, thereby increasing production of target product (e.g., sebacic or dodecanedioic acid) due to increased carbon flux through the pathway. In certain embodiments, the 3-hydroxyacyl-CoA dehydrogenase gene can be isolated from any suitable organism. Non-limiting examples of organisms that include, or can be used as donors for, 3-hydroxyacyl-CoA dehydrogenase enzymes include *Candida*, *Saccharomyces*, or *Yarrowia*.

10

Beta-oxidation – Acetyl-CoA C-Acyltransferase

An Acetyl-CoA C-acyltransferase (e.g., beta-ketothiolase) enzyme catalyzes the formation of a fatty acyl-CoA shortened by 2 carbons by cleavage of the 3-ketoacyl-CoA by the thiol group of another molecule of CoA. The thiol is inserted between C-2 and C-3, which yields an acetyl CoA molecule and an acyl CoA molecule that is two carbons shorter. An Acetyl-CoA C-acyltransferase sometimes is encoded by the host organism and sometimes can be added to generate an engineered organism. In certain embodiments, the acetyl-CoA C-acyltransferase activity is unchanged in a host or engineered organism. In some embodiments, the host acetyl-CoA C-acyltransferase activity can be increased by increasing the number of copies of an acetyl-CoA C-acyltransferase gene, or by increasing the activity of a promoter that regulates transcription of an acetyl-CoA C-acyltransferase gene, thereby increasing the production of target product (e.g., sebacic or dodecanedioic acid) due to increased carbon flux through the pathway. In certain embodiments, the acetyl-CoA C-acyltransferase gene can be isolated from any suitable organism. Non-limiting examples of organisms that include, or can be used as donors for, acetyl-CoA C-acyltransferase enzymes include *Candida*, *Saccharomyces*, or *Yarrowia*.

Beta-oxidation – Enoyl CoA Isomerase

Feedstocks, such as fatty acid distillates and soapstocks can comprise unsaturated fatty acids, for example, such as oleic acid (C18:1), linoleic acid (C18:2), and linolenic acid (C18:3). In some embodiments, unsaturated fatty acids are converted to dicarboxylic acids that maintain the position and orientation of the double bonds. Cells can employ additional enzymes to allow the oxidation (e.g., beta oxidation) of these types of unsaturated fatty acids or diacids. In some instances, an

enzyme enoyl-CoA isomerase (ECI) is required for the beta-oxidation of substrates with double bonds at odd numbered positions. In some instances, the enzyme dienoyl-CoA reductase (DCR) is required for the beta-oxidation of substrates with double bonds at even numbered positions).

- 5 Enoyl CoA Isomerase (ECI) can also be known as enoyl-CoA delta isomerase 1, dodecenoyl-CoA isomerase, 3,2 trans-enoyl-CoA isomerase, acetylene-allene isomerase, delta3, delta2-enoyl-CoA isomerase, dodecenoyl-CoA delta isomerase, and EC 5.3.3.8 (in human for example). Several alternatively spliced transcript variants are also known. ECI is a member of the hydratase/isomerase superfamily. ECI can be a key mitochondrial enzyme involved in beta-
- 10 oxidation of unsaturated fatty acids. This enzyme can isomerize both 3-cis and 3-trans double bonds into the 2-trans form in a range of ECI enzymes from different species. ECI can catalyze the transformation of 3-cis and 3-trans-enoyl-CoA esters arising during the stepwise degradation of cis-, mono-, and polyunsaturated fatty acids to the 2-trans-enoyl-CoA intermediates.
- 15 In some embodiments, ECI is a critical enzyme because of its activity and the normal position of double bonds in some feedstocks (e.g., soapstocks and fatty acid distillates). Many unsaturated fatty acids have a cis- $\Delta 9$ double bond. During the beta-oxidation of an 18-carbon diacid with a cis- $\Delta 9$ double bond, the double bond is encountered when it has been chain shortened to 12 carbons. At this stage the 12-carbon molecule can have a cis- $\Delta 3$ double bond that is not a substrate for an acyl-CoA oxidase. Therefore, in some embodiments, ECI is required to convert the cis- $\Delta 3$ double bond to a trans- $\Delta 2$ double bond. In some instances, the product of the ECI reaction is a substrate for the second step in beta-oxidation, so ECI can effectively bypasses acyl-CoA oxidase in this particular round of beta-oxidation. In some instances, this is important because even if a strain comprises an acyl-CoA oxidase that is not active on feedstocks of $\leq C12$ (i.e., 12 carbons), an
- 20 active ECI can effect the shortening of one more round of beta-oxidation which can produce a 10-carbon product for substrates with a cis- $\Delta 9$ double bond. Therefore, in some embodiments the ECI gene is disrupted (e.g., knocked out or deleted) in a yeast (e.g., in a *Candida* strain) to prevent chain shortening past a desired chain-length (e.g., in this instance, 12 carbons). In some embodiments, disrupting the expression (e.g. knocking out the expression) of an ECI gene can
- 25 result in an increase in the production of a fatty dicarboxylic acid comprising 10 to 18 carbons. In some embodiments, disrupting the expression (e.g. knocking out the expression) of an ECI gene can result in an increase in the production of a fatty dicarboxylic acid comprising 10, 12, 14, 16 or 18 carbons. In some embodiments, disrupting the expression of an enoyl CoA isomerase can
- 30

increase the production of fatty dicarboxylic acid comprising 10, 12, 14, 16 or 18 carbons when using certain feedstocks (e.g., certain soapstocks or fatty acid distillates).

5 In some embodiments, an ECI knock out (i.e. *eciΔ*) strain is able to produce DDDA from oleic acid even in the presence of acyl-CoA oxidase with activity on substrates of chain-length less than 12 carbons.

10 In some embodiments, a 12 carbon dicarboxylic acids produced from fatty acid feedstocks comprising unsaturated fatty acids require hydrogenation to arrive at the fully saturated DDDA product.

Omega oxidation activities

15 The term “omega oxidation activity” refers to any of the activities in the omega oxidation pathway utilized to metabolize alkanes, fatty alcohols, fatty acids, dicarboxylic acids, or sugars. The activities utilized to metabolize fatty alcohols, fatty acids, or dicarboxylic acids include, but are not limited to, monooxygenase activity (e.g., cytochrome P450 activity), monooxygenase reductase activity (e.g., cytochrome P450 reductase activity), alcohol dehydrogenase activity (e.g., fatty alcohol dehydrogenase activity, or long-chain alcohol dehydrogenase activity), fatty alcohol
20 oxidase activity, fatty aldehyde dehydrogenase activity, and thioesterase activity.

Omega Oxidation - Monooxygenases

25 A cytochrome P450 enzyme (e.g., monooxygenase activity) often catalyzes the insertion of one atom of oxygen into an organic substrate (RH) while the other oxygen atom is reduced to water. Insertion of the oxygen atom near the omega carbon of a substrate yields an alcohol derivative of the original starting substrate (e.g., yields a fatty alcohol). A cytochrome P450 sometimes is encoded by the host organism and sometimes can be added to generate an engineered organism.

30 In certain embodiments, the monooxygenase activity is unchanged in a host or engineered organism. In some embodiments, the host monooxygenase activity can be increased by increasing the number of copies of a cytochrome P450 gene, or by increasing the activity of a promoter that regulates transcription of a cytochrome P450 gene, thereby increasing the production of target product (e.g., sebacic or dodecanedioic acid) due to increased carbon flux

through the pathway. In certain embodiments, the cytochrome P450 gene can be isolated from any suitable organism. Non-limiting examples of organisms that include, or can be used as donors for, cytochrome P450 enzymes include yeast (e.g., *Candida*, *Saccharomyces*, *Debaryomyces*, *Meyerozyma*, *Lodderomyces*, *Scheffersomyces*, *Clavispora*, *Yarrowia*, *Pichia*, *Kluyveromyces*,
 5 *Eremothecium*, *Zygosaccharomyces*, *Lachancea*, *Nakaseomyces*), animals (e.g., *Homo*, *Rattus*), bacteria (e.g., *Escherichia*, *Pseudomonas*, *Bacillus*), or plants (e.g., *Arabidopsis*, *Nicototania*, *Cuphea*).

Omega Oxidation - Monooxygenase Reductases

10

A cytochrome P450 reductase (e.g., monooxygenase reductase activity) catalyzes the reduction of the heme-thiolate moiety in cytochrome P450 by transferring an electron to the cytochrome P450. A cytochrome P450 reductase sometimes is encoded by the host organism and sometimes can be added to generate an engineered organism. In certain embodiments, the monooxygenase
 15 reductase activity is unchanged in a host or engineered organism. In some embodiments, the host monooxygenase reductase activity can be increased by increasing the number of copies of a cytochrome P450 reductase gene, or by increasing the activity of a promoter that regulates transcription of a cytochrome P450 reductase gene, thereby increasing the production of target product (e.g., sebacic or dodecanedioic acid) due to increased carbon flux through the pathway. In
 20 certain embodiments, the cytochrome P450 reductase gene can be isolated from any suitable organism. Non-limiting examples of organisms that include, or can be used as donors for, cytochrome P450 reductase enzymes include yeast (e.g., *Candida*, *Saccharomyces*, *Debaryomyces*, *Meyerozyma*, *Lodderomyces*, *Scheffersomyces*, *Clavispora*, *Yarrowia*, *Pichia*, *Kluyveromyces*, *Eremothecium*, *Zygosaccharomyces*, *Lachancea*, *Nakaseomyces*), animals (e.g.,
 25 *Homo*, *Rattus*), bacteria (e.g., *Escherichia*, *Pseudomonas*, *Bacillus*), or plants (e.g., *Arabidopsis*, *Nicototania*, *Cuphea*).

Omega Oxidation - Alcohol Dehydrogenases

30 An alcohol dehydrogenase (e.g., fatty alcohol dehydrogenase, long-chain alcohol dehydrogenase) catalyzes the removal of a hydrogen from an alcohol to yield an aldehyde or ketone and a hydrogen atom and NADH, in the endoplasmic reticulum of a cell. An alcohol dehydrogenase sometimes is encoded by the host organism and sometimes can be added to generate an engineered organism. In certain embodiments, the alcohol dehydrogenase activity is unchanged in

a host or engineered organism. In some embodiments, the host alcohol dehydrogenase activity can be increased by increasing the number of copies of an alcohol dehydrogenase gene, or by increasing the activity of a promoter that regulates transcription of an alcohol dehydrogenase gene, thereby increasing the production of target product (e.g., sebacic or dodecanedioic acid) due to increased carbon flux through the pathway. In certain embodiments, the alcohol dehydrogenase gene can be isolated from any suitable organism. Non-limiting examples of organisms that include, or can be used as donors for, alcohol dehydrogenase enzymes include yeast (e.g., Candida, Saccharomyces, Debaryomyces, Meyerozyma, Lodderomyces, Scheffersomyces, Clavispora, Yarrowia, Pichia, Kluyveromyces, Eremothecium, Zygosaccharomyces, Lachancea, Nakaseomyces), animals (e.g., Homo, Rattus), bacteria (e.g., Escherichia, Pseudomonas, Bacillus), or plants (e.g., Arabidopsis, Nictotania, Cuphea).

Omega Oxidation - Fatty Alcohol Oxidases

A fatty alcohol oxidase (e.g., long-chain alcohol oxidase, EC 1.1.3.20) enzyme catalyzes the addition of oxygen to two molecules of a long-chain alcohol to yield 2 long chain aldehydes and 2 molecules of water, in the peroxisome of a cell. A fatty alcohol oxidase sometimes is encoded by the host organism and sometimes can be added to generate an engineered organism. In certain embodiments, the fatty alcohol oxidase activity is unchanged in a host or engineered organism. In some embodiments, the host fatty alcohol oxidase activity can be increased by increasing the number of copies of a fatty alcohol oxidase gene, or by increasing the activity of a promoter that regulates transcription of a fatty alcohol oxidase gene, thereby increasing the production of target product (e.g., sebacic or dodecanedioic acid) due to increased carbon flux through the pathway. In certain embodiments, the fatty alcohol oxidase gene can be isolated from any suitable organism. Non-limiting examples of organisms that include, or can be used as donors for, fatty alcohol oxidase enzymes include yeast (e.g., Candida, Saccharomyces, Debaryomyces, Meyerozyma, Lodderomyces, Scheffersomyces, Clavispora, Yarrowia, Pichia, Kluyveromyces, Eremothecium, Zygosaccharomyces, Lachancea, Nakaseomyces), animals (e.g., Homo, Rattus), bacteria (e.g., Escherichia, Pseudomonas, Bacillus), or plants (e.g., Arabidopsis, Nictotania, Cuphea).

Omega Oxidation - Aldehyde Dehydrogenases

A fatty aldehyde dehydrogenase (e.g., long chain aldehyde dehydrogenase) enzyme catalyzes the oxidation of long chain aldehydes to a long chain carboxylic acid, NADH and H⁺. A fatty aldehyde

dehydrogenase sometimes is encoded by the host organism and sometimes can be added to generate an engineered organism. In certain embodiments, the fatty aldehyde dehydrogenase activity is unchanged in a host or engineered organism. In some embodiments, the host fatty aldehyde dehydrogenase activity can be increased by increasing the number of copies of a fatty aldehyde dehydrogenase gene, or by increasing the activity of a promoter that regulates transcription of a fatty aldehyde dehydrogenase gene, thereby increasing the production of target product (e.g., sebacic or dodecanedioic acid) due to increased carbon flux through the pathway. In certain embodiments, the fatty aldehyde dehydrogenase gene can be isolated from any suitable organism. Non-limiting examples of organisms that include, or can be used as donors for, fatty aldehyde dehydrogenase enzymes include yeast (e.g., *Candida*, *Saccharomyces*, *Debaryomyces*, *Meyerozyma*, *Lodderomyces*, *Scheffersomyces*, *Clavispora*, *Yarrowia*, *Pichia*, *Kluyveromyces*, *Eremothecium*, *Zygosaccharomyces*, *Lachancea*, *Nakaseomyces*), animals (e.g., *Homo*, *Rattus*), bacteria (e.g., *Escherichia*, *Pseudomonas*, *Bacillus*), or plants (e.g., *Arabidopsis*, *Nicototania*, *Cuphea*).

15

Omega Oxidation - Thioesterases

A thioesterase enzyme (e.g., acyl-CoA thioesterase activity, acyl-ACP thioesterase activity) catalyzes the removal of Coenzyme A or acyl carrier protein (e.g., ACP) from a fatty acid including acyl-CoA or acyl carrier protein (e.g., esterified fatty acid) to yield a fatty acid and an alcohol. The reaction occurs in the presence of water and Coenzyme A or acyl carrier protein is specifically removed at a thiol group. A thioesterase sometimes is encoded by the host organism and sometimes can be added to generate an engineered organism. In certain embodiments, the thioesterase activity is unchanged in a host or engineered organism. In some embodiments, the host thioesterase activity can be increased by increasing the number of copies of a thioesterase gene, or by increasing the activity of a promoter that regulates transcription of a thioesterase gene, thereby increasing the production of target product (e.g., sebacic or dodecanedioic acid) due to increased carbon flux through the pathway. In certain embodiments, a thioesterase gene can be isolated from any suitable organism. Non-limiting examples of organisms that include, or can be used as donors for, thioesterase enzymes include yeast (e.g., *Candida*, *Saccharomyces*, *Debaryomyces*, *Meyerozyma*, *Lodderomyces*, *Scheffersomyces*, *Clavispora*, *Yarrowia*, *Pichia*, *Kluyveromyces*, *Eremothecium*, *Zygosaccharomyces*, *Lachancea*, *Nakaseomyces*), animals (e.g., *Homo*, *Rattus*), bacteria (e.g., *Escherichia*, *Pseudomonas*, *Bacillus*), or plants (e.g., *Arabidopsis*, *Nicototania*, *Cuphea*).

30

Engineered Pathways

Figures 1-8 depict embodiments of biological pathways for making sebacic acid and dodecanedioic acid, using various alkanes, fatty acids, fatty alcohols or combinations thereof. Any suitable
5 alkane, fatty acid, fatty alcohol, plant based oil, seed based oil, non-petroleum derived soap stock or the like can be used as the feedstock for the organism (e.g., dodecane, methyl laurate, lauric acid, carbon sources having 10 or greater carbons (e.g. for sebacic acid production) or carbon sources having 12 or greater carbons (e.g. for dodecanedioic acid production). In some
10 embodiments, carbon sources with greater than 12 carbons can be metabolized using naturally occurring and/or engineered pathways to yield molecules that can be further metabolized using the beta oxidation pathway shown in the lower portion of Figures 5-8. In some embodiments, the activities in the pathways depicted in Figures 1-8 can be engineered, as described herein, to enhance metabolism and target product formation.

15 In certain embodiments, one or more activities in one or more metabolic pathways can be engineered to increase carbon flux through the engineered pathways to produce a desired product (e.g., sebacic or dodecanedioic acid). The engineered activities can be chosen to allow increased production of metabolic intermediates that can be utilized in one or more other engineered
20 pathways to achieve increased production of a desired product with respect to the unmodified host organism. The engineered activities also can be chosen to allow decreased activity of enzymes that reduce production of a desired intermediate or end product (e.g., reverse activities). This “carbon flux management” can be optimized for any chosen feedstock, by engineering the appropriate activities in the appropriate pathways. Non-limiting examples are given herein using
25 pure alkanes (e.g., single chain length alkanes, dodecane or example), mixed chain-length alkanes, long-chain alkanes, pure fatty acids (e.g., single chain length fatty acids, capric acid for example) and mixed chain length fatty acids (see Figures 1-8). The process of “carbon flux management” through engineered pathways produces a dicarboxylic acid (e.g. sebacic acid or dodecanedioic acid) at a level and rate closer to the calculated maximum theoretical yield for any
30 given feedstock, in certain embodiments. The terms “theoretical yield” or “maximum theoretical yield” as used herein refer to the yield of product of a chemical or biological reaction that can be formed if the reaction went to completion. Theoretical yield is based on the stoichiometry of the reaction and ideal conditions in which starting material is completely consumed, undesired side reactions do not occur, the reverse reaction does not occur, and there no losses in the work-up procedure.

A microorganism may be modified and engineered to include or regulate one or more activities in a fatty dicarboxylic acid (e.g., octanedioic acid, decanedioic acid, dodecanedioic acid, tetradecanedioic acid, hexadecanedioic acid, octadecanedioic acid, eicosanedioic acid) pathway. The term "activity" as used herein refers to the functioning of a microorganism's natural or engineered biological pathways to yield various products including a fatty dicarboxylic acid (e.g., octanedioic acid, decanedioic acid, dodecanedioic acid, tetradecanedioic acid, hexadecanedioic acid, octadecanedioic acid, eicosanedioic acid) and its precursors. A fatty dicarboxylic acid (e.g., octanedioic acid, decanedioic acid, dodecanedioic acid, tetradecanedioic acid, hexadecanedioic acid, octadecanedioic acid, eicosanedioic acid) producing activity can be provided by any non-mammalian source in certain embodiments. Such sources include, without limitation, eukaryotes such as yeast and fungi and prokaryotes such as bacteria. In some embodiments, a reverse activity in a pathway described herein can be altered (e.g., disrupted, reduced) to increase carbon flux through a beta oxidation pathway, an omega oxidation pathway, or a beta oxidation and omega oxidation pathway, towards the production of target product (e.g., sebacic or dodecanedioic acid). In some embodiments, a genetic modification disrupts an activity in the beta oxidation pathway, or disrupts a polynucleotide that encodes a polypeptide that carries out a forward reaction in the beta oxidation pathway, which renders beta oxidation activity undetectable. The term "undetectable" as used herein refers to an amount of an analyte that is below the limits of detection, using detection methods or assays known (e.g., described herein). In certain embodiments, the genetic modification partially reduces beta oxidation activity. The term "partially reduces beta oxidation activity" as used here refers to a level of activity in an engineered organism that is lower than the level of activity found in the host or starting organism.

In some embodiments, a beta-oxidation activity can be modified to alter the catalytic specificity of the chosen activity. In certain embodiments, an acyl-CoA oxidase activity can be altered by modifying a catalytic domain associated with carbon chain length preference and/or specificity. In some embodiments, the altered catalytic specificity can be found by screening naturally occurring variant or mutant populations of a host organism. In certain embodiments, the altered catalytic can be generated by various mutagenesis techniques in conjunction with selection and/or screening for the desired activity. In some embodiments, the altered catalytic activity can be generated by generating chimeric acyl-CoA oxidases using a mix and match approach, followed by selection and/or screening for the desired catalytic activity. Examples of experiments performed to generate acyl-CoA oxidases with altered catalytic activity are described herein.

An activity within an engineered microorganism provided herein can include one or more (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 or all) of the following activities: 6-oxohexanoic acid dehydrogenase activity; 6-hydroxyhexanoic acid dehydrogenase activity; cytochrome P450 activity; cytochrome P450 reductase activity; fatty alcohol oxidase activity; acyl-CoA ligase activity, acyl-CoA oxidase activity; enoyl-CoA hydratase activity, 3-hydroxyacyl-CoA dehydrogenase activity, fatty acid synthase activity, lipase activity, acetyl-CoA carboxylase activity, acyltransferase activity (diacylglycerol acyl transferase, lecithin-cholesterol acyltransferase, phospholipid:diacylglycerol acyltransferase) and thioesterase activity (e.g., acyl-CoA hydrolase, acyl-CoA thioesterase, acyl-ACP thioesterase, acetyl-CoA C-acyltransferase, beta-ketothiolase, and the like). In certain

embodiments, one or more (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or all) of the foregoing activities is altered by way of a genetic modification. In some embodiments, one or more (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 or all) of the foregoing activities is altered by way of (i) adding a heterologous polynucleotide that encodes a polypeptide having the activity, and/or (ii) altering or adding a regulatory sequence that regulates the expression of a polypeptide having the activity. In

certain embodiments, one or more of the foregoing activities is altered by way of (i) disrupting an endogenous polynucleotide that encodes a polypeptide having the activity (e.g., insertional mutagenesis), (ii) deleting a regulatory sequence that regulates the expression of a polypeptide having the activity, and/or (iii) deleting the coding sequence that encodes a polypeptide having the activity (e.g., knock out mutagenesis).

20

The term "omega hydroxyl fatty acid dehydrogenase activity" as used herein refers to conversion of an omega hydroxyl fatty acid to an omega oxo fatty acid. The omega hydroxyl fatty acid dehydrogenase activity can be provided by a polypeptide. In some embodiments, the polypeptide is encoded by a heterologous nucleotide sequence introduced to a host microorganism. In certain

embodiments, an endogenous polypeptide having the omega hydroxyl fatty acid dehydrogenase activity is identified in the host microorganism, and the host microorganism is genetically altered to increase the amount of the polypeptide produced (e.g., a heterologous promoter is introduced in operable linkage with a polynucleotide that encodes the polypeptide; the copy number of a polynucleotide that encodes the polypeptide is increased (e.g., by introducing a plasmid that

includes the polynucleotide)). Nucleic acid sequences conferring omega hydroxyl fatty acid dehydrogenase activity can be obtained from a number of sources, including Actinobacter, Norcardia, Pseudomonas and Xanthobacter bacteria. Examples of an amino acid sequence of a polypeptide having omega hydroxyl fatty acid dehydrogenase activity and a nucleotide sequence of a polynucleotide that encodes the polypeptide, are presented herein. Presence, absence or

30

amount of omega hydroxyl fatty acid dehydrogenase activity can be detected by any suitable method known in the art. In some embodiments, omega hydroxyl fatty acid dehydrogenase activity is not altered in a host microorganism, and in certain embodiments, the activity is added or increased in the engineered microorganism relative to the host microorganism.

5

The term "monooxygenase activity" as used herein refers to inserting one atom of oxygen from O₂ into an organic substrate (RH) and reducing the other oxygen atom to water. In some embodiments, monooxygenase activity refers to incorporation of an oxygen atom onto a six-carbon organic substrate. In certain embodiments, monooxygenase activity refers to conversion of

10

hexanoate to 6-hydroxyhexanoic acid. Monooxygenase activity can be provided by any suitable polypeptide, such as a cytochrome P450 polypeptide (hereafter "CYP450") in certain embodiments. Nucleic acid sequences conferring CYP450 activity can be obtained from a number of sources, including *Bacillus megaterium* and may be induced in organisms including but not limited to *Candida tropicalis*, *Yarrowia lipolytica*, *Aspergillus nidulans*, and *Aspergillus parasiticus*.

15

Examples of oligonucleotide sequences utilized to isolate a polynucleotide sequence encoding a polypeptide having CYP450 activity (e.g., CYP52A12 polynucleotide, a CYP52A13 polynucleotide, a CYP52A14 polynucleotide, a CYP52A15 polynucleotide, a CYP52A16 polynucleotide, a CYP52A17 polynucleotide, a CYP52A18 polynucleotide, a CYP52A19 polynucleotide, a CYP52A20 polynucleotide, a CYP52D2 polynucleotide, and/or a BM3 polynucleotide) are

20

presented herein. In some embodiments, monooxygenase activity is not altered in a host microorganism, and in certain embodiments, the activity is added or increased in the engineered microorganism relative to the host microorganism. In some embodiments, the altered monooxygenase activity is an endogenous activity, and in certain embodiments, the altered monooxygenase activity is an exogenous activity. In some embodiments, the exogenous activity is

25

a single polypeptide with both monooxygenase and monooxygenase reductase activities (e.g., *B. megaterium* cytochrome P450:NADPH P450 reductase).

Presence, absence or amount of cytochrome P450 activity can be detected by any suitable method known in the art. For example, detection can be performed by assaying a reaction containing

30

cytochrome P450 (CYP52A family) and NADPH - cytochrome P450 reductase (see Appl Environ Microbiol 69: 5983 and 5992). Briefly, cells are grown under standard conditions and harvested for production of microsomes, which are used to detect CYP activity. Microsomes are prepared by lysing cells in Tris-buffered sucrose (10mM Tris-HCl pH 7.5, 1mM EDTA, 0.25M sucrose). Differential centrifugation is performed first at 25,000xg then at 100,000xg to pellet cell debris then

microsomes, respectively. The microsome pellet is resuspended in 0.1M phosphate buffer (pH 7.5), 1mM EDTA to a final concentration of approximately 10mg protein/mL. A reaction mixture containing approximately 0.3mg microsomes, 0.1mM sodium hexanoate, 0.7mM NADPH, 50mM Tris-HCl pH 7.5 in 1mL is initiated by the addition of NADPH and incubated at 37°C for 10 minutes.

5 The reaction is terminated by addition of 0.25mL 5M HCl and 0.25mL 2.5ug/mL 10-hydroxydecanoic acid is added as an internal standard (3.3 nmol). The mixture is extracted with 4.5mL diethyl ether under NaCl-saturated conditions. The organic phase is transferred to a new tube and evaporated to dryness. The residue is dissolved in acetonitrile containing 10mM 3-bromomethyl-7-methoxy-1,4-benzoxazin-2-one (BrMB) and 0.1mL of 15mg/mL 18-crown-6 in

10 acetonitrile saturated with K₂CO₃. The solution is incubated at 40°C for 30 minutes before addition of 0.05mL 2% acetic acid. The fluorescently labeled omega-hydroxy fatty acids are resolved via HPLC with detection at 430nm and excitation at 355nm (Yamada et al., 1991, AnalBiochem 199: 132-136). Optionally, specifically induced CYP gene(s) may be detected by Northern blotting and/or quantitative RT-PCR. (Craft et al., 2003, AppEnvironMicro 69: 5983-5991).

15

The term "monooxygenase reductase activity" as used herein refers to the transfer of an electron from NAD(P)H, FMN, or FAD by way of an electron transfer chain, reducing the ferric heme iron of cytochrome P450 to the ferrous state. The term "monooxygenase reductase activity" as used herein also can refer to the transfer of a second electron via the electron transport system,

20 reducing a dioxygen adduct to a negatively charged peroxo group. In some embodiments, a monooxygenase activity can donate electrons from the two-electron donor NAD(P)H to the heme of cytochrome P450 (e.g., monooxygenase activity) in a coupled two-step reaction in which NAD(P)H can bind to the NAD(P)H-binding domain of the polypeptide having the monooxygenase reductase activity and electrons are shuttled from NAD(P)H through FAD and FMN to the heme of

25 the monooxygenase activity, thereby regenerating an active monooxygenase activity (e.g., cytochrome P450). Monooxygenase reductase activity can be provided by any suitable polypeptide, such as a cytochrome P450 reductase polypeptide (hereafter "CPR") in certain embodiments. Nucleic acid sequences conferring CPR activity can be obtained from and/or induced in a number of sources, including but not limited to *Bacillus megaterium*, *Candida tropicalis*, *Yarrowia lipolytica*, *Aspergillus nidulans*, and *Aspergillus parasiticus*. Examples of oligonucleotide sequences utilized to isolate a polynucleotide sequence encoding a polypeptide having CPR activity are presented herein. In some embodiments, monooxygenase reductase activity is not altered in a host microorganism, and in certain embodiments, the activity is added or

30 increased in the engineered microorganism relative to the host microorganism. In some

embodiments, the altered monooxygenase reductase activity is an endogenous activity, and in certain embodiments, the altered monooxygenase reductase activity is an exogenous activity. In some embodiments, the exogenous activity is a single polypeptide with both monooxygenase and monooxygenase reductase activities (e.g., *B. megaterium* cytochrome P450:NADPH P450 reductase).

Presence, absence or amount of CPR activity can be detected by any suitable method known in the art. For example, an engineered microorganism having an increased number of genes encoding a CPR activity, relative to the host microorganism, could be detected using quantitative nucleic acid detection methods (e.g., southern blotting, PCR, primer extension, the like and combinations thereof). An engineered microorganism having increased expression of genes encoding a CPR activity, relative to the host microorganism, could be detected using quantitative expression based analysis (e.g., RT-PCR, western blot analysis, northern blot analysis, the like and combinations thereof). Alternately, an enzymatic assay can be used to detect Cytochrome P450 reductase activity, where the enzyme activity alters the optical absorbance at 550 nanometers of a substrate solution (Masters, B.S.S., Williams, C.H., Kamin, H. (1967) *Methods in Enzymology*, X, 565-573).

Acyl-CoA Oxidases

The term "acyl-CoA oxidase activity" as used herein refers to the oxidation of a long chain fatty-acyl-CoA to a trans-2,3-dehydroacyl-CoA fatty alcohol. In some embodiments, the acyl-CoA activity is from a peroxisome. In certain embodiments, the acyl-CoA oxidase activity is a peroxisomal acyl-CoA oxidase (POX) activity, carried out by a POX polypeptide. In some embodiments the acyl-CoA oxidase activity is encoded by the host organism and sometimes can be altered to generate an engineered organism. Acyl-CoA oxidase activity is encoded by the POX4 and POX5 genes of *Candida* strain ATCC20336. In certain embodiments, endogenous acyl-CoA oxidase activity can be increased. In some embodiments, acyl-CoA oxidase activity of the POX4 polypeptide or the POX5 polypeptide can be altered independently of each other (e.g., increase activity of POX4 alone, POX5 alone, increase one and disrupt the other, and the like). Increasing the activity of one POX activity, while disrupting the activity of another POX activity, may alter the specific activity of acyl-CoA oxidase with respect to carbon chain length, while maintaining or increasing overall flux through the beta oxidation pathway, in certain embodiments.

In certain embodiments, host acyl-CoA oxidase activity of one of the POX genes can be increased by genetically altering (e.g., increasing) the amount of the polypeptide produced (e.g., a strongly transcribed or constitutively expressed heterologous promoter is introduced in operable linkage with a polynucleotide that encodes the polypeptide; the copy number of a polynucleotide that
5 encodes the polypeptide is increased (e.g., by introducing a plasmid that includes the polynucleotide, integration of additional copies in the host genome)). In some embodiments, the host acyl-CoA oxidase activity can be decreased by disruption (e.g., knockout, insertion mutagenesis, the like and combinations thereof) of an acyl-CoA oxidase gene, or by decreasing the activity of the promoter (e.g., addition of repressor sequences to the promoter or 5'UTR) which
10 transcribes an acyl-CoA oxidase gene.

As noted above, disruption of nucleotide sequences encoding POX4, POX 5, or POX4 and POX5 sometimes can alter pathway efficiency, specificity and/or specific activity with respect to metabolism of carbon chains of different lengths (e.g., carbon chains including fatty alcohols, fatty
15 acids, paraffins, dicarboxylic acids of between about 1 and about 60 carbons in length). In some embodiments, the nucleotide sequence of POX4, POX5, or POX4 and POX5 is disrupted with a URA3 nucleotide sequence encoding a selectable marker, and introduced to a host microorganism, thereby generating an engineered organism deficient in POX4, POX5 or POX4 and POX5 activity. Nucleic acid sequences encoding POX4 and POX5 can be obtained from a number
20 of sources, including *Candida tropicalis*, for example. Examples of POX4 and POX5 amino acid sequences and nucleotide sequences of polynucleotides that encode the polypeptides, are presented herein. Described in the examples are experiments conducted to amplify the activity encoded by the POX5 gene.

Also as noted above, catalytic specificity of acyl-CoA oxidases (e.g., POX4, POX5) can be altered by a variety of methods. Altering the binding and/or catalytic specificity of acyl-CoA oxidases may prove advantageous for generating novel acyl-CoA oxidases with altered chain length recognition, altered chain length catalytic activity, and/or generation of an acyl-CoA oxidase activity with a narrow or specific chain length specificity, thereby allowing further increases in pathway efficiency,
25 specificity and/or specific activity with respect to metabolism of carbon chains of different lengths or metabolism of carbon chain distributions found in a particular chosen feedstock. In some embodiments the altered acyl-CoA oxidase sequences are identified and/or generated by; (i) screening naturally occurring variant populations; (ii) mutagenesis of endogenous sequences; (iii) introduction of heterologous sequences having a desired specificity; (iv) generation of chimeric
30

sequences having a portion of the coding sequence from one polynucleotide source (e.g., gene, organism) and a portion of the coding sequence from another source and/or (v) intelligent design using nucleotide sequences and three dimensional structure analysis from an acyl-CoA oxidase having a desired specificity to remodel an endogenous acyl-CoA oxidase, thereby generating a novel specificity enzyme. In some embodiments a chimeric acyl-CoA oxidase sequence can have polynucleotide sequence contributions from two or more sources. In some embodiments, a chimeric acyl-CoA oxidase sequence comprises a portion of the coding sequences from an endogenous polynucleotide and a portion of the coding sequence from a heterologous polynucleotide. Described in the examples are methods utilized to identify and/or generate acyl-CoA oxidases with novel catalytic and binding specificities.

Introduction of heterologous Acyl CoA Oxidase sequences having a desired specificity

Thousands of Acyl CoA Oxidases and Acyl CoA-like Oxidases have been cloned, sequenced and isolated from a variety of organisms (SEQ ID NO. 51 through SEQ ID NO. 3673 and SEQ ID NO. 3810 through SEQ ID NO. 3882). Many of these enzymes have reported catalytic activity with selective substrate specificity. For example, some Acyl CoA Oxidases (e.g., Pox5p from a Candida strain) display optimal activity on substrates of 12 to 18 carbons (FIG. 21). In some embodiments, an organism (e.g., a yeast) or a genetically modified organism (e.g., a genetically modified yeast, e.g., a yeast in which β -oxidation activity is blocked) is engineered to express a heterologous Acyl-CoA Oxidase with selective substrate specificity. In some embodiments, an organism (e.g., a yeast) or a genetically modified organism (e.g., a genetically modified yeast, e.g., a yeast in which β -oxidation activity is blocked) is engineered to express an Acyl-CoA Oxidase or Acyl CoA-like Oxidase selected from SEQ ID NO. 51 to SEQ ID NO. 3673. In some embodiments, an organism (e.g., a yeast) or a genetically modified organism (e.g., a genetically modified yeast, e.g., a yeast in which β -oxidation activity is blocked) is engineered to express an Acyl-CoA Oxidase or Acyl CoA-like Oxidase selected from SEQ ID NO. 3810 through SEQ ID NO. 3882.

Presence, absence or amount of acyl-coA oxidase activity can be detected by any suitable method known in the art. For example, using enzymatic assays as described in Shimizu et al, 1979, and as described herein in the Examples, POX4, POX5 and other acyl-coA oxidase activities can be assessed. Alternatively, nucleic acid sequences representing native and/or disrupted POX4 and POX5 sequences also can be detected using nucleic acid detection methods (e.g., PCR, primer extension, nucleic acid hybridization, the like and combinations thereof), or quantitative expression

based analysis (e.g., RT-PCR, western blot analysis, northern blot analysis, the like and combinations thereof), where the engineered organism exhibits decreased RNA and/or polypeptide levels as compared to the host organism.

5 *Genetic Modification of Acyl-CoA Oxidases*

A rate-limiting step for β -oxidation is the first step in the pathway carried out by the enzyme acyl CoA oxidase. Different Acyl-CoA oxidases can display different chain-length substrate specificities. Some acyl CoA oxidases display broad chain-length specificity and can accept any
10 fatty acyl CoA (or diacyl-CoA) as a substrate. However, some acyl CoA oxidases can display narrow chain-length specificity.

For example the Pox5 enzyme from *Candida* strain ATCC20336 displays a decrease in activity on substrates below C10 (FIG. 21) and has low activity on C6 and C8 substrates. In a cell with Pox5
15 as the only functional acyl CoA oxidase, long chain fatty acyl-CoA or diacyl-CoA substrates can be shortened to about 8 carbons and do not typically enter another cycle of β -oxidation. The shorter substrates (e.g., a C8 fatty dicarboxylic acid) are not typically recognized as a substrate by Pox5, the CoA is removed by peroxisomal thioesterases and the fatty dicarboxylic acid (e.g., an α,ω - dicarboxylic acid) product is secreted from the cell. In this embodiment, the acyl CoA oxidase
20 chain-length substrate specificity effectively controls the chain length of a diacid produced.

In some embodiments, a β -oxidation pathway in a yeast is active and includes a genetically modified acyl CoA oxidase. In some embodiments, an acyl CoA oxidase is genetically modified to prevent complete oxidation of fatty acyl-CoA or diacyl-CoA substrates. Genetic modification of an
25 acyl CoA oxidase can increase the production yield of a desired fatty acid or fatty dicarboxylic acid product. Therefore, in some embodiments, metabolic degradation of a fatty acid of a specified chain length (e.g., the chain length of a desired or target fatty acid or fatty dicarboxylic acid product) is reduced significantly, when an acyl CoA oxidase is genetically modified. In some embodiments, metabolic degradation of a fatty dicarboxylic acid product (e.g., DDDA) by beta-
30 oxidation is reduced significantly, when an acyl CoA oxidase is genetically modified. This can be accomplished by modifying the substrate specificity of an acyl CoA oxidase such that the enzyme has low activity (e.g., enzymatic activity) with chain lengths less than that of a desired product.

In some embodiments, the substrate specificity of an acyl CoA oxidase is modified such that the enzyme has low activity for aliphatic molecules with chain lengths less than C24 (i.e., 24 carbons). In some embodiments, the substrate specificity of an acyl CoA oxidase is modified such that the enzyme has very low activity with chain lengths less than 24, 22, 20, 18, 16, 14, 12, 10, 8, 6 or 4 carbons. In some embodiments, the substrate specificity of an acyl CoA oxidase is modified such that the enzyme has very low activity with chain lengths less than 18, 16, 14, 12, 10 or 8 carbons. In some embodiments, the substrate specificity of an acyl CoA oxidase is modified such that the enzyme has very low activity with chain lengths less than C12. In some embodiments, the substrate specificity of an acyl CoA oxidase is modified such that the enzyme has very low activity with chain lengths less than C10.

In some embodiments, genes encoding a genetically modified acyl CoA oxidase are engineered and expressed in a suitable organism (e.g., a bacteria (e.g., E.coli) or a yeast) to test the substrate specificity of the modified enzyme in vitro. In some embodiments, genes encoding a genetically modified acyl CoA oxidase are engineered and expressed in a suitable yeast and the substrate specificity is tested. In some embodiments, yeast that express a modified acyl CoA oxidase are tested for production of the desired fatty acid or fatty dicarboxylic acid product. A modified acyl CoA oxidase can be generated in any suitable manner, non limiting examples of which are provided hereafter.

20

Random Mutagenesis of Acyl-CoA Oxidase

A library of genetically modified acyl CoA oxidases can be generated using several methods known in the art (e.g., site-directed mutagenesis). Genetically modified acyl CoA oxidase genes can then be transformed into a β -oxidation blocked strain of a suitable yeast strain (e.g., *Candida* spp. (e.g., *Candida viswanathii* or *Candida tropicalis*)). In some embodiments, a genetically modified acyl CoA oxidase is expressed under the control of the POX4 promoter or another strong constitutive or inducible promoter in a *pox4 Δ /pox4 Δ pox5 Δ /pox5 Δ* (e.g., an organism that lacks endogenous acyl CoA oxidase activity) background. In some embodiments, the genetically modified acyl CoA oxidase is expressed under the control of endogenous promoter. In some embodiments, the genetically modified acyl CoA oxidase is expressed under the control of a heterologous promoter. The transformants can be selected by growth in a fatty acid or methyl-derivate fatty acid containing fatty acids with two more carbons than the diacid product of interest. For example, for adipic acid, the transformants can be grown in caprylic acid or methyl-caprylate.

30

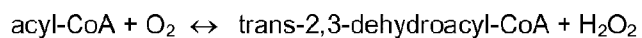
For example, for dodecanedioic acid, the transformants can be grown in tetradecanedioic acid. The group of transformants can then be moved to a medium with a carbon source of a fatty acid of interest (for example dodecanedioic acid) in the presence of an agent that kills growing cells (e.g., Nystatin) and cells that cannot metabolize the carbon source (e.g., dodecanedioic acid in this example) can be selected. The resulting modified strains can then be further characterized for acyl CoA oxidase activity. This method can be used to select for any modified acyl CoA oxidase (e.g., those listed and/or described in TABLES 9 through 26). In addition, this method can be used to select for any heterologous acyl CoA oxidase (e.g., those listed in SEQ ID NO. 51 through 3273 and SEQ ID NO. 3728 through 3810) expressed in a suitable organism.

Rational Mutagenesis of Acyl-CoA Oxidase

Structural and sequence information and experimental data can be combined to determine specific mutations to be tested in a acyl-CoA oxidase for altered specificity. For example, primary sequences of acyl-CoA oxidases tested can be compared and correlated with substrate specificity. Based on such an analysis, single amino-acids, small numbers of contiguous amino acids and/or domains can be proposed for providing a desired substrate specificity. Those amino acids positions can be targeted for specific or random mutations for improve specificity.

Acyl CoA oxidase structure also can be modeled against a known tertiary structure using modeling methods known in the art. The models can be used to propose amino acids and regions pertaining to substrate selectivity. For example, biochemical, structure and sequence data suggest that the N-terminus of acyl CoA oxidases often, in part, determines substrate specificity. Mutations or region replacements can be introduced based on such analyses and the specificity of the new acyl CoA oxidase tested as described before. The resulting information can be used to go back to the models to postulate new potential mutations. As for random mutagenesis, any suitable acyl CoA oxidase can be modified to alter substrate specificity (e.g., those listed in SEQ ID NO. 51 through 3273 and SEQ ID NO. 3728 through 3810)

The term "acyl CoA oxidase activity" as used herein refers to the enzymatic activity (e.g., catalytic activity) of a acyl CoA oxidase. An acyl CoA oxidase can catalyze the following chemical reaction:



In some embodiments, acyl CoA oxidase activity refers to oxidation of a long chain fatty-acyl-CoA to a trans-2,3-dehydroacyl-CoA fatty alcohol. In some embodiments, acyl CoA oxidase activity refers to its enzyme activity (or lack thereof) on a selective set of substrates. The activity of an acyl CoA oxidase can be affected by its ability to bind a substrate, oxidize a substrate and/or release a product. In some embodiments, an acyl CoA oxidase is active in one compartment of a cell and not in another compartment of the cell. In some embodiments, the acyl CoA oxidase activity is from a peroxisome. In certain embodiments, the acyl CoA oxidase activity is a peroxisomal acyl CoA oxidase (POX) activity, carried out by a POX polypeptide. In some embodiments the acyl CoA oxidase activity is encoded by the host organism and sometimes can be altered to generate an engineered organism. Acyl-CoA oxidase activity can be encoded by the POX4 and POX5 genes of *Candida* spp.. In certain embodiments, endogenous acyl CoA oxidase activity can be increased. In some embodiments, acyl CoA oxidases in an organism, containing one or more acyl CoA oxidases, can be independently modified (e.g., one or more acyl CoA oxidases can be modified). In some embodiments, acyl CoA oxidase activity of a POX4 polypeptide or a POX5 polypeptide can be altered independently of each other (e.g., increase activity of POX4 alone, POX5 alone, increase one and disrupt the other, and the like). Increasing the activity of one POX activity, while disrupting the activity of another POX activity, may alter the specific activity of acyl CoA oxidase with respect to carbon chain length, while maintaining or increasing overall flux through the beta oxidation pathway, in certain embodiments.

20

In certain embodiments, host activity of one or more acyl CoA oxidase genes can be increased by genetically altering (e.g., increasing) the amount of a polypeptide produced (e.g., a strongly transcribed or constitutively expressed heterologous promoter is introduced in operable linkage with a polynucleotide that encodes a polypeptide; the copy number of a polynucleotide that encodes the polypeptide is increased (e.g., by introducing a plasmid that includes the polynucleotide, integration of additional copies in the host genome)). In some embodiments, host activity of one or more acyl CoA oxidases can be decreased by disruption (e.g., knockout, insertion mutagenesis, the like and combinations thereof) of an acyl CoA oxidase gene, or by decreasing the activity of the promoter (e.g., addition of repressor sequences to the promoter or 5'UTR) which transcribes an acyl CoA oxidase gene.

30

As noted above, disruption of nucleotide sequences encoding one or more acyl CoA oxidases (e.g., POX4, POX 5, or POX4 and POX5) sometimes can alter pathway efficiency, specificity and/or specific activity with respect to metabolism of carbon chains of different lengths (e.g.,

carbon chains including fatty alcohols, fatty acids, paraffins, dicarboxylic acids, aliphatic molecules of between about 1 and about 60 carbons in length). In some embodiments, the nucleotide sequence of one or more acyl CoA oxidases (e.g., POX4, POX 5, or POX4 and POX5) is disrupted with a URA3 nucleotide sequence encoding a selectable marker, and introduced to a host
5 microorganism, thereby generating an engineered organism deficient in an acyl CoA oxidase activity.

Also as noted above, catalytic specificity of acyl CoA oxidases (e.g., POX4, POX5) can be altered by a variety of methods. Altering the binding and/or catalytic specificity of acyl CoA oxidases may
10 prove advantageous for generating novel acyl CoA oxidases with altered chain length recognition, altered chain length catalytic activity, and/or generation of an acyl CoA oxidase activity with a narrow or specific chain length specificity, thereby allowing further increases in pathway efficiency, specificity and/or specific activity with respect to metabolism of carbon chains of different lengths or metabolism of carbon chain distributions found in a particular chosen feedstock. In some
15 embodiments the altered acyl CoA oxidase sequences are identified and/or generated by: (i) screening naturally occurring variant populations; (ii) mutagenesis of endogenous sequences; (iii) introduction of heterologous sequences having a desired specificity (e.g., introduction of one or more unmodified or modified acyl CoA oxidases from another organism into a host organism in which one or more endogenous acyl-CoA oxidases are optionally disrupted); (iv) generation of
20 chimeric sequences having a portion of the coding sequence from one polynucleotide source (e.g., gene, organism) and a portion of the coding sequence from another source and/or (v) intelligent design using nucleotide sequences and three dimensional structure analysis from an acyl CoA oxidase having a desired specificity to remodel an endogenous acyl CoA oxidase, thereby generating a novel specificity enzyme. In some embodiments a chimeric acyl CoA oxidase
25 sequence can have polynucleotide sequence contributions from two or more sources. In some embodiments, a chimeric acyl CoA oxidase sequence comprises a portion of the coding sequences from an endogenous polynucleotide and a portion of the coding sequence from a heterologous polynucleotide. Described in the examples are methods utilized to identify and/or generate acyl CoA oxidases with novel catalytic and binding specificities.

30

Nucleic acid sequences encoding acyl CoA oxidases (e.g., POX4 and POX5) can be obtained from any suitable source, including any animal (e.g., mammals, fish, reptiles, amphibians, etc.), any plant, fungus, yeast, protozoan, bacteria, virus, phage, and the like). Non-limiting examples of suitable yeast sources include Yarrowia yeast (e.g., Y. lipolytica (formerly classified as Candida

lipolytica)), *Candida* yeast (e.g., *C. revkaui*, *C. viswanathii*, *C. pulcherrima*, *C. tropicalis*, *C. utilis*), *Rhodotorula* yeast (e.g., *R. glutinus*, *R. graminis*), *Rhodospiridium* yeast (e.g., *R. toruloides*), *Saccharomyces* yeast (e.g., *S. cerevisiae*, *S. bayanus*, *S. pastorianus*, *S. carlsbergensis*), *Cryptococcus* yeast, *Trichosporon* yeast (e.g., *T. pullans*, *T. cutaneum*), *Pichia* yeast (e.g., *P. pastoris*) and *Lipomyces* yeast (e.g., *L. starkeyii*, *L. lipoferus*). In some embodiments, a suitable yeast is of the genus *Arachniotus*, *Aspergillus*, *Aureobasidium*, *Auxarthron*, *Blastomyces*, *Candida*, *Chrysosporium*, *Chrysosporium* *Debaryomyces*, *Coccidiodes*, *Cryptococcus*, *Gymnoascus*, *Hansenula*, *Histoplasma*, *Issatchenkia*, *Kluyveromyces*, *Lipomyces*, *Issatchenkia*, *Microsporum*, *Myxotrichum*, *Myxozyma*, *Oidiodendron*, *Pachysolen*, *Penicillium*, *Pichia*, *Rhodospiridium*, *Rhodotorula*, *Rhodotorula*, *Saccharomyces*, *Schizosaccharomyces*, *Scopulariopsis*, *Sepedonium*, *Trichosporon*, or *Yarrowia*. In some embodiments, a suitable yeast is of the species *Arachniotus* *flavoluteus*, *Aspergillus* *flavus*, *Aspergillus* *fumigatus*, *Aspergillus* *niger*, *Aureobasidium* *pullulans*, *Auxarthron* *thaxteri*, *Blastomyces* *dermatitidis*, *Candida* *albicans*, *Candida* *dublinsiensis*, *Candida* *famata*, *Candida* *glabrata*, *Candida* *guilliermondii*, *Candida* *kefyr*, *Candida* *krusei*, *Candida* *lambica*, *Candida* *lipolytica*, *Candida* *lustritanae*, *Candida* *parapsilosis*, *Candida* *pulcherrima*, *Candida* *revkaui*, *Candida* *rugosa*, *Candida* *tropicalis*, *Candida* *utilis*, *Candida* *viswanathii*, *Candida* *xestobii*, *Chrysosporium* *keratinophilum*, *Coccidiodes* *immitis*, *Cryptococcus* *albidus* var. *diffuens*, *Cryptococcus* *laurentii*, *Cryptococcus* *neofomans*, *Debaryomyces* *hansenii*, *Gymnoascus* *dugwayensis*, *Hansenula* *anomala*, *Histoplasma* *capsulatum*, *Issatchenkia* *occidentalis*, *Issatchenkia* *orientalis*, *Kluyveromyces* *lactis*, *Kluyveromyces* *marxianus*, *Kluyveromyces* *thermotolerans*, *Kluyveromyces* *waltii*, *Lipomyces* *lipoferus*, *Lipomyces* *starkeyii*, *Microsporum* *gypseum*, *Myxotrichum* *deflexum*, *Oidiodendron* *echinulatum*, *Pachysolen* *tannophilis*, *Penicillium* *notatum*, *Pichia* *anomala*, *Pichia* *pastoris*, *Pichia* *stipitis*, *Rhodospiridium* *toruloides*, *Rhodotorula* *glutinus*, *Rhodotorula* *graminis*, *Saccharomyces* *cerevisiae*, *Saccharomyces* *kluyveri*, *Schizosaccharomyces* *pombe*, *Scopulariopsis* *acremonium*, *Sepedonium* *chrysospermum*, *Trichosporon* *cutaneum*, *Trichosporon* *pullans*, *Yarrowia* *lipolytica*, or *Yarrowia* *lipolytica* (formerly classified as *Candida* *lipolytica*). In some embodiments, a suitable yeast is a *Y. lipolytica* strain that includes, but is not limited to, ATCC20362, ATCC8862, ATCC18944, ATCC20228, ATCC76982 and LGAM S(7)1 strains (Papanikolaou S., and Aggelis G., *Bioresour. Technol.* 82(1):43-9 (2002)). In certain embodiments, a suitable yeast is a *Candida* species (i.e., *Candida* spp.) yeast. Any nucleic acid sequence encoding an acyl CoA oxidase, acyl CoA oxidase-like activity or acyl-CoA dehydrogenase activity can be used to alter the substrate specificity of a yeast as described herein. Non-limiting examples of acyl CoA oxidase, acyl CoA oxidase-like and acyl CoA dehydrogenase amino acid sequences and nucleotide sequences are provided herein and in

SEQ ID NO. 51 through 3810. Described in the examples are experiments conducted to modify and amplify the activity of an acyl CoA oxidase gene (e.g., the POX5 gene).

5 Presence, absence or amount of acyl CoA oxidase activity can be detected by any suitable method known in the art. For example, enzymatic assays as described in Shimizu et al, 1979, and as described herein in the Examples can be used to assess acyl CoA oxidase activity. Nucleic acid sequences representing native and/or disrupted acyl CoA oxidase sequences also can be detected using nucleic acid detection methods (e.g., PCR, primer extension, nucleic acid hybridization, the like and combinations thereof), or quantitative expression based analysis (e.g., RT-PCR, western blot analysis, northern blot analysis, the like and combinations thereof), where the engineered
10 organism exhibits decreased RNA and/or polypeptide levels as compared to the host organism.

Acyl CoA Dehydrogenase

15 Acyl-CoA dehydrogenases (ACADs) are a class of enzymes that can function to catalyze the initial step in each cycle of fatty acid β -oxidation in the mitochondria of cells. They can be very similar in structure and function to Acyl CoA oxidases. Their action results in the introduction of a trans double-bond between C2 and C3 of an acyl-CoA thioester substrate. FAD is a required co-factor in the mechanism in order for the enzyme to bind to its appropriate substrate.

20 Acyl-CoA dehydrogenases can be categorized into four distinct groups based on their specificity for short-, medium-, or long-chain fatty acid, and very long-chain fatty acid acyl-CoA substrates. While different dehydrogenases target fatty acids of varying chain length, all types of acyl-CoA dehydrogenases can be mechanistically similar. Differences in ACADs can occur based on the
25 location of the active site along the amino acid sequence.

The medium chain acyl-CoA dehydrogenase is a homotetramer with each subunit containing roughly 400 amino acids and one equivalent of FAD. The tetramer is classified as a "dimer of dimers".

30 The interface between the two monomers of a single dimer of an acyl-CoA dehydrogenase contains the FAD binding sites and has extensive bonding interactions. In contrast, the interface between the two dimers has few interactions. There are a total of 4 active sites within the tetramer,

each of which contains a single FAD molecule and an acyl-CoA substrate. This gives a total of four FAD molecules and four acyl-CoA substrates per enzymatic molecule.

5 FAD is bound between the three domains of the monomer, where only the nucleotide portion is accessible. FAD binding contributes significantly to overall enzyme stability. The acyl-CoA substrate is bound completely within each monomer of the enzyme. In some ACADs, the active site is lined with the residues F252, T255, V259, T96, T99, A100, L103, Y375, Y375, and E376. The area of interest within the substrate can become wedged between Glu 376 and FAD, lining up the molecules into an ideal position for the reaction.

10

Some ACAD sequences are presented in SEQ ID NO.s 3728 through 3810.

Thioesterase

15 The term "thioesterase activity" as used herein refers to removal of Coenzyme A from hexanoate. The term "thioesterase activity" as used herein also refers to the removal of Coenzyme A from an activated fatty acid (e.g., fatty-acyl-CoA). A Non-limiting example of an enzyme with thioesterase activity includes acyl-CoA hydrolase (e.g., EC 3.1.2.20; also referred to as acyl coenzyme A thioesterase, acyl-CoA thioesterase, acyl coenzyme A hydrolase, thioesterase B, thioesterase II, 20 lecithinase B, lysophospholipase L1, acyl-CoA thioesterase 1, and acyl-CoA thioesterase). Thioesterases that remove Coenzyme A from fatty-acyl-CoA molecules catalyze the reaction,

acyl-CoA + H₂O → CoA + a carboxylate,

25 where the carboxylate often is a fatty acid. The released Coenzyme A can then be reused for other cellular activities.

The thioesterase activity can be provided by a polypeptide. In certain embodiments, the polypeptide is an endogenous nucleotide sequence that is increased in copy number, operably 30 linked to a heterologous and/or endogenous promoter, or increased in copy number and operably linked to a heterologous and/or endogenous promoter. In some embodiments, the polypeptide is encoded by a heterologous nucleotide sequence introduced to a host microorganism. Nucleic acid sequences conferring thioesterase activity can be obtained from a number of sources, including *Cuphea lanceolata*, *C. tropicalis* (e.g., see SEQ ID NOS: 33 and 35), and *E. coli* (e.g., see SEQ ID

- NO: 37). Additional organisms that can be used as thioesterase polynucleotide sequence donors are given herein. Examples of such polypeptides include, without limitation, acyl-(ACP) thioesterase type B from *Cuphea lanceolata* (see SEQ ID NO: 1), acyl-CoA hydrolase (e.g., ACHA and ACHB, see SEQ ID NOS: 34 and 36)) from *C. tropicalis*, acyl-CoA thioesterase (e.g., TESA, see SEQ ID NO: 38) from *E. coli*. A non-limiting example of a thioesterase polynucleotide sequences is referenced by accession number CAB60830 at the World Wide Web Uniform Resource Locator (URL) ncbi.nlm.nih.gov of the National Center for Biotechnology Information (NCBI).
- 10 Presence, absence or amount of thioesterase activity can be detected by any suitable method known in the art. An example of such a method is described Chemistry and Biology 9: 981-988. In some embodiments, thioesterase activity is not altered in a host microorganism, and in certain embodiments, the activity is added or increased in the engineered microorganism relative to the host microorganism. In some embodiments, a polypeptide having thioesterase activity is linked to
- 15 another polypeptide (e.g., a hexanoate synthase A or hexanoate synthase B polypeptide). Non-limiting examples of polynucleotide sequences encoding thioesterase activities and polypeptides having thioesterase activity are provided in Example 33.

Reducing Omega Fatty Acid Conversion - General

20

- The term "a genetic modification that reduces omega hydroxyl fatty acid conversion" as used herein refer to genetic alterations of a host microorganism that reduce an endogenous activity that converts an omega hydroxyl fatty acid to another product. In some embodiments, an endogenous omega hydroxyl fatty acid dehydrogenase activity is reduced. Such alterations can
- 25 advantageously increase the amount of a dicarboxylic acid, which can be purified and further processed.

Reducing Beta Oxidation - General

- 30 The term "a genetic modification that reduces beta-oxidation activity" as used herein refers to a genetic alteration of a host microorganism that reduces an endogenous activity that oxidizes a beta carbon of carboxylic acid containing organic molecules. In certain embodiments, the organic molecule is a ten or twelve carbon molecule, and sometimes contains one or two carboxylic acid moieties located at a terminus of the molecule (e.g., sebacic or dodecanedioic acid). Such

alterations can advantageously increase yields of end products, such as a fatty dicarboxylic acid (e.g., octanedioic acid, decanedioic acid, dodecanedioic acid, tetradecanedioic acid, hexadecanedioic acid, octadecanedioic acid, eicosanedioic acid).

5 *Increasing Fatty Acid Synthesis - General*

The term "a genetic modification that results in increased fatty acid synthesis" as used herein also refers to a genetic alteration of a host microorganism that reduces an endogenous activity that converts fatty acids into fatty-acyl-CoA intermediates. In some embodiments, an endogenous
 10 activity that converts fatty acids into fatty-acyl-CoA intermediates is reduced. In certain embodiments, an acyl-CoA synthetase activity is reduced. Such alterations can advantageously increase yields of end products, such as a fatty dicarboxylic acid (e.g., octanedioic acid, decanedioic acid, dodecanedioic acid, tetradecanedioic acid, hexadecanedioic acid, octadecanedioic acid, eicosanedioic acid).

15

Acyl-CoA Synthetase

Fatty acids can be converted into fatty-acyl-CoA intermediates by the activity of an acyl-CoA synthetase (e.g., ACS1, ACS2; EC 6.2.1.3; also referred to as acyl-CoA synthetase, acyl-CoA
 20 ligase), in many organisms. Acyl-CoA synthetase has six isoforms encoded by ACS1, FAT1, ACS2A, ACS2B, ACS2C and ACS2D, respectively, in *Candida* spp. (e.g., homologous to FAA1, FAT1, and FAA2 in *S. cerevisiae*). Acyl-CoA synthetase is a member of the ligase class of enzymes and catalyzes the reaction,

25 $\text{ATP} + \text{Fatty Acid} + \text{CoA} \rightleftharpoons \text{AMP} + \text{Pyrophosphate} + \text{Fatty-Acyl-CoA}.$

Fatty acids and Coenzyme A often are utilized in the activation of fatty acids to fatty-acyl-CoA intermediates for entry into various cellular processes. Without being limited by theory, it is believed that reduction in the amount of fatty-acyl-CoA available for various cellular processes can
 30 increase the amount of fatty acids available for conversion into a fatty dicarboxylic acid (e.g., a sebacic or dodecanedioic acid) by other engineered pathways in the same host organism (e.g., omega oxidation pathway, beta oxidation pathway, omega oxidation pathway and beta oxidation pathway). Acyl-CoA synthetase can be inactivated by any suitable means. Described herein are gene knockout methods suitable for use to disrupt the nucleotide sequence that encodes a

polypeptide having ACS1 activity. A nucleotide sequence of ACS1 is provided in Example 33, SEQ ID NO: 39. An example of an integration/disruption construct, configured to generate a deletion mutant for ACS1 is also provided in the Examples.

- 5 The presence, absence or amount of acyl-CoA synthetase activity can be detected by any suitable method known in the art. Non-limiting examples of suitable detection methods include enzymatic assays (e.g., Lageweg et al "A Fluorimetric Assay for Acyl-CoA Synthetase Activity", Analytical Biochemistry, 197(2):384-388 (1991)), PCR based assays (e.g., qPCR, RT-PCR), immunological detection methods (e.g., antibodies specific for acyl-CoA synthetase), the like and combinations thereof.

The term "a genetic modification that results in increased fatty acid synthesis" as used herein also refers to a genetic alteration of a host microorganism that reduces an endogenous activity that converts long chain and very long chain fatty acids into activated fatty-acyl-CoA intermediates. In some embodiments, an endogenous activity that converts long chain and very long chain fatty acids into activated fatty-acyl-CoA intermediates is reduced. In certain embodiments, a long chain acyl-CoA synthetase activity is reduced. Such alterations can advantageously increase yields of end products, such as a fatty dicarboxylic acid (e.g., octanedioic acid, decanedioic acid, dodecanedioic acid, tetradecanedioic acid, hexadecanedioic acid, octadecanedioic acid, eicosanedioic acid).

Long chain fatty acids (e.g., C12-C18 chain lengths) and very long chain fatty acids (e.g., C20-C26) often are activated and/or transported by the thioesterification activity of a long-chain acyl-CoA synthetase (e.g., FAT1; EC 6.2.1.3; also referred to as long-chain fatty acid-CoA ligase, acyl-CoA synthetase; fatty acid thiokinase (long chain); acyl-activating enzyme; palmitoyl-CoA synthase; lignoceroyl-CoA synthase; arachidonyl-CoA synthetase; acyl coenzyme A synthetase; acyl-CoA ligase; palmitoyl coenzyme A synthetase; thiokinase; palmitoyl-CoA ligase; acyl-coenzyme A ligase; fatty acid CoA ligase; long-chain fatty acyl coenzyme A synthetase; oleoyl-CoA synthetase; stearoyl-CoA synthetase; long chain fatty acyl-CoA synthetase; long-chain acyl CoA synthetase; fatty acid elongase (ELO); LCFA synthetase; pristanoyl-CoA synthetase; ACS3; long-chain acyl-CoA synthetase I; long-chain acyl-CoA synthetase II; fatty acyl-coenzyme A synthetase; long-chain acyl-coenzyme A synthetase; and acid:CoA ligase (AMP-forming)), in some organisms. Fatty acids also can be transported into the host organism from feedstocks by the activity of long chain acyl-CoA synthetase.

Long-chain acyl-CoA synthetase catalyzes the reaction,

ATP + a long-chain carboxylic acid + CoA = AMP + diphosphate + an acyl-CoA,

- 5 where “an acyl-CoA” refers to a fatty-acyl-CoA molecule. As noted herein, activation of fatty acids is often necessary for entry of fatty acids into various cellular processes (e.g., as an energy source, as a component for membrane formation and/or remodeling, as carbon storage molecules). Deletion mutants of FAT1 have been shown to accumulate very long chain fatty acids and exhibit decreased activation of these fatty acids. Without being limited by theory, it is believed that
- 10 reduction in the activity of long-chain acyl-CoA synthetase may reduce the amount of long chain fatty acids converted into fatty-acyl-CoA intermediates, thereby increasing the amount of fatty acids available for conversion into a fatty dicarboxylic acid (e.g., octanedioic acid, decanedioic acid, dodecanedioic acid, tetradecanedioic acid, hexadecanedioic acid, octadecanedioic acid, eicosanedioic acid) by other engineered pathways in the same host organism (e.g., omega
- 15 oxidation pathway, beta oxidation pathway, omega oxidation pathway and beta oxidation pathway). Long-chain-acyl-CoA synthetase activity can be reduced or inactivated by any suitable means. Described herein are gene knockout methods suitable for disrupting the nucleotide sequence that encodes the polypeptide having FAT1 activity. The nucleotide sequence of FAT1 is provided in Example 33, SEQ ID NO: 41. DNA vectors suitable for use in constructing “knockout” constructs
- 20 are described herein.

- The presence, absence or amount of long-chain-acyl-CoA synthetase activity can be detected by any suitable method known in the art. Non-limiting examples of suitable detection methods include enzymatic assays, binding assays (e.g., Erland et al, Analytical Biochemistry 295(1):38-44 (2001)),
- 25 PCR based assays (e.g., qPCR, RTPCR), immunological detection methods (e.g., antibodies specific for long-chain-acyl-CoA synthetase), the like and combinations thereof.

Selective modification of ACS activity

- 30 In some embodiments, a beta-oxidation pathway is functional and is modified for selective substrate specificity. In some embodiments a beta-oxidation pathway is selective for only diacyl-CoA thioesters and in some embodiments only on diacyl-CoA's of a chain length greater than 6, 8, 10, 12, 14, 16, 18 or 20 carbons. Beta-oxidation selectivity can be achieved by: 1) utilizing the difference in transport of acyl-CoA's and diacids across the peroxisomal membrane, 2) selectively

knocking out acyl-CoA synthetase (ACS) activity in the cytosolic compartment, 3) knocking out ACS activity in the peroxisomal compartment for isozymes with substrate specificity for short chain substrates, and/or 4) engineering a beta-oxidation pathway that will work only on substrates longer than 6, 8, 10, 12, 14, 16, 18 or 20 carbons.

5

In *S. cerevisiae*, cytoplasmic ACS activity is encoded by FAA1, FAA3, FAA4 and FAT1, while peroxisomal activity is encoded by FAA2. Homologs for FAA1 and FAT1 were identified in *Candida* strains however there were no identified homologs for FAA3 or FAA4. As many as five homologs for the *S. cerevisiae* peroxisomal FAA2 were identified in *Candida* strains. Two of the
10 five homologs display 95% identity to one another and are most likely alleles of the same gene. Four FAA2 homologs were identified in *Candida* strain ATCC20336 (e.g., ACS2A through ACS2D).

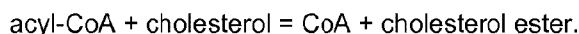
In some embodiments, one strategy is to control the subcellular location of ACS enzyme activity so that it is only present in the peroxisome. FAA1 and FAT1 mutants, *faa1Δ* and *fat1Δ* were
15 constructed and should have very little ACS activity targeted to the cytoplasm. In these mutant strains, exogenously supplied long-chain free fatty acids accumulate in the cytoplasm since they cannot be transported into the peroxisome unless they are activated to the acyl-CoA thioester. High concentrations of free fatty acid can be toxic, so the cell acts to detoxify itself by oxidizing the free fatty acids to dicarboxylic acids that are much less toxic. Unlike long-chain fatty acids, long-
20 chain dicarboxylic acids are able to diffuse into the peroxisomal compartment where they can then be activated to diacyl-CoA thioesters, which is required for entry into the beta-oxidation pathway. With multiple peroxisomal ACS isozymes it may be that each isozyme has different substrate specificity. In some embodiments, it is desired to retain those peroxisomal ACS enzymes with substrate specificity matching the chain-length of the fatty acid feedstock but without activity (or low
25 activity) on diacids of chain-length $\leq 6, 8, 10, 12, 14, 16, 18$ or 20 carbons. With this strategy any long-chain dicarboxyl-CoA that is chain-shortened by beta-oxidation to 12 carbons, for example, that is subsequently hydrolyzed to a dicarboxylic acid and free CoA cannot be reactivated to a dicarboxyl-CoA for re-entry into beta-oxidation for further chain shortening. In some embodiments, in combination with controlling the substrate chain-length specificity of the peroxisomal ACS, a
30 peroxisomal thioesterase activity is amplified with maximum activity at the desired chain-length of our product. This strategy can control the chain-length of the dicarboxylic acid produced by beta-oxidation.

In some embodiments, the flow of fatty acids into the peroxisome is controlled by knocking out the genes PXA1 and PXA2. These genes encode subunits of an ATP binding cassette transporter that is responsible for transporting long-chain fatty acyl-CoA's from the cytoplasm across the peroxisomal membrane into the peroxisomal matrix. Even though, in some embodiments, the genes encoding the cytoplasmic ACS's are knocked out, there may still be some residual ACS activity in the cytoplasm from the peroxisomal ACS's. The ACS isozymes destined for the peroxisome are translated in the cytoplasm and fully folded prior to import into the peroxisome. Therefore the peroxisomal ACS's may contribute to a small amount of cytoplasmic ACS activity. Deletion of the Pxa1p/Pxa2p transporter can prevent any of the long-chain fatty acids activated to acyl-CoA thioesters from being transported into the peroxisome for degradation.

Acyl-CoA sterol acyltransferase

The term "a genetic modification that results in increased fatty acid synthesis" as used herein also refers to a genetic alteration of a host microorganism that reduces an endogenous activity that converts fatty acids into cholesterol esters. In some embodiments, an endogenous activity that converts fatty acids into cholesterol esters is reduced. In certain embodiments, an acyl-CoA sterol acyltransferase activity is reduced. Such alterations can advantageously increase yields of end products, such as a fatty dicarboxylic acid (e.g., octanedioic acid, decanedioic acid, dodecanedioic acid, tetradecanedioic acid, hexadecanedioic acid, octadecanedioic acid, eicosanedioic acid).

Fatty acids can be converted into a cholesterol-ester by the activity of acyl-CoA sterol acyltransferase (e.g., ARE1, ARE2, EC 2.3.1.26; also referred to as sterol O-acyltransferase; cholesterol acyltransferase; sterol-ester synthase; sterol-ester synthetase; sterol-ester synthase; acyl coenzyme A-cholesterol-O-acyltransferase; acyl-CoA:cholesterol acyltransferase; ACAT; acylcoenzyme A:cholesterol O-acyltransferase; cholesterol ester synthase; cholesterol ester synthetase; and cholesteryl ester synthetase), in many organisms. Without being limited by any theory, cholesterol esterification may be involved in directing fatty acids away from incorporation into cell membranes and towards storage forms of lipids. Acyl-CoA sterol acyltransferase catalyzes the reaction,



The esterification of cholesterol is believed to limit its solubility in cell membrane lipids and thus promotes accumulation of cholesterol ester in the fat droplets (e.g., a form of carbon storage molecule) within cytoplasm. Therefore, without being limited by any theory esterification of cholesterol may cause the accumulation of lipid storage molecules, and disruption of the activity of acyl-CoA sterol acyltransferase may cause an increase in acyl-CoA levels that can be converted into a fatty dicarboxylic acid (e.g., octanedioic acid, decanedioic acid, dodecanedioic acid, tetradecanedioic acid, hexadecanedioic acid, octadecanedioic acid, eicosanedioic acid) by other engineered pathways in the same host organism (e.g., omega oxidation pathway, beta oxidation pathway, omega oxidation pathway and beta oxidation pathway). Acyl-CoA sterol acyltransferase can be inactivated by any suitable means. Described herein are gene knockout methods suitable for disrupting nucleotide sequences that encode polypeptides having ARE1 activity, ARE2 activity or ARE1 activity and ARE2 activity. The nucleotide sequences of ARE1 and ARE2 are provided in Example 33, SEQ ID NOS: 43 and 45. DNA vectors suitable for use in constructing "knockout" constructs are described herein.

The presence, absence or amount of acyl-CoA sterol acyltransferase activity can be detected by any suitable method known in the art. Non-limiting examples of suitable detection methods include enzymatic assays (e.g., Chen et al, Plant Physiology 145:974-984 (2007)), binding assays, PCR based assays (e.g., qPCR, RT-PCR), immunological detection methods (e.g., antibodies specific for long-chain-acyl-CoA synthetase), the like and combinations thereof.

Diacylglycerol acyltransferase & Acyltransferases

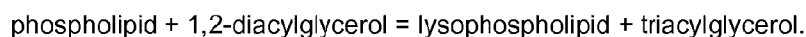
The term "a genetic modification that results in increased fatty acid synthesis" as used herein also refers to a genetic alteration of a host microorganism that reduces an endogenous activity that catalyzes diacylglycerol esterification (e.g., addition of acyl group to a diacylglycerol to form a triacylglycerol). In some embodiments, an endogenous activity that converts diacylglycerol into triacylglycerol is reduced. In certain embodiments, an acyltransferase activity is reduced. In some embodiments a diacylglycerol acyltransferase activity is reduced. In some embodiments a diacylglycerol acyltransferase (e.g., DGA1, EC 2.3.1.20) activity and an acyltransferase (e.g., LRO1) activity are reduced. Such alterations can advantageously increase yields of end products, such as a fatty dicarboxylic acid (e.g., octanedioic acid, decanedioic acid, dodecanedioic acid, tetradecanedioic acid, hexadecanedioic acid, octadecanedioic acid, eicosanedioic acid).

Diacylglycerol can be converted into triacylglycerol by the activity of diacylglycerol acyltransferase (e.g., DGA1; EC 2.3.1.20; also referred to as diglyceride acyltransferase; 1,2-diacylglycerol acyltransferase; diacylglycerol acyltransferase; diglyceride O-acyltransferase; palmitoyl-CoA-sn-1,2-diacylglycerol acyltransferase; acyl-CoA:1,2-diacylglycerol O-acyltransferase and acyl-CoA:1,2-diacyl-sn-glycerol O-acyltransferase), in many organisms. Diacylglycerol acyltransferase catalyzes the reaction,



and is generally considered the terminal and only committed step in triglyceride synthesis. The product of the DGA1 gene in yeast normally is localized to lipid particles.

In addition to the diacylglycerol esterification activity described for DGA1, many organisms also can generate triglycerides by the activity of other acyltransferase activities, non-limiting examples of which include lecithin-cholesterol acyltransferase activity (e.g., LRO1; EC 2.3.1.43; also referred to as phosphatidylcholine-sterol O-acyltransferase activity; lecithin-cholesterol acyltransferase activity; phospholipid-cholesterol acyltransferase activity; LCAT (lecithin-cholesterol acyltransferase) activity; lecithin:cholesterol acyltransferase activity; and lysolecithin acyltransferase activity) and phospholipid:diacylglycerol acyltransferase (e.g., EC 2.3.1.158; also referred to as PDAT activity and phospholipid:1,2-diacyl-sn-glycerol O-acyltransferase activity). Acyltransferases of the families EC 2.3.1.43 and EC 2.3.1.58 catalyze the general reaction,



Triacylglycerides often are utilized as carbon (e.g., fatty acid or lipid) storage molecules. Without being limited by any theory, it is believed that reducing the activity of acyltransferase may reduce the conversion of diacylglycerol to triacylglycerol, which may cause increased accumulation of fatty acid, in conjunction with additional genetic modifications (e.g., lipase to further remove fatty acids from the glycerol backbone) that can be converted into a fatty dicarboxylic acid (e.g., octanedioic acid, decanedioic acid, dodecanedioic acid, tetradecanedioic acid, hexadecanedioic acid, octadecanedioic acid, eicosanedioic acid) by other engineered pathways in the same host organism (e.g., omega oxidation pathway, beta oxidation pathway, omega oxidation pathway and beta oxidation pathway). Acyltransferases can be inactivated by any suitable means. Described herein are gene knockout methods suitable for disrupting nucleotide sequences that encode

polypeptides having DGA1 activity, LRO1 activity or DGA1 activity and LRO1 activity. The nucleotide sequence of DGA1 is provided in Example 33, SEQ ID NO: 47. The nucleotide sequence of LRO1 is provided in Example 33, SEQ ID NO: 49. DNA vectors suitable for use in constructing "knockout" constructs are described herein.

5

The presence, absence or amount of acyltransferase activity can be detected by any suitable method known in the art. Non-limiting examples of suitable detection methods include enzymatic assays (e.g., Geelen, Analytical Biochemistry 322(2):264-268 (2003), Dahlqvist et al, PNAS 97(12):6487-6492 (2000)), binding assays, PCR based assays (e.g., qPCR, RTPCR),

10 immunological detection methods (e.g., antibodies specific for a DGA1 or LRO1 acyltransferase), the like and combinations thereof.

Polynucleotides and Polypeptides

15 A nucleic acid (e.g., also referred to herein as nucleic acid reagent, target nucleic acid, target nucleotide sequence, nucleic acid sequence of interest or nucleic acid region of interest) can be from any source or composition, such as DNA, cDNA, gDNA (genomic DNA), RNA, siRNA (short inhibitory RNA), RNAi, tRNA or mRNA, for example, and can be in any form (e.g., linear, circular, supercoiled, single-stranded, double-stranded, and the like). A nucleic acid can also comprise

20 DNA or RNA analogs (e.g., containing base analogs, sugar analogs and/or a non-native backbone and the like). It is understood that the term "nucleic acid" does not refer to or infer a specific length of the polynucleotide chain, thus polynucleotides and oligonucleotides are also included in the definition. Deoxyribonucleotides include deoxyadenosine, deoxycytidine, deoxyguanosine and deoxythymidine. For RNA, the uracil base is uridine.

25

A nucleic acid sometimes is a plasmid, phage, autonomously replicating sequence (ARS), centromere, artificial chromosome, yeast artificial chromosome (e.g., YAC) or other nucleic acid able to replicate or be replicated in a host cell. In certain embodiments a nucleic acid can be from a library or can be obtained from enzymatically digested, sheared or sonicated genomic DNA (e.g.,

30 fragmented) from an organism of interest. In some embodiments, nucleic acid subjected to fragmentation or cleavage may have a nominal, average or mean length of about 5 to about 10,000 base pairs, about 100 to about 1,000 base pairs, about 100 to about 500 base pairs, or about 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000 or 10000 base pairs.

Fragments can be generated by any suitable method in the art, and the average, mean or nominal length of nucleic acid fragments can be controlled by selecting an appropriate fragment-generating procedure by the person of ordinary skill. In some embodiments, the fragmented DNA can be size selected to obtain nucleic acid fragments of a particular size range.

- 5 Nucleic acid can be fragmented by various methods known to the person of ordinary skill, which include without limitation, physical, chemical and enzymic processes. Examples of such processes are described in U.S. Patent Application Publication No. 20050112590 (published on May 26, 2005, entitled "Fragmentation-based methods and systems for sequence variation detection and discovery," naming Van Den Boom et al.). Certain processes can be selected by the person of
- 10 ordinary skill to generate non-specifically cleaved fragments or specifically cleaved fragments. Examples of processes that can generate non-specifically cleaved fragment sample nucleic acid include, without limitation, contacting sample nucleic acid with apparatus that expose nucleic acid to shearing force (e.g., passing nucleic acid through a syringe needle; use of a French press); exposing sample nucleic acid to irradiation (e.g., gamma, x-ray, UV irradiation; fragment sizes can
- 15 be controlled by irradiation intensity); boiling nucleic acid in water (e.g., yields about 500 base pair fragments) and exposing nucleic acid to an acid and base hydrolysis process.

- Nucleic acid may be specifically cleaved by contacting the nucleic acid with one or more specific cleavage agents. The term "specific cleavage agent" as used herein refers to an agent, sometimes
- 20 a chemical or an enzyme that can cleave a nucleic acid at one or more specific sites. Specific cleavage agents often will cleave specifically according to a particular nucleotide sequence at a particular site. Examples of enzymic specific cleavage agents include without limitation endonucleases (e.g., DNase (e.g., DNase I, II); RNase (e.g., RNase E, F, H, P); Cleavase™ enzyme; Taq DNA polymerase; E. coli DNA polymerase I and eukaryotic structure-specific
- 25 endonucleases; murine FEN-1 endonucleases; type I, II or III restriction endonucleases such as Acc I, Afl III, Alu I, Alw44 I, Apa I, Asn I, Ava I, Ava II, BamH I, Ban II, Bcl I, Bgl I, Bgl II, Bln I, Bsm I, BssH II, BstE II, Cfo I, Cla I, Dde I, Dpn I, Dra I, EclX I, EcoR I, EcoR I, EcoR II, EcoR V, Hae II, Hae II, Hind II, Hind III, Hpa I, Hpa II, Kpn I, Ksp I, Mlu I, MluN I, Msp I, Nci I, Nco I, Nde I, Nde II, Nhe I, Not I, Nru I, Nsi I, Pst I, Pvu I, Pvu II, Rsa I, Sac I, Sal I, Sau3A I, Sca I, ScrF I, Sfi I, Sma I,
- 30 Spe I, Sph I, Ssp I, Stu I, Sty I, Swa I, Taq I, Xba I, Xho I); glycosylases (e.g., uracil-DNA glycosylase (UDG), 3-methyladenine DNA glycosylase, 3-methyladenine DNA glycosylase II, pyrimidine hydrate-DNA glycosylase, FaPy-DNA glycosylase, thymine mismatch-DNA glycosylase, hypoxanthine-DNA glycosylase, 5-Hydroxymethyluracil DNA glycosylase (HmUDG), 5-Hydroxymethylcytosine DNA glycosylase, or 1,N6-etheno-adenine DNA glycosylase);

exonucleases (e.g., exonuclease III); ribozymes, and DNazymes. Sample nucleic acid may be treated with a chemical agent, or synthesized using modified nucleotides, and the modified nucleic acid may be cleaved. In non-limiting examples, sample nucleic acid may be treated with (i) alkylating agents such as methylnitrosourea that generate several alkylated bases, including N3-methyladenine and N3-methylguanine, which are recognized and cleaved by alkyl purine DNA-glycosylase; (ii) sodium bisulfite, which causes deamination of cytosine residues in DNA to form uracil residues that can be cleaved by uracil N-glycosylase; and (iii) a chemical agent that converts guanine to its oxidized form, 8-hydroxyguanine, which can be cleaved by formamidopyrimidine DNA N-glycosylase. Examples of chemical cleavage processes include without limitation alkylation, (e.g., alkylation of phosphorothioate-modified nucleic acid); cleavage of acid lability of P3'-N5'-phosphoroamidate-containing nucleic acid; and osmium tetroxide and piperidine treatment of nucleic acid.

As used herein, the term "complementary cleavage reactions" refers to cleavage reactions that are carried out on the same nucleic acid using different cleavage reagents or by altering the cleavage specificity of the same cleavage reagent such that alternate cleavage patterns of the same target or reference nucleic acid or protein are generated. In certain embodiments, nucleic acids of interest may be treated with one or more specific cleavage agents (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or more specific cleavage agents) in one or more reaction vessels (e.g., nucleic acid of interest is treated with each specific cleavage agent in a separate vessel).

A nucleic acid suitable for use in the embodiments described herein sometimes is amplified by any amplification process known in the art (e.g., PCR, RT-PCR and the like). Nucleic acid amplification may be particularly beneficial when using organisms that are typically difficult to culture (e.g., slow growing, require specialized culture conditions and the like). The terms "amplify", "amplification", "amplification reaction", or "amplifying" as used herein refer to any in vitro processes for multiplying the copies of a target sequence of nucleic acid. Amplification sometimes refers to an "exponential" increase in target nucleic acid. However, "amplifying" as used herein can also refer to linear increases in the numbers of a select target sequence of nucleic acid, but is different than a one-time, single primer extension step. In some embodiments, a limited amplification reaction, also known as pre-amplification, can be performed. Pre-amplification is a method in which a limited amount of amplification occurs due to a small number of cycles, for example 10 cycles, being performed. Pre-amplification can allow some amplification, but stops amplification prior to the exponential phase, and typically produces about 500 copies of the desired nucleotide sequence(s).

Use of pre-amplification may also limit inaccuracies associated with depleted reactants in standard PCR reactions.

5 In some embodiments, a nucleic acid reagent sometimes is stably integrated into the chromosome of the host organism, or a nucleic acid reagent can be a deletion of a portion of the host chromosome, in certain embodiments (e.g., genetically modified organisms, where alteration of the host genome confers the ability to selectively or preferentially maintain the desired organism carrying the genetic modification). Such nucleic acid reagents (e.g., nucleic acids or genetically modified organisms whose altered genome confers a selectable trait to the organism) can be
10 selected for their ability to guide production of a desired protein or nucleic acid molecule. When desired, the nucleic acid reagent can be altered such that codons encode for (i) the same amino acid, using a different tRNA than that specified in the native sequence, or (ii) a different amino acid than is normal, including unconventional or unnatural amino acids (including detectably labeled amino acids). As described herein, the term "native sequence" refers to an unmodified nucleotide
15 sequence as found in its natural setting (e.g., a nucleotide sequence as found in an organism).

A nucleic acid or nucleic acid reagent can comprise certain elements often selected according to the intended use of the nucleic acid. Any of the following elements can be included in or excluded from a nucleic acid reagent. A nucleic acid reagent, for example, may include one or more or all of
20 the following nucleotide elements: one or more promoter elements, one or more 5' untranslated regions (5'UTRs), one or more regions into which a target nucleotide sequence may be inserted (an "insertion element"), one or more target nucleotide sequences, one or more 3' untranslated regions (3'UTRs), and one or more selection elements. A nucleic acid reagent can be provided with one or more of such elements and other elements may be inserted into the nucleic acid before
25 the nucleic acid is introduced into the desired organism. In some embodiments, a provided nucleic acid reagent comprises a promoter, 5'UTR, optional 3'UTR and insertion element(s) by which a target nucleotide sequence is inserted (i.e., cloned) into the nucleotide acid reagent. In certain embodiments, a provided nucleic acid reagent comprises a promoter, insertion element(s) and optional 3'UTR, and a 5' UTR/target nucleotide sequence is inserted with an optional 3'UTR. The
30 elements can be arranged in any order suitable for expression in the chosen expression system (e.g., expression in a chosen organism, or expression in a cell free system, for example), and in some embodiments a nucleic acid reagent comprises the following elements in the 5' to 3' direction: (1) promoter element, 5'UTR, and insertion element(s); (2) promoter element, 5'UTR,

and target nucleotide sequence; (3) promoter element, 5'UTR, insertion element(s) and 3'UTR; and (4) promoter element, 5'UTR, target nucleotide sequence and 3'UTR.

