(12) STANDARD PATENT (11) Application No. **AU 2009231744 B2**
(19) AUSTRALIAN PATENT OFFICE

(54) Title
Delta4 desaturase and its use in making polyunsaturated fatty acids

(51) International Patent Classification(s)
A01H 5/00 (2006.01) ***C12N 15/09*** (2006.01)

(21) Application No: **2009231744** (22) Date of Filing: **2009.04.01**

(87) WIPO No: **WO09/124101**

(30) Priority Data

(31) Number	(32) Date	(33) Country
61/041,716	**2008.04.02**	**US**

(43) Publication Date: **2009.10.08**
(44) Accepted Journal Date: **2014.05.08**

(71) Applicant(s)
E. I. du Pont de Nemours and Company

(72) Inventor(s)
Zhu, Quinn Qun;Xue, Zhixiong

(74) Agent / Attorney
Houlihan2, Level 1 70 Doncaster Road, BALWYN NORTH, VIC, 3104

(56) Related Art
WO 2007/096387
JP 2005 525780

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
8 October 2009 (08.10.2009)

PCT

(10) International Publication Number
WO 2009/124101 A1

(51) International Patent Classification:
A01H 5/00 (2006.01) *C12N 15/09* (2006.01)

(21) International Application Number:
PCT/US2009/039062

(22) International Filing Date:
1 April 2009 (01.04.2009)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/041,716 2 April 2008 (02.04.2008) US

(71) Applicant *(for all designated States except US)*: **E. I. DU PONT DE NEMOURS AND COMPANY** [US/US]; 1007 Market Street, Wilmington, DE 19898 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **XUE, Zhixiong** [CN/US]; 111 Harvey Lane, Chadds Ford, PA 19317-9728 (US). **ZHU, Quinn, Qun** [US/US]; 544 Revere Road, West Chester, PA 19382 (US).

(74) Agent: **SMITH, Loretta, F.**; E. I. du Pont de Nemours and Company, Legal Patent Records Center, 4417 Lancaster Pike, Wilmington, DE 19805 (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, 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, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, 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, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (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, SE, SI, SK, 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))*

— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*

— *with sequence listing part of description (Rule 5.2(a))*

WO 2009/124101 A1

(54) **Title**: Δ4 DESATURASE AND ITS USE IN MAKING POLYUNSATURATED FATTY ACIDS

(57) **Abstract**: Described here are Δ4 desaturases that convert all-*cis*-7,10,13,16,19-docosapentaenoic acid ["DPA"; 22:5 ω-3] to docosahexaenoic acid ["DHA"; 22:6 ω-3], with secondary activity in converting docosatetraenoic acid ["DTA"; 22:4 ω-6] to all-*cis*-4,7,10,13,16-docosapentaenoic acid ["DPAn-6"; 22:5 ω-6]. Also, described here are isolated nuclei acid fragments and recombinant constructs comprising such fragments encoding Δ4 desaturases as well as methods of making long chain polyunsaturated fatty acids ["PUFAs"] using this Δ4 desaturase in oleaginous yeast.

TITLE

Δ4 DESATURASE AND ITS USE IN MAKING POLYUNSATURATED FATTY ACIDS

[0001] This application claims the benefit of priority of U.S. Provisional Application No. 61/041716, filed April 2, 2008, currently pending, and hereby incorporated herein by reference in its entirety.

FIELD OF THE INVENTION

[0002] This invention is in the field of biotechnology. More specifically, this invention pertains to the identification of nucleic acid fragments encoding a Δ4 fatty acid desaturase enzyme and the use of this desaturase in making long chain polyunsaturated fatty acids ["PUFAs"].

BACKGROUND OF THE INVENTION

[0003] The health benefits associated with polyunsaturated fatty acids ["PUFAs"], especially ω-3 and ω-6 PUFAs, have been well documented. In order to find ways to produce large-scale quantities of ω-3 and ω-6 PUFAs, researchers have directed their work toward the discovery of genes and the understanding of the encoded biosynthetic pathways that result in lipids and fatty acids.

[0004] A variety of different hosts including plants, algae, fungi, stramenopiles and yeast are being investigated as means for commercial PUFA production. Genetic engineering has demonstrated that the natural abilities of some hosts, even those natively limited to linoleic acid ["LA"; 18:2 ω-6] or α-linolenic acid ["ALA"; 18:3 ω-3] fatty acid production, can be substantially altered to result in high-level production of various long-chain ω-3/ω-6 PUFAs. Whether this is the result of natural abilities or recombinant technology, production of docosahexaenoic acid ["DHA"; 22:6 ω-3] from docosapentaenoic acid ["DPA"; 22:5 ω-3] may require expression of a Δ4 desaturase. More specifically, most Δ4 desaturase enzymes identified so far have the primary ability to convert DPA to DHA,

with secondary activity in converting docosatetraenoic acid ["DTA"; 22:4 ω-6] to ω-6 docosapentaenoic acid ["DPAn-6"; 22:5 ω-6].

[0005] Based on the role Δ4 desaturase enzymes play in the synthesis of DHA, there has been considerable effort to identify and characterize these enzymes from various sources. Numerous Δ4 desaturases have been disclosed in both the open literature and the patent literature. Some examples include: *Euglena gracilis* (SEQ ID NO:13; GenBank Accession No. AY278558; Meyer et al., *Biochemistry,* 42(32):9779-9788 (2003)); *Thalassiosira pseudonana* (SEQ ID NO:37; GenBank Accession No. AAX14506; Tonon et al., *FEBS J.,* 272(13):3401-3412 (2005)); *Thraustochytrium aureum* (SEQ ID NO:14; GenBank Accession No. AAN75707); *Thraustochytrium* sp. (GenBank Accession No. CAD42496; U.S. Pat. No. 7,087,432); *Schizochytrium aggregatum* (SEQ ID NO:41; Int'l. App. Pub. No. WO 2002/090493); *Pavlova lutheri* (SEQ ID NO:42; GenBank Accession No. AAQ98793); and, *Isochrysis galbana* (SEQ ID NO:43; GenBank Accession No. AAV33631; Pereira et al., *Biochem. J.,* 384(2):357-366 (2004); Int'l. App. Pub. No. WO 2002/090493). There is need for the identification and isolation of additional genes encoding Δ4 desaturases that will be suitable for heterologous expression in a variety of host organisms for use in the production of ω-3/ω-6 fatty acids.

[0006] Applicants have solved the stated problem by isolating genes encoding Δ4 fatty acid desaturases from *Eutreptiella cf_gymnastica* CCMP1594.

SUMMARY OF THE INVENTION

[0007] Described herein are new genetic constructs encoding polypeptides having Δ4 desaturase activity, and their use in algae, bacteria, yeast, euglenoids, stramenopiles, fungi, plants and animals for the production of polyunsaturated fatty acids ["PUFAs"].