Promoters

5

A promoter element typically is required for DNA synthesis and/or RNA synthesis. A promoter element often comprises a region of DNA that can facilitate the transcription of a particular gene, by providing a start site for the synthesis of RNA corresponding to a gene. Promoters generally are located near the genes they regulate, are located upstream of the gene (e.g., 5' of the gene), and are on the same strand of DNA as the sense strand of the gene, in some embodiments. In some embodiments, a promoter element can be isolated from a gene or organism and inserted in functional connection with a polynucleotide sequence to allow altered and/or regulated expression. A non-native promoter (e.g., promoter not normally associated with a given nucleic acid sequence) used for expression of a nucleic acid often is referred to as a heterologous promoter. In certain embodiments, a heterologous promoter and/or a 5'UTR can be inserted in functional connection with a polynucleotide that encodes a polypeptide having a desired activity as described herein. The terms "operably linked" and "in functional connection with" as used herein with respect to promoters, refer to a relationship between a coding sequence and a promoter element. The promoter is operably linked or in functional connection with the coding sequence when expression from the coding sequence via transcription is regulated, or controlled by, the promoter element. The terms "operably linked" and "in functional connection with" are utilized interchangeably herein with respect to promoter elements.

A promoter often interacts with a RNA polymerase. A polymerase is an enzyme that catalyses synthesis of nucleic acids using a preexisting nucleic acid reagent. When the template is a DNA template, an RNA molecule is transcribed before protein is synthesized. Enzymes having polymerase activity suitable for use in the present methods include any polymerase that is active in the chosen system with the chosen template to synthesize protein. In some embodiments, a promoter (e.g., a heterologous promoter) also referred to herein as a promoter element, can be operably linked to a nucleotide sequence or an open reading frame (ORF). Transcription from the promoter element can catalyze the synthesis of an RNA corresponding to the nucleotide sequence or ORF sequence operably linked to the promoter, which in turn leads to synthesis of a desired peptide, polypeptide or protein.

Promoter elements sometimes exhibit responsiveness to regulatory control. Promoter elements also sometimes can be regulated by a selective agent. That is, transcription from promoter elements sometimes can be turned on, turned off, up-regulated or down-regulated, in response to a change in environmental, nutritional or internal conditions or signals (e.g., heat inducible promoters, light regulated promoters, feedback regulated promoters, hormone influenced promoters, tissue specific promoters, oxygen and pH influenced promoters, promoters that are responsive to selective agents (e.g., kanamycin) and the like, for example). Promoters influenced by environmental, nutritional or internal signals frequently are influenced by a signal (direct or indirect) that binds at or near the promoter and increases or decreases expression of the target sequence under certain conditions.

Non-limiting examples of selective or regulatory agents that can influence transcription from a promoter element used in embodiments described herein include, without limitation, (1) nucleic acid segments that encode products that provide resistance against otherwise toxic compounds (e.g., antibiotics); (2) nucleic acid segments that encode products that are otherwise lacking in the recipient cell (e.g., essential products, tRNA genes, auxotrophic markers); (3) nucleic acid segments that encode products that suppress the activity of a gene product; (4) nucleic acid segments that encode products that can be readily identified (e.g., phenotypic markers such as antibiotics (e.g., β -lactamase), β -galactosidase, green fluorescent protein (GFP), yellow fluorescent protein (YFP), red fluorescent protein (RFP), cyan fluorescent protein (CFP), and cell surface proteins); (5) nucleic acid segments that bind products that are otherwise detrimental to cell survival and/or function; (6) nucleic acid segments that otherwise inhibit the activity of any of the nucleic acid segments described in Nos. 1-5 above (e.g., antisense oligonucleotides); (7) nucleic acid segments that bind products that modify a substrate (e.g., restriction endonucleases); (8) nucleic acid segments that can be used to isolate or identify a desired molecule (e.g., specific protein binding sites); (9) nucleic acid segments that encode a specific nucleotide sequence that can be otherwise non-functional (e.g., for PCR amplification of subpopulations of molecules); (10) nucleic acid segments that, when absent, directly or indirectly confer resistance or sensitivity to particular compounds; (11) nucleic acid segments that encode products that either are toxic or convert a relatively non-toxic compound to a toxic compound (e.g., Herpes simplex thymidine kinase, cytosine deaminase) in recipient cells; (12) nucleic acid segments that inhibit replication, partition or heritability of nucleic acid molecules that contain them; and/or (13) nucleic acid segments that encode conditional replication functions, e.g., replication in certain hosts or host cell strains or under certain environmental conditions (e.g., temperature, nutritional conditions, and the

like). In some embodiments, the regulatory or selective agent can be added to change the existing growth conditions to which the organism is subjected (e.g., growth in liquid culture, growth in a fermentor, growth on solid nutrient plates and the like for example).

- 5 In some embodiments, regulation of a promoter element can be used to alter (e.g., increase, add, decrease or substantially eliminate) the activity of a peptide, polypeptide or protein (e.g., enzyme activity for example). For example, a microorganism can be engineered by genetic modification to express a nucleic acid reagent that can add a novel activity (e.g., an activity not normally found in the host organism) or increase the expression of an existing activity by increasing transcription
- 10 from a homologous or heterologous promoter operably linked to a nucleotide sequence of interest (e.g., homologous or heterologous nucleotide sequence of interest), in certain embodiments. In some embodiments, a microorganism can be engineered by genetic modification to express a nucleic acid reagent that can decrease expression of an activity by decreasing or substantially eliminating transcription from a homologous or heterologous promoter operably linked to a
- 15 nucleotide sequence of interest, in certain embodiments.

In some embodiments the activity can be altered using recombinant DNA and genetic techniques known to the artisan. Methods for engineering microorganisms are further described herein. Tables herein provide non-limiting lists of yeast promoters that are up-regulated by oxygen, yeast

20 promoters that are down-regulated by oxygen, yeast transcriptional repressors and their associated genes, DNA binding motifs as determined using the MEME sequence analysis software. Potential regulator binding motifs can be identified using the program MEME to search intergenic regions bound by regulators for overrepresented sequences. For each regulator, the sequences of intergenic regions bound with p-values less than 0.001 were extracted to use as

25 input for motif discovery. The MEME software was run using the following settings: a motif width ranging from 6 to 18 bases, the “zoops” distribution model, a 6th order Markov background model and a discovery limit of 20 motifs. The discovered sequence motifs were scored for significance by two criteria: an E-value calculated by MEME and a specificity score. The motif with the best score using each metric is shown for each regulator. All motifs presented are derived from datasets

30 generated in rich growth conditions with the exception of a previously published dataset for epitope-tagged Gal4 grown in galactose.

In some embodiments, the altered activity can be found by screening the organism under conditions that select for the desired change in activity. For example, certain microorganisms can

be adapted to increase or decrease an activity by selecting or screening the organism in question on a media containing substances that are poorly metabolized or even toxic. An increase in the ability of an organism to grow a substance that is normally poorly metabolized may result in an increase in the growth rate on that substance, for example. A decrease in the sensitivity to a toxic substance might be manifested by growth on higher concentrations of the toxic substance, for example. Genetic modifications that are identified in this manner sometimes are referred to as naturally occurring mutations or the organisms that carry them can sometimes be referred to as naturally occurring mutants. Modifications obtained in this manner are not limited to alterations in promoter sequences. That is, screening microorganisms by selective pressure, as described above, can yield genetic alterations that can occur in non-promoter sequences, and sometimes also can occur in sequences that are not in the nucleotide sequence of interest, but in a related nucleotide sequences (e.g., a gene involved in a different step of the same pathway, a transport gene, and the like). Naturally occurring mutants sometimes can be found by isolating naturally occurring variants from unique environments, in some embodiments.

Homology and Identity

In addition to the regulated promoter sequences, regulatory sequences, and coding polynucleotides provided herein, a nucleic acid reagent may include a polynucleotide sequence 80% or more identical to the foregoing (or to the complementary sequences). That is, a nucleotide sequence that is at least 80% or more, 81% or more, 82% or more, 83% or more, 84% or more, 85% or more, 86% or more, 87% or more, 88% or more, 89% or more, 90% or more, 91% or more, 92% or more, 93% or more, 94% or more, 95% or more, 96% or more, 97% or more, 98% or more, or 99% or more identical to a nucleotide sequence described herein can be utilized. The term "identical" as used herein refers to two or more nucleotide sequences having substantially the same nucleotide sequence when compared to each other. One test for determining whether two nucleotide sequences or amino acids sequences are substantially identical is to determine the percent of identical nucleotide sequences or amino acid sequences shared.

Calculations of sequence identity can be performed as follows. 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). The length of a reference sequence aligned for comparison purposes is sometimes 30% or more, 40% or more, 50% or more, often 60% or more,

- and more often 70% or more, 80% or more, 90% or more, or 100% of the length of the reference sequence. The nucleotides or amino acids at corresponding nucleotide or polypeptide positions, respectively, are then compared among the two sequences. When a position in the first sequence is occupied by the same nucleotide or amino acid as the corresponding position in the second
5 sequence, the nucleotides or amino acids are deemed to be identical at that position. 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, introduced for optimal alignment of the two sequences.
- 10 Comparison of sequences and determination of percent identity between two sequences can be accomplished using a mathematical algorithm. Percent identity between two amino acid or nucleotide sequences can be determined using the algorithm of Meyers & Miller, *CABIOS* 4: 11-17 (1989), which has been incorporated into the ALIGN program (version 2.0), using a PAM120 weight residue table, a gap length penalty of 12 and a gap penalty of 4. Also, percent identity
15 between two amino acid sequences can be determined using the Needleman & Wunsch, *J. Mol. Biol.* 48: 444-453 (1970) algorithm which has been incorporated into the GAP program in the GCG software package (available at the [http](http://www.gcg.com) address www.gcg.com), 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. Percent identity between two nucleotide sequences can be determined using the GAP
20 program in the GCG software package (available at [http](http://www.gcg.com) address www.gcg.com), 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 set of parameters often used is a Blossum 62 scoring matrix with a gap open penalty of 12, a gap extend penalty of 4, and a frameshift gap penalty of 5.
- 25 Sequence identity can also be determined by hybridization assays conducted under stringent conditions. As use herein, the term "stringent conditions" refers to conditions for hybridization and washing. Stringent conditions are known to those skilled in the art and can be found in *Current Protocols in Molecular Biology*, John Wiley & Sons, N.Y., 6.3.1-6.3.6 (1989). Aqueous and non-aqueous methods are described in that reference and either can be used. An example of stringent
30 hybridization conditions is hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2X SSC, 0.1% SDS at 50°C. Another example of stringent hybridization conditions are hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2X SSC, 0.1% SDS at 55°C. A further example of stringent hybridization conditions is hybridization in 6X sodium chloride/sodium citrate (SSC) at

about 45°C, followed by one or more washes in 0.2X SSC, 0.1% SDS at 60°C. Often, stringent hybridization conditions are hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2X SSC, 0.1% SDS at 65°C. More often, stringency conditions are 0.5M sodium phosphate, 7% SDS at 65°C, followed by one or more washes at 0.2X
 5 SSC, 1% SDS at 65°C.

UTRs

As noted above, nucleic acid reagents may also comprise one or more 5' UTR's, and one or more
 10 3'UTR's. A 5' UTR may comprise one or more elements endogenous to the nucleotide sequence from which it originates, and sometimes includes one or more exogenous elements. A 5' UTR can originate from any suitable nucleic acid, such as genomic DNA, plasmid DNA, RNA or mRNA, for example, from any suitable organism (e.g., virus, bacterium, yeast, fungi, plant, insect or mammal). The artisan may select appropriate elements for the 5' UTR based upon the chosen expression
 15 system (e.g., expression in a chosen organism, or expression in a cell free system, for example). A 5' UTR sometimes comprises one or more of the following elements known to the artisan: enhancer sequences (e.g., transcriptional or translational), transcription initiation site, transcription factor binding site, translation regulation site, translation initiation site, translation factor binding site, accessory protein binding site, feedback regulation agent binding sites, Pribnow box, TATA
 20 box, -35 element, E-box (helix-loop-helix binding element), ribosome binding site, replicon, internal ribosome entry site (IRES), silencer element and the like. In some embodiments, a promoter element may be isolated such that all 5' UTR elements necessary for proper conditional regulation are contained in the promoter element fragment, or within a functional subsequence of a promoter element fragment.

25

A 5' UTR in the nucleic acid reagent can comprise a translational enhancer nucleotide sequence. A translational enhancer nucleotide sequence often is located between the promoter and the target nucleotide sequence in a nucleic acid reagent. A translational enhancer sequence often binds to a ribosome, sometimes is an 18S rRNA-binding ribonucleotide sequence (i.e., a 40S ribosome
 30 binding sequence) and sometimes is an internal ribosome entry sequence (IRES). An IRES generally forms an RNA scaffold with precisely placed RNA tertiary structures that contact a 40S ribosomal subunit via a number of specific intermolecular interactions. Examples of ribosomal enhancer sequences are known and can be identified by the artisan (e.g., Mignone et al., Nucleic Acids Research 33: D141-D146 (2005); Paulous et al., Nucleic Acids Research 31: 722-733

(2003); Akbergenov et al., *Nucleic Acids Research* 32: 239-247 (2004); Mignone et al., *Genome Biology* 3(3): reviews0004.1-0001.10 (2002); Gallie, *Nucleic Acids Research* 30: 3401-3411 (2002); Shaloiko et al., [http address www.interscience.wiley.com](http://www.interscience.wiley.com), DOI: 10.1002/bit.20267; and Gallie et al., *Nucleic Acids Research* 15: 3257-3273 (1987)).

5

A translational enhancer sequence sometimes is a eukaryotic sequence, such as a Kozak consensus sequence or other sequence (e.g., hydroid polyp sequence, GenBank accession no. U07128). A translational enhancer sequence sometimes is a prokaryotic sequence, such as a Shine-Dalgarno consensus sequence. In certain embodiments, the translational enhancer
 10 sequence is a viral nucleotide sequence. A translational enhancer sequence sometimes is from a 5' UTR of a plant virus, such as Tobacco Mosaic Virus (TMV), Alfalfa Mosaic Virus (AMV); Tobacco Etch Virus (ETV); Potato Virus Y (PVY); Turnip Mosaic (poty) Virus and Pea Seed Borne Mosaic Virus, for example. In certain embodiments, an omega sequence about 67 bases in length from TMV is included in the nucleic acid reagent as a translational enhancer sequence (e.g.,
 15 devoid of guanosine nucleotides and includes a 25 nucleotide long poly (CAA) central region).

A 3' UTR may comprise one or more elements endogenous to the nucleotide sequence from which it originates and sometimes includes one or more exogenous elements. A 3' UTR may originate from any suitable nucleic acid, such as genomic DNA, plasmid DNA, RNA or mRNA, for example,
 20 from any suitable organism (e.g., a virus, bacterium, yeast, fungi, plant, insect or mammal). The artisan can select appropriate elements for the 3' UTR based upon the chosen expression system (e.g., expression in a chosen organism, for example). A 3' UTR sometimes comprises one or more of the following elements known to the artisan: transcription regulation site, transcription initiation site, transcription termination site, transcription factor binding site, translation regulation site,
 25 translation termination site, translation initiation site, translation factor binding site, ribosome binding site, replicon, enhancer element, silencer element and polyadenosine tail. A 3' UTR often includes a polyadenosine tail and sometimes does not, and if a polyadenosine tail is present, one or more adenosine moieties may be added or deleted from it (e.g., about 5, about 10, about 15, about 20, about 25, about 30, about 35, about 40, about 45 or about 50 adenosine moieties may
 30 be added or subtracted).

In some embodiments, modification of a 5' UTR and/or a 3' UTR can be used to alter (e.g., increase, add, decrease or substantially eliminate) the activity of a promoter. Alteration of the promoter activity can in turn alter the activity of a peptide, polypeptide or protein (e.g., enzyme

activity for example), by a change in transcription of the nucleotide sequence(s) of interest from an operably linked promoter element comprising the modified 5' or 3' UTR. For example, a microorganism can be engineered by genetic modification to express a nucleic acid reagent comprising a modified 5' or 3' UTR that can add a novel activity (e.g., an activity not normally found in the host organism) or increase the expression of an existing activity by increasing transcription from a homologous or heterologous promoter operably linked to a nucleotide sequence of interest (e.g., homologous or heterologous nucleotide sequence of interest), in certain embodiments. In some embodiments, a microorganism can be engineered by genetic modification to express a nucleic acid reagent comprising a modified 5' or 3' UTR that can decrease the expression of an activity by decreasing or substantially eliminating transcription from a homologous or heterologous promoter operably linked to a nucleotide sequence of interest, in certain embodiments.

Target Nucleotide Sequence

A nucleotide reagent sometimes can comprise a target nucleotide sequence. A "target nucleotide sequence" as used herein encodes a nucleic acid, peptide, polypeptide or protein of interest, and may be a ribonucleotide sequence or a deoxyribonucleotide sequence. A target nucleic acid sometimes is an untranslated ribonucleic acid and sometimes is a translated ribonucleic acid. An untranslated ribonucleic acid may include, but is not limited to, a small interfering ribonucleic acid (siRNA), a short hairpin ribonucleic acid (shRNA), other ribonucleic acid capable of RNA interference (RNAi), an antisense ribonucleic acid, or a ribozyme. A translatable target nucleotide sequence (e.g., a target ribonucleotide sequence) sometimes encodes a peptide, polypeptide or protein, which are sometimes referred to herein as "target peptides," "target polypeptides" or "target proteins."

Any peptides, polypeptides or proteins, or an activity catalyzed by one or more peptides, polypeptides or proteins may be encoded by a target nucleotide sequence and may be selected by a user. Representative proteins include enzymes (e.g., acetyl-CoA carboxylase, acyl-CoA oxidase, thioesterase, monooxygenase, monooxygenase reductase, fatty alcohol oxidase, acyltransferase and the like, for example), antibodies, serum proteins (e.g., albumin), membrane bound proteins, hormones (e.g., growth hormone, erythropoietin, insulin, etc.), cytokines, etc., and include both naturally occurring and exogenously expressed polypeptides. Representative activities (e.g., enzymes or combinations of enzymes which are functionally associated to provide an activity) include thioesterase activity, monooxygenase activity, monooxygenase reductase activity,

acyltransferase activity, omega hydroxyl fatty acid dehydrogenase activity, beta-oxidation activity, omega-oxidation activity and the like, for example. The term “enzyme” as used herein refers to a protein which can act as a catalyst to induce a chemical change in other compounds, thereby producing one or more products from one or more substrates.

5

Specific polypeptides (e.g., enzymes) useful for embodiments described herein are listed herein. The term “protein” as used herein refers to a molecule having a sequence of amino acids linked by peptide bonds. This term includes fusion proteins, oligopeptides, peptides, cyclic peptides, polypeptides and polypeptide derivatives, whether native or recombinant, and also includes
10 fragments, derivatives, homologs, and variants thereof. A protein or polypeptide sometimes is of intracellular origin (e.g., located in the nucleus, cytosol, or interstitial space of host cells in vivo) and sometimes is a cell membrane protein in vivo. In some embodiments (described above, and in further detail hereafter in Engineering and Alteration Methods), a genetic modification can result in a modification (e.g., increase, substantially increase, decrease or substantially decrease) of a
15 target activity.

A translatable nucleotide sequence generally is located between a start codon (AUG in ribonucleic acids and ATG in deoxyribonucleic acids) and a stop codon (e.g., UAA (ochre), UAG (amber) or UGA (opal) in ribonucleic acids and TAA, TAG or TGA in deoxyribonucleic acids), and sometimes
20 is referred to herein as an “open reading frame” (ORF). A translatable nucleotide sequence (e.g., ORF) sometimes is encoded differently in one organism (e.g., most organisms encode CTG as leucine) than in another organism (e.g., *C. tropicalis* encodes CTG as serine). In some embodiments, a translatable nucleotide sequence is altered to correct alternate genetic code (e.g., codon usage) differences between a nucleotide donor organism and an nucleotide recipient
25 organism (e.g., engineered organism). In certain embodiments, a translatable nucleotide sequence is altered to improve; (i) codon usage, (ii) transcriptional efficiency, (iii) translational efficiency, (iv) the like, and combinations thereof.

Nucleic Acid Reagents & Tools

30

A nucleic acid reagent sometimes comprises one or more ORFs. An ORF may be from any suitable source, sometimes from genomic DNA, mRNA, reverse transcribed RNA or complementary DNA (cDNA) or a nucleic acid library comprising one or more of the foregoing, and is from any organism species that contains a nucleic acid sequence of interest, protein of interest,

or activity of interest. Non-limiting examples of organisms from which an ORF can be obtained include bacteria, yeast, fungi, human, insect, nematode, bovine, equine, canine, feline, rat or mouse, for example.

- 5 A nucleic acid reagent sometimes comprises a nucleotide sequence adjacent to an ORF that is translated in conjunction with the ORF and encodes an amino acid tag. The tag-encoding nucleotide sequence is located 3' and/or 5' of an ORF in the nucleic acid reagent, thereby encoding a tag at the C-terminus or N-terminus of the protein or peptide encoded by the ORF. Any tag that does not abrogate in vitro transcription and/or translation may be utilized and may be
10 appropriately selected by the artisan. Tags may facilitate isolation and/or purification of the desired ORF product from culture or fermentation media.

- A tag sometimes specifically binds a molecule or moiety of a solid phase or a detectable label, for example, thereby having utility for isolating, purifying and/or detecting a protein or peptide encoded
15 by the ORF. In some embodiments, a tag comprises one or more of the following elements: FLAG (e.g., DYKDDDDKG), V5 (e.g., GKIPNPLLGLDST), c-MYC (e.g., EQKLISEEDL), HSV (e.g., QPELAPEDPED), influenza hemagglutinin, HA (e.g., YPYDVPDYA), VSV-G (e.g., YTDIEMNRLGK), bacterial glutathione-S-transferase, maltose binding protein, a streptavidin- or avidin-binding tag (e.g., pcDNA™6 BioEase™ Gateway® Biotinylation System (Invitrogen)), thioredoxin, β -
20 galactosidase, VSV-glycoprotein, a fluorescent protein (e.g., green fluorescent protein or one of its many color variants (e.g., yellow, red, blue)), a polylysine or polyarginine sequence, a polyhistidine sequence (e.g., His6) or other sequence that chelates a metal (e.g., cobalt, zinc, copper), and/or a cysteine-rich sequence that binds to an arsenic-containing molecule. In certain embodiments, a cysteine-rich tag comprises the amino acid sequence CC-Xn-CC, wherein X is any amino acid and
25 n is 1 to 3, and the cysteine-rich sequence sometimes is CCPGCC. In certain embodiments, the tag comprises a cysteine-rich element and a polyhistidine element (e.g., CCPGCC and His6).

- A tag often conveniently binds to a binding partner. For example, some tags bind to an antibody (e.g., FLAG) and sometimes specifically bind to a small molecule. For example, a polyhistidine tag
30 specifically chelates a bivalent metal, such as copper, zinc and cobalt; a polylysine or polyarginine tag specifically binds to a zinc finger; a glutathione S-transferase tag binds to glutathione; and a cysteine-rich tag specifically binds to an arsenic-containing molecule. Arsenic-containing molecules include LUMIO™ agents (Invitrogen, California), such as FIAsH™ (EDT2[4',5'-bis(1,3,2-dithioarsolan-2-yl)fluorescein-(1,2-ethanedithiol)2]) and ReAsH reagents (e.g., U.S. Patent

5,932,474 to Tsien et al., entitled "Target Sequences for Synthetic Molecules;" U.S. Patent 6,054,271 to Tsien et al., entitled "Methods of Using Synthetic Molecules and Target Sequences;" U.S. Patents 6,451,569 and 6,008,378; published U.S. Patent Application 2003/0083373, and published PCT Patent Application WO 99/21013, all to Tsien et al. and all entitled "Synthetic
5 Molecules that Specifically React with Target Sequences"). Such antibodies and small molecules sometimes are linked to a solid phase for convenient isolation of the target protein or target peptide.

A tag sometimes comprises a sequence that localizes a translated protein or peptide to a
10 component in a system, which is referred to as a "signal sequence" or "localization signal sequence" herein. A signal sequence often is incorporated at the N-terminus of a target protein or target peptide, and sometimes is incorporated at the C-terminus. Examples of signal sequences are known to the artisan, are readily incorporated into a nucleic acid reagent, and often are selected according to the organism in which expression of the nucleic acid reagent is performed. A
15 signal sequence in some embodiments localizes a translated protein or peptide to a cell membrane. Examples of signal sequences include, but are not limited to, a nucleus targeting signal (e.g., steroid receptor sequence and N-terminal sequence of SV40 virus large T antigen); mitochondrial targeting signal (e.g., amino acid sequence that forms an amphipathic helix); peroxisome targeting signal (e.g., C-terminal sequence in YFG from *S.cerevisiae*); and a secretion
20 signal (e.g., N-terminal sequences from invertase, mating factor alpha, PHO5 and SUC2 in *S.cerevisiae*; multiple N-terminal sequences of *B. subtilis* proteins (e.g., Tjalsma et al., *Microbiol.Molec. Biol. Rev.* 64: 515-547 (2000)); alpha amylase signal sequence (e.g., U.S. Patent No. 6,288,302); pectate lyase signal sequence (e.g., U.S. Patent No. 5,846,818); precollagen signal sequence (e.g., U.S. Patent No. 5,712,114); OmpA signal sequence (e.g., U.S. Patent No.
25 5,470,719); lam beta signal sequence (e.g., U.S. Patent No. 5,389,529); *B. brevis* signal sequence (e.g., U.S. Patent No. 5,232,841); and *P. pastoris* signal sequence (e.g., U.S. Patent No. 5,268,273)).

A tag sometimes is directly adjacent to the amino acid sequence encoded by an ORF (i.e., there is
30 no intervening sequence) and sometimes a tag is substantially adjacent to an ORF encoded amino acid sequence (e.g., an intervening sequence is present). An intervening sequence sometimes includes a recognition site for a protease, which is useful for cleaving a tag from a target protein or peptide. In some embodiments, the intervening sequence is cleaved by Factor Xa (e.g., recognition site I (E/D)GR), thrombin (e.g., recognition site LVPRGS), enterokinase (e.g.,

recognition site DDDDK), TEV protease (e.g., recognition site ENLYFQG) or PreScission™ protease (e.g., recognition site LEVLFQGP), for example.

- 5 An intervening sequence sometimes is referred to herein as a “linker sequence,” and may be of any suitable length selected by the artisan. A linker sequence sometimes is about 1 to about 20 amino acids in length, and sometimes about 5 to about 10 amino acids in length. The artisan may select the linker length to substantially preserve target protein or peptide function (e.g., a tag may reduce target protein or peptide function unless separated by a linker), to enhance disassociation of a tag from a target protein or peptide when a protease cleavage site is present (e.g., cleavage
10 may be enhanced when a linker is present), and to enhance interaction of a tag/target protein product with a solid phase. A linker can be of any suitable amino acid content, and often comprises a higher proportion of amino acids having relatively short side chains (e.g., glycine, alanine, serine and threonine).
- 15 A nucleic acid reagent sometimes includes a stop codon between a tag element and an insertion element or ORF, which can be useful for translating an ORF with or without the tag. Mutant tRNA molecules that recognize stop codons (described above) suppress translation termination and thereby are designated “suppressor tRNAs.” Suppressor tRNAs can result in the insertion of amino acids and continuation of translation past stop codons (e.g., U.S. Patent Application No.
20 60/587,583, filed July 14, 2004, entitled “Production of Fusion Proteins by Cell-Free Protein Synthesis,”; Eggertsson, et al., (1988) Microbiological Review 52(3):354-374, and Engleberg-Kukla, et al. (1996) in *Escherichia coli* and *Salmonella* Cellular and Molecular Biology, Chapter 60, pps 909-921, Neidhardt, et al. eds., ASM Press, Washington, DC). A number of suppressor tRNAs are known, including but not limited to, supE, supP, supD, supF and supZ suppressors, which
25 suppress the termination of translation of the amber stop codon; supB, glT, supL, supN, supC and supM suppressors, which suppress the function of the ochre stop codon and glyT, trpT and Su-9 suppressors, which suppress the function of the opal stop codon. In general, suppressor tRNAs contain one or more mutations in the anti-codon loop of the tRNA that allows the tRNA to base pair with a codon that ordinarily functions as a stop codon. The mutant tRNA is charged with its
30 cognate amino acid residue and the cognate amino acid residue is inserted into the translating polypeptide when the stop codon is encountered. Mutations that enhance the efficiency of termination suppressors (i.e., increase stop codon read-through) have been identified. These include, but are not limited to, mutations in the uar gene (also known as the prfA gene), mutations

in the ups gene, mutations in the sueA, sueB and sueC genes, mutations in the rpsD (ramA) and rpsE (spcA) genes and mutations in the rplL gene.

Thus, a nucleic acid reagent comprising a stop codon located between an ORF and a tag can yield
5 a translated ORF alone when no suppressor tRNA is present in the translation system, and can
yield a translated ORF-tag fusion when a suppressor tRNA is present in the system. Suppressor
tRNA can be generated in cells transfected with a nucleic acid encoding the tRNA (e.g., a
replication incompetent adenovirus containing the human tRNA-Ser suppressor gene can be
transfected into cells, or a YAC containing a yeast or bacterial tRNA suppressor gene can be
10 transfected into yeast cells, for example). Vectors for synthesizing suppressor tRNA and for
translating ORFs with or without a tag are available to the artisan (e.g., Tag-On-Demand™ kit
(Invitrogen Corporation, California); Tag-On-Demand™ Suppressor Supernatant Instruction
Manual, Version B, 6 June 2003, at [http address www.invitrogen.com/content/sfs/
manuals/tagondemand_supernatant_man.pdf](http://www.invitrogen.com/content/sfs/manuals/tagondemand_supernatant_man.pdf); Tag-On-Demand™ Gateway® Vector Instruction
15 Manual, Version B, 20 June, 2003 at [http address www.invitrogen.com/content/sfs/
manuals/tagondemand_vectors_man.pdf](http://www.invitrogen.com/content/sfs/manuals/tagondemand_vectors_man.pdf); and Capone et al., Amber, ochre and opal suppressor
tRNA genes derived from a human serine tRNA gene. EMBO J. 4:213, 1985).

Any convenient cloning strategy known in the art may be utilized to incorporate an element, such
20 as an ORF, into a nucleic acid reagent. Known methods can be utilized to insert an element into
the template independent of an insertion element, such as (1) cleaving the template at one or more
existing restriction enzyme sites and ligating an element of interest and (2) adding restriction
enzyme sites to the template by hybridizing oligonucleotide primers that include one or more
suitable restriction enzyme sites and amplifying by polymerase chain reaction (described in greater
25 detail herein). Other cloning strategies take advantage of one or more insertion sites present or
inserted into the nucleic acid reagent, such as an oligonucleotide primer hybridization site for PCR,
for example, and others described herein. In some embodiments, a cloning strategy can be
combined with genetic manipulation such as recombination (e.g., recombination of a nucleic acid
reagent with a nucleic acid sequence of interest into the genome of the organism to be modified,
30 as described further herein). In some embodiments, the cloned ORF(s) can produce (directly or
indirectly) a fatty dicarboxylic acid (e.g., octanedioic acid, decanedioic acid, dodecanedioic acid,
tetradecanedioic acid, hexadecanedioic acid, octadecanedioic acid, eicosanedioic acid), by
engineering a microorganism with one or more ORFs of interest, which microorganism comprises
one or more altered activities selected from the group consisting of omega hydroxyl fatty acid

dehydrogenase activity, acyl-CoA oxidase activity, acyltransferase activity, thioesterase activity, monooxygenase activity and monooxygenase reductase activity.

In some embodiments, the nucleic acid reagent includes one or more recombinase insertion sites.

- 5 A recombinase insertion site is a recognition sequence on a nucleic acid molecule that participates in an integration/recombination reaction by recombination proteins. For example, the recombination site for Cre recombinase is loxP, which is a 34 base pair sequence comprised of two 13 base pair inverted repeats (serving as the recombinase binding sites) flanking an 8 base pair core sequence (e.g., Figure 1 of Sauer, B., Curr. Opin. Biotech. 5:521-527 (1994)). Other
- 10 examples of recombination sites include attB, attP, attL, and attR sequences, and mutants, fragments, variants and derivatives thereof, which are recognized by the recombination protein λ Int and by the auxiliary proteins integration host factor (IHF), FIS and excisionase (Xis) (e.g., U.S. Patent Nos. 5,888,732; 6,143,557; 6,171,861; 6,270,969; 6,277,608; and 6,720,140; U.S. Patent Appln. Nos. 09/517,466, filed March 2, 2000, and 09/732,914, filed August 14, 2003, and in U.S.
- 15 patent publication no. 2002-0007051-A1; Landy, Curr. Opin. Biotech. 3:699-707 (1993)).

- Examples of recombinase cloning nucleic acids are in Gateway® systems (Invitrogen, California), which include at least one recombination site for cloning a desired nucleic acid molecules in vivo or in vitro. In some embodiments, the system utilizes vectors that contain at least two different site-
- 20 specific recombination sites, often based on the bacteriophage lambda system (e.g., att1 and att2), and are mutated from the wild-type (att0) sites. Each mutated site has a unique specificity for its cognate partner att site (i.e., its binding partner recombination site) of the same type (for example attB1 with attP1, or attL1 with attR1) and will not cross-react with recombination sites of the other mutant type or with the wild-type att0 site. Different site specificities allow directional cloning or
- 25 linkage of desired molecules thus providing desired orientation of the cloned molecules. Nucleic acid fragments flanked by recombination sites are cloned and subcloned using the Gateway® system by replacing a selectable marker (for example, ccdB) flanked by att sites on the recipient plasmid molecule, sometimes termed the Destination Vector. Desired clones are then selected by transformation of a ccdB sensitive host strain and positive selection for a marker on the recipient
- 30 molecule. Similar strategies for negative selection (e.g., use of toxic genes) can be used in other organisms such as thymidine kinase (TK) in mammals and insects.

A recombination system useful for engineering yeast is outlined briefly. The system makes use of the URA3 gene (e.g., for *S. cerevisiae* and *C. albicans*, for example) or URA4 and URA5 genes

(e.g., for *S. pombe*, for example) and toxicity of the nucleotide analogue 5-Fluoroorotic acid (5-FOA). The URA3 or URA4 and URA5 genes encode orotidine-5'-monophosphate (OMP) dicarboxylase. Yeast with an active URA3 or URA4 and URA5 gene (phenotypically Ura⁺) convert 5-FOA to fluorodeoxyuridine, which is toxic to yeast cells. Yeast carrying a mutation in the
5 appropriate gene(s) or having a knock out of the appropriate gene(s) can grow in the presence of 5-FOA, if the media is also supplemented with uracil.

A nucleic acid engineering construct can be made which may comprise the URA3 gene or cassette (for *S. cerevisiae*), flanked on either side by the same nucleotide sequence in the same
10 orientation. The URA3 cassette comprises a promoter, the URA3 gene and a functional transcription terminator. Target sequences which direct the construct to a particular nucleic acid region of interest in the organism to be engineered are added such that the target sequences are adjacent to and abut the flanking sequences on either side of the URA3 cassette. Yeast can be transformed with the engineering construct and plated on minimal media without uracil. Colonies
15 can be screened by PCR to determine those transformants that have the engineering construct inserted in the proper location in the genome. Checking insertion location prior to selecting for recombination of the *ura3* cassette may reduce the number of incorrect clones carried through to later stages of the procedure. Correctly inserted transformants can then be replica plated on minimal media containing 5-FOA to select for recombination of the URA3 cassette out of the
20 construct, leaving a disrupted gene and an identifiable footprint (e.g., nucleic acid sequence) that can be used to verify the presence of the disrupted gene. The technique described is useful for disrupting or "knocking out" gene function, but also can be used to insert genes or constructs into a host organism's genome in a targeted, sequence specific manner.

25 In certain embodiments, a nucleic acid reagent includes one or more topoisomerase insertion sites. A topoisomerase insertion site is a defined nucleotide sequence recognized and bound by a site-specific topoisomerase. For example, the nucleotide sequence 5'-(C/T)CCTT-3' is a topoisomerase recognition site bound specifically by most poxvirus topoisomerases, including vaccinia virus DNA topoisomerase I. After binding to the recognition sequence, the topoisomerase
30 cleaves the strand at the 3'-most thymidine of the recognition site to produce a nucleotide sequence comprising 5'-(C/T)CCTT-PO₄-TOPO, a complex of the topoisomerase covalently bound to the 3' phosphate via a tyrosine in the topoisomerase (e.g., Shuman, J. Biol. Chem. 266:11372-11379, 1991; Sekiguchi and Shuman, Nucl. Acids Res. 22:5360-5365, 1994; U.S. Pat. No. 5,766,891; PCT/US95/16099; and PCT/US98/12372). In comparison, the nucleotide sequence 5'-

GCAACTT-3' is a topoisomerase recognition site for type IA E. coli topoisomerase III. An element to be inserted often is combined with topoisomerase-reacted template and thereby incorporated into the nucleic acid reagent (e.g., World Wide Web URL invitrogen.com/downloads/F-13512_Topo_Flyer.pdf; World Wide Web URL invitrogen.com/content/sfs/brochures/710_021849%20_B_TOPOCloning_bro.pdf; TOPO TA Cloning® Kit and Zero Blunt® TOPO® Cloning Kit product information).

A nucleic acid reagent sometimes contains one or more origin of replication (ORI) elements. In some embodiments, a template comprises two or more ORIs, where one functions efficiently in one organism (e.g., a bacterium) and another functions efficiently in another organism (e.g., a eukaryote, like yeast for example). In some embodiments, an ORI may function efficiently in one species (e.g., *S. cerevisiae*, for example) and another ORI may function efficiently in a different species (e.g., *S. pombe*, for example). A nucleic acid reagent also sometimes includes one or more transcription regulation sites.

A nucleic acid reagent can include one or more selection elements (e.g., elements for selection of the presence of the nucleic acid reagent, and not for activation of a promoter element which can be selectively regulated). Selection elements often are utilized using known processes to determine whether a nucleic acid reagent is included in a cell. In some embodiments, a nucleic acid reagent includes two or more selection elements, where one functions efficiently in one organism and another functions efficiently in another organism. Examples of selection elements include, but are not limited to, (1) nucleic acid segments that encode products that provide resistance against otherwise toxic compounds (e.g., antibiotics); (2) nucleic acid segments that encode products that are otherwise lacking in the recipient cell (e.g., essential products, tRNA genes, auxotrophic markers); (3) nucleic acid segments that encode products that suppress the activity of a gene product; (4) nucleic acid segments that encode products that can be readily identified (e.g., phenotypic markers such as antibiotics (e.g., β -lactamase), β -galactosidase, green fluorescent protein (GFP), yellow fluorescent protein (YFP), red fluorescent protein (RFP), cyan fluorescent protein (CFP), and cell surface proteins); (5) nucleic acid segments that bind products that are otherwise detrimental to cell survival and/or function; (6) nucleic acid segments that otherwise inhibit the activity of any of the nucleic acid segments described in Nos. 1-5 above (e.g., antisense oligonucleotides); (7) nucleic acid segments that bind products that modify a substrate (e.g., restriction endonucleases); (8) nucleic acid segments that can be used to isolate or identify a desired molecule (e.g., specific protein binding sites); (9) nucleic acid segments that encode a

specific nucleotide sequence that can be otherwise non-functional (e.g., for PCR amplification of subpopulations of molecules); (10) nucleic acid segments that, when absent, directly or indirectly confer resistance or sensitivity to particular compounds; (11) nucleic acid segments that encode products that either are toxic or convert a relatively non-toxic compound to a toxic compound (e.g.,
 5 Herpes simplex thymidine kinase, cytosine deaminase) in recipient cells; (12) nucleic acid segments that inhibit replication, partition or heritability of nucleic acid molecules that contain them; and/or (13) nucleic acid segments that encode conditional replication functions, e.g., replication in certain hosts or host cell strains or under certain environmental conditions (e.g., temperature, nutritional conditions, and the like).

10

A nucleic acid reagent is of any form useful for in vivo transcription and/or translation. A nucleic acid sometimes is a plasmid, such as a supercoiled plasmid, sometimes is a yeast artificial chromosome (e.g., YAC), sometimes is a linear nucleic acid (e.g., a linear nucleic acid produced by PCR or by restriction digest), sometimes is single-stranded and sometimes is double-stranded. A
 15 nucleic acid reagent sometimes is prepared by an amplification process, such as a polymerase chain reaction (PCR) process or transcription-mediated amplification process (TMA). In TMA, two enzymes are used in an isothermal reaction to produce amplification products detected by light emission (see, e.g., Biochemistry 1996 Jun 25;35(25):8429-38 and [http address www.devicelink.com/ivdt/archive/00/11/007.html](http://www.devicelink.com/ivdt/archive/00/11/007.html)). Standard PCR processes are known (e.g., U. S.
 20 Patent Nos. 4,683,202; 4,683,195; 4,965,188; and 5,656,493), and generally are performed in cycles. Each cycle includes heat denaturation, in which hybrid nucleic acids dissociate; cooling, in which primer oligonucleotides hybridize; and extension of the oligonucleotides by a polymerase (i.e., Taq polymerase). An example of a PCR cyclical process is treating the sample at 95 °C for 5 minutes; repeating forty-five cycles of 95 °C for 1 minute, 59 °C for 1 minute, 10 seconds, and 72 °C
 25 for 1 minute 30 seconds; and then treating the sample at 72 °C for 5 minutes. Multiple cycles frequently are performed using a commercially available thermal cycler. PCR amplification products sometimes are stored for a time at a lower temperature (e.g., at 4 °C) and sometimes are frozen (e.g., at -20 °C) before analysis.

30 In some embodiments, a nucleic acid reagent, protein reagent, protein fragment reagent or other reagent described herein is isolated or purified. The term "isolated" as used herein refers to material removed from its original environment (e.g., the natural environment if it is naturally occurring, or a host cell if expressed exogenously), and thus is altered "by the hand of man" from its original environment. The term "purified" as used herein with reference to molecules does not

refer to absolute purity. Rather, "purified" refers to a substance in a composition that contains fewer substance species in the same class (e.g., nucleic acid or protein species) other than the substance of interest in comparison to the sample from which it originated. "Purified," if a nucleic acid or protein for example, refers to a substance in a composition that contains fewer nucleic acid species or protein species other than the nucleic acid or protein of interest in comparison to the sample from which it originated. Sometimes, a protein or nucleic acid is "substantially pure," indicating that the protein or nucleic acid represents at least 50% of protein or nucleic acid on a mass basis of the composition. Often, a substantially pure protein or nucleic acid is at least 75% on a mass basis of the composition, and sometimes at least 95% on a mass basis of the composition.

Engineering and Alteration Methods

Methods and compositions (e.g., nucleic acid reagents) described herein can be used to generate engineered microorganisms. As noted above, the term "engineered microorganism" as used herein refers to a modified organism that includes one or more activities distinct from an activity present in a microorganism utilized as a starting point for modification (e.g., host microorganism or unmodified organism). Engineered microorganisms typically arise as a result of a genetic modification, usually introduced or selected for, by one of skill in the art using readily available techniques. Non-limiting examples of methods useful for generating an altered activity include, introducing a heterologous polynucleotide (e.g., nucleic acid or gene integration, also referred to as "knock in"), removing an endogenous polynucleotide, altering the sequence of an existing endogenous nucleic acid sequence (e.g., site-directed mutagenesis), disruption of an existing endogenous nucleic acid sequence (e.g., knock outs and transposon or insertion element mediated mutagenesis), selection for an altered activity where the selection causes a change in a naturally occurring activity that can be stably inherited (e.g., causes a change in a nucleic acid sequence in the genome of the organism or in an epigenetic nucleic acid that is replicated and passed on to daughter cells), PCR-based mutagenesis, and the like. The term "mutagenesis" as used herein refers to any modification to a nucleic acid (e.g., nucleic acid reagent, or host chromosome, for example) that is subsequently used to generate a product in a host or modified organism. Non-limiting examples of mutagenesis include, deletion, insertion, substitution, rearrangement, point mutations, suppressor mutations and the like. Mutagenesis methods are known in the art and are readily available to the artisan. Non-limiting examples of mutagenesis methods are described herein and can also be found in Maniatis, T., E. F. Fritsch and J. Sambrook (1982) *Molecular Cloning: a Laboratory Manual*; Cold Spring Harbor Laboratory, Cold Spring

Harbor, N.Y. Another non-limiting example of mutagenesis can be conducted using a Stratagene (San Diego, CA) "QuickChange" kit according to the manufacturer's instructions.

5 The term "genetic modification" as used herein refers to any suitable nucleic acid addition, removal or alteration that facilitates production of a target fatty dicarboxylic acid product (e.g., sebacic or dodecanedioic acid) in an engineered microorganism. Genetic modifications include, without limitation, insertion of one or more nucleotides in a native nucleic acid of a host organism in one or more locations, deletion of one or more nucleotides in a native nucleic acid of a host organism in one or more locations, modification or substitution of one or more nucleotides in a native nucleic
10 acid of a host organism in one or more locations, insertion of a non-native nucleic acid into a host organism (e.g., insertion of an autonomously replicating vector), and removal of a non-native nucleic acid in a host organism (e.g., removal of a vector).

15 The term "heterologous polynucleotide" as used herein refers to a nucleotide sequence not present in a host microorganism in some embodiments. In certain embodiments, a heterologous polynucleotide is present in a different amount (e.g., different copy number) than in a host microorganism, which can be accomplished, for example, by introducing more copies of a particular nucleotide sequence to a host microorganism (e.g., the particular nucleotide sequence may be in a nucleic acid autonomous of the host chromosome or may be inserted into a
20 chromosome). A heterologous polynucleotide is from a different organism in some embodiments, and in certain embodiments, is from the same type of organism but from an outside source (e.g., a recombinant source).

In some embodiments, an organism engineered using the methods and nucleic acid reagents
25 described herein can produce a fatty dicarboxylic acid (e.g., octanedioic acid, decanedioic acid, dodecanedioic acid, tetradecanedioic acid, hexadecanedioic acid, octadecanedioic acid, eicosanedioic acid). In certain embodiments, an engineered microorganism described herein that produces a fatty dicarboxylic acid (e.g., octanedioic acid, decanedioic acid, dodecanedioic acid, tetradecanedioic acid, hexadecanedioic acid, octadecanedioic acid, eicosanedioic acid) may
30 comprise one or more altered activities selected from the group consisting of omega oxo fatty acid dehydrogenase activity, omega hydroxyl fatty acid dehydrogenase activity, fatty acid synthase activity, acetyl CoA carboxylase activity, acyl-CoA oxidase activity, monooxygenase activity and monooxygenase reductase activity. In some embodiments, an engineered microorganism as described herein may comprise a genetic modification that adds or increases the omega oxo fatty

acid dehydrogenase activity, omega hydroxyl fatty acid dehydrogenase activity, fatty acid synthase activity, acetyl CoA carboxylase activity, acyl-CoA oxidase activity, monooxygenase activity and monooxygenase reductase activity.

- 5 In certain embodiments, an engineered microorganism described herein can comprise an altered thioesterase activity. In some embodiments, the engineered microorganism may comprise a genetic alteration that adds or increases a thioesterase activity. In some embodiments, the engineered microorganism comprising a genetic alteration that adds or increases a thioesterase activity, may further comprise a heterologous polynucleotide encoding a polypeptide having
- 10 thioesterase activity.

The term "altered activity" as used herein refers to an activity in an engineered microorganism that is added or modified relative to the host microorganism (e.g., added, increased, reduced, inhibited or removed activity). An activity can be altered by introducing a genetic modification to a host

15 microorganism that yields an engineered microorganism having added, increased, reduced, inhibited or removed activity.

An added activity often is an activity not detectable in a host microorganism. An increased activity generally is an activity detectable in a host microorganism that has been increased in an

20 engineered microorganism. An activity can be increased to any suitable level for production of a target fatty dicarboxylic acid product (e.g., sebacic or dodecanedioic acid), including but not limited to less than 2-fold (e.g., about 10% increase to about 99% increase; about 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90% increase), 2-fold, 3-fold, 4-fold, 5-fold, 6-fold, 7-fold, 8-fold, 9-fold, or 10-fold increase, or greater than about 10-fold increase. A reduced or inhibited activity generally is an

25 activity detectable in a host microorganism that has been reduced or inhibited in an engineered microorganism. An activity can be reduced to undetectable levels in some embodiments, or detectable levels in certain embodiments. An activity can be decreased to any suitable level for production of a target fatty dicarboxylic acid product (e.g., sebacic or dodecanedioic acid), including but not limited to less than 2-fold (e.g., about 10% decrease to about 99% decrease;

30 about 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90% decrease), 2-fold, 3-fold, 4-fold, 5-fold, 6-fold, 7-fold, 8-fold, 9-fold, or 10-fold decrease, or greater than about 10-fold decrease.

An altered activity sometimes is an activity not detectable in a host organism and is added to an engineered organism. An altered activity also may be an activity detectable in a host organism and

is increased in an engineered organism. An activity may be added or increased by increasing the number of copies of a polynucleotide that encodes a polypeptide having a target activity, in some embodiments. In some embodiments, the activity of a native polypeptide can be increased by increasing in the modified organism the number of copies of a polynucleotide that encodes the polypeptide (e.g., introducing 1 to about 100 additional copies of the polynucleotide (e.g., introducing 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 22, 24, 26, 28, 30 or more additional copies of the polynucleotide). In certain embodiments an activity can be added or increased by inserting into a host microorganism a polynucleotide that encodes a heterologous polypeptide having the added activity or encodes a modified endogenous polypeptide. In such embodiments, 1 to about 100 copies of the polynucleotide can be introduced (e.g., introducing 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 22, 24, 26, 28, 30 copies). A "modified endogenous polypeptide" often has an activity different than an activity of a native polypeptide counterpart (e.g., different catalytic activity and/or different substrate specificity), and often is active (e.g., an activity (e.g., substrate turnover) is detectable). In certain embodiments, an activity can be added or increased by inserting into a host microorganism a heterologous polynucleotide that is (i) operably linked to another polynucleotide that encodes a polypeptide having the added activity, and (ii) up regulates production of the polynucleotide. Thus, an activity can be added or increased by inserting or modifying a regulatory polynucleotide operably linked to another polynucleotide that encodes a polypeptide having the target activity. In certain embodiments, an activity can be added or increased by subjecting a host microorganism to a selective environment and screening for microorganisms that have a detectable level of the target activity. Examples of a selective environment include, without limitation, a medium containing a substrate that a host organism can process and a medium lacking a substrate that a host organism can process.

25 An altered activity sometimes is an activity detectable in a host organism and is reduced, inhibited or removed (i.e., not detectable) in an engineered organism. An activity may be reduced or removed by decreasing the number of copies of a polynucleotide that encodes a polypeptide having a target activity, in some embodiments. In some embodiments, an activity can be reduced or removed by (i) inserting a polynucleotide within a polynucleotide that encodes a polypeptide having the target activity (disruptive insertion), and/or (ii) removing a portion of or all of a polynucleotide that encodes a polypeptide having the target activity (deletion or knock out, respectively). In certain embodiments, an activity can be reduced or removed by inserting into a host microorganism a heterologous polynucleotide that is (i) operably linked to another polynucleotide that encodes a polypeptide having the target activity, and (ii) down regulates

production of the polynucleotide. Thus, an activity can be reduced or removed by inserting or modifying a regulatory polynucleotide operably linked to another polynucleotide that encodes a polypeptide having the target activity.

- 5 An activity also can be reduced or removed by (i) inhibiting a polynucleotide that encodes a polypeptide having the activity or (ii) inhibiting a polynucleotide operably linked to another polynucleotide that encodes a polypeptide having the activity. A polynucleotide can be inhibited by a suitable technique known in the art, such as by contacting an RNA encoded by the polynucleotide with a specific inhibitory RNA (e.g., RNAi, siRNA, ribozyme). An activity also can be
- 10 reduced or removed by contacting a polypeptide having the activity with a molecule that specifically inhibits the activity (e.g., enzyme inhibitor, antibody). In certain embodiments, an activity can be reduced or removed by subjecting a host microorganism to a selective environment and screening for microorganisms that have a reduced level or removal of the target activity.
- 15 In some embodiments, an untranslated ribonucleic acid, or a cDNA can be used to reduce the expression of a particular activity or enzyme. For example, a microorganism can be engineered by genetic modification to express a nucleic acid reagent that reduces the expression of an activity by producing an RNA molecule that is partially or substantially homologous to a nucleic acid sequence of interest which encodes the activity of interest. The RNA molecule can bind to the
- 20 nucleic acid sequence of interest and inhibit the nucleic acid sequence from performing its natural function, in certain embodiments. In some embodiments, the RNA may alter the nucleic acid sequence of interest which encodes the activity of interest in a manner that the nucleic acid sequence of interest is no longer capable of performing its natural function (e.g., the action of a ribozyme for example).
- 25 In certain embodiments, nucleotide sequences sometimes are added to, modified or removed from one or more of the nucleic acid reagent elements, such as the promoter, 5'UTR, target sequence, or 3'UTR elements, to enhance, potentially enhance, reduce, or potentially reduce transcription and/or translation before or after such elements are incorporated in a nucleic acid reagent. In
- 30 some embodiments, one or more of the following sequences may be modified or removed if they are present in a 5'UTR: a sequence that forms a stable secondary structure (e.g., quadruplex structure or stem loop stem structure (e.g., EMBL sequences X12949, AF274954, AF139980, AF152961, S95936, U194144, AF116649 or substantially identical sequences that form such stem loop stem structures)); a translation initiation codon upstream of the target nucleotide sequence

start codon; a stop codon upstream of the target nucleotide sequence translation initiation codon; an ORF upstream of the target nucleotide sequence translation initiation codon; an iron responsive element (IRE) or like sequence; and a 5' terminal oligopyrimidine tract (TOP, e.g., consisting of 5-15 pyrimidines adjacent to the cap). A translational enhancer sequence and/or an internal
 5 ribosome entry site (IRES) sometimes is inserted into a 5'UTR (e.g., EMBL nucleotide sequences J04513, X87949, M95825, M12783, AF025841, AF013263, AF006822, M17169, M13440, M22427, D14838 and M17446 and substantially identical nucleotide sequences).

An AU-rich element (ARE, e.g., AUUUA repeats) and/or splicing junction that follows a non-sense
 10 codon sometimes is removed from or modified in a 3'UTR. A polyadenosine tail sometimes is inserted into a 3'UTR if none is present, sometimes is removed if it is present, and adenosine moieties sometimes are added to or removed from a polyadenosine tail present in a 3'UTR. Thus, some embodiments are directed to a process comprising: determining whether any nucleotide sequences that increase, potentially increase, reduce or potentially reduce translation efficiency
 15 are present in the elements, and adding, removing or modifying one or more of such sequences if they are identified. Certain embodiments are directed to a process comprising: determining whether any nucleotide sequences that increase or potentially increase translation efficiency are not present in the elements, and incorporating such sequences into the nucleic acid reagent.

20 In some embodiments, an activity can be altered by modifying the nucleotide sequence of an ORF. An ORF sometimes is mutated or modified (for example, by point mutation, deletion mutation, insertion mutation, PCR based mutagenesis and the like) to alter, enhance or increase, reduce, substantially reduce or eliminate the activity of the encoded protein or peptide. The protein or peptide encoded by a modified ORF sometimes is produced in a lower amount or may not be
 25 produced at detectable levels, and in other embodiments, the product or protein encoded by the modified ORF is produced at a higher level (e.g., codons sometimes are modified so they are compatible with tRNA's preferentially used in the host organism or engineered organism). To determine the relative activity, the activity from the product of the mutated ORF (or cell containing it) can be compared to the activity of the product or protein encoded by the unmodified ORF (or cell
 30 containing it).

In some embodiments, an ORF nucleotide sequence sometimes is mutated or modified to alter the triplet nucleotide sequences used to encode amino acids (e.g., amino acid codon triplets, for example). Modification of the nucleotide sequence of an ORF to alter codon triplets sometimes is

used to change the codon found in the original sequence to better match the preferred codon usage of the organism in which the ORF or nucleic acid reagent will be expressed. The codon usage, and therefore the codon triplets encoded by a nucleic acid sequence, in bacteria may be different from the preferred codon usage in eukaryotes, like yeast or plants for example. Preferred
5 codon usage also may be different between bacterial species. In certain embodiments an ORF nucleotide sequences sometimes is modified to eliminate codon pairs and/or eliminate mRNA secondary structures that can cause pauses during translation of the mRNA encoded by the ORF nucleotide sequence. Translational pausing sometimes occurs when nucleic acid secondary structures exist in an mRNA, and sometimes occurs due to the presence of codon pairs that slow
10 the rate of translation by causing ribosomes to pause. In some embodiments, the use of lower abundance codon triplets can reduce translational pausing due to a decrease in the pause time needed to load a charged tRNA into the ribosome translation machinery. Therefore, to increase transcriptional and translational efficiency in bacteria (e.g., where transcription and translation are concurrent, for example) or to increase translational efficiency in eukaryotes (e.g., where
15 transcription and translation are functionally separated), the nucleotide sequence of a nucleotide sequence of interest can be altered to better suit the transcription and/or translational machinery of the host and/or genetically modified microorganism. In certain embodiments, slowing the rate of translation by the use of lower abundance codons, which slow or pause the ribosome, can lead to higher yields of the desired product due to an increase in correctly folded proteins and a reduction
20 in the formation of inclusion bodies.

Codons can be altered and optimized according to the preferred usage by a given organism by determining the codon distribution of the nucleotide sequence donor organism and comparing the distribution of codons to the distribution of codons in the recipient or host organism. Techniques
25 described herein (e.g., site directed mutagenesis and the like) can then be used to alter the codons accordingly. Comparisons of codon usage can be done by hand, or using nucleic acid analysis software commercially available to the artisan.

Modification of the nucleotide sequence of an ORF also can be used to correct codon triplet
30 sequences that have diverged in different organisms. For example, certain yeast (e.g., *C. tropicalis* and *C. maltosa*) use the amino acid triplet CUG (e.g., CTG in the DNA sequence) to encode serine. CUG typically encodes leucine in most organisms. In order to maintain the correct amino acid in the resultant polypeptide or protein, the CUG codon must be altered to reflect the organism in which the nucleic acid reagent will be expressed. Thus, if an ORF from a bacterial

donor is to be expressed in either Candida yeast strain mentioned above, the heterologous nucleotide sequence must first be altered or modified to the appropriate leucine codon. Therefore, in some embodiments, the nucleotide sequence of an ORF sometimes is altered or modified to correct for differences that have occurred in the evolution of the amino acid codon triplets between
5 different organisms. In some embodiments, the nucleotide sequence can be left unchanged at a particular amino acid codon, if the amino acid encoded is a conservative or neutral change in amino acid when compared to the originally encoded amino acid.

In some embodiments, an activity can be altered by modifying translational regulation signals, like
10 a stop codon for example. A stop codon at the end of an ORF sometimes is modified to another stop codon, such as an amber stop codon described above. In some embodiments, a stop codon is introduced within an ORF, sometimes by insertion or mutation of an existing codon. An ORF comprising a modified terminal stop codon and/or internal stop codon often is translated in a system comprising a suppressor tRNA that recognizes the stop codon. An ORF comprising a stop
15 codon sometimes is translated in a system comprising a suppressor tRNA that incorporates an unnatural amino acid during translation of the target protein or target peptide. Methods for incorporating unnatural amino acids into a target protein or peptide are known, which include, for example, processes utilizing a heterologous tRNA/synthetase pair, where the tRNA recognizes an amber stop codon and is loaded with an unnatural amino acid (e.g., World Wide Web URL
20 iupac.org/news/prize/2003/wang.pdf).

Depending on the portion of a nucleic acid reagent (e.g., Promoter, 5' or 3' UTR, ORI, ORF, and the like) chosen for alteration (e.g., by mutagenesis, introduction or deletion, for example) the modifications described above can alter a given activity by (i) increasing or decreasing feedback
25 inhibition mechanisms, (ii) increasing or decreasing promoter initiation, (iii) increasing or decreasing translation initiation, (iv) increasing or decreasing translational efficiency, (v) modifying localization of peptides or products expressed from nucleic acid reagents described herein, or (vi) increasing or decreasing the copy number of a nucleotide sequence of interest, (vii) expression of an anti-sense RNA, RNAi, siRNA, ribozyme and the like. In some embodiments, alteration of a
30 nucleic acid reagent or nucleotide sequence can alter a region involved in feedback inhibition (e.g., 5' UTR, promoter and the like). A modification sometimes is made that can add or enhance binding of a feedback regulator and sometimes a modification is made that can reduce, inhibit or eliminate binding of a feedback regulator.

In certain embodiments, alteration of a nucleic acid reagent or nucleotide sequence can alter sequences involved in transcription initiation (e.g., promoters, 5' UTR, and the like). A modification sometimes can be made that can enhance or increase initiation from an endogenous or heterologous promoter element. A modification sometimes can be made that removes or disrupts sequences that increase or enhance transcription initiation, resulting in a decrease or elimination of transcription from an endogenous or heterologous promoter element.

In some embodiments, alteration of a nucleic acid reagent or nucleotide sequence can alter sequences involved in translational initiation or translational efficiency (e.g., 5' UTR, 3' UTR, codon triplets of higher or lower abundance, translational terminator sequences and the like, for example). A modification sometimes can be made that can increase or decrease translational initiation, modifying a ribosome binding site for example. A modification sometimes can be made that can increase or decrease translational efficiency. Removing or adding sequences that form hairpins and changing codon triplets to a more or less preferred codon are non-limiting examples of genetic modifications that can be made to alter translation initiation and translation efficiency.

In certain embodiments, alteration of a nucleic acid reagent or nucleotide sequence can alter sequences involved in localization of peptides, proteins or other desired products (e.g., a sebatic acid or dodecanedioic acid, for example). A modification sometimes can be made that can alter, add or remove sequences responsible for targeting a polypeptide, protein or product to an intracellular organelle, the periplasm, cellular membranes, or extracellularly. Transport of a heterologous product to a different intracellular space or extracellularly sometimes can reduce or eliminate the formation of inclusion bodies (e.g., insoluble aggregates of the desired product).

In some embodiments, alteration of a nucleic acid reagent or nucleotide sequence can alter sequences involved in increasing or decreasing the copy number of a nucleotide sequence of interest. A modification sometimes can be made that increases or decreases the number of copies of an ORF stably integrated into the genome of an organism or on an epigenetic nucleic acid reagent. Non-limiting examples of alterations that can increase the number of copies of a sequence of interest include, adding copies of the sequence of interest by duplication of regions in the genome (e.g., adding additional copies by recombination or by causing gene amplification of the host genome, for example), cloning additional copies of a sequence onto a nucleic acid reagent, or altering an ORI to increase the number of copies of an epigenetic nucleic acid reagent. Non-limiting examples of alterations that can decrease the number of copies of a sequence of

interest include, removing copies of the sequence of interest by deletion or disruption of regions in the genome, removing additional copies of the sequence from epigenetic nucleic acid reagents, or altering an ORI to decrease the number of copies of an epigenetic nucleic acid reagent.

- 5 In certain embodiments, increasing or decreasing the expression of a nucleotide sequence of interest can also be accomplished by altering, adding or removing sequences involved in the expression of an anti-sense RNA, RNAi, siRNA, ribozyme and the like. The methods described above can be used to modify expression of anti-sense RNA, RNAi, siRNA, ribozyme and the like.
- 10 The methods and nucleic acid reagents described herein can be used to generate genetically modified microorganisms with altered activities in cellular processes involved in a fatty dicarboxylic acid (e.g., octanedioic acid, decanedioic acid, dodecanedioic acid, tetradecanedioic acid, hexadecanedioic acid, octadecanedioic acid, eicosanedioic acid) synthesis. In some
15 embodiments, an engineered microorganism described herein may comprise an increased number of copies of an endogenous polynucleotide encoding a polypeptide having omega oxo fatty acid dehydrogenase activity. In certain embodiments, an engineered microorganism described herein may comprise an increased number of copies of an endogenous polynucleotide encoding a polypeptide having omega hydroxyl fatty acid dehydrogenase activity. In some embodiments, an
20 engineered microorganism described herein may comprise a heterologous polynucleotide encoding a polypeptide having omega oxo fatty acid dehydrogenase activity. In some embodiments, an engineered microorganism described herein may comprise a heterologous polynucleotide encoding a polypeptide having omega hydroxyl fatty acid dehydrogenase activity. In some embodiments, the heterologous polynucleotide can be from a bacterium. In some
25 embodiments, the bacterium can be an Acinetobacter, Nocardia, Pseudomonas or Xanthobacter bacterium.

In some embodiments, an engineered microorganism described herein may comprise a heterologous polynucleotide encoding a polypeptide having monooxygenase activity. In certain
30 embodiments, the heterologous polynucleotide can be from a bacterium. In some embodiments, the bacterium can be a Bacillus bacterium. In certain embodiments, the Bacillus bacterium is B. megaterium.

In certain embodiments, an engineered microorganism described herein may comprise a genetic modification that reduces omega hydroxyl fatty acid conversion. In some embodiments, the

genetic modification can reduce omega hydroxyl fatty acid dehydrogenase activity. In certain embodiments, an engineered microorganism described herein may comprise a genetic modification that reduces beta-oxidation activity. In some embodiments, the genetic modification can reduce a target activity described herein.

5

Engineered microorganisms that produce a fatty dicarboxylic acid (e.g., octanedioic acid, decanedioic acid, dodecanedioic acid, tetradecanedioic acid, hexadecanedioic acid, octadecanedioic acid, eicosanedioic acid), as described herein, can comprise an altered monooxygenase activity, in certain embodiments. In some embodiments, the engineered microorganism described herein may comprise a genetic modification that alters the monooxygenase activity. In certain embodiments, the engineered microorganism described herein can comprise an increase number of copies of an endogenous polynucleotide encoding a polypeptide having monooxygenase activity. In some embodiments, the engineered microorganism described herein can comprise a heterologous polynucleotide encoding a polypeptide having monooxygenase activity. In certain embodiments, the heterologous polynucleotide can be from a bacterium. In some embodiments, the bacterium can be a *Bacillus* bacterium. In certain embodiments, the *Bacillus* bacterium is *B. megaterium*. In some embodiments, the genetic modification can reduce a polyketide synthase activity.

20 Engineered microorganisms that produce a fatty dicarboxylic acid (e.g., octanedioic acid, decanedioic acid, dodecanedioic acid, tetradecanedioic acid, hexadecanedioic acid, octadecanedioic acid, eicosanedioic acid), as described herein, can comprise an altered thioesterase activity, in certain embodiments. In some embodiments, the engineered microorganism may comprise a genetic modification that adds or increases the thioesterase activity. In certain embodiments, the engineered microorganism may comprise a heterologous polynucleotide encoding a polypeptide having thioesterase activity.

In some embodiments, the engineered microorganism with an altered thioesterase activity may comprise an altered omega oxo fatty acid dehydrogenase activity. In certain embodiments, the engineered microorganism with an altered thioesterase activity may comprise a genetic modification that adds or increases omega oxo fatty acid dehydrogenase activity. In some embodiments, the engineered microorganism may comprise a heterologous polynucleotide encoding a polypeptide having altered omega oxo fatty acid dehydrogenase activity. In certain

embodiments, the heterologous polynucleotide can be from a bacterium. In some embodiments, the bacterium can be an *Acinetobacter*, *Nocardia*, *Pseudomonas* or *Xanthobacter* bacterium.

Engineered microorganisms that produce a fatty dicarboxylic acid (e.g., octanedioic acid, decanedioic acid, dodecanedioic acid, tetradecanedioic acid, hexadecanedioic acid, octadecanedioic acid, eicosanedioic acid), as described herein, can comprise an altered omega hydroxyl fatty acid dehydrogenase activity. In certain embodiments, the engineered microorganism may comprise a genetic modification that adds or increases the omega hydroxyl fatty acid dehydrogenase activity. In certain embodiments, the engineered microorganism may comprise a heterologous polynucleotide encoding a polypeptide having altered omega hydroxyl fatty acid dehydrogenase activity. In some embodiments, the heterologous polynucleotide is from a bacterium. In certain embodiments, the bacterium can be an *Acinetobacter*, *Nocardia*, *Pseudomonas* or *Xanthobacter* bacterium. In some embodiments, the engineered microorganism can be a eukaryote. In certain embodiments, the eukaryote can be a yeast. In some embodiments, the eukaryote may be a fungus. In certain embodiments, the yeast can be a *Candida* yeast. In some embodiments, the *Candida* yeast may be *C. tropicalis*. In certain embodiments, the fungus can be a *Yarrowia* fungus. In some embodiments the *Yarrowia* fungus may be *Y. lipolytica*. In certain embodiments, the fungus can be an *Aspergillus* fungus. In some embodiments, the *Aspergillus* fungus may be *A. parasiticus* or *A. nidulans*. In some embodiments, an engineered microorganism as described above may comprise a genetic modification that reduces omega hydroxyl fatty acid conversion. In certain embodiments, the genetic modification can reduce omega hydroxyl fatty acid dehydrogenase activity. In some embodiments the genetic may reduce beta-oxidation activity. In certain embodiments, the genetic modification may reduce a target activity described herein.

Engineered microorganisms can be prepared by altering, introducing or removing nucleotide sequences in the host genome or in stably maintained epigenetic nucleic acid reagents, as noted above. The nucleic acid reagents use to alter, introduce or remove nucleotide sequences in the host genome or epigenetic nucleic acids can be prepared using the methods described herein or available to the artisan.

Nucleic acid sequences having a desired activity can be isolated from cells of a suitable organism using lysis and nucleic acid purification procedures described in a known reference manual (e.g., Maniatis, T., E. F. Fritsch and J. Sambrook (1982) *Molecular Cloning: a Laboratory Manual*; Cold

Spring Harbor Laboratory, Cold Spring Harbor, N.Y.) or using commercially available cell lysis and DNA purification reagents and kits. In some embodiments, nucleic acids used to engineer microorganisms can be provided for conducting methods described herein after processing of the organism containing the nucleic acid. For example, the nucleic acid of interest may be extracted, isolated, purified or amplified from a sample (e.g., from an organism of interest or culture containing a plurality of organisms of interest, like yeast or bacteria for example). The term "isolated" as used herein refers to nucleic acid removed from its original environment (e.g., the natural environment if it is naturally occurring, or a host cell if expressed exogenously), and thus is altered "by the hand of man" from its original environment. An isolated nucleic acid generally is provided with fewer non-nucleic acid components (e.g., protein, lipid) than the amount of components present in a source sample. A composition comprising isolated sample nucleic acid can be substantially isolated (e.g., about 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or greater than 99% free of non-nucleic acid components). The term "purified" as used herein refers to sample nucleic acid provided that contains fewer nucleic acid species than in the sample source from which the sample nucleic acid is derived. A composition comprising sample nucleic acid may be substantially purified (e.g., about 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or greater than 99% free of other nucleic acid species). The term "amplified" as used herein refers to subjecting nucleic acid of a cell, organism or sample to a process that linearly or exponentially generates amplicon nucleic acids having the same or substantially the same nucleotide sequence as the nucleotide sequence of the nucleic acid in the sample, or portion thereof. As noted above, the nucleic acids used to prepare nucleic acid reagents as described herein can be subjected to fragmentation or cleavage.

Amplification of nucleic acids is sometimes necessary when dealing with organisms that are difficult to culture. Where amplification may be desired, any suitable amplification technique can be utilized. Non-limiting examples of methods for amplification of polynucleotides include, polymerase chain reaction (PCR); ligation amplification (or ligase chain reaction (LCR)); amplification methods based on the use of Q-beta replicase or template-dependent polymerase (see US Patent Publication Number US20050287592); helicase-dependant isothermal amplification (Vincent et al., "Helicase-dependent isothermal DNA amplification". EMBO reports 5 (8): 795-800 (2004)); strand displacement amplification (SDA); thermophilic SDA nucleic acid sequence based amplification (3SR or NASBA) and transcription-associated amplification (TAA). Non-limiting examples of PCR amplification methods include standard PCR, AFLP-PCR, Allele-specific PCR, Alu-PCR, Asymmetric PCR, Colony PCR, Hot start PCR, Inverse PCR (IPCR), In

situ PCR (ISH), Intersequence-specific PCR (ISSR-PCR), Long PCR, Multiplex PCR, Nested PCR, Quantitative PCR, Reverse Transcriptase PCR (RT-PCR), Real Time PCR, Single cell PCR, Solid phase PCR, combinations thereof, and the like. Reagents and hardware for conducting PCR are commercially available.

5

Protocols for conducting the various type of PCR listed above are readily available to the artisan. PCR conditions can be dependent upon primer sequences, target abundance, and the desired amount of amplification, and therefore, one of skill in the art may choose from a number of PCR protocols available (see, e.g., U.S. Pat. Nos. 4,683,195 and 4,683,202; and PCR Protocols: A

10 Guide to Methods and Applications, Innis et al., eds, 1990. PCR often is carried out as an automated process with a thermostable enzyme. In this process, the temperature of the reaction mixture is cycled through a denaturing region, a primer-annealing region, and an extension reaction region automatically. Machines specifically adapted for this purpose are commercially available. A non-limiting example of a PCR protocol that may be suitable for embodiments
15 described herein is, treating the sample at 95°C for 5 minutes; repeating forty-five cycles of 95°C for 1 minute, 59°C for 1 minute, 10 seconds, and 72°C for 1 minute 30 seconds; and then treating the sample at 72°C for 5 minutes. Additional PCR protocols are described in the example section. Multiple cycles frequently are performed using a commercially available thermal cycler. Suitable isothermal amplification processes known and selected by the person of ordinary skill in the art
20 also may be applied, in certain embodiments. In some embodiments, nucleic acids encoding polypeptides with a desired activity can be isolated by amplifying the desired sequence from an organism having the desired activity using oligonucleotides or primers designed based on sequences described herein.

25 Amplified, isolated and/or purified nucleic acids can be cloned into the recombinant DNA vectors described in Figures herein or into suitable commercially available recombinant DNA vectors. Cloning of nucleic acid sequences of interest into recombinant DNA vectors can facilitate further manipulations of the nucleic acids for preparation of nucleic acid reagents, (e.g., alteration of nucleotide sequences by mutagenesis, homologous recombination, amplification and the like, for
30 example). Standard cloning procedures (e.g., enzymic digestion, ligation, and the like) are known (e.g., described in Maniatis, T., E. F. Fritsch and J. Sambrook (1982) *Molecular Cloning: a Laboratory Manual*; Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.).

In some embodiments, nucleic acid sequences prepared by isolation or amplification can be used, without any further modification, to add an activity to a microorganism and thereby create a genetically modified or engineered microorganism. In certain embodiments, nucleic acid sequences prepared by isolation or amplification can be genetically modified to alter (e.g., increase or decrease, for example) a desired activity. In some embodiments, nucleic acids, used to add an activity to an organism, sometimes are genetically modified to optimize the heterologous polynucleotide sequence encoding the desired activity (e.g., polypeptide or protein, for example). The term "optimize" as used herein can refer to alteration to increase or enhance expression by preferred codon usage. The term optimize can also refer to modifications to the amino acid sequence to increase the activity of a polypeptide or protein, such that the activity exhibits a higher catalytic activity as compared to the "natural" version of the polypeptide or protein.

Nucleic acid sequences of interest can be genetically modified using methods known in the art. Mutagenesis techniques are particularly useful for small scale (e.g., 1, 2, 5, 10 or more nucleotides) or large scale (e.g., 50, 100, 150, 200, 500, or more nucleotides) genetic modification. Mutagenesis allows the artisan to alter the genetic information of an organism in a stable manner, either naturally (e.g., isolation using selection and screening) or experimentally by the use of chemicals, radiation or inaccurate DNA replication (e.g., PCR mutagenesis). In some embodiments, genetic modification can be performed by whole scale synthetic synthesis of nucleic acids, using a native nucleotide sequence as the reference sequence, and modifying nucleotides that can result in the desired alteration of activity. Mutagenesis methods sometimes are specific or targeted to specific regions or nucleotides (e.g., site-directed mutagenesis, PCR-based site-directed mutagenesis, and in vitro mutagenesis techniques such as transplacement and in vivo oligonucleotide site-directed mutagenesis, for example). Mutagenesis methods sometimes are non-specific or random with respect to the placement of genetic modifications (e.g., chemical mutagenesis, insertion element (e.g., insertion or transposon elements) and inaccurate PCR based methods, for example).

Site directed mutagenesis is a procedure in which a specific nucleotide or specific nucleotides in a DNA molecule are mutated or altered. Site directed mutagenesis typically is performed using a nucleic acid sequence of interest cloned into a circular plasmid vector. Site-directed mutagenesis requires that the wild type sequence be known and used a platform for the genetic alteration. Site-directed mutagenesis sometimes is referred to as oligonucleotide-directed mutagenesis because the technique can be performed using oligonucleotides which have the desired genetic

modification incorporated into the complement a nucleotide sequence of interest. The wild type sequence and the altered nucleotide are allowed to hybridize and the hybridized nucleic acids are extended and replicated using a DNA polymerase. The double stranded nucleic acids are introduced into a host (e.g., *E. coli*, for example) and further rounds of replication are carried out in vivo. The transformed cells carrying the mutated nucleic acid sequence are then selected and/or screened for those cells carrying the correctly mutagenized sequence. Cassette mutagenesis and PCR-based site-directed mutagenesis are further modifications of the site-directed mutagenesis technique. Site-directed mutagenesis can also be performed in vivo (e.g., transplacement “pop-in pop-out”, In vivo site-directed mutagenesis with synthetic oligonucleotides and the like, for example).

PCR-based mutagenesis can be performed using PCR with oligonucleotide primers that contain the desired mutation or mutations. The technique functions in a manner similar to standard site-directed mutagenesis, with the exception that a thermocycler and PCR conditions are used to replace replication and selection of the clones in a microorganism host. As PCR-based mutagenesis also uses a circular plasmid vector, the amplified fragment (e.g., linear nucleic acid molecule) containing the incorporated genetic modifications can be separated from the plasmid containing the template sequence after a sufficient number of rounds of thermocycler amplification, using standard electrophoretic procedures. A modification of this method uses linear amplification methods and a pair of mutagenic primers that amplify the entire plasmid. The procedure takes advantage of the *E. coli* Dam methylase system which causes DNA replicated in vivo to be sensitive to the restriction endonucleases DpnI. PCR synthesized DNA is not methylated and is therefore resistant to DpnI. This approach allows the template plasmid to be digested, leaving the genetically modified, PCR synthesized plasmids to be isolated and transformed into a host bacteria for DNA repair and replication, thereby facilitating subsequent cloning and identification steps. A certain amount of randomness can be added to PCR-based sited directed mutagenesis by using partially degenerate primers.

Recombination sometimes can be used as a tool for mutagenesis. Homologous recombination allows the artisan to specifically target regions of known sequence for insertion of heterologous nucleotide sequences using the host organisms natural DNA replication and repair enzymes. Homologous recombination methods sometimes are referred to as “pop in pop out” mutagenesis, transplacement, knock out mutagenesis or knock in mutagenesis. Integration of a nucleic acid sequence into a host genome is a single cross over event, which inserts the entire nucleic acid

reagent (e.g., pop in). A second cross over event excises all but a portion of the nucleic acid reagent, leaving behind a heterologous sequence, often referred to as a "footprint" (e.g., pop out). Mutagenesis by insertion (e.g., knock in) or by double recombination leaving behind a disrupting heterologous nucleic acid (e.g., knock out) both serve to disrupt or "knock out" the function of the gene or nucleic acid sequence in which insertion occurs. By combining selectable markers and/or auxotrophic markers with nucleic acid reagents designed to provide the appropriate nucleic acid target sequences, the artisan can target a selectable nucleic acid reagent to a specific region, and then select for recombination events that "pop out" a portion of the inserted (e.g., "pop in") nucleic acid reagent.

10

Such methods take advantage of nucleic acid reagents that have been specifically designed with known target nucleic acid sequences at or near a nucleic acid or genomic region of interest. Popping out typically leaves a "foot print" of left over sequences that remain after the recombination event. The left over sequence can disrupt a gene and thereby reduce or eliminate expression of that gene. In some embodiments, the method can be used to insert sequences, upstream or downstream of genes that can result in an enhancement or reduction in expression of the gene. In certain embodiments, new genes can be introduced into the genome of a host organism using similar recombination or "pop in" methods. An example of a yeast recombination system using the *ura3* gene and 5-FOA were described briefly above and further detail is presented herein.

15
20

A method for modification is described in Alani et al., "A method for gene disruption that allows repeated use of URA3 selection in the construction of multiply disrupted yeast strains", *Genetics* 116(4):541-545 August 1987. The original method uses a Ura3 cassette with 1000 base pairs (bp) of the same nucleotide sequence cloned in the same orientation on either side of the URA3 cassette. Targeting sequences of about 50 bp are added to each side of the construct. The double stranded targeting sequences are complementary to sequences in the genome of the host organism. The targeting sequences allow site-specific recombination in a region of interest. The modification of the original technique replaces the two 1000 bp sequence direct repeats with two 200 bp direct repeats. The modified method also uses 50 bp targeting sequences. The modification reduces or eliminates recombination of a second knock out into the 1000 bp repeat left behind in a first mutagenesis, therefore allowing multiply knocked out yeast. Additionally, the 200 bp sequences used herein are uniquely designed, self-assembling sequences that leave behind identifiable footprints. The technique used to design the sequences incorporate design

25
30

features such as low identity to the yeast genome, and low identity to each other. Therefore a library of the self-assembling sequences can be generated to allow multiple knockouts in the same organism, while reducing or eliminating the potential for integration into a previous knockout.

- 5 As noted above, the URA3 cassette makes use of the toxicity of 5-FOA in yeast carrying a functional URA3 gene. Uracil synthesis deficient yeast are transformed with the modified URA3 cassette, using standard yeast transformation protocols, and the transformed cells are plated on minimal media minus uracil. In some embodiments, PCR can be used to verify correct insertion into the region of interest in the host genome, and certain embodiments the PCR step can be
- 10 omitted. Inclusion of the PCR step can reduce the number of transformants that need to be counter selected to "pop out" the URA3 cassette. The transformants (e.g., all or the ones determined to be correct by PCR, for example) can then be counter-selected on media containing 5-FOA, which will select for recombination out (e.g., popping out) of the URA3 cassette, thus rendering the yeast *ura3* deficient again, and resistant to 5-FOA toxicity. Targeting sequences
- 15 used to direct recombination events to specific regions are presented herein. A modification of the method described above can be used to integrate genes in to the chromosome, where after recombination a functional gene is left in the chromosome next to the 200bp footprint.

- In some embodiments, other auxotrophic or dominant selection markers can be used in place of
- 20 URA3 (e.g., an auxotrophic selectable marker), with the appropriate change in selection media and selection agents. Auxotrophic selectable markers are used in strains deficient for synthesis of a required biological molecule (e.g., amino acid or nucleoside, for example). Non-limiting examples of additional auxotrophic markers include; *HIS3*, *TRP1*, *LEU2*, *LEU2-d*, and *LYS2*. Certain auxotrophic markers (e.g., URA3 and *LYS2*) allow counter selection to select for the second
- 25 recombination event that pops out all but one of the direct repeats of the recombination construct. *HIS3* encodes an activity involved in histidine synthesis. *TRP1* encodes an activity involved in tryptophan synthesis. *LEU2* encodes an activity involved in leucine synthesis. *LEU2-d* is a low expression version of *LEU2* that selects for increased copy number (e.g., gene or plasmid copy number, for example) to allow survival on minimal media without leucine. *LYS2* encodes an
- 30 activity involved in lysine synthesis, and allows counter selection for recombination out of the *LYS2* gene using alpha-amino adipate (α -amino adipate).

Dominant selectable markers are useful because they also allow industrial and/or prototrophic strains to be used for genetic manipulations. Additionally, dominant selectable markers provide the advantage that rich medium can be used for plating and culture growth, and thus growth rates are

markedly increased. Non-limiting examples of dominant selectable markers include; Tn903 kan^r, Cm^r, Hyg^r, CUP1, and DHFR. Tn903 kan^r encodes an activity involved in kanamycin antibiotic resistance (e.g., typically neomycin phosphotransferase II or NPTII, for example). Cm^r encodes an activity involved in chloramphenicol antibiotic resistance (e.g., typically chloramphenicol acetyl transferase or CAT, for example). Hyg^r encodes an activity involved in hygromycin resistance by phosphorylation of hygromycin B (e.g., hygromycin phosphotransferase, or HPT). CUP1 encodes an activity involved in resistance to heavy metal (e.g., copper, for example) toxicity. DHFR encodes a dihydrofolate reductase activity which confers resistance to methotrexate and sulfanilamide compounds.

10

In contrast to site-directed or specific mutagenesis, random mutagenesis does not require any sequence information and can be accomplished by a number of widely different methods. Random mutagenesis often is used to generate mutant libraries that can be used to screen for the desired genotype or phenotype. Non-limiting examples of random mutagenesis include; chemical mutagenesis, UV-induced mutagenesis, insertion element or transposon-mediated mutagenesis, DNA shuffling, error-prone PCR mutagenesis, and the like.

15

Chemical mutagenesis often involves chemicals like ethyl methanesulfonate (EMS), nitrous acid, mitomycin C, N-methyl-N-nitrosourea (MNU), diepoxybutane (DEB), 1, 2, 7, 8- diepoxyoctane (DEO), methyl methane sulfonate (MMS), N-methyl- N'-nitro-N-nitrosoguanidine (MNNG), 4-nitroquinoline 1-oxide (4-NQO), 2-methyloxy-6-chloro-9(3-[ethyl-2-chloroethyl]-aminopropylamino)-acridinedihydrochloride (ICR-170), 2-amino purine (2AP), and hydroxylamine (HA), provided herein as non-limiting examples. These chemicals can cause base-pair substitutions, frameshift mutations, deletions, transversion mutations, transition mutations, incorrect replication, and the like. In some embodiments, the mutagenesis can be carried out in vivo. Sometimes the mutagenic process involves the use of the host organisms DNA replication and repair mechanisms to incorporate and replicate the mutagenized base or bases.

25

Another type of chemical mutagenesis involves the use of base-analogs. The use of base-analogs cause incorrect base pairing which in the following round of replication is corrected to a mismatched nucleotide when compared to the starting sequence. Base analog mutagenesis introduces a small amount of non-randomness to random mutagenesis, because specific base analogs can be chose which can be incorporated at certain nucleotides in the starting sequence. Correction of the mispairing typically yields a known substitution. For example, Bromo-

30

deoxyuridine (BrdU) can be incorporated into DNA and replaces T in the sequence. The host DNA repair and replication machinery can sometime correct the defect, but sometimes will mispair the BrdU with a G. The next round of replication then causes a G-C transversion from the original A-T in the native sequence.

5

Ultra violet (UV) induced mutagenesis is caused by the formation of thymidine dimers when UV light irradiates chemical bonds between two adjacent thymine residues. Excision repair mechanism of the host organism correct the lesion in the DNA, but occasionally the lesion is incorrectly repaired typically resulting in a C to T transition.

10

Insertion element or transposon-mediated mutagenesis makes use of naturally occurring or modified naturally occurring mobile genetic elements. Transposons often encode accessory activities in addition to the activities necessary for transposition (e.g., movement using a transposase activity, for example). In many examples, transposon accessory activities are antibiotic resistance markers (e.g., see Tn903 kan^r described above, for example). Insertion elements typically only encode the activities necessary for movement of the nucleic acid sequence. Insertion element and transposon mediated mutagenesis often can occur randomly, however specific target sequences are known for some transposons. Mobile genetic elements like IS elements or Transposons (Tn) often have inverted repeats, direct repeats or both inverted and direct repeats flanking the region coding for the transposition genes. Recombination events catalyzed by the transposase cause the element to remove itself from the genome and move to a new location, leaving behind a portion of an inverted or direct repeat. Classic examples of transposons are the "mobile genetic elements" discovered in maize. Transposon mutagenesis kits are commercially available which are designed to leave behind a 5 codon insert (e.g., Mutation Generation System kit, Finnzymes, World Wide Web URL finnzymes.us, for example). This allows the artisan to identify the insertion site, without fully disrupting the function of most genes.

DNA shuffling is a method which uses DNA fragments from members of a mutant library and reshuffles the fragments randomly to generate new mutant sequence combinations. The fragments are typically generated using DNaseI, followed by random annealing and re-joining using self priming PCR. The DNA overhanging ends, from annealing of random fragments, provide "primer" sequences for the PCR process. Shuffling can be applied to libraries generated by any of the above mutagenesis methods.

30

Error prone PCR and its derivative rolling circle error prone PCR uses increased magnesium and manganese concentrations in conjunction with limiting amounts of one or two nucleotides to reduce the fidelity of the Taq polymerase. The error rate can be as high as 2% under appropriate conditions, when the resultant mutant sequence is compared to the wild type starting sequence.

- 5 After amplification, the library of mutant coding sequences must be cloned into a suitable plasmid. Although point mutations are the most common types of mutation in error prone PCR, deletions and frameshift mutations are also possible. There are a number of commercial error-prone PCR kits available, including those from Stratagene and Clontech (e.g., World Wide Web URL stratagene.com and World Wide Web URL clontech.com, respectively, for example). Rolling circle
10 error-prone PCR is a variant of error-prone PCR in which wild-type sequence is first cloned into a plasmid, then the whole plasmid is amplified under error-prone conditions.

- As noted above, organisms with altered activities can also be isolated using genetic selection and screening of organisms challenged on selective media or by identifying naturally occurring variants
15 from unique environments. For example, 2-Deoxy-D-glucose is a toxic glucose analog. Growth of yeast on this substance yields mutants that are glucose-deregulated. A number of mutants have been isolated using 2-Deoxy-D-glucose including transport mutants, and mutants that ferment glucose and galactose simultaneously instead of glucose first then galactose when glucose is depleted. Similar techniques have been used to isolate mutant microorganisms that can
20 metabolize plastics (e.g., from landfills), petrochemicals (e.g., from oil spills), and the like, either in a laboratory setting or from unique environments.

- Similar methods can be used to isolate naturally occurring mutations in a desired activity when the activity exists at a relatively low or nearly undetectable level in the organism of choice, in some
25 embodiments. The method generally consists of growing the organism to a specific density in liquid culture, concentrating the cells, and plating the cells on various concentrations of the substance to which an increase in metabolic activity is desired. The cells are incubated at a moderate growth temperature, for 5 to 10 days. To enhance the selection process, the plates can be stored for another 5 to 10 days at a low temperature. The low temperature sometimes can
30 allow strains that have gained or increased an activity to continue growing while other strains are inhibited for growth at the low temperature. Following the initial selection and secondary growth at low temperature, the plates can be replica plated on higher or lower concentrations of the selection substance to further select for the desired activity.

A native, heterologous or mutagenized polynucleotide can be introduced into a nucleic acid reagent for introduction into a host organism, thereby generating an engineered microorganism. Standard recombinant DNA techniques (restriction enzyme digests, ligation, and the like) can be used by the artisan to combine the mutagenized nucleic acid of interest into a suitable nucleic acid reagent capable of (i) being stably maintained by selection in the host organism, or (ii) being integrating into the genome of the host organism. As noted above, sometimes nucleic acid reagents comprise two replication origins to allow the same nucleic acid reagent to be manipulated in bacterial before final introduction of the final product into the host organism (e.g., yeast or fungus for example). Standard molecular biology and recombinant DNA methods are known (e.g., described in Maniatis, T., E. F. Fritsch and J. Sambrook (1982) *Molecular Cloning: a Laboratory Manual*; Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.).

Nucleic acid reagents can be introduced into microorganisms using various techniques. Non-limiting examples of methods used to introduce heterologous nucleic acids into various organisms include; transformation, transfection, transduction, electroporation, ultrasound-mediated transformation, particle bombardment and the like. In some instances the addition of carrier molecules (e.g., bis-benzimidazolyl compounds, for example, see US Patent 5595899) can increase the uptake of DNA in cells typically though to be difficult to transform by conventional methods. Conventional methods of transformation are known (e.g., described in Maniatis, T., E. F. Fritsch and J. Sambrook (1982) *Molecular Cloning: a Laboratory Manual*; Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.).

Modified Activities

Certain activities in a genetically modified organism can be modified by techniques known in the art. An acyl-coA oxidase activity or acyl-coA dehydrogenase activity, or acyl-coA oxidase activity and acyl-coA dehydrogenase activity, can be modified in an organism in certain embodiments. In some embodiments, a modified endogenous acyl-coA oxidase polypeptide, modified endogenous acyl-coA dehydrogenase polypeptide, modified heterologous acyl-coA oxidase polypeptide, and/or modified heterologous acyl-coA dehydrogenase polypeptide may be introduced into an organism. A modified polypeptide can be expressed by a host organism that includes a modified polynucleotide encoding the modified polypeptide. Modified polypeptides often have an activity different than the activity of an unmodified counterpart. A modified activity sometimes is a different catalytic activity or a different substrate specificity, or a different catalytic activity and a different

substrate specificity. A different activity sometimes is an activity that is higher or lower than the activity of an unmodified counterpart polypeptide. In some embodiments, the catalytic activity of a modified polypeptide is higher or lower than the catalytic activity of the unmodified counterpart for a particular substrate. In certain embodiments, the substrate specificity of a modified polypeptide is higher or lower than the substrate specificity of the unmodified counterpart for a particular substrate. A modified polypeptide often is active and an activity of a modified polypeptide often can be detected (e.g., substrate turnover can be detected). A desired activity for a particular polypeptide sometimes is referred to as a "target activity."

- 10 In some embodiments a genetic modification in a genetically modified organism alters a substrate specificity of an acyl-coA oxidase polypeptide produced in the organism. Sometimes the substrate specificity is reduced for a substrate having a particular chain length. In some embodiments, a modified acyl-coA oxidase substrate specificity is reduced for a C8, C10, C12, C14, C16, C18, C20 substrate or combination thereof. In certain embodiments, a modified acyl-coA oxidase substrate specificity is reduced for a C10, C12, or C18 substrate.

In some embodiments a genetic modification in a genetically modified organism alters a substrate specificity of an acyl-coA dehydrogenase polypeptide produced in the organism. Sometimes a co-factor specificity is modified, and in some embodiments the modified polypeptide prefers to utilize oxygen as a co-factor.

One or more particular modifications can be selected to generate a modified polypeptide having a target activity. Modifications often are amino acid modifications (e.g., deletion, insertion of one or more amino acids). Amino acid modifications sometimes are amino acid substitutions. Amino acid substitutions sometimes are conservative, non-limiting examples of which include substitution of an amino acid containing an acidic moiety to another amino acid containing an acidic moiety (e.g., D, E), substitution of an amino acid containing a basic moiety to another amino acid containing a basic moiety (e.g., H, K, R), substitution of an amino acid containing an aliphatic chain moiety to another amino acid containing an aliphatic chain moiety (V, L, I, A), substitution of an amino acid containing a cyclic moiety to another amino acid containing a cyclic moiety (e.g., W, F, Y), and substitution of an amino acid containing a polar moiety to another amino acid containing a polar moiety (e.g., S, T). Amino acid substitutions sometimes are non-conservative, non-limiting examples of which include substitution of an amino acid containing an acidic moiety to an amino acid containing a basic moiety, substitution of an amino acid containing a basic moiety to an amino

acid containing an acidic moiety, substitution of an amino acid containing relatively small moiety (e.g., G, A) to another amino acid containing a relatively large moiety (e.g., Y, W, F, I, L), and substitution of an amino acid containing a relatively large moiety to another amino acid containing an relatively small moiety.

5

Particular modifications can be selected using any suitable method known in the art. In certain embodiments, a reference structure is known for a related polypeptide with a known activity, and modifications to a target polypeptide can be guided by alignment of the target polypeptide structure to the reference structure. A reference structure sometimes is a primary structure (e.g.,

10 polynucleotide or polypeptide sequence) and the primary structure of a target can be aligned to the reference structure using an alignment method known in the art. Particular amino acids in the target that align with (e.g., are identical to or homologous to) or do not align with (e.g., are not identical to or not homologous to) particular amino acids in the reference can be selected for modification. Selections can be made by inspection of an alignment or by software known in the art that identifies, scores and/or ranks amino acids for modification based on an alignment.

A reference structure sometimes is a secondary structure, tertiary structure or quaternary structure, each of which are three dimensional structures pertaining to a polypeptide. A primary structure of a target polypeptide can be modeled to a secondary, tertiary or quaternary reference structure using three-dimensional modeling software known in the art. A secondary, tertiary or quaternary structure of a target polypeptide can be compared to a secondary, tertiary or quaternary reference structure using three-dimensional comparative software known in the art. Particular structures (e.g., a particular individual amino acid; a particular group of contiguous or non-contiguous amino acids) in the target that align with or map to, or do not align with or map to, particular structures in the reference can be selected for modification. Also, particular structures in the target that are in proximity to a substrate or co-factor can be selected for modification. Selections can be made by inspection of an alignment or map or by software known in the art that identifies, scores and/or ranks amino acids and/or structures for modification based on an alignment and map.

30 After particular amino acids and/or structures are selected for modification in a first polypeptide, amino acids and structures in a second polypeptide that align with the selected amino acids and structures in the first polypeptide may be identified. In a non-limiting example, particular amino acid substitutions and structural modifications (e.g., loop amino acid deletion/insertion) for *Candida* spp. POX4 and POX5 polypeptides are disclosed herein. A primary structure of another acyl-coA

oxidase polypeptide can be aligned with the amino acid sequence or modeled structure of a POX4 or POX5 polypeptide and some or all amino acids of the other polypeptide that align with those selected for modification in the POX4 or POX5 polypeptide also can be selected for modification. Certain criteria for selecting acyl-coA dehydrogenase modifications also are described herein.

5

One or more activities of a modified polypeptide can be characterized using any suitable assay known in the art. A modified polypeptide can be expressed in an organism other than a target organism in which a target product will be produced, for assaying activity. For example, a modified polypeptide can be expressed in a bacterium (e.g., *E. coli*), assayed and then introduced into a yeast (e.g., *Candida* spp. yeast) for production of a target diacid.

10

Feedstocks, Media, Supplements & Additives

Engineered microorganisms often are cultured under conditions that optimize yield of a fatty dicarboxylic acid (e.g., an eight to eighteen-carbon fatty dicarboxylic acid). Non-limiting examples of fatty dicarboxylic acids include suberic acid (i.e., octanedioic acid, 1,8-octanedioic acid, octanedioic acid, octane-1,8-dioic acid, 1,6-hexanedicarboxylic acid, capryllic diacids), sebacic acid (i.e., 1,10-decanedioic acid, decanedioic acid, decane-1,10-dioic acid, 1,8-octanedicarboxylic acid, capric diacid), dodecanedioic acid (i.e., DDDA, 1,12-dodecanedioic acid, dodecanedioic acid, dodecane-1,12-dioic acid, 1,10-decanedicarboxylic acid, decamethylenedicarboxylic acid, 1,10-dicarboxydecane, lauric diacid), tetradecanedioic acid (i.e., TDDA, 1,14-tetradecanedioic acid, tetradecanedioic acid, tetradecane-1,14-dioic acid, 1,12-dodecanedicarboxylic acid, myristic diacid), thapsic acid (i.e., hexadecanedioic acid, 1,16-hexadecanedioic acid, hexadecanedioic acid, hexadecane-1,16-dioic acid, 1,14-tetradecanedicarboxylic acid, palmitic diacid), cis-9-hexadecenedioic acid (i.e., palmitoleic diacids), octanedioic acid (i.e., 1,18-octadecanedioic acid, octadecanedioic acid, octadecane-1,18-dioic acid, 1,16-hexadecanedicarboxylic acid, stearic diacid), cis-9-octadecenedioic acid (i.e., oleic diacids), cis-9,12-octadecenedioic acid (i.e., linoleic diacids), cis-9,12,15-octadecenedioic acid (i.e., linolenic diacids), arachidic diacid (i.e., eicosanoic diacid, icosanoic diacid), 11-eicosenoic diacid (i.e., cis-11-eicosenedioic acid), 13-eicosenoic diacids (i.e., cis-13-eicosenedioic acid), arachidonic diacid (i.e., cis-5,8,11,14-eicosatetraenedioic acid). Culture conditions often optimize activity of one or more of the following activities: omega oxo fatty acid dehydrogenase activity, omega hydroxyl fatty acid dehydrogenase activity, acetyl CoA carboxylase activity, monooxygenase activity, monooxygenase reductase activity, fatty alcohol oxidase, acyl-CoA ligase, acyl-CoA oxidase, enoyl-CoA hydratase, 3-hydroxyacyl-CoA

30

dehydrogenase, and/or acyltransferase (e.g., acetyl-CoA C-acyltransferase) activities. In general, non-limiting examples of conditions that may be optimized include the type and amount of carbon source, the type and amount of nitrogen source, the carbon-to-nitrogen ratio, the oxygen level, growth temperature, pH, length of the biomass production phase, length of target product
5 accumulation phase, and time of cell harvest.

Culture media generally contain a suitable carbon source. Carbon sources useful for culturing microorganisms and/or fermentation processes sometimes are referred to as feedstocks. The term "feedstock" as used herein refers to a composition containing a carbon source that is provided to
10 an organism, which is used by the organism to produce energy and metabolic products useful for growth. A feedstock may be a natural substance, a "man-made substance," a purified or isolated substance, a mixture of purified substances, a mixture of unpurified substances or combinations thereof. A feedstock often is prepared by and/or provided to an organism by a person, and a feedstock often is formulated prior to administration to the organism. A carbon source may
15 comprise, but is not limited to including, one or more of the following substances: alkanes, alkenes, mono-carboxylic acids, di-carboxylic acids, monosaccharides (e.g., also referred to as "saccharides," which include 6-carbon sugars (e.g., glucose, fructose), 5-carbon sugars (e.g., xylose and other pentoses) and the like), disaccharides (e.g., lactose, sucrose), oligosaccharides (e.g., glycans, homopolymers of a monosaccharide), polysaccharides (e.g., starch, cellulose,
20 heteropolymers of monosaccharides or mixtures thereof), sugar alcohols (e.g., glycerol), and renewable feedstocks (e.g., cheese whey permeate, cornsteep liquor, sugar beet molasses, barley malt).

Carbon sources also can be selected from one or more of the following non-limiting examples:
25 paraffin (e.g., saturated paraffin, unsaturated paraffin, substituted paraffin, linear paraffin, branched paraffin, or combinations thereof); alkanes (e.g., dodecane), alkenes or alkynes, each of which may be linear, branched, saturated, unsaturated, substituted or combinations thereof (described in greater detail below); linear or branched alcohols (e.g., dodecanol); fatty acids (e.g., about 1 carbon to about 60 carbons, including free fatty acids, soap stock, for example); esters of fatty
30 acids; monoglycerides; diglycerides; triglycerides, phospholipids. Non-limiting commercial sources of products for preparing feedstocks include plants, plant oils or plant products (e.g., vegetable oils (e.g., almond oil, canola oil, cocoa butter, coconut oil, corn oil, cottonseed oil, flaxseed oil, grape seed oil, illipe, olive oil, palm oil, palm olein, palm kernel oil, safflower oil, peanut oil, soybean oil, sesame oil, shea nut oil, sunflower oil, walnut oil, the like and combinations thereof) and animal fats

(e.g., beef tallow, butterfat, lard, cod liver oil). A carbon source may include a petroleum product and/or a petroleum distillate (e.g., diesel, fuel oils, gasoline, kerosene, paraffin wax, paraffin oil, petrochemicals). In some embodiments, a feedstock comprises petroleum distillate. A carbon source can be a fatty acid distillate (e.g., a palm oil distillate or corn oil distillate). Fatty acid
5 distillates can be by-products from the refining of crude plant oils. In some embodiments, a feedstock comprises a fatty acid distillate.

In some embodiments, a feedstock comprises a soapstock (i.e. soap stock). A widely practiced method for purifying crude vegetable oils for edible use is the alkali or caustic refining method. This
10 process employs a dilute aqueous solution of caustic soda to react with the free fatty acids present which results in the formation of soaps. The soaps together with hydrated phosphatides, gums and prooxidant metals are typically separated from the refined oil as the heavy phase discharge from the refining centrifuge and are typically known as soapstock.

15 A carbon source also may include a metabolic product that can be used directly as a metabolic substrate in an engineered pathway described herein, or indirectly via conversion to a different molecule using engineered or native biosynthetic pathways in an engineered microorganism. In certain embodiments, metabolic pathways can be preferentially biased towards production of a desired product by increasing the levels of one or more activities in one or more metabolic
20 pathways having and/or generating at least one common metabolic and/or synthetic substrate. In some embodiments, a metabolic byproduct (e.g., fatty acid) of an engineered activity (e.g., omega oxidation activity) can be used in one or more metabolic pathways selected from gluconeogenesis, pentose phosphate pathway, glycolysis, fatty acid synthesis, beta oxidation, and omega oxidation, to generate a carbon source that can be converted to a fatty dicarboxylic acid (e.g., octanedioic
25 acid, decanedioic acid, dodecanedioic acid, tetradecanedioic acid, hexadecanedioic acid, octadecanedioic acid, eicosanedioic acid).

The term "paraffin" as used herein refers to the common name for alkane hydrocarbons, independent of the source (e.g., plant derived, petroleum derived, chemically synthesized,
30 fermented by a microorganism), or carbon chain length. A carbon source sometimes comprises a paraffin, and in some embodiments, a paraffin is predominant in a carbon source (e.g., about 75%, 80%, 85%, 90% or 95% paraffin). A paraffin sometimes is saturated (e.g., fully saturated), sometimes includes one or more unsaturations (e.g., about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 unsaturations) and sometimes is substituted with one or more non-hydrogen substituents. Non-

limiting examples of non-hydrogen substituents include halo, acetyl, =O, =N-CN, =N-OR, =NR, OR, NR₂, SR, SO₂R, SO₂NR₂, NRSO₂R, NRCONR₂, NRCOOR, NRCOR, CN, COOR, CONR₂, OOCR, COR, and NO₂, where each R is independently H, C1-C8 alkyl, C2-C8 heteroalkyl, C1-C8 acyl, C2-C8 heteroacyl, C2-C8 alkenyl, C2-C8 heteroalkenyl, C2-C8 alkynyl, C2-C8 heteroalkynyl, C6-C10 aryl, or C5-C10 heteroaryl, and each R is optionally substituted with halo, =O, =N-CN, =N-OR', =NR', OR', NR'₂, SR', SO₂R', SO₂NR'₂, NR'SO₂R', NR'CONR'₂, NR'COOR', NR'COR', CN, COOR', CONR'₂, OOCR', COR', and NO₂, where each R' is independently H, C1-C8 alkyl, C2-C8 heteroalkyl, C1-C8 acyl, C2-C8 heteroacyl, C6-C10 aryl or C5-C10 heteroaryl.

10 In some embodiments a feedstock is selected according to the genotype and/or phenotype of the engineered microorganism to be cultured. For example, a feedstock rich in 12-carbon fatty acids, 12-carbon dicarboxylic acids or 12-carbon paraffins, or a mixture of 10, 12 and 14-carbon compounds can be useful for culturing yeast strains harboring an alteration that partially blocks beta oxidation by disrupting POX4 activity, as described herein. Non-limiting examples of carbon sources having 10 to 14 carbons include fats (e.g., coconut oil, palm kernel oil), paraffins (e.g., alkanes, alkenes, or alkynes) having 10 to 14 carbons, (e.g., dodecane (also referred to as adakane¹², bihexyl, dihexyl and duodecane); tetradecane), alkene and alkyne derivatives), fatty acids (dodecanoic acid, tetradecanoic acid), fatty alcohols (dodecanol, tetradecanol), the like, non-toxic substituted derivatives or combinations thereof.

20

A carbon source sometimes comprises an alkyl, alkenyl or alkynyl compound or molecule (e.g., a compound that includes an alkyl, alkenyl or alkynyl moiety (e.g., alkane, alkene, alkyne)). In certain embodiments, an alkyl, alkenyl or alkynyl molecule, or combination thereof, is predominant in a carbon source (e.g., about 75%, 80%, 85%, 90% or 95% of such molecules). As used herein, the terms "alkyl," "alkenyl" and "alkynyl" include straight-chain (referred to herein as "linear"), branched-chain (referred to herein as "non-linear"), cyclic monovalent hydrocarbyl radicals, and combinations of these, which contain only C and H atoms when they are unsubstituted. Non-limiting examples of alkyl moieties include methyl, ethyl, isobutyl, cyclohexyl, cyclopentylethyl, 2-propenyl, 3-butenyl, and the like. An alkyl that contains only C and H atoms and is unsubstituted sometimes is referred to as "saturated." An alkenyl or alkynyl generally is "unsaturated" as it contains one or more double bonds or triple bonds, respectively. An alkenyl can include any number of double bonds, such as 1, 2, 3, 4 or 5 double bonds, for example. An alkynyl can include any number of triple bonds, such as 1, 2, 3, 4 or 5 triple bonds, for example. Alkyl, alkenyl and alkynyl molecules sometimes contain between about 2 to about 60 carbon atoms (C). For

example, an alkyl, alkenyl and alkynyl molecule can include about 1 carbon atom, about 2 carbon atoms, about 3 carbon atoms, about 4 carbon atoms, about 5 carbon atoms, about 6 carbon atoms, about 7 carbon atoms, about 8 carbon atoms, about 9 carbon atoms, about 10 carbon atoms, about 12 carbon atoms, about 14 carbon atoms, about 16 carbon atoms, about 18 carbon atoms, about 20 carbon atoms, about 22 carbon atoms, about 24 carbon atoms, about 26 carbon atoms, about 28 carbon atoms, about 30 carbon atoms, about 32 carbon atoms, about 34 carbon atoms, about 36 carbon atoms, about 38 carbon atoms, about 40 carbon atoms, about 42 carbon atoms, about 44 carbon atoms, about 46 carbon atoms, about 48 carbon atoms, about 50 carbon atoms, about 52 carbon atoms, about 54 carbon atoms, about 56 carbon atoms, about 58 carbon atoms or about 60 carbon atoms. In some embodiments, paraffins can have a mean number of carbon atoms of between about 8 to about 18 carbon atoms (e.g., about 8 carbon atoms, about 9 carbon atoms, about 10 carbon atoms, about 11 carbon atoms, about 12 carbon atoms, about 13 carbon atoms, about 14 carbon atoms, about 15 carbon atoms, about 16 carbon atoms, about 17 carbon atoms and about 18 carbon atoms). A single group can include more than one type of multiple bond, or more than one multiple bond. Such groups are included within the definition of the term "alkenyl" when they contain at least one carbon-carbon double bond, and are included within the term "alkynyl" when they contain at least one carbon-carbon triple bond. Alkyl, alkenyl and alkynyl molecules include molecules that comprise an alkyl, alkenyl and/or alkynyl moiety, and include molecules that consist of an alkyl, alkenyl or alkynyl moiety (i.e., alkane, alkene and alkyne molecules).

Alkyl, alkenyl and alkynyl substituents sometimes contain 1-20C (alkyl) or 2-20C (alkenyl or alkynyl). They can contain about 8-20C or about 10-20C in some embodiments. A single group can include more than one type of multiple bond, or more than one multiple bond. Such groups are included within the definition of the term "alkenyl" when they contain at least one carbon-carbon double bond, and are included within the term "alkynyl" when they contain at least one carbon-carbon triple bond.

Alkyl, alkenyl and alkynyl groups or compounds sometimes are substituted to the extent that such substitution can be synthesized and can exist. Typical substituents include, but are not limited to, halo, acetyl, =O, =N-CN, =N-OR, =NR, OR, NR₂, SR, SO₂R, SO₂NR₂, NRSO₂R, NRCONR₂, NRCOOR, NRCOR, CN, COOR, CONR₂, OOCR, COR, and NO₂, where each R is independently H, C1-C8 alkyl, C2-C8 heteroalkyl, C1-C8 acyl, C2-C8 heteroacyl, C2-C8 alkenyl, C2-C8 heteroalkenyl, C2-C8 alkynyl, C2-C8 heteroalkynyl, C6-C11 aryl, or C5-C11 heteroaryl, and each R

is optionally substituted with halo, =O, =N-CN, =N-OR', =NR', OR', NR'₂, SR', SO₂R', SO₂NR'₂, NR'SO₂R', NR'CONR'₂, NR'COOR', NR'COR', CN, COOR', CONR'₂, OOCR', COR', and NO₂, where each R' is independently H, C1-C8 alkyl, C2-C8 heteroalkyl, C1-C8 acyl, C2-C8 heteroacyl, C6-C10 aryl or C5-C10 heteroaryl. Alkyl, alkenyl and alkynyl groups can also be substituted by

5 C1-C8 acyl, C2-C8 heteroacyl, C6-C10 aryl or C5-C10 heteroaryl, each of which can be substituted by the substituents that are appropriate for the particular group.

"Acetylene" or "acetyl" substituents are 2-10C alkynyl groups that are optionally substituted, and are of the formula -C≡C-Ri, where Ri is H or C1-C8 alkyl, C2-C8 heteroalkyl, C2-C8 alkenyl, C2-C8 heteroalkenyl, C2-C8 alkynyl, C2-C8 heteroalkynyl, C1-C8 acyl, C2-C8 heteroacyl, C6-C10 aryl, C5-C10 heteroaryl, C7-C12 arylalkyl, or C6-C12 heteroarylalkyl, and each Ri group is optionally substituted with one or more substituents selected from halo, =O, =N-CN, =N-OR', =NR', OR', NR'₂, SR', SO₂R', SO₂NR'₂, NR'SO₂R', NR'CONR'₂, NR'COOR', NR'COR', CN, COOR', CONR'₂, OOCR', COR', and NO₂, where each R' is independently H, C1-C6 alkyl, C2-C6 heteroalkyl, C1-C6 acyl, C2-C6 heteroacyl, C6-C10 aryl, C5-C10 heteroaryl, C7-12 arylalkyl, or C6-12 heteroarylalkyl, each of which is optionally substituted with one or more groups selected from halo, C1-C4 alkyl, C1-C4 heteroalkyl, C1-C6 acyl, C1-C6 heteroacyl, hydroxy, amino, and =O; and where two R' can be linked to form a 3-7 membered ring optionally containing up to three heteroatoms selected from N, O and S. In some embodiments, Ri of -C≡C-Ri is H or Me.

20

A carbon source sometimes comprises a heteroalkyl, heteroalkenyl and/or heteroalkynyl molecule or compound (e.g., comprises heteroalkyl, heteroalkenyl and/or heteroalkynyl moiety (e.g., heteroalkane, heteroalkene or heteroalkyne)). "Heteroalkyl", "heteroalkenyl", and "heteroalkynyl" and the like are defined similarly to the corresponding hydrocarbyl (alkyl, alkenyl and alkynyl) groups, but the 'hetero' terms refer to groups that contain one to three O, S or N heteroatoms or combinations thereof within the backbone; thus at least one carbon atom of a corresponding alkyl, alkenyl, or alkynyl group is replaced by one of the specified heteroatoms to form a heteroalkyl, heteroalkenyl, or heteroalkynyl group. The typical and sizes for heteroforms of alkyl, alkenyl and alkynyl groups are generally the same as for the corresponding hydrocarbyl groups, and the

25 substituents that may be present on the heteroforms are the same as those described above for the hydrocarbyl groups. For reasons of chemical stability, it is also understood that, unless otherwise specified, such groups do not include more than two contiguous heteroatoms except where an oxo group is present on N or S as in a nitro or sulfonyl group.

30

The term "alkyl" as used herein includes cycloalkyl and cycloalkylalkyl groups and compounds, the term "cycloalkyl" may be used herein to describe a carbocyclic non-aromatic compound or group that is connected via a ring carbon atom, and "cycloalkylalkyl" may be used to describe a carbocyclic non-aromatic compound or group that is connected to a molecule through an alkyl
5 linker. Similarly, "heterocyclyl" may be used to describe a non-aromatic cyclic group that contains at least one heteroatom as a ring member and that is connected to the molecule via a ring atom, which may be C or N; and "heterocyclalkyl" may be used to describe such a group that is connected to another molecule through a linker. The sizes and substituents that are suitable for the cycloalkyl, cycloalkylalkyl, heterocyclyl, and heterocyclalkyl groups are the same as those
10 described above for alkyl groups. As used herein, these terms also include rings that contain a double bond or two, as long as the ring is not aromatic.

A carbon source sometimes comprises an acyl compound or moiety (e.g., compound comprising an acyl moiety). As used herein, "acyl" encompasses groups comprising an alkyl, alkenyl, alkynyl,
15 aryl or arylalkyl radical attached at one of the two available valence positions of a carbonyl carbon atom, and heteroacyl refers to the corresponding groups where at least one carbon other than the carbonyl carbon has been replaced by a heteroatom chosen from N, O and S. Thus heteroacyl includes, for example, $-C(=O)OR$ and $-C(=O)NR_2$ as well as $-C(=O)$ -heteroaryl.

20 Acyl and heteroacyl groups are bonded to any group or molecule to which they are attached through the open valence of the carbonyl carbon atom. Typically, they are C1-C8 acyl groups, which include formyl, acetyl, pivaloyl, and benzoyl, and C2-C8 heteroacyl groups, which include methoxyacetyl, ethoxycarbonyl, and 4-pyridinoyl. The hydrocarbyl groups, aryl groups, and heteroforms of such groups that comprise an acyl or heteroacyl group can be substituted with the
25 substituents described herein as generally suitable substituents for each of the corresponding component of the acyl or heteroacyl group.

A carbon source sometimes comprises one or more aromatic moieties and/or heteroaromatic moieties. "Aromatic" moiety or "aryl" moiety refers to a monocyclic or fused bicyclic moiety having
30 the well-known characteristics of aromaticity; examples include phenyl and naphthyl. Similarly, "heteroaromatic" and "heteroaryl" refer to such monocyclic or fused bicyclic ring systems which contain as ring members one or more heteroatoms selected from O, S and N. The inclusion of a heteroatom permits aromaticity in 5 membered rings as well as 6 membered rings. Typical heteroaromatic systems include monocyclic C5-C6 aromatic groups such as pyridyl, pyrimidyl,

pyrazinyl, thienyl, furanyl, pyrrolyl, pyrazolyl, thiazolyl, oxazolyl, and imidazolyl and the fused bicyclic moieties formed by fusing one of these monocyclic groups with a phenyl ring or with any of the heteroaromatic monocyclic groups to form a C8-C10 bicyclic group such as indolyl, benzimidazolyl, indazolyl, benzotriazolyl, isoquinolyl, quinolyl, benzothiazolyl, benzofuranyl, pyrazolopyridyl, quinazolinyl, quinoxalyl, cinnolyl, and the like. Any monocyclic or fused ring bicyclic system which has the characteristics of aromaticity in terms of electron distribution throughout the ring system is included in this definition. It also includes bicyclic groups where at least the ring which is directly attached to the remainder of the molecule has the characteristics of aromaticity. Typically, the ring systems contain 5-12 ring member atoms. The monocyclic heteroaryls sometimes contain 5-6 ring members, and the bicyclic heteroaryls sometimes contain 8-10 ring members.

Aryl and heteroaryl moieties may be substituted with a variety of substituents including C1-C8 alkyl, C2-C8 alkenyl, C2-C8 alkynyl, C5-C12 aryl, C1-C8 acyl, and heteroforms of these, each of which can itself be further substituted; other substituents for aryl and heteroaryl moieties include halo, OR, NR₂, SR, SO₂R, SO₂NR₂, NRSO₂R, NRCONR₂, NRCOOR, NRCOR, CN, COOR, CONR₂, OOCR, COR, and NO₂, where each R is independently H, C1-C8 alkyl, C2-C8 heteroalkyl, C2-C8 alkenyl, C2-C8 heteroalkenyl, C2-C8 alkynyl, C2-C8 heteroalkynyl, C6-C10 aryl, C5-C10 heteroaryl, C7-C12 arylalkyl, or C6-C12 heteroarylalkyl, and each R is optionally substituted as described above for alkyl groups. The substituent groups on an aryl or heteroaryl group may be further substituted with the groups described herein as suitable for each type of such substituents or for each component of the substituent. Thus, for example, an arylalkyl substituent may be substituted on the aryl portion with substituents typical for aryl groups, and it may be further substituted on the alkyl portion with substituents as typical or suitable for alkyl groups.

Similarly, "arylalkyl" and "heteroarylalkyl" refer to aromatic and heteroaromatic ring systems, which are stand-alone molecules (e.g., benzene or substituted benzene, pyridine or substituted pyridine), or which are bonded to an attachment point through a linking group such as an alkylene, including substituted or unsubstituted, saturated or unsaturated, cyclic or acyclic linkers. A linker often is C1-C8 alkyl or a hetero form thereof. These linkers also may include a carbonyl group, thus making them able to provide substituents as an acyl or heteroacyl moiety. An aryl or heteroaryl ring in an arylalkyl or heteroarylalkyl group may be substituted with the same substituents described above for aryl groups. An arylalkyl group sometimes includes a phenyl ring optionally substituted with the groups defined above for aryl groups and a C1-C4 alkylene that is unsubstituted or is substituted

with one or two C1-C4 alkyl groups or heteroalkyl groups, where the alkyl or heteroalkyl groups can optionally cyclize to form a ring such as cyclopropane, dioxolane, or oxacyclopentane. Similarly, a heteroarylalkyl group often includes a C5-C6 monocyclic heteroaryl group optionally substituted with one or more of the groups described above as substituents typical on aryl groups and a C1-C4 alkylene that is unsubstituted. A heteroarylalkyl group sometimes is substituted with one or two C1-C4 alkyl groups or heteroalkyl groups, or includes an optionally substituted phenyl ring or C5-C6 monocyclic heteroaryl and a C1-C4 heteroalkylene that is unsubstituted or is substituted with one or two C1-C4 alkyl or heteroalkyl groups, where the alkyl or heteroalkyl groups can optionally cyclize to form a ring such as cyclopropane, dioxolane, or oxacyclopentane.

Where an arylalkyl or heteroarylalkyl group is described as optionally substituted, the substituents may be on the alkyl or heteroalkyl portion or on the aryl or heteroaryl portion of the group. The substituents optionally present on the alkyl or heteroalkyl portion sometimes are the same as those described above for alkyl groups, and the substituents optionally present on the aryl or heteroaryl portion often are the same as those described above for aryl groups generally.

“Arylalkyl” groups as used herein are hydrocarbyl groups if they are unsubstituted, and are described by the total number of carbon atoms in the ring and alkylene or similar linker. Thus a benzyl group is a C7-arylalkyl group, and phenylethyl is a C8-arylalkyl.

“Heteroarylalkyl” as described above refers to a moiety comprising an aryl group that is attached through a linking group, and differs from “arylalkyl” in that at least one ring atom of the aryl moiety or one atom in the linking group is a heteroatom selected from N, O and S. The heteroarylalkyl groups are described herein according to the total number of atoms in the ring and linker combined, and they include aryl groups linked through a heteroalkyl linker; heteroaryl groups linked through a hydrocarbyl linker such as an alkylene; and heteroaryl groups linked through a heteroalkyl linker. Thus, for example, C7-heteroarylalkyl includes pyridylmethyl, phenoxy, and N-pyrrolylmethoxy.

“Alkylene” as used herein refers to a divalent hydrocarbyl group. Because an alkylene is divalent, it can link two other groups together. An alkylene often is referred to as $-(CH_2)_n-$ where n can be 1-20, 1-10, 1-8, or 1-4, though where specified, an alkylene can also be substituted by other groups, and can be of other lengths, and the open valences need not be at opposite ends of a chain. Thus $-CH(Me)-$ and $-C(Me)_2-$ may also be referred to as alkenes, as can a cyclic group such as

cyclopropan-1,1-diyl. Where an alkylene group is substituted, the substituents include those typically present on alkyl groups as described herein.

- In some embodiments, a feedstock includes a mixture of carbon sources, where each carbon source in the feedstock is selected based on the genotype of the engineered microorganism. In certain embodiments, a mixed carbon source feedstock includes one or more carbon sources selected from sugars, cellulose, alkanes, fatty acids, triacylglycerides, paraffins, the like and combinations thereof.
- 5 Nitrogen may be supplied from an inorganic (e.g., (NH₄)₂SO₄) or organic source (e.g., urea or glutamate). In addition to appropriate carbon and nitrogen sources, culture media also can contain suitable minerals, salts, cofactors, buffers, vitamins, metal ions (e.g., Mn^{sup.+2}, Co^{sup.+2}, Zn^{sup.+2}, Mg^{sup.+2}) and other components suitable for culture of microorganisms.
- 10 Engineered microorganisms sometimes are cultured in complex media (e.g., yeast extract-peptone-dextrose broth (YPD)). In some embodiments, engineered microorganisms are cultured in a defined minimal media that lacks a component necessary for growth and thereby forces selection of a desired expression cassette (e.g., Yeast Nitrogen Base (DIFCO Laboratories, Detroit, Mich.)).
- 15 Culture media in some embodiments are common commercially prepared media, such as Yeast Nitrogen Base (DIFCO Laboratories, Detroit, Mich.). Other defined or synthetic growth media may also be used and the appropriate medium for growth of the particular microorganism are known. A variety of host organisms can be selected for the production of engineered microorganisms. Non-limiting examples include yeast (e.g., *Candida tropicalis* (e.g., ATCC20336, ATCC20913,
- 20 ATCC20962), *Yarrowia lipolytica* (e.g., ATCC20228)) and filamentous fungi (e.g., *Aspergillus nidulans* (e.g., ATCC38164) and *Aspergillus parasiticus* (e.g., ATCC 24690)). In specific embodiments, yeast are cultured in YPD media (10 g/L Bacto Yeast Extract, 20 g/L Bacto Peptone, and 20 g/L Dextrose). Filamentous fungi, in particular embodiments, are grown in CM (Complete Medium) containing 10 g/L Dextrose, 2 g/L Bacto Peptone, 1g/L Bacto Yeast Extract, 1 g/L
- 25 Casamino acids, 50 mL /L 20X Nitrate Salts (120 g/L NaNO₃, 10.4 g/L KCl, 10.4 g/L MgSO₄·7 H₂O), 1 mL/L 1000X Trace Elements (22 g/L ZnSO₄·7 H₂O, 11 g/L H₃BO₃, 5 g/L MnCl₂·7 H₂O, 5 g/L FeSO₄·7 H₂O, 1.7 g/L CoCl₂·6 H₂O, 1.6 g/L CuSO₄·5 H₂O, 1.5 g/L Na₂MoO₄·2 H₂O, and 50 g/L Na₄EDTA), and 1 mL/L Vitamin Solution (100 mg each of Biotin, pyridoxine, thiamine, riboflavin, p-aminobenzoic acid, and nicotinic acid in 100 mL water).

Growth Conditions & Fermentation

A suitable pH range for the fermentation often is between about pH 4.0 to about pH 8.0, where a pH in the range of about pH 5.5 to about pH 7.0 sometimes is utilized for initial culture conditions.

5 Depending on the host organism, culturing may be conducted under aerobic or anaerobic conditions, where microaerobic conditions sometimes are maintained. A two-stage process may be utilized, where one stage promotes microorganism proliferation and another state promotes production of target molecule. In a two-stage process, the first stage may be conducted under aerobic conditions (e.g., introduction of air and/or oxygen) and the second stage may be conducted

10 under anaerobic conditions (e.g., air or oxygen are not introduced to the culture conditions). In some embodiments, the first stage may be conducted under anaerobic conditions and the second stage may be conducted under aerobic conditions. In certain embodiments, a two-stage process may include two more organisms, where one organism generates an intermediate product in one stage and another organism processes the intermediate product into a target fatty dicarboxylic acid

15 product (e.g., sebacic or dodecanedioic acid) in another stage, for example.

A variety of fermentation processes may be applied for commercial biological production of a target fatty dicarboxylic acid product. In some embodiments, commercial production of a target fatty dicarboxylic acid product from a recombinant microbial host is conducted using a batch, fed-batch

20 or continuous fermentation process, for example.

A batch fermentation process often is a closed system where the media composition is fixed at the beginning of the process and not subject to further additions beyond those required for maintenance of pH and oxygen level during the process. At the beginning of the culturing process

25 the media is inoculated with the desired organism and growth or metabolic activity is permitted to occur without adding additional sources (i.e., carbon and nitrogen sources) to the medium. In batch processes the metabolite and biomass compositions of the system change constantly up to the time the culture is terminated. In a typical batch process, cells proceed through a static lag phase to a high-growth log phase and finally to a stationary phase, wherein the growth rate is

30 diminished or halted. Left untreated, cells in the stationary phase will eventually die.

A variation of the standard batch process is the fed-batch process, where the carbon source is continually added to the fermentor over the course of the fermentation process. Fed-batch processes are useful when catabolite repression is apt to inhibit the metabolism of the cells or

where it is desirable to have limited amounts of carbon source in the media at any one time. Measurement of the carbon source concentration in fed-batch systems may be estimated on the basis of the changes of measurable factors such as pH, dissolved oxygen and the partial pressure of waste gases (e.g., CO.sub.2).

5

Batch and fed-batch culturing methods are known in the art. Examples of such methods may be found in Thomas D. Brock in *Biotechnology: A Textbook of Industrial Microbiology*, 2nd ed., (1989) Sinauer Associates Sunderland, Mass. and Deshpande, Mukund V., *Appl. Biochem. Biotechnol.*, 36:227 (1992).

10

In continuous fermentation process a defined media often is continuously added to a bioreactor while an equal amount of culture volume is removed simultaneously for product recovery. Continuous cultures generally maintain cells in the log phase of growth at a constant cell density. Continuous or semi-continuous culture methods permit the modulation of one factor or any number of factors that affect cell growth or end product concentration. For example, an approach may limit the carbon source and allow all other parameters to moderate metabolism. In some systems, a number of factors affecting growth may be altered continuously while the cell concentration, measured by media turbidity, is kept constant. Continuous systems often maintain steady state growth and thus the cell growth rate often is balanced against cell loss due to media being drawn off the culture. Methods of modulating nutrients and growth factors for continuous culture processes, as well as techniques for maximizing the rate of product formation, are known and a variety of methods are detailed by Brock, *supra*.

15

20

25

30

In some embodiments involving fermentation, the fermentation can be carried out using two or more microorganisms (e.g., host microorganism, engineered microorganism, isolated naturally occurring microorganism, the like and combinations thereof), where a feedstock is partially or completely utilized by one or more organisms in the fermentation (e.g., mixed fermentation), and the products of cellular respiration or metabolism of one or more organisms can be further metabolized by one or more other organisms to produce a desired target product (e.g., sebacic acid, dodecanedioic acid, hexanoic acid). In certain embodiments, each organism can be fermented independently and the products of cellular respiration or metabolism purified and contacted with another organism to produce a desired target product. In some embodiments, one or more organisms are partially or completely blocked in a metabolic pathway (e.g., beta oxidation, omega oxidation, the like or combinations thereof), thereby producing a desired product that can

be used as a feedstock for one or more other organisms. Any suitable combination of microorganisms can be utilized to carry out mixed fermentation or sequential fermentation.

Target Product Production, Isolation and Yield

5

In various embodiments a fatty dicarboxylic acid (e.g., octanedioic acid, decanedioic acid, dodecanedioic acid, tetradecanedioic acid, hexadecanedioic acid, octadecanedioic acid, eicosanedioic acid) is isolated or purified from the culture media or extracted from the engineered microorganisms. In some embodiments, fermentation of feedstocks by methods described herein
 10 can produce a target fatty dicarboxylic acid product (e.g., sebacic or dodecanedioic acid) at a level of about 10% to about 100% of theoretical yield (e.g., about 15%, about 20%, about 25% or more of theoretical yield (e.g., 25% or more, 26% or more, 27% or more, 28% or more, 29% or more, 30% or more, 31% or more, 32% or more, 33% or more, 34% or more, 35% or more, 36% or more, 37% or more, 38% or more, 39% or more, 40% or more, 41% or more, 42% or more, 43% or more,
 15 44% or more, 45% or more, 46% or more, 47% or more, 48% or more, 49% or more, 50% or more, 51% or more, 52% or more, 53% or more, 54% or more, 55% or more, 56% or more, 57% or more, 58% or more, 59% or more, 60% or more, 61% or more, 62% or more, 63% or more, 64% or more, 65% or more, 66% or more, 67% or more, 68% or more, 69% or more, 70% or more, 71% or more, 72% or more, 73% or more, 74% or more, 75% or more, 76% or more, 77% or more, 78% or more,
 20 79% or more, 80% or more, 81% or more, 82% or more, 83% or more, 84% or more, 85% or more, 86% or more, 87% or more, 88% or more, 89% or more, 90% or more, 91% or more, 92% or more, 93% or more, 94% or more, 95% or more, 96% or more, 97% or more, 98% or more, or 99% or more of theoretical yield). The term "theoretical yield" as used herein refers to the amount of product that could be made from a starting material if the reaction is 100% complete. Theoretical
 25 yield is based on the stoichiometry of a reaction and ideal conditions in which starting material is completely consumed, undesired side reactions do not occur, the reverse reaction does not occur, and there are no losses in the work-up procedure. Culture media may be tested for target product (e.g., sebacic or dodecanedioic acid) concentration and drawn off when the concentration reaches a predetermined level. Detection methods are known in the art, including but not limited to
 30 chromatographic methods (e.g., gas chromatography) or combined chromatographic/mass spectrometry (e.g., GC-MS) methods. Target product (e.g., sebacic or dodecanedioic acid) may be present at a range of levels as described herein.

A target fatty dicarboxylic acid product sometimes is retained within an engineered microorganism after a culture process is completed, and in certain embodiments, the target product is secreted out of the microorganism into the culture medium. For the latter embodiments, (i) culture media may be drawn from the culture system and fresh medium may be supplemented, and/or (ii) target
5 product may be extracted from the culture media during or after the culture process is completed. Engineered microorganisms may be cultured on or in solid, semi-solid or liquid media. In some embodiments media is drained from cells adhering to a plate. In certain embodiments, a liquid-cell mixture is centrifuged at a speed sufficient to pellet the cells but not disrupt the cells and allow extraction of the media, as known in the art. The cells may then be resuspended in fresh media.
10 Target product may be purified from culture media according to known methods known in the art.

Provided herein are non-limiting examples of methods useful for recovering target product from fermentation broth and/or isolating/partially purifying a target fatty dicarboxylic acid product from non-target products when utilizing mixed chain length feedstocks. Recovery of a fatty dicarboxylic
15 acid (e.g., octanedioic acid, decanedioic acid, dodecanedioic acid, tetradecanedioic acid, hexadecanedioic acid, octadecanedioic acid, eicosanedioic acid) from fermentation broth can be accomplished using a variety of methods. Optionally, one can first employ a centrifugation step to separate cell mass and a fatty dicarboxylic acid (e.g., octanedioic acid, decanedioic acid, dodecanedioic acid, tetradecanedioic acid, hexadecanedioic acid, octadecanedioic acid,
20 eicosanedioic acid) from the aqueous phase. A fatty dicarboxylic acid (e.g., octanedioic acid, decanedioic acid, dodecanedioic acid, tetradecanedioic acid, hexadecanedioic acid, octadecanedioic acid, eicosanedioic acid) has limited solubility in water under fermentation conditions, and has a density similar to that of water. Upon centrifugation, the majority of fatty dicarboxylic acid (e.g., octanedioic acid, decanedioic acid, dodecanedioic acid, tetradecanedioic
25 acid, hexadecanedioic acid, octadecanedioic acid, eicosanedioic acid) will be pulled away from the water stream, and be concentrated in the cell mass stream. The concentrated fatty dicarboxylic acid stream will then be further concentrated via filtration steps (e.g., solid dodecanedioic acid will be retained on a filter, allowing water to pass through, concentrating the product). Once the fatty dicarboxylic acid (e.g., octanedioic acid, decanedioic acid, dodecanedioic acid, tetradecanedioic
30 acid, hexadecanedioic acid, octadecanedioic acid, eicosanedioic acid) is concentrated to the desired level, the temperature will be increased to above its melting point of 130 C. After the fatty dicarboxylic acid is melted, the remaining impurities are removed via filtration; the final product is recovered by decreasing the temperature, allowing the fatty dicarboxylic acid to solidify, and collecting the solid product.

Alternatively, a fatty dicarboxylic acid (e.g., octanedioic acid, decanedioic acid, dodecanedioic acid, tetradecanedioic acid, hexadecanedioic acid, octadecanedioic acid, eicosanedioic acid) can be recovered from fermentation broth by first extracting the broth with an organic solvent in which a fatty dicarboxylic acid (e.g., octanedioic acid, decanedioic acid, dodecanedioic acid, tetradecanedioic acid, hexadecanedioic acid, octadecanedioic acid, eicosanedioic acid) is soluble (e.g., ethanol). The organic solvent phase can then be filtered through various membranes to further purify the fatty dicarboxylic acid. Subsequent extractions with the same or a different organic solvent can then be performed and each round of extraction can be followed by membrane filtration to further concentrate the fatty dicarboxylic acid. The organic solvent can be evaporated, leaving the fatty dicarboxylic acid behind as a residue and the residue can be dried to provide the fatty dicarboxylic acid (e.g., octanedioic acid, decanedioic acid, dodecanedioic acid, tetradecanedioic acid, hexadecanedioic acid, octadecanedioic acid, eicosanedioic acid) in solid form.

In certain embodiments, target product is extracted from the cultured engineered microorganisms. The microorganism cells may be concentrated through centrifugation at a speed sufficient to shear the cell membranes. In some embodiments, the cells may be physically disrupted (e.g., shear force, sonication) or chemically disrupted (e.g., contacted with detergent or other lysing agent). The phases may be separated by centrifugation or other method known in the art and target product may be isolated according to known methods.

Commercial grade target product sometimes is provided in substantially pure form (e.g., 90% pure or greater, 95% pure or greater, 99% pure or greater or 99.5% pure or greater). In some embodiments, target product may be modified into any one of a number of downstream products.

For example, a fatty dicarboxylic acid (e.g., octanedioic acid, decanedioic acid, dodecanedioic acid, tetradecanedioic acid, hexadecanedioic acid, octadecanedioic acid, eicosanedioic acid) may be polycondensed with hexamethylenediamine to produce nylon. Nylon may be further processed into fibers for applications in carpeting, automobile tire cord and clothing. A fatty dicarboxylic acid (e.g., octanedioic acid, decanedioic acid, dodecanedioic acid, tetradecanedioic acid, hexadecanedioic acid, octadecanedioic acid, eicosanedioic acid) is also used for manufacturing plasticizers, lubricant components and polyester polyols for polyurethane systems. Various esters of food grade fatty dicarboxylic acids (e.g., octanedioic acid, decanedioic acid, dodecanedioic acid, tetradecanedioic acid, hexadecanedioic acid, octadecanedioic acid, eicosanedioic acid) are used as components in fragrance manufacture, gelling aids, flavorings, acidulant, leavening and

buffering agent. A fatty dicarboxylic acid (e.g., octanedioic acid, decanedioic acid, dodecanedioic acid, tetradecanedioic acid, hexadecanedioic acid, octadecanedioic acid, eicosanedioic acid) has two carboxylic acid (-COOH) groups, which can yield two kinds of salts. Its derivatives, acyl halides, anhydrides, esters, amides and nitriles, are used in making a variety of downstream products through further reactions of substitution, catalytic reduction, metal hydride reduction, diborane reduction, keto formation with organometallic reagents, electrophile bonding at oxygen, and condensation.

Target product may be provided within cultured microbes containing target product, and cultured microbes may be supplied fresh or frozen in a liquid media or dried. Fresh or frozen microbes may be contained in appropriate moisture-proof containers that may also be temperature controlled as necessary. Target product sometimes is provided in culture medium that is substantially cell-free. In some embodiments target product or modified target product purified from microbes is provided, and target product sometimes is provided in substantially pure form. In certain embodiments crystallized or powdered target product is provided. Dodecanedioic acid (1,12 dodecanedioic acid; DDDA) is a white powder or crystal with a melting point of between 260°F and 266°F. Sebacic acid (1,8 octanedicarboxylic acid) is also a white powder or crystal with a melting point of between 268°F and 274°F. A crystallized or powdered fatty dicarboxylic acid (e.g., octanedioic acid, decanedioic acid, dodecanedioic acid, tetradecanedioic acid, hexadecanedioic acid, octadecanedioic acid, eicosanedioic acid) may be transported in a variety of containers including one ton cartons, drums, 50 pound bags and the like.

In certain embodiments, a fatty dicarboxylic acid target product (e.g., dodecanedioic acid or sebacic acid) is produced with a yield of about 0.50 grams of target product per gram of feedstock added, or greater; 0.51 grams of target product per gram of feedstock added, or greater; 0.52 grams of target product per gram of feedstock added, or greater; 0.53 grams of target product per gram of feedstock added, or greater; 0.54 grams of target product per gram of feedstock added, or greater; 0.55 grams of target product per gram of feedstock added, or greater; 0.56 grams of target product per gram of feedstock added, or greater; 0.57 grams of target product per gram of feedstock added, or greater; 0.58 grams of target product per gram of feedstock added, or greater; 0.59 grams of target product per gram of feedstock added, or greater; 0.60 grams of target product per gram of feedstock added, or greater; 0.61 grams of target product per gram of feedstock added, or greater; 0.62 grams of target product per gram of feedstock added, or greater; 0.63 grams of target product per gram of feedstock added, or greater; 0.64 grams of target product per

gram of feedstock added, or greater; 0.65 grams of target product per gram of feedstock added, or greater; 0.66 grams of target product per gram of feedstock added, or greater; 0.67 grams of target product per gram of feedstock added, or greater; 0.68 grams of target product per gram of feedstock added, or greater; 0.69 grams of target product per gram of feedstock added, or greater; 0.70 grams of target product per gram of feedstock added or greater; 0.71 grams of target product per gram of feedstock added, or greater; 0.72 grams of target product per gram of feedstock added, or greater; 0.73 grams of target product per gram of feedstock added, or greater; 0.74 grams of target product per gram of feedstock added, or greater; 0.75 grams of target product per gram of feedstock added, or greater; 0.76 grams of target product per gram of feedstock added, or greater; 0.77 grams of target product per gram of feedstock added, or greater; 0.78 grams of target product per gram of feedstock added, or greater; 0.79 grams of target product per gram of feedstock added, or greater; 0.80 grams of target product per gram of feedstock added, or greater; 0.81 grams of target product per gram of feedstock added, or greater; 0.82 grams of target product per gram of feedstock added, or greater; 0.83 grams of target product per gram of feedstock added, or greater; 0.84 grams of target product per gram of feedstock added, or greater; 0.85 grams of target product per gram of feedstock added, or greater; 0.86 grams of target product per gram of feedstock added, or greater; 0.87 grams of target product per gram of feedstock added, or greater; 0.88 grams of target product per gram of feedstock added, or greater; 0.89 grams of target product per gram of feedstock added, or greater; 0.90 grams of target product per gram of feedstock added, or greater; 0.91 grams of target product per gram of feedstock added, or greater; 0.92 grams of target product per gram of feedstock added, or greater; 0.93 grams of target product per gram of feedstock added, or greater; 0.94 grams of target product per gram of feedstock added, or greater; 0.95 grams of target product per gram of feedstock added, or greater; 0.96 grams of target product per gram of feedstock added, or greater; 0.97 grams of target product per gram of feedstock added, or greater; 0.98 grams of target product per gram of feedstock added, or greater; 0.99 grams of target product per gram of feedstock added, or greater; 1.0 grams of target product per gram of feedstock added, or greater; 1.1 grams of target product per gram of feedstock added, or greater; 1.2 grams of target product per gram of feedstock added, or greater; 1.3 grams of target product per gram of feedstock added, or greater; 1.4 grams of target product per gram of feedstock added, or greater; or about 1.5 grams of target product per gram of feedstock added, or greater.

In certain embodiments, a fatty dicarboxylic acid (e.g., octanedioic acid, decanedioic acid, dodecanedioic acid, tetradecanedioic acid, hexadecanedioic acid, octadecanedioic acid,

eicosanedioic acid) is produced with a yield of greater than about 0.15 grams per gram of the feedstock (e.g., dodecane, mixed chain length alkanes, lauric acid, mixed chain length fatty acids, oil, the like or combinations of the foregoing). In some embodiments, a fatty dicarboxylic acid (e.g., octanedioic acid, decanedioic acid, dodecanedioic acid, tetradecanedioic acid, hexadecanedioic acid, octadecanedioic acid, eicosanedioic acid) is produced at between about 10% and about 100% of maximum theoretical yield of any introduced feedstock ((e.g., about 15%, about 20%, about 25% or more of theoretical yield (e.g., 25% or more, 26% or more, 27% or more, 28% or more, 29% or more, 30% or more, 31% or more, 32% or more, 33% or more, 34% or more, 35% or more, 36% or more, 37% or more, 38% or more, 39% or more, 40% or more, 41% or more, 42% or more, 43% or more, 44% or more, 45% or more, 46% or more, 47% or more, 48% or more, 49% or more, 50% or more, 51% or more, 52% or more, 53% or more, 54% or more, 55% or more, 56% or more, 57% or more, 58% or more, 59% or more, 60% or more, 61% or more, 62% or more, 63% or more, 64% or more, 65% or more, 66% or more, 67% or more, 68% or more, 69% or more, 70% or more, 71% or more, 72% or more, 73% or more, 74% or more, 75% or more, 76% or more, 77% or more, 78% or more, 79% or more, 80% or more, 81% or more, 82% or more, 83% or more, 84% or more, 85% or more, 86% or more, 87% or more, 88% or more, 89% or more, 90% or more, 91% or more, 92% or more, 93% or more, 94% or more, 95% or more, 96% or more, 97% or more, 98% or more, or 99% or more of theoretical maximum yield). In certain embodiments, a fatty dicarboxylic acid (e.g., octanedioic acid, decanedioic acid, dodecanedioic acid, tetradecanedioic acid, hexadecanedioic acid, octadecanedioic acid, eicosanedioic acid) is produced in a concentration range of between about 50 g/L to about 1000g/L of culture media (e.g., about 50 g/L, about 55 g/L, about 60 g/L, about 65 g/L, about 70 g/L, about 75 g/L, about 80 g/L, about 85 g/L, about 90 g/L, about 95 g/L, about 100 g/L, about 110 g/L, about 120 g/L, about 130 g/L, about 140 g/L, about 150 g/L, about 160 g/L, about 170 g/L, about 180 g/L, about 190 g/L, about 200 g/L, about 225 g/L, about 250 g/L, about 275 g/L, about 300 g/L, about 325 g/L, about 350 g/L, about 375 g/L, about 400 g/L, about 425 g/L, about 450 g/L, about 475 g/L, about 500 g/L, about 550 g/L, about 600 g/L, about 650 g/L, about 700 g/L, about 750 g/L, about 800 g/L, about 850 g/L, about 900 g/L, about 950 g/L, or about 1000 g/L).

In some embodiments, a fatty dicarboxylic acid (e.g., octanedioic acid, decanedioic acid, dodecanedioic acid, tetradecanedioic acid, hexadecanedioic acid, octadecanedioic acid, eicosanedioic acid) is produced at a rate of between about 0.5 g/L/hour to about 5 g/L/hour (e.g., about 0.5 g/L/hour, about 0.6 g/L/hour, about 0.7 g/L/hour, about 0.8 g/L/hour, about 0.9 g/L/hour, about 1.0 g/L/hour, about 1.1 g/L/hour, about 1.2 g/L/hour, about 1.3 g/L/hour, about 1.4 g/L/hour,

about 1.5 g/L/hour, about 1.6 g/L/hour, about 1.7 g/L/hour, about 1.8 g/L/hour, about 1.9 g/L/hour, about 2.0 g/L/hour, about 2.25 g/L/hour, about 2.5 g/L/hour, about 2.75 g/L/hour, about 3.0 g/L/hour, about 3.25 g/L/hour, about 3.5 g/L/hour, about 3.75 g/L/hour, about 4.0 g/L/hour, about 4.25 g/L/hour, about 4.5 g/L/hour, about 4.75 g/L/hour, or about 5.0 g/L/hour.) In certain, 5
embodiments, the engineered organism comprises between about a 5-fold to about a 500-fold increase in a fatty dicarboxylic acid (e.g., octanedioic acid, decanedioic acid, dodecanedioic acid, tetradecanedioic acid, hexadecanedioic acid, octadecanedioic acid, eicosanedioic acid) production when compared to wild-type or partially engineered organisms of the same strain, under identical fermentation conditions (e.g., about a 5-fold increase, about a 10-fold increase, about a 15-fold 10
increase, about a 20-fold increase, about a 25-fold increase, about a 30-fold increase, about a 35-fold increase, about a 40-fold increase, about a 45-fold increase, about a 50-fold increase, about a 55-fold increase, about a 60-fold increase, about a 65-fold increase, about a 70-fold increase, about a 75-fold increase, about a 80-fold increase, about a 85-fold increase, about a 90-fold increase, about a 95-fold increase, about a 100-fold increase, about a 125-fold increase, about a 150-fold increase, about a 175-fold increase, about a 200-fold increase, about a 250-fold increase, 15
about a 300-fold increase, about a 350-fold increase, about a 400-fold increase, about a 450-fold increase, or about a 500-fold increase).

In certain embodiments, the maximum theoretical yield (Y_{max}) of dodecanedioic acid in a fully beta-oxidation blocked engineered microorganism is about 1.15 grams of dodecanedioic acid produced 20
per gram of lauric acid added. In some embodiments, the maximum theoretical yield (Y_{max}) of dodecanedioic acid in a fully beta-oxidation blocked engineered microorganism is about 1.07 grams of dodecanedioic acid produced per gram of methyl laurate added. In certain embodiments, the maximum theoretical yield (Y_{max}) of dodecanedioic acid in a partially beta-oxidation blocked 25
engineered microorganism is about 0.82 grams of dodecanedioic acid produced per gram of oleic acid added. In some embodiments, the maximum theoretical yield (Y_{max}) of dodecanedioic acid in a partially beta-oxidation blocked engineered microorganism is about 0.95 grams of dodecanedioic acid produced per gram of coconut oil added. The percentage of Y_{max} for the engineered microorganism under conditions in which dodecanedioic acid is produced is calculated as ($\%Y_{max}$) 30
= $Y_{p/s} / Y_{max} * 100$, where ($Y_{p/s}$) = [dodecanedioic acid (g/L)] * final volume of culture in flask (L)] / [feedstock added to flask (g)]. In some embodiments, the engineered microorganism produces dodecanedioic acid at about 10% to about 100% of maximum theoretical yield.

In certain embodiments, the maximum theoretical yield ($Y_{\max}$) of sebacic acid in a fully beta-oxidation blocked engineered microorganism is about 1.42 grams of sebacic acid produced per gram of decane added. In some embodiments, the maximum theoretical yield ($Y_{\max}$) of sebacic acid in a fully beta-oxidation blocked engineered microorganism is about 1.17 grams of sebacic acid produced per gram of capric acid added. In certain embodiments, the maximum theoretical yield ($Y_{\max}$) of sebacic acid in a partially beta-oxidation blocked engineered microorganism is about 0.83 grams of sebacic acid produced per gram of coconut oil added. In some embodiments, the maximum theoretical yield ($Y_{\max}$) of sebacic acid in a partially beta-oxidation blocked engineered microorganism is about 0.72 grams of sebacic acid produced per gram of oleic acid added. The percentage of $Y_{\max}$ for the engineered microorganism under conditions in which sebacic acid is produced is calculated as $(\%Y_{\max}) = Y_{p/s} / Y_{\max} * 100$, where $(Y_{p/s}) = [\text{sebacic acid (g/L)}] * \text{final volume of culture in flask (L)} / [\text{feedstock added to flask (g)}]$. In some embodiments, the engineered microorganism produces sebacic acid at about 10% to about 100% of maximum theoretical yield.

15

Examples

The examples set forth below illustrate certain embodiments and do not limit the technology. Certain examples set forth below utilize standard recombinant DNA and other biotechnology protocols known in the art. Many such techniques are described in detail in Maniatis, T., E. F. Fritsch and J. Sambrook (1982) *Molecular Cloning: a Laboratory Manual*; Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y. DNA mutagenesis can be accomplished using the Stratagene (San Diego, CA) "QuickChange" kit according to the manufacturer's instructions.

Non-limiting examples of recombinant DNA techniques and genetic manipulation of microorganisms are described herein. In some embodiments, strains of engineered organisms described herein are mated to combine genetic backgrounds to further enhance carbon flux management through native and/or engineered pathways described herein, for the production of a desired target product (e.g., sebacic or dodecanedioic acid).

30

Example 1: Conversion of decane to sebacic acid in shake flask fermentation

50mL of SP92 medium (6.7g/L yeast nitrogen base, 3.0 g/L yeast extract, 3.0 g/L $(\text{NH}_4)_2\text{SO}_4$, 1.0 g/L K_2HPO_4 , 1.0 g/L KH_2PO_4 , 75 g/L dextrose) was inoculated with a single colony of a completely

beta-oxidation blocked strain of *Candida tropicalis* (ATCC20962) and the culture was grown overnight at 30°C, with shaking at about 300rpm. Cells were pelleted by centrifugation for 10 minutes at 4°C and 1,050 x g and the supernatant discarded. Cells were resuspended in 20mL TB- low nitrogen (low-N) media (1.7g/L yeast nitrogen base without ammonium sulfate, 3.0 g/L yeast extract, 1.0 g/L K₂HPO₄, 1.0 g/L KH₂PO₄) and transferred to a new sterile 250mL glass baffled flask and incubated at 30C, with shaking at about 250rpm, utilizing the following feeding schedule: dextrose fed to 0.1% at 0, 1, 2, 3, 4, and 5 hours, dextrose fed to 5% at 30 hours, decane fed to 0.7% at 0, 5, 30, and 48 hours. Samples were removed for gas chromatographic (GC) analysis at 0, 4, 30, and 72 hours. The GC profile showed that the culture accumulated the C10 dicarboxylic acid (sebacic acid) with very little accumulation of the C10 monocarboxylic acid (capric acid), as shown in FIG. 9. After 72 hours of incubation the concentration of sebacic acid was 0.94 g/L and the capric acid concentration was 0.01 g/L. There was no significant accumulation of any other monoacid or diacid.

15 *Example 2: Conversion of capric acid to sebacic acid in shake flask fermentation*

5 mL of SP92-glycerol medium (6.7 g/L yeast nitrogen base, 3.0 g/L yeast extract, 3.0 g/L (NH₄)₂SO₄, 1.0 g/L K₂HPO₄, 1.0 g/L KH₂PO₄, 75 g/L glycerol) was inoculated with a single colony of *Candida tropicalis* (ATCC20962) and the culture was grown overnight at 30°C, with shaking at about 250rpm. This starter culture was then used to inoculate 25 mL cultures in the same medium to an initial OD_{600nm} of 0.4 and grown overnight at 30°C, with shaking at about 300 rpm. Cells were pelleted by centrifugation for 10 minutes at 4°C and 1,050 x g and the supernatant discarded. Cells were resuspended in 12.5 mL TB-lowN media + glycerol (1.7 g/L yeast nitrogen base without ammonium sulfate, 3.0 g/L yeast extract, 1.0 g/L K₂HPO₄, 1.0 g/L KH₂PO₄, 75 g/L glycerol) and transferred to a new sterile 250 mL glass baffled flask. Cultures were fed 0.05% or 0.1% capric acid and incubated at 30°C, with shaking at about 300 rpm. After 24 hours incubation cultures were fed glycerol to 75 g/L and incubation continued before sampling for GC at 48 hours. GC analysis showed that nearly all capric acid was converted to sebacic acid under both starting concentrations of capric acid, as shown in FIG. 10.

30

Example 3: Fermentation procedure for conversion of decane to sebacic acid

Filter sterilized modified SP92-glycerol fermentation medium (6.7 g/L yeast nitrogen base, 3.0 g/L yeast extract, 3.0 g/L (NH₄)₂SO₄, 1.0 g/L K₂HPO₄, 1.0 g/L KH₂PO₄, 20 g/L glycerol) is

transferred to a sterile fermentation vessel. Growth of *Candida tropicalis* (ATCC20962) is inoculated to an initial OD_{600nm} of about 1.0 with a 5% inoculum and growth carried out under the following conditions: 30°C with shaking at about 1000 rpm, 1 volume per volume per minute aeration (vvm), pH 5.8 and initial volume of 0.3 L. Growth proceeds for approximately 8 hours and
5 the conversion phase is initiated by the addition of decane to 2 g/L. Continuous feeds for decane (1 g/L-h) and glucose (1.5 g/L-h) are initiated at the same time as the addition of the decane bolus. Fermentation conditions are maintained at 30°C, 1000 rpm, 1 vvm, and pH 5.8 for 44 hours.

Samples were collected for GC analysis at 44 hours after initiating the conversion phase. The
10 data, presented in FIG. 16, show that the decane was converted exclusively to the C10 dicarboxylic acid, sebacic acid. Significant evaporative losses from the decane feed bottles prevented an accurate determination of product yield.

*Example 4: Conversion of mixed fatty acid feedstock to mixed diacid products containing sebacic
15 acid in shake flask fermentation*

5 mL of SP92-glycerol medium (6.7 g/L yeast nitrogen base, 3.0 g/L yeast extract, 3.0 g/L (NH₄)₂SO₄, 1.0 g/L K₂HPO₄, 1.0 g/L KH₂PO₄, 75 g/L glycerol) is inoculated with a single colony of *Candida tropicalis* (ATCC20962) and grown as described in Example 2. 25 mL of the same
20 media is inoculated using overnight cultures to an initial OD_{600nm} of 0.4 and grown overnight at 30°C, with shaking at about 300 rpm. Cells are pelleted by centrifugation for 10 minutes at 4°C and 1,050 x g and the supernatant discarded. Cells are resuspended in 12.5 mL TB-lowN media without carbon source (1.7 g/L yeast nitrogen base without ammonium sulfate, 3.0 g/L yeast extract, 1.0 g/L K₂HPO₄, 1.0 g/L KH₂PO₄) and transferred to a new sterile 250 mL glass baffled
25 flask. Cultures are fed 0.05% capric acid, 0.05% methyl laurate, and 30 g/L glycerol and incubated at 30°C, 300 rpm. After 24 hours of incubation cultures are sampled for GC analysis.

The results, presented in FIG. 17, show that the C12 and C10 fatty acids were converted to dicarboxylic acids of the same chain length (e.g., C12 and C10 dicarboxylic acids), with no
30 evidence of chain shortening of the diacids (e.g., no significant levels of monocarboxylic acids were detected).

Example 5: Conversion of long chain fatty acids to mixed diacids

SP92 fermentation medium was filter sterilized and transferred to a sterile fermentation vessel. Growth of a partially beta-oxidation blocked strain of *Candida tropicalis* (sAA106) was initiated with a 10% inoculum (initial OD_{600nm}=3.0) and grown under the following conditions: of 30°C with shaking at about 1200 rpm, 1 vvm, pH 6.1 and initial volume of 0.3 L. Growth continued until the glucose concentration dropped to less than 2 g/L at which time the conversion phase was initiated by increasing the pH to 8.0 by the addition of 6N KOH and by the addition of methyl myristate to 30 g/L. Immediately following the methyl myristate bolus a continuous feed of glucose was initiated at a rate of 1.5 g/L-h. Fermentation conditions were maintained at 30°C, 1200 rpm, 1 vvm, and pH 8.0 for 90 hours with boluses of 30 g/L methyl myristate at 24, 48, and 72 hours after initiation of conversion. Samples for GC were collected at 24, 48, 72, and 90 hours. The diacid profile graphically illustrated in FIG. 11 shows an accumulation of dicarboxylic acids ranging in chain-length from 6 to 14 carbons long, including sebacic acid. The methyl myristate substrate (methyl ester of myristic acid) is first converted to the C14 dicarboxylic acid via the ω -oxidation pathway before being shortened by two carbon increments via the cyclic β -oxidation pathway. The glucose co-feed employed during the fermentation represses the β -oxidation pathway such that all chain-lengths of diacid accumulate. Manipulation of diacid chain-length distribution is being investigated by altering the glucose co-feed rate in the fermentation medium, thereby allowing growth under varying glucose concentrations.

Example 6: Fermentation procedure for conversion of mixed long-chain fatty acids to mixed diacids of shorter chain length

SP92 fermentation medium without glycerol was filter sterilized and transferred to a sterile fermentation vessel. Autoclaved virgin coconut oil was added to the vessel to a final concentration of 80 g/L. A partially beta-oxidation blocked *Candida tropicalis* strain (sAA496) was inoculated to an initial OD_{600nm} of 1.0 with a 5% inoculum and grown under the following conditions: 30°C with shaking at about 1200 rpm, 1 vvm, initial pH 6.5 and initial volume of 1.0 L. The effect of pH on the distribution of fatty acid chain lengths was determined by manipulating the pH of the fermentation media. The pH of the fermentation was either 1) increased to pH 7.5 and controlled at that pH for the entire run, 2) allowed to drop naturally due to the growth of the culture before controlling at pH 6.0 for the rest of the run, or 3) allowed to drop naturally due to the growth of the culture before controlling at pH 4.5 for the rest of the run. Samples were collected for GC analysis after 140

hours of fermentation time. The product diacid composition was shown to shift to longer chain diacids with increasing pH, as shown in the table below.

	Diacid composition (fraction of total diacids)			
	C12 Diacid	Sebacic Acid	Suberic Acid	Adipic Acid
pH 4.5	0.00	0.00	0.68	0.32
pH 6.0	0.03	0.10	0.75	0.12
pH 7.5	0.16	0.17	0.62	0.05

5

Example 7: Conversion of capric acid to sebacic acid in shake flask fermentations using fully beta-oxidation blocked strains having additional genetic modifications in the omega oxidation pathway.

- 10 Various genetically modified strains of *Candida tropicalis* were inoculated into 5 mL of SP92 medium (6.7 g/L yeast nitrogen base, 3.0 g/L yeast extract, 3.0 g/L $(\text{NH}_4)_2\text{SO}_4$, 1.0 g/L K_2HPO_4 , 1.0 g/L KH_2PO_4 , 75 g/L glycerol). The strains included a completely beta-oxidation blocked strain of *Candida tropicalis* (sAA003), as well as derivatives of sAA003 with amplified components of the omega-oxidation pathway (e.g., various cytochrome P450s, cytochrome P450 reductase or
- 15 combinations thereof) and the cultures grown overnight at 30°C, with shaking at about 250 rpm. These starter cultures were then used to inoculate 25 mL cultures in the same medium and grown overnight at 30°C, with shaking at about 250 rpm. Cells were pelleted by centrifugation for 10 minutes at 4°C and 1,050 x g and the supernatant discarded. Cells were resuspended in 12.5 mL TB-lowN media + glycerol (1.7 g/L yeast nitrogen base without ammonium sulfate, 3.0 g/L yeast
- 20 extract, 1.0 g/L K_2HPO_4 , 1.0 g/L KH_2PO_4 , 75 g/L glycerol) and transferred to a new sterile 250 mL glass baffled flask. Cultures were fed 0.05% from a 5% capric acid solution in ethanol and incubated at 30°C, with shaking at about 300 rpm. After 24 hours incubation cultures were fed glycerol to 30 g/L and an additional bolus of 0.05% capric acid. Incubation continued before
- 25 sampling for GC at 24, 48, and 72 hours. The results are shown in FIG. 12. GC analysis showed that a greater proportion of capric acid was converted to sebacic acid when particular elements of the omega-oxidation pathway are amplified. The data are presented as % of theoretical maximum yield. Strains which include genetic modifications to CYP18 and CYP19 achieve approximately 80% of theoretical maximum yield in conversion of capric acid to sebacic acid. The strain

designated +CPR+A18 has about 30 copies of CYPA18, whereas the strain designated +CPR+A19 has about 7 copies of CYPA19.

Example 8: Conversion of methyl-laurate to dodecanedioic acid in shake flask fermentation

5

5 mL of SP92 glycerol medium (6.7 g/L yeast nitrogen base, 3.0 g/L yeast extract, 3.0 g/L (NH₄)₂SO₄, 1.0 g/L K₂HPO₄, 1.0 g/L KH₂PO₄, 75 g/L glycerol) was inoculated with a single colony of a completely beta-oxidation blocked strain of *Candida tropicalis* (ATCC20962), as well as, modified derivatives of this strain with amplified components of the omega-oxidation pathway, and the cultures grown overnight at 30°C, with shaking at about 250rpm. The starter cultures were then used to inoculate 25 mL cultures of the same medium and grown overnight at 30°C, with shaking at about 250 rpm. Cells were pelleted by centrifugation for 10 minutes at 4°C and 1,050 x g and the supernatant discarded. Cells were resuspended in 12.5 mL SP92 glycerol medium and transferred to a sterile 250 mL glass baffled flask. Cultures were fed 2% (v/v) methyl laurate and incubated at 30°C, with shaking at about 300 rpm. After 24 hours incubation, cultures were fed glycerol to 60 g/L and incubation continued before sampling for GC at 48 hours. GC analysis showed that amplification of certain components of the omega oxidation pathway allow for increased conversion to dodecanedioic acid (FIG. 13).

20 *Example 9: Alteration of Acyl CoA Oxidase Substrate Specificity*

The substrate specificity of the peroxisomal acyl-CoA oxidase enzymes POX4 and POX5 have been shown to be involved in the control of the diacid product chain-length in fermentations of *Candida tropicalis* fed a mixed chain-length fatty acid feedstock. Reduction or elimination of POX4 activity, POX5 activity or POX4 activity and POX5 activity, effects the carbon chain-length distribution of dicarboxylic acids produced in *Candida* spp.. Acyl-CoA oxidase is the first enzyme in the cyclic beta-oxidation pathway that shortens a substrate by two carbons each cycle. Thus the acyl-CoA oxidase activity serves as the pathway entry point for substrates entering into the beta-oxidation pathway. Altering the substrate specificity an acyl-CoA oxidase activity such that it is not active on substrate carbon chains shorter than a desired carbon chain length (e.g., C8, C10, C12, C14 and the like), can inhibit shortening of carbon chains below a chosen threshold, allowing accumulation of a desired target chain length and product (e.g., C12, dodecanedioic acid).

The native acyl-CoA oxidase isozymes in *Candida strain ATCC20336*, Pox4p and Pox5p have different substrate specificities. The Pox4p isozyme has a broad substrate specificity while the Pox5p isozyme has a narrow substrate specificity. In strains that are Pox4⁺, Pox5⁺ the chain length of the diacid product is determined by the substrate specificity of the Pox5p isozyme and the main product is adipic acid.

To maximize production of desired diacid products of longer chain lengths (e.g., C12) in fermentations, genetically modified organisms containing an acyl-CoA oxidase activity with a substrate chain-length specificity appropriate for the chain-length of the desired diacid product can be engineered, in some embodiments. The source of the acyl-CoA oxidase activity or the method of engineering the acyl-CoA oxidase activity may vary. Non-limiting examples of organisms which can be used to provide polynucleotide sequences suitable for use in engineering altered substrate specificity acyl-CoA oxidase activities include; plants (e.g., *Arabidopsis*, *Cucurbita* (e.g., pumpkin, squash), *Oryza* (e.g., rice)); animals (e.g., *Bos* (e.g., bovine), *Cavia* (e.g., guinea pig), *Mus* (e.g., mouse), *Rattus* (e.g., rat), *Phascolarctos* (e.g., Koala), primates (e.g., orangutans)); molds (e.g., *Dictyostelium* (e.g., slime molds)); insects (e.g., *Drosophila*); Yeast (e.g., *Yarrowia lipolytica*, *Candida maltosa*, *Candida glabrata*, *Ashbya gossypii*, *Debaryomyces hansenii*, *Kluyveromyces lactis*, *Pichia pastoris*, *Saccharomyces cerevisiae*); bacteria (e.g., *Escherichia coli*); cyanobacteria; nematodes (e.g., *Caenorhabditis*); and humans.

20

Acyl-CoA oxidase activities with different substrate chain-length specificities can be identified by:

- 1) Selecting acyl-CoA oxidase genes from heterologous organisms that contain different substrate chain-length specificities. The identified genes can be transferred into a *Candida* strain deleted for all acyl-CoA oxidase activity. The only acyl-CoA oxidase activity detectable in such a genetically modified organism may be that imparted by the heterologous gene.
- 2) Engineering an acyl-CoA oxidase gene library by domain swapping from multiple acyl-CoA oxidase genes to produce a library of non-native chimeric acyl-CoA oxidase genes. The library of chimeric genes can be transferred into a strain of *Candida* deleted for all acyl-CoA oxidase activity. The only detectable acyl-CoA oxidase activity may be that imparted by an engineered gene from the library of non-native chimeric acyl-CoA oxidase genes.

- 3) Engineering an acyl-CoA oxidase gene library by random mutagenesis. A naturally occurring or engineered acyl-CoA oxidase activity with a substrate chain-length specificity close to that desired can be used as the basis for random mutagenesis, followed by screening and/or selection in an effort to generate and identify an altered activity with the desired substrate chain-length specificity. The library of genes can be transferred into a *Candida* strain deleted for all acyl-CoA oxidase activity. The only detectable acyl-CoA oxidase activity may be that imparted by the gene from the randomly mutagenized library.
- 4) Engineering an acyl-CoA oxidase gene by intelligent design and directed mutation using protein structural information to guide the position and identity of the amino acid(s) to be replaced. The engineered gene(s) can be transferred into a *Candida* strain deleted for all acyl-CoA oxidase activity. The only detectable acyl-CoA oxidase activity may be that imparted by the engineered gene(s).
- A non-limiting example of a post-engineering method for selecting genes that impart the desired substrate chain-length specificity is provided herein. Selection is performed by growth on substrates of different chain lengths that are provided as the only carbon source. Growth of the cells on certain substrates but not others often reflects the substrate chain-length specificity of the acyl-CoA oxidase enzyme present in the strain. *Candida tropicalis* can utilize alkanes provided in the gas phase as its sole carbon source for growth. Alkanes of different chain lengths are provided by soaking a filter paper in the appropriate alkane, and inverting a solid growth media without a carbon source over the filter paper, with each specific carbon source (e.g., specific chain length alkane) provided in a different petri dish. Serially diluted *Candida* carrying the altered specificity acyl-CoA oxidase genes are spotted on the solid growth media as a growth selection for the chain-length specificity of the acyl-CoA oxidase enzyme in each strain. Shown in FIGS. 14 and 15 are a schematic representation of the selection process, which provides an alkane as a gas phase carbon source, as described herein. The solid growth media is an agar medium containing yeast nitrogen base without amino acids or any other carbon source. The plated cells are inverted over a lid containing a filter paper soaked with an alkane of appropriate chain length that evaporates and provides the carbon source through the gas phase, as shown in FIG. 14.

Candida strains containing altered acyl-CoA oxidase activities generated as described herein are selected and/or screened using the method described herein. Strains carrying different altered acyl-CoA oxidase activities (e.g., strain 1 (S1), strain 2 (S2), strain 3 (S3), strain 4 (S4)) are grown

overnight in a rich medium (e.g., YPD). Overnight cultures are centrifuged and washed to remove any traces of residual rich medium and serial dilutions of the cells are prepared in a phosphate buffered solution. The serial dilutions of each strain are spotted onto multiple YNB agar plates (growth medium having no amino acids or other carbon sources), the individual plates inverted
5 over filter papers soaked in the appropriate chain length alkane, and the plate incubated at 30°C. The growth of the strains is dependent upon the chain-length specificity of the acyl-CoA oxidase. In order to utilize the particular alkane for growth the provided chain-length must be able to enter the beta-oxidation pathway. The shortest chain-length at which a certain strain is able to grow indicates the shortest chain-length of the acyl-CoA oxidase isozymes substrate specificity. An
10 example is provided in FIG. 15. FIG. 15 illustrates that strain S4 can grow on decane, but is unable to grow on octane. Therefore the modified acyl-CoA oxidase activity of strain S4 has a substrate chain-length specificity that inhibits the utilization of 8 carbon molecules and the diacid product from fermentations with this strain typically result in an 8 carbon diacid. Acyl-CoA oxidase activities with any desired specificity can be selected and/or screened using the method described
15 herein.

It will be understood that the example presented herein is a generalize method used to describe the selection/screening process. The feedstocks used for the selection and screening process are altered to suit the acyl-CoA oxidase activity being sought. For example, for acyl-CoA oxidases
20 having specificity for longer chain substrates, feedstocks having longer carbon chain lengths could be substituted to allow selection and or screening for acyl-CoA oxidase activities with specificities for longer carbon chain lengths.

Example 10: Transformation of Candida spp. procedure

25 5mL YPD start cultures were inoculated with a single colony of Candida strain ATCC20336 and incubated overnight at 30°C, with shaking at about 200rpm. The following day, fresh 25mL YPD cultures, containing 0.05% Antifoam B, were inoculated to an initial OD_{600nm} of 0.4 and the culture incubated at 30°C, with shaking at about 200rpm until an OD_{600nm} of 1.0-2.0 was reached. Cells
30 were pelleted by centrifugation at 1,000 x g, 4°C for 10 minutes. Cells were washed by resuspending in 10mL sterile water, pelleted, resuspended in 1mL sterile water and transferred to a 1.5mL microcentrifuge tube. The cells were then washed in 1mL sterile TE/LiOAC solution, pH 7.5, pelleted, resuspended in 0.25mL TE/LiOAC solution and incubated with shaking at 30°C for 30 minutes.

The cell solution was divided into 50uL aliquots in 1.5mL tubes to which was added 5-8ug of linearized DNA and 5uL of carrier DNA (boiled and cooled salmon sperm DNA, 10mg/mL). 300uL of sterile PEG solution (40% PEG 3500, 1X TE, 1X LiOAC) was added, mixed thoroughly and incubated at 30°C for 60 minutes with gentle mixing every 15 minutes. 40uL of DMSO was added, mixed thoroughly and the cell solution was incubated at 42°C for 15 minutes. Cells were then pelleted by centrifugation at 1,000 x g 30 seconds, resuspended in 500uL of YPD media and incubated at 30°C with shaking at about 200rpm for 2 hours. Cells were then pelleted by centrifugation and resuspended in 1mL 1X TE, cells were pelleted again, resuspended in 0.2mL 1X TE and plated on selective media. Plates were incubated at 30°C for growth of transformants.