[0008] Described herein are isolated nucleic acid molecules selected from the group consisting of:

(a) an isolated nucleotide sequence encoding a Δ4 desaturase enzyme selected from the group consisting of SEQ ID NO:2 and SEQ ID NO:4;

(b) an isolated nucleotide sequence that hybridizes with (a) under the following hybridization conditions: 0.1X SSC, 0.1% SDS, 65°C and washed with 2X SSC, 0.1% SDS followed by 0.1X SSC, 0.1% SDS; and,

(c) an isolated nucleotide sequence that is completely complementary to (a) or (b).

Other isolated nucleic acid molecules described herein comprise a first nucleotide sequence encoding a Δ4 desaturase enzyme of at least 514 amino acids that has at least 68% identity based on the Clustal W method of alignment when compared to a polypeptide having the sequence as set forth in SEQ ID NO:2; or a second nucleotide sequence comprising the complement of the first nucleotide sequence.

[0009] Also described herein are genetic chimera of the nucleic acid molecules described herein and transformed host cells comprising them. In addition, described herein are methods for the production of docosahexaenoic acid is provided herein, comprising:

a) providing a host cell comprising:

(i) an isolated nucleotide molecule encoding a Δ4 desaturase polypeptide having at least 68% identity when compared to a polypeptide having the amino acid sequence as set forth in SEQ ID NO:2, based on the Clustal W method of alignment; and,

(ii) a source of all-*cis*-7,10,13,16,19-docosapentaenoic acid (22:5, ω3);

b) growing the host cell of step (a) under conditions wherein the nucleic acid fragment encoding the Δ4 desaturase polypeptide is expressed and the all-*cis*-7,10,13,16,19-

docosapentaenoic acid (22:5, ω3) is converted to docosahexaenoic acid; and,

c) optionally recovering the docosahexaenoic acid of step (b).

Similarly, a method for the production of all-*cis*-4,7,10,13,16-docosapentaenoic acid (22:5, ω–6) is provided, comprising:

a) providing a host cell comprising:

(i) an isolated nucleotide molecule encoding a Δ4 desaturase polypeptide having at least 68% identity when compared to a polypeptide having the amino acid sequence as set forth in SEQ ID NO:2, based on the Clustal W method of alignment; and,

(ii) a source of docosatetraenoic acid;

b) growing the host cell of step (a) under conditions wherein the nucleic acid fragment encoding the Δ4 desaturase polypeptide is expressed and the docosatetraenoic acid is converted to all-*cis*-4,7,10,13,16-docosapentaenoic acid (22:5, ω–6); and,

c) optionally recovering the all-*cis*-4,7,10,13,16-docosapentaenoic acid (22:5, ω–6) of step (b).

BIOLOGICAL DEPOSITS

[0010] The following biological material was made under the terms of the Budapest Treaty on the International Recognition of the Deposit of Microorganisms for the Purposes of Patent Procedure:

Biological Material	Accession Number	Date of Deposit
Yarrowia lipolytica Y4127	ATCC PTA-8802	November 29, 2007

As used herein, "ATCC" refers to the American Type Culture Collection International Depository Authority located at ATCC, 10801 University Blvd., Manassas, VA 20110-2209, U.S.A. The listed deposit will be maintained in the indicated international depository for at least 30 years and will be made available to the public upon the grant of a patent disclosing it. The availability of a deposit does not constitute a license to

practice the subject invention in derogation of patent rights granted by government action.

BRIEF DESCRIPTION OF THE DRAWINGS AND SEQUENCE LISTINGS

[0011] The invention can be more fully understood from the following detailed description and the accompanying sequence descriptions, which form a part of this application.

FIG. 1 includes FIG. 1A and FIG 1B, which together illustrate the ω-3/ω-6 fatty acid biosynthetic pathway, and should be viewed together when considering the description of this pathway.

FIG. 2 includes FIG. 2A, FIG. 2B and FIG. 2C, which together show an alignment between and among the *Euglena gracilis* Δ4 fatty acid desaturase (SEQ ID NO:13; GenBank Accession No. AY278558), *Thalassiosira pseudonana* Δ4 fatty acid desaturase (SEQ ID NO:37; GenBank Accession No. AAX14506), *Thraustochytrium* sp. FJN-10 Δ4 fatty acid desaturase (SEQ ID NO:38; GenBank Accession No. AAZ43257), and *Pavlova lutheri* Δ4 fatty acid desaturase (SEQ ID NO:42; GenBank Accession No. AAQ98793), using a Clustal W analysis (MegAlign™ program of DNASTAR software). Degenerate primers were designed to correspond to the boxed regions.

FIG. 3 consists of FIG. 3A, FIG. 3B, FIG. 3C, FIG. 3D and FIG. 3E, which together show a comparison of the DNA sequence of the *Eutreptiella cf_gymnastica* CCMP1594 Δ4 desaturase gene (designated as E1594D4; SEQ ID NO:1) and the synthetic gene (designated as E1594D4S; SEQ ID NO:3) codon-optimized for expression in *Yarrowia lipolytica*.

FIG. 4 provides plasmid maps for the following: (A) p1594D4S; and, (B) pZKL4-220ESC4.

FIG. 5 diagrams the development of *Yarrowia lipolytica* strain Y4184U, producing about 31% EPA in the total lipid fraction.

[0012] The following sequences comply with 37 C.F.R. §1.821-1.825 ("Requirements for Patent Applications Containing Nucleotide Sequences and/or Amino Acid Sequence Disclosures - the Sequence Rules") and are consistent with World Intellectual Property Organization (WIPO) Standard ST.25 (1998) and the sequence listing requirements of the EPO and PCT (Rules 5.2 and 49.5(a-bis), and Section 208 and Annex C of the Administrative Instructions). The symbols and format used for nucleotide and amino acid sequence data comply with the rules set forth in 37 C.F.R. §1.822.

SEQ ID NOs:1-9, 13-14, 37-47 are ORFs encoding genes or proteins (or portions thereof), or plasmids, as identified in Table 1.