Example 11: Procedure for recycling of the URA3 marker

The URA3 gene was obtained from genomic DNA of Candida yeast culture ATCC20336. Candida strain ATCC20336 has a limited number of selectable marker, as compared to *S. cerevisiae*, therefore, the URA3 marker is "recycled" to allow multiple rounds of selection using URA3. To reutilize the URA3 marker for subsequent engineering of Candida spp., a single colony having the Ura⁺ phenotype was inoculated into 3 mL YPD and grown overnight at 30°C with shaking. The overnight culture was then harvested by centrifugation and resuspended in 1 mL YNB+YE (6.7 g/L Yeast Nitrogen Broth, 3g/L Yeast Extract). The resuspended cells were then serially diluted in YNB+YE and 100 uL aliquots plated on YPD plates (incubation overnight at 30°C) to determine titer of the original suspension. Additionally, triplicate 100 uL aliquots of the undiluted suspension were plated on SC Dextrose (Bacto Agar 20g/L, Uracil 0.3 g/L, Dextrose 20 g/L, Yeast Nitrogen Broth 6.7 g/L, Amino Acid Dropout Mix 2.14 g/L) and 5-FOA at 3 different concentrations (0.5, 0.75, 1 mg/mL).

Plates were incubated for at least 5 days at 30°C. Colonies arising on the SC Dextrose + 5-FOA plates were resuspended in 50 uL sterile, distilled water and 5 uL utilized to streak on to YPD and SC-URA (SC Dextrose medium without Uracil) plates. Colonies growing only on YPD and not on SC-URA plates were then inoculated into 3 mL YPD and grown overnight at 30°C with shaking. Overnight cultures were harvested by centrifugation and resuspended in 1.5 mL YNB (6.7 g/L Yeast Nitrogen Broth). The resuspended cells were serially diluted in YNB and 100 uL aliquots plated on YPD plates and incubation overnight at 30°C to determine initial titer. 1 mL of each undiluted cell suspension also was plated on SC-URA and incubated for up to 7 days at

30°C. Colonies on the SC-URA plates are revertants and the isolate with the lowest reversion frequency ($< 10^{-7}$) was used for subsequent strain engineering.

Example 12: Cloning and Analysis of Candida Fatty alcohol oxidase (FAO) alleles

5

Isolation of fatty alcohol oxidase genes from Candida

Candida strain (ATCC20336) fatty alcohol oxidase genes were isolated by PCR amplification using primers generated to amplify the sequence region covering promoter, fatty alcohol oxidase gene
10 (FAO) and terminator of the FAO1 sequence (GenBank accession number of FAO1 AY538780). The primers used to amplify the fatty alcohol oxidase nucleotide sequences from Candida strain ATCC20336 strain ATCC20336, are showing in the table below.

Oligonucleotides for cloning FAO alleles

Oligo	Sequence
oAA0144	AACGACAAGATTAGATTGGTTGAGA
oAA0145	GTCGAGTTTGAAGTGTGTGTCTAAG
oAA0268	AGATCTCATATGGCTCCATTTTTGCCCGACCAGGTCGACTACAAACACGTC
oAA0269	ATCTGGATCCTCATTACTACAACCTTGGCTTTGGTCTTCAAGGAGTCTGCCAAACCTAAC
oAA0282	ACATCTGGATCCTCATTACTACAACCTTGGCCTTGGTCT
oAA0421	CACACAGCTCTTCTAGAAATGGCTCCATTTTTGCCCGACCAGGTCGAC
oAA0422	CACACAGCTCTTCCTTTCTACAACCTTGGCTTTGGTCTTCAAGGAGTCTGC
oAA0429	GTCTACTGATTCCCCTTTGTC
oAA0281	TTCTCGTTGTACCCGTCGCA

15

PCR reactions contained 25uL 2X master mix, 1.5 uL of oAA0144 and oAA0145 (10uM), 3.0uL genomic DNA, and 19uL sterile H₂O. Thermocycling parameters used were 98°C for 2 minutes, 35 cycles of 98°C 20 seconds, 52°C 20 seconds, 72°C 1 minute, followed by 72°C 5 minutes and a 4°C hold. PCR products of the correct size were gel purified, ligated into pCR-Blunt II-TOPO
20 (Invitrogen) and transformed into competent TOP10 E. coli cells (Invitrogen). Clones containing PCR inserts were sequenced to confirm correct DNA sequence. Four FAO alleles were identified from sequence analysis and designated as FAO-13, FAO-17, FAO-18 and FAO-20. The sequence of the clone designated FAO-18 had a sequence that was substantially identical to the sequence of FAO1 from GenBank. The resulting plasmids of the four alleles were designated pAA083, pAA084,

pAA059 and pAA085, respectively. Sequence identity comparisons of FAO genes isolated as described herein are shown in the tables below.

DNA sequence identity							
	FAO1	FAO-18	FAO-17	FAO-13	FAO-20	FAO2a	FAO2b
FAO1	100	100	98	96	95	83	82
FAO-18		100	98	96	95	83	82
FAO-17			100	98	98	83	82
FAO-13				100	99	83	83
FAO-20					100	83	83
FAO2a						100	96
FAO2b							100

5

Protein sequence identity							
	FAO1	FAO-18	FAO-17	FAO-13	FAO-20	FAO2a	FAO2b
FAO1	100	100	99	98	98	81	80
FAO-18		100	99	98	98	81	80
FAO-17			100	99	99	82	81
FAO-13				100	99	82	81
FAO-20					100	82	81
FAO2a						100	97
FAO2b							100

Amino acid differences in FAO alleles

	32	75	89	179	185	213	226	352	544	590
FAO1	E	M	G	L	Y	T	R	H	S	P
FAO-13	Q	T	A	L	Y	A	K	Q	A	A
FAO-20	Q	T	A	M	D	A	K	Q	A	A

10 Expression of FAO alleles in E. coli

To determine the levels of FAO enzyme activity with respect to various carbon sources, the four isolated FAO alleles were further cloned and over-expressed in E. coli. The FAOs were amplified using the plasmids mentioned above as DNA template by PCR with primers oAA0268 and

oAA0269 for FAO-13 and FAO-20 and oAA0268 and oAA0282 for FAO-17 and FAO-18, using conditions as described herein. PCR products of the correct size were gel purified and ligated into pET11a vector between NdeI and BamHI sites and transformed into BL21 (DE3) E. coli cells. The colonies containing corresponding FAOs were confirmed by DNA sequencing. Unmodified

5 pET11a vector also was transformed into BL21 (DE3) cells, as a control. The resulting strains and plasmids were designated sAA153 (pET11a), sAA154 (pAA079 containing FAO-13), sAA155 (pAA080 containing FAO-17), sAA156 (pAA081 containing FAO-18) and sAA157 (pAA082 containing FAO-20), respectively. The strains and plasmids were used for FAO over-expression in E. coli. One colony of each strain was transferred into 5 mL of LB medium containing 100 µg/mL

10 ampicillin and grown overnight at 37°C, 200 rpm. The overnight culture was used to inoculate a new culture to OD_{600nm} 0.2 in 25 ml LB containing 100 µg/ml ampicillin. Cells were induced at OD_{600nm} 0.8 with 0.3 mM IPTG for 3 hours and harvested by centrifugation at 4°C 1,050xg for 10 minutes. The cell pellet was stored at -20 °C.

15 Expression of FAOs in a Candida strain

Two alleles, FAO-13 and FAO-20, were chosen for amplification in Candida based on their substrate specificity profile, as determined from enzyme assays of soluble cell extracts of E. coli with over expressed FAOs. DNA fragments containing FAO-13 and FAO-20 were amplified using

20 plasmids pAA079 and pAA082 as DNA templates, respectively, by PCR with primers oAA0421 and oAA0422. PCR products of the correct sizes were gel purified and ligated into pCR-Blunt II-TOPO (Invitrogen), transformed into competent TOP10 E. coli cells (Invitrogen) and clones containing FAO inserts were sequenced to confirm correct DNA sequence. Plasmids containing FAO-13 and FAO-20 were digested with SapI and ligated into vector pAA105, which includes the Candida strain

25 ATCC20336 PGK promoter and terminator. The resulting plasmids were confirmed by restriction digestion and DNA sequencing and designated as pAA115 (FAO-13) and pAA116 (FAO-20), respectively. Plasmids pAA115 and pAA116 were linearized with SpeI, transformed into competent Candida spp. Ura⁻ strains sAA002 (SU-2, ATCC20913) and sAA103. The integration of FAO-13 and FAO-20 was confirmed by colony PCR using primers oAA0429 and oAA0281. The

30 resulting strains were designated as sAA278 (pAA115 integrated in strain sAA002), sAA280 (pAA116 integrated in sAA002), sAA282 (pAA115 integrated in sAA103), and sAA284 (pAA116 integrated in sAA103), and were used for fatty alcohol oxidase over-expression in Candida spp..

One colony of each strain was inoculated into 5 ml YPD and grown overnight as described herein. The overnight culture was used to inoculate a new 25 mL YPD culture to about OD_{600nm} 0.5. FAO over-expression was regulated by the PGK promoter/terminator, induced with glucose in the medium and expressed constitutively. Strains sAA002 and sAA103 (e.g., untransformed starting strains) were included as negative controls for FAO over-expression. Cells were harvested at early log phase (OD_{600nm} = in the range of between about 3 to about 5) by centrifugation at 4°C for 10 minutes at 1,050 x g. Cell pellets were stored at -20 °C.

Cell extract preparation from E. coli

Cell pellets from 25 mL of FAO expressing E. coli cultures were resuspended in 10 mL phosphate-glycerol buffer containing 50 mM potassium phosphate buffer (pH7.6), 20% glycerol, 1 mM Phenylmethylsulfonyl fluoride (PMSF), 2uL Benzonase™ 25U/uL, 20uL 10mg/mL lysozyme. The cells were then lysed by incubation at room temperature for 50 minutes on a rotating shaker, and the cell suspension centrifuged for 30 minutes at 4°C using 15,000 x g for. The supernatant was aliquoted in 1.5 ml microcentrifuge tubes and stored at -20 °C for FAO enzyme activity assays.

Cell extract preparation from Candida

Frozen C. tropicalis cell pellets were resuspended in 1.2 ml of phosphate-glycerol buffer containing 50 mM potassium phosphate buffer (pH7.6), 20% glycerol, 1 mM Phenylmethylsulfonyl fluoride (PMSF). Resuspended cells were transferred to 1.5mL screw-cap tubes containing about 500uL of zirconia beads on ice. The cells were lysed with a Bead Beater (Biospec) using 2 minute pulses and 1 minute rest intervals on ice. The process was repeated 3 times. The whole cell extract was then transferred to a new 1.5 ml tube and centrifuged at 16,000 x g for 15 minutes at 4°C. The supernatant was transferred into a new tube and used for FAO enzyme activity assays.

Protein concentration determination

Protein concentration of the cell extracts was determined using the Bradford Reagent following manufacturers' recommendations (Cat# 23238, Thermo scientific).

FAO enzyme activity assay

FAO enzyme activity assays were performed using a modification of Eirich et al., 2004). The assay utilizes a two-enzyme coupled reaction (e.g., FAO and horse radish peroxidase (HRP)) and can be

monitored by spectrophotometry. 1-Dodecanol was used as a standard substrate for fatty alcohol oxidase enzymatic activity assays. FAO oxidizes the dodecanol to dodecanal while reducing molecular oxygen to hydrogen peroxide simultaneously. HRP reduces (2,2'-azino-bis 3-ethylbenzthiazoline-6-sulfonic acid; ABTS) in the two-enzyme coupled reaction, where the electron
 5 obtained from oxidizing hydrogen peroxide to ABTS, which can be measured by spectrometry at 405 nm. The assay was modified using aminotriazole (AT) to prevent the destruction of H₂O₂ by endogenous catalase, thus eliminating the need for microsomal fractionation. The final reaction mixture (1.0 mL) for FAO enzyme assay consisted of 500 μ L of 200 mM HEPES buffer, pH 7.6; 50 μ L of a 10 mg/mL ABTS solution in deionized water; 10 μ L of 5 mM solution of dodecanol in
 10 acetone; 40 μ L of 1M AT and 5 μ L of a 2 mg/mL horseradish peroxidase solution in 50 mM potassium phosphate buffer, pH 7.6. Reaction activity was measured by measuring light absorbance at 405 nm for 10 minutes at room temperature after adding the extract. The amount of extract added to the reaction mixture was varied so that the activity fell within the range of 0.2 to 1.0 $\Delta A_{405nm}/min$. The actual amounts of extract used were about 1.69 U/mg for E. coli expressed
 15 FAO-13, 0.018U/mg for E. coli expressed FAO-17, 0.35U/mg for E. coli expressed FAO-18 (e.g., FAO1), 0.47U/mg E. coli expressed FAO-20, 0.036U/mg Candida (strain sAA278) expressed FAO-13, 0.016U/mg Candida (strain sAA282) expressed FAO-13, 0.032U/mg Candida (strain sAA280) expressed FAO-20 and 0.029U/mg Candida (strain sAA284) expressed FAO-20. FAO activity was reported as activity units/mg of total protein (1 unit = 1 μ mole substrate oxidized/min). An
 20 extinction coefficient at 405 nm of 18.4 was used for ABTS and was equivalent to 0.5 mM oxidized substrate. The results of the activity assays are shown in the tables below.

FAO activity (units/mg total protein) on primary alcohols

	1- Butanol	1- Pentanol	1- Hexanol	1- Octanol	1- Decanol	1- Dodecanol	1- Tetradecanol	Hexadecanol
FAO-13	0.01	0.09	1.17	82.67	70.94	100	79.35	58.88
FAO-17	0.72	0.26	1.06	66.23	22.00	100	47.86	60.98
FAO-18	0.07	0.11	0.26	60.56	54.56	100	114.47	50.65
FAO-20	0.07	0.11	0.91	55.96	74.57	100	89.52	42.59

FAO activity (units/mg total protein) on omega hydroxy fatty acids

	1-Dodecanol	6-OH-HA	10-OH-DA	12-OH-DDA	16-OH-HDA
FAO-13	100	4.18	4.14	6.87	8.57
FAO-17	100	1.18	0.00	0.59	0.94
FAO-18	100	0.00	0.00	4.87	2.94
FAO-20	100	0.03	0.04	2.25	7.46

Example 13: Construction of Candida shuttle vector pAA061

Vector pAA061 was constructed from a pUC19 backbone to harbor the selectable marker URA3 from Candida strain ATCC20336 as well as modifications to allow insertion of Candida promoters and terminators. A 1,507bp DNA fragment containing the promoter, ORF, and terminator of URA3 from Candida ATCC20336 was amplified using primers oAA0124 and oAA0125, shown in the table below. The URA3 PCR product was digested with NdeI / MluI and ligated into the 2,505bp fragment of pUC19 digested with NdeI / BsmBI (an MluI compatible overhang was produced by BsmBI). In order to replace the lac promoter with a short 21bp linker sequence, the resulting plasmid was digested with SphI / SapI and filled in with a linker produced by annealing oligos oAA0173 and oAA0174. The resulting plasmid was designated pAA061.

Oligonucleotides for construction of pAA061

Oligos	Sequence	PCR product (bp)
oAA0124	cacacacatatgCGACGGGTACAACGAGAATT	1507
oAA0125	cacacaacgcgtAGACGAAGCCGTTCTTCAAG	
oAA0173	ATGATCTGCCATGCCGAATC	21 (linker)
oAA0174	AGCGAGTTTCGGCATGGCAGATCATCATG	

Example 14: Cloning of Candida PGK promoter and terminator

Vector pAA105 was constructed from base vector pAA061 to include the phosphoglycerate kinase (PGK) promoter and terminator regions from Candida ATCC20336 with an intervening multiple cloning site (MCS) for insertion of open reading frames (ORF's). The PGK promoter region was amplified by PCR using primers oAA0347 and oAA0348, shown in the table below. The 1,029bp DNA fragment containing the PGK promoter was digested with restriction enzymes PstI / XmaI. The PGK terminator region was amplified by PCR using primers oAA0351 and oAA0352, also

- shown in the table below. The 396bp DNA fragment containing the PGK terminator was digested with restriction enzymes XmaI / EcoRI. The 3,728bp PstI / EcoRI DNA fragment from pAA061 was used in a three piece ligation reaction with the PGK promoter and terminator regions to produce pAA105. The sequence between the PGK promoter and terminator contains restriction sites for incorporating ORF's to be controlled by the functionally linked constitutive PGK promoter.

Oligonucleotides for cloning *Candida* PGK promoter and terminator

Oligos	Sequence	PCR product (bp)
oAA0347	CACACACTGCAGTTGTCCAATGTAATAATTTT	1028
oAA0348	CACACATCTAGACCCGGGCTCTTCTTCTGAATAGGCAATTGATA AACTTACTTATC	
oAA0351	GAGCCCGGGTCTAGATGTGTGCTCTTCCAAAGTACGGTGTGTG TGACA	396
oAA0352	CACACACATATGAATTCTGTACTGGTAGAGCTAAATT	

Example 15: Cloning of the *POX4* locus

- Primers oAA0138 and oAA0141 (shown in the table below) were generated to amplify the entire sequence of NCBI accession number M12160 for the YSAPOX4 locus from genomic DNA prepared from *Candida* strain ATCC20336. The 2,845bp PCR product was cloned into the vector, pCR-BluntII-TOPO (Invitrogen), sequenced and designated pAA052.

Oligonucleotides for cloning of *POX4*

Oligos	Sequence	PCR product (bp)
oAA0138	GAGCTCCAATTGTAATATTTTCGGG	2845
oAA0141	GTCGACCTAAATTCGCAACTATCAA	

15

Example 16: Cloning of the *POX5* locus

- Primers oAA0179 and oAA0182 (shown in the table below) were generated to amplify the entire sequence of NCBI accession number M12161 for the YSAPOX5 locus from genomic DNA prepared from *Candida* strain ATCC20336. The 2,624bp PCR product was cloned into the vector, pCR-BluntII-TOPO (Invitrogen), sequenced and designated pAA049.

Oligonucleotides for cloning of *POX5*

Oligos	Sequence	PCR product (bp)
oAA0179	GAATTCACATGGCTAATTTGGCCTCGGTTCCACAACGCACTCAGC ATTAAGAAA	2624
oAA0182	GAGCTCCCTGCAAACAGGGAAACACTTGTCTATCTGATTT	

Example 17: Construction of strains with amplified CPR and CYP52 genes

Strains having an increased number of copies of cytochrome P450 reductase (CPR) and/or for
 5 cytochrome P450 monooxygenase (CYP52) genes were constructed to determine how over
 expression of CPR and CYP52 affected diacid production.

Cloning and integration of the CPR gene.

10 A 3,019bp DNA fragment encoding the CPR promoter, ORF, and terminator from *Candida*
 ATCC750 was amplified by PCR using primers oAA0171 and oAA0172 (see table below)
 incorporating unique SapI and SphI sites. The amplified DNA fragment was cut with the indicated
 restriction enzymes and ligated into plasmid pAA061, (described in Example 13) to produce
 plasmid pAA067. Plasmid pAA067 was linearized with ClaI and transformed into *Candida* Ura⁻
 15 strain sAA103 (ura3/ura3, pox4::ura3/pox4::ura3, pox5::ura3/pox5::ura3). Transformations were
 performed with plasmid pAA067 alone and in combination with plasmids harboring the CYP52A15
 or CYP52A16 genes, described below.

Cloning and integration of CYP52A15 gene.

20 A 2,842bp DNA fragment encoding the CYP52A15 promoter, ORF, and terminator from *Candida*
 ATCC20336 was amplified by PCR using primers oAA0175 and oAA0178 (see table below) and
 cloned into pCR-BluntII-TOPO for DNA sequence verification. The cloned CYP52A15 DNA
 fragment was isolated by restriction digest with XbaI / BamHI (2,742bp) and ligated into plasmid
 25 pAA061, (described in Example 13), to produce plasmid pAA077. Plasmid pAA077 was linearized
 with PmlI and transformed into *Candida* Ura⁻ strain sAA103 (ura3/ura3, pox4::ura3/pox4::ura3,
 pox5::ura3/pox5::ura3). pAA077 was cotransformed with plasmid pAA067 harboring the CPR
 gene.

Cloning and integration of CYP52A16 gene.

A 2,728bp DNA fragment encoding the CYP52A16 promoter, ORF, and terminator from *Candida* ATCC20336 was amplified by PCR using primers oAA0177 and oAA0178 (see table below) and
 5 cloned into pCR-BluntII-TOPO for DNA sequence verification. The cloned CYP52A16 DNA fragment was amplified with primers oAA0260 and oAA0261 (see table below) which incorporated unique *SacI* / *XbaI* restriction sites. The amplified DNA fragment was digested with *SacI* and *XbaI* restriction enzymes and ligated into plasmid pAA061 to produce plasmid pAA078. Plasmid
 10 pAA078 was linearized with *Clal* and transformed into *Candida* Ura⁻ strain sAA103 (*ura3/ura3*, *pox4::ura3/pox4::ura3*, *pox5::ura3/pox5::ura3*). pAA078 was cotransformed with plasmid pAA067 harboring the CPR gene.

Oligonucleotides for cloning of CPR, CYP52A15 and CYP52A16

Oligos	Sequence	PCR product (bp)
oAA0171	cacctgctcttccAGCTGTCATGCTATTCAATGCTTCGA	3019
oAA0172	cacacagcatgcTAATGTTTATATCGTTGACGGTGAAA	
oAA0175	cacaaagcgggaagagcAAATTTTGTATTCTCAGTAGGATTT CATC	2842
oAA0178	cacacagcatgCAAACTTAAGGGTGTGTTGTAGATATCCC	2772
oAA0177	cacacacccgggATCGACAGTCGATTACGTAATCCATATT ATTT	
oAA0178	cacacagcatgCAAACTTAAGGGTGTGTTGTAGATATCCC	2772
oAA0260	cacacagagctcACAGTCGATTACGTAATCCAT	
oAA0261	cacatctagaGCATGCAAACTTAAGGGTGTGTA	

Preparation of genomic DNA.

15 Genomic DNA was prepared from transformants for PCR verification and for Southern blot analysis. Isolated colonies were inoculated into 3 mL YPD and grown overnight at 30 °C with shaking. Cells were pelleted by centrifugation. To each pellet, 200 uL Breaking Buffer (2% Triton X-100, 1% SDS, 100 mM NaCl, 10 mM Tris pH 8 and, 1 mM EDTA) was added, and the pellet
 20 resuspended and transferred to a fresh tube containing 200 uL 0.5 mm Zirconia/Silica Beads. 200 uL Phenol:Chloroform:Isoamyl Alcohol (25:24:1) was added to each tube, followed by vortexing for 1 minute. Sterile distilled water was added (200 uL) to each tube and the tubes were centrifuged at 13000 rpm for 10 minutes. The aqueous layer was ethanol precipitated and washed with 70% ethanol. The pellet was resuspended in 100-200 µl 10 mM Tris, after drying. Genomic DNA

preparation for southern blot analysis was performed using the same procedure on 25 mL cultures for each colony tested.

Characterization of strains with amplified CPR and CYP52 genes.

5

Verification of integrants was performed by PCR using primers oAA0252 and oAA0256 (CPR), oAA0231 and oAA0281 (CYP52A15), and oAA0242 and oAA0257 (CYP52A16). The primers used for verification are shown in the table below.

Oligonucleotides for PCR verification of CPR, CYP52A15 and CYP52A16

Oligos	Sequence	PCR product (bp)
oAA0252	TTAATGCCTTCTCAAGACAA	743
oAA0256	GGTTTTCCCAGTCACGACGT	
oAA0231	CCTTGCTAATTTTCTTCTGTATAGC	584
oAA0281	TTCTCGTTGTACCCGTCGCA	
oAA0242	CACACAACCTTCAGAGTTGCC	974
oAA0257	TCGCCACCTCTGACTTGAGC	

10

Southern blot analysis was used to determine the copy number of the CPR, CYP52A15 and CYP52A16 genes. Biotinylated DNA probes were prepared with gene specific oligonucleotides using the NEBlot Phototope Kit from New England BioLabs (Catalog #N7550S) on PCR products generated from each gene target as specified in the table below. Southern Hybridizations were performed using standard methods (e.g., Sambrook, J. and Russell, D. W. (2001) Molecular Cloning: A Laboratory Manual, (3rd ed.), pp. 6.33-6.64. Cold Spring Harbor Laboratory Press). Detection of hybridized probe was performed using the Phototope-Star Detection Kit from New England BioLabs (Catalog #N7020S). Copy number was determined by densitometry of the resulting bands.

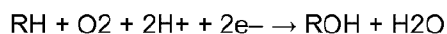
15

Oligonucleotides for Probe Template PCR of CPR, CYP52A15 and CYP52A16

Oligos	Sequence	Gene	Template	PCR product (bp)
oAA0250	AATTGAACATCAGAAGAGGA	CPR	pAA067	1313
oAA0254	CCTGAAATTTCCAAATGGTGTCTAA			
oAA0227	TTTTTTGTGCGCAAGTACAC	CYP52A15	pAA077	905
oAA0235	CAACTTGACGTGAGAAACCT			
oAA0239	AGATGCTCGTTTACACCCT	CYP52A16	pAA078	672
oAA0247	ACACAGCTTTGATGTTCTCT			

Example 18: Addition and/or Amplification of Monooxygenase and Monooxygenase reductase activities.

- 5 Cytochrome P450's often catalyze a monooxygenase reaction, e.g., insertion of one atom of oxygen into an organic substrate (RH) while the other oxygen atom is reduced to water:



- 10 The substrates sometimes are of a homogeneous carbon chain length. Enzymes with monooxygenase activity sometimes recognize substrates of specific carbon chain lengths, or a subgroup of carbon chain lengths with respect to organic substrates of homogenous carbon chain length. Addition of novel cytochrome activities (e.g., B. megaterium BM3) and/or amplification of certain or all endogenous or heterologous monooxygenase activities (e.g., CYP52A12
- 15 polynucleotide, CYP52A13 polynucleotide, CYP52A14 polynucleotide, CYP52A15 polynucleotide, CYP52A16 polynucleotide, CYP52A17 polynucleotide, CYP52A18 polynucleotide, CYP52A19 polynucleotide, CYP52A20 polynucleotide, CYP52D2 polynucleotide, BM3 polynucleotide) can contribute to an overall increase in carbon flux through native and/or engineered metabolic pathways, in some embodiments. In certain embodiments, adding a novel monooxygenase or
- 20 increasing certain or all endogenous or heterologous monooxygenase activities can increase the flux of substrates of specific carbon chain length or subgroups of substrates with mixtures of specific carbon chain lengths. In some embodiments, the selection of a monooxygenase activity for amplification in an engineered strain is related to the feedstock utilized for growth of the engineered strain, pathways for metabolism of the chosen feedstock and the desire end product
- 25 (e.g., dodecanedioic acid).

Strains engineered to utilize plant-based oils for conversion to dodecanedioic acid can benefit by having one or more monooxygenase activities with substrate specificity that matches the fatty acid chain-length distribution of the oil. For example, the most prevalent fatty acid in coconut oil is lauric acid (12 carbons long), therefore, the monooxygenase activity chosen for a coconut oil-
 5 utilizing strain can have a substrate preference for C12 fatty acids. For strains engineered to utilize other plant based oils with different fatty acid chain-length distributions it may be desirable to amplify a monooxygenase activity that has a matching substrate preference. In some embodiments, a genetic modification that alters monooxygenase activity increases the activity of one or more monooxygenase activities with a substrate preference for feedstocks having carbon
 10 chain lengths of between about 12 and about 24 carbons (e.g., mixed chain length alkanes, mixed chain length fatty acids, soapstocks, the like and combinations thereof). In certain embodiments, the genetic modification increases the activity of a monooxygenase activity with a preference for fatty acids having a carbon chain-length distribution of between about 10 carbons and about 16 carbons.

15 As mentioned previously, the enzymes that carry out the monooxygenase activity are reduced by the activity of monooxygenase reductase, thereby regenerating the enzyme. Selection of a CPR for amplification in an engineered strain depends upon which P450 is amplified, in some embodiments. A particular CPR may interact preferentially with one or more monooxygenase
 20 activities, in some embodiments, but not well with other monooxygenases. A monooxygenase reductase from *Candida* strain ATCC750, two monooxygenase reductase activities from *Candida* strain ATCC20336 and a monooxygenase reductase activity from *Bacillus megaterium* are being evaluated for activity with the added and/or amplified monooxygenases described herein. Provided in the tables below are nucleotide sequences used to add or amplify monooxygenase
 25 and monooxygenase reductase activities.

Example 19: Amplification of Selected Beta Oxidation Activities

Described herein are methods of amplifying a POX5 beta oxidation activity. Substantially similar
 30 methods can be utilized to amplify different beta oxidation activities including non-POX (e.g., acyl-CoA oxidase) activities and/or acyl-CoA oxidase activities with altered substrate specificities, as described herein.

Construction of POX5 amplified strainsPlasmid pAA166 ($P_{POX4}POX5T_{POX4}$)

- 5 A PCR product containing the nucleotide sequence of POX5 was amplified from Candida 20336 genomic DNA using primers oAA540 and oAA541. The PCR product was gel purified and ligated into pCR-Blunt II-TOPO (Invitrogen), transformed into competent TOP10 E. coli cells (Invitrogen) and clones containing PCR inserts were sequenced to confirm correct DNA sequence. One such plasmid was designated, pAA165. Plasmid pAA165 was digested with BspQI and a 2-kb fragment
- 10 was isolated. Plasmid pAA073 which contained a POX4 promoter and POX4 terminator was also digested with BspQI and gel purified. The isolated fragments were ligated together to generate plasmid pAA166. Plasmid pAA166 contains a $P_{POX4}POX5T_{POX4}$ fragment.

Plasmid pAA204 (thiolase deletion construct)

- 15 A PCR product containing the nucleotide sequence of a short-chain thiolase (e.g., acetyl-coA acetyltransferase) was amplified from Candida 20336 genomic DNA using primers oAA640 and oAA641. The PCR product was gel purified and ligated into pCR-Blunt II-TOPO (Invitrogen), transformed into competent TOP10 E. coli cells (Invitrogen) and clones containing PCR inserts
- 20 were sequenced to confirm correct DNA sequence. One such plasmid was designated, pAA184. A URA3 PCR product was amplified from pAA061 using primers oAA660 and oAA661. The PCR product was gel purified and ligated into pCR-Blunt II-TOPO (Invitrogen), transformed as described and clones containing PCR inserts were sequenced to confirm the correct DNA sequence. One such plasmid was designated pAA192. Plasmid pAA184 was digested with BglII/Sall and gel
- 25 purified. Plasmid pAA192 was digested with BglII/Sall and a 1.5 kb fragment was gel purified. The isolate fragments were ligated together to generate pAA199. An alternative P_{URA3} PCR product was amplified from plasmid pAA061 using primers oAA684 and oAA685. The PCR product was gel purified and ligated into pCR-Blunt II-TOPO (Invitrogen), transformed as described and clones containing PCR inserts were sequenced. One such plasmid was designated, pAA201. Plasmid
- 30 pAA199 was digested with Sall and gel purified. Plasmid pAA201 was digested with Sall and a 0.43 kb P_{URA3} was gel purified. The isolated fragments were ligated to generate plasmid pAA204 that contains a direct repeat of P_{URA3} .

Plasmid pAA221 ($P_{POX4}POX5T_{POX4}$ in thiolase deletion construct)

A PCR product containing the nucleotide sequence of $P_{POX4}POX5T_{POX4}$ was amplified from plasmid pAA166 DNA using primers oAA728 and oAA729. The PCR product was gel purified and ligated
 5 into pCR-Blunt II-TOPO, transformed as described and clones containing PCR inserts were sequenced to confirm the sequence of the insert. One such plasmid was designated, pAA220. Plasmid pAA204 was digested with BglII, treated with shrimp alkaline phosphatase (SAP), and a 6.5 kb fragment was gel purified. Plasmid pAA220 was digested with BglII and a 2.7 kb fragment containing $P_{POX4}POX5T_{POX4}$ was gel purified. The isolated fragments were ligated to generate
 10 plasmid pAA221.

Strain sAA617 ($P_{POX4}POX5T_{POX4}$ in sAA451)

Strain sAA451 is a ura-, partially β -oxidation blocked strain ($ura3/ura3\ pox4a::ura3/pox4b::ura3$
 15 $POX5/POX5$). Plasmid pAA221 was digested with EcoRI to release a DNA fragment containing $P_{POX4}POX5T_{POX4}$ in a thiolase deletion construct. The DNA was column purified and transformed to strain sAA451 to plate on SCD-ura plate. After two days, colonies were streaked out on YPD plates, single colonies selected and again streaked out on YPD plates. Single colonies were selected from the second YPD plates and characterized by colony PCR. The insertion of
 20 $P_{POX4}POX5T_{POX4}$ in strain sAA451, disrupting the short-chain thiolase gene, was confirmed by PCR and one such strain was designated sAA617.

Strain sAA620

25 Strain sAA617 was grown overnight on YPD medium and plated on SCD+URA+5-FOA, to select for loop-out of URA3. Colonies were streaked out onto YPD plates twice as described for strain sAA617, and single colonies characterized by colony PCR. The loop-out of URA3 by direct repeats of PURA3 was confirmed by PCR. One such strain was designated sAA620. Strain sAA620 has one additional copy of POX5 under control of the POX4 promoter.

30

Plasmid pAA156

A PCR product containing the nucleotide sequence of CYP52A19 was amplified from Candida strain 20336 genomic DNA, using primers oAA525 and oAA526. The PCR product was gel

purified and ligated into pCR-Blunt II-TOPO, transformed as described, and clones containing PCR inserts were sequenced to confirm correct DNA sequence. One such plasmid was designated, pAA144. Plasmid pAA144 was digested with BspQI and a 1.7-kb fragment was isolated. Plasmid pAA073, which includes a POX4 promoter and POX4 terminator, also was digested with BspQI
 5 and gel purified. The isolated fragments were ligated together to generate plasmid, pAA156. Plasmid pAA156 included $P_{POX4}CYP52A19T_{POX4}$ fragment and URA3.

Strain sAA496

10 Plasmid pAA156 was digested with ClaI and column purified. Strain sAA451 was transformed with this linearized DNA and plated on SCD-ura plate. Colonies were checked for CYP52A19 integration. Colonies positive for plasmid integration were further analyzed by qPCR to determine the number of copies of CYP52A19 integrated. One such strain, designated sAA496 contained about 13 copies of the monooxygenase activity encoded by CYP52A19.

15

Strains sAA632 and sAA635

Strain sAA620 was transformed with linearized pAA156 DNA and plated on SCD-ura plates. Several colonies were checked for CYP52A19 integration. Colonies positive for plasmid
 20 integration were further analyzed by qPCR to determine the number of copies of CYP52A19 integrated. One such strain, designated sAA632 contained about 27 copies of the monooxygenase activity encoded by CYP52A19. Another strain, designated sAA635, contained about 12 copies of the monooxygenase activity encoded by CYP52A19.

25 *Example 20: Cloning of Candida ACH genes*

ACH PCR product was amplified from *Candida strain ATCC20336* genomic DNA using primers oAA1095 and oAA1096, shown in the table below. The PCR product was gel purified and ligated into pCR-Blunt II-TOPO (Invitrogen), transformed into competent TOP10 *E. coli* cells (Invitrogen)
 30 and clones containing PCR inserts were sequenced to confirm correct DNA sequence.

Sequence analysis of multiple transformants revealed the presence of allelic sequences for the ACH gene, which were designated ACHA and ACHB. A vector containing the DNA sequence for

the ACHA allele was generated and designated pAA310. A vector containing the DNA sequence for the ACHB allele was generated and designated pAA311.

Primer	sequence
oAA1095	CACACACCCGGGATGATCAGAACCGTCCGTTATCAAT
oAA1096	CACACATCTAGACTCTCTTCTATTCTTAATTGCCGCTTCCACTAAACGG CAAAGTCTCCACG

5 *Example 21: Cloning of Candida FAT1 gene*

FAT1 PCR product was amplified from *Candida* 20336 genomic DNA using primers oAA1023 and oAA1024, shown in the table below. The PCR product was gel purified and ligated into pCR-Blunt II-TOPO (Invitrogen), transformed into competent TOP10 *E. coli* cells (Invitrogen) and clones containing PCR inserts were sequenced to confirm correct DNA sequence. A vector containing the DNA sequence for the *FAT1* gene was designated pAA296.

Primer	sequence
oAA1023	GATATTATTCCACCTTCCCTTCATT
oAA1024	CCGTTAAACAAAAATCAGTCTGTAAA

15 *Example 22: Cloning of Candida ARE1 and ARE2 genes*

ARE1 and *ARE2* PCR products were amplified from *Candida* 20336 genomic DNA using primers oAA2006/oAA2007 and oAA1012/oAA1018, respectively, shown in the table below. The PCR products were gel purified and ligated into pCR-Blunt II-TOPO (Invitrogen), transformed into competent TOP10 *E. coli* cells (Invitrogen) and clones containing PCR inserts were sequenced to confirm correct DNA sequence. A vector containing the DNA sequence for the *ARE1* gene was designated pAA318. A vector containing the DNA sequence for the *ARE2* gene was designated pAA301.

Primer	sequence
oAA1012	ATGTCCGACGACGAGATAGCAGGAATAGTCAT
oAA1018	TCAGAAGAGTAAATACAACGCACTAACCAAGCT
oAA2006	ATGCTGAAGAGAAAGAGACAACTCGACAAG
oAA2007	GTGGTTATCGGACTCTACATAATGTCAACG

Example 23: Construction of an optimized TESA gene for expression in Candida

- The gene sequence for the *E. coli* *TESA* gene was optimized for expression in *Candida* by codon replacement. A new *TESA* gene sequence was constructed using codons from *Candida* with similar usage frequency for each of the codons in the native *E. coli* *TESA* gene (avoiding the use of the CTG codon due to the alternative yeast nuclear genetic code utilized by *Candida*). The optimized *TESA* gene was synthesized with flanking BspQI restriction sites and provided in vector pIDTSMART-Kan (Integrated DNA Technologies). The vector was designated as pAA287.
- Plasmid pAA287 was cut with BspQI and the 555bp DNA fragment was gel purified. Plasmid pAA073 also was cut with BspQI and the linear DNA fragment was gel purified. The two DNA fragments were ligated together to place the optimized *TESA* gene under the control of the *Candida* *POX4* promoter. The resulting plasmid was designated pAA294.

Example 24: Cloning of Candida DGA1 gene

- DGA1* PCR product was amplified from *Candida* 20336 genomic DNA using primers oAA996 and oAA997, shown in the table below. The PCR product was gel purified and ligated into pCR-Blunt II-TOPO (Invitrogen), transformed into competent TOP10 *E. coli* cells (Invitrogen) and clones containing PCR inserts were sequenced to confirm correct DNA sequence. A vector containing the DNA sequence of the *DGA1* gene was designated pAA299.

Primer	Sequence
oAA996	ATGACTCAGGACTATAAAGACGATAGTCCTACGTCCACTGAGTTG
oAA997	CTATTCTACAATGTTTAATTCAACATCACCGTAGCCAAACCT

Example 25: Cloning of Candida LRO1 gene

LRO1 PCR product was amplified from *Candida* 20336 genomic DNA using primers oAA998 and oAA999, shown in the table below. The PCR product was gel purified and ligated into pCR-Blunt II-TOPO (Invitrogen), transformed into competent TOP10 *E. coli* cells (Invitrogen) and clones containing PCR inserts were sequenced to confirm correct DNA sequence. A vector containing the DNA sequence of the *LRO1* gene was designated pAA300.

Primer	sequence
oAA998	ATGTCGTCTTTAAAGAACAGAAAATC
oAA999	TTATAAATTTATGGCCTCTACTATTTCT

10 *Example 26: Cloning of Candida ACS1 gene and construction of deletion cassette*

ACS1 PCR product was amplified from *Candida* 20336 genomic DNA using primers oAA951 and oAA952, shown in the table below. The PCR product was gel purified and ligated into pCR-Blunt II-TOPO (Invitrogen), transformed into competent TOP10 *E. coli* cells (Invitrogen) and clones containing PCR inserts were sequenced to confirm the DNA sequence. One such plasmid was designated pAA275. Plasmid pAA280 was digested with BamHI to release a 2.0 kb $P_{URA3}URA3T_{URA3}P_{URA3}$ cassette. Plasmid pAA275 was digested with BglII and gel purified. The two pieces were ligated together to generate plasmid pAA276 and pAA282. Plasmid pAA276 and pAA282 have the $P_{URA3}URA3T_{URA3}P_{URA3}$ cassette inserted into the *ACS* gene in opposite orientations.

Primer	sequence
oAA951	CCTACTTCCACAGCTTTAATCTACTATCAT
oAA952	TTTAAGAAAACAACCTAAGAGAAGCCAC

Example 27: Construction of Strain sAA722 (pox4a::ura3/pox4b::ura3 POX5/POX5 ACS1/acs1:: $P_{URA3}URA3T_{URA3}P_{URA3}$)

25 Plasmid pAA276 was digested with BamHI/XhoI and column purified. Strain sAA329 (*ura3/ura3 pox4a::ura3/pox4b::ura3 POX5/POX5*) was transformed with the linearized DNA and plated on

SCD-ura plate. Several colonies were checked for *ACS1* disruption. One such strain was designated sAA722.

Example 28: Construction of Strain sAA741 (pox4a::ura3/pox4b::ura3 POX5/POX5

5 *ACS1/acs1::P_{URA3})*

Strain sAA722 was grown in YPD media overnight and plated on 5-FOA plate. Colonies that grew in the presence of 5-FOA were PCR screened for the looping out of the *URA3* gene leaving behind only the *URA3* promoter (*P_{URA3}*) in the *ACS1* site. Out of 30 colonies analyzed, only one strain
10 showed the correct genetic modification. The strain was designated sAA741.

Example 29: Construction of Strain sAA776 (pox4a::ura3/pox4b::ura3 POX5/POX5

acs1::P_{URA3}URA3T_{URA3}P_{URA3}/acs1::P_{URA3})

15 Plasmid pAA282 was digested with BamHI/XhoI and column purified. Strain sAA741 (see Example 28) was transformed with the linearized DNA and plated on SCD-ura plate. Several colonies were checked for double *ACS1* knockout by insertional inactivation. One such strain was designated sAA776.

20 *Example 30: Construction of Strain sAA779 (pox4a::ura3/pox4b::ura3 POX5/POX5 acs1::P_{URA3}/acs1::P_{URA3})*

Strain sAA776 (see Example 29) was grown in YPD media overnight and plated on 5-FOA plates. Colonies that grew in the presence of 5-FOA were PCR screened for the looping out of the *URA3*
25 gene leaving behind only the *URA3* promoter (*P_{URA3}*) in both *ACS1* copies. One such strain was designated sAA779.

Example 31: Construction of Strain sAA811 (pox4a::ura3/pox4b::ura3 POX5/POX5 acs1::

P_{URA3}/acs1::P_{URA3} ura3::3xP_{POX4}P450A19)

30

Plasmid pAA156 containing a *P450A19* integration cassette was digested with ClaI and column purified. Strain sAA779 (see Example 30) was transformed with the linearized DNA and plated on SCD-ura plate. Several colonies were checked for *P450A19* integration. From those colonies,

qPCR was performed to check the copy number of *P450A19* integration. One strain, designated sAA811, contained 3 copies of *P450A19*.

Example 32: Construction of Strain sAA810 (pox4a::ura3/pox4b::ura3 POX5/POX5 acs1::

5 *P_{URA3}/acs1::P_{URA3} ura3::5xP_{POX4}P450A19 ura3::8xP_{POX4}TESA)*

Plasmid pAA156 containing a P450-A19 integration cassette was digested with Clal and column purified. Plasmid pAA294 containing a *TESA* integration cassette also was digested with Clal and column purified. Strain sAA779 was cotransformed with both linearized DNAs and plated on SCD-
10 ura plate. Several colonies were checked for both *P450A19* integration and *TESA* integration. Colonies that were positive for both *TESA* and *P450A19* were further analyzed by qPCR. qPCR was performed to check the copy number of the *P450A19* and *TESA* integration events. One strain, designated sAA810, contained 5 copies of *P450A19* and 8 copies of *TESA*.

15 *Example 33. General Techniques & Methods.*

Growth Media, Reagents and Conditions

YPD, ScD-ura media and plates, and 5-FOA containing plates were made as described in Methods
20 in Yeast Genetics: a Cold Spring Harbor Laboratory Manual /David C. Amberg, Daniel J. Burke, Jeffrey Strathern, - 2005 ed.).

SP92 + glycerol was made by adding 6.7 g of Bacto yeast nitrogen base without amino acids (BD, Franklin Lakes, NJ, USA), 3.0 g of Bacto yeast extract (BD, Franklin Lakes, NJ, USA), 3.0 g of
25 ammonium sulfate, 1.0 g of potassium phosphate monobasic, 1.0 g of potassium phosphate dibasic, and 75 g of glycerol to water to a final volume of one liter. The media was then filtered sterilized.

TB-low N Media was made by adding 1.7 g Bacto yeast nitrogen base without ammonium sulfate,
30 3 g of Bacto yeast extract, 1 g of potassium phosphate monobasic and 1g potassium phosphate dibasic per liter of water. The media was filtered sterilized.

Overnight cultures were typically grown in 2 to 5 ml of either ScD-ura media or YPD media in standard culture tubes at 30 C on a shaker at about 250 rpm.

Molecular Methods

Gel purifications of DNA fragments were done as recommended by the manufacturer using either the GeneJET Gel Extraction Kit (Fermentas Inc, Glen Burnie, MD USA) or the Zymoclean Gel DNA
5 Recovery Kit (ZymoResearch, Irvine, CA).

PCR was performed using either PFU Ultra II DNA Polymerase (Agilent Technologies, Santa Clara, CA), Taq DNA polymerase (New England Biolabs, Ipswich, MA, USA), DreamTaq PCR Master Mix (Fermentas Inc, Glen Burnie, MD USA) or Quick Load Midas Mix (Monserate, San Diego, CA
10 USA). Each enzyme was used according to the manufacturer's instructions.

Restriction enzyme digestions were conducted as recommended by each manufacturer (New England Biolabs, Ipswich, MA, USA or Fermentas Inc, Glen Burnie, Maryland USA). DNA ligations were conducted using either the Rapid Ligation Kit (Fermentas Inc, Glen Burnie, Maryland USA) or
15 using T4 DNA Ligase (New England Biolabs, Ipswich, MA, USA) according to the manufacturer's instructions.

Yeast transformations were performed as described in Example 10.

20 Genomic DNA Preparation

The URA3 gene was obtained from genomic DNA of Candida yeast culture ATCC20336.

Genomic DNA from Candida strain ATCC20336 was prepared as follows: About 1.5 ml of an
25 overnight culture of cells was and the pellet was resuspended in about 200 µl of a solution containing 2% Triton X-100, 1% SDS, 100 mM NaCl, 10 MM Tris pH 8.0, and 1 mM EDTA. About 200 µl of acid washed glass beads were added with about 200 µl of phenol:chloroform:isoamyl alcohol (25:24:1) at a pH of about 8.0. The sample was vortexed for about 2 minutes after which about 200 µl of water was added. The sample was then centrifuged at 13000 rpm for about 10
30 minutes. The aqueous layer was transferred to a new microcentrifuge tube and an equal volume of chloroform:isoamyl alcohol (24:1) solution was added. This sample was vortexed for 10 seconds and then centrifuged at 13000 rpm for about 2 minutes. The aqueous layer was transferred to a new microfuge tube and 1 ml of ethanol was added. The tube was then placed at -80 °C for about 15 minutes and then spun at 13000 rpm for 15 minutes to pellet the DNA. The DNA

was washed with 70% ethanol and air-dried. The DNA was then resuspended in about 500 µl of water.

Genomic DNA for *Kluyveromyces lactis* (ATCC8585) was purchased from the American Type
5 Culture Collection (Manassas, VA, USA).

To calculate gene copy number, a qPCR method was used as described by Jin et al (Appl. Environ. Microbiol. January 2003 vol. 69, no. 1, 495-503). qPCR was performed according to the manufacturer's instructions using either the Brilliant III Ultra-Fast SYBR® Green QPCR Master Mix
10 (Agilent Technologies, Englewood, CO USA) or the QuantiTect Multiplex PCR NoROX Kit (Qiagen). Genomic DNA from *Candida* strain ATCC20336 or plasmid DNA containing the actin gene from ATCC20336 and the gene of interest were used as standards.

Primers and probes used throughout these Examples were made via standard DNA synthesis
15 techniques by Integrated DNA Technologies (Coralville, IA, USA).

Example 34: Construction of Cloning Plasmid AA073

The plasmid pAA073 was designed to contain the POX4 promoter and terminator from *Candida*
20 strain ATCC20336 (this strain is also referred to herein as strain sAA001). This plasmid was derived from the publicly available plasmid pUC19 which contains an ampicillin resistance marker. pAA073 was designed to have two Sapl restriction enzyme sites located between the POX4 promoter and terminator which allows unidirectional cloning of any gene of interest in tandem with the POX4 promoter. The *Candida* strain ATCC20336 *URA3* gene including the open reading
25 frame and the endogenous regulatory regions was also placed into pAA073 as a selection marker for transformants. Plasmid pAA073 allows the direct integration of multiple copies of any gene of interest by digesting the plasmid with a unique restriction enzyme such as SpeI, ClaI or BstZ171. These multiple cloning sites for are contained in the *URA3* auxotrophic marker region and can be
30 selectively be used to avoid cutting the gene of interest (i.e., the DNA sequence for the gene of interest can be searched for particular restriction enzyme cut sites and those enzymes can be avoided). In addition, this plasmid can serve as a template to generate an antibiotic free-DNA cassette containing the gene of interest and the POX 4 regulatory regions inserted between the 3' and 5' regions of the *URA3* gene; this cassette can be PCR amplified using the plasmid as a template, and the isolated PCR product can be inserted into any microorganism strain.

A diagram of pAA073 is set forth in **FIG. 18** and the sequence of pAA073 is set forth as SEQ ID NO. 3704.

5

Example 35. *Cloning enoyl-CoA isomerase (Eci) genes from ATCC 20336*

The amino acid sequence for Eci1 (i.e. Eci) from *S. cerevisiae* S288c (SEQ ID NO. 3705) was used to identify homologs from *Candida* species ATCC MYA-3404 and ATCC20336. The BLAST search revealed two Eci1p homologs in each strain of *Candida*, which have been named Eci1p and Eci2p (see TABLE 1). The percent amino acid identities for the homologs are shown below:

TABLE 1

	Amino acid percent identity				
	Eci2p_MYA-3404 SEQ ID NO. 3707	Eci1p_MYA-3404 SEQ ID NO. 3706	Eci1p_S.c. SEQ ID NO. 3705	Eci2p_20336 SEQ ID NO. 3709	Eci1p_20336 SEQ ID NO. 3708
Eci2p_MYA-3404 SEQ ID NO. 3707		58	36	84	57
Eci1p_MYA-3404 SEQ ID NO. 3706			39	57	92
Eci1p_S.c. SEQ ID NO. 3705				37	40
Eci2p_20336 SEQ ID NO. 3709					57
Eci1p_20336 SEQ ID NO. 3708					

10

The ECI1 gene encoding the N-terminal 241 residues of SEQ ID NO. 3708 was amplified from genomic DNA (ATCC 20336) using oligonucleotides oAA2835 (SEQ ID NO. 3712) and oAA2836 (SEQ ID NO. 3713) that also incorporated unique SapI restriction sites. The 770 bp PCR product was gel purified and ligated into pCR-Blunt II-TOPO (Life Technologies), transformed into competent TOP10 *E. coli* cells (Life Technologies) and clones containing PCR inserts were sequenced to confirm the correct DNA sequence. One such plasmid was named pAA574 (SEQ ID NO. 3710).

15

The full length ECI2 gene encoding Eci2p (SEQ ID NO. 3709) was amplified from genomic DNA (ATCC 20336) using oligonucleotides oAA2837 (SEQ ID NO. 3714) and oAA2838 (SEQ ID NO. 3715) that also incorporated unique SapI restriction sites. The 851 bp PCR product was gel purified and ligated into pCR-Blunt II-TOPO (Life Technologies), transformed into competent TOP10 E. coli cells (Life Technologies) and clones containing PCR inserts were sequenced to confirm the correct DNA sequence. One such plasmid was named pAA575 (SEQ ID NO. 3711).

Example 36 - Generation of Strain sAA1764 (ura3/ura3 pox4a::ura3/pox4b::ura3 POX5/POX5 acs1:: PURA3/acs1:: PURA3 fat1-Δ1::PURA3/fat1-Δ2::PURA3 eci1-Δ1::URA3/ECI1)

10

Deletion of the first allele of ECI1 was achieved by transforming cells (strain sAA886 (pox4a::ura3/pox4b::ura3 POX5/POX5 acs1:: PURA3/acs1:: PURA3 fat1-Δ1::PURA3/fat1-Δ2::PURA3 ura3/ura3)) with linear DNA cassettes constructed by overlap extension PCR (OE-PCR). A deletion cassette for the first ECI1 allele in strain sAA886 was generated from three DNA fragments. A first DNA fragment (ECI1 5' homology) was amplified from ATCC20336 gDNA using primers oAA3085 (SEQ ID NO. 3716) and oAA3086 (SEQ ID NO. 3717). A second DNA fragment (PURA3URA3TURA3PURA3) was amplified from plasmid pAA298 (FIG. 29, and SEQ ID NO: 3784) using primers oAA3087 (SEQ ID NO. 3718) and oAA3088 (SEQ ID NO. 3719). The third DNA fragment (ECI1 3' homology) was amplified from ATCC20336 gDNA using primers oAA3089 (SEQ ID NO. 3720) and oAA3090 (SEQ ID NO. 3721). All three DNA fragments were combined in the same reaction to generate the full-length deletion cassette (FIG. 19) by OE-PCR using primers oAA3085 (SEQ ID NO. 3716) and oAA3090 (SEQ ID NO. 3721).

Strain sAA886 was transformed with the full-length deletion cassette and plated on SCD-Ura plate. Several colonies were screened by PCR for integration of the deletion cassette at the first ECI1 allele. One such strain was named sAA1764.

Example 37. Generation of Strain sAA1860 (ura3/ura3 pox4a::ura3/pox4b::ura3 POX5/POX5 acs1:: PURA3/acs1:: PURA3 fat1-Δ1::PURA3/fat1-Δ2::PURA3 eci1-Δ1::PURA3/ECI1)

Strain sAA1764 was grown in YPD media overnight and plated on 5-FOA plate. Colonies that grew in the presence of 5-FOA were PCR screened for the looping out of the URA3 gene leaving behind only the URA3 promoter (PURA3) in the first ECI1 allele. One such strain was named sAA1860.

Example 38. Construction of a double *ECI1* knockout strain (*ura3/ura3 pox4a::ura3/pox4b::ura3 POX5/POX5 acs1:: PURA3/acs1:: PURA3 fat1-Δ1::PURA3/fat1-Δ2::PURA3 eci1-Δ1::PURA3/eci1-Δ2::URA3*)

- 5 Deletion of the second allele of *ECI1* is achieved by transforming cells with linear DNA cassettes constructed by overlap extension PCR (OE-PCR). A deletion cassette for the second *ECI1* allele in sAA1860 (*ura3/ura3 pox4a::ura3/pox4b::ura3 POX5/POX5 acs1:: PURA3/acs1:: PURA3 fat1-Δ1::PURA3/fat1-Δ2::PURA3 eci1-Δ1::PURA3/ECI1*) was generated from three DNA fragments. A first DNA fragment (*ECI1* 5' homology) was amplified from ATCC20336 gDNA using primers
- 10 oAA3212 (SEQ ID NO. 3722) and oAA3213 (SEQ ID NO. 3723). A second DNA fragment (*PURA3URA3TURA3PURA3*) was amplified from plasmid pAA298 (FIG. 29, and SEQ ID NO: 3784) using primers oAA3214 (SEQ ID NO. 3724) and oAA3215 (SEQ ID NO. 3725). A third DNA fragment (*ECI1* 3' homology) was amplified from ATCC20336 gDNA using primers oAA3216 (SEQ ID NO. 3726) and oAA3217 (SEQ ID NO. 3727). All three DNA fragments were combined in the
- 15 same reaction to generate the full-length deletion cassette (FIG. 20) by OE-PCR using primers oAA3212 (SEQ ID NO. 3722) and oAA3217 (SEQ ID NO. 3727).

To generate a double *ECI1* knockout strain, sAA1860 is transformed with the full-length deletion cassette and plated on SCD-Ura plate. Several colonies are screened by PCR for integration of

20 the deletion cassette at the second *ECI1* allele.

EXAMPLE 39 - Cloning of acyl CoA oxidase proteins

Acyl-CoA oxidases from a range of organisms were cloned into the *E. coli* expression vector

25 pET26b (EMD4Biosciences, Darmstadt, Germany), which contains a kanamycin resistance cassette. The source of the acyl-CoA oxidase, the name of the gene, the primers and restriction enzymes used to clone the acyl CoA oxidase coding sequence into pET26b and the coding sequence are described herein. The acyl CoA oxidase coding sequences were amplified by PCR using the appropriate primers designed from cDNA libraries, published cDNA or genomic DNA

30 sequences corresponding to the organism. In the event that a template source was not available, the coding sequences were synthesized as gBlocks (IDT) and stitched together by standard overlap extension PCR. The PCR products were then cloned into pCRII-Blunt TOPO vector (Life Sciences) and the products were sequenced to verify that they did not contain undesired mutations. The coding sequences were released from the TOPO vector using the appropriate

restriction enzymes and ligated into pET26b that had been digested with the same restriction enzymes. The resulting expression plasmids were then transformed into Rosetta II BL21 cells (Novagen).

5 *Example 40 - Expression of acyl-CoA oxidases in E. coli*

To express an enzyme, a colony from each transformation of Rosetta cells was used to start a 5 ml overnight culture of LB containing the antibiotics kanamycin (to select for pET26b) and chloramphenicol (to select for a second plasmid found in Rosetta II cells that mediates improved translation of eukaryotic proteins expressed in *E. coli*) grown at 37°C. The next morning, the
 10 overnight culture was used to seed 30 ml of LB containing kanamycin and chloramphenicol to an OD₆₀₀ reading of 0.1. The 30 ml cultures were grown at 37°C for 2 hours and then placed on ice for 10 minutes. To induce expression, isopropyl-beta-D-thiogalactopyranoside (IPTG) was added to the culture to a final concentration of 0.1 mM. In some cases induction was performed using
 15 Novagen Overnight Express Autoinduction System 1 (CAT# 71300-3). The cells were then shaken at 15°C overnight to express the acyl CoA oxidase.

Example 41 - Acyl-CoA Oxidase Activity Assay

To test the activity of the acyl CoA oxidase, cells from the overnight induction were pelleted at 1046 x g at 4°C in 50 ml conical tubes. The cell pellets were resuspended in 1 ml of 50mM KPO₄, pH 7.6, 50 µM FAD buffer and then transferred to a 2 ml centrifuge tube. A Misonix Sonicator 3000 (QSonica, Newtown, CT) was used to lyse the cells, which were sonicated at a power setting of 2 for 2 pulses of 20 seconds each. The lysates were placed on ice for 30 seconds in between
 25 each pulse. To obtain a supernatant, cell debris was pelleted at 16,100 x g for 10 mins in a 4°C microcentrifuge. The supernatant was transferred to a 1.5 ml centrifuge tube and Bradford assays (Thermoscientific) were performed on cell lysates according to manufacturer's specifications to determine protein concentration in preparation for the acyl-CoA oxidase assays. A Beckman Coulter™ DTX-800 Multimode Detector spectrophotometer was used for the assays. The
 30 spectrophotometer was set to read for 5 minutes at 500 nm, 30°C. Each reaction was 200 µl in volume and contained 10 µg of cell lysate in 50mM KPO₄, pH 7.6, 200 µg/ml BSA, 0.05% Triton X-100, 250µM fatty acyl-coA substrate, 50µM FAD, 10U horseradish peroxidase, 25mM p-hydroxybenzoic acid and 1mM 4-aminoantipyrine. Fatty acyl-CoA substrates covered a range from hexanoyl CoA (six carbon chain length) to oleoyl CoA (eighteen carbon chain length).

In TABLE 2 unshaded blocks indicated the sample was not tested. Dark shading indicates that no activity was detected. Light shading indicates that minimal activity (i.e. poor activity) was detected at less than or equal to 0.1 umol/min/ug (umol substrate/minute/ug total protein). Medium shading indicates that good activity was detected at > 0.1 umol/min/ug.

5

The results in TABLE 2 indicated that several enzymes are not functional when expressed in E.coli. Furthermore, the remaining enzymes that are functional when expressed in E.coli showed broad substrate specificity or were similar in their substrate specificity to Pox5 from Candida strain ATCC20336 (i.e. not very active on a C6 substrate, show peak activity on a C12 substrate and are

10

TABLE 2 – Activity of Heterologous Acyl CoA Oxidases

E. coli Strain number	Description	C6	C8	C10	C12	C14	C16	C18:1
1570	<i>Aspergillus nidulans</i> (An)AOXa in pET26b							
1557	AnAOXa (N-term. GST fusion) in pET26b							
1573	<i>Arabidopsis thaliana</i> (At)ACX2L (N-term. GST fusion) in pET26b							
1514	AtACX2L in pGEX							
1503	AtACX2L in pET26b							
1508	AtACX2L (N-term. GST fusion) in pET26b							
1510	AtACX1 in pET26b							
1511	AtACX3 in pET26b							
1366	AtACX2S in pET26b							
1507	AtACX2S (N-term. GST fusion) in pET26b							
1525	AtACX2S (N-term. GST fusion) in pET26b							
1663	AtACX2 N terminus-Pox5 C-terminus fusion in pET26b							
1369	AtACX5 in pET26b							
1362	Candida strain ATCC20336 Pox5 in pET26b							
1888	Candida strain ATCC20336 Pox4 in pET26b							
1365	<i>Debaryomyces hansenii</i> (Dh)7248p in pET26b							
1572	<i>Glycine max</i> (Gm)Aco2 pET26b							
1558	GmAco2 (N-term. GST fusion) in pET26b							
1515	GmAco2 in pGEX							
1367	<i>Pichia stipitis</i> (Ps)Acox2 in pET26b							
1504	<i>Rattus norvegicus</i> (Rn)Aco1 in pET26b							
1505	RnAco1b in pET26b							
1574	RnAco1 (N-term. GST fusion) in pET26b							

E. coli Strain number	Description	C6	C8	C10	C12	C14	C16	C18:1
1575	RnAco1b (N-term. GST fusion) in pET26b							
1501	<i>Yarrowia lipolytica</i> (Yl)Aco1 in pET26b							
1361	YlAco2 in pET26b							
1502	YlAco3 in pET26b							
1499	YlAco4 in pET26b							
1509	YlAco4 (N-term. GST fusion) in pET26b							
1500	YlAco5 in pET26b							
1368	YlAco6 in pET26b							
1363	<i>Zea mays</i> (Zm)Aco1 in pET26b							
1552	ZmAco2 in pET26b							
1555	ZmAco2 (N-term. GST fusion) in pET26b							
1364	ZmAco4 pET26b							
1506	ZmAco4 (N-term. GST fusion) in pET26b							
1795	<i>Cucumis sativus</i> Aco in pET26b							
1775	<i>Arthrobacter ureafaciens</i> Aco in pET26b							
1776	<i>Salinobacter ruber</i> Aco in pET26b							
1777	<i>Thermobifida fusca</i> Aco in pET26b							

Example 42 - Genetic Modification of *Candida Pox4*, *Pox5* and *R. norvegicus* VLCAD

The objective was to design mutations in 1) the *Pox4* and *Pox5* acyl CoA oxidases of *Candida* strain ATCC20336 (*Pox4* and *Pox5*, respectively) to alter their respective substrate specificities and 2) the *R. norvegicus* very long chain acyl-CoA dehydrogenase (VLCAD) to convert it into an acyl CoA oxidase. When introduced into *Candida*, these mutant enzymes may mediate selective conversion of fatty acid substrates to sebacic, dodecanedioic acid or longer chain diacids by beta oxidation.

10

Site-directed mutagenesis of *Pox4* and *Pox5*- Methodology

Several approaches were used to identify regions and/or residues of *Pox4* and *Pox5* of *Candida* strain ATCC20336 that determine the substrate specificities of these enzymes. In rat liver, a single gene with an alternatively spliced third exon produces two spliceoforms, AcoI (acyl CoA oxidase-I, *R. norvegicus*, *RnAcoI*) and AcoII (acyl CoA oxidase-II, *R. norvegicus*, *RnAcoII*), which are identical in amino acid length and differ in amino acid sequence only at the region encoded by the differentially spliced exon (Miyazawa, S., Hayashi, H., Hijikata, M., Ishii, N., Furuta, S.,

15

Kagamiyama, H., Osumi, T., Hashimoto, T. (1987) *Complete nucleotide sequence of cDNA and predicted amino acid sequence of rat acyl-CoA oxidase*. J. Biol. Chem. 262(17):8138-43.; Osumi, T., Ishii, N., Miyazawa, S., Hashimoto, T. (1987) *Isolation and structural characterization of the rat acyl-CoA oxidase gene*. J. Biol. Chem. 262 (17):8138-43; Setoyama, C., Tamaoki, H., Nishina, Y., Shiga, K., Miura, R. (1995) *Functional expression of two forms of rat acyl-CoA oxidase and their substrate specificities*. Biochem. Biophys. Res. Commun. 217(2):482-7). A comparison of the primary amino acid sequences of Acol and Acoll revealed differences in residues 90 to 133 as a result of the alternatively spliced exon (underlined residues, FIG. 23). The splicing event resulted in two enzymes, Acol and Acoll, that display different substrate activity profiles. RnAcol prefers substrates with few carbons (e.g., fatty acids with 8 or 10 carbons). RnAcoll prefers substrates with longer carbon chains (e.g., 14 carbons). The crystal structure of RnAcoll has been solved (PDB: 1IS2 (without substrate); PDB: 2DDH (with dodecanoate substrate)) and the region encoded by the alternatively spliced exon ends at the boundary between the N-terminal alpha helical domain and the subsequent beta sheet domain, both of which are characteristic structural features of acyl CoA oxidases (Acyl-CoA dehydrogenase (ACAD) superfamily, NCBI Conserved Domains Accession cl0993) and have been identified as a region that may determine substrate specificity.

To verify that this region of an acyl CoA oxidase plays a role in determining substrate specificity, the HotSpot Wizard algorithm was utilized (Pavelka, A., Chovancova, E., Damborsky, J. HotSpot Wizard: a Web Server for Identification of Hot Spots in Protein Engineering, Nucleic Acids Research 37: W376-W383, 2009. <http://loschmidt.chemi.muni.cz/hotspotwizard/>). HotSpot Wizard is a program that identifies regions of a protein for engineering of substrate specificity or activity. The program utilizes structural, functional and sequence homology data from numerous databases, such as PDB, UniProt and NCBI, to identify regions and/or residues that are "hot spots" for mutagenesis. The search relies on a PDB file corresponding to a crystal structure of the enzyme of interest. In the case of Pox4 or Pox5, no such structure was available. Therefore, the structures of both proteins were determined by modeling with the crystal structure of *R. norvegicus* Acoll as the template (PDB:1IS2). The SWISS-MODEL program was used to generate these models (Arnold K., Bordoli L., Kopp J., and Schwede T. (2006). *The SWISS-MODEL Workspace: A web-based environment for protein structure homology modelling*. Bioinformatics, 22,195-201; Kiefer F, Arnold K, Künzli M, Bordoli L, Schwede T (2009) The SWISS-MODEL Repository and associated resources. Nucleic Acids Research. 37, D387-D392. Peitsch, M. C. (1995) Protein modeling by E-mail Bio/Technology 13: 658-660; <http://swissmodel.expasy.org/>). The resulting models, which were PDB files, were entered in HotSpot Wizard as the "Query structure". The

results of a HotSpot Wizard analysis are summarized in FIG 24A, 24B and 25). Residues highlighted in grey are proposed mutagenic “hot spots”. Dark grey shading indicates residues with greater variability than those with light grey shading. Residues shown in bold are found within or close to the substrate binding pocket (discussed below).

5

All three enzymes were aligned to show areas of homology (FIG. 26). In FIG. 26 light grey shading indicates identity between all three enzymes. Darker shades of grey indicate partial identity or homology between the three enzymes (e.g., a dark grey shading may indicate identity for two of the three proteins in the alignment). In some cases, dark grey shading indicates sequence
10 similarity (i.e., the residues are similar because they are acidic, basic, polar or non-polar). The underlined region (FIG. 23 and FIG. 26) indicates the alternatively spliced exon of AcoII.

Molecular modeling alignments were used to identify residues in Pox4 and Pox5 that are found within or close to the substrate binding pocket (residues shown in bold, FIG. 24A and 24B). The
15 molecular structure of AcoII complexed with its substrate dodecanoate (PDB:2DDH) as determined from its crystal structure, was aligned with the predicted molecular models of Pox4 and Pox5. Residues located in the N-terminal loop and first part of alpha helix D (TABLE 3) appear at the surface and lining of the substrate entry/exit channel. In Pox4 these residues correspond to the sequence IDTFNK (a.a. 95-100 of Pox4 from Candida strain ATCC20336) and PDQQAQ (a.a. 80-
20 85 of Pox5 from Candida strain ATCC20336). According to HotSpot Wizard this entire sequence is category 9, which means highly variable.

TABLE 3 - *N-terminal loop and first part of alpha helix D*

Protein	Sequence	Residue(s)
1IS2 (ACOII)	ISDPEE	79-84
ACOI	ISDPEE	79-84
Pox4	IDTFNK	95-100
Pox5	PDQQAQ	80-85

25

Residues located in the loop between alpha helices D and E' form part of the substrate binding pocket. None of these residues were identified as contact residues for the 12 carbon substrate but may be contact residues for longer substrates. This stretch of four amino acids is located within the divergent exon splice site of ACO-I and ACO-II. In Pox4 and Pox5, three of these amino acids are
30 highly conserved (TABLE 4). The fourth amino acid is different (113G in Pox4 & F98 in Pox5). Of the four amino acids in this region, the divergent residue is closest to the substrate and the amino

acid character of this residue is drastically different between Pox4 and Pox5. For this reason, this residue is of particular interest. Additionally, according to HotSpot this residue is highly variable.

TABLE 4 - *Loop between alpha helices D and E'*

5

Protein	Sequence	Residue(s)
1IS2 (ACOI)	RGHP	94-97
ACOI	ANFV	94-97
Pox4	PQVG	110-113
Pox5	PQVF	95-98

The residue D101 is a contact residue for the substrate carbons 6 through 9 in the 2DDH crystal structure of RnACOI. This residue is located at the beginning of alpha helix E' which is part of the substrate binding pocket. Since this is a contact residue and is located in the small region of sequence that differs between ACO-I and ACO-II, the corresponding amino acid in Pox4 and Pox5 (TABLE 5) is of interest. This residue is of interest since it contacts the substrate at carbons 6-9 and ACO-II has lower activity on substrates of chain-length 6-12 compared to ACO-I. If either Pox4 or Pox5 is modified at this position to aspartate, it is expected that there would be a decrease in activity on adipic acid and lead to increases in yield of larger diacids. This residue has a score of 6 from HotSpot.

10

15

TABLE 5 - *Residue making contact with substrate carbons 6-9*

Protein	Sequence	Residue(s)
1IS2 (ACOI)	D	101
ACOI	G	101
Pox4	G	117
Pox5	G	102

Residue F284 is a contact residue for the substrate carbons 10 through 12 in the 2DDH crystal structure of RnACOI. This residue is conserved between RnACOI and RnACOI. The corresponding amino acid in Pox4 and Pox5 (TABLE 6) is one of the very few substrate contact residues that differ between Pox4 and Pox5. The ACOI, ACOI, and Pox4 enzymes all have a large hydrophobic residue at this location whereas the Pox5 enzyme has a small polar residue. The HotSpot score for this residue is 9.

20

TABLE 6 - *Residue making contact with substrate carbons 10-12*

Protein	Sequence	Residue(s)
1IS2 (ACOI)	F	284
ACOI	F	284
Pox4	M	302
Pox5	T	287

The loop C-terminal to alpha helix L is much smaller in Pox5 than it is in Pox4 or ACO-I/ACO-II (TABLE 7). This loop appears to display structural flexibility and may have implications for the structure of the substrate pocket and how much the substrate-binding pocket “breathes”. The residues in this region vary in HotSpot analysis, however the residues just downstream of this region in Pox5 are all highly variable (shown next).

10 TABLE 7 - *Alpha helix L and loop C-terminal to alpha helix L*

Protein	Sequence	Residue(s)
1IS2 (ACOI)	IYDQVRSGKLVGGMVSYLNDLPSQRIQPQQVA	438-469
ACOI	IYDQVRSGKLVGGMVSYLNDLPSQRIQPQQVA	438-469
Pox4	QVISIEDAGKTVRGSTAFLNQLKDYTGSSSKV	473-505
Pox5	DLLKEPEQKGL	453-463

The loop between alpha helix L and M does not appear to be as variable between Pox4 and Pox5 (TABLE 8), although HotSpot analysis assigns this stretch of residues with scores of 9 with high variability. It is expected that this loop, including the previous section mentioned above, is a target for mutagenesis.

TABLE 8 - *Loop between α -helix L and M*

Protein	Sequence	Residue(s)
1IS2 (ACOI)	VWPTMV	470-475
ACOI	VWPTMV	470-475
Pox4	VLNTVA	506-511
Pox5	VLSSVA	464-469

20 For both Pox4 and Pox5, the HotSpot Wizard analyses, combined with molecular modeling alignments, determined that residues within the same approximate regions are good targets for

mutagenesis. The multiple sequence alignment shows that the alternatively spliced exon of AcoII overlaps with hot spot residues in all three acyl CoA oxidases (FIG. 26).

Site-directed mutagenesis of Candida Pox4 and Pox5 to alter substrate specificity - Method

5

Using the HotSpot Wizard and molecular modeling results as a guide, specific amino acids in Pox4 and Pox5 were mutated (i.e. added, deleted or substituted) by converting primarily polar or charged residues in the hot spot regions to alanine. TABLES 9A and 9B below show a summary of Candida strain ATCC20336 Pox5 and Pox4 mutations that were made and tested. The summary of the acyl CoA activity profile associated with some of the mutants in TABLES 9A and 9B are shown in FIG. 27 (Pox5) and FIG. 28 (Pox4). The number of carbons in each substrate tested is shown below each bar in FIG. 27 and FIG. 28. Pox5 Mutant I (grey highlight in TABLE 9A) results from "ACAD-based mutagenesis" (see discussion below).

15 **TABLE 9A**

Pox5			
MUTANT	POSITION	AMINO ACID(S)	MUTATION
A	81, 82	DQ	AA
B	86, 88	RLS	ALA
C	93, 94	FD	AA
D	291, 292	DS	AA
E	95, 96	PQ	AA
F	294, 295	RM	AA
G	287	T	A
H	290, 291	MD	AA
I	284/436	GE	EG
J	291	D	G
K	292	S	A
L	93	F	A
M	94	D	G
CT1	102	G	D
N	86	R	A
O	88	S	A
P	98	F	G
Q	83, 85	QAQ	AAA
CT2	453-463	DLLKEPEQKGL	QVISIEDAGKTVRGSTAFLNQLKDYTGSSNSSKV

TABLE 9B**Pox4**

MUTANT	POSITION	AMINO ACID(S)	MUTATION
	98, 99,		
A	100	FNK	AAA
B	102, 103	LS	AA
C	96	D	A
D	90	R	A
E	88	R	A
F	302	M	A
G	309, 310	RM	A
H	98	F	A
I	99	N	A
J	100	K	A
K	102	L	A
L	103	S	A
CT3	473-505	QVISIEDAGKTVRGSTAFLNQLKDYTGSSSKV	DLLKEPEQKGL

Pox4 and Pox5 from *Candida* strain ATCC20336 were cloned into pET26b for expression in *E. coli* and assayed for acyl CoA oxidase activity in vitro. The activity profiles of the genetically modified

5 Pox4 and Pox5 were compared to the activity profile of the wild type enzymes. To alter their substrate activity profile, site directed mutagenesis was performed on several locations in Pox4 and Pox5. Complementary primers encoding the point mutation(s) were used to amplify the coding sequences of Pox4 or Pox5 generating two to four PCR products that were then "stitched" together to regenerate the entire coding region using overlap extension PCR (FIG. 22). As shown in FIG.

10 22, overlap extension PCR was performed using primers A, B, C and D. Primers B and C are complementary and contain the introduced genetic modifications (e.g. mutations). PCR was performed using oligonucleotides A and B to produce a product with overlap to a PCR product generated using oligonucleotides C and D. The A-B product was used as a primer for the C-D product, and vice versa, for overlap extension. Several mutagenic primer pairs, for example, like

15 the B-C primer pair, were used to produce mutations at different locations that were "stitched" together, i.e. A-B, C-D, E-F, etc. to generate an intact, full length coding region. To produce more of the final product containing the mutation(s), a PCR using the A primer and the most 3' reverse primer was performed. Primers A and D were used to amplify the entire coding sequence of Pox4 and Pox5 and to incorporate the restriction enzyme sites (RE1 and RE2) for cloning into an *E. coli*

20 expression vector. The primers used for the site-directed mutagenesis for Pox5 (*Candida* strain ATCC20336) are listed in TABLE 10 to TABLE 19. The primers used for the site-directed mutagenesis for Pox4 (*Candida* strain ATCC20336) are listed in TABLE 20 to TABLE 26.

TABLE 10 - Pox5 (Candida strain ATCC20336)

Mutant Name	Position	Native Amino Acid(s)	Introduced Mutation	Primer	Primer Sequence (5'-3')	Restriction Sites
A				PRIMER A	GTTCACTGCCATATGCCTACCGAACTTCAAAAAGAAAG AGAACTC	Nde1
A	81, 82	DQ	AA	PRIMER B	GATCGACAATCTCTGGCCTGAGCAGCTGGTACTCGT GCTCAAAG	
A	81, 82	DQ	AA	PRIMER C	CTTTGAGCAGAGTACCCAGCTGCTCAGGCCAGAGAT TGTCGATC	
A				PRIMER D	CTTCGAGATGCGGCCGCTTAAGTGGACAAGATTTCAGC AGCTTCTTCG	Not1
B				PRIMER A	GTTCACTGCCATATGCCTACCGAACTTCAAAAAGAAAG AGAACTC	Nde1
B	86, 88	RLS	ALA	PRIMER B	GTGGGTCAAAGACACCGAGGATAGCCAAAGCCTGGGC CTGTTGGTCTGGGTAC	
B	86, 88	RLS	ALA	PRIMER C	GTACCCAGACCAACAGGCCAGGCTTTGGCTATCCTCG GTGTCTTTGACCCAC	
B				PRIMER D	CTTCGAGATGCGGCCGCTTAAGTGGACAAGATTTCAGC AGCTTCTTCG	Not1

TABLE 11 - Pox5 (Candida strain ATCC20336)

Mutant Name	Position	Native Amino Acid(s)	Introduced Mutation	Primer	Primer Sequence (5'-3')	Restriction Sites
C				PRIMER A	GTTCACTGCCATATGCCTACCGAACTTCAAAAAGAAAG AGAACTC	Nde1
C	93, 94	FD	AA	PRIMER B	GATTCTGGTGAAGACTTGTGGAGCAGCGACACCGAGG ATCGACAATC	
C	93, 94	FD	AA	PRIMER C	GATTGTCGATCCTCGGTGTCGCTGCTCCACAAGTCTTC ACCAGAATC	
C				PRIMER D	CTTCGAGATGCGGCCGCTTAAGTGGACAAGATTTCAGC AGCTTCTTCG	Not1
D				PRIMER A	GTTCACTGCCATATGCCTACCGAACTTCAAAAAGAAAG AGAACTC	Nde1
D	291, 292	DS	AA	PRIMER B	GAATCTACTGGTCATTCTGTAAGCAGCCATCATCATGG TGACTCTACC	
D	291, 292	DS	AA	PRIMER C	GGTAGAGTCACCATGATGATGGCTGCTTACAGAATGA CCAGTAGATTC	
D				PRIMER D	CTTCGAGATGCGGCCGCTTAAGTGGACAAGATTTCAGC AGCTTCTTCG	Not1

TABLE 12 - Pox5 (Candida strain ATCC20336)

Mutant Name	Position	Native Amino Acid(s)	Introduced Mutation	Primer	Primer Sequence (5'-3')	Restriction Sites
E				PRIMER A	GTTCAGTGCATATGCCTACCGAACTTCAAAAAGAAAG AGAACTC	Nde1
E	95, 96	PQ	AA	PRIMER B	CACCGATTCTGGTGAAGACAGCAGCGTCAAAGACACC GAGGATCG	
E	95, 96	PQ	AA	PRIMER C	CGATCCTCGGTGTCTTTGACGCTGCTGTCTTCACCAGA ATCGGTG	
E				PRIMER D	CTTCGAGATGCGGCCGCTTAACTGGACAAGATTTTCAGC AGCTTCTTCG	Not1
F				PRIMER A	GTTCAGTGCATATGCCTACCGAACTTCAAAAAGAAAG AGAACTC	Nde1
F	294, 295	RM	AA	PRIMER B	GGTGATGAATCTACTGGTCGCGCGTAGGAGTCCATC ATCATG	
F	294, 295	RM	AA	PRIMER C	CATGATGATGGACTCCTACGCCGCGACCAGTAGATTCA TCACC	
F				PRIMER D	CTTCGAGATGCGGCCGCTTAACTGGACAAGATTTTCAGC AGCTTCTTCG	Not1

TABLE 13 - Pox5 (Candida strain ATCC20336)

Mutant Name	Position	Native Amino Acid(s)	Introduced Mutation	Primer	Primer Sequence (5'-3')	Restriction Sites
G				PRIMER A	GTTCAGTGCATATGCCTACCGAACTTCAAAAAGAAAG AGAACTC	Nde1
G	287	T	A	PRIMER B	GAGTCCATCATCATGGCGACTCTACCACCAATC	
G	287	T	A	PRIMER C	GATTGGTGGTAGAGTCGCCATGATGATGGACTC	
G				PRIMER D	CTTCGAGATGCGGCCGCTTAACTGGACAAGATTTTCAGC AGCTTCTTCG	Not1
H				PRIMER A	GTTCAGTGCATATGCCTACCGAACTTCAAAAAGAAAG AGAACTC	Nde1
H	290, 291	MD	AA	PRIMER B	CTGGTCATTCTGTAGGAGGCTGCCATCATGGTGACTCT ACC	
H	290, 291	MD	AA	PRIMER C	GGTAGAGTCACCATGATGGCAGCCTCTACAGAATGAC CAG	
H				PRIMER D	CTTCGAGATGCGGCCGCTTAACTGGACAAGATTTTCAGC AGCTTCTTCG	Not1

TABLE 14 - Pox5 (Candida strain ATCC20336)

Mutant Name	Position	Native Amino Acid(s)	Introduced Mutation	Primer	Primer Sequence (5'-3')	Restriction Sites
I				PRIMER A	GTTCACTGCCATATGCCTACCGAACTTCAAAAAGAAAG AGAACTC	Nde1
I	284/436	GE	EG	PRIMER B	CATCATCATGGTGACTCTTTACCAATCAAAGCCGAG	
I	284/436	GE	EG	PRIMER C	CTCGGCTTTGATTGGTGAAAGAGTCACCATGATGATG	
I	284/436	GE	EG	PRIMER D	GTTGTTGTACCTCCCCAGGTACATTGG	
I	284/436	GE	EG	PRIMER E	CCAATGTACCTGGGGAGGTGACAACAAC	
I				PRIMER F	CTTCGAGATGCGGCCGCTTAACTGGACAAGATTTTCAG CAGCTTCTTCG	Not1
J				PRIMER A	GTTCACTGCCATATGCCTACCGAACTTCAAAAAGAAAG AGAACTC	Nde1
J	291	D	G	PRIMER B	GGTCATTCTGTAGGAGCCCATCATCATGGTGAC	
J	291	D	G	PRIMER C	GTCACCATGATGATGGGCTCCTACAGAATGACC	
J				PRIMER D	CTTCGAGATGCGGCCGCTTAACTGGACAAGATTTTCAG CAGCTTCTTCG	Not1

5 TABLE 15 - Pox5 (Candida strain ATCC20336)

Mutant Name	Position	Native Amino Acid(s)	Introduced Mutation	Primer	Primer Sequence (5'-3')	Restriction Sites
K				PRIMER A	GTTCACTGCCATATGCCTACCGAACTTCAAAAAGAAAG AGAACTC	Nde1
K	292	S	A	PRIMER B	CTGGTCATTCTGTAGGCGTCCATCATCATGGTG	
K	292	S	A	PRIMER C	CACCATGATGATGGACGCCTACAGAATGACCAG	
K				PRIMER D	CTTCGAGATGCGGCCGCTTAACTGGACAAGATTTTCAGC AGCTTCTTCG	Not1
L				PRIMER A	GTTCACTGCCATATGCCTACCGAACTTCAAAAAGAAAG AGAACTC	Nde1
L	93	F	A	PRIMER B	GTGAAGACTTGTGGGTCTGCGACACCGAGGATCGAC	
L	93	F	A	PRIMER C	GTCGATCCTCGGTGTCGAGACCCACAAGTCTTCAC	
L				PRIMER D	CTTCGAGATGCGGCCGCTTAACTGGACAAGATTTTCAGC AGCTTCTTCG	Not1

TABLE 16 - Pox5 (Candida strain ATCC20336)

Mutant Name	Position	Native Amino Acid(s)	Introduced Mutation	Primer	Primer Sequence (5'-3')	Restriction Sites
M				PRIMER A	GTTCAGTCCATATGCCTACCGAACTTCAAAAAGAAAG AGAACTC	Nde1
M	94	D	G	PRIMER B	GGTGAAGACTTGTGGGCCAAAGACACCGAGGATC	
M	94	D	G	PRIMER C	GATCCTCGGTGCTTTGGCCACAAGTCTTCACC	
M				PRIMER D	CTTCGAGATGCGGCCGCTTAACTGGACAAGATTTCAGC AGCTTCTTCG	Not1
CT1*				PRIMER A	CACACAAGGGGAATTGTGAGCGGATAAC	XbaI
CT1*	102	G	D	PRIMER B	AACCCAAGTTGACGTCGATTCTGG	
CT1*	102	G	D	PRIMER C	CCAGAATCGACGTCAACTTGGGTT	
CT1*				PRIMER D	CACACAACTGGATCCAACCGTTATCG	BamHI

* Mutant CT1 produces a smaller fragment that is used to replace the sequence in between the XbaI and BamHI sites of wildtype Pox5 cloned into pET26b.

5

TABLE 17 - Pox5 (Candida strain ATCC20336)

Mutant Name	Position	Native Amino Acid(s)	Introduced Mutation	Primer	Primer Sequence (5'-3')	Restriction Sites
N				PRIMER A	GTTCAGTCCATATGCCTACCGAACTTCAAAAAGAAAG AGAACTC	Nde1
N	86	R	A	PRIMER B	CAAAGACACCGAGGATCGACAAAGCCTGGGCTGTG GTCTGGGTAC	
N	86	R	A	PRIMER C	GTACCCAGACCAACAGGCCAGGCTTTGTCGATCCTCG GTGTCTTTG	
N				PRIMER D	CTTCGAGATGCGGCCGCTTAACTGGACAAGATTTCAGC AGCTTCTTCG	Not1
O				PRIMER A	GTTCAGTCCATATGCCTACCGAACTTCAAAAAGAAAG AGAACTC	Nde1
O	88	S	A	PRIMER B	CAAAGACACCGAGGATCGCCAATCTCTGGGCTGTG	
O	88	S	A	PRIMER C	CAACAGGCCAGAGATTGGCGATCCTCGGTGTCTTTG	
O				PRIMER D	CTTCGAGATGCGGCCGCTTAACTGGACAAGATTTCAGC AGCTTCTTCG	Not1

TABLE 18 - Pox5 (Candida strain ATCC20336)

Mutant Name	Position	Native Amino Acid(s)	Introduced Mutation	Primer	Primer Sequence (5'-3')	Restriction Sites
P				PRIMER A	GTTCAGTGCATATGCCTACCGAACTTCAAAAAGAAAG AGAACTC	Nde1
P	98	F	G	PRIMER B	GTTGACACCGATTCTGGTCCGACTTGTGGGTCAAAGAC	
P	98	F	G	PRIMER C	GTCTTTGACCCACAAGTCGGAACCAGAATCGGTGTCAAC	
P				PRIMER D	CTTCGAGATGCGGCCGCTTAAGTGGACAAGATTTTCAGC AGCTTCTTCG	Not1
Q				PRIMER A	GTTCAGTGCATATGCCTACCGAACTTCAAAAAGAAAG AGAACTC	Nde1
Q	83, 85	QAQ	AAA	PRIMER B	CAAAGACACCGAGGATCGACAATCTAGCCGCAGCTTG GTCTGGGTACTCGTGCTCAAAG	
Q	83, 85	QAQ	AAA	PRIMER C	CTTTGAGCAGCAGTACCCAGACCAAGCTGCCGCTAGAT TGTCGATCCTCGGTGTCTTTG	
Q				PRIMER D	CTTCGAGATGCGGCCGCTTAAGTGGACAAGATTTTCAGC AGCTTCTTCG	Not1

TABLE 19 - Pox5 (Candida strain ATCC20336)

Mutant Name	Position	Native Amino Acid(s)	Introduced Mutation	Primer	Primer Sequence (5'-3')	Restriction Sites
CT2				PRIMER A	GTTCAGTGCATATGCCTACCGAACTTCAAAAAGAAAG AGAACTC	Nde1
CT2	453-463	DLLKE PEQK GL	QVISIEDAG KTVRGSTAF LNQLKDYT GSNSSKV	PRIMER B	CGGCATCTTCAATGCTGATAACTTGCTCTAACCATTG GCTTGGCA	
CT2	453-463	DLLKE PEQK GL	QVISIEDAG KTVRGSTAF LNQLKDYT GSNSSKV	PRIMER C	TGCCAAGCCAATGGTTAGAGCAAGTTATCAGCATTGA AGATGCC	
CT2	453-463	DLLKE PEQK GL	QVISIEDAG KTVRGSTAF LNQLKDYT GSNSSKV	PRIMER D	TCGGCAACGCTGGAGAGACAACCTTGAGCTGTTG GAACCAAGTGT	
CT2	453-463	DLLKE PEQK GL	QVISIEDAG KTVRGSTAF LNQLKDYT GSNSSKV	PRIMER E	ACACTGGTTCCAACAGCTCCAAGGTTGTTCTCTCCAG CGTTGCCGA	
CT2				PRIMER F	CTTCGAGATGCGGCCGCTTAAGTGGACAAGATTTTCAG CAGCTTCTTCG	Not1

TABLE 20 - Pox4 (Candida strain ATCC20336)

Mutant Name	Position	Native Amino Acid(s)	Introduced Mutation	Primer	Primer Sequence (5'-3')	Restriction Sites
A				PRIMER A	GTTCACTGCCATATGACTTTTACAAAGAAAAACGTT AGTGTA	Nde1
A	98, 99, 100	FNK	AAA	PRIMER B	CAAAGATACCAATCAAGGACAATCTAGCAGCAGCA GTGTCGATGGATTCTTGTCTCTG	
A	98, 99, 100	FNK	AAA	PRIMER C	CAGAGAACAAGAATCCATCGACACTGCTGCTGCTA GATTGTCCTTGATTGGTATCTTTG	
A				PRIMER D	CTTCGAGATGCGGCCGCTTATTACTTGGACAAGAT AGCAGCGGTTTCATCAGA	Not1
B				PRIMER A	GTTCACTGCCATATGACTTTTACAAAGAAAAACGTT AGTGTA	Nde1
B	102, 103	LS	AA	PRIMER B	GTGGGTCAAAGATACCAATCAAAGCAGCTCTCTTG TTGAAAGTTCGATG	
B	102, 103	LS	AA	PRIMER C	CATCGACACTTTCAACAAGAGAGCTGCTTTGATTG GTATCTTGACCCAC	
B				PRIMER D	CTTCGAGATGCGGCCGCTTATTACTTGGACAAGAT AGCAGCGGTTTCATCAGA	Not1

5 TABLE 21 - Pox4 (Candida strain ATCC20336)

Mutant Name	Position	Native Amino Acid(s)	Introduced Mutation	Primer	Primer Sequence (5'-3')	Restriction Sites
C				PRIMER A	GTTCACTGCCATATGACTTTTACAAAGAAAAACGTT AGTGTA	Nde1
C	96	D	A	PRIMER B	GGACAATCTCTTGTGAAAGTAGCGATGGATTCTT GTTCTCTG	
C	96	D	A	PRIMER C	CAGAGAACAAGAATCCATCGCTACTTTCAACAAGA GATTGTCC	
C				PRIMER D	CTTCGAGATGCGGCCGCTTATTACTTGGACAAGAT AGCAGCGGTTTCATCAGA	Not1
D				PRIMER A	GTTCACTGCCATATGACTTTTACAAAGAAAAACGTT AGTGTA	Nde1
D	90	R	A	PRIMER B	GTGTCGATGGATTCTTGTTCAGCGTATCTGGCGAT TCTGTTG	
D	90	R	A	PRIMER C	CAACAGAATCGCCAGATACGCTGAACAAGAATCCA TCGACAC	
D				PRIMER D	CTTCGAGATGCGGCCGCTTATTACTTGGACAAGAT AGCAGCGGTTTCATCAGA	Not1

TABLE 22 - Pox4 (Candida strain ATCC20336)

Mutant Name	Position	Native Amino Acid(s)	Introduced Mutation	Primer	Primer Sequence (5'-3')	Restriction Sites
E				PRIMER A	GTTCACTGCCATATGACTTTTACAAAGAAAAACGTT AGTGTA	Nde1
E	88	R	A	PRIMER B	GATGGATTCTTGTTCTCTGTAAGCGGCGATTCTGTT GATCTTGAC	
E	88	R	A	PRIMER C	GTCAAGATCAACAGAATCGCCGCTTACAGAGAACA AGAATCCATC	
E				PRIMER D	CTTCGAGATGCGGCCGCTTATTACTTGGACAAGAT AGCAGCGGTTTCATCAGA	Not1
F				PRIMER A	GTTCACTGCCATATGACTTTTACAAAGAAAAACGTT AGTGTA	Nde1
F	302	M	A	PRIMER B	GAGTCCAAAACCATCGCGACTCTACCACCCAAC	
F	302	M	A	PRIMER C	GTTGGGTGGTAGAGTCGCGATGGTTTTGGACTC	
F				PRIMER D	CTTCGAGATGCGGCCGCTTATTACTTGGACAAGAT AGCAGCGGTTTCATCAGA	Not1

TABLE 23 - Pox4 (Candida strain ATCC20336)

Mutant Name	Position	Native Amino Acid(s)	Introduced Mutation	Primer	Primer Sequence (5'-3')	Restriction Sites
G				PRIMER A	GTTCACTGCCATATGACTTTTACAAAGAAAAACGTT AGTGTA	Nde1
G	309, 310	RM	A	PRIMER B	GTGGACATTCTAGCCAACGCGGCGTAGGAGTCCA AAACCATC	
G	309, 310	RM	A	PRIMER C	GATGGTTTTGGACTCCTACGCCGCTTGGCTAGAA TGCCAC	
G				PRIMER D	CTTCGAGATGCGGCCGCTTATTACTTGGACAAGAT AGCAGCGGTTTCATCAGA	Not1
H				PRIMER A	GTTCACTGCCATATGACTTTTACAAAGAAAAACGTT AGTGTA	Nde1
H	98	F	A	PRIMER B	CAAGGACAATCTCTTGTTGGCAGTGTCGATGGATT CTTG	
H	98	F	A	PRIMER C	CAAGAATCCATCGACACTGCCAACAAGAGATTGTC CTTG	
H				PRIMER D	CTTCGAGATGCGGCCGCTTATTACTTGGACAAGAT AGCAGCGGTTTCATCAGA	Not1

TABLE 24 - Pox4 (Candida strain ATCC20336)

Mutant Name	Position	Native Amino Acid(s)	Introduced Mutation	Primer	Primer Sequence (5'-3')	Restriction Sites
I				PRIMER A	GTTCACTGCCATATGACTTTTACAAAGAAAAACGTT AGTGTA	Nde1
I	99	N	A	PRIMER B	CAATCAAGGACAATCTCTTCGCGAAAGTGTGATG GATTC	
I	99	N	A	PRIMER C	GAATCCATCGACACTTTCGCGAAGAGATTGTCCTT GATTG	
I				PRIMER D	CTTCGAGATGCGGCCGCTTATTACTTGGACAAGAT AGCAGCGGTTTCATCAGA	Not1
J				PRIMER A	GTTCACTGCCATATGACTTTTACAAAGAAAAACGTT AGTGTA	Nde1
J	100	K	A	PRIMER B	CAATCAAGGACAATCTCGCGTTGAAAGTGTGATG	
J	100	K	A	PRIMER C	CATCGACACTTTC AACGCGAGATTGTCCTTGATTG	
J				PRIMER D	CTTCGAGATGCGGCCGCTTATTACTTGGACAAGAT AGCAGCGGTTTCATCAGA	Not1

TABLE 25 - Pox4 (Candida strain ATCC20336)

Mutant Name	Position	Native Amino Acid(s)	Introduced Mutation	Primer	Primer Sequence (5'-3')	Restriction Sites
K				PRIMER A	GTTCACTGCCATATGACTTTTACAAAGAAAAACGTT AGTGTA	Nde1
K	102	L	A	PRIMER B	CAAAGATACCAATCAAGGAGGCTCTCTTGTTGAAA GTGTCG	
K	102	L	A	PRIMER C	CGACACTTCAACAAGAGAGCCTCCTTGATTGGTA TCTTTG	
K				PRIMER D	CTTCGAGATGCGGCCGCTTATTACTTGGACAAGAT AGCAGCGGTTTCATCAGA	Not1
L				PRIMER A	GTTCACTGCCATATGACTTTTACAAAGAAAAACGTT AGTGTA	Nde1
L	103	S	A	PRIMER B	GTCAAAGATACCAATCAAGGCAATCTCTTGTTGA AAGTG	
L	103	S	A	PRIMER C	CACTTCAACAAGAGATTGGCCTTGATTGGTATCTT TGAC	
L				PRIMER D	CTTCGAGATGCGGCCGCTTATTACTTGGACAAGAT AGCAGCGGTTTCATCAGA	Not1

TABLE 26 - Pox4 (Candida strain ATCC20336)

Mutant Name	Position	Native Amino Acid(s)	Introduced Mutation	Primer	Primer Sequence (5'-3')	Restriction Sites
CT3				PRIMER A	GTTCAGTCCATATGACTTTTACAAAGAAAAACGTT AGTGTA	Nde1
CT3	473-505	QVISIE DAGKT VRGST AFLNQL KDYTGS NSSKV	DLLKEPEQK GL	PRIMER B	CAATCCCTTTTGTCTGGCTCCTCAACAAGTCCTT GACAATTGGCTTACCAA	
CT3	473-505	QVISIE DAGKT VRGST AFLNQL KDYTGS NSSKV	DLLKEPEQK GL	PRIMER C	GACTTGTGAAGGAGCCAGAACAAAAGGGATTGG TTTTGAACACTGTTGCTGA	
CT3				PRIMER D	CTTCGAGATGCGGCCGCTTATTACTTGGACAAGAT AGCAGCGGTTTCATCAGA	Not1

In vitro acyl CoA oxidase Assay

5

E.coli lysates were tested for acyl CoA oxidase activity as described in Example 41.

In vitro Activity Assay for Pox4 Mutants

- 10 TABLE 27 shows the acyl CoA oxidase activity profile associated with Pox4 mutants and TABLE 28 shows the acyl CoA oxidase activity profile associated with Pox5 mutants. The carbon length of the substrates tested is indicated above the data as C6 (6 carbons), C8 (8 carbons), C10 (10 carbons), C12 (twelve carbons), C14 (fourteen carbons), C16 (sixteen carbons) and C18:1 (eighteen carbons). In TABLE 27 and 28, unshaded blocks indicated the sample was not tested.
- 15 Dark shading indicates that no activity was detected. Light shading indicates that minimal activity (i.e. poor activity) was detected at less than or equal to 0.1 umol/min/ug (umol substrate/minute/ug total protein). Medium shading indicates that good activity was detected at > 0.1 umol/min/ug.

- 20 The Pox4 Mutant C, although displaying good activity across all substrates tested, demonstrated reduced overall activity for all substrates (TABLE 27). Pox4 Mutant D showed a similar result. Activity on C12 and C18:1 substrates was abolished in Pox4 Mutants B, A, E and G (TABLE 27) and CT3 (not shown).

TABLE 27 - *In vitro* Activity Assay for Pox4 Mutants

Mutant Name	SDM Candida strain ATCC20336 Pox4 Mutants (all in pET26b)	C6	C8	C10	C12	C14	C16	C18:1
WT	Pox4 WT							
B	Pox4 with L102A, S103A							
A	Pox4 with F98A, N99A, K100A							
C	Pox4 with D96A							
C*	Pox4 with D96A, L692S							
D	Pox4 with R90A							
E	Pox4 with R88A							
F	Pox4 with M302A							
G	Pox4 with R309A, M310A							

* Indicates a secondary mutation (I692S) that occurred during PCR for cloning Candida strain ATCC20336 POX5

5 *In vitro* Activity Assay for Pox5 Mutants - Results

Acyl CoA oxidase activity was abolished in Pox5 Mutants B, C, F and M at least on substrates C6, C12 and C18:1 (TABLE 28). Mutants CT1 and CT2 were also inactive (not shown). Mutants A, E and I showed no change when compared to the activity of the wild type protein. However, Pox5 Mutants D, H, G, and J displayed altered substrate specificity when compared to wild type Pox5. Pox5 mutants D, H and J demonstrated reduced activity on C6 and/or C8 substrates. Pox5 mutant G displayed increased activity on C18:1 substrates.

TABLE 28 - *In vitro* Activity Assay for Pox5 Mutants

Mutant Name	SDM Candida strain ATCC20336 Pox5 Mutants (all in pET26b)	C6	C8	C10	C12	C14	C16	C18:1
WT	Pox5 WT							
A	Pox5 with D81A, Q82A							
B	Pox5 with R86A, S88A							
C	Pox5 with F93A, D94A							
D	Pox5 with D291A, S292A							
G	Pox5 with T287A							
H	Pox5 with M290A, D291A							
E	Pox5 with PQ95AA							
F	Pox5 with R294A, M295A							
J	Pox5 with D291G							
K	Pox5 with S292A							
L	Pox5 with F93A							
M	Pox5 with D94G							

Acyl-CoA dehydrogenase-based mutagenesis

Acyl-CoA oxidases and acyl-CoA dehydrogenases (ACAD) both utilize similar but distinct mechanisms to catalyze dehydrogenation of an acyl-CoA substrate to produce a 2-*trans*-enoyl-CoA, the first step in β -oxidation (Arent, S., Pye, V.E., Henriksen, A. (2008). Structure and function of plant acyl CoA oxidases. *Plant Phys. Biochem.* 46:292-301). There are acyl-CoA dehydrogenases of different classes and they are grouped according to their substrate specificities: very long, long, medium and short chain (VLCAD, LCAD, MCAD, SCAD, respectively) (Kim, J.J., Miura, R. (2004). Acyl-CoA dehydrogenases and acyl CoA oxidases. Structural basis for mechanistic similarities and differences. *Eur. J. Biochem.*, 271(3):483-93.). The crystal structures of several of these enzymes have been solved and these data show structural differences that very likely contribute to their respective differences in substrate specificity. The crystal structure of VLCAD (PDB: 3B96) has revealed regions and amino acid residues of the protein that make it structurally, and more than likely functionally, different from MCAD (PDB: 3MDE) (McAndrew, R.P., Wang, Y., Mohsen, A.W., He, M., Vockley, J., Kim, J.J. (2008). Structural basis for substrate fatty acyl chain specificity: crystal structure of human very-long-chain acyl-CoA dehydrogenase. *J. Biol. Chem.* 283(14):9435-43). In some cases, a more significant difference is the location of the catalytic residue. In MCAD, the catalytic glutamate is located at position 376 on the loop connecting helix J and K while in LCAD, the catalytic glutamate is at position 255 on the adjacent helix G (Nandy, A., Kieweg, V., Kräutle, F.G., Vock, P., Küchler, B., Bross, P., Kim, J.J., Rasched, I., Ghisla, S. (1996). *Medium-long-chain chimeric human Acyl-CoA dehydrogenase: medium-chain enzyme with the active center base arrangement of long-chain Acyl-CoA dehydrogenase*. *Biochemistry*, 35(38):12402-11; Lee, H.J., Wang, M., Paschke, R., Nandy, A., Ghisla, S., Kim, J.J. (1996). *Crystal structures of the wild type and the Glu376Gly/Thr255Glu mutant of human medium-chain acyl-CoA dehydrogenase: influence of the location of the catalytic base on substrate specificity*. *Biochemistry*, 35(38):12412-20).

The crystal structure of VLCAD (PDB: 3B96) has also revealed regions and amino acid residues of the protein that make it structurally, and more than likely functionally, different from MCAD (PDB: 3MDE) (McAndrew et al., 2008). VLCAD is larger than other acyl-CoA dehydrogenase proteins and forms a dimer, like a typical acyl CoA oxidase. Its substrate binding cavity is larger compared to other acyl-CoA dehydrogenase proteins and resembles an acyl CoA oxidase substrate binding pocket. The larger and more spacious pocket is necessary for accommodating the longer fatty acyl-CoA substrates that it acts upon. However, the crystal structures of rat Acoil and Arabidopsis

thaliana ACX1 (PDB ID: 1W07) also reveal large substrate binding pockets and this feature does not necessarily explain the substrate specificities of each enzyme (Arent et al., 2008). Structural differences between MCAD and VLCAD offer some insight. At the base of the MCAD substrate binding pocket, there are two polar/charged residues (Q95 and E99) that are different from the
 5 analogous residues in VLCAD (G175 and G178). The increased hydrophobicity of the base of the VLCAD substrate binding pocket may be a factor, in addition to pocket size and depth, which contributes to substrate specificity. The corresponding residues in *Candida* strain ATCC20336 Pox5 are F98 and G102. Mutant P (F98G) is the mutation that should more closely reproduce the base of the VLCAD substrate binding pocket.

10

A double mutation in MCAD (e.g., E376G, T255E) can change its substrate specificity profile. This double mutation produced somewhat of a chimeric enzyme (MLCAD)(Nandy, et.al.1996). MCAD has a broad substrate profile (C4 – C18) with peak activity at C6 and C8. LCAD has a similarly broad profile with peak activity at C10 and C12. MLCAD has a more defined substrate profile (C10
 15 – C18) compared to MCAD or LCAD with peak activity at C12. However, the overall enzymatic activity of the MLCAD was also reduced (V_{max} of MLCAD for C12 substrate is approximately 25% of V_{max} of LCAD for C12 substrate).

Based in part on the results of the above studies, Pox5 was mutated as described in TABLE 14
 20 (Mutant I) to shift its substrate profile to preferentially act on longer chain substrates.

VLCAD mutagenesis

VLCAD has a substrate profile that is appropriate for production of longer chain diacids, such as
 25 sebacic or dodecanedioic acid. Activity of the enzyme ranges from acyl substrates that are 10 carbons to 22 carbons long and peak activity is on a C16 substrate. However, the enzymatic mechanism of VLCAD differs from that of a typical acyl CoA oxidase with respect to the final electron acceptor; in VLCAD, the enzyme is reoxidized by electron transfer ferroprotein (ETF) and AOXs are reoxidized by oxygen to produce hydrogen peroxide (Arent et al., 2008; Kim and Miura,
 30 2004). To accommodate the difference in mechanisms, the substrate binding pocket of an acyl CoA oxidase, such as *A. thaliana* ACX1, is more spacious than that of VLCAD to allow oxygen into the pocket to act as the final electron acceptor and reoxidize the flavine adenine dinucleotide, or FAD, cofactor required for dehydrogenation of the acyl-CoA substrate. ETF performs this function in a typical acad and reoxidation of FAD by oxygen is inhibited while substrate is bound (Kumar,

- N.R., Srivastava, D.K. (1995). Facile and restricted pathways for the dissociation of octenoyl-CoA from the medium-chain fatty acyl-CoA dehydrogenase (MCAD)-FADH₂-octenoyl-CoA charge-transfer complex: energetics and mechanism of suppression of the enzyme's oxidase activity. *Biochemistry*, 34(29): 9434-43). This is reflected in the shape of the substrate binding pocket with respect to FAD. In an acyl CoA oxidase, FAD is more solvent exposed, but in MCAD, the entire flavin ring is embedded in the protein and is only accessible to solvent when substrate is not present (Kim and Miura, 2004). In order for an acad to have oxidase activity, the substrate binding pocket must become more solvent accessible and permit oxidation of the reduced FAD cofactor by oxygen. Mutagenesis studies of MCAD have identified a residue that can achieve this result.
- 5 Tyrosine 375 in MCAD, when changed to a lysine, confers significantly increased (~200-fold increase relative to wild type MCAD) acyl CoA oxidase activity (Zeng, J., Liu, Y., Wu, L., Li, D. (2007). Mutation of Tyr375 to Lys375 allows medium-chain acyl-CoA dehydrogenase to acquire acyl CoA oxidase activity. *Biochim. Biophys. Acta*, 1774(12): 1628-34).
- 10
- 15 Molecular modeling suggests that the mutation increases solvent accessibility near the FAD moiety in the active site. In order for VLCAD to function as an acyl CoA oxidase with the appropriate substrate specificity profile, an analogous mutation in VLCAD is made. Tyrosine 375 in MCAD corresponds to phenylalanine 461 in human and rat VLCAD. A F461K mutation in VLCAD is tested to see if it will now have acyl CoA oxidase activity.
- 20

Example 43: Nucleotide and Amino Acid Sequences Used for Manipulations described herein

SEQ ID NO:	Description	Sequence
SEQ ID NO:1	Thioesterase activity Cuphea lanceolata Amino acid (A.A. Seq)	MVAAAATSAFFPVPAPGTSPKPGKSGNWPSSLSPTEFKPKSIPNAGFQVKA NASHPKANGSAVNLSKSGSLNTQEDTSSPPPPRAFLNQLPDWSMLLTAIT TVFVAAEKQWTMLDRKSKRPDMLVDSVGLKSIVRDGLVSRQSFILRSYEI GADRTASIELTMNHLQETSINHCKSLGLLNDGFGRTPGMCKNDLIWVLTK MQIMVNRYPTWGDTEINTWFSQSGKIGMASDWLISDCNTGEILIRATSV WAMMNQKTRRFSRLPYEVRQELTPHFVDSPHVIEDNDQKLHKFDVKTGD SIRKGLTPRWNDLDVNQHVSNVKYIGWILESMPIEVLETQELCSTVEYRR ECGMDSVLESVTAVDPSENGGRSQYKHLRLLEDGTDIVKSRTIEWRPKNA GTNGAISTSTAKTSNGNSAS
SEQ ID NO:2	FAO-13 (fatty alcohol oxidase activity) C. Tropicalis Nucleotide (Nuc. Seq)	atggctccattttgcccgaccaggtcgactacaaacacgtcgacacccttatgtattatgtgacgggatc atccacgaaacaccgtcgaccaaataaagacgtattgctcctgacttccctgctgacaagtacgaa gagtacgtcaggacattcaccaaacccctccgaaacccaggggtcagggaacccgtctacaacacag tcaacgcaaacaccacggacgcaatccaccagttcattatcttgaccaatgtttggcatccagggctctg gctccagcttgaccaactcgttgacgcctatcaaggacatgagcttgggaagaccgtgaaaaattgttg cctcgtggcgcgactcccaatcgctgccaaaaggaagtgttcaggtgtttctacgcttacctgtgca

SEQ ID NO:	Description	Sequence
		cggtcacgagattggccaatgagttgcattgaaagccattcattatccaggaagagaagaccgtgaaa aggcttatgaaaccaggagattgaccttttaagaccagtttttgaaaaaccgaagtttaacggcgct gagttgacttgccagatatgatgatcattatggatctggcgccgtgctggtgttggtgcccacacttg gccaacgatggcctcaagagtttggttttgaaaaagggcaatacttagcaactcagagttgaactttgat gacaaggacggcggtcaagaattataccaaagtgagggtactttgactacagcaaaaccaacagttgtg ttctgctggtccacttttggtggcggtaccactgtcaattggcagcctgtcttaagacgccattcaagggtg cgtaaggaatggatgatgagttgggttgactttgctgctgatgaagcatacgataaagcgcaggattat gtttggcagcaaatgggagcttctaccgaaggcatcaccactctttggctaacgagattattatgaagg tggttaagaaattagggtacaaggccaagggtattagaccaaaacagcggtggtcatcctcagcacagat gcggtttctgtatttggtgctgaagcacgggtatcaagcagggttctgtaataactggttagagacgcagc tgcccacgggtcccagttcatgcaacagggttagagttttgcaaaacttaacaagaagggtacgcttacg gtatctgtgtgaggatgttgtaaccggcgccaagttcaccattactggccccaaaagttgtgtgctgc cggtgcttgaacatccatctgtgttggtcaactccggctcaagaacaagaacatcggtgaagaacttaa ctttgacccagttctgtctgtgttggtgattttggcaagacgttcaagcagaccactccacaactccatc atgactgccccttggcagaagccgctgatttagacggcaagggccatggatgcagaattgaaacatctt gaacgctccattcatccaggcttcatcttaccatggagaggttagtaacgaggctagacgagactgttgc gttaacaacaatggtggcgatgttgccttagctgtgacaccaccagtggttcggttctgtctcatccaac caaacctgaagcttgggtgctgagtagcagcgtgaacaagttgacagaaaactcgatcttcaggcattgt tggctactgctgactgtgtatataccaagggtccaagagaatccttagtccacaggcatgggtgccaat tgaatccgacaagccaaggataagagatcaatcaaggacgaggactatgtcgaatggagagccaa gggtgccaagattccttgcacacctaagggtcaccctatgggttcggcacaatgattcttgcggtatgt cagggaagggtcctaaatccggtgctgttgacaccgatggtagattgttgaatgttgcgaatgttatgtg gatgcaagcttttgccaactgaagcgggtgccaaccctatgggtcaccaccatgactctgcccagacatgt tgcttaggttggcagactccttgaagaccaagccaagttgtag
SEQ ID NO:3	FAO-13 (fatty alcohol oxidase activity) C. Tropicalis A.A. Seq	MAPFLPDQVDYKHVDTLMLLCDGIIHETTVDDQIKDVIAPDFPADKYEEYVR TFTKPSETPGFRET VYNTVNANTTD AIHQFIILTNVLASRLAPAL TNSLTPI KDMSLEDREKLLASWRDSPIAAKRKLFRVLSTLTLVTFTRLANELHLKAIHY PGREDREKAYETQEIDPFKYQFLEKPKFYGAEL YLPDIDVIIIGSGAGAGVV AHTLANDGFKSLVLEKGKYSNSELNFDDKDG VQELYQSGGTLTTVNQQ LFVLGSGTFGGGTTVNWSACLKTPFKVRKEWYDEFGVDFAADEAYDKAQ DYVWQQMGASTEGITHSLANEIIIIEGGKLG YKAKVLDQNSGGHPQHRC GFCYL GCKHGIKQGSVNNWFRDAAAHGSQFMQQRVLQILNKKGIAYGIL CEDVVTGAKFTITGPKKFVVAAGALNTPSVLVNSGFKNKNIGKNLTLHPVS VFGDFGKDQVQADHFHNSIMTALCSEAADLDGKGHGCRIETILNAPFIQAS FLPWRGSNEARRDLLRYNNMVAMLLLSRDTTSGSVSAHPTKPEALVVEY DVNKFDRNSILQALLVTADLLYIQGAKRILSPQAWVPFESDKPKDKRSIKD EDYVEWRAKVAKIPFDYGS PYGSAHQMSCRMMSGKGP KYGAVD TDGR LFEC SNVYVADASLLPTASGANPMVTTMTLARHVALGLADSLKTKAKL
SEQ ID NO:4	FAO-17(fatty alcohol oxidase activity) C. Tropicalis Nuc. Seq	atggctccattttgcccggaccaggtgcactacaacacgtgcacacccttatgtattatgtgacgggac atccacgaaccacggtggacgaatcaaagacgtcattgcccctgacttccccgccgacaaatag aggagtacgtcaggacattcaccaaaacctccgaaccccagggttcagggaacacgctcacaacac cgtcaacgcaaacaccatggatgcaatccaccagttcattatctgaccaatgttttggatcaagggtctt ggcaccagcttgcacaaactcgttgactcctatcaaggacatgagcttgaagaccgtgaaaagttgta gcctcgtggcgtagctcccctattgctgctaaaaggaagttgtcagggtgttctacgcttaccttggtcac gttcacgagattggccaatgagttgactttgaaagccattcattatccaggaagagaagaccgtgaaaa ggcttatgaaaccaggagattgaccttttaagtaaccagtttttgaaaaaacgaagtttacggcgctg agttgtacttgccagatatgatgatcattatggatctggtgcgggtgctggtgtgtgtgcccacacttg ccaacgatggctcaagagttgtgttttgaaaaagggcaatacttagcaactccgagttgaactttgatg acaaggacggcggtcaagaattataccaaagtgagggtactttgactacagtaaccaacagttgtgtgt ctgtgtgttccactttgtggcggtaccactgtcaattggcagcctgtcttaagacgccattcaagggtgc gaaggaaatggatgatgagttgtgtgactttgctgctgatgaagcatacgataaagcgcaggattatg

SEQ ID NO:	Description	Sequence
		ttggcagcaaatgggagcttctaccgaaggcatcaccactcttggctaacgagattattattgaaggt ggtaagaaaattaggtta caaggccaaggtattagacaaaacagcgggtggtcaacctcagcacagatg cggtttctgttatttgggtgtaagcacgggtatcaagcagggctctgttaataactggttagagacgcagct gccacgggttctcagttcatgcaacaggttagagtttggcaaatccttaacaagaagggcacgccttatggt atctgtgtgaggatgttgaaccgggtgccaaggtcaccattactggcccaaaaaagttgtgtgcccgcg gcgccttaaacactccatctgtgttggtcaactccggattcaagaacaagaacalcgggtaagaacttaac ttgtcaccagtttctgctgtgttggtagtttggcaagaacgttcaagcagaccatccacaactccatca tgactgcccttgttcagaagcgcgtgatttagacggccaagggccatggatgcagaattgaaccatcttg aacgtcccatcatccaggcttcattcttccatggagaggtagtaacagggtctagacgagactgtgtgcg tacaacaacatgggtggcgatgtgtctccttagctgtgacaccaccagtggttccggttctgtctatccaacc aaacctgaagcttgggtgtcgagtagacgtgaacaagttgacagaaaactcgatcttgaggcattgtt ggtcactgctgactgtgtatataccaaggtgccaagagaaactcttagtccacaggcatgggtgccaatttt gaatccgacaagccaaggataagagatcaatcaaggacgaggactatgtcgaatggagagccaa gggtccagaagttcttcgacacatcgggtcaccttatggttcacatcaaatgtctctgtccgtagt caggtaagggttccaaatacgggtgctgttgacaccgatggtagattgttgaatgttgaatgtttatgttgc gatgcaagctctttgccactgcaagcgggtgccaacccctatggtcaccacatgactcttgaagacatgt tgcgtaggttggcagactcctgaagaccaaggccaagttgatg
SEQ ID NO:5	FAO-17(fatty alcohol oxidase activity) C. Tropicalis A.A. Seq	MAPFLPDQVDYKHVDTLMLLCDGIIHETTVDIHKDVIAPDFPADKYEYVVRT FTKPSETPGFRET VYNTVNANTMDA I HQFIILTNVLGSRVLAPAL TNSLTPI KDMSLEDREKLLASWRDSPAIAKRKLFRLVSTLTLVTFTRLANELHLKAIHY PGHREDEKAYETQEIDPFKYQFLEKPKFYGAELYLPDIDVVIIGSGAGAGV AATRLANDGFKSLVLEKGYFSNSELNFDDKDGVDQELYSGGTLTTVNQQ LFLVLGSTFGGGTTVNWSACLKTPFKVRKEWYDEFGVDFAADEAYDKAQ DYVWQQMGASTEGITHSLANEIIIIEGGKKLGYKAKVLDQNSGGHPQHRC GFCYL GCKHGIKQGSVNNWFRDAAA HGSQFMQQVRVLQILNKKGIAYGIL CEDVVTGAKFTITGPKKFVVAAGALNTPSVLVNSGFKNKNIGKNLTLPVPS VVF GDFGKDVQADHFHNSIMTALCSEAADLDGKGHGCRITILNAPFIQAS FLPWRGSNEARRDLLRYNNMVAMLLLSRDTTSGSVSAHPTKPEALVVEY DVNKFDRNSILQALLVTADLLYIQGAKRILSPQAWVPFIESDKPKDKRSIKD EDYVEWRKAKIPFDITYGSPYGS AHQMSSCRM SGKGP KYGAVDTDGR LFEC SNVYVADASLLPTASGANPMVTTMTLARHVALGLADSLKTKAKI
SEQ ID NO:6	FAO-20(fatty alcohol oxidase activity) C. Tropicalis Nuc. Seq	atggctccattttgcccgaaccaggtcgactacaaaacgctcgacacccattatgtattatgtgacgggac atccacgaaacacgctcgaccaaatcaaagacgtattgtctcctgacttccctgctgacaagtacgaa gagtacgtcaggacattcaccaaacccctcgaaaccccagggttcagggaacacgtctacaacacag taacgcgaacacacacgagcgcaatccaccagttcattatcttgaccaatgttttggcatccagggtcttg gtccagctttgacaaactcgttgacgcctatcaaggacatgagcttggaagaacgtgaaaaattgttgg cctcgttggcgcgaccccccaatcgctgccaaggaaattgttcagggtgtttccacgcttacccttggtta ctttcagcgagattggccaatgagttgcaittgaagccattcactacaggaaggaagacgctgaaa aggcttatgaaacccaggagattgaccccttcaagtaaccggttatggaaaagccaagtttgacggcg ctgagttgtacttgcagataitgatgttatcattattggatctgtgtccggtgtcgtgtgttggcccacactt ggccaacgatgtgctcaagagtttgggttggaaaagggcaaaatacttagcaactccgagttgaactttg atgacaaggacggcggttcaagaattataccaaaaggtaggtactttgactacagtcacaaacagttgtt tgctctgtcgtgtccacttttgggtggcggttaccactgtcaattggtcagcctgtcttaagacgccaatcaagg gcgtgaagaaatgtatgatgattgtgtgtgtgacttgcgtgatgaagcatcagataaagcgaggtatt atgtttggcagcaaatgggagcttctaccgaaggcatcaccactcttggctaacgagattattattgaa gggtgtaagaaattaggttacaaggccaaggtattagacaaaacagcgggtgtcalcctcagcacag atgcggtttctgttatttgggtgtaagcacggtatcaagcaggttctgttaataactggttagagacgca gtgcccacgggttccagttcatgcaacaggttagagtttggcaatacttaacaagaaggggcatgcctta cggtaacttctgttgaggatgtgttaacgggcgccaaggtcaccattacggccccaaaaagttgtgtgtgt gcgggtcttgaacacactccatctgtgtgtgtaactccgggtcctcaagacaagaacatcggtgaagaact aaccttgcaaccagtttctgtcgtgttgggtatgttggcaagacgttcaagcagaccacttccacaactcc

SEQ ID NO:	Description	Sequence
		atcatgactgcccttgttcagaagccgctgatttagacggcaaggccatggatgcagaattgaaacca tcttgaacgctccattcatccaggctcattctaccatggagaggttagtaacgaggctagacgagactgt igcgttacaacaacatggtggcgatgtgtccttagctgtacaccaccagtggtccgttctgtcatcc aaccaaacctgaagcttgggtgtagtacgacgtgaacaagttgacagaactcgatcttgaggc attgttggtcactgctgactgtgttatccaagggtccaagagaatccttagtccaaggcatgggtgcc aatgttgaatcgacaagccaaaggataagagatcaatcaaggacgaggactatgtcgaatggaga gccaagggtgccaagattccttgcacacacacgtcactatgttgggcacatcaaatgtctcttgcc gtatgtcaggtaagggtcctaaatacgggtgctgtgacaccgatggttagattgttgaatgtcgaatgttat gttgccgatgcaagcttcttgcacactgcaagcgggtgccaacctatggtcaccacatgactcttgccag acatgttgcgtaggttggcagactccttgaagaccaagccaagttgtag
SEQ ID NO:7	FAO-20(fatty alcohol oxidase activity) C. Tropicalis A.A. Seq	MAPFLPDQVDYKHVDTLMLLCDGIIHETTVDDQIKDVIAPDFPADKYEEYVR TFTKPSETPGFRET VYNTVNANTTD AIHQFIILTNVLASRV LAPAL TNSLTPI KDMSLEDREKLLASWRDSPIAAKRKL FRLVSTLTLVTFTRLANELHLKAIHY PGREDREKAYETQEIDPFKYQFMEKPKFDGAELYLPDIDVIIIGSGAGAGV VAHTLANDGFKSLVLEKGYFSNSELNFDDKDG VQELYQSGGTLTTVNQ QLFVLGASTFGGGTTVNWSACLKTPFKVRKEWYDEF GVDFAADEAYDKA QDYVWQQMGASTE GITHSLANEIIIIEGGKKLGYKAKVLDQNSGGHPQHR CGFCYL GCKHGIKQGSVNNWFRDAAAHGSQFMQQVRVLQILNKKGIAYG ILCEDVVTGAKFTITGPKKFVVAAGALNTPSVLVNSGFKNKNIGKNLTLHPV SVVFGDFGKDVQADHFHNSIMTALCSEAADLDGKGHGCR IETILNAPFIQA SFLPWRGSNEARRDLLRYNNMVAMLLLSRD TTS GSVSAHPTKPEALVVE YDVNKFDRNSILQALLVTADLLYIQGAKRILSPQAWVPIFESDKPKDKRSIK DEDYVEWRKAKIPFD TYGSPY GSAHQMS CRMMSGKGP KYGAVDTDG RLFEC SNVYVADASLLPTASGANPMVTTMTLARHVALGLADSLKTKAKL
SEQ ID NO:8	FAO-2a(fatty alcohol oxidase activity) C. Tropicalis Nuc. Seq	atgaataccttctgacagctgctgaatacaaacacgtcgacacccttgttattgtgtgacgggac atccacgaaacccacgtcgatcagatcaaggacgccattgctccgaactccctgaggacccagtaaga ggagatctcaagacctcaccagccatctgagacccctgggtcagagaagccgtctacgacacgat caacgccaccccaaccgatgccgtgacatgtgtattgtcttgaccacccgattggactccagaatcttg gccccacgttgaccaactcgttgacgcctatcaaggatatgacctgaaggagcgtgaacaattgttg cctctggcgtagtccccgattgcggcaagagaagattgtcagattgattcctcgctaccttgacgac gtttacgagattggccagcgaattgcattgaaagccatccactaccctggcagagacttgcgtgaaaa ggcgatgaaacccagggtggtgaccttaccgttcttattggagaaaccaaagttgacggcgcc gaattgtacttgccagatatcgacgtcatcattggatcaggcgccgggtgctggtgctatgcccacac tctcgccaacgacgggtcaagaccttggtttgaaaagggaagatttcagcaactccgagttgaac ttaatgacgctgatggcgtaagaggtgtaccaaggtaaggtgcttggccaccacacatcagcag atgtttattcttgcgggtccacttggcggtggtaccatgtcaactggtctgcttcttaaaacaccatt aaagtgcgtaaggagtgtgacgacgagttgttctgaattgtgctcgatgaagcctacgacaagcg caggattatgttgaaacaaatgggtgcttcaacagatggaaatcaatcactccttggccaaagcgaagttg ggtgaaggaggaagaagttgggtctacaagagcaaggaaattgagcagaacaacgggtggccacc tgaccacccatgtgttctgttacttgggtgtaagtagcgtattaaacagggtctgtgaataactggtta gagacgcagctgccacgggtccaagtcatgcaacaagtcagagttgtgcaaatcctcaacaagaat ggcgctcgcttattgttctgtgtaggagtgctgaacccggagtcaggttcaactatttagtgccccaaaa gtttgttctgtgctgttcttgaacacgccaaactgtgtgaccaactccggattcaagaacaagcacattg gtaagaacttgacgttgacccagtttccaccgtgttgggtgacttggcagagacgtgcaagccgacat ttccacaaatcattatgacttgccttgaaggtgtgacttggacggcaagggccacggatgcaga atcgaacacatcttgacgtccattcatccaagcttcttgtgcatggagagggaagtgacgaggtcag aagagactgttgcgttacacaacatgttgccatgttgcattacacgctgataccaccagtggtcagt tctgctgacccaaagaagcccgacgcttgaattgtgactatgagattaacaagttgacaagaatgcca tcttgaagcttcttgatcacttccgacatgtgtacattgaagggtccaagagaatcctcagtcacagc catgggtgccaatcttgaagtcgaacaagccaaggagcaagaacgatcaaggacaaggactatgt tgagtgagagcgaaggctgctaagataccttgcacacctacgggttcgcatatgggtcgcacatcaa

SEQ ID NO:	Description	Sequence
		atgtccaccgtctgtatgtccggaagggtcctaaatacggctgtgtgatactatggttagattgttgaat gttcgaatgtctatgttctgtatgctagtgtttgctactgccagcgggtgccaaccaatgatataccacat gaccttgcctagacagattgcgttaggtttggctgactcctgaagaccaaccaagttgtag
SEQ ID NO:9	FAO-2a(fatty alcohol oxidase activity) C. Tropicalis A.A. Seq	MNTFLPDVLEYKHVDTLILLCDGIIHETTVDQIKDAIAPDFPEDQYEEYLKT FTKPSETPGFREAVYDTINATPTDAVHMCIVLTALDSRILAPTLTNSLTPIK DMTLKEREQLLASWRDSPIAAKRRLFRLISSLTTLTFTRLASELHLKAIHYP GRDLREKAYETQVVDPPFRYSFMEKPKFDGAELYLPDIDVIIIIGSGAGAGVM AHTLANDGFKTLVLEKGYFSNSELNFNDADGVKELYQGKGALATTNQQ MFILAGSTLGGGTTVNWSACLKTPFKVRKEWYDEFGLEFAADEAYDKAQ DYVWKQMGASTDGIHSLANEVVEGGKKLGYKSKEIEQNNGGHPDHPC GFCYLGCKYGIKQGSVNNWFRDAAAHGSKFMQQVRVQILNKNGVAYGI LCEDVETGVRFTISGPKKFVVSAGSLNTPVLTNSGFKNKHIGKNLTLPV STVFGDFGRDVQADHFHKSIMTSLCYEVADLDGKGHGCRIETILNAPFIQA SLLPWRRGSDEVRRDLLRYNNMVAMLLITRDTTSGSVSDPKPDALIVDY EINKFDKNAILQAFILTSMDLYIEGAKRILSPQPWVPIFESNPKPEQRTIKDK DYVEWRAKAAKIPFDITYGSAYGSAHQMSTCRMSGKGPYGAVDTDGRL FECNVYVADASVLTASGANPMISTMTFARQIALGLADSLKTKPKL
SEQ ID NO:10	FAO-2b(fatty alcohol oxidase activity) C. Tropicalis Nuc. Seq	atgaatacctcttccagacgtgctcgaatacaaacacgtcgataccctttgtattatgtgacgggatca tccacgaaaccacagtcgaccagatcagggaacgcatctcccgacttccctgaagaccagtaacga ggagtatctcaagaccttcccaagccatctgagacccctgggtcagagaagcgtctacgacacgat caacagcaccaccaaccgaggtctgtcacatgtgtattgtattgaccaccgcatggactcgagaatcttg gccccacgttgaccaactcgttgacgcctatcaaggatatgacctgaaagagcgtgaacaattgttg ctgcctggcgtgattcccgatcgcggccaagagaagattgttcagattgatitcctcacttacctgaaga cctttacgagattggccagcagcttgacgtgagagccatccactccctggcagagacttgcgtgaaa aggcatatgaaccacaggtgggtgaccttccaggtacctgttatggaaaaacaaagttgacggcac cgagttgacttgcagatatcgaagtcateatcattggatccggtgcgggtgctggtgcatggcccaca cttagccaaacgacgggtacaagacctgtgtttggaaaaaggaaagtatttcagcaactccgagttgaa cttaatatgctccgatggtatgaagagttgtaccaaggtaaatgtgcgttgaccaccacgaaccagca gatgttattcttgcgggttccacttggcgggtgttaccatgttaactgctgctgttcttaaaacaccattt aaagtgcgaaggagtgtagcagcagattgttctgaattgtgcgacgaagcctaagcaaaagca caagactatgttggaaacaaatggcgctctaccgaaggaaatcactccttggcgaacgcggtgtg gggtgaaggaggaagaagttgggttacaagagcaaggaaatcgagcagaacaatgggtggccatcct gaccacccctgtggttctgttacttgggtctgtaagtacggtattaagcagggttctgtgaataactggttag agacgcagctgcccacgggtccaagttcatgcaacaagtcagagttgtgcaaatcctccacaataaag gcgtgcctatggcatctgtgtgaggtgtcgagaccggagtcgaattcactatcagtgccccaaaaa gtttgtgttctgcagggtcttgaacacgccaacgggttgaccaactccggttcaagaacaacacat cggttaagaacttgacgttgacccagtttgcagcgtgttgggtgacttggcagagacgtgcaagccgac catttccacaaatctattatgacttgcctgttgcgaagtcgctgaacttggacggcaagggccaaggatg cagaatcgagaccatcttgaacgctccattcatccaagcttcttggccatggagaggaagcgacgag gtcagaagagacttgtgcgttacaacaacatgggtggccatgtgtctatcaccggtgacaccaacgtg gttcagtttctgctgacccaaagaagcccagcgttgaattgtcgactatgacatcaacaagttgacaag aatgccatcttgcagcttcttgatcacctccgacatgtgtacatcgaagggtgccaagagaatcctcagt ccacaggcatgggtgccaacttcttgatcgaacaagccaaggagcaagaacaatcaaggacaa ggactatgcaaatggagagccaaggctgccaagataccttgcacacctaagcttctgctatgggtgc gcacatcaaatgtccactgtctatgtccggaaagggtcctaatacggcgccgttgataccgatgga gattgttgaatgtcgaatgtctatgtctgatgctagtgtttgcctactgccagcgtgccaacccaatg atctccaccatgacgttgcctagacagattgcgttaggttggctgactcttgaagaccaaccaagttg ag
SEQ ID NO:11	FAO-2b(fatty alcohol oxidase activity)	MNTFLPDVLEYKHVDTLILLCDGIIHETTVDQIRDAIAPDFPEDQYEEYLKT FTKPSETPGFREAVYDTINSTPTEAVHMCIVLTALDSRILAPTLTNSLTPIK DMTLKEREQLLAWRDSPIAAKRRLFRLISSLTTLTFTRLASDLHLRAIHYP

SEQ ID NO:	Description	Sequence
	C. Tropicalis A.A. Seq	GRDLREKAYETQVVDPFYRYSFMEKPKFDGTELYLPDIDVIIIGSGAGAGVM AHTLANDGYKTLVLEKGKYFSNSELNFNDADGMKELYQGKCALTTTNQQ MFILAGSTLGGGTTVNWWSACKLTPFKVRKEWYDEFGLEFAADEAYDKAQ DYVWKQMGASTEGITHSLANAVVVEGGKKLGYKSKEIQNNGGHPDHP GFCYLGCKYGIKQGSVNNWFRDAAAHGSKFMQQVRVVLHNNKGVAYGI LCEDVETGVKFTISGPKKFVVSAGSLNTPVLTNSGFKNKHIGKNLTLHPV STVFGDFGRDVQADHFHKSIMTSLCYEVADLDGKGHGCRIETILNAPFIQA SLLPWRGSDEVRDRLRYNNMVAMLLITRDTTSGSVSADPKKPDALIVDY DINKFDKNAILQAFLITSDMLYIEGAKRILSPQAWVPIFESNKPKEQRTIKDK DYVEWRAKAAKIPFDYGSAYGSAHQMST
SEQ ID NO:12	FAO-18(fatty alcohol oxidase activity) C. Tropicalis Nuc. Seq	atggctccattttgcccagaccaggtcgactacaacacgctcgacacccttatgtattatgtacgggac atccacgaaaccacggtgacgaaatcaaagacgtcattgcccctgactcccgcgacaaatac aggagtagtcaggacattcacaaacccctcgaaacccaggggtcagggaacggtctacaacac cgtcaacgcaaacacccatggatgcaatccaccagttcattatctgaccaatgtttgggatcaagggtct ggcaccagcttgcaccaactcgttgactcctcaaggacatgagcttggaaagaccgtgaaaagtgtta gcctcgtggcgtgactcccctattgctgctaaagggaagtgttcaggttggtttctacgctaccttggtcac gttcacgagattggccaatgagttgattgaaagccattcattaccaggaagagaagaccgtgaaaa ggcttatgaaacccaggagattgacccctttaagtaacagttttgaaaaacggaagtgttacggcgctg agttgactctgcagatattgatgtacatttggatctggggccgggtgctggttgcgtggcccacacttg accaacgacggctcaagagttggtttgaaaagggcagatacttagcaactccgagttgaacttga tgacaaggacgggttcaagaattataccaaagtggaggtacttggaccacccgtaaccagcagttggtt gttctgctggttccacttttgggtggttaccactgtcaattggtcggcctgtcttaaacgcactcaagggtg cgtaaggatggtatgatgagttggcgttgacttgcctgctgaagcctacgacaaagcacaggatt atgtttggcagcaaatgggagctctaccgaaggcatcaccactcttggctaacgagattatttgaa ggtggcaagaaattaggttacaaggccaaggtattagacaaaaacagcgggtgctatcctcatcacag atggcgtttctgtatttgggtgtaagcacggtatcaagcagggctctgtaataactggttagagacgca gctgcccacgggtctcagttcatgcaacaggtagagtttgcataatcctaacaagaagggtcagcttat ggatcttgtgtgaggtgtgttaacgggtgccaagttcaccattactggcccaaaaaagtgttgttgcg cgggogccttaaacactccatctgtgttgtaactcgggttcaagaacaagaacatcgglaagaact aactttgcatccagttctgtcgtgttgggtgatttggcaagacgttcaagcagatcactccacaactcca tcatgactgctcttgcagaagccgctgatttagacggcaggggtcaggtgagaaatgaaacatc tgaacgctccatcatccagggttcatcttccatgagaggttagtaacgagggtcagagactgtgtg cgttacaacaacatggtggccatgttactcttagtctgataaccaccagtggttccgttctccatcca ctaaacctgaagcattagtgtcgagtacgacgtgaacaagttgacagaaactccatctgcaggcatt gttggtcactgctgactgtgtgacattcaagggtgccaagagaalccctagtccccaaacatgggtgccaat tttgaatccgacaagccaaggataagagatcaatcaaggacgaggactatgtogaatggagagcc aagggtgccaagattcctttgacacctaaggctcgcttatggttcggcgcatcaaatgtcttctgtctg gtcaggtgaagggtcctaaatacgggtgctgtgataccgatggtagattgttgaatgttcgaatgttattgtg ctgacgctagtctttgccaaactgtagcgggtgctaactcctatggtcaccaccatgactctgcaagacatg tgcttaggtttggcagactccttgaagaccaaggccaagtttag
SEQ ID NO:13	FAO-1(fatty alcohol oxidase activity) C. Tropicalis A.A. Seq	MAPFLPDQVDYKHVDLMLLQDGIHETTVEIKDVIAPDFPADKYEEYVRT FTKPSETPGFRETIVNTVNANTMDAIHQFIILTNVLGSRVLAPALTNLSLTP KDMSLEDREKLLASWRDSPIAAKRKLRLVSTLTLVTFTRLANELHLKAIHY PGREDREKAYETQEIDPFKYQFLEKPKFYGAELYLPDIDVIIIGSGAGAGV AHTLTNDGFKSLVLEKGGRYFSNSELNFDDKDGVLQELYQSGGTLTTVNQQ LFVLGASTFGGGTTVNWWSACKLTPFKVRKEWYDEFGVDFAADEAYDKAQ DYVWQMGASTEGITHSLANEIIIIEGGKKLGYKAKVLDQNSGGHPHHRGG FCYLGCKHGIKQGSVNNWFRDAAAHGSQFMQQVRVLQILNKKGIAYGILC EDVVTGAKFTITGPKKFVVAAGALNTPSVLVNSGFKNKNIGKNLTLHPVSV VFGDFGKDVQADHFHNSIMTALCSEAADLDGKGHGCRIETILNAPFIQASF LPWRGSNEARRDRLRYNNMVAMLLLSRDTTSGSVSSHPTKPEALVVEYD

SEQ ID NO:	Description	Sequence
		VNKFDRNSILQALLVTADLLYIQGAKRILSPQPWWPIFESDKPKDKRSIKDE DYVEWRKAVAKIPFDITYGSPYGSAHQMSSCRMMSGKGPYGAVDTDGRL FEC SNVYVADASLLPTASGANPMVTTMTLARHVALGLADSLKTKAKL
SEQ ID NO:14	cytochrome P450 A12 (CYP52A12) Nuc. Seq	atggccacacagaatcatcgattctgtactccgtacttgaccaatggtacactgtgattactgcagc agtattagctcttctatctccacaaacatcaagaactacgtcaaggcaagaaatgaaatgtgcgac caccatactgaagatgccggtctcactggtattctgtcttgatgcgcatcaaggccaagaaacgac ggttagattgctaacttgcgcatgaagtttgcagagtaaccaaacacaccttactgtctgtgcg gtgcttgaagattgtcactgttgacccagaaaacatcaaggctgtcttgccaccaattcactgact tctcttggtaccagacagcccacttgccttgggtgacggatcttcaacttggaaggagaag gttgaagcactccagactatgttgagaccacagttgctagagaccagattggacacgttaaagcctt ggaaccacacatccaaatcatggctaagcagatcaagttgaaccagggaagacattcgatccaag aattgttcttagattaccgtcgacacccgctactgagttctgttggtaacccgttcaactctgtacgatga aaaattgggcatcccaactccaaacgaaatccagggaagagaaaacttgcgctgttcaacgtttc ccaacactacttggccaccagaagtactccagacttttacttttgaccaaccctaaagaaatcagaga ctgaacgccaaggctccacttggccaagtacttgcacaaggccgtgaacttactctgaagaac tcgaagagaaatccaagtcgggtacgtttctgtacgaattggtaagcaaacagagatccaaaggt cttgaagatcaattgtgaacattatggtgacggaagagacaccactgcgggttggctctgtctgtt tgaattggctagacaccagagatgtgttccaagttgagagaagaaatcgaagttaacttgggtgtgt gaagactcccggtgaagaaatcacttgaagcctgaagagatgtgaatacttgaaggctatcctta acgaaaccttgcgtatgtacccatctgttctgtcaactttagaacggccaccagagacaccacttgcga agaggtgtgtgtcgaacggtaaccgacccaatctacattcctaaggctccactgtgtcactgtgtctac aagaccacccgttgaagaatactacggttaaggacgtaacgacttcagaccagaaagatgggttga accatctactaagaagtgggtgtgtgtatgtccattcaacgggtgtccaagagtctgttgggtcaaca attcgcttactgaagcttctatgtgatcactagattggccagatgttgaactgtctcatctgatccag gtctgaataccctccaccaaagtgattcacttgaccatgagtcacaacgatggtgtcttgaagatgt aa
SEQ ID NO:15	cytochrome P450 A13 (CYP52A13) Nuc. Seq	atgactgtacacgaattatcgccacatacttcacaaatggtacgtgatagaccactgcgttattgctt atagagtccctgactacttctatggcagatacttgatgataagcttggtgtaaacattttccagaaaca gacagacggctgttccgattcaaaagctcggctgaattgttgaagaagaagagcgacggtaacctcat agacttcaactccagcgtatccagatctcgatcgtcccgatacccaacttcaacttccggctctttcc atcaacctgtcaataccttgagccggagaaacatcaaggccatcttggccactcagttcaacgatttctc cttgggtaccagacactcgcacttgccttgggtgatggtatcttactgttgatggcgccgctgga agcacagcagatctatgttgagaccacagtttgcagagaacagatttccacgtcaagttgtggagcc acacgtcaggtgttctcaaacagtcagaaaggcacagggcaagacttttgacatccaggaattgttt tcagattgacggctgactccgccaacggagtttgggtgaatccgttgagtccttgagagatgaatctatc ggcatgtccatcaatgcgcttgacttgaacggcaaggctggcttgcgtatgtttaaactatcgagaatta ttggcttcgagagcggttatgcaacaattgtactgggtgtgaacgggaaaaagtttaaggagtgcaac gctaaagtgcacaagttgtgactactacgtcaacaaggcttgaactgacgctgaacaattggaaa agcaggatggttatgttttgaagaaatggtcaagcaaacagagacaagcaagttgagagacc aattgtgaacatcatggtgtggttagagacaccacggccggttggctgttggcttgaattggcca gaaaccagaagttaaccaaaagttgagagaagaaatgaggacaagtttgactcggtagaatgc tagtgtgaagacatttcttgaagtcgtgaagtctgtgaatactgaaggctgttctcaacgaacacttga gattgtacccatccgtgccaagaatttcagagttgccaccaagaacactacccccaagaggtgggtg taaggacgggtgtctcctgttttggtagaaagggtcagacggctatttgcggtgtcagcagccac agaaaccagctgttgcgtaaggacgctctgagtttagaccagagagatggtttagccagagaca aagaagcttggctgggctctcctccattcaacgggtgtccaagaatctgttgggacagcagttgccttg acagaagcttctgtatgtcactgacgggtgtccaggagtttgacactgtctatggaccagacaccga atatccacctaaagaaatgtcgcaattgaccatgtcgttctgcagcggtccaatattgagatgtattag
SEQ ID NO:16	cytochrome P450 A14 (CYP52A14) Nuc. Seq	atgactgcacaggatattatcgccacatacatcaccaaatggtacgtgatagaccactgcgttattgctt atagggtcctcgactactttacggcagatacttgatgataagcttggtgtaaacgggttttccagaac aaacagacgggtatttccgattcaaaagctccactgaattgttaaaaaagaagagtacggtaacctcat

SEQ ID NO:	Description	Sequence
		agacttcactctcgcgcatccaagcgcctaactcgtccagatatcccaactttacattcccaatctttcc atcaaccttatcagcacccttgagccggagaaacatcaaggctatcttggccaccaggtcaacgattctc ctgggaccagacactcgcactttgctcctttgtggcgatggatctttaccttggacggtgccggctgg aagcacagcagatctatgttgagaccacagtttgcagagaaacagatttccacgctcaagttgttggagc cacacatgcagggtgtcttcaagcacgtcagaaaggcacagggcaagaatttgacatccaagaattgtt ttccagattgacggtcgactccgccactgagttttgttggatccgttgagtccttgagagatgaatcta tgggatgtccatcaatgcacttgaacttgcggaaggctggcttgcgtgattgttttaactctgcagaaac tatttggtctcagagcgggtatgcaacaaitgtactgggtgttgaaagggaataaaggagtgacaa cgtctaaagtgcacaagttgtcgtactattacgtcagcaaggcttggactgacacctgaacaattggaa aagcaggatggttaigtcttctgtacgagttgttcaagcaaacagagacaggcaagtttgagagac cagttgttgaacatcatggttgcggtagagacaccaccgctgggttgtgtcgttgtttcttgaattggcc agaaaccagagagtgaccaacaagttgagagaagaatcaggagacaagtttggcttgggtgagaatg ctctgttgaagaacttcttctgagtcgttgaagtcagtgaataactgaaggctgttctcaacgaacttgg agattgtaccactcgtgcacagaaittcagagttgccaccaaaaactactcctccaaggggaggt ggtaaggacgggttatctcctgttttggtcagaaagggtcaaacctgtatgtacggtgtctacgtctgccac agaaaccagctgtctacggttaaggacgccccttgaattagaccagagaggtgggtttagccagagac aaagaagcttggctggcccttcttccattcaacggtgtgtccaagaatttgcgttgggacagcagttgcct gacagaagcttgcgtatgtcactgtcagattgtcctcaagagtttggacactgtctatggacccaacaccg aatatccacctaggaaaaatgtcgcatttgaccatgtcccttttcgacgggtgccacattgagatgattag
SEQ ID NO:17	cytochrome P450 A15 (CYP52A15) Nuc. Seq	atgtcgtcttctccatcgttggccaagaggttctcgtaccactagtccttaccatcgagtactttcttgacaac tacaccagatggttactacttataccttgggtctcttctgttgaaacttataagtttgcctccacacaagggtact tggaaacgcaggttccacgccaagccactcggtaactttgcagggaacctcactgttggatcgcctactccg ttgtcttgcactactgaagtcgaaggtagcgtgaagtgttgcgttgggctcctgtgaaacacagactc atcgtcagagacccaagtagcaagacaactgggtcaggattgttggcctccattgattgaaccatg gaccagagaaacatcaaggctgttttggctactcagttcaatgatttcttcttgggaaccagacacgatttct tgtactccttgttgggtgacggtattttcccttggacgggtgtcgtgtggaacatagtagaactatgtttag accacagttgttagagaacaggttctcactcagtcagttgttggagccacacgttcagggtgttctcaagca cgttagaaagcacgcggtcaaacgttcgacatccaagaattgttctcagggttgacctgcactccgc accgagttcttgttgggtgagtcgtcgtgaatccttgagggaacgaalcttggattgacccaaccacaag gatttgcagtggaagagatttccgtgcagcttcaactlctgcagacttaccaggcctacagattttgtt gcaacaaatgtactggtacttgaatggctcggaaatcagaagtcgattgtcgtcgtgcacaagttgtcgt accactaigtgcaaaaggcttggagtgacccgacgatgactgcagaaacaagacggctatgtgttctt gtacgagttggctaagcaaacagagacccaaggcttgagagaccagttatgaacattttgttggcc ggtagagacacgaccgcgggttgttcttcttctcagagttgtcaagaacctcagggtgttggcta agttgagagaggttggaacacagatttgactcggtagaagaagctcgtgttgaagagatctcgttgg agtccctgaagctctgtgagacttgaaggctgtcatcaatgaacacctgagattgtaccatcgggtccaca caactttagagttgtaccagaaacactaccctcccaagaggtgtgtgtgaagatggatactcgccaatt gtcgtcaagaagggtcaagttgtcatgtactgttattgtaccacagagacccaagtatctacggtgc cgacgctgacgttctcagaccagaaagatggttgaaccagaaactagaagaattgggctgggcatcgt tccattcaalgtgttgccaagaatctgttgggtcaacagtttgcctgacccaagctcatcactcgtgtc agattgtccaggagttgcacactgtctatggacccagacccaagaatccacaaaattgcagaac acctgacctgtcgtcttctgaggtgtcgtatgtagaattgactaa
SEQ ID NO:18	cytochrome P450 A16 (CYP52A16) Nuc. Seq	atgtcgtcttctccatcgttggctcaggaggttctcgtaccactagtccttaccatcgagtactttcttgacaact acaccagatggttactacttcatcccttgggtcttcttctgttgaaactatcagcttgcctccacacaagggtact tggaaacgcaggttccacgccaagccgctcggtaactcgtgttggatcctacgttggatcgcctactccgt tgaactctgactactaaagtcgaaggtagcgtcatgaagtttgccttgagcttctggaacaacaagtac attgtcaagaaccacaagtacaagcaactggccttagaattgtcggcctccattgattgaaccatag accagagaaacatcaaaagctgtgttggctactcagttcaacgatttctccttgggaaactagacacgatttct tgtactccttgttggcgatggtattttaccttggacgggtgtcgtgtggaacacagtagaactatgtttag accagattgtctagagaacaggttccacagtttggcaagttgttgggaacacacagctcaggtgttctcaagc acgttagaaaaacaccgcggtcagactttgacatccaagaattgttctcagattgacctgcactccgc

SEQ ID NO:	Description	Sequence
		accgagttctgttggtagtctgctgaatccttgagagacgactctgttggttgaccccaaccaag gatttcgaaggcagaggagattcgcctgacgcttcaactactcgagacttacaggcctacagattttg tgcacaacaaatgtactggatttgaatggcgcggaatcagaaagtcgattgccatcggtcacaagttgct gaccactatgtgcaaaaggcttggagttgaccgacgatgacttgagaaaacagacggctatgtgtct tgtacgagttggctaaagcaaaactagagacccaaaggcttgagagaccagttgtgaaacttttggtgcc ggtagagacacgacccgcgggttgttgcgttctacgagttgtcgagaaacccgaagtggttgc aagttgagagaggaggttgaaaaacagatttgactcggcgaaggctcgtgttgaaagatctcttt agtcctctgaagtcctgtgacttgaaggctgcatcaatgaagccttgagattgtaccactgttccac acaacttcagagttgccaccagaaaactacccctccaaggaggcggttgtaaagacggatgctcgcc aattgtgtcaagaagggtcaagttgtcatgtacactgtcattggtaccacagagacccaagtatctacg gtgccgacgcccagctctcagaccagaaagatggttcgagccagaaactagaaagtgggctgggc atatgtccattcaatggtgtccaagaatctgtttgggtcagcagttgcttgactgaagctcatacgtca ctgtcagattgtctcaagagtttgaaactgtccctggatccaacgctgagtaccaccacaaaattgca gaacaccttgaccttgctactcttgaagtgctgacgttagaattgtctaa
SEQ ID NO:19	cytochrome P450 A17 (CYP52A17) Nuc. Seq	atgattgaacaaactcctagaatattggtatgtcgttgtccagttgtgtacatcatcaaaacactcctgcat acacaaagactcgcgtcttgatgaaaaagttgggtgctgtcctcagtcacaaacaaagttgtacgacaacg ctttcgttatcgtcaatggatggaaggctcaccagttcaagaaggaggcagggtcacaagagtacaac gattacaagttgaccactccaagaaaccaagcgctgggcacclacgtcagtatcttttcggcaccagga tcgtcgtgaccaaagatccagagaatatcaaagctatttggcaacccagtttgggtgattttctttgggcaa gaggcacactctttttaagcctttgttaggtgatggatcttcacattggacggtggaaggtggaagcaca gcagagccatgttgagaccacagttgcccagagaaacaaagttgctcatgtgacgctcgttggaaacacactt ccagttgttgaagaagcatattcttaagcacaagggtgaatacttcatatccaggaattgtcttttagattta ccgttgactcggccacgggacttctatttggtagtcctgcctcttgaaggagcaatctatttggtatcaa ccaagacgatalagattttgctggtagaaaggacittgctgagtcgttcaacaaaggccagggaatacttg gctattagaaccttgggtgcagacgttctactggttggtaacaacaaggagtttagagactgtaccaagct gggtgcacaagttcaccaactactatgttcagaaaggcttggatgtcagccagaagagcttgaaaagca aagtgggtatgtgtcttgaacgagctgtcaagcagacaagagaccccaatgtgttgctgacacagctctt gaacatctgttggccggaagagacacactgctgggtgtgtgtcgttctgtcgttcttgagttggccagacac ccagagatctgggccaagttgagagaggaaattgaacaacaglttggcttggagaagactcgtgttg aagagattacctttgagagctgaagagatgtgacttgaaagcgttcccttaatgaaacottgctgtattt acccaagtgtcccaagaaactcagaatcgccaccaagaaacagcattgccaaggggctgggtgtc agacggttacctcgccaactctgatccaaaagggaagaagctgtgtcgtatgtgatcaactctactcatttg accctgtctattacggccctgatgctgctgagttcagaccagagagatgggttgagccatcaacccaaaaa gctcgtgctgggcttacttgccattcaacgggtgttccaagaaatctgtttgggtcagcagtttgccttgacgga agctggctatgtgttggcttagattgggtcaagagttctccacgttaggtcggaccagcagaggtgtacc cgccaaagagggtgaccaacttgaccatgtgttgcaggatggtgtctattgtcaagtttgactag
SEQ ID NO:20	cytochrome P450 A18 (CYP52A18) Nuc. Seq	atgattgaacaaatcctagaatattggtatattgtgtcctgtgtgtacatcatcaaaacactcattgccta cagaagaactcgcgtcttgatgaaaacagttgggtgtcgtctcaatcaacaaacagttgtacgacaacggt ttcgttatcgtcaacggatggaaggctctccagttcaagaaaggaggcagagctcaagagtacaacg atcacaagtttgacagctccaagaacccaagcgctcgccacctatgtcagtatctttttggcaccagatt gtcgtgaccaaggatccagagaaatatcaaagctatttggcaacccagtttggcgattttctttgggcaag agacacgctcttttaaaccttgttaggtgatggatcttcaccttggacggcggaaggctggaagcatag cagatccatgttaagaccagtttgcagagaacaagttgctcatgtgacgtcgttggaaaccaacttc cagttgttgaagaagcatatccttaaacacaagggtgagtaattgtatcaggaattgtctttagatttac tgtcgaactcggccaaggagttcttatttggtagtcctgtcactccttaaggagcgaactatcggatcaaa ccaagacgatalagattttgctggtagaaaggacittgctgagtcgttcaacaaaggccaggagatttgt ctattagaattttgtgtgcagaccttactggtgatcaacaacaaggagtttagagactgtaccaagctgg tgcaacaagtttaccactactatgttcagaaaggcttggatgtcaccagagggaacttgaagcaagcaag gctgggtatgttcttgaatgagctgtcaagcagacgagagaccccaagggtgtgtcgttgaccagcttttg aacactctgttggcaggaagagacacactgctgggtgtgtccttgcgtgtgtttagagcttgcagaaac ccacacatctgggccaagttgagagaggaaattgaacagcagtttggctcgttggaagactctcgtgtgtg

SEQ ID NO:	Description	Sequence
		aagagattacctttgagagctgaagagatgtgagtaactgaaagcgttcctaacgaaaccttgcgtgttt acccaagtgcccaagaaactcagaatcgccaccaagaatacaacattgccaggggtggtgtcc agacgggtaccagccaatcttgatccaaaaggagagaggtgtgctgatgtatcaactctaccactta gatcctgtctattatggcctgatgctgctgagttcagaccagagagatggtttgagccatcaaccagaaa gctcggctgggcttactgccaatcaacgggtggccacgaatctgttgggtcagcagttgcttgaccga agctgttacgttttggtcagattggtcaagagttctccacattaggctggaccagatgaagtgtatcc accaaagaggttgaccaactgaccaigtgttgcaggatggtgctattgtcaagttgactag
SEQ ID NO:21	cytochrome P450 A19 (CYP52A19) Nuc. Seq	atgctcgatcagatcttactactggtacattgtcttgcattgttggccattatcaaccagatcgtggctcat gtcaggaccaattattgtagaagaaattgggtgctaagccattcacacogtccaacgtgacgggtggt igggcttcaaattcggcgtgaattctcctcaaaagcaaaaagtgtgggagactggttattatcatctccc gtttccacgataatgaggacactttctccagctatgctttggcaacctgtggttccaccagggaacccg agaatatcaaggcgttttggcaaccagtttggatgtttcattgggacagcagggtcaagttctcaaac cattattgggtacgggtatctcattgacgacgagggctggaagcacagcagagccatgttagac cacagtttgcagagaaacagttgctcatgtgacgtcgttggaaacacactccagttgtgaagaagcat atccttaaacacaagggtgagtaacttggatccagggaattgttcttagattactgtcgtactcggccacgg agttctatttggtagtccgtgactcctaaaggacgaggaaattggtacgacacgaaagacatgtct gaagaaagacgcagatttgcgacgcgttcaacaagtcgaagtcacgtggccaccagagttgcttca cagaactgtactgttgggtcaacaacaaagagttcaaggagtgcaatgacattgtccaagttacca actactatgttcagaaagccttggatgctaccccagaggaaactgaaaagcaaggcgggtatgttctt gtatgagcttgcagcagacgagagaccccaagggttgcgtgaccagcttgaacatcttgttggca ggaagagacaccactgctgggttgttcttgcgtgtttagtggccagaaacccacacatctgggc caagttgagagaggaaattgaacagcagtttggcttggagaagactcgttgaagaattaccttgg agagctgaagagatgtgacttgaaggcgttgaacgaaactttagattacaccaagtgctcc aagaaaacgcaagatttgcgattaaagacacgactttaccaagaggcgggtggccccaacggcaagga tcctatctgatcaggaaggatgaggtggtagtactccatctcggcaactcagacaaatcctgcttatta igggcgcgtagtctgctgatttagaccgaaagatggttgaacatcaactagaaacttgggatgggctt icttgcattcaacgggtggtccaagaatcgttgggacaaacagtttgccttgactgaagcgggtacgtttt gttagactgttcaggagttccaaactgtcacaagaccccgaaaccaagttaccaccacctagattgg cacactgacgatgtgctgttgcaggtgcacacgtcaagatgtcatag
SEQ ID NO:22	cytochrome P450 A20 (CYP52A20) Nuc. Seq	atgctcgaccagatcttccattactggtacattgtcttgcattgttggctcattcaagcagatcgtggctcat gccaggaccaattattgtagaagaagttggcgctaagccattcacacatgtccaactagacgggtgg ttggcttcaaatttggcgtgaattctcctcaaaagctaaaagtgtgggagggcagggttattatcatctccc gtttccacgataatgaggacactttctccagctatgctttggcaacctgtggttccaccagggaacccg agaatatcaaggcgttttggcaaccagtttggatgtttcattgggaagcagggtcaaatcttcaaac cattgttgggtacgggtatctcacttggacgagcgaaggctggaagcacagcagagccatgttagac cacagtttgcagagagcaagttgctcatgtgacgtcgttggaaacacattccagttgtgaagaagcat attcttaagcacagggtgaatacttgaatccagggaattgttcttagattaccgttgaatcagcgaaggga gtttatttggtagtccgtgcaacttaagggaagcgaaggaaattggtacgatacgaaggacatggctg aagaaagacgcaaatggcagcgttcaacaagtcgaagtcatttgcaccagagttgctttaca gacattgtactgttggtaacaacaaagagttcaaggagtgcaacgacattgtccacaagttcaccaa ctactatgttcagaaagccttggatgctaccccagaggaaactgaaaaaacaggcgggtatgttcttgt acgagcttgcacagcagaaagaccccaatgttgcgtgaccagcttggaaatcttgttggctgg aagggaacaccactgctgggttgttcttgcgtgttgcgttggcagggaacccacacatctgggcca agttgagagaggaagaaatgaatcacacttggcgtgggtgaggactcgtgttgaagagattactttag agcttgaagagatgtgacttgaagcgttgaacgaaacgttgaattacaccaagtgctcca agaaacgcaagatttgcattaaagacacgactttaccaagaggcgggtggcccaacggcaaggatc ctatcttgatcagaaagaatgaggtgggtcaatactccatctcggcaactcagacaaatcctgcttattg gcgccgatgctgctgatttagaccggaaagatggtttagccatcaactagaaacttgggatgggctta cttgcattcaacgggttccaagaatcgttgggacaacagtttgccttgaccgaagcgggttacgtttt gttagactgttcaggaaatccctagctgtcagcaggaccccgaaactgagtaccaccacctagattggc acactgacgatgtgctgttgcaggtgcacacgtcaagatgtcatag

SEQ ID NO:	Description	Sequence
SEQ ID NO:23	cytochrome P450 D2 (CYP52D2) Nuc. Seq	atggctatatctagttgctatcgtaggagtgatctgtgctgtcttcaattgctgttggttatfctgggtatgaat attgttataactaaatactttagtcacaaacatggcgctcgagaaatcgagaatgtatcaacgatgggttc ittgggttccgcttaccttgcactcatgtagcgagccagcaatgagggccgacttatcgagttcagtgcaag agattcgagtcggcgccacatccacagaacaagacattggtcaaccgggcaatgagcgttccgtgtgata ctaccaaggacccagtgaaatcaaaagcgatgctatcgacccagtttgatgactttccctgggttgag actacaccagtttgcgctgttggggaaaggcatcttacttggacggcccgagtggaagcagagc cgatctatgttgcgtcgcaatttgcacaaagtcgggttctcatatcctggatctagaacgcatttggttg cttcggaagcacattgtaggccacaatggagactcttcgacatccaggagctctacttccgggttctcgat ggatgtggcgacgggggtttgttggcgagtcgttgggtcggtgaagacgaagatgcgaggttcttg aagcattcaatgagtcgagaagtatttggcaactagggcaacggtgcacgagttgacttcttggtagc ggtttaggttccgagtaacaacaggttggcgaaagtctgcagccagtggtccacaaggcgttagat gttgacccggaagacacagcgagtagctgttttccgagtggtggtcaaacacactcgagatccggtt gtttacaagaccaagcgttgacgtctgtctggacgcgacaccccgctgttattatcggttgc acatttgagctagccgggaatgaccacatgtggaggaagctacgagaggaggttatcctgacgatggg accgtccagtgatgaaataaccgtggcggttgaaagttgcgttacctcaaacgaatcctaaacga aactcttcgactataccaagtgtgcttaggaacgcgagattgtacgaggaatcgcagcttctcgt ggcgaggtccagatggatcggttccgatttgataagaaaggccagccagtggggtatttcaattgtgt acacactggaatgaaaggataggggaatgataccatgttgcgacccgagagatgggctgcgtta gagggcaagagttgggtgtgtatcttcaatcaacggcgcccgagaagctgcctgtgcagcagt tgcatacctgaagcttcgtatgtttggctcgtattgacacagtgctacacgacgatacagctagaactac cgagtacccaccaaagaaactcgttcactcacgatgagcttctcaacgggggtgtacatccgaactag aactga
SEQ ID NO:24	cytochrome P450:NADPH P450 reductase (Bacillus megaterium) nucleotide Nuc. Seq	atgacaattaaagaaatgcctcagccaaaaacgttggagagcttaaaaattaccgltattaaacacag ataaacgggttcaagctttagtgaataatgaggaataatcttaaatcgaggcgctggt cgtgtaacgcgctactatcaagtcagcgtctaataaagaagcatgcatgaatcagcttgcataaaa acttaagtaagcgcttaaaattgtacgtgatttgcaggagacgggtatttacaagctggacgcgatgaaa aaaattgaaaaaagcgcataatcttacttcaagctcagtcagcaggcaatgaaaggctatcatg cgatgatggtcgatatcgccgtgcagcttgcataaaagtgaggcgctctaaatgcagatgagcatattga agtaccggaagacatgacacgtttaaagcttgatacaattggtcttgcggctttaaactatcgctttaaagc ittaccgagatcagcctcatcatttatacaagtatggtcggtgcactggatgaagcaatgaacaagctg cagcgagcaaatccagacgacccagcttatgatgaaaacaagcgccagtttcaagaagatatcaagg tgatgaacgacctagtagataaaatttgcagatcgcaaaagcaagcggtgaacaagcgatgattat taacgcataatgaacggaagaaatccagaaacgggtgagccgcttgatgacgagaacattcgtat caaatattatcattctaaattgcgggacacgaaaacaacagtggtctttatcatcttgcgtgatttctatg aaaaatccacatgtattacaacagcagcagaagaagcagcagcaggttctagtagatccgttccaag ctacaacaaagtaaacagcttaaatatgtcggcatggtcttaaacgaagcgctgcgttatgccaact gctcctgcgtttccctatagcaaaagaagatacgggtcttgaggagaataatccttagaaaaaggcg acgaactaatggttctgattcctcagcttcacggtgataaaacaatttggggagacgatgtggaagagttc cgtccagagcggttgaataccaagtgcatccgcagcatgcgtttaaaccggttggaaacgggtcagc gtgcgtgatcggtcagcagttcgcttcatgaagcaacgctggtacttggtatgctaaaaactttg acttgaagatcatacaaaactacgagctgataaaagaaacttaacgltaaaacctgaaggcttgtg gtaaaagcaaaatcgaacaaatccgcttggcggtattccttcaactagcactgaacagctctgctaaaa aagtagcaaaaaggcagaaaaacgctcataatacgcgctgctgtgtatcaggttcaaatatggga acagctgaaggaaacggcgctgatttagcagataatgcaatgagcaaaaggaattgcacgcgaggtcgc aacgctgattcacacgcccgaatcttccgcggaaggagctglatattgtaacggcgcttataacg gtcatccgctgataacgcaaaatgtgcagtggttagaccaagcgtctgctgatgaagtaaaag gcgttcgctactcgttattgtagtcggcgataaaaactgggctactacgtatcaaaaagtgctgctttat cgatgaaaacgcttgcgctaaaggggcagaaaaacatcgctgacggcggtgaagcagatgaagcga cgacttgaaggcacaatgaagaatggcgtaacatatgtggagtgacgtagcagcctaattaaactc gacattgaaaacagtgagataataaatctacttcttcaatttgcagcagcgccgagataatgcc gcttgcgaaaatgcacggtgcgttttcaacgaacgctgtagcaagcaaaagaaactcaacagccaggca gtgcacgaagcagcgacatcttgaattgaacttcaaaaagaagcttctatcaagaaggagatcattt

SEQ ID NO:	Description	Sequence
		aggtgttattcctcgcaactgaaggaaatagtaaaccgtgtaacagcaaggttcggcctagatgcatca cagcaaatccgtctggagcagaagaagaaaaattagctcattgccaactcgtaaaacagtatccgta gaagagcttctgcaatacgtggagctcaagatcctgttacgcgcacgcagcttcgcgaatggctgcta aaacgggtctgcccgcgcataaagtagagcttgaagcctgtctgaaaagcaagcctacaaagaaca agtgcctggcaaaacgtttaacaatgcttgaactgctgaaaaatacccgcgctgtaaatgaaattcagc gaatttatcgccctctgccaagcatacgcccgcgtattactcgatttctcatcacctcgtgctgatgaaa aacaagcaagcatcacggtcagcgtgtctcaggagaagcgtggagcggatattggagaataataaagg aattgcgtcgaaactatctgcccagctgcaagaaggagatacgattacgtgcttatttccacaccgcagt cagaatttacgtcgcaaaaagaccctgaaacgcgcgttatcatggtcggaccgggaacaggcgctgc gccgttttagaggcttctgtaggcgcgcaaacagctaaaagaacaaggacagtcacttggagaagca catttatacttcggctgcgcgttcacactcatgaagactatctgtatcaagaagagcttgaaaacgcctaaag cgaaggcatcattacgtctcataccgcttttctcgatgccaatcagccgaaaacatacgttcagcagc taatggaaacagacggcaagaaatgattgaactcttgatcaaggagcgcacttctatatttgcggaga cggagccaaatggcacctgcctgtagcaacgcttatgaaaagctatgctgacgttcaccaagtgag gtgaagcagacgctcgcttatggctgcagcagctagaagaaaaaggccgatacgcaaaagacgtgt gggctgggtaa
SEQ ID NO:25	NADPH cytochrome P450 reductase, CPR (C. tropicalis strain ATCC750) Nuc. Seq	atggcattagataagttagatttatagttattataacattgggtggtgcaattgcagcttatttgcagaagaatc agtttcttgaccaacaacaagataccgggttccttaatactgatagtgagatggttaattcaagagatatct tacaagcttgaagaagaacaataaaaaatcgttattatttggatcccaaacagggtacagcagaag attatgccaacaattgtcaagagaattgcattcaagatttgggttgaaaacatgggttgcgtattcgtgat tatgatctcgaaaaactcggagataattactgaagatactctggtttcttattgttgctacttatggtagaggtga accaacggataatgctgacgaatttcaacttgggtgactgaagaagctgacaccttgacttctgaaat atactgttttgggttggttaattcaacttatgaattctcaatgctatggtagaaaaattgacagattgttggga gaaaaagggtgtagcagatttgcgaatacggtagaaggtgacgatggtagctttagatgaagattt ctggcctggaaaggataacgtgttgccttaagaagaatgattgaatttgaagaaaaagagttgaaata cgaaccaaatgttaattgactgaagagatgatttatctggcaatgatccagatgtctcctgggtgaac caaatgtcaaatacattaatactgaaggtgtgacttaactaaaggctcatttgatcactcatccattttg gctagaattgttaaaactaaagaattgttactctgaagacagacattgttctcatgttgaaattgatttctg aatcaaatgaaalataccaccgggtgatctctgcaatctggccatctaactctgatgaaaacattaag caatttgcctaaatgtttgtttagaagacaaactgatactgttattgaattgaagctttggattccactatt ccatcccatccctaataccaatcattatggagctgtatttagacaccatttggaaattcaggtcctgtttcta gacaatttttctatctattgctgatttgcctctgatgaagaaactaaaaagcatttactagaatttgggtg ataagcaagaatttctagtaaaagtcacccgtagaaaattcaacattgcgatgcttattatttgcctcaa caacagaccatggtcggatgttccattcgaattccttatgaaaatgtccaacacttaactcctcgttattact ccatttcttctcctcattaaagtgaagcaaaccaatattgtactgctgttgaagccgaagaagaag ctgatggaagaccagttactggtgtgtcaccaactgttgaagaattgaattgaacaaaacaaact ggtgaaacccaatggtcattatgattgaatggtccaagaggcaaaattagcaagttcagattgccagtt cacgttagaagatcattcaaatccaagaatagcactaccccgatttattgattgtccagggtacc ggtgttgaccattgagagggtttagagaaagaggtcaacaagttaaaaatggtgttaattgttgaag actgtattgtttatggatgtagaaattccgaacaagaattctgttacaacaagaattggagtgaaatgcct cagtattgggagaaaaattcgaaatgttaattgccttctcaagacaagatccaactaagaagtttatgttc aagataagattttagaaaatagtgctctgttgatgagttattatctagtgagcaattatttattgttgggtga tgcagtagaatggctagagatgttcaagctgcaattgccaagattgttgcctaaagtagagatatccac gaagataaagctgctgaattggttaaatcttggaaagttcaaatagataccaagaagatgtctggttaa
SEQ ID NO:26	NADPH cytochrome P450 reductase A, CPRA (Candida strain ATCC20336) Nuc. Seq	atggcttagacaagttagattgtatgtatcataacattgggtgctgtagccgctatttgcctaaagaa ccagttccttgatcagcccaggacacgggttcctcaacacggacagcgggaagcaactccagagac gtcttgctgacattgaagaagaataataaaaacacgtgtgttgggtccagacgggtacggcaga agattacgccaacaaattgtccagagaattgcactccagatttggcttgaaaacgatggttgagatttgc ctgattacgattgggataactcggagatacaccgaagacatctggtgttttcttattgttgcacctaaggt gaggtgaaacctaccgataatgcgacgagttccacacctggtgactgaagaagctgacactttgagt acctgaaatacacccgtgttgggtgggttaactccacgtacgagttctcaatgccaattgtagaaaagttt