Table 1

Summary Of Nucleic Acid And Protein SEQ ID Numbers

Description and Abbreviation	**Nucleic acid SEQ ID NO.**	**Protein SEQ ID NO.**
Eutreptiella cf_gymnastica CCMP1594 Δ4 desaturase ("E1594D4")	1 (1545 bp)	2 (514 AA)
Synthetic Δ4 desaturase derived from *Eutreptiella cf_gymnastica* CCMP1594, codon-optimized for expression in *Yarrowia lipolytica* ("E1594D4S")	3 (1548 bp)	4 (515 AA)
Eutreptiella cf_gymnastica CCMP1594 Δ4 desaturase partial fragment	5 (847 bp)	--
Eutreptiella cf_gymnastica CCMP1594 Δ4 desaturase E1594D4-5'-A fragment	6 (359 bp)	--
Eutreptiella cf_gymnastica CCMP1594 Δ4 desaturase E1594D4-5'-B fragment	7 (395 bp)	--
Eutreptiella cf_gymnastica CCMP1594 Δ4 desaturase E1594D4-3' fragment	8 (873 bp)	--
Eutreptiella cf_gymnastica CCMP1594 Δ4 desaturase E1594D4-cDNA fragment	9 (2070 bp)	--
Euglena gracilis Δ4 fatty acid desaturase (GenBank Accession No. AY278558)	--	13 (541 AA)
Thraustochytrium aureum Δ4 desaturase (GenBank Accession No. AAN75707)	--	14 (515 AA)
Thalassiosira pseuduonana Δ4 fatty acid desaturase (GenBank Accession No. AAX14506)	--	37 (550 AA)
Thraustochytrium sp. FJN-10 Δ4 fatty acid desaturase (GenBank Accession No. AAZ43257)	--	38 (519 AA)
Plasmid p1594D4S	39 (4275 bp)	--
Plasmid pZKL4-220ESC4	40	--

	(13175 bp)	
Schizochytrium aggregatum Δ4 desaturase (Int'l. App. Pub. No. WO 2002/090493)	--	41 (509 AA)
Pavlova lutheri Δ4 desaturase (GenBank Accession No. AAQ98793)	--	42 (445 AA)
Isochrysis galbana Δ4 desaturase (GenBank Accession No. AAV33631)	--	43 (433 AA)
Synthetic C20 elongase derived from *Euglena gracilis*, codon-optimized for expression in *Yarrowia lipolytica* (U.S. Pat. Appl. Pub. No. 2008-0254191) ("EgC20ES")	44 (912 bp)	45 (303 AA)
Synthetic C20 elongase derived from *Euglena anabaena*, codon-optimized for expression in *Yarrowia lipolytica* (U.S. Pat. Appl. Pub. No. 2008-0254191) ("EaC20ES")	46 (900 bp)	47 (299 AA)

SEQ ID NOs:10-12 correspond to SMART™ IV oligonucleotide primer, CDSIII/3' PCR primer and 5' CDSIII PCR primer, respectively, used for *Eutreptiella cf_gymnastica* CCMP1594 cDNA synthesis.

SEQ ID NOs:15-17 correspond to degenerate oligonucleotide primers D4-F1, D4-F2 and D4-F3, respectively, all of which encode the peptide set forth in SEQ ID NO:18.

SEQ ID NO:19 corresponds to degenerate oligonucleotide primer D4-F4, which encodes the peptide set forth in SEQ ID NO:20.

SEQ ID NO:21 corresponds to degenerate oligonucleotide primer D4-F5, which encodes the peptide set forth in SEQ ID NO:22.

SEQ ID NOs:23-25 correspond to degenerate oligonucleotide primers D4-F6, D4-F7 and D4-F8, respectively, all of which encode the peptide set forth in SEQ ID NO:26.

SEQ ID NOs:27 and 28 correspond to degenerate oligonucleotide primers D4-R1 and D4-R2, both of which encode the peptide set forth in SEQ ID NO:29.

SEQ ID NOs:30-34 correspond to primers 1594D4-5-1, 1594D4-5-2, DNR CDS 5-2, 1594D4-5-4 and 1594D4-5-5, respectively, used to amplify the 5' coding region of the *Eutreptiella cf_gymnastica* CCMP1594 Δ4 desaturase gene.

SEQ ID NOs:35 and 36 correspond to primers 1594D4-3-1 and 1594D4-3-2, respectively, used to amplify the 3' coding region of the *Eutreptiella cf_gymnastica* CCMP1594 Δ4 desaturase gene.

DETAILED DESCRIPTION OF THE INVENTION

[0013] Applicants have identified a novel *Eutreptiella cf_gymnastica* CCMP1594 Δ4 desaturase enzyme and gene encoding the same that may be used for the manipulation of biochemical pathways for the production of healthful PUFAs. Thus, the subject invention finds many applications.

[0014] PUFAs, or derivatives thereof, are used as dietary substitutes, or supplements, particularly infant formulas, for patients undergoing intravenous feeding or for preventing or treating malnutrition. Alternatively, the purified PUFAs and derivatives thereof may be incorporated into cooking oils, fats or margarines and ingested as part of a consumer's typical diet, thereby giving the consumer desired dietary supplementation. Further, PUFAs may also be incorporated into infant formulas, nutritional supplements or other food products and may find use as anti-inflammatory or cholesterol lowering agents. Optionally, the compositions may be used for pharmaceutical use, either human or veterinary.

Definitions

[0015] In this disclosure, a number of terms and abbreviations are used. The following definitions are provided.

"Open reading frame" is abbreviated as "ORF".

"Polymerase chain reaction" is abbreviated as "PCR".

"American Type Culture Collection" is abbreviated as "ATCC".

"Polyunsaturated fatty acid(s)" is abbreviated as "PUFA(s)".

"Triacylglycerols" are abbreviated as "TAGs".

"Total fatty acids" are abbreviated as "TFAs".

"Dry cell weight" is abbreviated as "DCW".

The term “invention” or “present invention” as used herein is not meant to be limiting but applies generally to any of the inventions in the claims or described herein.

The term “fatty acids” refers to long chain aliphatic acids (alkanoic acids) of varying chain lengths, from about C_{12} to C_{22}, although both longer and shorter chain-length acids are known. The predominant chain lengths are between C_{16} and C_{22}. The structure of a fatty acid is represented by a simple notation system of “X:Y”, where X is the total number of carbon ["C"] atoms in the particular fatty acid and Y is the number of double bonds. Additional details concerning the differentiation between “saturated fatty acids” versus “unsaturated fatty acids”, “monounsaturated fatty acids” versus “polyunsaturated fatty acids” [“PUFAs”], and “omega-6 fatty acids” [“ω-6” or “*n*-6”] versus “omega-3 fatty acids” [“ω-3” or “*n*-3”] are provided in U.S. Pat. No. 7,238,482.

Nomenclature used to describe PUFAs herein is given in Table 2. In the column titled “Shorthand Notation”, the omega-reference system is used to indicate the number of carbons, the number of double bonds and the position of the double bond closest to the omega carbon, counting from the omega carbon, which is numbered 1 for this purpose. The remainder of the Table summarizes the common names of ω-3 and ω-6 fatty acids and their precursors, the abbreviations that are used throughout the specification and the chemical name of each compound.