SEQ ID NO:	Description	Sequence
		gacagattgtgagcgaaggggtgacagggttgcgaatacgcgtgaagggtgatgcggactggc accttggacgaagattcatggcctggaaggacaatgtcttgacgccttgaagaatgatttgaacttga gaaaagggaattgaagtacgaaccaaactgaaattgactgagagagacgacttctgctgctgactc ccaagttccttgggtgagccaaacaagaagtacatcaactccgagggcacgacttgaccaagggtc cattcgaccacacccaccatactggccagaatcacgagacgagagagtgttcagctccaaggac agacactgtatccagttgaatttgacatttctgaatcgaactgaaatcacaccacgggtgaccatctagct atctggccatccaactccgacgaaaaattaagcaatttccaagtgttcggattggaagataaactcg acactgttattgaattgaaggcgttggactccactacaccatccattcccaacccaattacctaagggtg ctgtcattagacaccatttagaaatctccgggtccagctctcgagacaattcttttgcattgctgggttgc ctgatgaagaaacaagaaggcctttaccagacttgggtggaagcaagaattcgcgcgaagggtc accgcgagaaagtcaacattgccgagcttctgtatattctccaacaacgcctcattggctcgatgttctt tgaattcctattgaaaacgttccacactgactccacgttactactcattctcgtctgctcattgagtgaaa agcaactcatcaacgttactgacgttgtgaagccgaagaagaagctgatggcagaccagtcactgggt gtgtccaccaactgttgaagaacgttgaattgtgcaaaacaagactggcgaaaagccactgttccact acgatttgagcggcccaaggaggcaagttcaacaagttcaagttgccagtcatgtgagaagatccaac ttaagttgcaaagaactccaccacccagttatctgtattgggtccagggtactgggtgccccattgaga ggttttgcagagaagaagttcaacaagtcaagaatgggtgcaatgttggcaagactgtgttttatgggtg cagaaactccaacgaggaacttttgcacgaagaatgggcgagtagcgtctctgtttgggtgaaaac ttgagatgtcaatgccttctccagacaagaccatccaagaaggttacgtccaggataagatttaga aaacagccaactgtgacgagttgtgactgaagggtccattatcagctctgtgtgtagccagtagaa tggctagagacgtgcagaccacaatttccaagattgtgtctaaagcagagaaattagtgaaagacaag gctgctgaattgtcaagtcctggaagggtccaaaatagataccaagaagatgttggtag
SEQ ID NO:27	NADPH cytochrome P450 reductase B, CPRB (Candida strain ATCC20336) Nuc. Seq	atggcttagacaagttgattgtatgtcatcataacattgggtgctgctggccgcctatttgcgaaga ccagttccttgatcagccccaggacacgggttctccaacaggacagcgggaagcaactccagagac gtcttgctgacattgaagaagaataataaaaacagttgttgttgggtcccagacgggtacggcaga agattacgccaacaattgtcaagagaattgcactccagatttggcttgaanaacatgggtgcagatttgc ctgattacgattgggataacttcggagatcacccaagatatctgtgttttcatcgttgcacctaagggt gaggggtgaacctaccgacaattgcccagcaggttccacacctggtgactgaagaagctgacacttgagt actttgagatatacctgttccgggttgggttaactccaactacaggttctcaatgctattggtagaagttga cagattgttgagtgaagaagggtggaacagatttgcgaatgtggaaggtagcagcggcactggcac cttgacgaagatttcatggccttgaaggataatgtcttgacgccttgaagaatgactgaactttgaaga aaagggaattgaagtacgaaccaaactgaaattgactgagagagatgacttctgctgctgcgactccc aagtttcttgggtgagccaaacaagaagtacatcaactccgagggcacgacttgaccaagggtccat tcgaccacacccaccatacttggccaggatcacgagaccagagagttgttcagctccaaggaaag acactgtattcacgttgaatttgacatttctgaatcgaacttgaatlacaccacgggtgaccatctagccat ctggccatccaactccgacgaaaaacatcaagcaatttccaagtgttcggattggaagataaactoga cactgttattgaattgaaggcattggactccacttaccattccattcccaactccaattactacggtgctg tatttagacaccatttagaaatctccgggtccagctctcgagacaattcttttgcgattgctgggttgcctc atgaagaaacaagaagacttccaccagacttgggtggaacaaagaattcgcaccaagggttac cgcgagaagttcaacattgcgagcttctgttatattctccaacaacactccatggctcgatgttctttg agttccttattgaaaacatccaacacttgactccacgttactactccatttcttctcgtcgttgagtga caactcatcaatgttactgacgtcgttggagccgaagaagaagcgcgattggcagaccagtcactggtgt gttaccacactgttgaagaacattgaatttgcgaacaaagactggcgaaaagccactgttactac gatttgagcggcccaaggaggcaagttcaacaagttcaagttccagtgacgtgagaagatccaacttt aagttgccaaagaactccaccacccagttatcttgattggtccaggtactggtgttgcctatgagagg ttcgttagagaaagagttcaacaagtcaagaatgggtgcaatgttggcaagacttgggtttatgggtgca gaaactccaacgaggacttttgcacgaagaatgggcgagtagcgtctgttttgggtgaaaacttt gagatgttcaatgccttcttagacaagaccatccaagaagggttacgtccaggaataagatttagaaa acagccaactgtgcacgaatttgaccgaagggtccattatcagctctgtgtgtagccagtagaatg gccagagacgtccagaccagatctccaagattgttgcacaaagcagagaaatcagtgaaagacaag gccgctgaattgtgcaagtcctggaaagtccaaaatagataccaagaagatgttggtag

SEQ ID NO:	Description	Sequence
SEQ ID NO:28	cytochrome P450:NADPH P450 reductase (Bacillus megaterium) amino acid [P450 activity shown in italics, P450 reductase activity shown in normal font] A.A. Seq	<i>mtikempqpkftgelknlpIntdkpvqalmkiadelgeifkfeapgrvtrylssqrlikeacdesrfdknls qalkfvrdagdgiftswtheknwkkahnillpsfsqqamkgyhammvdiaqvqlvqkwerlnadehie vpedmtrltldtligcgfnrfsfyrdqphpfitsmvrasdeamnksqranpddpaydenkrqfqedik vmndlvdkiiadrkasgeqsddllthmlngkdpetgeplddeniryqiitfliaghettsqllsfasyflvkn hvlqkaaeearvlvpvpsykqvkqlkyvgmvlneasrlwptapafsllyakedtvlgggeyplekgdel mvsipqlhrdktiwdvvefrperfenpsaiphafkpfngqracigqqfalheatsvlgmmlkhfdf edhtnyesdikettllkpegfivvkakskkiplggipspssteqsakvrkkaenahntpslvlygsnmgt egtardladiamskgfapqvattldshagnlpregavlivtasynghppdnakqfvdwldqasadevkg vrysvfgcgdknwattyqkvpafidellaakgaeniadrgeadasddfegtyeewrehmwsvdaayf nldiensednkslslqfvdsaadmplakmhgafstnvaskelqpggsarstrhleilpkeasyqeg dhlgviprmyegivnrvtarfgldasqirseaeeeklahlplaktvsveelsqyvelqdpvtrtqlramaak tvcpphkveleallekqaykeqsakrltmleslekypacemkfselialspsirprysissprvdekq asitvsvvsgeawsgygeykgyasnylaesqegdtitctfstpqseftspkdpetplimvgpgtgvapfrg ivqarkqlkeqqslgeahlyfgcrsphedysyqeelenaqsegitlhtafrmpnqpktyvqhvmeg dgkklieIdqgahfyicgdgsqmapaveatlmksyadvhqvseadarlwsqleekgryakdvwaag *</i>
SEQ ID NO:29	acyl CoA oxidase, POX4 (Candida strain ATCC20336) nucleotide	ATGACTTTTACAAAGAAAAACGTTAGTGTATCACAAAGGTCCTGACCCTA GATCATCCATCCAAAAGGAAAGAGACAGCTCCAAATGGAACCCCTCAAC AAATGAACACTACTTCTTGAAGGCTCCGTCGAAAGAAGTGAGTTGATGA AGGCTTTGGCCCAACAAATGGAAGAGACCCCAATCTTGTACAGACG GCTCCTACTACGACTTGACCAAGGACCAACAAAGAGAATTGACCGCCG TCAAGATCAACAGAATCGCCAGATACAGAGAACAAGAATCCATCGACA CTTTCAACAAGAGATTGTCCTTGATTGGTATCTTTGACCCACAGGTGG GTACCAGAATTGGTGTCAACCTCGGTTTGTTCCTTTCTTGATCAGAGG TAACGGTACCACTTCCCAATTGAACTACTGGGCTAACGAAAAGGAAAC CGCTGACGTTAAAGGTATCTACGGTTGTTTCGGTATGACCGAATTGGC CCACGGTTCCAACGTTGCTGGTTTGGAAACCACCGCCACATTTGACAA GGAATCTGACGAGTTTGTCAACACCCACACATTGGTGCCACCAA GTGGTGGATTGGTGGTGTCTGCTCACTCCGCCACCCACTGTTCTGTCTA CGCCAGATTGATTGTTGACGGTCAAGATTACGGTGTCAAGACTTTTGT TGTCCCATTGAGAGACTCCAACCACGACCTCATGCCAGGTGTCACTGT TGGTGACATTGGTGCCAAGATGGGTAGAGATGGTATCGATAACGGTTG GATCCAATTCTCCAACGTCAGAATCCCAAGATTCTTTATGTTGCAAAAG TTCTGTAAGGTTTCTGCTGAAGGTGAAGTCACCTTGCCACCTTTGGAA CAATTGTCTTACTCCGCCTTGTGGGTGGTAGAGTCATGATGGTTTTG GACTCCTACAGAATGTTGGCTAGAATGTCCACCATTGCCTTGAGATAC GCCATTGGTAGAAGACAATTCAAGGGTGACAATGTCGATCCAAAAGAT CCAAACGCTTTGGAACCCAATTGATAGATTACCCATTGCACCAAAAG AGATTGTTCCCATACTTGGCTGCTGCCTACGTCATCTCCGCTGGTGCC CTCAAGGTTGAAGACACCATCCATAACACCTTGGCTGAATTGGACGCT GCCGTTGAAAAGAACGACACCAAGGCTATCTTTAAGTCTATTGACGAC ATGAAGTCATTGTTTGTGACTCTGGTTCCTTGAAGTCCACTGCCACTT GGTTGGGTGCTGAAGCCATTGACCAATGTAGACAAGCCTGTGGTGGT CACGGTTACTCGTCTACAACGGCTTCGGTAAAGCCTACAACGATTGG GTTGTCCAATGTACTTGGGAAGGTGACAACAATGTCTTGGCCATGAGT GTTGGTAAGCCAATTGTCAAGCAAGTTATCAGCATTGAAGATGCCGGC AAGACCGTCAGAGGTTCCACCGCTTTCTTGAACCAATTGAAGGACTAC ACTGGTTCCAACAGCTCCAAGGTTGTTTTGAACACTGTTGCTGACTTG GACGACATCAAGACTGTCATCAAGGCTATTGAAGTTGCCATCATCAGA TTGTCCCAAGAAGCTGCTTCTATTGTCAAGAAGGAATCTTTCGACTATG TCGGCGCTGAATTGGTTCAACTCTCCAAGTTGAAGGCTCACCCTACT

SEQ ID NO:	Description	Sequence
		TGTTGACTGAATACATCAGAAGAATTGACACCTTTGACCAAAAGGACTT GGTTCCATACTTGATCACCCCTCGGTAAGTTGTACGCTGCCACTATTGT CTTGGACAGATTTGCCGGTGTCTTCTTGACTTTCAACGTTGCCTCCAC CGAAGCCATCACTGCTTTGGCCTCTGTGCAAATTCAAAAGTTGTGTGC TGAAGTCAGACCAAACGTTGTTGCTTACACCGACTCCTTCCAACAATC CGACATGATTGTCAATTCTGCTATTGGTAGATACGATGGTGACATCTAT GAGAACTACTTTGACTTGGTCAAGTTGCAGAACCCACCATCCAAGACC AAGGCTCCTTACTCTGATGCTTTGGAAGCCATGTTGAACAGACCAACC TTGGACGAAAGAGAAAAGATTGAAAAGTCTGATGAAACCGCTGCTATC TTGTCCAAGTAA
SEQ ID NO:30	acyl CoA oxidase, POX4 (Candida strain ATCC20336) amino acid	MTFTKKNVSVSQGPDPRSSIQKERDSSKWNPQQMNYFLEGSVERSELM KALAQQMERDPIFTDGSYYDLTKDQQREL TAVKINRIARYREQESIDTFN KRLSLIGIFDPQVGTRIGVNLGLFLSCIRNGTTSQNLNYWANEKETADVKG YGCFGMTLHAGSNVAGLETTATFDKESDEFVINTPHIGATKWWIGGAH SATHCSVYARLIVDGQDYGVKTFVPLRDSNHDLMPGVTVGDIGAKMGR DGIDNGWIQFSNVIRPRFFMLQKFKVSAEGEVLPPLEQLSYSALLGGR VMMVLD SYRMLARMSTIALRYAIGRRQFKGDNVDPKDPNALETQLIDYPL HQKRLFPYLAAYVISAGALKVEDTIHNTLAELDAAVEKNDTKAIFKSIDDM KSLFVDSGLKSTATWLGAEAIDQCRQACGGHGYSSYNGFGKAYNDWV VQCTWEGDNNVLAMSVGKPIVKQVISIEDAGKTVRGSTAFNLQKDYTGS NSSKVVLNTVADLDDIKTVIKAIEVAIRLSQEAASIVKESFDYVGAELVQL SKLKAHHYLLTEYIRRIDTFDQKDLVPYLITLGKLYAATIVLDRFAGVFLTFN VASTEAITALASVQIPKLCAEVRPNVVAYTDSFQQSDMIVNSAIGRYDGDY ENYFDLVKLQNPSPKTKAPYSDALEAMLNRPTLDERERFEKSDETAAILSK *
SEQ ID NO:31	acyl CoA oxidase, POX5 (Candida strain ATCC20336) nucleotide	ATGCCTACCGA/ACTTC/AAAAAG/AAAGAGAACTCACC/AGTTCA/ACCCA AAGGAGTTGAACTACTTCTTGGAAGGTTCCCAAGAAAGATCCGAGATC ATCAGCAACATGGTCAACAAATGCAAAAAGACCCATCTTGAAGGTC GACGCTTCATACTACAACCTTGACCAAAGACCAACAAAGAGAAGTCACC GCCAAGAAGATTGCCAGACTCTCCAGATACTTTGAGCAGAGTACCCA GACCAACAGGCCAGAGATTGTGATCCTCGGTGTCTTTGACCCACAA GTCTTCACCAGAATCGGTGTCAACTTGGGTTTGTGTTTCTGTGTCC GTGGTAACGGTACCAACTCCCAGTTCTTCTACTGGACCATAAATAAGG GTATCGACAAGTTGAGAGGTATCTATGGTTGTTTTGGTATGACTGAGTT GGCCACGGTTCACAGTCCAAGGTATTGAAACCACCGCCACTTTTGA CGAAGACACTGACGAGTTTGTCAACACCCACACATTGGTGCCAC CAAGTGGTGGATCGGTGGTGTCTGCGCACTCCGCCACCCACTGCTCCG TCTACGCCAGATTGAAGGTCAAAGGAAAGGACTACGGTGTCAAGACCT TTGTTGTCCCATTGAGAGACTCCAACCACGACCTCGAGCCAGGTGTGA CTGTTGGTGACATTGGTGCCAAGATGGGTAGAGACGGTATCGATAAC GGTTGGATCCAGTTCTCCAACGTGAGAATCCCAAGATTCTTTATGTTGC AAAAGTACTGTAAGGTTTCCCGTCTGGGTGAAGTCACCATGCCACCAT CTGAACAATTGTCTTACTCGGCTTTGATTGGTGGTAGAGTACCATGAT GATGGACTCCTACAGAATGACCAGTAGATTCAACCATGCTTGCCTTGAG ATACGCCATCCACAGAAGACAATTCAAGAAGAAGGACACCGATACCAT TGAAACCAAGTTGATTGACTACCCATTGCATCAAAGAGATTGTTCCCA TTCTTGGCTGCCGCTTACTTGTCTCCCAAGGTGCCTTGACTTAGAAC AAACCATGAACGCAACCAACGACAAGTTGGACGAAGCTGTCAAGTGT GGTGAAGGAAAGCCATTGACGCTGCCATTGTGCAATCCAAGAAATTG TTCGTGCTTCCGGTTGTTGAAGTCCACCTGTACCTGTTGACTGCT GAAGCCATTGACGAAGCTCGTCAAGCTTGTGGTGGTCACGGTACTC

SEQ ID NO:	Description	Sequence
		GTCTTACAACGGTTTCGGTAAAGCCTACTCCGACTGGGTTGTCCAATG TACCTGGGAAGGTGACAACAACATCTTGGCCATGAACGTTGCCAAGCC AATGGTTAGAGACTTGTGTAAGGAGCCAGAACAAAAGGGATTGGTTCT CTCCAGCGTTGCCGACTTGGACGACCCAGCCAAGTTGGTTAAGGCTTT CGACCACGCCCTTTCCGGCTTGGCCAGAGACATTGGTGCTGTTGCTG AAGACAAGGGTTTCGACATTACCGGTCCAAGTTTGGTTTTGGTTCCA AGTTGAACGCTCACAGATTCTTGATTGACGGTTTCTTCAAGCGTATCAC CCCAGAATGGTCTGAAGTCTTGAGACCTTTGGGTTTCTTGATGCCGA CTGGATCTTGACCAACTTTGGTGCCACCTTCTTGACGTACGGTATCATT ACCCCAGATGTCAGCAGAAAGATTTCTCCGAGCACTTCCAGCCTTG TGTGCCAAGGTTAGACCAAACGTTGTTGGTTTGACTGATGGTTTCAAC TTGACTGACATGATGACCAATGCTGCTATTGGTAGATATGATGGTAAC GTCTACGAACACTACTTCGAACTGTCAAGGCTTTGAACCCACCAGAA AACACCAAGGCTCCATACTCCAAGGCTTTGGAAGACATGTTGAACCGT CCAGACCTTGAAGTCAGAGAAAGAGGTGAAAAGTCCGAAGAAGCTGC TGAAATCTTGTCCAGTTAA
SEQ ID NO:32	acyl CoA oxidase, POX5 (Candida strain ATCC20336) amino acid	MPTELQKERELTKFNPKELNYFLEGSQERSEIISNMVEQMOKDPIKVDAS YYNLTKDQQREVTAKKIARLSRYFEHEYDPDQQAQRLSILGVFDPQVFRIG VNLGLFVSCVRNGTNSQFFYWTINKGIDKLRGIYGCFGMTTELAHGSNVQ GIETTATFDEDTDEFVINTPHIGATKWWIGGAHSAHCSVYARLKVKGKD YGVKTFVPLRDSNHDLEPGVTVDIGAKMGRDGDIDNGWIQFSNVRIPRF FMLQKYCKVSRSGEVTMPPEQLSYALIGGRVTMMMDSYRMTSRFITIA LRYAIHRRQFKKKTDTIETKLIDYPLHQKRLFPFLAAAYLFSQALYLEQT MNATNDKLDEAVSAGEKEAIDAAIVESKFLFVASGCLKSTCTWLTAEAIDE ARQACGGHGYSSYNGFGKAYS DWVWVQCTWEGDNNILAMNVAKPMVRD LLKEPEQKGLVLSSVADLDDPAKLKAFDHALSGLARDIGAVAEDKGFDT GPSLVLVSKLNAHRFLIDGFFKRITPEWSEVLRPLGFLYADWLTNFGATFL QYGIITPDVSRKISSEHFPALCAKVRPNVGLTDGFNLDTMMTNAAIGRYD GNVYEHYFETVKALNPPENTKAPYSKALEDMLNRPDLEVRERGEKSEEA AEILSS*
SEQ ID NO:33	Acyl-CoA Hydrolase (ACHA) Nucleotide Seq	atgatcagaaccgtccgttatcaatccctcaagagggtcagacctctggcttctctctgtttctgtccacg ctacaactcccagaaggccaattccaccgtccagaccaccctgggtccgacgagccagctgaagcc gccgacgccgcgcacgatcctgcgcgagttgcgagacaagcagacgaacccgaacaaggccac ctggctcgatgcgttaacgagcgggagaagttgcgtgccgagggcaagacgattgacagttcagct acgttgaccccaagacgacccgtcgtggggagaagacacgcagtgactcgttctcgttctgtgtgccc gttaaggacgacaagtgtgtgtgacgcgtacatcaatgcgttggccggtgctgtlagcgcagttgtt ccaggacttggacgccttggcggggcgcatcgctacaggcactgttcccgaaggagcccgtaagt tcaaggcagcgtggataggggtacatggtgaagaaagtggaogagattaa caattacaattcgtgtt ggccgggtccgtgacgtggaacgggagatcgtcgatggagatcacggtgaagggtatgctttgaag acgccgtgccggatataacgaacgaggagtcctgccggcagagaatgtgttttgctgctaattcacc ttctgtggcacggaaccacttacacacaagtccttgcattacaagattgttgcctgactgagaagga ctgggtcgactatcccggtcgtgagtcaccaacgcgaagaagtgatgg caagaacaagaag atcttgagcctaccgcggaagagtcgaagttgatctacga catgtggagatcgtccaagtccttacga acatcgagagggccaacgatgggacgttcggtcatgaaggacacgaccatgaagtcacctgttcatg cagccccagtaccgtaacagacactatacatgattttcggagggactgtttaagacaaacttcgaatt ggctactgtaccgcggcaacgtttccctggccgggcccgttctcagcttggactccaccacgttca agaaccccggtcccgtgggtcggtgctcaccatggactcgtcgatctcgtacacggagcagctgcac gaggagtgaggagattgacgcggactcaccgttcaacttcagcttgcctgccacgaacaagatctc gaagaaccccgaggcgttctgtcggaacccggcacgttgaltcaagtcaggctgacacatacatcc aggagttagagcagagtgtaagaagcccgcggtacgttcactactcgttctatgttgataaagaag cgttactgttgatggaaggcgtcgtttgttcagttatcccgacagcgtactccgagatgatgactatgtg

SEQ ID NO:	Description	Sequence
		ggcgggagaagaagagcccaggatacgtactacgtggagacttgcggttagtggaagcggcaattaa
SEQ ID NO:34	Acyl-CoA Hydrolase (ACHA) Amino Acid Seq	MIRTVRYQSLKRFRPSALSVPFRPRYNSQKANFHRPDHPGSDEPAEAAADAAATILAEIRDQKQTNPNKATWLDALTEREKLRAEGKTIDSFSYVDPKTTVVGEKTRSDSFSFLLLPFKDDKWLC DAYINAFGRRLVAQLFQDL DALAGRIAYRHCSPAEPVNV TASVDRVY MVKKVDEINNYNFVLAGSVTWTGRSSMEITVKGYAFEDAVPDITNEESLPAENVFLAANFTFVARNPLTHKSFAINRLLPVT EKDWVDYRRAESHNAKKKLMKNKKILEPTAESKLIYDMWRSSKSLQNI ERANDGIAFMKDTTMMKSTLFMQPQYRNRHSYMIFGGYLLRQTFELAYCTAATFSSAGPRFVSLDSTTFKNPVPVGSVLTMDSSISYTEHVHEGV EIEDADS PFNFSLPATNKISK NPEAFLSEPGLTIQVKVD TYIQELEQSVKKPAGTFIYSFYVDKESVTV DKGASFCSVIPQTYSEMPTYVGGRRRAQDTANYVETLPFSGSGN
SEQ ID NO:35	Acyl-CoA Hydrolase (ACHB) Nucleotide Seq	atgatcagaacccgtccgttatcaatcctcaagaggttcaaacctctgactttatcccccggtttccgtccacgctacaactcccagaaggccaatttccacccgtccagaccacgctgggtccgacgagccagccgaagccgccgacgcccgtccacgatcctcgccgaggttgcgagacaagcagacgaacccgaacaaggcca cctggctcgatgcgttaacggagcgggagaagtgtcggtccgagggcaagacaatcgacagcttcagctacgttgaccccaagacaacccgtcggtggggagaagacacgcagcgcgactcgttctcgttctgttgccgttcaaggacgacaagtgtgtgtgacgcgtacatcaatgcgtttggccggttgcgtgtagcgcagttgtccaggacttggacgccttggcgggcgcgcgtacaggaactgtcccccgctgagcccgtaaigtacacggcgcgagcgttgatagagtgtatgttgaagaaagtggaagcagagataataattacaatttccgtgtggcgggggtccgtgacgtggacgggagatcgtcgatggagatcacgggtcaaagggtatgctttgaagacgcgcgtccggagataactaacgaggagtcctgcgggcagagaatgtgttcttgcgtgtaattcaccctcggtgcacgttaacccactcacacaagtccttcgtatctaacagattgttgcctgacgtgagaaggactgggtcgattatcgctgctgagtcacacaacgccaagaagaagtgatggcaaaagaaacagaa gatcttggagcctaccccggaagagtcacagttgatctacgacatgtggagatcgtccagtcctacagaacatcgagaaggccaacgacgggatcgcttcgatgaaggacacgataatgaagtcacccctgttcacgcagccccagtacgtaacagacactcatacatgatttccggtgggtattgttaagacaaacttgcgaattggcctattgtaccgcagcaacgttttccctggcgggaccccggttctgcagcttgactccaccacgttcaagaaccccggtcccggtgggtcggtgtcaccatggactcgtcgatctcgtaacggagcagcgtccacgatggcgttgaggagattgacgccgactcccggttcaacttcagcttgcctg ccacgaacaagatcgcgaagaaccccgaggcgttctgtcggagcccggaacgttgatccaagtaaggtegacagtcacccag gagttagagcaaagtgtaagaagcctgcgggaacgttcactcgttctatgttgataaagagagcgctactgtggatggaaaggcgtgtttgttcagttatcccgacagcgtactccgagatgatgacttatgtgggcgggagaagaagagcccagatactgctaattacgtggagacttgcggttagtggaagcggcaattaa
SEQ ID NO:36	Acyl-CoA Hydrolase (ACHB)	MIRTVRYQSFKRFRKPLT LSPVFRPRYNSQKANFHRPDHAGSDEPAEAAADAAATILAEIRDQKQTNPNKATWLDALTEREKLRAEGKTIDSFSYVDPKTTVVGEKTRSDSFSFLLLPFKDDKWLC DAYINAFGRRLVAQLFQDL DALAGRIAYRHCSPAEPVNV TASVDRVY MVKKVDEINNYNFVLAGSVTWTGRSSMEITVKGYAFEDAVPEITNEESLPAENVFLAVNFTFVARNPLTHKSFAINRLLPVT EKDWVDYRRAESHNAKKKLMKNKKILEPTPEESKLIYDMWRSSKSLQNI EKANDGIAFMKDTIMKSTLFMQPQYRNRHSYMIFGGYLLRQTFELAYCTAATFSLAGPRFVSLDSTTFKNPVPVGSVLTMDSSISYTEHVHDGV EIEDADS PFNFSLPATNKISK NPEAFLSEPGLTIQVKVD TYIQELEQSVKKPAGTFIYSFYVDKESVTV DKGASFCSVIPQTYSEMPTYVGGRRRAQDTANYVETLPFSGSGN
SEQ ID NO:37	<i>E. coli</i> Acyl-CoA Thioesterase (TESA) gene	atggccgatacattgctcatcttgggtgactcttctgcagggatcggtatccgcactcgcgcgcatggcctgcactcctcaatgacaaatggcaagcaagacatcggtcgtgaatgcactatctcgtggcgataccctgcagcaggggttggcccgctcccagcctgttgaagcaacatcaaccacgttgggtcttggctgaattg

SEQ ID NO:	Description	Sequence
	without signal peptide sequence optimized for <i>C. tropicalis</i> Nucleotide Seq	ggcggcaatgatggtctcagaggtttcaacctcaacagaccgagcagacattgcgtcaaatctccaa gacgtgaaggcagcaaacgccgaacctctcttgatgcagataagattgcctgccaatatggtcgtaga tacaatgaagcctttctgcaatctacccgaagcttgcaaggagttgacgtcccatgttgccgttttgat ggaagagggtgacctaagcctcagtggtgcaagacgatggtatccatccgaaccgtgatgcacaac cattcatcgacagattggatggccaaacaactccaaccttggccaatcatgatagctaa
SEQ ID NO:38	<i>E. coli</i> Acyl-CoA Thioesterase (TESA) without signal peptide Amino Acid Seq	MADTLLILGDSLSAGYRMSASAAWPALLNDKWQSKTSVVNASISGDTSQ QGLARLPALLKQHQPWWVLVELGGNDGLRGFQPQQTEQTLRQILQDVKA ANAEPLLMQIRLPANYGRRYNEAFSAIYPKLAKEFDVPLLPFLMEEVYLKP QWMQDDGIHPNRDAQPFADWMAKQLQLVNHDS
SEQ ID NO:39	Acyl-CoA Synthetase (ACS1) Nuc. Seq	atggggtgcccccttaacagtcgcgttgccgaagcaaaaccaggcgaaaccgclccaagaagaaaa gccgctcaaaaaatggcctctgtcgaaacgcccaacagactcaaggcaaccatttgccagactcatt gaagagtggtttgcagaaacggcaccagagatgccatggcctggagagacttggtcgaaatccacgt cgaaaccaaacagggtaccaaaatcattgacggcgaacagaaaaaggtcgataaggactggatcta ctacgaaatgggtccttacaactacatctctaccccaagttgtgacgttggcaagaactactccaagg gtttgttgagttgggttgccccagatcaagaatccaagttgatgactttgacagaccccccacaagt ggatgcagaccttctagcctccagtttccaaggtatccccgttgacacgcctacgacacctgggtgagt cgggcttgacccactccttggtgcaaacggaatccgatgccgtgttaccgacaaccaattgtgtctcc tggatcgctccttggaagggccacctccgcaagtatgtcatccacggggaaaagaltgacctaacg acaagagacagggcggcaaaatctaccaggatgcggaaaaggccaaggagaagatttacaattta gaccagatattaaattatcttctgcagcaggttgttcattgggtgaacaatcgtccaagaattgcatttc ccaaaaccagaagaccaatctgatcatgtacacctcgggttccacgggtgctccaaagggtgtggt atcacaatgccaaactgttgccgcggtgggtggtatctcccaatgctactagagactgggtagaac tgcgacagagtgatgcatgttgcattggccacatttgcagttggccttgagttggtactctctggtgg ggggctccattgggttacgccaatgtcaagaacttgacccaagcctcctgcagaaactgtacccaagac ttgattgaatcaaaccaaccatcatggttggtgtgtgctgctggttggaatcggtcagaaagggtgtctgt ctaaattgaacaggcttctccaatccaacaaaagatctctgggtgcatcaatgccaagtctacttga accgttatggcttgccaggcgggtgggtgtgtgacgctgtctcaagaagggttaaagccgacctggtggc caatgctgtatgttgatgtggtgggtcccaatctctgttgatgcccaagtggttatctccacctgtgtgcg ccaatgtgttggttacgggttgactgaaacctgtgccaataccaccattgtcgacacacgcgcttcca gattggtacttgggtaccttggttgatctgtcactgccaagttggttgatgtgtgatgtggatactacgc caagaacaaccagggtgaatctggttgaaaggcgggtccagttgtcaaggaatactacaagaacgaa gaagaaccaaggctgcatcaccgaagatggctggttcaagactgggtgatattgtgaatggacgcg cgacgggtggttgaacatcatgacgtaagaagaactgggtcaagacttgaatggtgaatacatgtctt ggagaaattggaagatttacagatccaaccactgatttgaactgtgtgttaccgtgaccaaaccaa ggtcaagccaattgctattgtcttccaattgaagccaactgaagctatgttgaaggacgaaaagattat ccagatgctgattcaagaattgagcagcttggttcaacaagaagggtgccaagctgtcttgaga cactgtctcaaacgggtaacaacaagggttgaaggatgaattgtgcagaatgtgtctgttgatg acgagtggaacccacagaatggtttgtactctgccccaaagtgcagagaaagaagatttagaaag ttgtaaaaaagaagttgaagggcatacaagtcgtcttag
SEQ ID NO:40	Acyl-CoA Synthetase (ACS1) A. A. Seq	MGAPLTVAVGEAKPGETAPRRKAAQKMASVERPTDSKATTLPDFIEECFA RNGTRDAMAWRDLVEIHVETKQVTKIIDGEQKKVDKDWIYYEMGPYNYIS YPKLLTLVKNYSKGLLELGLAPDQESKLMIFASTSHKWMQTFCLASSFQGI VVTAYDTLGEGLTHSLVQTESDAVFTDNQLLSSLRPLEKATSVKYVIH EKIDPNDKRQGGKIYQDAEKAKEKILQIRPDIKFISFDEVVALGEQSSKELH FPKPEDPICIMYTSGSTGAPKGVVITNANIVAAGGISTNATRDLVRTDRV IAFLPLAHIFELAFELVTFWWGAPLGYANVKLTLEASCRNCQPDLEFKPTI MVGVAWVWESVRKGVLSKLKQASPIQQKIFWAAFNAKSTLNRYGLPGGG LFDVAFKKVKAATGGQLRYVLNNGSPISVDAQVFISTLLAPMLLGYGLTET CANTTIVEHTRFQIGTLGLTVGSVTAKLVDVADAGYYAKNNQGEIWLKGG

SEQ ID NO:	Description	Sequence
		PVVKEYYKNEEETKAAFTEDGWFKTGDIGEWADGGLNIIDRKKNLVKT NGEYIALEKLESIYRSNHLILNLCVYADQTKVKPIAIVLPIEANLKSMLKDEKII PDADSQLSSLVHNKKVAQAVLRHLLQTGKQQGLKGIELLQNVVLLDDEW TPQNGFVTSQAKLQRKKILESCKKEVEEAYKSS
SEQ ID NO:41	Long-chain Acyl-CoA Synthetase (FAT1) Nuc. Seq	atgtcaggattagaaatagccgctgctgcatccttggtagtcagttattggaagccaaatatttaattgcc gacgacgtgctgttagccaagacagtcgctgtcaatgccctccatactgtggaaagccagcagaggt aaggcatcactgtgtacttttcgagcagtcggttcaagaacccaacaacaaagcgttggcgttcc caagaccaagaaagaatgccccaccccccaagaccgacgacgaggttccagatctacgacgat cagttgacctagaagaatacacctacaaggaattgtacgacatggtttgaagtactatacatcttgaa gaacgagtacggcgtcactgccaacgacacatcggtgttctgtatgaacaaagccgttttcattgtctt gtggttggcattgtggaacattgtgtcctgctgctgtcttgaaattcaacaccaaaggacaagccattgat ccactgtcttaagattgtcaaagccttcgaagtttcgttgaccggagctgtgattccccaatcagagatacc gaggctcagatcagagagggaattgccacatgtgcaaataaactacattgacgagttgctgtttgaca gattgagactcaagtcgactccaaaacacagagccgaggacaagaccagaacccaacccgatact gactcctccgctgtgcatgtattacacctcggtaccacgggttgcacaaagccggtatcatgtcctgg agaaaagcctcatggcctcggttttcttgccacatcatgaagattgactcgaaatcgaacgtctgacc gcatgcccctgtaccactccacggcggtcattgtgggtgtgtctactttgattgtcggtggctgtgtctc cgtgtccagaaaattctcgtactctgtctgacccagccagattatgtggtgacccacgtgcaat acgtcggtgaggtctgtctgtactgttgaaactcaagcctcatccagaccaagacagacacaatgtcag aattgcctacggtaacgggtgtcgtccagatatatgtgtcgtgagttcaagcgcagattccacattgaaggta tcggtgagttctacgcccacccagagtcctctatcgccaccaccaactgacgtacggtgagtagcggtg cggcgctgtcgaagtagcgggtccctcatcagctgttattgtctaccagcagaaattggccaagatgg accagaagacgagagtgaaatctacaaggaccccaagacgggttctgtaccgagggcgttaca acgagccaggtgagttgtgatgagaattctgaaccctaacgacgtgcagaaatcctccaggggtattat ggtaacaagtcgcccacacagcaaaatcctcacaatgttttcaaaaagggtgacgcgtgtgtacag atccgggtgactgttgaagatggacgaggacaaattgtgtactttgtcgacagattaggtgacacttccgt tggaagtccgaaaacgtctccgccaacgaggtcgagaacgaattgatgggtccaaaggccttgaaagc agtccgtcgtgtgtcgtgtcaaggtgccaacacgaaggtagagcctgtttgcccgtctgtgaagccaa ggacgagttgagccatgaagaaatctgaattgattcactctcagtgaccaagtcctgtgtgtatgc tcaacctgctgtcatcaagattggcaccattgaggcttcgcacacacacagggttctaagaaacaaattc aagaacaaaagttgccaagggtgaagacggcaaggattgatctactggttgaaatggcgacaagt accaggagttgactgaagacgattggtcttattgtaccgtaagccaaattg
SEQ ID NO:42	Long-chain Acyl-CoA Synthetase (FAT1) A.A. Seq	MSGLEIAAAAILGSQLLEAKYLIADDVSLAKTVAVNALPYLWKASRGKASY WYFFEQSVFKNPNNKALAFPRPRKNAPTPKTDAGFQIYDDQFDLEEYTY KELYDMVLKYSYILKNEYGVTANDTIGVSCMNKPLFIVLWLALWNIGALPA FLNFNTKDKPLIHCLKIVNASQVFVDPDCDSPIRDTEAQIREELPHVQINYID EFALFDRLRLKSTPKHRAEDKTRRPTDSDSSACALIYTSGTTGLPKAGIMS WRKAFMASVFFGHIMKIDSKSNVLTAMPLYHSTAAMLGLCPTLIVGGCVS VSQKFSATSFWTQARLCGATHVQYVGEVCRYLLNSKPHPDQDRHNVRIA YGNGLRPDIWSEFKRRFHIEGIGEFYAATESPIATTNLQYGEYGVGACRKY GSLISLLLSTQQKLAKMDPEDESEIYKDPKTGFCTEAAAYNEPGELLMRILN PNDVQKSFQGYGKNSATNSKILTNVFKKGDWYRSGDLLKMDNEKLLY FVDRLGDTFRWKSENVSAEVENELMGSKALKQSVVGVKVPNEHGRA CFAYCEAKDELSSHEILKLIHSHVTKSLPVYAQPAFIKIGTIEASHNHKVPKN QFKNQKLPGGEDGKDLIYWLNGDKYQELTEDDWSLICTGKAKL
SEQ ID NO:43	Acyl-CoA Sterol acyl transferase (ARE1) Nuc. Seq	atgtccgacgacgagatagcaggaatagtcattgaatcgacgatgacgtgaaatccacgtcttcttcc aggaagaactagtcgaggttgaaatgtccaactcgtccattacgaatccacgacccgatgagtcgtacc gtcctgaagaaacctcattgcattacaggaggaagtccacaggaacccgacagaggagtcgttcc gagatcaccaagaacgtgaatgatccgatctagttccaagattgagaacctaaaggggcaaagtaag ccaacgggaagacaggttgagggaagcactaccttcacacctccaggacgtcaagttctgtccgggt caacgacatcaagtcaagctgaactccgcgaagattctagattcggtgctttcaagagtgaaat

SEQ ID NO:	Description	Sequence
		itggagtcctgaccatctctgggtgggtatcgcactctacatattgtcaacgtgtcagatgttactttggcat ggccaagcccttacggactggatcatcataggaatgtcaagcaggactgggtgaaagtgcactcgtt gatcttgccatgtacctatccctgattttccttattcttgagggtgcatgcaaacggggtgatgtatctggc atggtcttgatgggcaatacagggggttacagcttggtgttttgacgttctggacggtagttccgcagga gttggccaatggatctctcgttgattgcaagaattttctgactctgcatgctgggtgttattatgaagatgcagt cgtatgggcattacaatggataccttgggatgtgatcaggaaggatggcctctgaggctgatctcagg gacctttctgagtatgatgaagatttcccttgatcacgtggagggtctagaacagagctgtgtgttcc aaacacgagttggagttcaatcgaatggaactgctgagagggaaggaccacatcacatgtattcga cgaaaaggatgtcaacaaaccaatacgtgtcttgcaagaaggagggaattatcaagttccggcaaaaca tcaactcaaggattttcgagtacagatgttcccaacgctagtctacacgttgagcttccccgaactcg acagattagatggacgtatgtgtgcagaagggttggaacatttgccttagtgtttgcatgattatcgctg ccgaagagagttctgcccctgtgatgcaagaagtgtatcagtcacacaaatggcaaccaaccaagggt tcccaaaatactctgctgttcttccactgtatattaccgctcggcaagcagtactgtctcattcatctca tctggaatgaaattcacaacggcatagcggagttagcaggtttggcgaccggcattttacggcgcttg gtggtcgagcgtcgattacatggactattcaagaaaatggaacaccatcgtgcaccgatctcctcgctgg caggtttacaattcgagcattcacatcctcgttattccaggacgcgaagcgcgtagltacatttgccttc tgcacaatccacgaactcgttatgtacgtcctatttggcaaatcagagggtacctattcctacgatgctt gtccagatcccatgacggtcacctccaagttcaacaaccgtgttggggcaacatcatgttctgttgac gtatttatctggccccagctgtgtgtgctgtatttactctctag
SEQ ID NO:44	Acyl-CoA Sterol acyl transferase (ARE1) A.A. Seq	MSDDEIAGIVIEIDDDVKSTSSSQEELVEVEMSNSSINESQTDSEYRPEETS LHYRRKSHRTPSEESFLEITKNVNDPDLVSKIENLRGKVSQREDRLRKHYL HTSQDVKFLSRFNDIKFKLNSATILDSDAFYKSEYFVLTIFVWVIALYILST LSDVYFGMAKPLLDWIIIGMFQDLVKVALVDLAMYLSSYFPYFLQVACKR GDVSWHGLGWAIQGVYSLVFLTFWTVVPQELAMDLPWIARIFLILHCLVFI MKMQSYGHHYNGYLWDVYQEGLADEADLRDLSEYDEDFPLDHVEVLEQS LWFAKHELEFQSNGTAEKRDHHHHVFDEKDVNKPVRVLQEEGIKFPANIN FKDYFEYSMFPTLVYTLSPRTRQIRWTYVLQKVLGTFALVFAMIIVAEESF CPLMQEVDQYTKLPTNQRFPKYFVLSHLILPLGKQYLLSFILIWNEILNGIA ELSRFGDRHFYGAWWSSVDYMDYSRKWNTIVHRFLRRHVYNSSIHLGIS RTQAAIVTLLLSATIHVLMYVLFGLRGYLFMTMLVQIPMTVTSKFNNRWV GNIMFWLTYLSGPSLVSALYLLF
SEQ ID NO:45	Acyl-CoA Sterol acyl transferase (ARE2) Nuc. Seq	atgtccgacgacgagatagcagggaatagtcattgaaatcgacgatgacgtgaaatctcgtctcgttcc aggaagacctagtcgaggttgagatgtccaactcgtccattaacgaatccagacggatgagttgctgt accgtcctgaagaaatctcattgcatcgagaaggaagtccacaaagacccgtcagatgagtcgttcc tagagatcaccaagaacgtgaatgacggatctagtctccaagattgagaactaaagggcgaagta agccaacgggaagacagggtgaggaaactactcctccacacatccaggacgtcaagttctgtctcg gttcaacgacatcaagtcaagctgaactcgcgcgacgattctagattcggatgcgtttacaagagcgag caacttggagcttgactatctctggtgttatcggaactctacataatgcaacgtgtgcagacatgtatttg gcatggccaagccctactggaactgataatcatagggaatgtcaagaaggatttgatgcaagttgact cgttgatctgtcatgtactatcctcgtattttccttatttctacagggtgcatgcaagacgggagctatctt ggcatggtcttgatgggcatagggggttacagcttggtgttttaactttctggcggtactcctcgtg gagctggccatggatctccttgattgcacgagttttctgacttgcatgcttggtgtttatgaagatgc aatcatatggacattacaatggatacctttgggagtatatcaggaaggattggtctcggaagctgactca cggctgtttcagatgatgatgatttccctggatcacggggagggtctagaaacagagctgtggtgtg ccaaacacgagttggagtttcaatcaatggaactacggagaggaaggatcacatcatcatgtattcg acgaaaaggatgtcaacaaaccaatgcgtgtctgcaagaagagggaattatcaaatccggcaaac atcaattcaaggattattcgagtacagtatgtcccacgctagtctacacattgaacttcccagaattc gacatattagatggcgatgtgttgcaaaaatttgggaacattgccttagtggttccatgattatcgtc gccgaagagagtttctgctccttgatgcaagaagtgaacagtacacaagatggcaaccaaccaag gttctcaaagtactcgtcgttcttccactgtatattgccctcggcaaacagactgtcctcgtttatcctc atttggaaacgaaattcacaacgggatagcggagttaagcagggttggggatcgcaatttctacggcgct

SEQ ID NO:	Description	Sequence
		ggtgggtcaagcgctgactacatggactatcaagaaaatggaacacgatcgctgcaccgattcctccgcc ggcacggttacattcgaccattcgcatcctcggtatttccaggaccaagccgcgataattacatttgc itcagccacaatccacgaactcggtatgtacatcctatttgaaaattacgagggtagctattcctacgatg ctgtccagatcccatgacagtcacgcgccaagttcaa caaccggttggtggggaacatcatgttctgggtg acgtatttatctggcccagcttggttagtgcgttgattactctctga
SEQ ID NO:46	Acyl-CoA Sterol acyl transferase (ARE2) A.A. Seq	MSDDEIAGIVIEIDDDVKSTSSFQEDLVEVEMSNSSINESQTDELSYRPEEI SLHSRRKSHKTPSDESFLAITKNVNDPDLVSKIENLRGKVSQREDRLRKH YLHTSQDVKFLSRFNDIKFKSNSATILDSDAFYKSEHFGVLTIFWV/VIGLYI MSTLSDMYFGMAKPLSDWIIIGMFKKDLMQVALVDLVMYLSSYFPYFLQV ACKTGAI SWHGLGWA IQGVYSLVFLTFWAVLPSESAMDLPWIARVFLILHC LVFIMKMQSYGHYNGYLWDVYQEGLVSEADLTAVSEYDDDFPSDHGEVL EQSLWFAKHELEFQSNGTTERKDHHHHVFDEKDVNKPMPVLQEEGIKFP ANINFKDYFEYSMFPTLVYTLNFPRI RIRHWAYVLQKVLGTFALVFAMIIA EESFCPLMQEVEQYTRLPNTQRF SKYFVVL SHLILPLGQYLLSFLI/WNEI LNGIAELSRFGDRHFYGAWWSSVDYMDYSRKWNTIVHRFLRRHVYNSTI RILGISRTQAAIITLLSATIHELVMYILFGKLRGYLFLTMLVQIPMTVTAKFN NRLWGNIMFWLTYLSGPSLSVALYLLF
SEQ ID NO:47	Diacylglycerol acyltransferase (DGA1) Nuc. Seq	atgactcaggactataaagacgatagtcctacgtccactgagttggacactaacatagaagaggtgga aagcactgcaaccctagagtcggaactcagacagagaaaacagaccaggaactccagcatcaa ccccaccacacccccaacaacacagcaggcgcataagaaagccctgaagaatggcaagaggaa gagaccattataaacgtggcgccgctcaacaccccgctggctcagaggctcgagactttggcgtgtgtt ggcactgtgtcagatcccgctcttattgtttgttcttgcttacggtctccatgggggtgtctgggtgttcttatc attttgccatattcattgtgtgtacggttogaattgcacactccatcgaatgtaagttgtctatcgtgtgc gcaactcgttcaagaatttcacatttgggactggtttgtcaagtatttcccgaltggaagacaaagcaggt cgagttggatcctacttttagcgaattgcctgtggaagagagcggcgacagttcggaacgacgacgaac aagacttggtgtctgagcaacagcagaaattgtgtgatcaaatctcaagttttcggttgaagaaacgctt gaatgacacctccctgggcaaaccagagacattcaagaatgtgcctacgggtccaaggataattttggg taccaccacacgagtgatttctatgggggagtggtgtttgccaacaacgccttgaggaaacgaa ccatatacgccaatttccaaatggttaaaacattcttcacgacagctccaagggcgagagattgtccc tggtattggcaatatctccattgacgctaccacacagtttgcgtccattttaccgtgactactgatggc ttggggatcactagtcacgcgctaaacattagaagcttgatcaacaatggagacaaactctgtgtgc tcgtcgttggcgggtgcacaagaaatcgttgtgaacaatatgattgccaagcagccagagtcgggtacg gttacaagagagcctagatattcatgtgcgaccagtcgaagaagaagaagaagaagagatgata ccaagcagctagagaacccaagtcctaaacgtgaagtgcaattggtcttgaacaaacgtaaaggtttg tgaagttggctatcgaaactagaaatgttctctgtgctcatttttcattcgagagaagctgatgttacagat tggcccagccagcaccaggtcgttctgtacaagttccagcaatggatgaaggcaacttttcaattcacc atccattgttttagtgcctcagggcgtgttcaatcatgattcggattgttgcattcagaacccaataaacatt tgcgtcggtagaccgctacattccgcacaaacgtcttgaagaatacaagcaaaagcaaccagagg agtttgcgaagaggaaacctgccaglaccccgatgaagaagcttgatcttcaccgatatttcaaagc tgggtgaaaagaagccaagactcaagatcaagactaaaatccacctgcattactagacaagtagc acaagctatacgtcgacaggtgaagaaggtctatgaagaagaacaggaaaggttggctacggtgat gttgaattaaacattgtagaatg
SEQ ID NO:48	Diacylglycerol acyltransferase (DGA1) A.A. Seq	MTQDYKDDSPSTELDTNIEEVESTATLESELQRKQTTETPASTPPPPP QQQQAHHKKASKNGKRKRPFINVAPLNTPLAHRLETLAVVWHCVSIPFFMF LFLTVMGMLLGWFFIILPYFIWWYGFDLHTPSNGKVYVRVNSFKNFIIW DWFKYFPIEVHKTVELDPTFSELPVEESGDSSDDDEQDLVSEHSRTLVD QIFKFFGLKKRLNDTSSGKPETFKNVPTGPRYIFGYHHPGVISMGA/VGLFA NNALRNEPYTPISKWLKPPFFHDSKGERLFPGIGNIFPLTLTTQFALPFYRD YLMALGITSASAKNIRSLINNGDNSVCLVVGGAQESLLNNMIAKHARVGYG YKESLDIHGDQSEEEEEEDDTKQLENPSPKREVQLVLNKRKGFKVLAIEL GNVSLVPIFAFGEADVRLAQAPAGSFLYKFQQWMKATFQFTIPLFSARG

SEQ ID NO:	Description	Sequence
		VFIYDFGLLPFRNPINICVGRPVYIPHNVLQEYKQKHPEEFAEEEPASTPM KKSQSFTDMFKAGEKKPKTSSIKTKIPPALLDKYHKLYVDELKKVYEENKE RFGYGDVELNIVE
SEQ ID NO:49	Diacylglycerol acyltransferase (LRO1) Nuc. Seq	atgtcgtctttaagaacagaaaatccgcaagcgtcgccaagcgatacagaagactcagaaaacag aggcagatctcctcaattgatcccaacggcaccataitgacgaccagtctacatgacgaacccccacc acagccatcaccaccacaacataactagaccagattggaggacgatggcagcatcctgggtccag aagatcgtcgatctccaaatccgacgacctgcaggcaaaagaaaagaagaagaacccaagaaga agatctggagtcgtcggtcatgtttatcttggtagccctcaitgggttaatcttgcgtgggcgtttaccac agacacgcatccttcaatggcgactggagaagtatatcaacttgaccagctcaacgggatcttgaag actggaagaactggaaggatatctgccaacagcatccagacgtacttgcaggaatcgggcaaggg cgaagataacgacgggtgcatgtgtggcggattcctctcgtcggtcctcgctgaagccagaa gaactcactgacaaccacaatgtcgtgtgttctcgtgtgtgagcaggggttgaatcgtgggga acaaccaccacgggtgattgccatctatcggaacttcaggaagagattgtggggaatctttatgtta aggacaatgatttggagaaaaagctgtgtgtgaagcatatccagttggacgagaagacgggttgat cctcccaatattaagggtccgtcggtcgaggggttcgaagcggcagatttcttatggcgtggatcggatc tggaaacagatcttgcagaactggcggttatgggtacggaccaaataacatggtgagtgctagtatg actggagattggcttacttgacttgagagaagagatggatatttgcgaactaaagcgagattgag tgaataacaagtgaacaacaagaagactgtgtgattggccaatcgatgggacccagattatttcta cttttgaatgggtcgaagccacggggaaccatactatggcaatggcggaccaaactgggtgaatg atcatatgagtcgattatgacatcagttgggtcgacttgggtaccccaagagtattcgtgtgtatctc ggggaatgaaagacacccgttaattgaacgcgttgccgtttacgggttgagcaattttcagcaggc gtgaaagagtcgatattgctgacattgggtggtgagcagtatgacccaaggggggagacaagat atggggcaacttgacgcatgcgcagatgatccaatttccacattcagtgatgacgaagtacggacag ccacgaacctaaagatcgttctttgtacgtttatccaattcaagaaccaaactagcgaagcgaagccat acagggagatcaccatggctgaaggtatcgatgaattgttgacaaatcaccagactggtattccaaga gagtcgtgagaactactctacggcattacagacagcaaggcgcaattagagaagaacaacaatga ccacctgaagtgtggaacccattagaagctgccttgccctaaagcaccgacatgaagatctattgttct acggagttggaaatcctaccgaaagggcatacaagtatgtgactgacgataaaaaagccacgaat ggactacataatagacgcgacgatgccaatggagtcataataggagacggagacggcactgttctgt attaaccacatcgatgtgccatgagtggtggccaaggagacaagtcgagatacaaccagccaactcg aaggttaccattgtgaaatcaagcagagccagacagattgattacgaggcgccgaagactgc ggaacatgtgatatttggggagtgccgagttgaacgagttgatttgaatgtgttagcggaacgggg acgagattgagaatagatatgtcagcaactaaaagaaatagtagagccataaattataa
SEQ ID NO:50	Diacylglycerol acyltransferase (LRO1) A.A. Seq	MSSLKNRKSASVATSDTEDSETEAVSSSIDPNGTILRPVLHDEPHSHHH HNITRPVLEDDGSISVSRRSSISKSDDSQAKQKKKKPKKKILESRRVMFIG TLIGLIFAWAFTTDTHPFNGDLEKFINFDQLNGIFDDWKNWKDILPNSIQTY LQESGKGEDNDGLHGSADSFVGLRLKAQKNFTDNHNVVLPVGVVSTGL ESWGTTTTGDGPSIGYFRKRLWGSFYMLRTMILEKTCWLKHIQLDEKTGL DPPNIKVRAAQGFEEADFFMAGYWIWNKILQNLAVIGYPNNMVSASYD WRLAYIDLERRDGYFSKLKAQIELNNKLNKKTVLIGHSMGTQIIFYFLKWV EATGKPYYGNGGPNWVNDHIESIIDISGSTLGTPKSIPVLISGEMKDTVQLN ALAVYGLEQFFSRRERVDMLRTFGGVASMLPKGGDKIWGNLTHAPDDPI STFSDDEVTDSEHPKDRSFGTFIQFKNQTSAPYREITMAEGIDELDKS PDWYSKRVENYSYGITDSKAQLEKNNNDHSHKWSNPLEALPKAPDMKI YCFYGVGNPTERAYKYVTADKKATKLDYIIDADDANGVILGDGDGTVSLLT HSMCHWAKGDKSRYPANSKVITVEIKHEPDRFDLRGGAKTAEHVDILG SAELNELILTVSGNGDEIENRYVSNLKEIVEAINL

SEQ ID NO:	Description	Sequence
SEQ ID NO: 3674	ECI1, ATCC20336 (Nucleic Acid Seq.)	atgtccgacgaggaatcagatatcttatacagaggtcagagacagaaccgccatcatcacctgaacatc cccaagagattgaacgcattgaacgggtgctcaatacttgaagttgggtaagttctggagagagccaaca acgaagaggacaccgtcttgacattgatccaggccctgggcagattcttccgccgggtccaattcgcc gacaacgatatggccaaggctgaatgtccaagttgtcagtcacgagtactggttgaaagattcgctg ccagaaacatctggtgaccaactgttcaacgaccaagaagatctggctgctgctgcaatggtcca gttatcggttgagcactggttgtgtgtgtgctgattggtctacgtccacgactgaacaagtctacctctg gccccattgccaactgggttgggtgccgaagggtctctctgccacttgttcaacagattgggtggtca aaggctctgagggcctgtgtgtggccaagccaatcgccggccaagactgttacaacgccgggttcatcaa caagcactacgacgtaagtttctccactgaagagttcaacgaacacgtctacaaggagttgaacgga agctttgaaaactgcacgatgactccatttgcagaacaagcaattgtgaagttgccagagaccaggc catcaactag
SEQ ID NO: 3675	ECI1, ATCC20336 (Amino Acid seq.)	MSDEESDILYEVRDRTAIIITLNIPKRLNALNGAQYKLKGLFLERANNEEDTVL TLIQASGRFFSAGANFADNDMAKVEMSKLFSHEYWLERFVARNIWLTNLF NDHKKILAAAVNGPVI GLSTGLLLVLDLVYVHDLNKFYLLAPFANLGLVAEG ASSATL FNRLGWSKASEALLAKPIGGQDCYNAGFINKHYDGKFSSTEEFN EHVYKELTEAFENLHDDSILQNKQLLKL SRDQAIN*
SEQ ID NO: 3676	ECI2, ATCC20336 (Nucleic Acid Seq.)	atgtccgacgaccttatcacctacgaagtcgaagaccgagctgccgtgatcacctgaacaaacccaag aagctcaacgccttatcgatcccgagtcagacaccatctgcaagctcttagaacgagccaacgcgga agaagacaccgtcatcacctgtctccagtccacggcgccgggtgtctctgcccgggccaacgcgactc catcggtgggcaggatgcgagctcgagacctggtgaacatgctgggtggccaagcagacgttctgtg cagacgttctcgcacacaagaagatcctgcccgtcgcttgaacggccccgtgatggcttatcgccg cggtcgtggcgctctcgactggtctacgtgcacaacgcgcgcaaagacgttctctgaccccgctcgcca acatcgggatccttgcgaggggcgccactcagccacgttgcctatgcgctgggggtgctcaggggcc gcggaagcgltgtgtgtgcaaaagaggttccgggagatgactgcagagagcggggtcttcaataagg actacaaggggcagttcaagtcgaggaggttaacgaggtcgtctgaaggagttgacgcccac ggaaaaattgcatgaggactcgatcatccagaacaaggagttgtgaaggctatttcaagccaagatc agtgaggtcaactcgaggaggtgtcaagaggtgtglaaagtgacaccttgggggtccaatggataga ttaaaaaattgctaagtggtgagttgaacataaattatag
SEQ ID NO: 3677	ECI2, ATCC20336 (Amino Acid seq.)	MSDDLITYEVKDRAAVITLNNPKKLNALSI PQYDTICKLLERANA EEDTVITLL QSTGRVFSAGANADSVGQDAELETWLNMSVAKQTFVLVQTF LAHKILAVA LNGPVI GLSAA FVALCDLVYVHNAAKTFFLTPFANIGILAEGGTSATLPMRV GWSRAAEALLLSKRISGDDLQRAGFFNKDYKGQFKSAEEFNEVV LKELL ATENLHEDSIIQNKELLKAIKPKISEVNSQEVSRGVYKWTSGVPMDRFKKL LNGELKHKL*
SEQ ID NO: 3678	>gi 50550800 refl XM_502873.1 Yarrowia lipolytica YALI0D15708p (YALI0D15708g) mRNA, complete cds(similar to uniprot P45954 Homo sapiens Acyl-CoA dehydrogenase short/branched chain specific mitochondrial precursor)	ATGTTGTCCATTGATCCATTACCCGATCTCTCCCCATTGGCAGCCGAA TCTGCCAGCAGAGTGCCATGAAGGCCTCTACTGTGCGACCTCTCGCCT TGAGAGCTTACTCCACCCGACCCCTGTCACTCACTTCTCCGAGGAGG AGGAGATGTTTCGTGACATGGTTAGCAAGTTTGCTGATGAGGTGATTGC TCCCAAGGTCCGTGAGATGGACGAGGCCGAGCAGATGGACAAGACAA TCATCCAGGACATGTTGACAATGGCCTTATGGGCATCGAGACTCCCG AGGAGTTCCGTTGGTGCAGGTGCCAACTTCACCTCTGCTATCATCGTGC TTGAGGAGCTTGCCAAAGGTGGACCCCTCAGTGTCTGTGATGAACGATG TCCACAATACCCTCGTCAACACCTGCATCCGATCCTGGGGATCCGACG CACTCCGAAACAAGTATCTCCCCAGCTTGCTGCCCAGAAGGTGGGAT CTTTCGCTCTTTCTGAGCCCTCTCCGGATCTGATGCCTTCGCCATGAA GTCTCGAGCCACAAAGACTGACGATGGATACATTTTGAACGGTTCCAA GATGTGGATCACCAACGCTGCCGAGGCTGAGCTTTTCATTGTTTTGCT AATCTCGATCCCAGCAAGGGCTACAAGGGTATTACTGCCTTTGTTGTCG AGAAGGACATGGGAGTGCAGATTGCTAAGAAGGAGCAGAGAGCTGGGT ATCCGAGCCTCTTCTACCTGCGTTCTCAACTTCGAGGACGTTTTTCATTC

SEQ ID NO:	Description	Sequence
		CTAAGGAGAACCTTCTTGGCGAGGAGGGCAAGGGCTACAAGATTGCTATCGAGTGCTTGAACGAGGGCCGAATCGGAATTGCGGCCAGATGCTTGGCCTTGCTGGTGGAGCTTTCAAGAAGGCTACCGGCTATGCTTTCAACGACAGAAAAGCAGTTCGGCCAGTACATCGGTGAGTTCAGGGTATGCAGCACCAGATTGGCCAGGCCGCTACTGAGATCGAGGCTGCTCGACTCCTGTCTACAACGCTGCCCCGACTCAAGGAGGCTGGCGTTCCTTTCACAAAGGAGGCTGCTATGGCAAAGCTCTATGCTTCCCAGGTTGCAGGAAACGTCGCATCCAAGGCTGTGCAATGGATGGGTGGTGTGCGGATTCACGAGAGGAGACTCTGGAGAAGTTCCTCCGAGATTCTAAGATCGGTGCCATTTACGAGGGAACTTCCAACATCCAGCTGCAGACTATTGCCAAGATCATCCAGAAAGGAGTCTGCCTAA
SEQ ID NO: 3679	>gi 49648741 emb CAG81061.1 YALI0D15708p [Yarrowia lipolytica CLIB122](similar to uniprot P45954 Homo sapiens Acyl-CoA dehydrogenase short/branched chain specific mitochondrial precursor)	MLSIRSITRSLPIGSRICQQSAMKASTVRPLALRAYSTRPPVTHFSEEEEMF RDMVSKFADEVIAPKVREMDAEQMDKTIQDMFDNGLMGIETPEEFGGA GANFTSAIIVVEELAKVDPSVSMNDVHNTLVNTCIRSWGSDALRNKYL PQ LAAQKVGSFALSEPSSGSDAFAMKSRATKTDG YILNGSKMWITNAEAE LFIVFANLDPSKGYKGITAFVVEKDMGVQIAKKEQKLGRASSTCandidaLNF EDVFIPKENLLGEEGKGYKIAIECLNEGRIGIAAQM LGLAGGAFKKATGYAF NDRKQFGQYIGEFQGMQHQIGQAATEIEARLLVYNAARLKEAGVPFTKE AAMAKLYASQVAGNVASKAVEWMGGVGFTREETLEKFFRDSKIGAIYEGT SNIQLQTIAKIIQKESA
SEQ ID NO: 3680	>gi 210075528 ref XM_501919.2 Yarrowia lipolytica YALI0C16797p (YALI0C16797g) mRNA, complete cds (similar to uniprot Q96VP9 Glomus intraradices Probable acyl-CoA dehydrogenase)	ATGAGCGAGCAGTACACCCCGAACAAGTTGCGGAGCACAACTCTCCC GAATCTCTGTGGATCATCATTGACGGTAACGTTTTTCGACCTCACTGAAT TCCAGAAAGAACACCCCGCGGAAAAAAGATTCTCAAACGAGTCGCAG GAAAAGACGCTACCAAGTTTTTCCACAAATACCACGACGCCCCCAAGAT TATGCGAAAGGTTGGACACAAGTTCAAGATCGGAACCCCTTAAAGACGC TGAAGCAAACCCCACTCGAGCCATGATTGCCCTAACAAGACCACCGC CCTCGAGCCCTACGGAGACCTTGTCCTTACGCCGACCCCAACTGGTA CCACGGCTACCACAACCCCTACTACAAGGAGTCCCACGCCAAGCTGCG TGACGAGGTCCGACAGTGGGTTGAGGAGAAGATTGAGCCCTTCGTTGA GGAATGGGATGAGGAGAAGGAGGTTCCCAAGGAGATCTTCCAGGAGA TGGGCAAGCGAGGTTACCTTGCCGGCTCTCTCGGCACCCCTACAAG GAGCTGGCCAAGTACACCAACGTCAAGCCCCGCTCTGTGCCCATTGAG GAGTACGACATGTTCCACGAGCTCATCATCACCAGCAGATCATGCGA GCTGGCTCCGGAGGTCTCACCTGGAACCTGCTTGGTGGCTACTGTATT GGTCTGCCTCCCGTGATCAAGTTCGCCAAGGAGCCCTTAAGGAGCGA ATCCTCCCCGGCCTGCTCGACGGTTCCAAGCGAATCTGTCTGTGTATC ACTGAGCCCCGACGCTGGCTCCGATGTTGCCAACATCACCACACTACCGC GAGAAGACCCCCGACGGAAGTTCTACATTGTCAACGGTATCAAGAAG TGGATCACCAACGGTATCTGGGCTGACTACTTCACTGTTGCCGTCCGA ACCGGTGGCCCCGGCTCTGGCATGAACGGTATCTCTGTTCTGCTGCTC GAGCGAGGCATGGAGGGTCTTGAGACCCGACGAATGAACACTCAGGG TATGCTGTCTTCCGGCTCTACCTGGGTCAACATGGAGGATGTCAAGGT CCCCCTGGAGAACCTGCTCGGCAAGGAGAACAAGGGTTTCAAGGTCAT CATGACCAACTTCAACCACGAGCGAGTTGGTATCATCATCCAGGCCAAA CCGAGCTTCTCGAGTTTGCTACGAGGAGGCCTGCAAGTACGCCACAA

SEQ ID NO:	Description	Sequence
		GCGAAAGACTTTTCGGCAAGCCTCTGATTGAGCACCCCGTCATCCGAGC CAAGCTCGCCAACATGGCCATTGGAATCGAGTCCACCCACGCCTGGCT CGAGAACCTGGTCTTCCAGTGCCAGATGTTCCCCGAGGAGGAGGCCAT GCTTCGACTTGGTGGTGCCATTGCTGGTTGCAAGGCCACGCCACCCA GACCCTCGAGCTGTGTGCCCGAGAGGCTTCCAGATCTTTGGTGGTCT TTCCTACACCCGAGGCGGTCTCGGAGGTAAGGTTGAGCGACTGTACCG AGAGGTCCGAGCCTACGCCATCCCCGGTGGATCCGAGGAGATTATGCT GGATCTGGCCATGCGACAGGCCCTCAAGGTCCACAAGGCTGTTGGCG CCAAGCTTTAA
SEQ ID NO: 3681	>gi 199425292 emb CAG82239.2 YALIO16797p [Yarrowia lipolytica CLIB122](similar to uniprot Q96VP9 Glomus intraradices Probable acyl-CoA dehydrogenase)	MSEQYTPEQVAEHNPSLESLWIIIDGNVFDLTFQKEHPGGKKILKRVAGKD ATKFFHKYHDAPKIMRKVGHKFKIGTLKDAEANPTRAMIAPNKTALPEY DLVPYADPNWYHGYHNPYYKESHAKLRDEVQRWVEEKIEPFVEEWDEEK EVPKEIFQEMKRGYLAGSLGTPYKELAKYTNVKPASVPIEYDMFHELIIT DEIMRAGSGGLTWNLLGGYCIPLPPVIKFAKEPLKERILPGLLDGSKRICI TEPDAGSDVANITTTAEKTPDGKFYIVNGIKKWITNGIWADYFTVAVRTGGP GSGMNGISVLLLRGMEGLETRRMNTQGMSSGSTWVTMEDVKVPVENL LGKENKGFVIMTNFNHERVGIIQANRASRVCEEACKYAHKRKTFGKPLI EHPVIRAKLANMAIRIESTHAWLENLVFQCMFPEEEAMRLRGLGAAGCKA QATQTLCLCAREASQIFGGLSYTRGGLGGKVERLYREVRAIIPGGSEEIM LDLAMRQALKVHKAVGAKL
SEQ ID NO: 3682	>gi 50556785 ref XM_505801.1 Yarrowia lipolytica YALIOF23749p (YALIOF23749g) mRNA, complete cds (highly similar to uniprot Q7S579 Neurospora crassa NCU02291.1 hypothetical protein probable Glutaryl-CoA dehydrogenase)	ATGCTTACCAGAATCTCCCGTTTGGCACCTGCTGCCCCGAGGCTTTGCT ACCTCCTCCGTCAACCGATCCACAGCCGCCATGGACTGGCAGGATCCC TTCCAGCTGGACTCTCTTCTCACCGAGGACGAGATTGCCGTGGCTGAG GCTGCTCGAGACTTCTGCCAGACAGAGCTCTACCCCAAGGTACTTGAG GGCTACCGAACCGAGGAGTTCCCCCGAAGCATCATGAAGCAGATGGG TGAGGTTGGTCTGCTCGGAACAACCGTCAAGAGCCACGGATGCCCCG GCATGTCTTCTGTGCTTACGGTCTCGTGGCCCCGAGAGTCCGAGAGG GTCGACTCCGGCTACCGATCTGCCATGTCTGTGCAGTCGTGCTGGTC ATGCACCCCATGAACAGTTTGGATCCCAGGAGCAGAAGGACCGGTTT CTGCCCAAATTGGCCTCCGGCGAGATGATCGGCTGCTTCGGTCTCACC GAGCCTAACCACGGTTCCGACCCTGGATCCATGGAGACCGTCGCCAA GATGCACCCCTACTAAGAAGGGCGTCATTGTGCTCAATGGAGCCAAGAA CTGGATCACTAATCTCTCTATTGCCGATCTCATGGTTGTGTGGGCCAAG TTGGACGGTAAGATCCGAGGCTTCTTGTGAGCGATCTCAGGTGCGCC TCCGGCCTCGCTACTCCCGCATCAAGAACAAGACCGCTCTGCGAGCC TCCATCACCGGTATGATCCAGATGGACGACGTTGAGATCCCTGTGGAG AACATGTTCCCGAGGTGACCGGTCTCAAGGGCCCCCTTACCTGCCTC AACTCTGCCCGATACGGTATCGCCTGGGGAACCATGGGCGCTCTGTCC GAGTCCATCAAGCTCGCTCGAGAGTACTCTCTGGACCGAAAGCAGTTT AAGGGCCAGCCTCTGGCCAAGTACCAGCTCATCCAGAAGAAGCTCGCT GACGCTCTGACCGATGCCACCTACGGACAGGTGCTGCCATTACGGTC GGCCGGCTCAAGGATGCCGGCAATTGTCTCCCGAGCTCATCTCCATG ATTAAGAGACAGAAGTGTGACCGAGCCCTCGCTGGCGCTCGAAACCTG ATGGAGATCTTTGGCGGTAACGCTGCCTCTGACGAGTACCACATTGGC CGAATTGCCGCCAACCTGTGGGTTGTCCAGACCTATGAGGGCCAGTCT GATATCCATGCTCTCATCCTGGGAAGAGCCATGACCGGCGTCCAGGCT TTTGCTTAA
SEQ ID	>gi 50556786 ref XP_505801.1	MLTRISRLAPAARGFATSSVNRSTAAMDWQDPFQLDSLSTEDEIAVAEAAAR DFCQTELYPKVLEGYRTEEFPRSIMKQMGVEVLLGTTVKSHGCPGMSSVA

SEQ ID NO:	Description	Sequence
NO: 3683	YALI0F23749p [Yarrowia lipolytica](highly similar to uniprot Q7S579 Neurospora crassa NCU02291.1 hypothetical protein probable Glutaryl-CoA dehydrogenase)	YGLVAREVERVDSGYRSAMSVQSSLYMHPIEQFGSQEQKDRFLPKLASG EMIGCFGLTEPNHGS DPGSMETVAKMHPTKKGVIVLNGAKNWITNSPIADL MVVWAKLDGKIRGFLVERSQVASGLATPAIKNKTALRASITGMIQMDDVEIP VENMFPEVTGLKGPFTCLNSARYGIAWGTMGALSESILKAREYSLDRKQF KGQPLAKYQLIQKKLADALTDATYQGQVAAIQVGR LKDAGNCPPELISMIKR QNCDRALAGARNLMEIFGGNAASDEYHIGRIANLWVVQTYEGQSDIHALI LGRAMTGVQAFA
SEQ ID NO: 3684	>gi 255723091 ref XM_002546434.1 Candida tropicalis MYA-3404 conserved hypothetical protein, mRNA (similar to C. albicans ACD99)	ATGTCAGTCAAAGAAGATATCCCAGCTGTTTTCTTTCCCAAATTTCTCC TCGTGGTCTTGAAGCTATCCAGAAAACCAAAGACTTTGTCAATGACTAC TGTATTCAGCCGATGAAATCTACTTCAAACAGGTCTCTACTGATCCTG CCAAAGATGGAACAATCCACCTATTATTGAGACATTGAAATCCAA AGCCAAAGAACTTGGTTTATGGAATATGTTTTATCCAAACATTATAAGG AAGGTCCACAATATACCAACTTAGAATATGGTTTGATGGCTAGATATTTG GGTGCTGCACACACTGCTCCTGAAGCTACTAATACTGCTGCTCCAGATA CTGGTAACATGGAATTACTTGTCTAAATACGGTACTCCATACCAAAAAGA AAAGTACTTACAACCATTTGTTAGATGGAAAGATCAGATCTGCTTTCTTGA TGACCGAAAAGGGCACATCATCTTCCAATGCATTAAATATCTCCACTAG TGCCAAAAGAATGCCAGTGGTAACTATGTTCTTGATGGTGTAAGTG G TTTGCTTCAGGTGCTGGTGATCCAAGGTGTTCTGTTTGGTTGGTCATGT GTAAACTGAAGACAATAAGAAGAAGCCATATGCAAACACACCGTGTT GGTCTTGATGCCAAGAGAGCATTGGCTAGCGGCAGGGCCAAATTAGT CAGACCTTTGCATGTTATTGGATATGATGATGCTCCTCATGGTCATTGT GAAATTTCTTTTGAAGTACGAAGTTCTGCTGACGAAATGCCAAATG CTGTTTTGGCCGGTATTGGAAGAGGATTTGAGTTGATTGAGTCTAGATT AGGACCTGGTAGAATTCATCATTGTATGAGAGCTATTGGTACAGGTGAA ATTGCATTGTTGATCATTGCTCATAGAGCTAACCACAGAATGATTTTTGG AAAACCAATGAAAGACAGAGAAGGATTTTTGTCTAAGTTCGGTCAGAGC AGAATTGATATTACCAGATGCTTGTATTGGTATTAAATGCTGCTCATAA AATTGATATTCCAACGCAAGGCTGCTCAGAAAGAGATTGCCATGGCT AAGATTGAAACACCAAGAACCATCTCTGATATCCTTGACTGGGGTATCC AAGTTTTTGGCGCAGAAGGGGTCTCACAAGACACAGACTTAGCTAGAA TGTATGCTCTCAACAGAACCCTTGAGAATTGCTGATGGTCCTGATGAAGC TCACTTGGCACAATTGGCAAGAAATGAGGCCAAAAAATCCAGAGGT CGATATCTTCTTTGAACATGTTGCTAGTCAACGTAATAAATTATAG
SEQ ID NO: 3685	>gi 240130997 gb EER30559.1 conserved hypothetical protein [Candida tropicalis MYA-3404] (similar to C. albicans ACD99)	MSVKEDIPAVFLSQISPRGLEAIQKTKDFVNDYCI PADEIYFKQVSTDPAKR WKTIPPIETLKS KAKELGLWNMFLSKHYKEGPQYTNLEYGLMARYLGRAH TAPEATNTAAPDTGNMELLAKYGT PYQKEKYLQPLLDGKIRSAFLMTEKGT SSSNALNISTS AKKNASGNYVLDGVKWFASGAGDPRCSVWVLMCKTEDN KKNPYANHTVLVLD AKRALASGRAKLVRPLHVIGYDDAPHGHCEISFENYE VPADEMPNAVLAGIGRGFELIQSRLGPGRIHHCMRAIGTGEIALLIIAHRANH RMIFGKPMKDREGFLSKFGQSRIDITRCLLLVLNAAHKIDISNAKAAQKEIAM AKIETPRTISDILDWGIQVFGAEGVSQD TDLARMYALNRTLRIADGPDEAHL AQLARNEAKKFPEVDIFFEHVASQRNKL
SEQ ID NO:	XP_716423.1 probable acyl-CoA	MSVKEDIPAVFLEKVS PRGLEAIQKTKDFVNDYCLPADQIYFEQLSDIPSER WKSVPPIETLKKKAKELGLWNMFLSKHYKEGPQYTNLEYGLMARYLGRS YTAPEATNTAAPDTGNMELFAKYGTTYQKDRYLKPLLNGEIRSAFLMTEKG

SEQ ID NO:	Description	Sequence
3686	dehydrogenase [Candida albicans SC5314]	VSSSNALNISTSAVKNSNGNYVLNGVKWFASGAGDPRCSVWVLMCKTDN NKQNPYQNHTVLIIDAKKALATGKAKLIRPLQVIGFDDAPHGHCEIQFDYE VPADEMPNVVMAGVGRGFELIQSRLGPGRIHHCRAIGAGEFALLRIAHR ANHRLIFGKPMNQREGFLSRYGQSKIDIERCLLLVLNAAHKIDISNAKEAQK EIAMAKIETPRITSDILDWGIQVYGAEGMSQDTELARMYAHNRTLRIADGPD EAHLAQLARNEAKKFAKVDDFFANMETQRSKL
SEQ ID NO: 3687	XM_711330.1 Candida albicans SC5314 probable acyl-CoA dehydrogenase (ACD99) mRNA, complete cds	ATGTCAGTTAAAGAAGACATTCTGCTGTTTTCTTGAAAAGGTTTCTCC TCGTGGTCTCGAAGCCATCCAGAAAAACCAAGATTTCTGTTAACGATTAT TGTCTTCCAGCCGATCAAATTTATTTGAACAACCTTCAGACATCCCATC AGAAAGATGGAAGAGTGTTCTCTGTCATTGAGACATTGAAGAAGAAA GCCAAGGAACCTTGTTTTATGGAACATGTTTTGTCAAAGCATTATAAGG AAGGTCCACAATATACAACTTAGAGTATGGATTGATGGCCAGATACTT GGGTGCTTCATACACAGCACCAGAGGCTACCAACACAGCTGCTCCAGA TACCGGTAATATGGAATTGTTTGCCAAATATGGAACCACTTATCAGAAA GATAGATACTTGAAACCTTGTTAAATGGGGAAATTAGATCAGCATTCTT GATGACAGAAAAAGGTGTTTCATCATCTAATGCTCTCAATATTTCTACAA GTGCTGTCAAGAATTCGAATGGAAATTACGTGCTCAATGGTGTCAAATG GTTTGCTTCAGGTGCAGGAGATCCAAGATGTTCCGTCTGGTTGGTGAT GTGCAAGACAGACAACAACAAGCAAAATCCATATCAAAACACACAGTT TTGATCATCGATGCCAAAAAGGCTTTGGCTACTGGAAAAACCAAAATTGA TCAGACCATTGCAGGTCATTGGTTTTGATGATGCTCCTCATGACATTG TGAGATTCAATTTCAAGATTACGAAGTTCCTGCCGATGAAATGCCTAAT GTTGTTATGGCTGGTGTGTTGGTAGAGGATTTGAGTTGATTCAATCCAGAT TGGGTCCAGGTAGAATCCACCATTGTATGAGAGCTATTGGTGTGGTG AATTTGCATTATTGAGAATTGCTCACAGAGCAAAATCACAGATTGATTTTT GGTAAACCTATGAACCAGAGGGAAGGATTCTTATCCAGATACGGACAA AGCAAAATCGACATTGAAAGATGTTTATTGTTGGTGTGTAATGCTGCTC ACAAAATTGATATTTCCAATGCCAAAGAAGCACAAAAGGAAATTGCTAT GGCTAAGATTGAGACCCCGAGAACTATCTCTGATATTCTCGATTGGGGT ATTCAAGTTTATGGTGCTGAGGGTATGTCACAAGATACCGAGTTGGCCA GAATGTATGCTCATAACAGAACATTGAGAATAGCTGATGGACCTGATGA AGCTCATTTGGCCCAATTGGCTAGAAATGAAGCTAAAAAGTTTGCAAAA GTTGACGACTTTTTCGCCAACATGGAACTCAACGTAGCAAATTATAA
SEQ ID NO: 3688	acyl-CoA dehydrogenase NM domain-like protein [Candida tenuis ATCC 10573] (similar to C. albicans ACD99)	ATGACAGACCTTGACATTCCAGCAGTATTCCTTGATAAGATCTCACCAC GTGGCCTCGAGGCGATCCGCAAGACCTACGACTTTGTGCATAACTACT GTATTCCTGCGGATGCTCTCTACTTTGACCAAATTTCCAGGATCCCGA ACAAAGGTGGAAAACCACTCTGAAGTCACTGAAAAATTGAAACAAAAG GCCAAACAATTAGGTTTGTTGGAACATGTTCTCTCTAAGCACTATACCG ATGGACCTGGCTACACAACTTGAGATATGGCCTTATGGCGCAATTCTT GGGCCGGTCTGTTCTGTCGCCACCCGAGGCCACCAATACAGGTGCACCCG ATACAGGTAACATGGAGATTCTCGCCAAGTTCGGCTCGGCCTATCACC GGGAGCAGTACCTCCTTCCATTGCTCCGCGGTGAGATCCGCTCGGCGT TCTTGATGACAGAAAAAGGCACTTCTTCATCCAATGCCTTGAACATCTC ATGCTCGGCCCCAGAAGAATTCACACGGCAACTACGTTCTCAATGGAGT CAAGTGGTTTGCCTCTGGTGCAGGTGATCCTCGGTGTCGCGTGTGGTT GGTGATGTGCAAAACCGAGTCTCTGGACAACATCTACCGTAATCACAGT GTGTTGGTGTGGATGCGAAAAAGGCTTTAGCTTCAGGAAAAGCCAAAT TGATCCGACCACTCAGCGTGTTTGGCTATGACGATGCTCCTCATGGAC ACTGTGAGGTGGAGTTCAACGACTTTGAGGTGCCAGCCGAGGATATGG ATAATTCTATCCTTGTAAGGTGGGTATGGGATTTGAGATCCAGATC TCGTTTAGGCCCTGGGCGTATTCACCACTGTATGCGTCTTATTGGTGCC

SEQ ID NO:	Description	Sequence
		GGAGAATATGCCTTAATGAGGGCGGTGCTGAGGGCTGCCGGCAGAGA CATTTTCGGCAAGCCCATGGTGAAGAGAGAATCATTTCTCAATGCTTAT GGAGAGCATAAGCTTTCACCTTCAGAAATGCCGTCTTTTGGTGCTTAATG CAGCTCATCAAATCGATATTTTGAATGCTAAGACTGCCAAAAGAGATAT AGCCATGGCCAAAATCGAGACTCCCAGAGCAGTATTGAAGATTCTTGA CTGGTGTATTGAGTTTATGGGGCTGAAGGAGTGTCTCAAGACACAGA GCTTGCAAAGATGTATGCTCACGCTCGGACTTTGAGAATCGCAGATGG ACCAGATGAAGCACACCTTGGACAGCTTGCACGGGACGAGTCAAAGAA GTTTGCGGAGGTGGTGAAGTACTTTGAGGGACACAAGGCACGTCAAGA CCAAGTCCTGAAGTTGTGA
SEQ ID NO: 3689	>gij344233800 g b EGV65670.1 acyl-CoA dehydrogenase NM domain-like protein [Candida tenuis ATCC 10573](similar to C. albicans ACD99)	MTDLDIPAVFLDKISPRGLEAIRKTYDFVHNYCIPADALYFDQISQDPEQRW KTTPEVTEKLKQKAKQLGLWNMFLSKHYTDGPGYTNLEYGLMAQFLGRSF VAPEATNTGAPDTGNMEILAKFGSAYHREQYLLPLLRGEIRSAFLMTEKGT SSSNALNISCQAQKNSHGNYVLNGVKWFASGAGDPRCRVWLVCMCKTESS DNIYRNHSLVLDAAKKALASGKAKLIRPLSVFGYDDAPHGHCEVEFNDFEV PAEDMDNSILGKVGMGFEIIQSRLGPGRIHHCMLRIGAGEYALMRAV/SRAA GRDIFGKPMVKRESFLNAYGEHKLQLKORLLVLNAAHQIDISNAKTAKRDI AMAKIETPRAVLKILDWCIQVYGAEGVSQDTELAKMYAHARTLRIADGPDE AHLGQLARDESKKFAEVVKYFEGHKARQDQVSKL
SEQ ID NO: 3690	>gij50309254 refl XM_454634.1 Kluyveromyces lactis NRRL Y- 1140 hypothetical protein partial mRNA (similar to C. albicans ACD99)	ATGCCTAATGTCAGTGATAGACCGCGGACATATAAGAAACCTGCTTTAG AAGATGTTGATCCCATCACAACTATATACCTGCCAGTGTTAGGGATAA ATTTGATGAGAGGCAGATGGATCGGTTCAAGAAGTTGCCGAAATTTGTT GAGTTTGAATGTTTGCCATTAGATACGGTGTATTTGCAAGAGAGTACCC TATTTGAGCATGAAAGCGATTTAGAGACGTGCCAGTCATTATTAATTTA AGGAAGAAATTGGAGGCATACCAGTTGCATAAAATGTTTGTTCCAATGG ATCAACGTGGGTACGACCATAGTTTCAACGATAATTGGGAAGTGGTGA GTATGGTTGAATTTGCTATGATCGCTTTCCTTGCTGGAAGATCTGTCATT GCCAGTTATTTGTTCCATTTGGATGATTTGATCGATTTAGGAACATACA AGTTTTGTTGAGAAATGTTTGTTCGAACCATGATTTGTGGGTACAAGTG ATAGATGAGTTAGTTTCTAATAATATGAAATCGTGTGTTGATGGTAAGTGA AAGAGATGTGTCTGGTCTGATGCGTTGAACGTTCAAACACCTGTAAA ATTGAAGGGGATGATCTAAACGAAGAGGAGGCTACTATGACACTTAAC GGTACTAAATGGTTTATCAAAGATGCAGGAGACTCAGATATTTGGTTAG TTTTATGTGTCACCTGAATTTGATGAGGGCAACATTTATAGAAAACATACA TTATGCCCTTGTTAACAGGAATGATTTACCACCAAATTCAACAAGAATTGA ACCTATAGAAAACAAATGAAGCGATTGGTAAATTTTATGAAGTACAATTTA AAGATTGTAAAGTACCGTTAAATATTATTGGTGAAAGAGGTGAAGGTGA TCAAATTTTACAAATGAAATCCTCTGTTACAAAATTATTTCAATGCTTAAA ACTTTGTGGTATGGGACAAGAATCCTTGAGACTTTCCTAATAAGAGAGCT GCTGAAAGGAAAGTGTGTTGTTCCAAATTACAGAAGAGTGAGTATTTCA AATTTGATCTTGCTCATTGGAGGATTAAGATTGAAACCTGTAAGCTGCTT TGTTTCAACGCGGCAATCAAATGTGATTACGAAGGTGTAAAAGCGGCAA GAGAAGAAATTGGGATGGTGAAAGCCGTGACACCAAAGGAAATCTCGT CACTGGTGGATTGGTCTATCCAGTTGCATGGATGTTACGGACTCTGTTC AACACAAACACCCCTTGTCACATATGTGGCAAGTGAGTCGATCGCTAAGA ATTAATGATACGCCGGACGAATCATTAATATCACAACCTGGGGAGGTTGG AAATCAGTAATTATAACAAATTTCAAAGACATACGATCAAGAATTAACG ACGCTCGCTGGCAAATGA