Table 2

Nomenclature of Polyunsaturated Fatty Acids And Precursors

Common Name	Abbreviation	Chemical Name	Shorthand Notation
Myristic	--	tetradecanoic	14:0
Palmitic	Palmitate	hexadecanoic	16:0
Palmitoleic	--	9-hexadecenoic	16:1
Stearic	--	octadecanoic	18:0
Oleic	--	*cis*-9-octadecenoic	18:1
Linoleic	LA	*cis*-9, 12-octadecadienoic	18:2 ω -6
˜Linolenic	GLA	*cis*-6, 9, 12-octadecatrienoic	18:3 ω -6
Eicosadienoic	EDA	*cis*-11, 14-eicosadienoic	20:2 ω -6

Dihomo- -Linolenic	DGLA	*cis*-8, 11, 14-eicosatrienoic	20:3 ω -6
Arachidonic	ARA	*cis*-5, 8, 11, 14-eicosatetraenoic	20:4 ω -6
˜Linolenic	ALA	*cis*-9, 12, 15-octadecatrienoic	18:3 ω -3
Stearidonic	STA	*cis*-6, 9, 12, 15-octadecatetraenoic	18:4 ω -3
Eicosatrienoic	ETrA	*cis*-11, 14, 17-eicosatrienoic	20:3 ω -3
Sciadonic	SCI	*cis*-5,11,14-eicosatrienoic	20:3b ω -6
Juniperonic	JUP	*cis*-5,11,14, 17-eicosatetraenoic	20:4b ω -3
Eicosa-tetraenoic	ETA	*cis*-8, 11, 14, 17-eicosatetraenoic	20:4 ω -3
Eicosa-pentaenoic	EPA	*cis*-5, 8, 11, 14, 17-eicosapentaenoic	20:5 ω -3
Docosatrienoic	DRA	*cis*-10, 13, 16-docosatrienoic	22:3 ω -6
Docosa-tetraenoic	DTA	*cis*-7, 10, 13, 16-docosatetraenoic	22:4 ω -6
Docosa-pentaenoic	DPAn-6	*cis*-4, 7, 10, 13, 16-docosapentaenoic	22:5 ω -6
Docosa-pentaenoic	DPA	*cis*-7, 10, 13, 16, 19-docosapentaenoic	22:5 ω -3
Docosa-hexaenoic	DHA	*cis*-4, 7, 10, 13, 16, 19-docosahexaenoic	22:6 ω -3

Although the ω -3/ ω-6 PUFAs listed in Table 2 are the most likely to be accumulated in the oil fractions of oleaginous yeast using the methods described herein, this list should not be construed as limiting or as complete.

The term “total lipid fraction” of cells herein refers to all esterified fatty acids of the cell. Various subfractions within the total lipid fraction can be isolated, including the triacylglycerol ["oil"] fraction, phosphatidylcholine fraction and the phosphatidyletanolamine fraction, although this is by no means inclusive of all sub-fractions.

The terms “triacylglycerols” [“TAGs”] and “oil” are interchangeable and refer to neutral lipids composed of three fatty acyl residues esterified to a glycerol molecule. TAGs can contain long chain PUFAs, as well as shorter saturated and unsaturated fatty acids and longer chain saturated fatty acids. The TAG fraction of cells is also referred to as the “oil

fraction", and "oil biosynthesis" generically refers to the synthesis of TAGs in the cell. The oil or TAG fraction is a sub-fraction of the total lipid fraction, although also it constitutes a major part of the total lipid content, measured as the weight of total fatty acids in the cell as a percent of the dry cell weight [see below], in oleaginous organisms. The fatty acid composition in the oil ["TAG"] fraction and the fatty acid composition of the total lipid fraction are generally similar. Thus, an increase or decrease in the concentration of PUFAs in the total lipid fraction will correspond with an increase or decrease in the concentration of PUFAs in the oil ["TAG"] fraction, and vice versa.

The term "total fatty acids" ["TFAs"] herein refer to the sum of all cellular fatty acids that can be derivitized to fatty acid methyl esters ["FAMEs"] by the base transesterification method (as known in the art) in a given sample, which may be the total lipid fraction or the oil fraction, for example. Thus, total fatty acids include fatty acids from neutral and polar lipid fractions, including the phosphatidylcholine fraction, the phosphatidyletanolamine fraction and the diacylglycerol, monoacylglycerol and triacylglycerol ["TAG or oil"] fractions but not free fatty acids.

The term "total lipid content" of cells is a measure of TFAs as a percent of the dry cell weight ["DCW"]. Thus, total lipid content ["TFAs % DCW"] is equivalent to, e.g., milligrams of total fatty acids per 100 milligrams of DCW.

Generally, the concentration of a fatty acid is expressed herein as a weight percent of TFAs ["% TFAs"], e.g., milligrams of the given fatty acid per 100 milligrams of TFAs. Unless otherwise specifically stated in the disclosure herein, reference to the percent of a given fatty acid with respect to total lipids is equivalent to concentration of the fatty acid as % TFAs, e.g., % DHA of total lipids is equivalent to DHA % TFAs.

In some cases, it is useful to express the content of a given fatty acid(s) in a cell as its percent of the dry cell weight ["% DCW"]. Thus, for example, docosahexaenoic acid % DCW would be determined according to the following formula: (docosahexaenoic acid % TFAs) * (TFA % DCW)]/100.

The terms “lipid profile” and “lipid composition” are interchangeable and refer to the amount of an individual fatty acid contained in a particular lipid fraction, such as in the total lipid fraction or the oil ["TAG"] fraction, wherein the amount is expressed as a percent of TFAs. The sum of each individual fatty acid present in the mixture should be 100.

A metabolic pathway, or biosynthetic pathway, in a biochemical sense, can be regarded as a series of chemical reactions occurring within a cell, catalyzed by enzymes, to achieve either the formation of a metabolic product to be used or stored by the cell, or the initiation of another metabolic pathway, which is termed "flux generating step". Many of these pathways are elaborate, and involve a step by step modification of the initial substance to shape it into a product having the exact chemical structure desired.

The term “PUFA biosynthetic pathway” refers to a metabolic process that converts oleic acid to ω-6 fatty acids such as LA, EDA, GLA, DGLA, ARA, DRA, DTA and DPAn-6 and ω-3 fatty acids such as ALA, STA, ETrA, ETA, EPA, DPA and DHA. This process is well described in the literature. See e.g., Int'. App. Pub. No. WO 2006/052870. Briefly, this process involves elongation of the carbon chain through the addition of carbon atoms and desaturation of the elongated molecule through the addition of double bonds, via a series of special elongation and desaturation enzymes termed “PUFA biosynthetic pathway enzymes” that are present in the endoplasmic reticulum membrane. More specifically, “PUFA biosynthetic pathway enzymes” refer to any of the following enzymes (and genes which encode said enzymes) associated with the biosynthesis of a PUFA, including: a Δ9 elongase, a $C_{14/16}$ elongase, a $C_{16/18}$ elongase, a $C_{18/20}$ elongase, a $C_{20/22}$ elongase. a Δ4 desaturase, a Δ5 desaturase, a Δ6 desaturase, a Δ8 desaturase, Δ9 desaturase, a Δ12 desaturase, a Δ 5 desaturase and/or a Δ17 desaturase.

The term “ω-3/ ω -6 fatty acid biosynthetic pathway” refers to a set of genes which, when expressed under the appropriate conditions encode enzymes that catalyze the production of either or both ω-3 and ω-6 fatty

acids. Typically the genes involved in the ω-3/ω-6 fatty acid biosynthetic pathway encode PUFA biosynthetic pathway enzymes. A representative pathway is illustrated in FIG. 1, providing for the conversion of myristic acid through various intermediates to DHA, which demonstrates how both ω-3 and ω-6 fatty acids may be produced from a common source. The pathway is naturally divided into two portions, such that one portion generates only ω-3 fatty acids and the other portion, only ω-6 fatty acids. That portion that only generates only ω-3 fatty acids is referred to herein as the ω-3 fatty acid biosynthetic pathway, whereas that portion that generates only ω-6 fatty acids is referred to herein as the ω-6 fatty acid biosynthetic pathway.

The term "functional" as used herein relating to the ω-3/ω-6 fatty acid biosynthetic pathway, means that some (or all) of the genes in the pathway express active enzymes, resulting in *in vivo* catalysis or substrate conversion. It should be understood that "ω-3/ω-6 fatty acid biosynthetic pathway" or "functional ω-3/ω-6 fatty acid biosynthetic pathway" does not imply that all of the genes listed in the above paragraph are required, as a number of fatty acid products require only the expression of a subset of the genes of this pathway.

The term "ω6 desaturase/ω6 elongase pathway" refers to a PUFA biosynthetic pathway that minimally includes at least one ω6 desaturase and at least one $C_{18/20}$ elongase, thereby enabling biosynthesis of DGLA and/or ETA from LA and ALA, respectively, with GLA and/or STA as intermediate fatty acids. With expression of other desaturases and elongases, ARA, DTA, DPAn-6, EPA, DPA and DHA may also be synthesized.

The term "Δ9 elongase/Δ8 desaturase pathway" refers to a PUFA biosynthetic pathway that minimally includes at least one Δ9 elongase and at least one Δ8 desaturase, thereby enabling biosynthesis of DGLA and/or ETA from LA and ALA, respectively, with EDA and/or ETrA as intermediate fatty acids. With expression of other desaturases and

elongases, ARA, DTA, DPAn-6, EPA, DPA and DHA may also be synthesized.

The term “intermediate fatty acid” refers to any fatty acid produced in a fatty acid metabolic pathway that can be further converted to an intended product fatty acid in this pathway by the action of other metabolic pathway enzymes. For instance, when EPA is produced using the Δ9 elongase/Δ8 desaturase pathway, EDA, ETrA, DGLA, ETA and ARA can be produced and are considered “intermediate fatty acids” since these fatty acids can be further converted to EPA via action of other metabolic pathway enzymes.

The term “desaturase" refers to a polypeptide that can desaturate adjoining carbons in a fatty acid by removing a hydrogen from one of the adjoining carbons and thereby introducing a double bond between them. Desaturation produces a fatty acid or precursor of interest. Despite use of the omega-reference system throughout the specification to refer to specific fatty acids, it is more convenient to indicate the activity of a desaturase by counting from the carboxyl end of the substrate using the delta-system. Of particular interest herein are Δ4 desaturases that catalyze the conversion of the substrate fatty acid, DPA, to DHA and/or or the conversion of the substrate fatty acid, DTA, to DPAn-6. Other desaturases include: 1) Δ17 desaturases that desaturate a fatty acid between the 17th and 18th carbon atom numbered from the carboxyl-terminal end of the molecule and which, for example, catalyze the conversion of the substrate fatty acid, ARA, to EPA and/or the conversion of the substrate fatty acid, DGLA, to ETA; 2) Δ6 desaturases that catalyze the conversion of the substrate fatty acid, LA, to GLA and/or the conversion of the substrate fatty acid, ALA, to STA; 3) Δ12 desaturases that catalyze the conversion of the substrate fatty acid, oleic acid, to LA; 4) Δ15 desaturases that catalyze the conversion of the substrate fatty acid, LA, to ALA and/or the conversion of the substrate fatty acid, GLA, to STA; 5) Δ5 desaturases that catalyze the conversion of the substrate fatty acid, DGLA, to ARA and/or the conversion of the substrate fatty acid, ETA, to

EPA; 6) Δ8 desaturases that catalyze the conversion of the substrate fatty acid, EDA, to DGLA and/or the conversion of the substrate fatty acid, ETrA, to ETA; and, 7) Δ9 desaturases that catalyze the conversion of the substrate fatty acid, palmitate, to palmitoleic acid (16:1) and/or the conversion of the substrate fatty acid, stearic acid, to oleic acid. In the art, Δ15 and Δ17 desaturases are also occasionally referred to as "omega-3 desaturases", "w-3 desaturases", and/or "ω-3 desaturases", based on their ability to convert ω-6 fatty acids into their ω-3 counterparts (e.g., conversion of LA into ALA and ARA into EPA, respectively). It may be desirable to empirically determine the specificity of a particular fatty acid desaturase by transforming a suitable host with the gene for the fatty acid desaturase and determining its effect on the fatty acid profile of the host.

The term "E1594D4" refers to a Δ4 desaturase enzyme (SEQ ID NO:2) isolated from *Eutreptiella cf_gymnastica* CCMP1594, encoded by SEQ ID NO:1 herein. Similarly, the term "E1594D4S" refers to a synthetic Δ4 desaturase derived from *Eutreptiella cf_gymnastica* CCMP1594 that is codon-optimized for expression in *Yarrowia lipolytica* (i.e., SEQ ID NOs:3 and 4).

The terms "conversion efficiency" and "percent substrate conversion" refer to the efficiency by which a particular enzyme, such as a desaturase, can convert substrate to product. The conversion efficiency is measured according to the following formula: ([product]/[substrate + product])*100, where 'product' includes the immediate product and all products in the pathway derived from it.