SEQ ID NO:	Description	Sequence
SEQ ID NO: 3691	>gij49643769 emb CAG99721.1 KLLA0E15181p [Kluyveromyces lactis] (similar to C. albicans ACD99)	MPNVSDRPRTYKKPALEDVDPITNYIPASVRDKFDERQMDRFKKLRKFVEF ECLPLDTVYLQESTLFEHESDLETCPVIINLRKKLEAYQLHKMFVPMQDRG YDHSFNDNWEVVSMEVFAMIAFLAGRSVIASYLHLDLIDLTGTVLLRNG CSNHDLWVQVIDELVSNMKSCLMVSESDVSGSDALNVQTTCKIEGDDLN EEEATMTLNGTKWFIKDAGDSDIWLVL Candida TEFDEGNIYRKHTLCLVN RNDLPPNSTRIEPIETNEAIGKFYEVQFKDCKVPLNIIGERGEYQILQMKS SVTKLFQCLKLCGMGQESRLSNKRAAERKVF GSKLQKSEYFKFDLAHWR IKIETCKLLCFNAIAIKCDYEGVKAAREEIGMVKAVTPKEISSLVDSIQLHGC YGLCSTQTPLSHMWQVSRSLRINDTPDESLISQLGRLEISNYNKFQKTYDQ ELTTLAGK
SEQ ID NO: 3692	>gij301507715 g b GU338397.1 Candida rugosa propionyl-CoA dehydrogenase mRNA, complete cds (similar to C. albicans ACD99)	ATGTCGATTAAGGACGACATCCCTGCCATCTTTACGAAAACTTTCCC CCCGCGGGCTTGAGGCTATCGCCAAAACCAAGGAATTCGTCGACACTT ACTGCTCCCCCGCCGACGAGATCTACTTCCAACAGGTGAGAAGTACG ACCGCGGTGGAAGGAAACGCCCCCATCACCGAGCACCTTGAAGAAG AAAGCTAAAGAGCTCGGGTTATGGAACATGTTCTTGCTGAAGCACTACG CCGAGGGCGCCGCTACACCAACTTGGAGTATGGGCTTATGGCCCAG TACCTTGCGCGCAGTCACATCGCCCCCTGAAGCTACCAACACCAATGCT CCTGACACCGGCAACATGGAGATCCTTGCCAAAGTACGGCAACGACTAC CACAAGCAGCGCTACCTCCAGCCGCTTCTCGACGGTAAATCCGCGCTG GCGTTCTTAATGACGGAAGGGGACGTCGCTGTCCAACGCCCTTAAC ATCTCGTGCCTGGCAAACTTAACCAAAATGGCAACTACGTCAATCAACG GCGTCAAGTGGTTCGCCCTGGGTGCCGGCGACCCCCGGTGCAAGGTG TGTTGACGATGTGCAAGACCAGCGACGACGACGCCAACCCATATTTT AACCACTCGTTGCTTGTGCTTGATGTGCGACAAGGCCCTCGCCCTGGGA CAGGCTCGTGTGTCCGCCCGTTGCACGTGTTTGGCTACGACGACGCT CCTCACGGTCACTGTGAAATTGAATTTAACAACACTACGAAGTGTCCAAAG AGGAAATGGCCAAACGTCATCCTCGGCCAGGTGGGCCAAGGATTTGCCA TCATCCAGCTGAGATTGGGGCCGGGGCGCATCCACCACTGCATGCGG ATGATTGGCGTCGGCGAATTCGCCTTGATGAGAGTGGCTCAGCGGGCT AACCAACGTATCATCTTCGGTAAGCCCATGGCCAAGCGCGAAGTGT TGAACGCCCTACGCTCAGGCAAGATCGACATCCAAAAGTGCCGCTTGT TTGTTCTTAATGCCGCCACACATCGACATTGCCGGAGCCAAAGCGG CGCAAGCCGACATCGCCATGGCCAAGATCGAGACCCCGAGAACCATC CTTCGCATCTTGGACTGGGGGATCCAGATGTTTGGCGCCGAAGGGT GTCTCAAGACACCGAGCTCTCGCGCATGTACGCGTTGGGGCGGACGT TACGCATTGCCGACGGCCCCGATGAAGCTCACTTGGGCCAATTGGCCC GTAAGGAGCTGAAGAAGTTCCTTACGTGATGAGTACTTTAAGCGGTT TGAAGAAAATAAGGCCAAGTTGGCCAAGTTGTAA
SEQ ID NO: 3693	>gij301507716 g b ADK77878.1 propionyl-CoA dehydrogenase [Candida rugosa] (similar to C. albicans ACD99)	MSIKDDIPAIFYEKLSPRGLEIAIKTEFVDITYCSPADEIFYQQVRTDDRRW KETPPITEHLKKKAKELGLWNMFLSKHYAEGAGYTNLEYGLMAQYLGRSHI APEATNTNAPDTGNMEILAKYGN DYHKQRYLQPLLDGKIRSAFLMTEKGT SSNALNISCSAKLNQNGNYVINGVKWFASGAGDPRCKVWLTMCCTSDDD ANPYFNHSLVLVDVKALASGQARVVRPLHVFGYDDAPHGHCEIEFN NYE VSKEEMANVILGQVGGFAIIQSRLGPGRIHHCMRMIGVGEFALMRVAQR ANHRIIFGKPMKRESFLNAYAAQAKIDIQKCRFLVNLAAHHIDIAKAAQA DIAMAKIETPRTILRILDWGIQMFAGVGSQDEL SRMYALGRTLRIADGPD EAHLGQLARKESKFPYVDEYFKRFEENKAKLAKL
SEQ ID NO: 3694	>gij380353348:2 14809-216140 Candida orthopsilosis Co	ATGTCAGTTAAAGACGATATCCCAGCTATCTTTTATAGATAAGGTTTCTCC AAGAGGTCTTGAAGCAATTCAAAAGACAAAGGACTTTGTCGACCAATAT TGTATCCCTGCTGATAAGATTTTCAAGGAGCAAATTCGCAAGACCCAA AAATAAGATGGAAACAATATCCAGCTATCATTGAACCATTGAAGAAAA

SEQ ID NO:	Description	Sequence
	90-125, chromosome 4 draft sequence (similar to <i>C. albicans</i> ACD99)	GGCTAGAGAGTTGGGTTTGTGGAACATGTTTTGTCCAAGCATTACAAA GAGGGTCCTCAATTTACCAATTTGGAATACGGATTAATGGCTAGGTATT TGGGAAGATGTCACACTGGACCAGAAGCAACCAACACCAGTGCCCCAG ACACAGGTAATATGGAATTGTTTGGCTAAATATGGTACAAAGGCGCAAAA GGATAAGTATTTAGTGCCCTTGATGGATGGTAAGATCAGATCGGCATTCT TTGATGACCGAAAAGGGGATTTTCATCGTCGAATGCATTAAACATTTCAA CCACTGCCATTAAGAATGCCCGTGGTAACATATGTGTTGAATGGAACAAA GTGGTTTGCCTCTGGTGCTGGAGATCCAAGAACTGCTGTTTGGTTGGT TATGTGCAAAACAGACAATGATGAAAGTAATATGTTGAGAAACCACTCC GTGTTAGTCATTGATGTCAAGCATGCATTAGCATCAGGTAAGGCTGAAG TTATCAGGCCTTTGAGTATTTTGGCTACGATGATGCACCCCATGGTCA TTGTGAAATCGTTTTCAAGGATTATGAAGTTTCATCTGAATTGATGCCAG AAACGATTTTGGCCGGTGTCGGTAGGGGATTTGAATTGATTCAATCCCG TTTGGGTCCAGGTAGAATCCATCATTGTATGAGAGCCATAGGTGCTGGT GAATTTGCCTTGTTGCGTATTGCTCACAGAGCAAATCACAGAACCATCT TTGGTAGGCCAATGAATAGAAGAGAAGGCTTCTTGATGCAGTATGCCAA GTACAGAATTGAAATCAAAAATGTTTATTATTGGTTTTGAATGCTGCTC ACAAGATTGACATCACTAATGCCAAACATGCACAAAGAGAAATTGCCAT GGCTAAAATTGAGACTCCAAAAACAATTTGCGATATTCTCGACTGGGGT ATTCAAGTCTTTGGAGCCGAAGGATTCTCTCAAGATACAGAATTGGCAC AAATGTATGCTTGAATAGAACTTTGAGAAATCGCTGATGGTCTCTGATGA AGCACATTTGGCTCAATTGTCAAGAAGAGAAGCTGCCAAGTTTCCAGAA GTTGATGAGTTTTTCAAGAGTGTTGAATCAAGAGTTGAAGCTATTAGTAA GTTATAA
SEQ ID NO: 3695	>gij380353467 e mb CCG22977.1 hypothetical protein CORT_0D01290 [Candida orthopsilosis](similar to <i>C. albicans</i> ACD99)	MSVKDDIPAIFLDKVS PRGLEAIQKTKDFVDQYCI PADKIFKEQISQDPKIRW KQYPAIIEPLKKKARELGLWNMFLSKHYKEGPQFTNLEYGLMARYLGRCH TGPEATNTSAPDTGNMELFAKYGTAKQKDKYLVPLMDGKIRSAFLMTEKGI SSSNALNISTTAIKNARGNYVLNGTKWFASGAGDPRTAVWLVLMCKTDNDE SNMFRNHSLVIDVKHALASGKAEVIRPLSIFGYDDAPHGHCEIVFKDYEV SELMPETILAGVGRGFELIQSRLGPGRIHHCRAIGAGEFALLRIAHRANHR TIFGRPMNRREGFLMQYAKYRIEQKCLLLVLNAAHKIDITNAKHAQREIAM AKIETPKTICDILDWGIQVFGAEGFSQDTELAQMYAWNRTLRIADGPDEAH LAQLSRREA AKFPEVDEFFKSVESRVEAISKL
SEQ ID NO: 3696	>gij354545630:2 25012-226343 Candida parapsilosis strain CDC317 annotated contig 005809(similar to <i>C. albicans</i> ACD99)	ATGTCAGTTAAGGACGATATTCAGCAATCTTTTATAGATAAGGTTTCCCC AAGAGGTCTTGAAGCTATTCAAAAGACAAAAGACTTTGTTGAGCAATAC TGTATTCCTGCCGATAAAGTTTTCAAGAAACAGATTTTCGACAGACCCAG CGGTAAGATGGAACAAATACCCTGCTATTATTGAACCATTGAAGAAAAA GGCTAGGGAATTGGGATTGTGGAACATGTTTTGTCCAAGCATTACAAA GAGGGTCCTCAATTTACCAACTTGAATATGGATTGATGGCTAGGTATC TAGGAAGATGCCACACTGGTCCTGAAGCCACTAACACTAGTGCACCAG ACACGGGTAATATGGAGTTGTTTGCAAAATATGGTACAAAGGCGCAAAA AGACAAATATTTGGTGCCCTTGATGGATGGTAAGATTAGATCAGCATTT TTGATGACTGAAAAGGGGATCTCATCGTCCAATGCGTTGAACATTTCCA CCACTGCAATTA AAAA ACTCACGTGGAACTATGTCTTGAATGGTAACAA GTGGTTTGCATCAGGCGCTGGTGATCCTAGAACTGCCGTTTGGTTGGT TATGTGTAAGACTGCCAACGATGAAAAGAATGCATTTAAAACCACTCA GTATTAGTGATTAATGTTAAGCATGCATTAGCATCAGGCAAGGCTGAAG TTATTAGACCTTTGGGAATTTTCGGATACGACGATGCTCCTCATGGACA TTGTGAAATTGTTTTCAAAGATTATGAAGTTTCATCAGAGTTGATGCCAG ATACCATTTTGGCTGGTGTGGTAAAGGATTTCGAATTGATTCAATCTAG ATTGGGCCCCGGGTAGAATCCATCATTGTATGAGAGCTATTGGTGCTGG

SEQ ID NO:	Description	Sequence
		TGAATTTGCATTGTTGCGTATCGCCACAGAGCTAATCACAGAATTATTTTGGTAAACCAATGAATAGAGAAGGCTTTTTGATGCAGTATGCCAAGTACAGAATCGAGATTCAAAAATGTTTATTGTTGGTTTTAAATGCTGCCCACAAGATAGATATCACTAATGCCAAAGAAGCTCAAAGAGAAATTGCAATGGCCAAGATTGAAACTCCAAAAACCATTGTGATATTCTTGATTGGGGTATTCAAGTGTGGAGCTGAGGGTTCTCTCAGGATACAGAATTGGCGCAATGTACGCTTGAACAGAACTTTGAGAATTGCAGATGGACCAGATGAGCACATTTGGCTCAATTAGCAAGAAGAGAAGCCGCAAAGTCCCTGACGTTGACGTGTTTTTAAAGATGTTGATTCAAGAGTTGAGGCTGTTAGTAAATTATAA
SEQ ID NO: 3697	>gij354545753 emb CCE42481.1 hypothetical protein CPAR2_201240 [Candida parapsilosis](similar to C. albicans ACD99)	MSVKDDIPAIFLDKVS PRGLEAIQKTKDFVEQYCI PADKVFKKQISTDPAVRWKQYPAIIEPLKKKARELG LWNMF LSKHYKEGPQFTNLEYGLMARYLGRC HTGPEATNTSAPDTGNMELFAKYGTAKQDKYLVPLMDGKIRSAFLMTEKGISSSNALNISTTAIKNSRGNYVLNGTKWFASGAGDPRTAVWLVMCKTANDEKNAFKNHNSVLVINVKHALASGKAEVIRPLGIFGYDDAPHGHCEIVFKDYEVSSELMPDTILAGVGKGFELIQSRLGPGRIHHC MRAIGAGEFALLRIAHRAN HRIIFGKPMNRREGFLMQYAKYRIEQKCLLLVLNAAHKIDITNAKEAQREIAMAKIETPKTICDILDWGIQVFGAEGFSQDTELAQMYAWNRTLRIADGPDEAHLAQLARREAAKFPD VDVFFKDVDSRVEAVSKL
SEQ ID NO: 3698	>gij241959309 ref XM_002422329.1 Candida dubliniensis CD36 acyl-coa dehydrogenase, putative (CD36_34410) mRNA, complete cds	ATGTCAGTTAAAGAAGATATTCCTGCTATTTTCCTTGAAAAGATTTCCCCTCGTGGTCTTGATGCTATCCAGAAAACCAAAGATTTCTGTAACGATTATTGTCTCCAGCAGATCAGATCTATTTTGAGCAGCTCTCTGACATCCCTTCAGAAAAGATGGAAGTGTTCCTCCTGTCATTGAGACATTGAAGAAGAAAGCCAAGGAACCTGGTTTATGGAACATGTTTTGTCAAAGCATTATAAAGAAGGTCCACAATACACAACTTAGAGTATGGGTTGATGGCCAGATACTTGGGTCTTCATACACTGCGCCAGAGGCTACCAATACTGCTGCTCCAGATACCGGTAATATGGAATTGTTTGCCAAATATGGTACCACTTATCAGAAA GATAGATACTTGAAACCCCTGTAAATGGGGAAATCAGATCGGCATTCTTGATGACCGAAAAGGGTGTTCATCATCCAATGCTCTCAATATTTCTACAGCGCTATCAAGAACTCTAATGGTAATTACGTGCTCAATGGTGTCAAATGGTTTGCTTCAGGGGCAGGAGATCCAAGATGCTCTGTATGGTTGGTAA TGTGCAAGACCGACAACAATAAGCAAAACCCCTTATCAGAACCACACTGT TTTGATTATCGATGCAAAAAAGGCTTTGGCTACCGGAAAAGCCAAATTGATCAGACCATTGCAAGTCATTGGTTTTGATGATGCTCCCATGGACATTGTGAAATCCAATTCAAAGACTACGAAGTTCCTGCTGATGAAATGCCTAATGTTGTAATGGCAGGTGTTGGTAGAGGATTTGAGTTGATTCAATCCAGATTGGGTCCAGGTAGAATCCACCATTGTATGAGAGCTATTGGTTCTGGTG AATTTGCTTTATTAAGAATTGCTCATAGAGCAAATCACAGATTAATTTTTGGTAAGCCCATGAACCAAAGAGAGGGGTTCTTATCCAGATACGGACAAAGCAAAATAGATATTGAAAGATGTTTGTTATTGGTGTGAATGCCGCTCACAAAATCGATATTTCCAATGCCAAAGAGGCACAAAGGGAAATTGCTATGGCCAAGATTGAAACCCCAAGAACTATTTCTGATATTCTCGATTGGGGTATTCAAGTTTATGGAGCTGAAGGTATGTCTCAAGACACTGAGTTGGCCAG AATGTATGCCCATAAACAGAACATTGAGAATAGCTGATGGACCTGATGAGCTCATTTGGCTCAATTGGCTAGAAACGAAGCTAAGAAGTTTCCAAAA GTTGACGCCTTCTTTACCAACATGGAACACAACGTAGCAAATTATAA

SEQ ID NO:	Description	Sequence
SEQ ID NO: 3699	>gil241959310 ref XP_002422374.1 acyl-coa dehydrogenase, putative [Candida dubliniensis CD36]	MSVKEDIPAIFLEKISPRGLDAIQKTKDFVNDYCLPADQIYFEQLSDIPSERW KSVPPVIETLKKKAKELGLWNMFLSKHYKEGPQYTNLEYGLMARYLGRSY TAPEATNTAAPDTGNMELFAKYGTTYQKDRLKPLLNGEIRSAFLMTEKGV SSSNALNISTSAIKNSNGNYVLNGVKWFASGAGDPRCSVWLVCKTDNN KQNPYQNHVLIIDAKKALATGKAKLIRPLQVIGFDDAPHGHCEIQFKDYEV PADEMPNVVMAGVGRGFELIQSRLGPGRIHHCMRAIGSGEFALLRIAHR NHRLIFGKPMNQREGFLSRYGQSKIDIERCLLLVLNAAHKIDISNAKEAQREI AMAKIETPRTISDILDWGIQVYGAEGMSQDTELARMYAHNRTLRIADGPDE AHLAQLARNEAKKFKPKVDAFFTNMETQRSKL
SEQ ID NO: 3700	>gil126138209 ref XM_001385591.1 Scheffersomyces stipitis CBS 6054 acetyl-coenzyme-A dehydrogenase partial mRNA	ATGTCCGCCAAAGACGATATCCCTGCCATTTTCTTGGACAAGATCTCTC CCAGAGGTCTTGAGGCCATTGAGAAGACCAAACGTTTCGTGGAAGACT ACTGTTTGCCAGCTGACGATATCTACTTCAAGCAGATCAAGACCGATCC CGCAGTTAGATGGAATATACTCCCGAAATCACGGAAGAGTTGAAGAAG AAAGCAAAGGAACTCGGGCTCTGGAACATGTTCTTGTCTAAGCACTACA AGGAAGGACCCAGTTCACTAACTTGGAGTACGGGTTGATGGCTGAGT ACTTGGGCAAATCCTTTGTTGCTCCAGAGGCTACCAACACTGCAGCTC CAGATACCGGAAACATGGAACTTTTTGCCAAATACGGAACCTCCATACCA AAAGGAGAAGTGGCTCAAGCCATTGTTGAACGGAGAAATCAGATCAGC TTTCTTGATGACAGAGAAGGGTGTTCCTTCATCGAATGCCTTGAACATTT CGACTAGTGCCATTAAAGAACGCCCCAAGGCAACTACGTTCTTAACGGTG TCAAGTGGTTTGCTTCTTGAGCTGGAGATCCAGATGTTCACTGCTGGC TTGTCATGTGTAAAACACCGACGACTCCAGCAAGCCATACTTCAACCA TTCTGTCTTGATTTTAGATCCCAAAGTCGCTATTGCTTCTGGAAAAGCCA GGGTGGTCAGACCTTTGCATGTGATTGGGTACGACGATGCGCCCCATG GCCATTGTGAAATCGAGTTCACCAACTACGAGGTTTCAGCTGAAGAAAT GAAGAACACCATTCTTGCTGGTGTGGCCGTGGTTTTGAGCTCATCCA GTCCCGTTTGGACCAGGCAGAATCCATCACTGTATGAGACTGATTGG TTCTGGCGAGTTTGCTTTGCTCAAGACAGCACACAGAGCCAACAACAG AATCATCTTTGGCAAGCCCTTGCCCAATAGAGAGTCCTTTATCACAGCT TTTGCTCAACATAAGATCGACATTCAGAAAGTGTGCTTTGTTGGTGTGA ACGCGGCCCCACAAGATTGACATCACC AATGCCAAGGGTGCCCGAGAAG GAAATTGCCATGGCAAAGATCGAGACTCCAAGGACAGTGTGCAAGATC ATAGATTGGGGCATGCAAATGTTTGGTGCCGAAGGGTTATCTCAAGAC ACTGAGCTTGCCAGAATTTATGCCATGACCAGAATATTGAGAATTGCCG ACGGTCCAGATGAAGCTCATTTGAACCAGTTAGGTAGAAACGAAGCAA AGAAATTCAACGAGGCTGATGCCTTCTTTGCTACCTATGAGGCAAGCAG AGCCAGATTGAAAAAATTGTAG
SEQ ID NO: 3701	>gil150866135 ref XP_001385628.2 acetyl-coenzyme-A dehydrogenase [Scheffersomyces stipitis CBS 6054]	MSAKDDIPAIFLDKISPRGLEAIEKTKRFVEDYCLPADDIYFKQIKTDPVRW KYTPETEKLKKKAKELGLWNMFLSKHYKEGPQFTNLEYGLMAEYLGKSFV APEATNTAAPDTGNMELFAKYGTPYQKEKWLKPLLNGEIRSAFLMTEKGV SSSNALNISTSAIKNAQGNVYVLNGVKWFASGAGDPRCSVWLVCKTTDDS SKPYFNHSLVLDPKVAIASGKARVVRPLHVIGYDDAPHGHCEIEFTNYEVS AEEMKNITLAGVGRGFELIQSRLGPGRIHHCMRSIGSGEFALLKTAHRANN RIIFGKPLANRESFITAFQHKIDIQKCRLLVLNAAHKIDITNAKGAQKEIAMA KIETPRTVCKIIDWGMQMFGEGLSQDTELARIYAMTRILRIADGPDEAHLN QLGRNEAKKFNEADAFFATYEASRARLEKL

SEQ ID NO:	Description	Sequence
SEQ ID NO: 3702	>gil146422929 ref XM_001487349.1 Meyerozyma guilliermondii ATCC 6260 hypothetical protein (PGUG_00776) partial mRNA	ATGTCTGTAAAGAGGATATTCCGGCTATTTTCTCGACAAGATTTTCGC CCAAAGGATTGGACGCGATCCAGAAATGTAAGGATTTTGTGAGCAATA CTGTCTTCCGGCGGATAAAATATACCTAGAGCAGCTTAGCCCTGACCC CACAAAAAGATGAAAATCTACCCACAAATCACTGAAAAATTGAAGAAA AAAGCCCAAGAATTGGGACTTTGGAACATGTTCTTGTCAAAACACTATG CTGAGGGTGACAGGGTACACCAACTTGAATATGGGCTCATGGCAGGTT ATTTAGGGCGGTCGTTGGTGGCCCCAGAAGCAACCAATACCAATGCAC CCGACACGGGCAATATGGAATTGCTTGCCAAATACGGCACTCAGTACC ATAAAGAACGTTGGCTCAAGCCATTGTTGAACGGAGAGATTCCGGTCGG CTTTTTTGATGACGGAAAAGGGTACTTCTTCGTCTAATGCGTTGAACATT TCTGTTTTCGGCCAAGAAAAAATGCCAATGGGAATTGGGTATTGAATGGTA TTAAGTGGTTCGCTTCTGGATCAGGAGACCCACGGTGTTTCAAGTTTGGTT GGTAATGTGCAAAACAGCCGAACTAAGGCGATTTATGAAAACCACTCG GTTCTCGTTATCGATGCCAAAAAGGCATTGGCTACAGGAAATGCCAAAT TGATCCGGCCATTACATGTTTTTGGCTATGACGATGCTCCTCACGGACA CTTTGAGGTGGAATTCAACAACATGAGATTCCAAGTGAAGATATGCCC CATTCCATATTGGCTTCTGAAGGTAGAGGATTGAGCTCATTCACTCGA GACTTGGTCCTGGTCGTATCCACCACTGTATGAGACTGATTGGTGCTG GAGAACAAGCGTTGTTGCGCGTGAGCCATCGTGCCAACAATCGGCTCA TTTTCGGTACACTTATGGCAAAGAGAGAATCATTATTACTGCATTTGCC CAGCACAAAGATCAACCTTCAGAAATGTAGATTGCTCGTTTTGAATGCTG CCCACAAAATTGACATCAGTAATGCCAAACAGGCACAACGGGAGATTG CTATGGCCAAAATTGAGACTCCAAGAACCGTTGGTAGGGTACTTGACT GGGGTATCCAAATGTTTGGAGCAGAGGGAGTTTCGCAAGACACCGAAT TGGCTCGTCTGTATGCTATCAACCGGACACTCCAGATTGCTGATGGCC CCGACGAAGCTCATTGAACCAATTGGGATTGAAAGAGGCCAAGAAATT TGCATTGCAAGTGAATTCTTTGCTCAACAAGAAGAAATACCGCAAACGA TTATCTAACCTCTAG
SEQ ID NO: 3703	>gil146422930 ref XP_001487399.1 hypothetical protein PGUG_00776 [Meyerozyma guilliermondii ATCC 6260]	MSVKEDIPAIFLDKISPKGLDAIQKCKDFVEQYCLPADKIYLEQLSPDPTKR WKSTPQITEKLKKKAQELGLWNMFLSKHYAEGAGYTNLEYGLMAGYLGR SLVAPEATNTNAPDTGNMELLAKYGTQYHKERWLKPLLNGEIRSAFLMTE KGTSSSNALNISVSAKKNANGNWWLNGIKWFASGSGDPRCSVWLVCMCKT AETKAIYENHSVLVIDAKKALATGNAKLIRPLHVFGYDDAPHGHFEVEFNKY EIPSEDMPHSILASEGRGFELIQSRLGPGRIHHCMLRIGAGEQALLRVSHRA NNRLIFGTLMAKRESFITAFQHKINLQKCRLLVLNAAHKIDISNAKQAQREI AMAKIETPRTVGRVLDWGIQMFAGVSGDTELARLYAINRTLQIADGPDE AHLNQLGLKEAKKFALASEFFAQQEYRKRLSNL
SEQ ID NO: 3704	oAA2835	CACACAGCTCTTCCATAATGTCCGACGAGGAATCAGA
SEQ ID NO: 3705	oAA2836	CACACAGCTCTTCCCTCTCTTCTATTCTAGTTGATGGCCTGGTCTC
SEQ ID NO: 3706	oAA2837	CACACAGCTCTTCCATAATGTCCGACGACCTTATCAC
SEQ ID NO: 3707	oAA2838	CACACAGCTCTTCCCTCTCTTCTATTCTATAATTTATGTTTCAACTCAC C
SEQ ID NO: 3708	oAA3085	ATCGTTACCACCATCCCTACAAT

SEQ ID NO:	Description	Sequence
SEQ ID NO: 3709	oAA3086	CCGAAACAACCGTAGATACCTTTAAGCTACAACACTATACACGATAATTCCC
SEQ ID NO: 3710	oAA3087	GGGAATTATCGTGTATAGTGTGTAGCTTAAAGGTATCTACGGTTGTTTCGG
SEQ ID NO: 3711	oAA3088	CTTGGACATTTGACCTTGGCGGTACCGAGCTCTGCGAATT
SEQ ID NO: 3712	oAA3089	AATTCGCAGAGCTCGGTACCGCCAAGGTCGAAATGTCCAAG
SEQ ID NO: 3713	oAA3090	GCTTGTTCTGCAAAATGGAGTCA
SEQ ID NO: 3714	oAA3212	GGGGGAGATCGTTACCACCA
SEQ ID NO: 3715	oAA3213	AATTCGCAGAGCTCGGTACCGCTGCTGCTGCTGCTGTTTT
SEQ ID NO: 3716	oAA3214	AAAACAGCAGCAGCAGCAGCGGTACCGAGCTCTGCGAATT
SEQ ID NO: 3717	oAA3215	TTCGTTGTTGGCTCTCTCCATTAAAGGTATCTACGGTTGTTTCGG
SEQ ID NO: 3718	oAA3216	CCGAAACAACCGTAGATACCTTTAATGGAGAGAGCCAACAACGAA
SEQ ID NO: 3719	oAA3217	CAAAGGCATCGGTCAACTCC
SEQ ID NO: 3720	Candida strain ATCC20336 POX4	ATGACTTTTACAAAGAAAAACGTTAGTGTATCACAAGGTCCTGACCCTA GATCATCCATCCAAAAGGAAAGAGACAGCTCCAAATGGAACCCTCAAC AAATGAACTACTTCTTGGAAGGCTCCGTCGAAAGAAGTGAGTTGATGAA GGCTTTGGCCCAACAAATGGAAAGAGACCCAATCTTGTTACAGACGCG CTCCTACTACGACTTGACCAAGGACCAACAAAGAGAATTGACCGCCGT CAAGATCAACAGAATCGCCAGATACAGAGAACAAGAATCCATCGACACT TTCAACAAGAGATTGTCCTTGATTGGTATCTTTGACCCACAGGTCGGTA CCAGAATTGGTGTC AACCTCGGTTTGTTCTTTCTTGATCAGAGGTAA CGGTACCACTTCCCAATTGAACTACTGGGCTAACGAAAAGGAAACCGC TGACGTTAAAGGTATCTACGGTTGTTTCGGTATGACCGAATTGGCCAC GGTTCCAACGTTGCTGGTTTGAAACCACCGCCACATTTGACAAGGAA TCTGACGAGTTTGTATCAACACCCACACATTGGTGCCACCAAGTGGT GGATTGGTGGTGTCTCACTCCGCCACCCACTGTTCTGTCTACGCCA GATTGATTGTTGACGGTCAAGATTACGGTGTCAAGACTTTTGTTGTCCC ATTGAGAGACTCCAACCACGACCTCATGCCAGGTGTCACTGTTGGTGA CATTGGTGCCAAGATGGGTAGAGATGGTATCGATAACGGTTGGATCCA ATTCTCCAACGTCAGAATCCCAAGATTCTTTATGTTGCCAAAAGTTCTGTA AGGTTTCTGCTGAAGGTGAAGTCACCTTGCCACCTTTGGAACAATTGTC TTACTCCGCCTTGTTGGGTGGTAGAGTCATGATGGTTTTGGACTCCTAC AGAATGTTGGCTAGAATGTCCACCATTCCTTGAGATACGCCATTGGTA GAAGACAATTCAAGGGTGACAATGTCGATCCAAAAGATCCAAACGCTTT GGAAACCCAATTGATAGATTACCCATTGCACCAAAAGAGATTGTTCCCA TACTTGGCTGCTGCCTACGTCATCTCCGCTGGTGCCCTCAAGGTTGAA GACACCATCCATAACACCTTGGCTGAATTGGACGCTGCCGTTGAAAAG

SEQ ID NO:	Description	Sequence
		AACGACACCAAGGCTATCTTTAAGTCTATTGACGACATGAAGTCATTGT TTGTTGACTCTGGTTCTTGAAGTCCACTGCCACTTGGTTGGGTGCTGA AGCCATTGACCAATGTAGACAAGCCTGTGGTGGTCACGGTTACTCGTC CTACAACGGCTTCGGTAAAGCCTACAACGATTGGGTTGTCCAATGTACT TGGGAAGGTGACAACAATGTCTTGCCCATGAGTGTGGTAAGCCAATT GTCAAGCAAGTTATCAGCATTGAAGATGCCGGAAGACCGTCAGAGGT TCCACCGCTTTCTTGAACCAATTGAAGGACTACACTGGTTCCAACAGCT CCAAGGTTGTTTTGAACACTGTTGCTGACTTGGACGACATCAAGACTGT CATCAAGGCTATTGAAGTTGCCATCATCAGATTGTCCAAGAAGCTGCT TCTATTGTCAAGAAGGAATCTTTCGACTATGTCGGCGCTGAATTGGTTC AACTCTCCAAGTTGAAGGCTCACCCTACTTGTGACTGAATACATCAG AAGAATTGACACCTTTGACCAAAAAGGACTTGGTTCCATACTTGATCACC CTCGGTAAGTTGTACGCTGCCACTATTGTCTTGACAGATTTGCCGGTG TCTTCTTGACTTTCAACGTTGCCTCCACCGAAGCCATCACTGCTTTGGC CTCTGTGCAAATTCCAAAGTTGTGTGCTGAAGTCAGACCAAACGTTGTT GCTTACACCGACTCCTTCCAACAATCCGACATGATTGTCAATTCTGCTA TTGGTAGATACGATGGTGACATCTATGAGAACTACTTTGACTTGGTCAA GTTGCAGAACCCACCATCCAAGACCAAGGCTCCTTACTCTGATGCTTTG GAAGCCATGTTGAACAGACCAACCTTGGACGAAAGAGAAAGATTTGAA AAGTCTGATGAAACCGCTGCTATCTTGTCCAAGTAA
SEQ ID NO: 3721	POX4	GTTCCTGCCATATGACTTTTACAAAGAAAAACGTTAGTGTATCACAAG G
		Candida strain ATCC20336, Fwd. Primer, NdeI
SEQ ID NO: 3722	POX4	CTTCGAGATGCGGCCGCTTACTTGGACAAGATAGCAGCGGTTTCATC
		Candida strain ATCC20336, Rev. Primer, NotI
SEQ ID NO: 3723	Candida strain ATCC20336	ATGCCTACCGAACTTCAAAAAGAAAGAGAACTCACCAAGTTCAACCCAA AGGAGTTGAACTACTTCTTGAAGGTTCCCAAGAAAGATCCGAGATCAT CAGCAACATGGTCCAACAATGCAAAAAGACCCATCTTGAAGGTGCA CGCTTCATACTACAACCTTGACCAAAAGACCAACAAGAGAAGTCACCGCC AAGAAGATTGCCAGACTCTCCAGATACTTTGAGCAGGATACCCAGAC CAACAGGCCCCAGAGATTGTCGATCCTCGGTGTCTTTGACCCACAAGTC TTCACCAGAATCGGTGTCAACTGGGTTTGTGTTTCTGTGTCCGTG GTAACGGTACCAACTCCCAGTTCTTCTACTGGACCATAAATAAGGGTAT CGACAAGTTGAGAGGTATCTATGGTTGTTTTGGTATGACTGAGTTGGCC CACGGTTCCAACGTCCAAGGTATTGAAACCAACCGCCACTTTTGACGAA GACACTGACGAGTTTGTATCAACACCCCAACATTGGTGCCACCAAG TGGTGGATCGGTGGTGCTGCGCACTCCGCCACCCACTGCTCCGTCTAC GCCAGATTGAAGGTCAAAGGAAAGGACTACGGTGTCAAGACCTTTGTT GTCCCATTTGAGAGACTCCAACCACGACCTCGAGCCAGGTGTGACTGTT GGTGACATTGGTGCCAAGATGGGTAGAGACGGTATCGATAACGGTTGG ATCCAGTTCTCCAACGTCAAGATCCCAAGATTCTTTATGTTGCAAAAGTA CTGTAAGGTTTCCCGTCTGGGTGAAGTCACCATGCCACCATCTGAACA ATTGTCTTACTCGGCTTTGATTGGTGGTAGAGTCACCATGATGATGGAC TCCTACAGAATGACCAGTAGATTTCATCACCATTGCCTTGAGATACGCCA TCCACAGAAGACAATTCAAGAAGAAGGACACCGATACCATTTGAAACCAA GTTGATTGACTACCATTTGCATCAAAAGAGATTGTTCCCATTTCTGGCT GCCGCTTACTTGTCTCCCAAGGTGCCTTGTACTTAGAACAAACCATGA ACGCAACCAACGACAAGTTGGACGAAGCTGTCAGTGCTGGTGAAAAGG
	POX5	

SEQ ID NO:	Description	Sequence
		AAGCCATTGACGCTGCCATTGTCTGAATCCAAGAAATTGTTGTCGCTTC CGGTTGTTTGAAGTCCACCTGTACCTGGTTGACTGCTGAAGCCATTGAC GAAGCTCGTCAAGCTTGTGGTGGTCACGGTTACTCGTCTTACAACGGT TTCGGTAAAGCCTACTCCGACTGGGTTGTCCAATGTACCTGGGAAGGT GACAACAACATCTTGCCATGAACGTTGCCAAGCCAATGGTTAGAGAC TTGTTGAAGGAGCCAGAACAAAAGGGATTGGTTCTCTCCAGCGTTGCC GACTTGGACGACCCAGCCAAGTTGGTTAAGGCTTTCGACCACGCCCTT TCCGGCTTGCCAGAGACATTGGTGCTGTTGCTGAAGACAAGGGTTTC GACATTACCGGTCCAAGTTTGGTTTTGGTTTCCAAGTTGAACGCTCACA GATTCTTGATTGACGGTTTCTTCAAGCGTATCACCCCAGAATGGTCTGA AGTCTTGAGACCTTTGGGTTTCTTGTATGCCGACTGGATCTTGACCAAC TTTGGTGCCACCTTCTTGCAGTACGGTATCATTACCCAGATGTCAGCA GAAAGATTTCTCCGAGCACTTCCCAGCCTTGTGTGCCAAGGTTAGAC CAAACGTTGTTGGTTTGAATGATGGTTTCACTTGAATGACATGATGAC CAATGCTGCTATTGGTAGATATGATGGTAACGTCTACGAACACTACTTC GAAACTGTCAAGGCTTTGAACCCACCAGAAAACACCAAGGCTCCATACT CCAAGGCTTTGGAAGACATGTTGAACCGTCCAGACCTTGAAGTCAGAG AAAGAGGTGAAAAGTCCGAAGAAGCTGCTGAAATCTTGTCAGTTAA
SEQ ID NO: 3724	POX5	GTTCACTGCCATATGCCTACCGAACTTCAAAAAGAAAGAGAACTC
		Candida strain ATCC20336, Fwd. Primer, NdeI
SEQ ID NO: 3725	POX5	CTTCGAGATGCGGCCGCTTAAGTGGACAAGATTCAGCAGCTTCTTCG
		Candida strain ATCC20336, Rev. Primer, NotI
SEQ ID NO: 3726	Aco1 (AJ001299.1)	GTTCACTGCCATATGACAACCAACACATTCACCGATCCTC
		Yarrowia lipolytica, Fwd. Primer, NdeI
SEQ ID NO: 3727	Aco1 (AJ001299.1)	CTTCGAGATCTCGAGTCACTCATCGAGATCGCAAATTTTCATCGTC
		Yarrowia lipolytica, Rev. Primer, XhoI
SEQ ID NO: 3728	Aco2 (XM_505264)	GTTCACTGCCATATGAACCCCAACAACACTGGCACC
		Yarrowia lipolytica, Fwd. Primer, NdeI
SEQ ID NO: 3729	Aco2 (XM_505264)	CTTCGAGATGCGGCCGCCTATTCTCATCAAGCTCGCAAATGTCATC
		Yarrowia lipolytica, Rev. Primer, NotI
SEQ ID NO: 3730	Aco3 (XM_503244)	GTTCACTGCCATATGATCTCCCCAACCTCACAGCTAAC
		Yarrowia lipolytica, Fwd. Primer, NdeI

SEQ ID NO:	Description	Sequence
SEQ ID NO: 3731	Aco3 (XM_503244)	CTTCGAGATGCGGCCGCCTATTCTCGTCCAGCTCGCAAATG
		Yarrowia lipolytica, Rev. Primer, NotI
SEQ ID NO: 3732	Aco4 (XM_504475)	GTTCACTGCCATATGATCACCCCAAACCCCGCTAAC
		Yarrowia lipolytica, Fwd. Primer, NdeI
SEQ ID NO: 3733	Aco4 (XM_504475)	CTTCGAGATCTCGAGTTACTGAATATCCTCGGGCTCCATGG
		Yarrowia lipolytica, Rev. Primer, XhoI
SEQ ID NO: 3734	Aco5 (XM_502199)	GTTCACTGCCATATGAACAACAACCCCAACCAACGTGATC
		Yarrowia lipolytica, Fwd. Primer, NdeI
SEQ ID NO: 3735	Aco5 (XM_502199)	CTTCGAGATGCGGCCGCCTACTCGTCCAGGTCGCAAATCTC
		Yarrowia lipolytica, Rev. Primer, NotI
SEQ ID NO: 3736	Aco6 (XM_503632)	GTTCACTGCCATATGCTCTCTCAACAGTCCCTCAACAC
		Yarrowia lipolytica, Fwd. Primer, NdeI / NcoI
SEQ ID NO: 3737	Aco6 (XM_503632)	CTTCGAGATCTCGAGCTACTCATCCTCAAGAGAGCAAATTTCTC
		Yarrowia lipolytica, Rev. Primer, NcoI / XhoI
SEQ ID NO: 3738	Aco1 (NM_001136902)	GTTCACTGCCATATGGACGCATCGGCGGAGGTGG
		Zea mays, Fwd. Primer, NdeI / EarI
SEQ ID NO: 3739	Aco1 (NM_001136902)	CTTCGAGATCTCGAGCTAGAGCCTGGAGAGCTTGAGCTGC
		Zea mays, Rev. Primer, EarI / XhoI
SEQ ID NO: 3740	Aco1b (NM_001175167)	GTTCACTGCCATATGGCGGAAGTGGAACCTCGC
		Zea mays, Fwd. Primer, NdeI / BstXI
SEQ ID NO: 3741	Aco1b (NM_001175167)	CTTCGAGATCTCGAGCTAGAGCCTGGAGAGCTTGAGCTGC
		Zea mays, Rev. Primer, BstXI / XhoI

SEQ ID NO:	Description	Sequence
SEQ ID NO: 3742	Aco2 (NM_001158552)	GTTCCTGCCATATGGACCTCACCTCGCCGTCGCC
		Zea mays, Fwd. Primer,
SEQ ID NO: 3743	Aco2 (NM_001158552)	CTTGCGGCCGCTCAGTGGCTCCCGGTTGACAGTGCA
		Zea mays, Rev. Primer,
SEQ ID NO: 3744	Aco4 (NM_001156834)	GTTCCTGCCATATGATGGCCGGGAAACGAGTTACGGG
		Zea mays, Fwd. Primer,
SEQ ID NO: 3745	Aco4 (NM_001156834)	CTTCGAGATCTCGAGTCACAGCCGGGCTTTCGCTGG
		Zea mays, Rev. Primer,
SEQ ID NO: 3746	ACOX2 (XM_001386762)	GTTCCTGCCATATGATCCTGTTGCCCAAAGAGCTCC
		Scheffersomyces stipitis, Fwd. Primer, NdeI / SalI
SEQ ID NO: 3747	ACOX2 (XM_001386762)	GTTCCTGCGCGGCCGCTAGCGGGACAATATCTTGGCAGCTTCG
		Scheffersomyces stipitis, Rev. Primer, SalI / NotI
SEQ ID NO: 3748	DEHA2D17248p (XM_459235)	GTTCCTGCCATATGGTTAGTGCTACTAATACAGTGAATTCAGG
		Debaryomyces hansenii, Fwd. Primer, NdeI
SEQ ID NO: 3749	DEHA2D17248p (XM_459235)	CTTCGAGATCTCGAGTTATTTGGATAAGATCTTAGCAGTTTCAGTAGAC TTTTC
		Debaryomyces hansenii, Rev. Primer, XhoI
SEQ ID NO: 3750	ACX1 (NM_117778)	GTTCCTGCGATTAATATGGAAGGAATTGATCACCTCGCCG
		Arabidopsis thaliana, Fwd. Primer, AseI
SEQ ID NO: 3751	ACX1 (NM_117778)	CTTCGAGATGTGCGACTCAGAGCCTAGCGGTACGAAGTTGC
		Arabidopsis thaliana, Rev. Primer, SalI
SEQ ID NO: 3752	ACX2 (NM_001037068)	GTTCCTGCCATATGGAATCGCGGCGAGAGAAGAATCC
		Arabidopsis thaliana, Fwd. Primer, NdeI

SEQ ID NO:	Description	Sequence
SEQ ID NO: 3753	ACX2 (NM_001037068)	CTTCGAGATGTGCGACTTATACAAGAAAACAAACCTTAGCTTTGTTAGGC GC
		Arabidopsis thaliana, Rev. Primer, Sall
SEQ ID NO: 3754	ACX2b (NM_125910)	GTTCAGTGCATATGGAATCGCGGCGAGAGAAGAATCC
		Arabidopsis thaliana, Fwd. Primer, NdeI
SEQ ID NO: 3755	ACX2b (NM_125910)	CTTCGAGATGTGCGACTTAGAATCCAACAACCTTGAGTATACTGGGAATAA G
		Arabidopsis thaliana, Rev. Primer, SstI
SEQ ID NO: 3756	ACX3 (NM_100511)	GTTCAGTGCATTAATATGTGCGGATAATCGTGCACTCCGACG
		Arabidopsis thaliana, Fwd. Primer, AseI
SEQ ID NO: 3757	ACX3 (NM_100511)	CTTCGAGATGTGCGACCTAAACTGAAGACCAAGCATTGGCTTCG
		Arabidopsis thaliana, Rev. Primer, SALI
SEQ ID NO: 3758	ACX5 (NM_129124)	GTTCAGTGCATATGGAGAGAGTTGATCACCTTGCCGATG
		Arabidopsis thaliana, Fwd. Primer, NdeI / EarI
SEQ ID NO: 3759	ACX5 (NM_129124)	GTTCAGTGCAGCGGCCGCTTAGAGTTTGGCAGAGCGGAAGCGTTG
		Arabidopsis thaliana, Rev. Primer, EarI / NotI
SEQ ID NO: 3760	aco2 (XM_003525015)	GTTCAGTGCATATGCAAACTCCGAAGTGTGAAGCA
		Glycine max, Fwd. Primer, NdeI
SEQ ID NO: 3761	aco2 (XM_003525015)	CTTGCGGCCGCTCAAAAACCGACGTATTGAGTGTAT
		Glycine max, Rev. Primer, NotI
SEQ ID NO: 3762	aoxA (XM_659264)	GTTCAGTGCATATGCCAAATCCACCTCCCGCCTGG
		Aspergillus nidulans, Fwd. Primer, NdeI

SEQ ID NO:	Description	Sequence
SEQ ID NO: 3763	aoxA (XM_659264)	CTTCGAGATCTCGAGTCACAGCTTGCTCTTAATCTCCCCCG
		Aspergillus nidulans, Rev. Primer, XhoI
SEQ ID NO: 3764	AcoI (NM_017340)	GTTCACTGCCATATGAACCCAGACTTGAGAAAGGAAAGAGC
		Rattus norvegicus, Fwd. Primer, NdeI
SEQ ID NO: 3765	AcoI (NM_017340)	CTTCGAGATCTCGAGCTACAACCTGGATTGCAATGGCTTCAAGTGC
		Rattus norvegicus, Rev. Primer, XhoI
SEQ ID NO: 3766	AcoII (1IS2_A)	GTTCACTGCCATATGAACCCAGACTTGAGAAAGGAAAGAGC
		Rattus norvegicus, Fwd. Primer, NdeI
SEQ ID NO: 3767	AcoII (1IS2_A)	CTTCGAGATCTCGAGCTACAACCTGGATTGCAATGGCTTCAAGTGC
		Rattus norvegicus, Rev. Primer, XhoI
SEQ ID NO: 3768	Aco (Cucsa.029560.1)	GTTCACTGCCATATGGCTTCGCCGCGCGAGTC
		Cucumis sativus, Fwd. Primer, NdeI
SEQ ID NO: 3769	Aco (Cucsa.029560.1)	CTTCGAGATCTCGAGTTAGAAGCCAACATACTGCGTATACTGCG
		Cucumis sativus, Rev. Primer, XhoI
SEQ ID NO: 3770	Aco (BAE47462)	GTTCACTGCCATATGACAGAAGTAGTGACCGCGCATC
		Arthrobacter ureafaciens, Fwd. Primer, NdeI
SEQ ID NO: 3771	Aco (BAE47462)	CTTCGAGATCTCGAGCTAGCGGGACTTGCCGGCC
		Arthrobacter ureafaciens, Rev. Primer, XhoI
SEQ ID NO: 3772	Aco (YP_003571780)	GTTCACTGCCATATGCTCGATACCGACTCGCCACG
		Salinobacter rubber, Fwd. Primer, NdeI
SEQ ID NO: 3773	Aco (YP_003571780)	CTTCGAGATCTCGAGCTCGAGTCATTTCTGGGCCGGG
		Salinobacter rubber, Rev. Primer, XhoI

SEQ ID NO:	Description	Sequence
SEQ ID NO: 3774	Aco (YP_290295.1)	GTTCACTGCCATATGCCCCAGCACGGCGATACAG
		Thermobifida fusca, Fwd. Primer, NdeI
SEQ ID NO: 3775	Aco (YP_290295.1)	CTTCGAGATCTCGAGCTCGAGTCACGTCTCCGCGC
		Thermobifida fusca, Rev. Primer, XhoI
SEQ ID NO: 3776	Aco (NC_009441.1)	GTTCACTGCCATATGAAACCAGCTAAACTTCAAGCCTTTACTCC
		Flavobacterium johnsoniae, Fwd. Primer, NdeI
SEQ ID NO: 3777	Aco (NC_009441.1)	CTTCGAGATCTCGAGCTAAACTGCAATTGGCGCTGCTAAACAG
		Flavobacterium johnsoniae, Rev. Primer, XhoI
SEQ ID NO: 3778	POX1 >gil50554589 ref XP_504703.1 YALI0E32835p [Yarrowia lipolytica]	MAKERGKTQFTVRDVTNFLNGGEEETQIVEKIMSSIERDPVLSVTADYDCN LQQARKQTMERVAALSPYLVTDTEKLSLWRAQLHGMVDMSTRTRLSIHNN LFIGSIRGSGTPEQFKYWVKKGAVAVKQFYGCFAMTELGHGSNLKGLETT ATYDQSDSQFIINTPHIGATKWWIGGAAHTSTHCVCFAKLIVHGKDYGTRN FVVPLRNVHDHSLKVGVSIGDIGKKMGRDGVNNGWIQFTNVRIQRQNMLM RYAKVSDTGWVTKPALDQLTYGALIRGRVSMIADSFHVSKRFLTIALRYACV RRQFGTSGDTKETKIIDYPYHQRRLLPLLAYCYAMKMGADAEAKTWIETTD RILALNPNDPAQKNDLEKAVTDTKELFAASAGMKAFTTWGCAKIIDEQRQA CGGHGYSGYNGFGQGYADWVQCTWEGDNNVLCLSMGRGLVQSALQIL AGKHV GASIQYVGDKSKISQNGQGTPREQLLSPEFLVEAFRTASRNILRT TDKYQELVKTLNPDQAFEELSQQRFQCARIHTRQHLISSFYARIATAKDDIK PHLLKLANLFALWSIEEDTGIFLRENILTPGDIDLINSLVDELCAVVRDQVIGL TDAFGLSDFFINAPIGSYDGNVYEKYFAKVNQQNPATNPRPPYYESTLKPFLFREEEDDEICDLDE
SEQ ID NO: 3779	POX2 >gil50555712 ref XP_505264.1 YALI0F10857p [Yarrowia lipolytica]	MNPNNGTGIEINGKEYNTFTEPPVAMAQERAKTSFPVREMTYFLDGGEKN TLKNEQIMEEIERDPLFNNDNYYDLNKEQIRELTMERVAKLSLFFVRDQPED DIKKRFALIGIADMGTYTRLGVHYGLFFGAVRGTGTAEQFGHWISKGAGDL RKFYGCFSMTELGHGSNLAGLETTAIYDEETDEFIINTPHIAATKWWIGGAA HTATHTVVFARLIVKGKDYGVKTFVVLRLNINDHSLKVGISIGDIGKKMGRD GIDNGWIQFTNVRIQRQNLLMKYTKVDREGNVTQPPLAQLTYGSLITGRVS MASDSHQVGKRFITIALRYACIRRFSTTPGQPETKIIDYPYHQRRLLPLLA YVYALKMTADEVGALFSRTMLKMDDLKPDDKAGLNEVVSDVKELFSVSAG LKAFSTWACADVIDKTRQACGGHGYSGYNGFGQAYADWVQCTWEGDN NILTLSAGRALIQSAVALRKGEVGNVAVSYLKRYKDLANAKLNGRSLTDPK VLVEAWEVAAGNIINRATDQYEKLIGELNADQAFEVLSQQRFQAAKVHTR RHLIAAFFSRIDTEAGEAIKQPLLNLALLFALWSIEEDSGLFLREGFLEPKDID TVTELVNKYCTTVREEVIGYDFAFNLSDYFINAPIGCYDGDAYRHYFQKVN EQNPARDPRPPYYASTLKPFLFREEEDDDICELDEE

SEQ ID NO:	Description	Sequence
SEQ ID NO: 3780	POX3 >gi 50551539 ref XP_503244.1 YALIOD24750p[Yarrowia lipolytica]	MISPNTANVEIDGKQYNTFTEPPKALAGERAKVKFPIKDMTEFLHGGEENVTMIERLMTELERDPVLNVSGDYDMPKEQLRETAVARIAALSGHWKDKTEKEALLRSQLHGIVDMGTRIRLGVHTGLFMGAIRGSGTKEQYDYWVRKGAA DVKGFYGCFAAMTELGHGNSVAGLETTATYIQDTEFIINTPNTGATKWWIGGAAHSATHTACFARLLVDGKDYGVKIFVVQLRDVSSHSLMPGIALGDIGKKMGRDAIDNGWQFTNVRIPRQNMLMKYAKVSSGKVSQPPLAQLTYGALIGGRVTMIADSFVSQRFITIALRYACVRRQFGTTPGQPETKIIDYPYHQRRLPLLAFTYAMKMAADQSQIQYDQTTDLLQTIDPKDKGALGKAIVDLKELFASAGLKAFTTWTCANIIDQCRQACGGHGYSGYNGFGQAYADWV/QCTWEGDNNVLCLSMGRGLIQSCLGHRKGKPLGSSVGYLANKGLEQATLSGRDLKDPKVLIEAWEKVANGAIQRATDKFVELTKGGLSPDQAFEELSQQRFQCAKIHTRKHLVTAFYERINASAKADV KPYLINLANLFTLWSIEEDSGLFLREGFLQPKDIDQVTELVNHYCKEVRDQVAGYTDAFGLSDWFINAPIGNYDGDVYKH YFAKVNQQNPAQNPRPPYYESTLRPFLFREDEDDDICELDEE
SEQ ID NO: 3781	POX4 >gi 50554133 ref XP_504475.1 YALIOE27654p[Yarrowia lipolytica]	MITPNPANDIVHDGKLYDTFTEPPKLMAQERAQLDFDPRDITYFLDGSKEETELLESMLMYERDPLFNNQNEYDESFETLRERSVKRIFQLSKSIAMDPEPMSFRKIGFLGILDMGYARLGVHYALFCNSIRGQGTDPQLMYWLDQGAMVKGIFYGCFAAMTEMGHGNSLRLETIATFDKETDEFIINTPHVGATKWWIGGAAHTATHTLAFARLQVDGKDYGVKSFVVPLRNLDHSLRPGIATGDIGKKMGRDAVDNGWQFTNVRVPRNYMLMKHTKVL RDGT V KQPPLAQLTYGSLITGRVQMTTDSHNVS KKF LTIALRYATIRRQFSSTPGEPETRLIDYLYHQRRLPLMAYSYAMKLAGDHVRELFASQEKAE SLKEDDKAGVESYVQDIKELFSVSAGLKAATTWACADIIDKARQACGGHGY SAYNGFGQAFQDWV/QCTWEGDNTVLTLSAGRALIQSALVYRKEGKLG NATKYLSRSKELANAKRNGRSLEDPKLLVEAWEAVSAGAINAATDAYEELSKQGVSVDECFEQVSQERFQAARIHTRRALIEAFYSRIATADEKVKPHLIPLANLFALWSIEEDSALFLAEGYFEPEDIIEVTSLVNKYCGIVRKNVIGYTDAFNLSDYFINAAIGRYDGDVYKNYFEKV KQQYPPEGGKPHYIEDVMKPFLLHRERIPDVPMEPEDIQ
SEQ ID NO: 3782	POX5 >gi 50549457 ref XP_502199.1 YALIOC23859p[Yarrowia lipolytica]	MNNNPTNVILGGKEYDTFTEPPAQMELERAKTQFKVRDVTNFLTGSEQETLLTERIMREIERDPVLNVAGDYDADLPTKRRQAVERIGALARYLPKDSEKEAILRGQLHGIVDMGTRTRIAVHYGLFMGAIRGSGTKEQDYWVAKGAATLHKFYGCFAAMTELGHGNSVAGLETTATLDKDTDEFIINTPNSGATKWWIGGAAHSATHTACLARLIVDGKDYGVKIFIVQLRDLNSHSLNLNGIAIGDIGKKMGRDAIDNGWQFTDVRIPRQNMLMRYDRVSRDGEVTTSELAQLTYGALLSGRVTMIAESHLLSARFLTIALRYACIRRQFGAVPDKPETKLIDYPYHQRRLPLLAYTYAMKMGADAEQQYQNSSFGALLKLNPKDAEKFAVATADLKALFASSAGMKAFTTWAAAIIIDECRQACGGHGYSGYNGFGQAYADWV/QCTWEGDNNVLCLSMGRSLIQSCIAMRKKKGHVGSVEYLQRRDELQNA RVDNKPLTDPAVLITAWEKVACEAINRATDSFIKLTQEGLSPDQAFEELSQQRFECARIHTRKHLITSFYARISKAKARVKPHLTVLANLFAVWSIEEDSGLFLREGCFEPAE MDEITALVDELCCAREQVIGFTDAFNLSDFFINAPIGRFDGDAYKHYMDEVKAANNPRNTHAPYYETKLRPFLFRPDEDEEICDLDE
SEQ ID NO: 3783	POX6 >gi 50552444 ref XP_503632.1 YALIOE06567p[Yarrowia lipolytica]	MLSQQSLNTFTEPPVEMARERNQTSFNPRLLTYFLDGGKENTLLMDRLMQEYERDPVFRNEG DYDITDVAQSRELAFKRIAKLIEYVHTDDEETYLYRCMLLGQIDMGAFARYAIHHGVWGGAIRGAGTPEQYEFVWKKGSLSVKKFYGSFSMTELGHGNSLVGLETTATLDKNADEFVINTPNVAATKWWIGGAADTATHTAVFARLIVDGEDHGKVFVQLRDVETHNLMPGIAIGDCGKKMGRGTDNGWQFTHVRI PRQNMLMRYCHVDS DGNVTEPMM AQMAYGALLAGRVGMAMDSYFTSRKFLTIALRYATIRRAFAAGGGQETKLIDYPYHQRRLPLMAQTYAIKCTADKVRDQFVKVTDMLLNLDVSDQEAVPKAIAEAKELFSVSAGVKATTTWACAHTIDQCRQACGGHGY SAYNGFGGRAYSDWVIQCTWEGDNNI

SEQ ID NO:	Description	Sequence
		LCLSAGRALVQSNRAVRAGKPIGGPTAYLAAPAGSPKLAGRNLYDPKVMIGAWETVSRALINRTTDEFVLAKKGLSTAQAYEELSQQRFCLTRIHTRLYMVKNFYERIAEEGTEFTKEPLTRLANLYAFWSVEEEAGIFLREGYITPQELKYISAEIRKQLLEVRKDVIGYTDFAFNVPDFFLNSAIGRADGDVYKNYFKVNTQNPPQDPRPPYYESVIRPFLFRKDEDEEICSLEDE
SEQ ID NO: 3784	pAA298 (Nucleic Acid Seq.)	gatctggaatccctcggtcggtcttgggggtgggggcattcttctggtcttgggaacgccaacgcttgt igtgtgggtctgaacacggactgctcgaaaaagtaccagatgatgcttacctctgctggcttccacaa gtatgggagggcattgacagcgactgcttggctaacagcagctgctggcaattaaatttggcttccaa taactgactaccaaggatggcagcagcggtatttctaactctgacatgttctcgtacgtagtagtaatga agggaagggtggaalaatacaaggcgaaattctgcagatatccatcacactggcgccgctcgagcatg catctagagggcccaattcgccctatagttagtctgtattacaaltcactggcgtgtgtttacaacgtctga ctgggaaaaccctggcgttaaccacctaactcgcttgcagcacatcccccttctcgccagctggcgttaata gcgaagaggcccgacccgatgccttcccaacagttgcgcagcctatacgtacggcagtttaagggtta cacctataaaagagagcggttatcgctgttgggtgacagagtgatattatgacacgcgggggc gacggatgggtatccccctggccagtgcaagctgctgctgcagataaagtctcccgtaactttacccgtg gtgcatactggggatgaaagctggcgcatgatgaccacgataggccagtgctgggtctcgttatcg gggaagaagtggcgtatctcagccacggcgaaaatgacatcaaaaacgccattacctgatgttctggg gaataataatgtcaggcatgagattatcaaaaaggatcttccctagatcttttaccgtagaagccagt ccgcagaaacggtgctgaccccgatgaatgtcagctactgggtatctctggacaagggaacacgcaa gcgcaagagaaaacaggtagctgtcagtggtgctacatggcgatagctagactggcggtttatgga cagcaagcgaaccggaattgccagctggggcgccctctgtaagggtgggaagccctgcaaaagtaac ctggatggccttctcgccgccaaggatctgatggcgaggggatcaagctctgatcaagagacaggatg aggatcgttctcgtatgaacaagatggattgcagcaggcttctcggccgctgggtggagaggctat icggctatgactgggcacaacagacaatcggctgctctgatgcccgtgttccggtctcagcgaggg gcgcccgttctttgtcaagaccgaacctgctcggtgccctgaatgaactgcaagacgagcgagcgcg ctatcgtggtcggcacgacggcggttcttgcgcagctgtgctcagctgttgcactgaagcggaaggg actggctgctattgggcgaagtgcgggggacaggatctcctgtcatctacacctgtcctcggagaaagtat ccatcatgctgatgcaatgcggcggtgcatacgtctgatccggctacctggccattgcaccacaagc gaaacatcgcatcgagcgagcagctacatcggtggaagccggtcttgcgtacaggatgatctggacga agagcatcaggggtcgcgccagccgaactgttgcagggcicaaggcgagcatgccgaaggcga ggatctcgtctgacccatggcgatgcctgttgcgaatatcatggtgaaaaatggcggtttctggattc atcgactgtggtcggcggtggtggcgacccgtatcaggacatagcgttggctacccgtgatattgtga agagcttggcggaatgggtgacccgttctcgtgttctacggtatcgccgtcggatctcagcagcgca icgcttctatcgcttctgaagatttctgaatttataacgcttacaatttctgatgcggtatttctcttac gcatctgtgcggtatttcaaccgcatacaggtggcacttttcggggaaatgtgcggaacccctatttgtt tatttttctaaatacattcaaatgtatccgctcatgagacaataaccctgataaaatgttcaataatagcac gtgaggagggccaccatggccaagttgaccagtgccgttccggtgtccacgcgcgcagctgcggg agcgtctgagttctggacggacggctcgggttctccgggactctgtgagagacgacttgcgggtgtg gtccgggacgagctgacccgttcatcagcgcggtccaggaccaggtgggtgcgggacaacaccctggc ctgggtgtgggtgcgcggcctggacgagctgtacgcgagtggttgcgaggtctgttccagaaacttcg ggacgcctcggggcgccatgaccgagatcggcgagcagcgtggggggcgggagttcgcctgcg cgacccggccggcaactgcgtgcactctgtggccgaggagcaggactgacacgtgtctaaaacttctttt taatttaaaggatctaggtgaagatcttttgataatctcatgacaaaatcccttaacgtgagtttctgtc cactgagcgtcagaccctgagaaaagatcaaaggatcttctgagatcctttttctgcgcgtaactgtct gcttgcaaaacaaaaaacaccgctaccagcgggtgtgtgttgcgggatcaagagctaccaactcttttc ogaaggtaactggctcagcagagcgagataccaaatactgtccttctagttagcgttaggcca ccactcaagaactctgtacacgcctacatacctcgtctgclaatctgttaccagtggtgtctgcaggt ggcgataagtcgtgtcttaccgggttgactcaagacgatagttaccggataaggcgagcgttcgggt gaacgggggggtctgtgcacacagcccagcttgagcgaacgacctacaccgaactgagatacctaca gcgtgagctatgagaaagcgccacgcttccgaaggagaaaggcggaacaggtatccgttaagcgg cagggctcggaacaggagagcgacagggagcttccaggggaaacgcctgtgtatctttagtctgt cgggttctgccacctctgactgagcgtcgattttgtgatgctcgtcagggggcgagccctatgaaaaa

SEQ ID NO:	Description	Sequence
		cgccagcaacgcgcccttttacggctcctgggctttgtctggcctttgtcacatgttcttctcggtatccc ctgattctgttgataaccgtaaccgctttgagtgagctgataccgctcgccgagccgaacgacccgag cgacgagagtcagtgagcgagggaagcggaagagcgcccaatacgcaaacgctctccccgcgcgt iggccgattcattaatgcagctggcagcagaggtttccgactggaagcgggcagtgagcgcaacgc aattaatgtgagttagctcactcattagcgaccccgagctttacactttatgcttccggctcgatgtgtgtg aattgtgagcgataacaatttcacacaggaacagctatgacctgattacgccaagctattaggga cactatagaatactcaagctatgcatcaagcttggtaccgagctcggaaccactagtaacggccgcag gtgctggaattcgccctccgttaaacaaaaatcagctctgtaaaaaagggtctaaataaatattctgtctagt gtacacattctcccaaatagtgaaatccagctctacaattggctttaccggtaacaataaagaccaat cgcttcagtcactcctgtgactgtgcattcaaccagtagatcaaatcctgtcggtcttcaccccttggca acttttggttctgaattggttcttaggaacctgtgtgtgtgcgaagcctcaatgtgccaatctgtgaaacg caggttgagcatacacaggcaagaactgtgtcagtgagagtgaaatcaatttcaagatttctcatgtgctc aactcgtcctggcttcacagacggcaaacaggctctacctcgtgtgttggcaccttgacaccgacaac gacggactgtctcaaggccttgagcccatcaattcgttctcgacctcggtggcgagacgttttcggactt ccaacggaaagtgtcacctaattctgtcgacaaagtacaacaattgtcctcgctcatcttcaacaagtca cggtatctgaccacgcgtcacctttttgaaaacattgtgtgaggtttgtgtgtgtgtgtgtgtgtgtgt aataaccctggaaggatttctgcagctgt agcgccctcggtacagaacccggctgt atttctgt caccgtactgcaagttgt gtggaatctgcgt tctgtctgt ggcaccacataatctggcctgggtccagaaacgaagtagcggagaatttctgggacagcgagacacag ccaccgacaataaagtaggacacaccccaacatggccggtgtgtgtgtgtgtgtgtgtgtgtgtgtgt gtcaagacgttctgaattcgaagtcaatctcatgatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt caggacatgataccgcttttggcaaacgggt gtcagtatcgggt aggcaaacctgtcaatgt gggaatcacagtcgggtcaacgaaaacttgcgaagcgttgacaatcttaagacagtggaatgaatggt tgtcctgt tgaaaagcggctgt atgtatgagtaactcaaaacacatgt atccactagtaacggccgcagtggt taaaggatctacggt atcaagaacatgt aggcgtcatgtgataaaatggatgagatttatcccaattgaagaaagagtttatgaaagtgtgtcaacc agaagctaaacagggaagaagcaaacgaagaggtgaacaagaagaagaaggtaataaagtattttg tattatatacaaaacaaagtaagggaatacagatttatcaataaattgccatactgacagtgagatatct catccattcccaactcccaagaaaaaaaagtgaaaaaaaatcaaaccaagatcaacct ccccatcatcatcgtcatcaaaccccgactcaattcgcaatggttagcacaacacatacacagaag ggcatcagcacacccctccnaggttgcacaacgtttattccgcttaattggagtcacaaaagaccaacct tgcgcctcgatcgacgtgacacaccccgagttccttctcgctcagacagctcggtccccacatctg tctcgtgaagacgcacatcgatnctcagactcagctacgagggcacgattgagccgtgtgtgtgtgtgt tgcagagcgccacgggttctgtatctcagggacaggaagtttctgtatctggaaacacgggtgtgtgt agtaacctcgggggtatagcagctcgcggt gaagggcgctgt tngggcgagttgtcagtaaaaggt agagtgatcgggagttcgt attggtatcatcagcgt gttgatgaggt cctgaggt agaataaattgttaataaattaggtctatatacatataactaagctttaggacgtcattgtgtgtgtgtgt

SEQ ID NO:	Description	Sequence
		gtctgctagtttagttctcatgatttcgaaaaccaataacgcaatggatgtagcagggatgggtgtagtgcg ttctgacaaaccagagtagcgcgcctcaaaccacgtcacattcgcccttgcttcatccgcatcactgc itgaaggatatccacgtacgagttgtaatacacctgaagaacggctcgtctacggtcgacgacgggtaca acgagaattgtattgaattgatcaagaacatgatcttggtgttacagaacatcaagttcttgaccagactg agaatgcacagatatacaaggcgtcatgtgataaaatggatgagatttatccacaattgaagaagagtt tatggaaagtggtaaccagaagctaaacaggaagaagcaaacgaagagggtgaaacaagaagaa gaaggtaaataagttttgtattatatacaaaacaaagtaaggaatacagattatacaataaattgccat actagtcacgtgagatatctcatccattccccaaactccaagaaaaaaaaaagtgaaaaaaaaaatc aaacccaaagatcaacctccccatcatcatcgtcatcaaacccccagctcaattcgagagctcggtac ccggg
SEQ ID NO: 3785	Candida strain ATCC20336 Thioesterase PTE1	MRGMEKPHSLFRRMSTAPFAIQPPISILSATLITSEFFVYSLYNFFHFIDLF YYYYTHIYPHRAMIENISGNGNYPQNHEVDLEKEFGVEKIGINLYRDKSPIP KPDRRSRGAYGGYLAGQALLVAMKSTPPEYRPHSFHSYFIKAVNDKETLE WRVEETSNGRNYANRSLQAFQAGNLVYTANVSLTKKNSAKKAEATGVK PFEFQGKPEHWEFEKHKRDDLPLATPSSSLLIYHKFFPEVVSLEASKEEESK PAADRELSWYFKWGINNEEGHHQPLVNLNSDYQYVGMAALTDVAYLNRL LRILRVEDADHTQLVHYFSVSLDHTMYFHDDDFDVTKWMGFTFKVTRFSH NRALCQGEVYNDKGVHVCTIVQEGLMMLNGLEEGAKL*
SEQ ID NO: 3786	Candida strain ATCC20336 Thioesterase PTE2	MICVFFPTSTFTTAHKFVSNLQSFFLSQQPHTTSYTMPTFNYKDGETIDVQ KEFGVVETAPNKYVGKPLVKPMFHVKGVFGGNLAGQALLVAMKSVGPD FSPHSLHSYFIRAGSDQTPVEWTVQAISDGNSFCNRFIKGVQNGQVIYIAN VSLTKRNSAADAMKKYEEYHAQIRQKKGKDGDADEEDEDDEDDNAPAK PFGFQTPSHKWIKDRDLKLPVSDMESNLLLYKLPPEFVSLKSSTEEESL PVSERRMGALAKWGIENEQGFNQPLTNLDKSFQYVGLANITDGLYLGTLN RILRIDDLTLDERATNYFSVSLDHVIYFHDDDFDVTKWMGFTFRCSRYSHN RVIFEGEIYSDKGVQVASIIQEGLVRFKDGYLKNAL

Example 44: Examples of certain non-limiting embodiments

A1. A genetically modified yeast, comprising an active, modified endogenous acyl-coA oxidase
5 polypeptide or an active, modified endogenous acyl-coA dehydrogenase polypeptide, which yeast
is capable of producing a diacid from a feedstock comprising one or more components from a
vegetable oil.

A2. The genetically modified yeast of embodiment A1, wherein the yeast is a genetically modified
10 Candida spp. yeast.

A2.1. The genetically modified yeast of embodiment A2, wherein the Candida spp. yeast is chosen
from *C. tropicalis* and *C. viswanathii*.

A2.2. The genetically modified yeast of embodiment A1, wherein the *Candida* spp. yeast is a genetically modified ATCC20336 yeast.

5 A2.3. The genetically modified yeast of any one of embodiments A2 to A2.2, wherein the endogenous acyl-coA oxidase polypeptide is a POX4 polypeptide.

A2.4. The genetically modified yeast of embodiment A2.3, wherein the POX4 polypeptide comprises a modified amino acid sequence of SEQ ID NO: 30.

10 A2.5. The genetically modified yeast of embodiment A2.3 or A2.4, wherein the POX4 polypeptide comprises an amino acid modification at one or more amino acid positions chosen from 88, 90, 96, 98, 99, 100, 102, 103, 302, 309, 310, 473, 474, 475, 476, 477, 478, 479, 480, 481, 482, 483, 484, 485, 486, 487, 488, 489, 490, 491, 492, 493, 494, 495, 496, 497, 498, 499, 500, 501, 502, 503, 504 and 505.

15 A2.6. The genetically modified yeast of any one of embodiments A2 to A2.2, wherein the endogenous acyl-coA oxidase polypeptide is a POX5 polypeptide.

20 A2.7. The genetically modified yeast of embodiment A2.6, wherein the POX5 polypeptide comprises a modified amino acid sequence of SEQ ID NO: 32.

A2.8. The genetically modified yeast of embodiment A2.6 or A2.7, wherein the POX5 polypeptide comprises an amino acid modification at one or more amino acid positions chosen from 81, 82, 83, 84, 85, 86, 88, 93, 94, 95, 96, 98, 102, 284, 287, 290, 291, 292, 294, 295, 436, 453, 454, 455, 456, 25 457, 458, 459, 460, 461, 462 and 463.

A2.9. The genetically modified yeast of any one of embodiments A2 to A2.2, wherein the acyl-coA dehydrogenase polypeptide is chosen from ACAD, VLCAD, LCAD, MCAD and SCAD polypeptides.

30 A2.10. The genetically modified yeast of embodiment A2.9, wherein the acyl-coA dehydrogenase polypeptide comprises a modified amino acid sequence of SEQ ID NO: 3685.

A2.11. The genetically modified yeast of embodiment A2.9 or A2.10, wherein the the acyl-coA dehydrogenase polypeptide comprises an amino acid modification at VLCAD position 461.

5 A2.12. The genetically modified yeast of any one of embodiments 2.5, 2.8 and 2.11, wherein at least one of the amino acid modifications is an amino acid substitution.

A2.13. The genetically modified yeast of embodiment A2.12, wherein at least one of the one or more amino acid substitutions is conservative.

10 A2.14. The genetically modified yeast of embodiment A2.12, wherein at least one of the one or more amino acid substitutions is not conservative.

A3. The genetically modified yeast of embodiment A1, wherein the yeast is a genetically modified *Yarrowia* spp. yeast.

15

A3.1. The genetically modified yeast of embodiment A3.1, wherein the *Yarrowia* spp. yeast is *Y. lipolytica*.

20 A3.2. The genetically modified yeast of embodiment A3 or A3.1, wherein the endogenous acyl-coA oxidase polypeptide is chosen from a POX1 polypeptide, POX2 polypeptide, POX3 polypeptide, POX4 polypeptide, POX5 polypeptide or POX6 polypeptide.

A3.3. The genetically modified yeast of embodiment 3.2, wherein the endogenous acyl-coA oxidase polypeptide is chosen from SEQ ID NOs: 3778 to 3783.

25

A4. The genetically modified yeast of embodiment A1, wherein the yeast is a genetically modified *Pichia* spp. yeast.

30 A4.1. The genetically modified yeast of embodiment A4.1, wherein the *Pichia* spp. yeast is chosen from *P. pastoris*, *P. membranifaciens*, *P. kluyveri*, *P. guilliermondii*, *P. heedii* and *P. subpelliculosa*.

A5. The genetically modified yeast of embodiment A1, wherein the yeast is a genetically modified *Saccharomyces* spp. yeast.

- A5.1. The genetically modified yeast of embodiment A5.1, wherein the *Saccharomyces* spp. yeast is chosen from *S. cerevisiae*, *S. bayanus*, *S. pastorianus* and *S. carlsbergensis*.
- 5 A6. The genetically modified yeast of embodiment A1, wherein the yeast is a genetically modified *Kluyveromyces* spp. yeast.
- A6.1. The genetically modified yeast of embodiment A6.1, wherein the *Kluyveromyces* spp. yeast is chosen from *K. lactis* and *K. marxianus*.
- 10 A7. The genetically modified yeast of any one of embodiments A1 to A6.1, wherein the modified endogenous acyl-coA oxidase polypeptide comprises an amino acid modification in the N-terminal loop.
- 15 A8. The genetically modified yeast of any one of embodiments A1 to A7, wherein the modified endogenous acyl-coA oxidase polypeptide comprises an amino acid modification in the D alpha helix.
- A9. The genetically modified yeast of any one of embodiments A1 to A8, wherein the modified endogenous acyl-coA oxidase polypeptide comprises an amino acid modification in the loop
20 between the D alpha helix and the E' alpha helix.
- A10. The genetically modified yeast of any one of embodiments A1 to A9, wherein the modified endogenous acyl-coA oxidase polypeptide comprises an amino acid modification to an amino acid in effective contact with carbons 6 to 9 in a feedstock component.
25
- A11. The genetically modified yeast of any one of embodiments A1 to A10, wherein the modified endogenous acyl-coA oxidase polypeptide comprises an amino acid modification to an amino acid in effective contact with carbons 10 to 12 in a feedstock component.
- 30 A12. The genetically modified yeast of any one of embodiments A1 to A11, wherein the modified endogenous acyl-coA oxidase polypeptide comprises an amino acid modification in the L alpha helix.

- A13. The genetically modified yeast of any one of embodiments A1 to A12, wherein the modified endogenous acyl-coA oxidase polypeptide comprises an amino acid modification in the loop C-terminal to the L alpha helix.
- 5 A14. The genetically modified yeast of any one of embodiments A1 to A13, wherein the modified endogenous acyl-coA oxidase polypeptide comprises an amino acid modification in the loop between the L alpha helix and the M alpha helix.
- A15. The genetically modified yeast of any one of embodiments A7 to A14, wherein the amino
10 acid modification comprises an amino acid substitution.
- A16. The genetically modified yeast of embodiment A15, wherein the amino acid substitution is conservative.
- 15 A17. The genetically modified yeast of embodiment A15, wherein the amino acid substitution is not conservative.
- A18. The genetically modified yeast of any one of embodiments A1 to A17, which comprises a genetic modification that reduces the activity of an enoyl coA isomerase polypeptide.
20
- A19. The genetically modified yeast of embodiment A18, wherein the genetic modification disrupts a polynucleotide that encodes the enoyl coA isomerase polypeptide.
- A20. The genetically modified yeast of embodiment A18 or A19, wherein the enoyl coA isomerase
25 polypeptide is a polypeptide native to the yeast.
- A21. The genetically modified yeast of embodiment A20, wherein the yeast is a Candida spp. yeast.
- 30 A22. The genetically modified yeast of embodiment A21, wherein the enoyl coA isomerase polypeptide comprises the amino acid sequence of SEQ ID NO: 3675 or 3677.

- A23. The genetically modified yeast of any one of embodiments A1 to A22, which comprises a genetic modification that reduces the cytoplasmic activity of an acyl-CoA synthetase (ACS) polypeptide.
- 5 A24. The genetically modified yeast of any one of embodiments A1 to A23, which comprises a genetic modification that reduces the peroxisomal activity of an acyl-CoA synthetase (ACS) polypeptide.
- A25. The genetically modified yeast of embodiment A23 or A24, wherein the genetic modification
10 disrupts a polynucleotide that encodes the acyl-CoA synthetase (ACS) polypeptide.
- A26. The genetically modified yeast of embodiment A25, wherein the genetic modification disrupts an ACS1 polypeptide or ACS2 polypeptide.
- 15 A27. The genetically modified yeast of any one of embodiments A23 to A26, wherein the genetic modification disrupts a polynucleotide that encodes a long-chain acyl-CoA synthetase polypeptide.
- A28. The genetically modified yeast of embodiment A27, wherein the genetic modification disrupts a FAT1 polypeptide.
20
- A29. The genetically modified yeast of any one of embodiments A23 to A28, wherein the acyl-CoA synthetase (ACS) polypeptide is a polypeptide native to the yeast.
- A30. The genetically modified yeast of embodiment A29, wherein the yeast is a *Candida* spp.
25 yeast.
- A31. The genetically modified yeast of embodiment A30, wherein the acyl-CoA synthetase (ACS) polypeptide comprises an amino acid sequence chosen from SEQ ID NOs: 80, 82, 84, 158 and 159.
30
- A32. The genetically modified yeast of embodiment A30, wherein the FAT1 polypeptide comprises the amino acid sequence of SEQ ID NO: 90.

- A33. The genetically modified yeast of any one of embodiments A1 to A32, which comprises a genetic modification that reduces the activity of a PXA polypeptide.
- A34. The genetically modified yeast of embodiment A33, wherein the genetic modification disrupts a polynucleotide that encodes the PXA polypeptide.
- A35. The genetically modified yeast of embodiment A33 or A34, wherein the PXA polypeptide is a PXA1 polypeptide or a PXA2 polypeptide, or a PXA1 polypeptide and a PXA2 polypeptide.
- A36. The genetically modified yeast of any one of embodiments A33 to A35, wherein the PXA polypeptide is native to the yeast.
- A37. The genetically modified yeast of embodiment A36, wherein the yeast is a *Candida* spp. yeast.
- A38. The genetically modified yeast of embodiment A37, wherein the PXA1 polypeptide comprises the amino acid sequence of SEQ ID NO: 92.
- A39. The genetically modified yeast of embodiment A37, wherein the PXA2 polypeptide comprises the amino acid sequence of SEQ ID NO: 94.
- A40. The genetically modified yeast of any one of embodiments A1 to A39, comprising an active, modified endogenous acyl-coA oxidase polypeptide and no active, modified endogenous acyl-coA dehydrogenase polypeptide.
- A41. The genetically modified yeast of any one of embodiments A1 to A39, comprising no active, modified endogenous acyl-coA oxidase polypeptide and an active, modified endogenous acyl-coA dehydrogenase polypeptide.
- B1. A genetically modified yeast, comprising a heterologous acyl-coA oxidase polypeptide or a heterologous acyl-coA dehydrogenase polypeptide, which yeast is capable of producing a diacid from a feedstock comprising one or more components from a vegetable oil.
- B2. The genetically modified yeast of embodiment B1, wherein the heterologous acyl-coA oxidase polypeptide is a native polypeptide.

- B3. The genetically modified yeast of embodiment B1, wherein the heterologous acyl-coA oxidase polypeptide is an active, modified polypeptide.
- B4. The genetically modified yeast of embodiment B1, wherein the heterologous acyl-coA dehydrogenase polypeptide is a native polypeptide.
- B5. The genetically modified yeast of embodiment B1, wherein the heterologous acyl-coA dehydrogenase polypeptide is an active, modified polypeptide.
- B6. The genetically modified yeast of embodiment B1 or B2, wherein the heterologous acyl-coA oxidase polypeptide is chosen from a polypeptide having an amino acid sequence set forth in SEQ ID NO: 51 to SEQ ID NO: 3673.
- B7. The genetically modified yeast of embodiment B1 or B4, wherein the heterologous acyl-coA dehydrogenase polypeptide is chosen from SEQ ID NOs: 3679 to 3683, 3686, 3689, 3691, 3693, 3695, 3697, 3699, 3701 and 3703.
- B8. The genetically modified yeast of any one of embodiments B1 to B7, which is chosen from a *Candida* spp. yeast, *Yarrowia* spp. yeast, *Pichia* spp. yeast, *Saccharomyces* spp. yeast and *Kluyveromyces* spp. yeast.
- B9. The genetically modified yeast of embodiment B8, which is chosen from *C. tropicalis*, *C. viswanathii*, *Y. lipolytica*, *P. pastoris*, *P. membranifaciens*, *P. kluyveri*, *P. guilliermondii*, *P. heedii*, *P. subpelliculosa*, *S. cerevisiae*, *S. bayanus*, *S. pastorianus*, *S. carlsbergensis*, *K. lactis* and *K. marxianus*.
- B10. The genetically modified yeast of any one of embodiments B1 to B9, which comprises a genetic modification that reduces the activity of an enoyl coA isomerase polypeptide.
- B11. The genetically modified yeast of embodiment B10, wherein the genetic modification disrupts a polynucleotide that encodes the enoyl coA isomerase polypeptide.
- B12. The genetically modified yeast of embodiment B10 or B11, wherein the enoyl coA isomerase polypeptide is a polypeptide native to the yeast.

- B13. The genetically modified yeast of embodiment B12, wherein the yeast is a *Candida* spp. yeast.
- 5 B14. The genetically modified yeast of embodiment B13, wherein the enoyl coA isomerase polypeptide comprises the amino acid sequence of SEQ ID NO: 3675 or 3677.
- B15. The genetically modified yeast of any one of embodiments B1 to B14, which comprises a genetic modification that reduces the cytoplasmic activity of an acyl-CoA synthetase (ACS) polypeptide.
- 10 B16. The genetically modified yeast of any one of embodiments B1 to B15, which comprises a genetic modification that reduces the peroxisomal activity of an acyl-CoA synthetase (ACS) polypeptide.
- 15 B17. The genetically modified yeast of embodiment B15 or B16, wherein the genetic modification disrupts a polynucleotide that encodes the acyl-CoA synthetase (ACS) polypeptide.
- B18. The genetically modified yeast of embodiment B17, wherein the genetic modification disrupts an ACS1 polypeptide or ACS2 polypeptide.
- 20 B19. The genetically modified yeast of any one of embodiments B15 to B18, wherein the genetic modification disrupts a polynucleotide that encodes a long-chain acyl-CoA synthetase polypeptide.
- B20. The genetically modified yeast of embodiment B19, wherein the genetic modification disrupts a FAT1 polypeptide.
- 25 B21. The genetically modified yeast of any one of embodiments B15 to B20, wherein the acyl-CoA synthetase (ACS) polypeptide is a polypeptide native to the yeast.
- 30 B22. The genetically modified yeast of embodiment B21, wherein the yeast is a *Candida* spp. yeast.

B23. The genetically modified yeast of embodiment B22, wherein the acyl-CoA synthetase (ACS) polypeptide comprises an amino acid sequence chosen from SEQ ID NOs: 80, 82, 84, 158 and 159.

- 5 B24. The genetically modified yeast of embodiment B21, wherein the FAT1 polypeptide comprises the amino acid sequence of SEQ ID NO: 90.

B25. The genetically modified yeast of any one of embodiments B1 to B24, which comprises a genetic modification that reduces the activity of a PXA polypeptide.

10

B26. The genetically modified yeast of embodiment B25, wherein the genetic modification disrupts a polynucleotide that encodes the PXA polypeptide.

- 15 B27. The genetically modified yeast of embodiment B25 or B26, wherein the PXA polypeptide is a PXA1 polypeptide or a PXA2 polypeptide, or a PXA1 polypeptide and a PXA2 polypeptide.

B28. The genetically modified yeast of any one of embodiments B25 to B27, wherein the PXA polypeptide is native to the yeast.

- 20 B29. The genetically modified yeast of embodiment B28, wherein the yeast is a *Candida* spp. yeast.

B30. The genetically modified yeast of embodiment B29, wherein the PXA1 polypeptide comprises the amino acid sequence of SEQ ID NO: 92.

25

B31. The genetically modified yeast of embodiment B29, wherein the PXA2 polypeptide comprises the amino acid sequence of SEQ ID NO: 94.

- 30 B32. The genetically modified yeast of any one of embodiments B1 to B31, comprising an active, modified endogenous acyl-coA oxidase polypeptide and no active, modified endogenous acyl-coA dehydrogenase polypeptide.

B33. The genetically modified yeast of any one of embodiments B1 to B31, comprising no active, modified endogenous acyl-coA oxidase polypeptide and an active, modified endogenous acyl-coA dehydrogenase polypeptide.

- 5 C1. The genetically modified yeast of any one of embodiments A1 to A41 and B1 to B33, comprising one or more genetic modifications that reduce the activity of one or more native endogenous acyl-coA oxidase polypeptides.

- 10 C2. The genetically modified yeast of embodiment C1, comprising genetic modifications that reduce the activity of all native endogenous acyl-coA oxidase polypeptides.

C3. The genetically modified yeast of embodiment C1 or C2, wherein the genetic modifications partially block beta oxidation activity.

- 15 C4. The genetically modified yeast of any one of embodiments A1 to A41, B1 to B33, and C1 to C3, wherein the diacid is a C4 to C24 diacid.

C5. The genetically modified yeast of embodiment C4, wherein the diacid is a C10, C12, C14, C16, C18 or C20 diacid.

20

C6. The genetically modified yeast of embodiment C5, wherein the diacid is a C10 diacid.

C7. The genetically modified yeast of embodiment C5, wherein the diacid is a C12 diacid.

- 25 C8. The genetically modified yeast of embodiment C5, wherein the diacid is a C18 diacid.

C9. The genetically modified yeast of any one of embodiments C4 to C8, wherein the diacid contains no unsaturation.

- 30 C10. The genetically modified yeast of any one of embodiments C4 to C8, wherein the diacid contains one or more unsaturations.

C10.1. The genetically modified yeast of any one of embodiments C4 to C10, wherein the diacid is the predominant diacid in a mixture of diacids.

- C11. The genetically modified yeast of any one of embodiments A1 to A41, B1 to B33, and C1 to C10.1, wherein the feedstock comprises a substantially pure oil.
- 5 C12. The genetically modified yeast of any one of embodiments A1 to A41, B1 to B33, and C1 to C10, wherein the feedstock comprises a plurality of fatty acids.
- C13. The genetically modified yeast of embodiment C12, wherein the feedstock comprises a soapstock.
- 10 C14. The genetically modified yeast of embodiment C12, wherein the feedstock comprises a fatty acid distillate.
- C15. The genetically modified yeast of any one of embodiments A1 to A41, B1 to B33, and C1 to C14, wherein the vegetable oil is from a plant chosen from palm, palm kernel, coconut, soy,
15 safflower, canola, palm, palm kernel or combination thereof.
- D1. A method for producing a diacid, comprising:
 contacting a genetically modified yeast of any one of embodiments 1 to A41, B1 to B33,
 and C1 to C15 with a feedstock comprising one or more components from a vegetable oil capable
20 of being converted by the yeast to a diacid; and
 culturing the yeast under conditions in which the diacid is produced from the feedstock.
- D2. The method of embodiment D1, wherein the diacid is a C4 to C24 diacid.
- 25 D3. The method of embodiment D2, wherein the diacid is a C10, C12, C14, C16, C18 or C20 diacid.
- D4. The method of embodiment D3, wherein the diacid is a C10 diacid.
- 30 D5. The method of embodiment D3, wherein the diacid is a C12 diacid.
- D6. The method of embodiment D3, wherein the diacid is a C18 diacid.

- D7. The method of any one of embodiments D1 to D6, wherein the diacid contains no unsaturation.
- D8. The method of any one of embodiments D1 to D6, wherein the diacid contains one or more unsaturations.
- D8.1. The method of any one of embodiments D2 to D8, wherein the diacid is the predominant diacid in a mixture of diacids.
- D9. The method of any one of embodiments D1 to D8.1, wherein the feedstock comprises a substantially pure oil.
- D10. The method of any one of embodiments D1 to D8, wherein the feedstock comprises a plurality of fatty acids.
- D11. The method of embodiment D10, wherein the feedstock comprises a soapstock.
- D12. The method of embodiment D10, wherein the feedstock comprises a fatty acid distillate.
- D13. The method of any one of embodiments D1 to D12, wherein the vegetable oil is from a plant chosen from palm, palm kernel, coconut, soy, safflower, canola, palm, palm kernel or combination thereof.
- E1. An isolated nucleic acid, comprising a polynucleotide that encodes a modified acyl-coA oxidase polypeptide from a yeast.
- E2. The isolated nucleic acid of embodiment E1, wherein the modified acyl-coA oxidase polypeptide comprises an amino acid modification in the N-terminal loop.
- E3. The isolated nucleic acid of embodiment E1, wherein the modified endogenous acyl-coA oxidase polypeptide comprises an amino acid modification in the D alpha helix.

- E4. The isolated nucleic acid of embodiment E1, wherein the modified endogenous acyl-coA oxidase polypeptide comprises an amino acid modification in the loop between the D alpha helix and the E' alpha helix.
- 5 E5. The isolated nucleic acid of embodiment E1, wherein the modified endogenous acyl-coA oxidase polypeptide comprises an amino acid modification to an amino acid in effective contact with carbons 6 to 9 in a feedstock component.
- E6. The isolated nucleic acid of embodiment E1, wherein the modified endogenous acyl-coA
10 oxidase polypeptide comprises an amino acid modification to an amino acid in effective contact with carbons 10 to 12 in a feedstock component.
- E7. The isolated nucleic acid of embodiment E1, wherein the modified endogenous acyl-coA
15 oxidase polypeptide comprises an amino acid modification in the L alpha helix.
- E8. The isolated nucleic acid of embodiment E1, wherein the modified endogenous acyl-coA oxidase polypeptide comprises an amino acid modification in the loop C-terminal to the L alpha helix.
- 20 E9. The isolated nucleic acid of embodiment E1, wherein the modified endogenous acyl-coA oxidase polypeptide comprises an amino acid modification in the loop between the L alpha helix and the M alpha helix.
- E10. The isolated nucleic acid of any one of embodiments E2 to E9, wherein the amino acid
25 modification comprises an amino acid substitution.
- E11. The isolated nucleic acid of embodiment E10, wherein the amino acid substitution is conservative.
- 30 E12. The isolated nucleic acid of embodiment E10, wherein the amino acid substitution is not conservative.

- E13. The isolated nucleic acid of any one of embodiments E1 to E12, wherein the yeast is chosen from a *Candida* spp. yeast, *Yarrowia* spp. yeast, *Pichia* spp. yeast, *Saccharomyces* spp. yeast and *Kluyveromyces* spp. yeast.
- 5 E14. The isolated nucleic acid of embodiment E13, wherein the yeast is chosen from *C. tropicalis*, *C. viswanathii*, *Y. lipolytica*, *P. pastoris*, *P. membranifaciens*, *P. kluyveri*, *P. guilliermondii*, *P. heedii*, *P. subpelliculosa*, *S. cerevisiae*, *S. bayanus*, *S. pastorianus*, *S. carlsbergensis*, *K. lactis* and *K. marxianus*.
- 10 E15. The isolated nucleic acid of embodiment E13, wherein the yeast is a *Candida* spp. yeast.
- E16. The isolated nucleic acid of embodiment E15, wherein the yeast is chosen from *C. tropicalis* and *C. viswanathii*.
- 15 E17. The isolated nucleic acid of embodiment E16, wherein the yeast is a genetically modified ATCC20336 yeast.
- E18. The isolated nucleic acid of any one of embodiments E15 to E17, wherein the endogenous acyl-coA oxidase polypeptide is a POX4 polypeptide.
- 20 E19. The isolated nucleic acid of embodiment E18, wherein the POX4 polypeptide comprises a modified amino acid sequence of SEQ ID NO: 30.
- E20. The isolated nucleic acid of embodiment E18 or E19, wherein the POX4 polypeptide
25 comprises amino acid modifications at one or more amino acid positions chosen from 88, 90, 96, 98, 99, 100, 102, 103, 302, 309, 310, 473, 474, 475, 476, 477, 478, 479, 480, 481, 482, 483, 484, 485, 486, 487, 488, 489, 490, 491, 492, 493, 494, 495, 496, 497, 498, 499, 500, 501, 502, 503, 504 and 505.
- 30 E21. The isolated nucleic acid of any one of embodiments E15 to E17, wherein the endogenous acyl-coA oxidase polypeptide is a POX5 polypeptide.
- E22. The isolated nucleic acid of embodiment E21, wherein the POX5 polypeptide comprises a modified amino acid sequence of SEQ ID NO: 32.

- E23. The isolated nucleic acid of embodiment E21 or E22, wherein the POX5 polypeptide comprises amino acid modifications at one or more amino acid positions chosen from 81, 82, 83, 84, 85, 86, 88, 93, 94, 95, 96, 98, 102, 284, 287, 290, 291, 292, 294, 295, 436, 453, 454, 455, 456, 457, 458, 459, 460, 461, 462 and 463.
- 5
- E24. The isolated nucleic acid of embodiment E20 or E23, wherein at least one of the amino acid modifications is an amino acid substitution.
- E25. The isolated nucleic acid of embodiment E24, wherein at least one of the amino acid
- 10 substitutions is conservative.
- E26. The isolated nucleic acid of embodiment E24, wherein at least one of the amino acid substitutions is non-conservative.
- 15 F1. The isolated nucleic acid of any one of embodiments E1 to E26, which is an expression vector.
- F2. A cell comprising a nucleic acid of any one of embodiments E1 to F1.
- 20 F3. The cell of embodiment F2, which is a bacterium.
- F4. The cell of embodiment F2, which is a yeast.
- F5. The cell of embodiment F4, which is a *Candida* spp. yeast.
- 25 F6. The cell of embodiment F5, wherein the *Candida* spp. yeast is chosen from *C. tropicalis* and *C. viswanathii*.
- F7. The cell of embodiment F6, wherein the *Candida* spp. yeast is a genetically modified
- 30 ATCC20336 yeast.
- F8. The cell of embodiment F4, which is chosen from a *Yarrowia* spp. yeast, *Pichia* spp. yeast, *Saccharomyces* spp. yeast and *Kluyveromyces* spp. yeast.

F9. The cell of embodiment F8, which is chosen from *Y. lipolytica*, *P. pastoris*, *P. membranifaciens*, *P. kluyveri*, *P. guilliermondii*, *P. heedii*, *P. subpelliculosa*, *S. cerevisiae*, *S. bayanus*, *S. pastorianus*, *S. carlsbergensis*, *K. lactis* and *K. marxianus*.

5

*

*

*

Citation of the above patents, patent applications, publications and documents is not an admission that any of the foregoing is pertinent prior art, nor does it constitute any admission as to the contents or date of these publications or documents.

10

Modifications may be made to the foregoing without departing from the basic aspects of the technology. Although the technology has been described in substantial detail with reference to one or more specific embodiments, those of ordinary skill in the art will recognize that changes may be made to the embodiments specifically disclosed in this application, yet these modifications and improvements are within the scope and spirit of the technology.

15

The technology illustratively described herein suitably may be practiced in the absence of any element(s) not specifically disclosed herein. Thus, for example, in each instance herein any of the terms "comprising," "consisting essentially of," and "consisting of" may be replaced with either of the other two terms. The terms and expressions which have been employed are used as terms of description and not of limitation, and use of such terms and expressions do not exclude any equivalents of the features shown and described or portions thereof, and various modifications are possible within the scope of the technology claimed. The term "a" or "an" can refer to one of or a plurality of the elements it modifies (e.g., "a reagent" can mean one or more reagents) unless it is contextually clear either one of the elements or more than one of the elements is described. The term "about" as used herein refers to a value within 10% of the underlying parameter (i.e., plus or minus 10%), and use of the term "about" at the beginning of a string of values modifies each of the values (i.e., "about 1, 2 and 3" refers to about 1, about 2 and about 3). For example, a weight of "about 100 grams" can include weights between 90 grams and 110 grams. Further, when a listing of values is described herein (e.g., about 50%, 60%, 70%, 80%, 85% or 86%) the listing includes all intermediate and fractional values thereof (e.g., 54%, 85.4%). Thus, it should be understood that although the present technology has been specifically disclosed by representative embodiments and optional features, modification and variation of the concepts herein disclosed

30

may be resorted to by those skilled in the art, and such modifications and variations are considered within the scope of this technology.

Certain embodiments of the technology are set forth in the claim(s) that follow(s).

What is claimed is:

1. A genetically modified yeast, comprising:
an active, modified endogenous POX5 polypeptide, wherein the POX5 polypeptide comprises an amino acid substitution as compared to the sequence shown in SEQ ID NO: 32 at one or more amino acid positions chosen from 81, 82, 83, 84, 85, 86, 88, 93, 94, 95, 96, 98, 102, 284, 287, 290, 291, 292, 294, 295, 436, 453, 454, 455, 456, 457, 458, 459, 460, 461, 462 and 463;
which yeast produces a diacid from a feedstock comprising one or more components from a vegetable oil capable of being converted by the yeast to a diacid.
2. The genetically modified yeast of claim 1, which comprises a genetic modification that reduces the activity of an enoyl coA isomerase polypeptide by disrupting a polynucleotide that encodes the enoyl coA isomerase polypeptide.
3. The genetically modified yeast of claim 2, wherein the enoyl coA isomerase polypeptide is a polypeptide native to the yeast.
4. The genetically modified yeast of claim 3, wherein the yeast is a *Candida* spp. yeast.
5. The genetically modified yeast of claim 2, wherein the enoyl coA isomerase polypeptide comprises the amino acid sequence of SEQ ID NO: 3675 or 3677.
6. The genetically modified yeast of any one of claims 1 to 5, wherein the diacid is a C4 to C24 diacid.
7. The genetically modified yeast of any one of claims 1 to 6, wherein the feedstock comprises a plurality of fatty acids.
8. The genetically modified yeast of claim 7, wherein the feedstock comprises a soapstock.
9. The genetically modified yeast of any one of claims 1 to 8, wherein the feedstock comprises substantially lauric acid and the diacid comprises substantially dodecanedioic acid.

10. The genetically modified yeast of claim 1, wherein the vegetable oil is from a plant chosen from palm, palm kernel, coconut, soy, safflower, canola, palm, palm kernel or combination thereof.

11. A method for producing a diacid, comprising:
contacting a genetically modified yeast of any one of claims 1 to 10 with a feedstock comprising one or more components from a vegetable oil capable of being converted by the yeast to a diacid; and
culturing the yeast under conditions in which the diacid is produced from the feedstock.

1 / 30

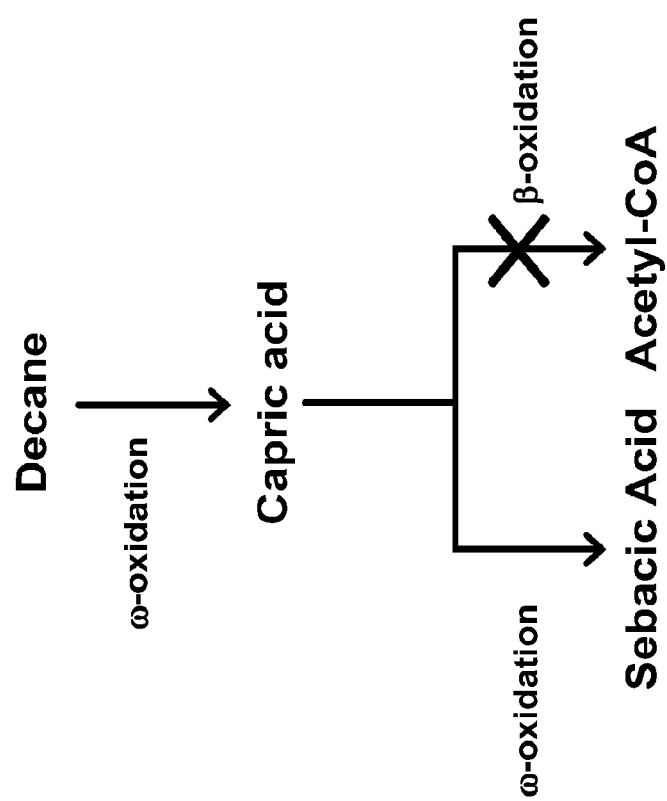


FIG. 1

2 / 30

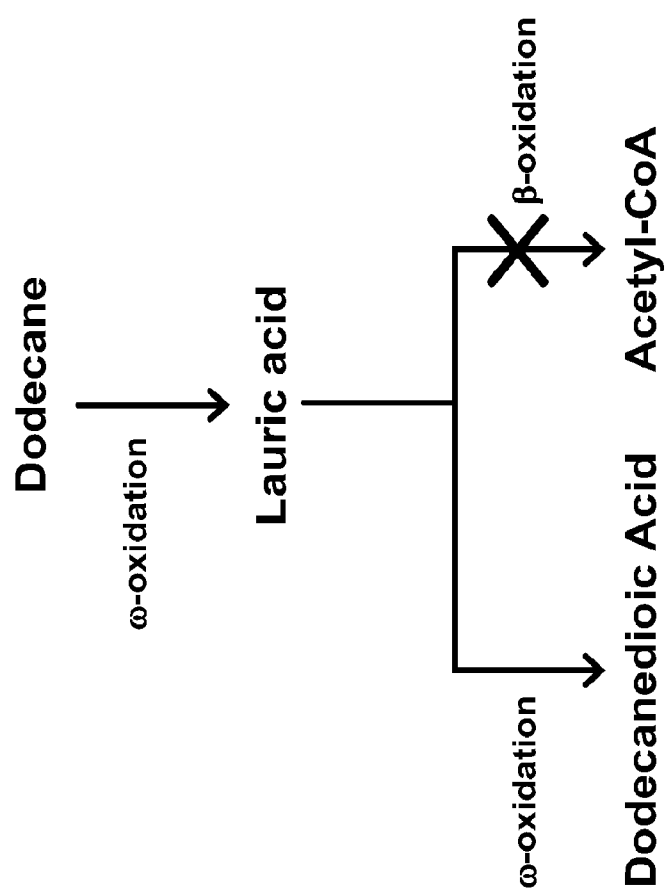


FIG. 2

3 / 30

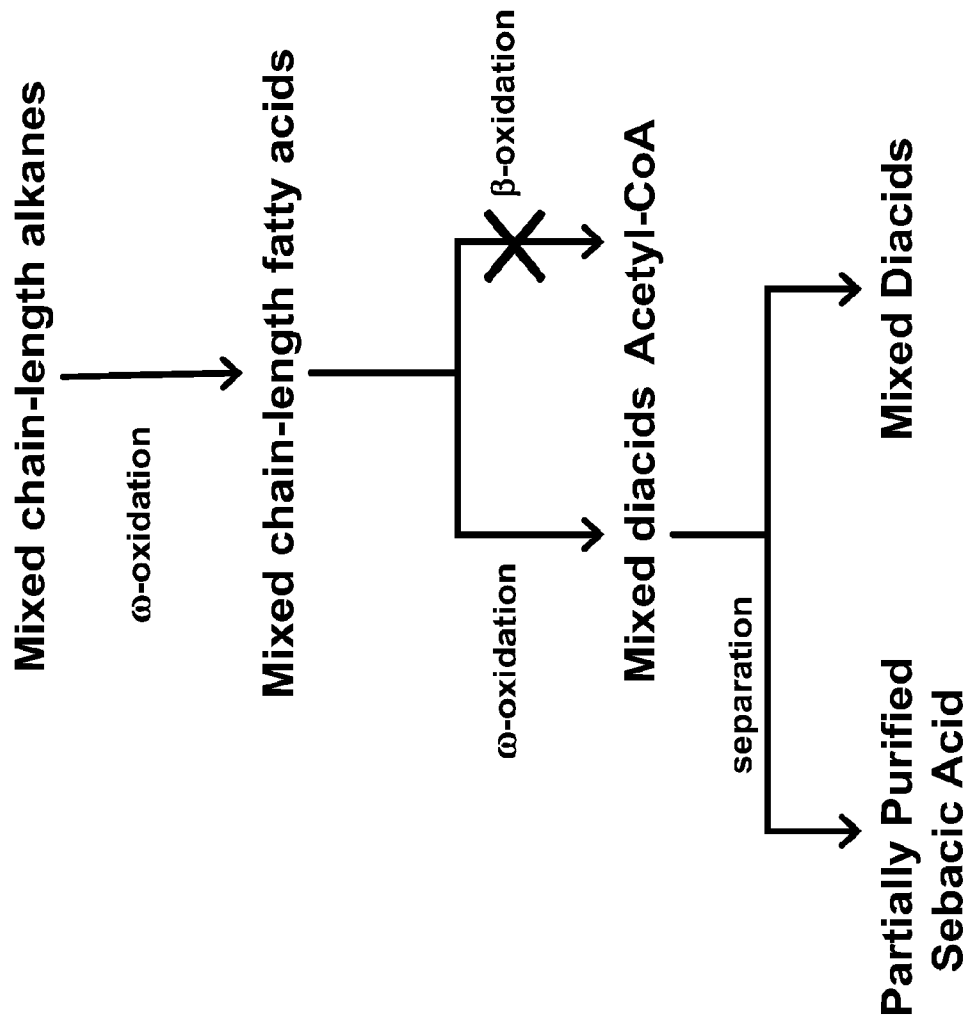


FIG. 3

4 / 30

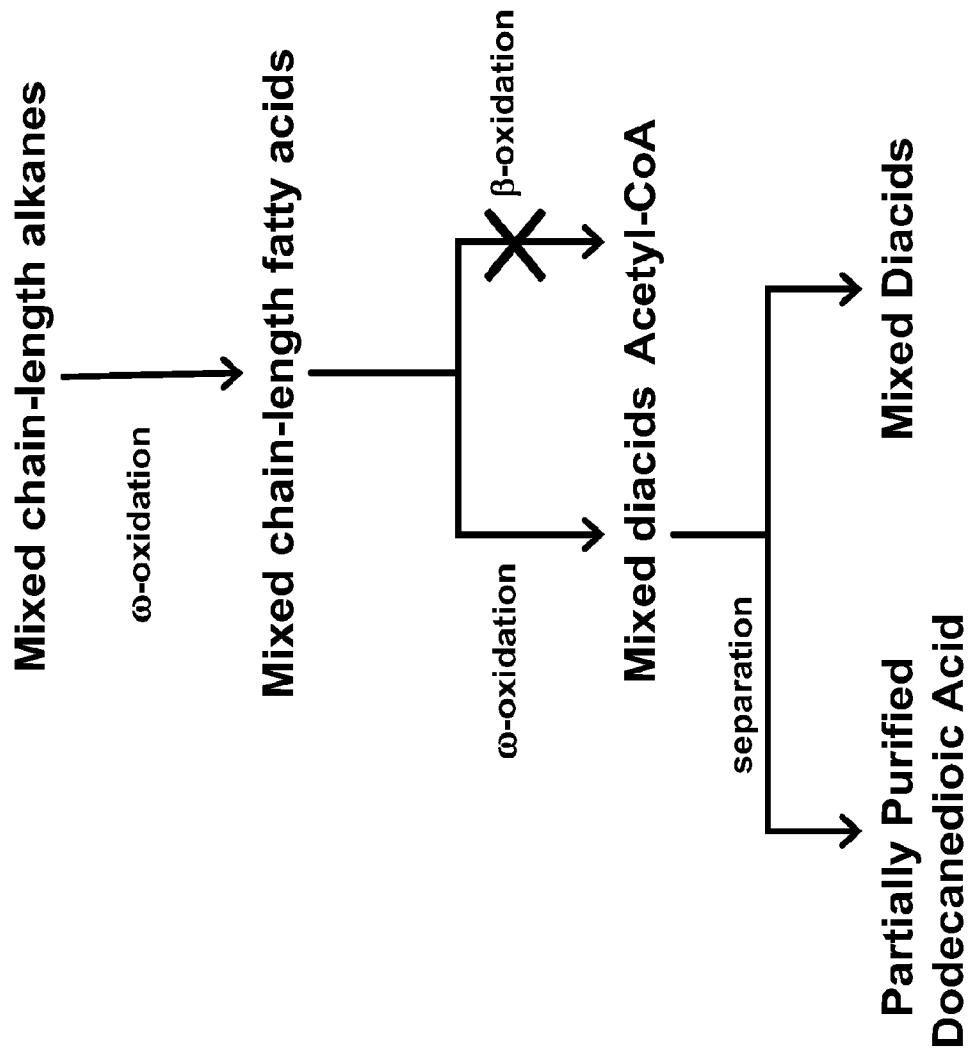


FIG. 4

5 / 30

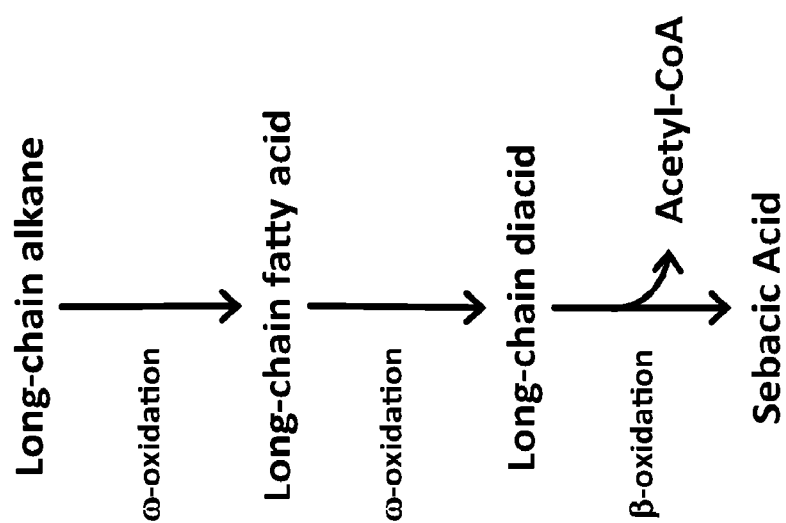


FIG. 5

6 / 30

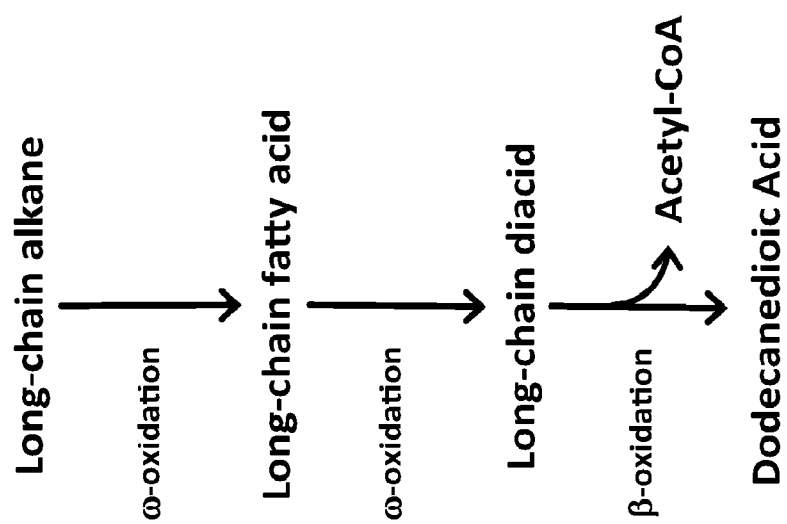


FIG. 6

7 / 30

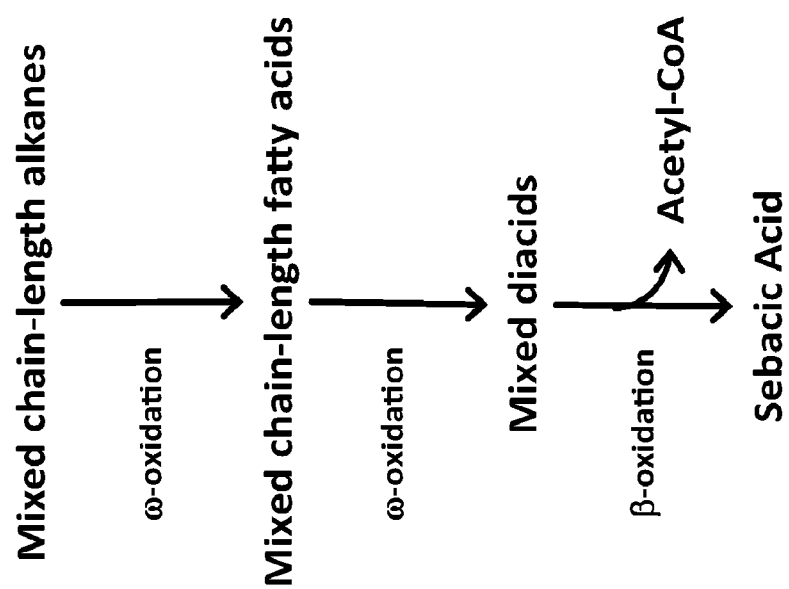


FIG. 7

8 / 30

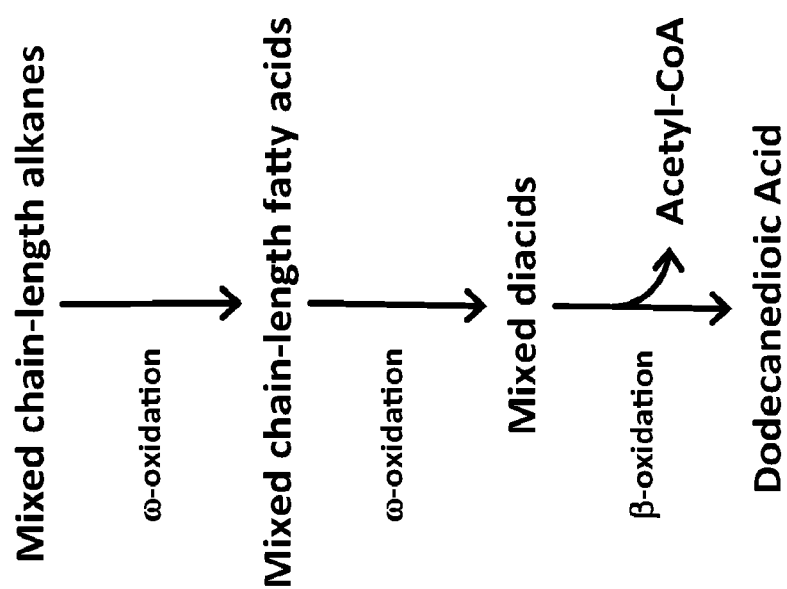


FIG. 8

9 / 30

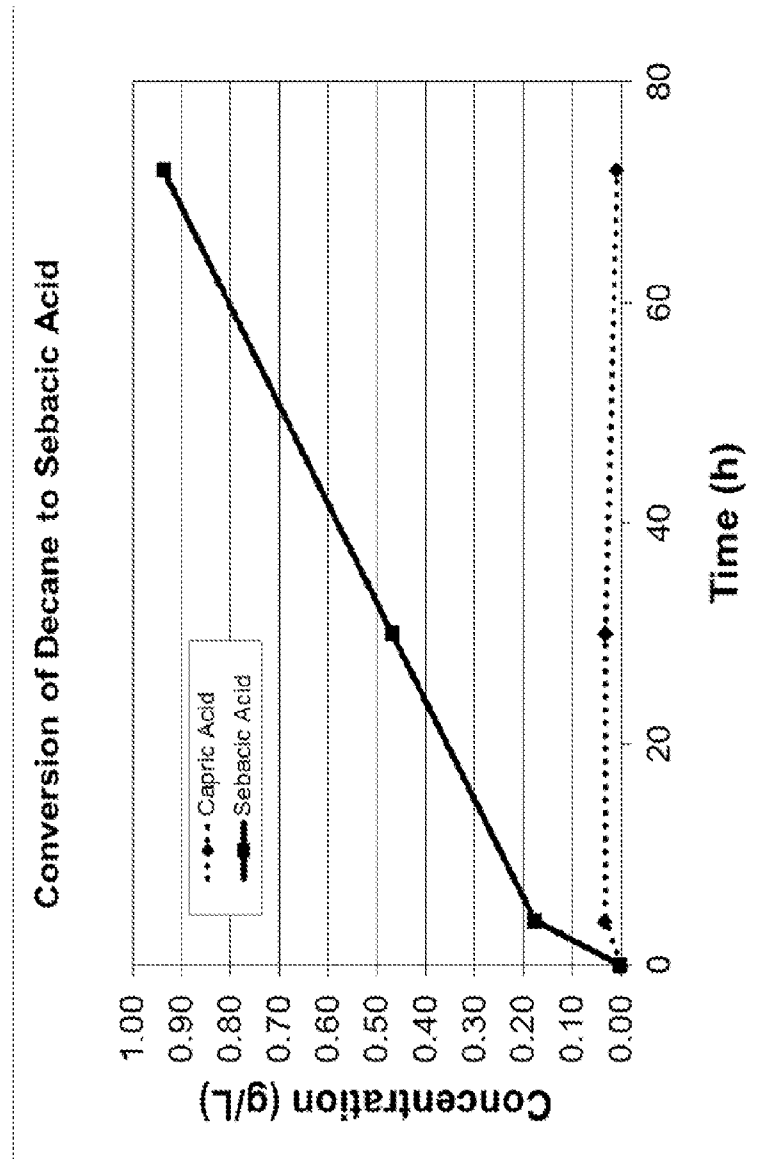


FIG. 9

10 / 30

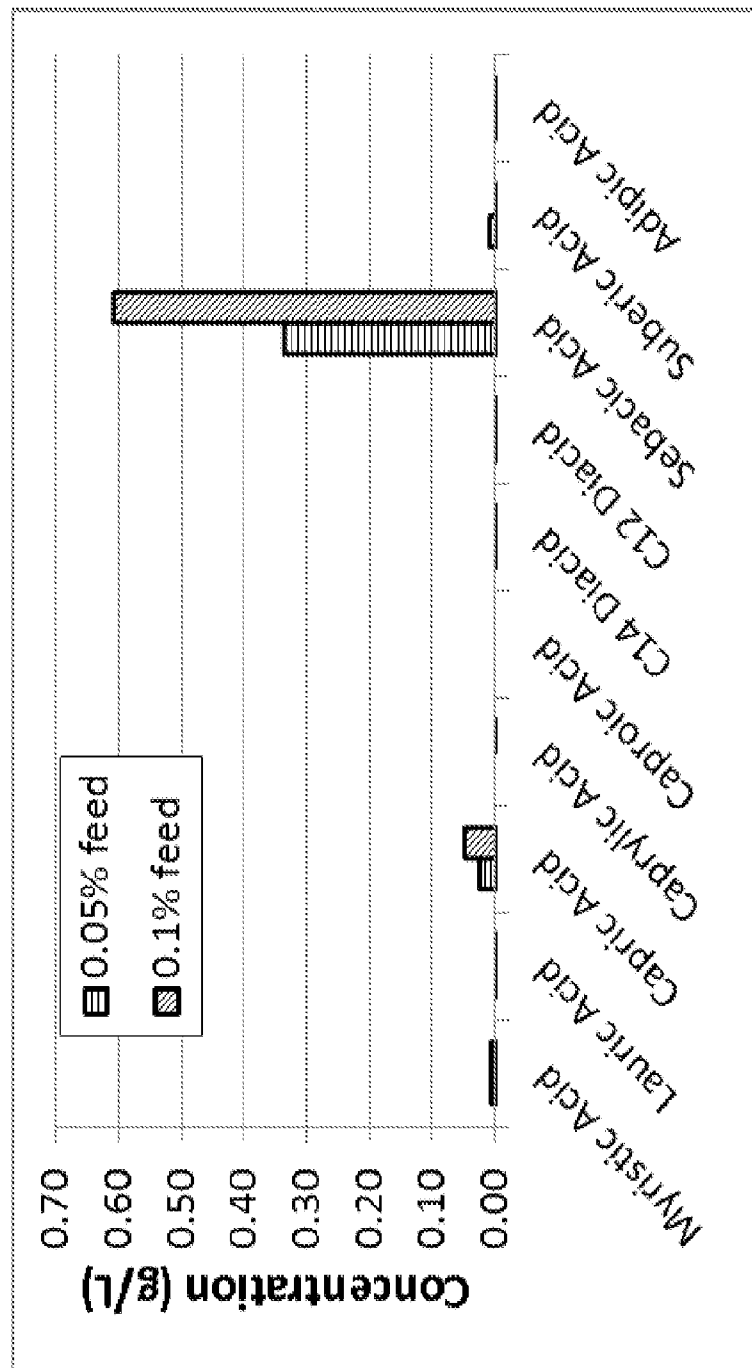


FIG. 10

11 / 30

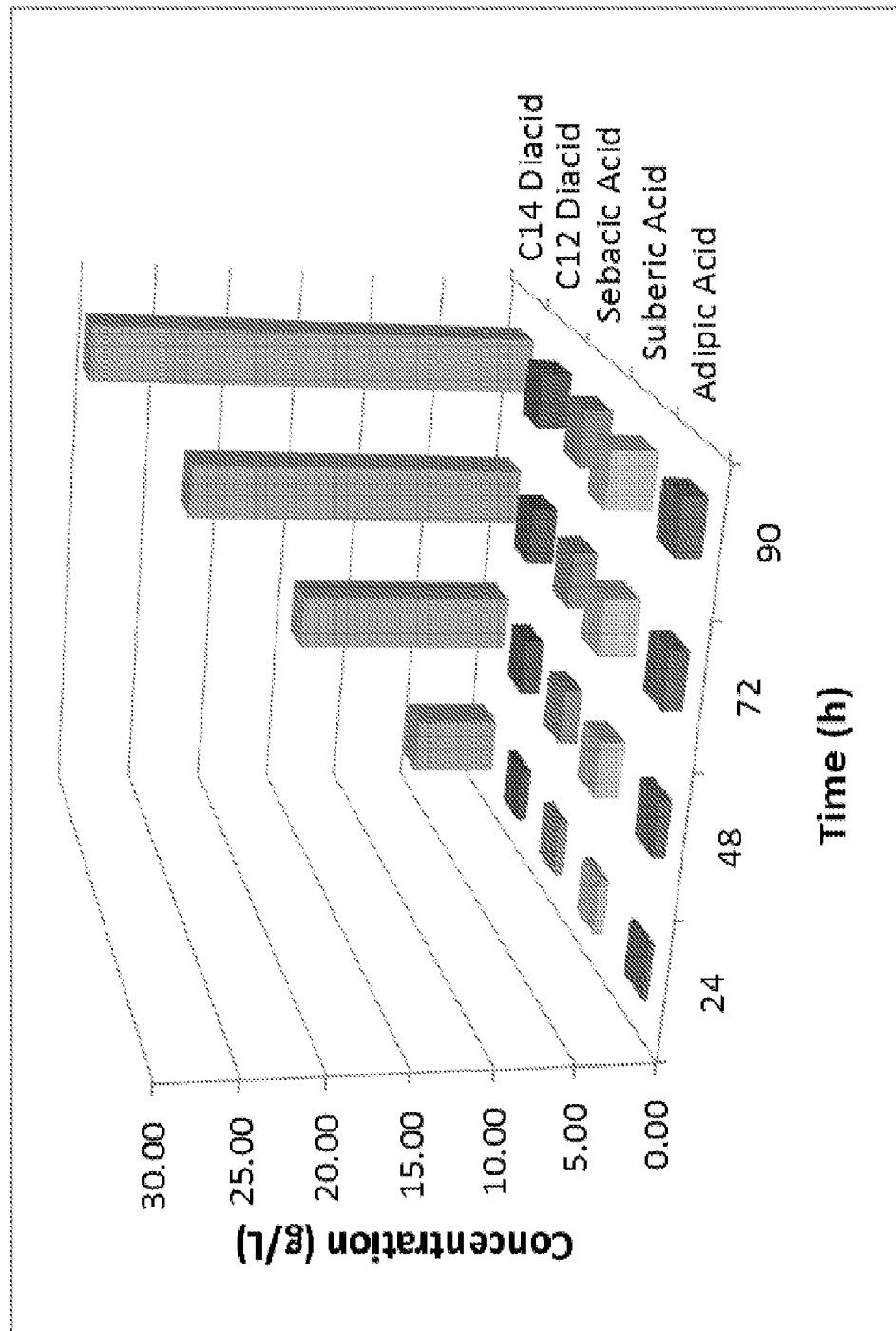


FIG. 11

12 / 30

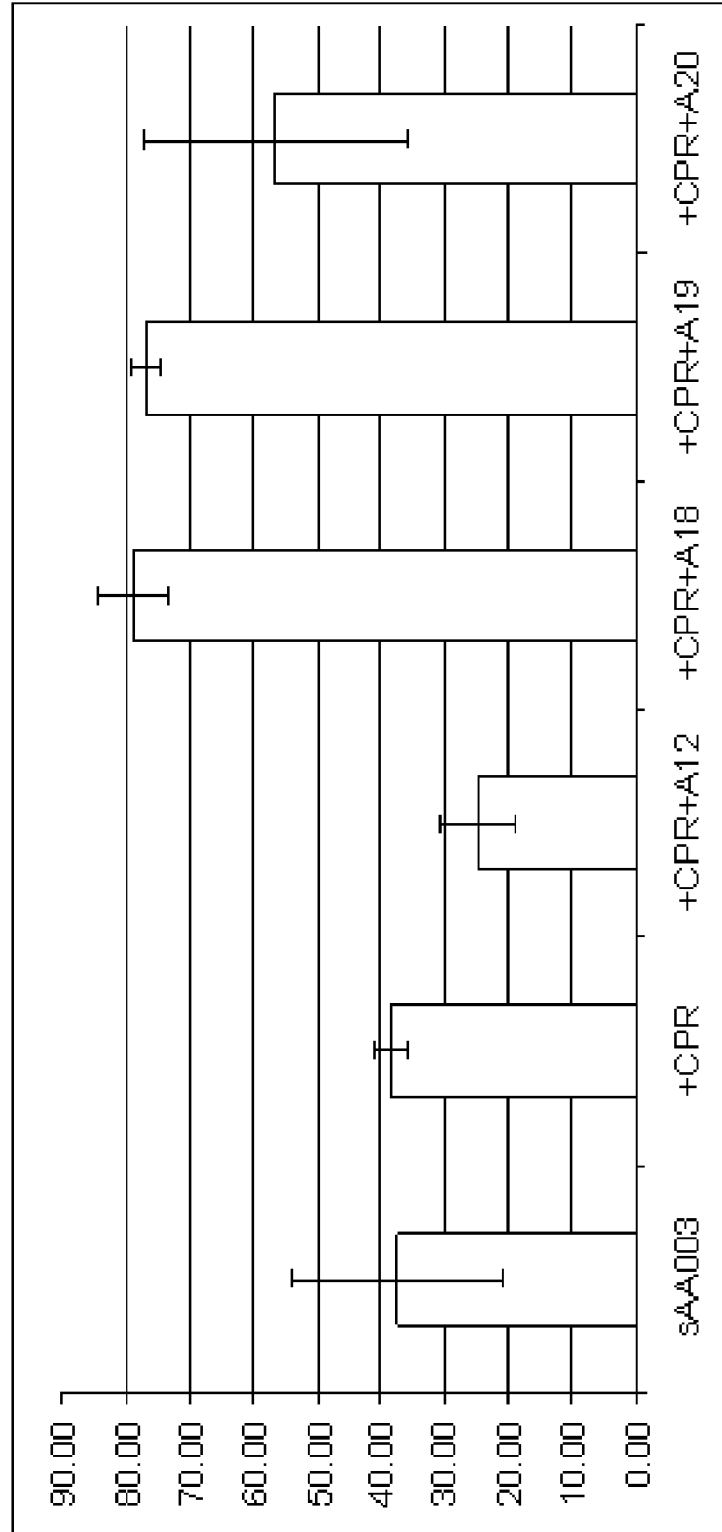


FIG. 12

13 / 30

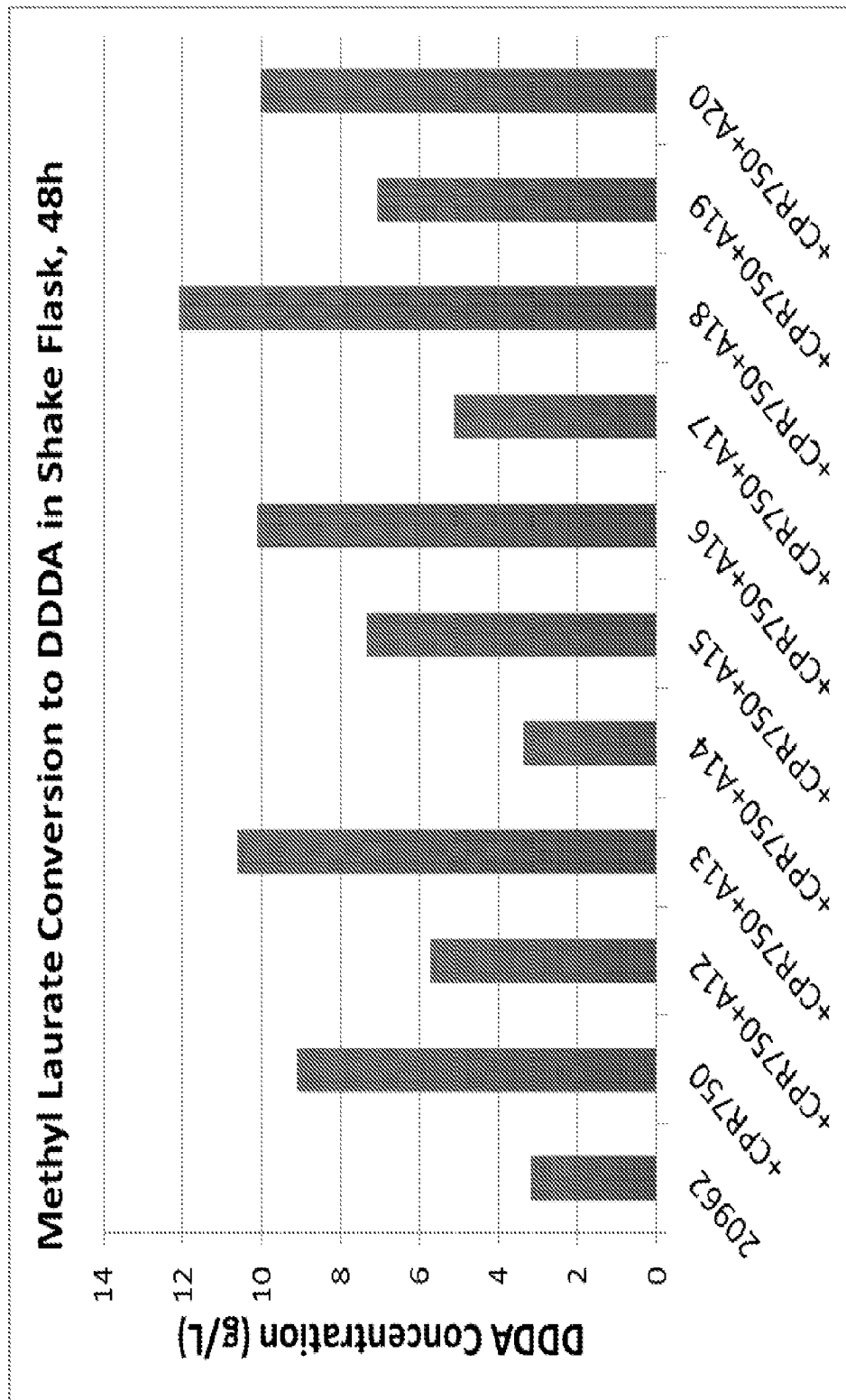


FIG. 13

14 / 30

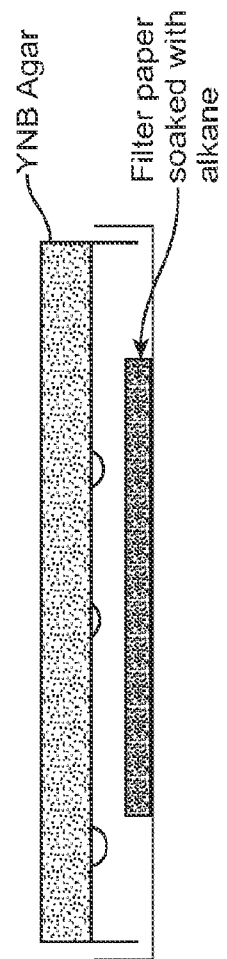


FIG. 14

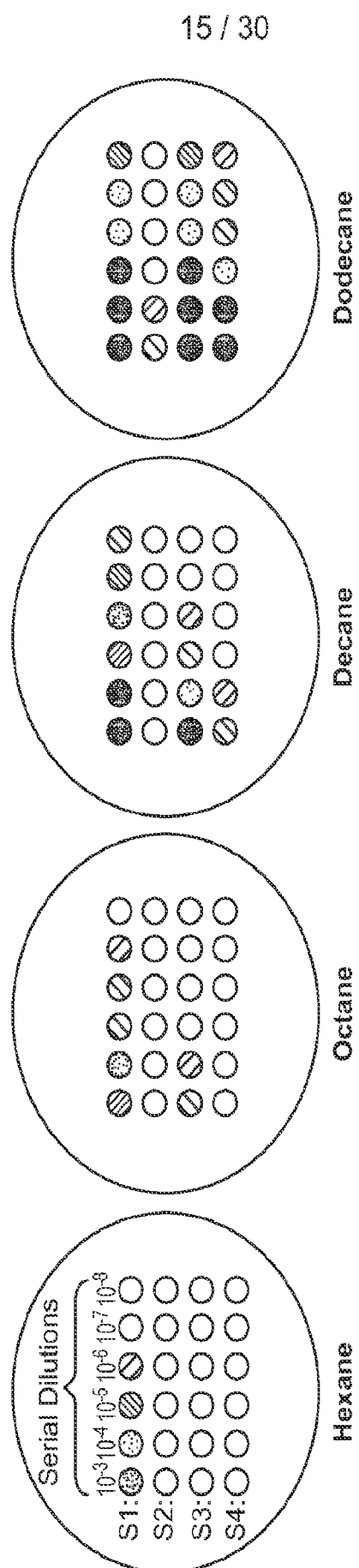
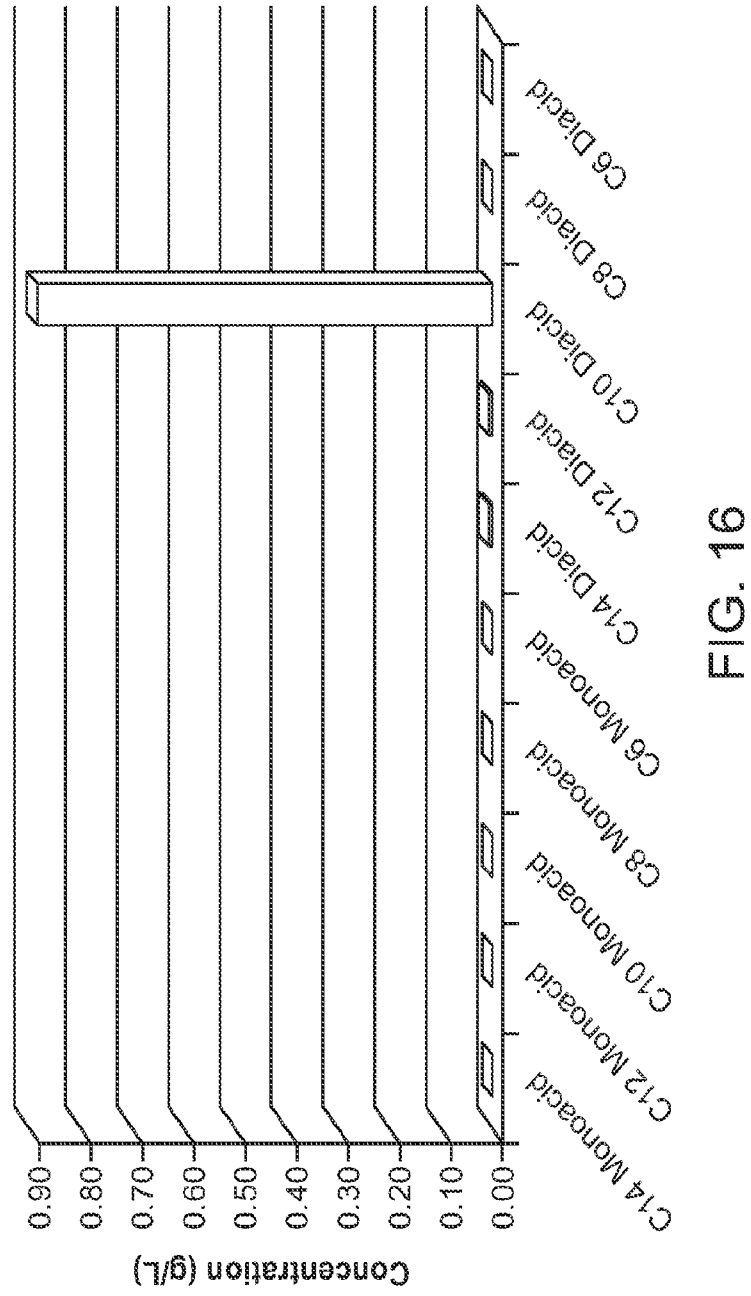


FIG. 15

16 / 30



17 / 30

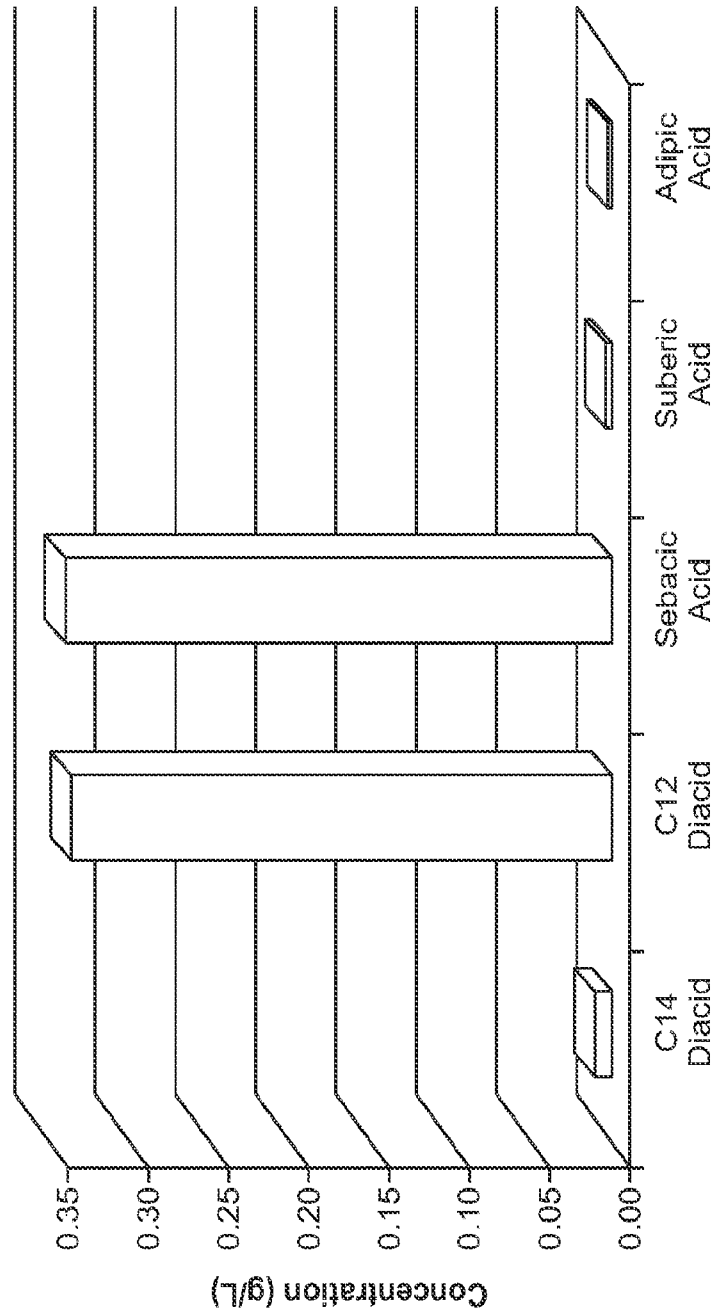


FIG. 17

18 / 30

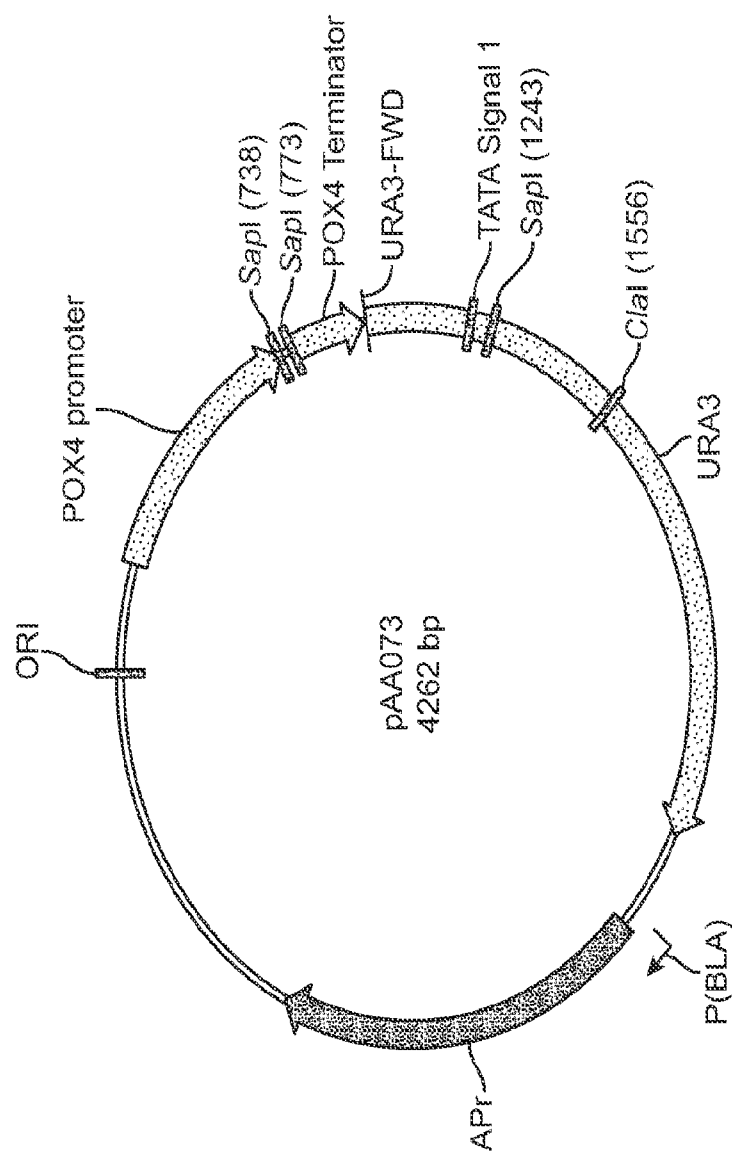


FIG. 18

19 / 30

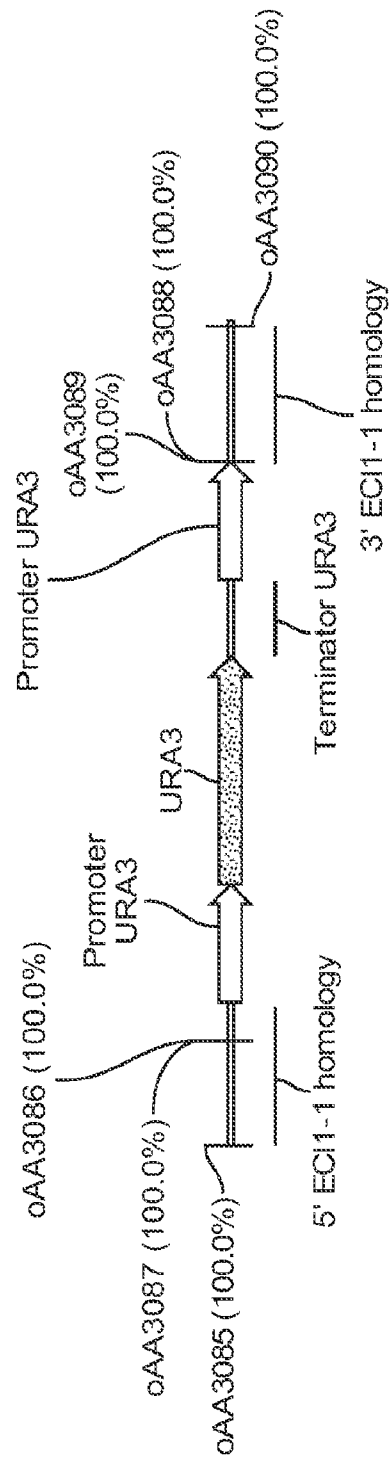


FIG. 19

20 / 30

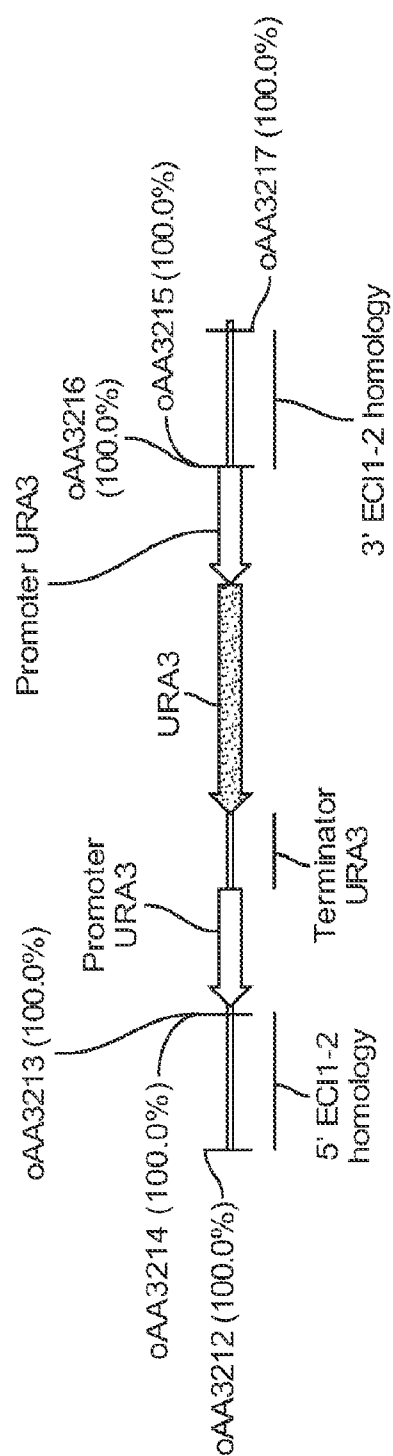


FIG. 20

21 / 30

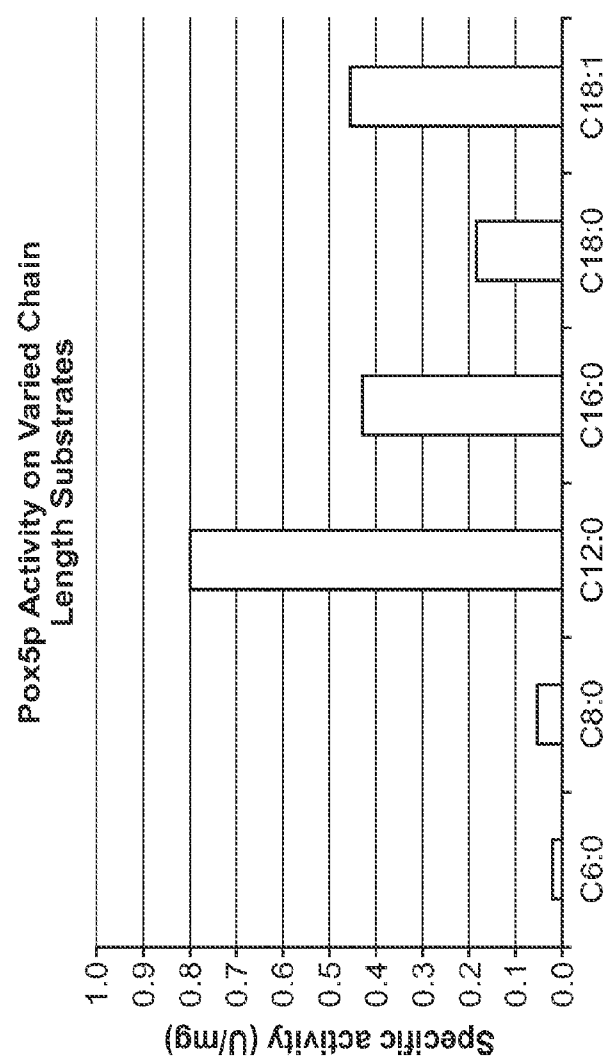


FIG. 21

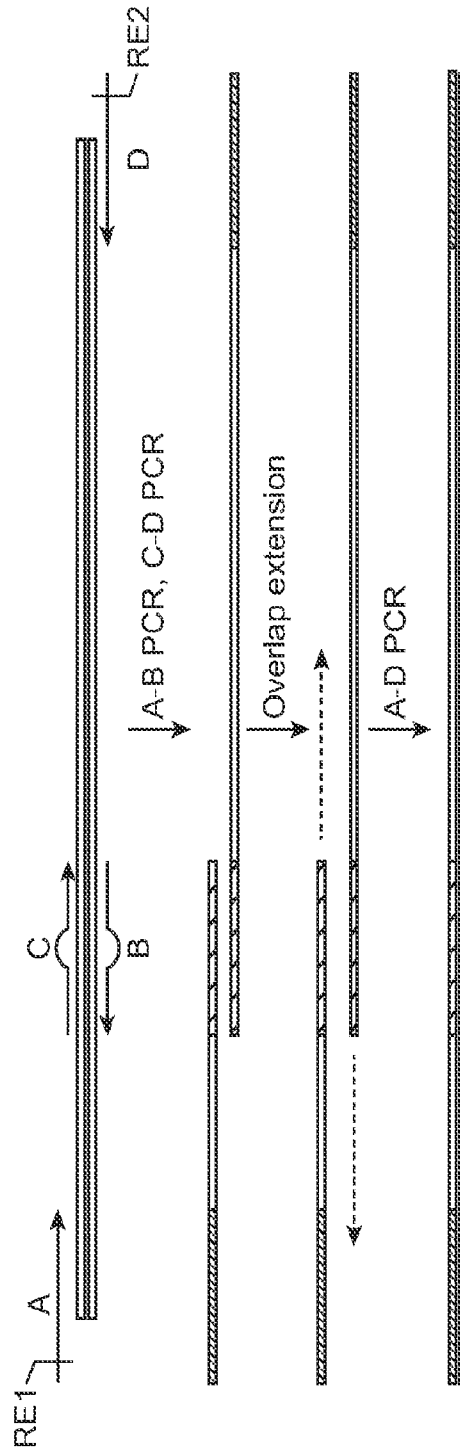


FIG. 22

23 / 30

AcII	1	MNPDLRKRASATFNPELITHLDGSPENTRRRRREIENLIINDPDFQHEDYNFELTRSQRY	60
AcI	1	MNPDLRKRASATFNPELITHLDGSPENTRRRRREIENLIINDPDFQHEDYNFELTRSQRY	60
AcII	61	EVAVKKSATMVKKKREYGISDPEEINMFKNSVHRGHPEPLDLHLGMFLPTLLHQATAEQQ	120
AcI	61	EVAVKKSATMVKKKREYGISDPEEINMFKNSVHRGHPEPLDLHLGMFLPTLLHQATAEQQ	120
AcII	121	ERREMPAMNLEITGYAQTENMGHGTHLRGLETATATYDPKTOEEFIINSPTVTSIKWMPGGL	180
AcI	121	EKMRFPSQELQIIGTYAQTENMGHGTHLRGLETATATYDPKTOEEFIINSPTVTSIKWMPGGL	180

FIG. 23

24A
G
L

MM T F T K K N V S V S Q G P D P R S S I Q K E R D S S K N N P Q Q M N Y L E G S V E R S S E L M K A L A Q Q M E R D P
II I L E D G S Y Y D L T K D Q Q R E L T A V K I N R I A D V K G I Y G C F G M T E L A H G S N V A G L E T T A T F D K
N L G L F S C I R G N G I C A T K W I G C A A H S A T H C S V Y A R L I V D G Q D Y C V K T F V V P L R D S N H D L M
E S D E F V I N T P H I G A K M G R D G I D N G W I Q F S N V R I P R F F M L Q K F C K V S A E G E V T L P P L E Q L S Y S A
P P G V T V G D I G A K M G R D G I D N G W I Q F S N V R I P R F F M L Q K F C K V S A E G E V T L P P L E Q L S Y S A
L L G G R V M V L D S Y R M L A R M S T I A L R Y A I G R R Q F F K G D N V D P K D P N A L E T Q L I D Y P L H Q K R
L L E P Y L A A A Y V I S G A L K V E D T I H N T L A E L D A A V E K N D T K A I F K S I D M K S L F V D S G S L K
S T A T W L G A E A I D Q C R Q A C G G H G Y S S Y N G F G K A Y N D W V V Q C T W E G D N N V L A M S V G K P I V K
Q Q V I S I E D A G K T V R G S T A F L N Q L K K D Y T G S N S S K V V L N T V A D L D D I K K T V I K A I E V A I I R L S
Q Q E A A S I V K K E S F D Y V G A E L V Q L S K K L K A H H Y L L T E Y I R R I D T D D Q K D L V P Y L I T L G K L Y A
A T I V L D R F A G V F L T F F N V A S T E A I T A L A S V Q I P K L C A E V R P N V V A Y T D S F Q Q S D D E T A A I
I G R Y D G D I Y E N Y F D V K L Q P S K T K A P Y S D A L E A M L N R P T L D E R E F E K S D E T A A I S

ELIGIBLE

52
G
F

26 / 30

RnAcoII	(1)	1	-----MM2DPRKESASATFNPELITTHLDICSPENTRRRRETEENLNDREDFQNSD--YMFETRSOR	75
CvPox5p	(1)		-----MPTELCQKREELTKFNPKEENVTTEIGSORESEITISNMVEQMCKDPLTKVDASTYNNENKDDQO	
CvPox4p	(1)		MTFTKXNVSVSQGPDPRSSHOKERDSKNNPQOMNYTTEIGSVSRSEIKRALAQOMERDRLFTDGSRYDNTKDDQO	
Consensus	(1)		MSDLQKERDSSKENEINVFLEGSERSEIIKIQNMKDPILDASYYLTKDQO	
RnAcoII	(60)	76	YIVAVKRSATMVVKKMRREYCHSDPREEIMWFKNSVVRCHPEPEDEHNGMPLPTILHQATARQQEPEFMP--ANNI-E	150
CvPox5p	(61)		REVTAKKIARLSRYTEHEYPDQQAQKLSLGLVDFDQVFTKIGVAGGLTSCVVRGNGRNSQTFXWTEKCIDKILRG	
CvPox4p	(76)		RELPVAKKINRTAKYREQESPDFTFNKNSLSEIGHDFDQVGTTRIGVYVGLTSCVVRGNGRTSGLNFWANEXETADVKG	
Consensus	(76)		REVTAKKIARIARYEEIDRLSIIGIFDPPQVTRIGVNLGLFLSCIRNGTSQYWIKLKG	
RnAcoII	(132)	151	NTGTYAQITRMGHGTHLRGLETNTAVDPKTOETELNSPIWTSTIKNWPPGGLGKTSNHAIVLAQDTTQGE CGLHAEV	225
CvPox5p	(136)		SYGCFGCTSLAHGGSNVQGLETTANFQEDDDEEVINTPEHIGATRWVHGGAAHSAHCSVYAAKVKYCKDQGVKTFV	
CvPox4p	(151)		SYGCFGCTSLAHGGSNVAGLETTANFQKESDENVINTPEHIGATRWVHGGAAHSAHCSVYAAKVKYCKDQGVKTFV	
Consensus	(151)		IYGCFCGMTSLAHGGSNVGLETTATFDTDEFVINTPHIGATKWNWIGGAAHSAHCSVYARLIVGDYGVKTFV	
RnAcoII	(207)	226	VRTRRIGTGKPLEGCTNGDNGPMTGYETEDNGELKMDNIRSEERENKDMNYAQNKPQCTYVKKQLSNKQTRGTMVYV	300
CvPox5p	(211)		VRDLRD-SI-NHDLIEPGVFNAGDNGAAMGRDGLDNGWLOFSGNVRSEERFTMDQRYCKVSRIGLEVIMPESS-EOGSSSAHICG	
CvPox4p	(226)		VRDLRD-SI-NHDLIMPGVFNAGDNGAAMGRDGLDNGWLOFSGNVRSEERFTMDQRYCKVSAEIGLEVTLPEPELOGSSSAHICG	
Consensus	(226)		VPLRDSNHDLIPGVTVGDIGAKMGRDGLDNGWIOFSGNVRIPREFMLQKYCKVS DGEVTLPPSEQLSYSALIGG	
RnAcoII	(282)	301	RISFEVGNRAQSLSRACTETRNISAVRRQSE-----IKQSEPRPQITLDEQTQQYKUTPSLNTTAAAFHFTVGRYMKR	375
CvPox5p	(285)		RVTMMMDSYFMTSPTFITNLRALTRRQEK-----KKQTDITETKLIDVPLHCKKATTFNNMLESQCALXYLEQ	
CvPox4p	(300)		RVTMMMDSYFMTLAEPMSTETLRALGRQEPKGDNDVDPKDPNARETQILIDVPLHCKKATTFNNMLESQCALXYLEQ	
Consensus	(301)		RV MVLDSYRMLSR TIALRYAI RRQEK XD IETQLIDVPLHQXRLFPFLAAAYLFS GALYLED	

FIG. 26

FIG. 26 (Cont.)

28 / 30

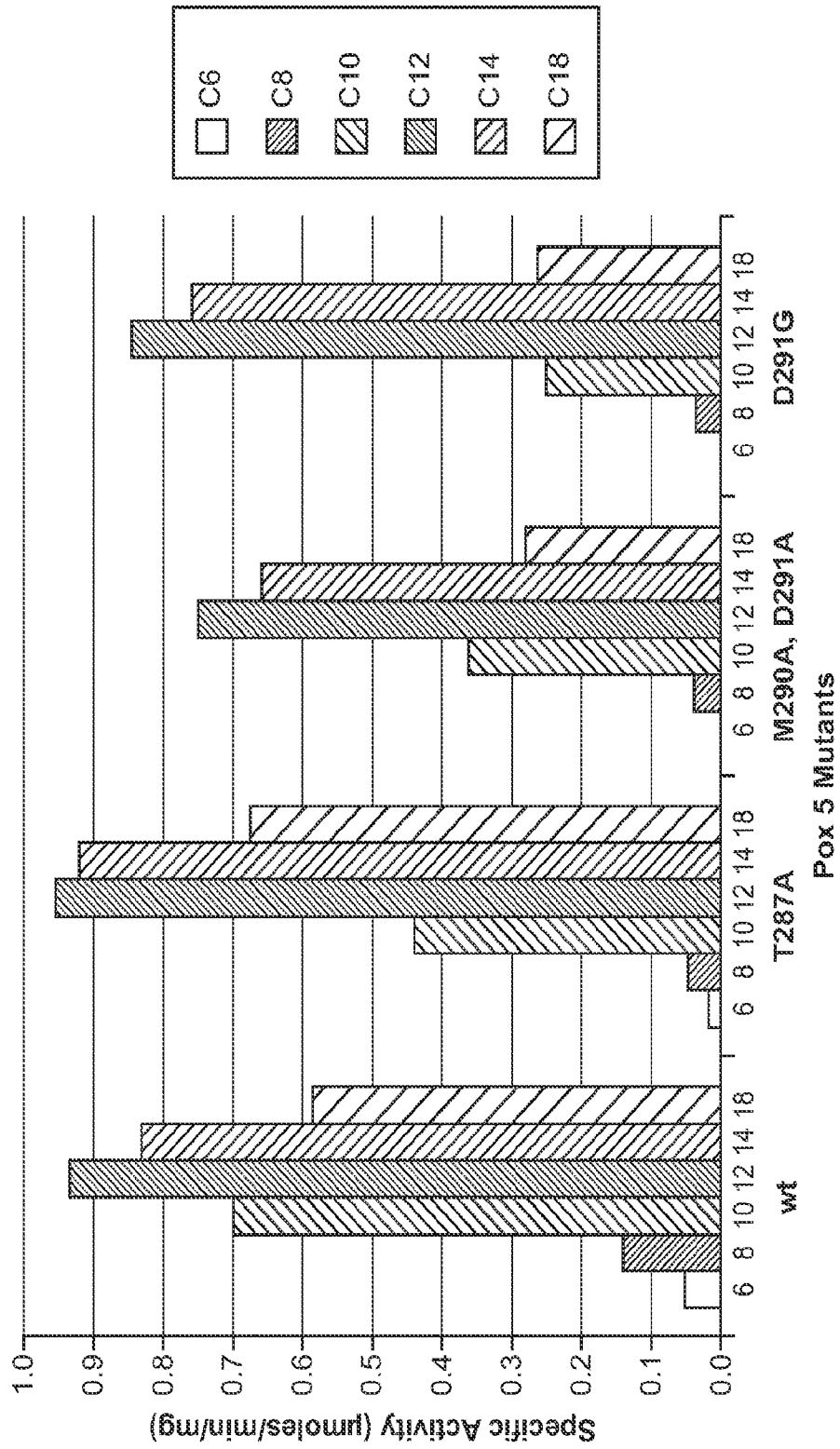


FIG. 27

29 / 30

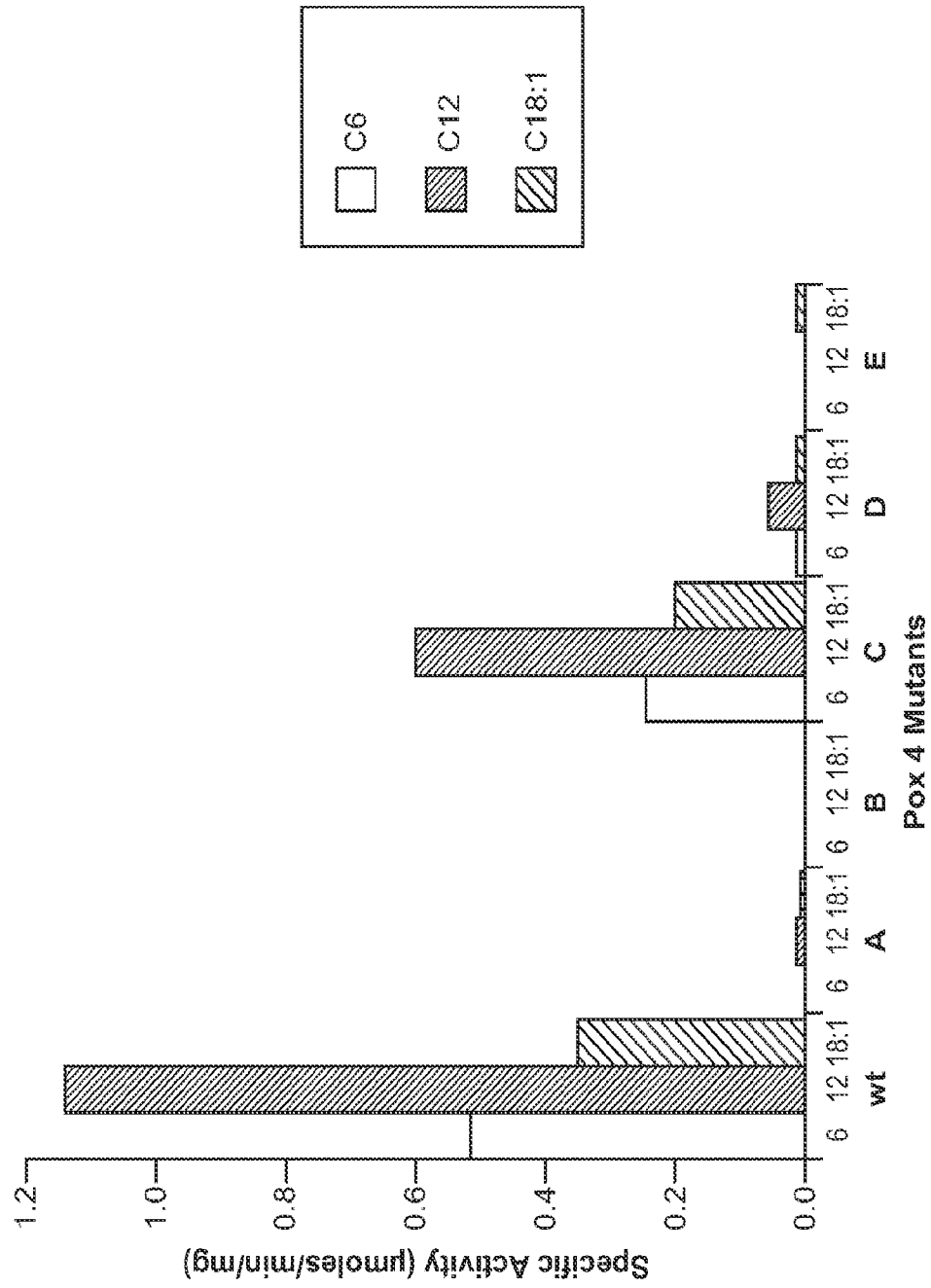


FIG. 28

30 / 30

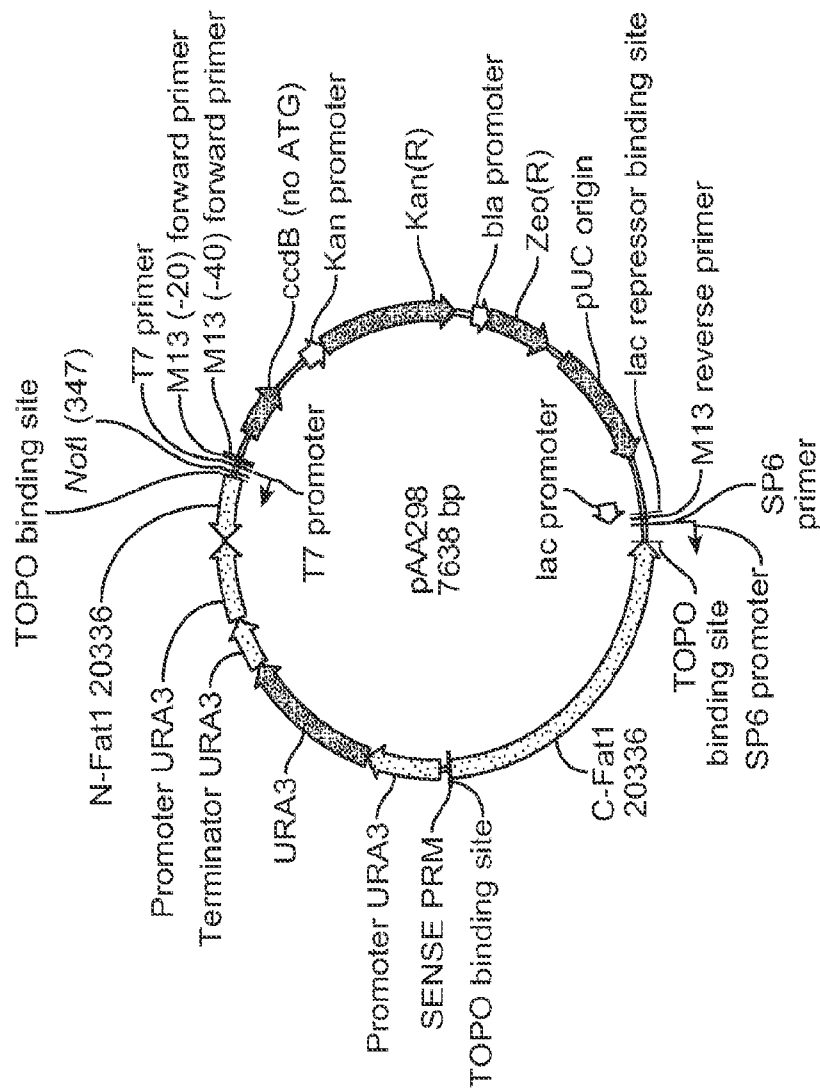


FIG. 29