The term "elongase" refers to a polypeptide that can elongate a fatty acid carbon chain to produce an acid that is 2 carbons longer than the fatty acid substrate that the elongase acts upon. This process of elongation occurs in a multi-step mechanism in association with fatty acid synthase, as described in Int'l App. Pub. No. WO 2005/047480. Examples of reactions catalyzed by elongase systems are the conversion of GLA to DGLA, ARA to DTA, STA to ETA and EPA to DPA. In general, the substrate selectivity of elongases is somewhat broad but segregated by both chain length and the degree and type of unsaturation. For example,

a $C_{14/16}$ elongase utilizes a C_{14} substrate e.g., myristic acid; a $C_{16/18}$ elongase utilizes a C_{16} substrate e.g., palmitate; a $C_{18/20}$ elongase [also known as a Δ6 elongase as the terms can be used interchangeably] utilizes a C_{18} substrate e.g., GLA, STA; and, a $C_{20/22}$ elongase [also known as a C20 elongase as the terms can be used interchangeably] utilizes a C_{20} substrate e.g., ARA, EPA. In like manner, a Δ9 elongase is able to catalyze the conversion of LA and ALA to EDA and ETrA, respectively.

It is important to note that some elongases have broad specificity and thus a single enzyme may be capable of catalyzing several elongase reactions. For example, a single enzyme may thus act as both a $C_{16/18}$ elongase and a $C_{18/20}$ elongase. It may be desirable to empirically determine the specificity of a fatty acid elongase by transforming a suitable host with the gene for the fatty acid elongase and determining its effect on the fatty acid profile of the host.

The term "oleaginous" refers to those organisms that tend to store their energy source in the form of lipid (Weete, In: Fungal Lipid Biochemistry, 2nd Ed., Plenum, 1980). The term "oleaginous yeast" refers to those microorganisms classified as yeasts that can make oil, that is, TAGs. Generally, the cellular oil or TAG content of oleaginous microorganisms follows a sigmoid curve, wherein the concentration of lipid increases until it reaches a maximum at the late logarithmic or early stationary growth phase and then gradually decreases during the late stationary and death phases (Yongmanitchai and Ward, *Appl. Environ. Microbiol.*, 57:419-25 (1991)). It is not uncommon for oleaginous microorganisms to accumulate in excess of about 25% of their dry cell weight as oil. Examples of oleaginous yeast include, but are no means limited to, the following genera: *Yarrowia, Candida, Rhodotorula, Rhodosporidium*, *Cryptococcus, Trichosporon* and *Lipomyces*.

The term "Euglenophyceae" refers to a group of unicellular colorless or photosynthetic flagellates ["euglenoids"] found living in freshwater, marine, soil and parasitic environments. The class is

characterized by solitary unicells, wherein most are free-swimming and have two flagella, one of which may be nonemergent, arising from an anterior invagination known as a reservoir. Photosynthetic euglenoids contain one to many chloroplasts, which vary from minute disks to expanded plates or ribbons. Colorless euglenoids depend on osmotrophy or phagotrophy for nutrient assimilation. About 1000 species have been described and classified into about 40 genera and 6 orders. Examples of Euglenophyceae include, but are no means limited to, the following genera: *Euglena, Eutreptiella* and *Tetruetreptia.*

As used herein the term "biomass" refers specifically to spent or used yeast cellular material from the fermentation of a recombinant production host producing PUFAs in commercially significant amounts, wherein the preferred production host is a recombinant strain of the oleaginous yeast, *Yarrowia lipolytica*. The biomass may be in the form of whole cells, whole cell lysates, homogenized cells, partially hydrolyzed cellular material, and/or partially purified cellular material e.g., microbially produced oil.

As used herein, the terms "isolated nucleic acid fragment" and "isolated nucleic acid molecule" are used interchangeably and refer to a polymer of RNA or DNA that is single- or double-stranded, optionally containing synthetic, non-natural or altered nucleotide bases. An isolated nucleic acid fragment in the form of a polymer of DNA may be comprised of one or more segments of cDNA, genomic DNA or synthetic DNA.

A nucleic acid fragment is "hybridizable" to another nucleic acid fragment, such as a cDNA, genomic DNA, or RNA molecule, when a single-stranded form of the nucleic acid fragment can anneal to the other nucleic acid fragment under the appropriate conditions of temperature and solution ionic strength. Hybridization and washing conditions are well known and exemplified in Sambrook, J., Fritsch, E. F. and Maniatis, T. Molecular Cloning: A Laboratory Manual, 2nd ed., Cold Spring Harbor Laboratory: Cold Spring Harbor, NY (1989).

The conditions of temperature and ionic strength determine the "stringency" of the hybridization. Stringency conditions can be adjusted to

screen for moderately similar fragments (such as homologous sequences from distantly related organisms), to highly similar fragments (such as genes that duplicate functional enzymes from closely related organisms). Post-hybridization washes determine stringency conditions. One set of preferred conditions uses a series of washes starting with 6X SSC, 0.5% SDS at room temperature for 15 min, then repeated with 2X SSC, 0.5% SDS at 45 °C for 30 min, and then repeated twice with 0.2X SSC, 0.5% SDS at 50 °C for 30 min. A more preferred set of stringent conditions uses higher temperatures in which the washes are identical to those above except for the temperature of the final two 30 min washes in 0.2X SSC, 0.5% SDS was increased to 60 °C. Another preferred set of highly stringent conditions uses two final washes in 0.1X SSC, 0.1% SDS at 65 °C. An additional set of stringent conditions include hybridization at 0.1X SSC, 0.1% SDS, 65 °C and washes with 2X SSC, 0.1% SDS followed by 0.1X SSC, 0.1% SDS, for example.

Hybridization requires that the two nucleic acids contain complementary sequences, although depending on the stringency of the hybridization, mismatches between bases are possible. The appropriate stringency for hybridizing nucleic acids depends on the length of the nucleic acids and the degree of complementation, variables well known in the art. The greater the degree of similarity or homology between two nucleotide sequences, the greater the value of T*m* for hybrids of nucleic acids having those sequences. The relative stability, corresponding to higher T*m*, of nucleic acid hybridizations decreases in the following order: RNA:RNA, DNA:RNA, DNA:DNA. For hybrids of greater than 100 nucleotides in length, equations for calculating T*m* have been derived (see Sambrook et al., *supra*, 9.50-9.51). For hybridizations with shorter nucleic acids, i.e., oligonucleotides, the position of mismatches becomes more important, and the length of the oligonucleotide determines its specificity (see Sambrook et al., *supra*, 11.7-11.8). In one embodiment the length for a hybridizable nucleic acid is at least about 10 nucleotides. Preferably a minimum length for a hybridizable nucleic acid is at least about 15 nucleotides; more preferably

at least about 20 nucleotides; and most preferably the length is at least about 30 nucleotides. Furthermore, the skilled artisan will recognize that the temperature and wash solution salt concentration may be adjusted as necessary according to factors such as length of the probe.

A "substantial portion" of an amino acid or nucleotide sequence is that portion comprising enough of the amino acid sequence of a polypeptide or the nucleotide sequence of a gene to putatively identify that polypeptide or gene, either by manual evaluation of the sequence by one skilled in the art, or by computer-automated sequence comparison and identification using algorithms such as BLAST (Basic Local Alignment Search Tool; Altschul, S. F., et al., *J. Mol. Biol.*, 215:403-410 (1993)). In general, a sequence of ten or more contiguous amino acids or thirty or more nucleotides is necessary in order to putatively identify a polypeptide or nucleic acid sequence as homologous to a known protein or gene. Moreover, with respect to nucleotide sequences, gene specific oligonucleotide probes comprising 20-30 contiguous nucleotides may be used in sequence-dependent methods of gene identification (e.g., Southern hybridization) and isolation (e.g., *in situ* hybridization of bacterial colonies or bacteriophage plaques). In addition, short oligonucleotides of 12-15 bases may be used as amplification primers in PCR in order to obtain a particular nucleic acid fragment comprising the primers. Accordingly, a "substantial portion" of a nucleotide sequence comprises enough of the sequence to specifically identify and/or isolate a nucleic acid fragment comprising the sequence. The disclosure herein teaches the complete amino acid and nucleotide sequence encoding particular euglenoid proteins. The skilled artisan, having the benefit of the sequences as reported herein, may now use all or a substantial portion of the disclosed sequences for purposes known to those skilled in this art. Accordingly, the complete sequences as reported in the accompanying Sequence Listing, as well as substantial portions of those sequences as defined above, are encompassed in the present disclosure.

The term "complementary" is used to describe the relationship between nucleotide bases that are capable of hybridizing to one another.

For example, with respect to DNA, adenosine is complementary to thymine and cytosine is complementary to guanine. Accordingly, isolated nucleic acid fragments that are complementary to the complete sequences as reported in the accompanying Sequence Listing, as well as those substantially similar nucleic acid sequences, are encompassed in the present disclosure.

The terms "homology" and "homologous" are used interchangeably herein. They refer to nucleic acid fragments wherein changes in one or more nucleotide bases do not affect the ability of the nucleic acid fragment to mediate gene expression or produce a certain phenotype. These terms also refer to modifications of the nucleic acid fragments described herein, such as deletion or insertion of one or more nucleotides that do not substantially alter the functional properties of the resulting nucleic acid fragment relative to the initial, unmodified fragment. It is therefore understood, as those skilled in the art will appreciate, that the invention encompasses more than the specific exemplary sequences.

Moreover, the skilled artisan recognizes that homologous nucleic acid sequences encompassed by this invention are also defined by their ability to hybridize, under moderately stringent conditions, such as 0.5 X SSC, 0.1% SDS, 60 °C, with the sequences exemplified herein, or to any portion of the nucleotide sequences disclosed herein and which are functionally equivalent thereto.

"Codon degeneracy" refers to the nature in the genetic code permitting variation of the nucleotide sequence without effecting the amino acid sequence of an encoded polypeptide. Accordingly, described herein is any nucleic acid fragment that encodes all or a substantial portion of the amino acid sequence encoding the euglenoid polypeptide as set forth in SEQ ID NO:2 and/or SEQ ID NO:4. The skilled artisan is well aware of the "codon-bias" exhibited by a specific host cell in usage of nucleotide codons to specify a given amino acid. Therefore, when synthesizing a gene for improved expression in a host cell, it is desirable to design the gene such that its frequency of codon usage approaches the frequency of preferred codon usage of the host cell.

"Synthetic genes" can be assembled from oligonucleotide building blocks that are chemically synthesized using procedures known to those skilled in the art. These oligonucleotide building blocks are annealed and then ligated to form gene segments that are then enzymatically assembled to construct the entire gene. Accordingly, the genes can be tailored for optimal gene expression based on optimization of nucleotide sequence to reflect the codon bias of the host cell. The skilled artisan appreciates the likelihood of successful gene expression if codon usage is biased towards those codons favored by the host. Determination of preferred codons can be based on a survey of genes derived from the host cell, where sequence information is available. For example, the codon usage profile for *Yarrowia lipolytica* is provided in U.S. Pat. No. 7,125,672.

"Gene" refers to a nucleic acid fragment that expresses a specific protein, and which may refer to the coding region alone or may include regulatory sequences preceding (5' non-coding sequences) and following (3' non-coding sequences) the coding sequence. "Native gene" refers to a gene as found in nature with its own regulatory sequences. "Chimeric gene" refers to any gene that is not a native gene, comprising regulatory and coding sequences that are not found together in nature. Accordingly, a chimeric gene may comprise regulatory sequences and coding sequences that are derived from different sources, or regulatory sequences and coding sequences derived from the same source, but arranged in a manner different than that found in nature. "Endogenous gene" refers to a native gene in its natural location in the genome of an organism. A "foreign" gene refers to a gene that is introduced into the host organism by gene transfer. Foreign genes can comprise native genes inserted into a non-native organism, native genes introduced into a new location within the native host, or chimeric genes. A "transgene" is a gene that has been introduced into the genome by a transformation procedure. A "codon-optimized gene" is a gene having its frequency of codon usage designed to mimic the frequency of preferred codon usage of the host cell.

"Coding sequence" refers to a DNA sequence that codes for a specific amino acid sequence. "Suitable regulatory sequences" refer to

nucleotide sequences located upstream (5' non-coding sequences), within, or downstream (3' non-coding sequences) of a coding sequence (or located within an intron thereof), and which influence the transcription, RNA processing or stability, or translation of the associated coding sequence. Regulatory sequences may include promoters, enhancers, silencers, 5' untranslated leader sequence (e.g., between the transcription start site and the translation initiation codon), introns, polyadenylation recognition sequences, RNA processing sites, effector binding sites and stem-loop structures.

"Promoter" refers to a DNA sequence capable of controlling the expression of a coding sequence or functional RNA. In general, a coding sequence is located 3' to a promoter sequence. Promoters may be derived in their entirety from a native gene, or be composed of different elements derived from different promoters found in nature, or even comprise synthetic DNA segments. It is understood by those skilled in the art that different promoters may direct the expression of a gene in different tissues or cell types, or at different stages of development, or in response to different environmental or physiological conditions. Promoters that cause a gene to be expressed in most cell types at most times are commonly referred to as "constitutive promoters". It is further recognized that since in most cases the exact boundaries of regulatory sequences have not been completely defined, DNA fragments of different lengths may have identical promoter activity.

The terms "3' non-coding sequences" and "transcription terminator" refer to DNA sequences located downstream of a coding sequence. This includes polyadenylation recognition sequences and other sequences encoding regulatory signals capable of affecting mRNA processing or gene expression. The polyadenylation signal is usually characterized by affecting the addition of polyadenylic acid tracts to the 3' end of the mRNA precursor. The 3' region can influence the transcription, RNA processing or stability, or translation of the associated coding sequence.

"RNA transcript" refers to the product resulting from RNA polymerase-catalyzed transcription of a DNA sequence. When the RNA

transcript is a perfect complementary copy of the DNA sequence, it is referred to as the primary transcript or it may be a RNA sequence derived from post-transcriptional processing of the primary transcript and is referred to as the mature RNA. "Messenger RNA" or "mRNA" refers to the RNA that is without introns and which can be translated into protein by the cell. "cDNA" refers to a double-stranded DNA that is complementary to, and derived from, mRNA. "Sense" RNA refers to RNA transcript that includes the mRNA and so can be translated into protein by the cell. "Antisense RNA" refers to a RNA transcript that is complementary to all or part of a target primary transcript or mRNA and that blocks the expression of a target gene (U.S. Pat. No. 5,107,065; Int'l. App. Pub. No. WO 99/28508). The complementarity of an antisense RNA may be with any part of the specific gene transcript, i.e., at the 5' non-coding sequence, 3' non-coding sequence, or the coding sequence. "Functional RNA" refers to antisense RNA, ribozyme RNA, or other RNA that is not translated and yet has an effect on cellular processes.

The term "operably linked" refers to the association of nucleic acid sequences on a single nucleic acid fragment so that the function of one is affected by the other. For example, a promoter is operably linked with a coding sequence when it is capable of affecting the expression of that coding sequence, i.e. the coding sequence is under the transcriptional control of the promoter. Coding sequences can be operably linked to regulatory sequences in sense or antisense orientation.

The term "expression", as used herein, refers to the transcription and stable accumulation of sense (mRNA) or antisense RNA derived from nucleic acid fragments. Expression may also refer to translation of mRNA into a polypeptide.

"Transformation" refers to the transfer of a nucleic acid molecule into a host organism, resulting in genetically stable inheritance. The nucleic acid molecule may be a plasmid that replicates autonomously, for example, or, it may integrate into the genome of the host organism. Host

organisms containing the transformed nucleic acid fragments are referred to as “transgenic” or “recombinant” or “transformed” organisms.

The terms “plasmid” and “vector” refer to an extra chromosomal element often carrying genes that are not part of the central metabolism of the cell, and usually in the form of circular double-stranded DNA fragments. Such elements may be autonomously replicating sequences, genome integrating sequences, phage or nucleotide sequences, linear or circular, of a single- or double-stranded DNA or RNA, derived from any source, in which a number of nucleotide sequences have been joined or recombined into a unique construction which is capable of introducing an expression cassette(s) into a cell.

The term “expression cassette” refers to a fragment of DNA comprising the coding sequence of a selected gene and regulatory sequences preceding (5' non-coding sequences) and following (3' non-coding sequences) the coding sequence that are required for expression of the selected gene product. Thus, an expression cassette is typically composed of: 1) a promoter sequence; 2) a coding sequence, i.e., open reading frame ["ORF"]; and, 3) a 3' untranslated region, i.e., a terminator that in eukaryotes usually contains a polyadenylation site. The expression cassette(s) is usually included within a vector, to facilitate cloning and transformation. Different expression cassettes can be transformed into different organisms including bacteria, yeast, plants and mammalian cells, as long as the correct regulatory sequences are used for each host.

The term “percent identity” refers to a relationship between two or more polypeptide sequences or two or more polynucleotide sequences, as determined by comparing the sequences. “ Percent identity” also means the degree of sequence relatedness between polypeptide or polynucleotide sequences, as the case may be, as determined by the percentage of match between compared sequences. “Percent identity” and “percent similarity” can be readily calculated by known methods, including but not limited to those described in: 1) Computational Molecular Biology (Lesk, A. M., Ed.) Oxford University: NY (1988); 2) Biocomputing: Informatics and Genome Projects (Smith, D. W., Ed.) Academic: NY

(1993); 3) Computer Analysis of Sequence Data, Part I (Griffin, A. M., and Griffin, H. G., Eds.) Humania: NJ (1994); 4) Sequence Analysis in Molecular Biology (von Heinje, G., Ed.) Academic (1987); and, 5) Sequence Analysis Primer (Gribskov, M. and Devereux, J., Eds.) Stockton: NY (1991).

Preferred methods to determine percent identity are designed to give the best match between the sequences tested. Methods to determine percent identity and percent similarity are codified in publicly available computer programs. Sequence alignments and percent identity calculations may be performed using the MegAlign™ program of the LASERGENE bioinformatics computing suite (DNASTAR Inc., Madison, WI). Multiple alignment of the sequences is performed using the "Clustal method of alignment" which encompasses several varieties of the algorithm including the "Clustal V method of alignment" and the "Clustal W method of alignment" (described by Higgins and Sharp, *CABIOS*, 5:151-153 (1989); Higgins, D.G. et al., *Comput. Appl. Biosci.,* 8:189-191 (1992)) and found in the MegAlign™ program of the LASERGENE bioinformatics computing suite (DNASTAR Inc.). After alignment of the sequences using either Clustal program, it is possible to obtain a "percent identity" by viewing the "sequence distances" table in the program.

For multiple alignments using the Clustal V method of alignment, the default values correspond to GAP PENALTY=10 and GAP LENGTH PENALTY=10. Default parameters for pairwise alignments and calculation of percent identity of protein sequences using the Clustal V method are KTUPLE=1, GAP PENALTY=3, WINDOW=5 and DIAGONALS SAVED=5. For nucleic acids these parameters are KTUPLE=2, GAP PENALTY=5, WINDOW=4 and DIAGONALS SAVED=4. Default parameters for multiple alignment using the Clustal W method of alignment correspond to GAP PENALTY=10, GAP LENGTH PENALTY=0.2, Delay Divergent Seqs(%)=30, DNA Transition Weight=0.5, Protein Weight Matrix=Gonnet Series, DNA Weight Matrix=IUB.

It is well understood by one skilled in the art that various measures of sequence percent identity are useful in identifying polypeptides, from other species, wherein such polypeptides have the same or similar function or activity. Suitable nucleic acid fragments, i.e., isolated polynucleotides according to the disclosure herein, encode polypeptides that are at least about 70% identical, preferably at least about 75% identical, and more preferably at least about 80% identical to the amino acid sequences reported herein. Preferred nucleic acid fragments encode amino acid sequences that are at least about 85% identical to the amino acid sequences reported herein. More preferred nucleic acid fragments encode amino acid sequences that are at least about 90% identical to the amino acid sequences reported herein. Most preferred are nucleic acid fragments that encode amino acid sequences that are at least about 95% identical to the amino acid sequences reported herein. Although preferred ranges are described above, any integer amino acid identity from 68% to 100% may be useful in describing the present invention, such as 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99%.

Suitable nucleic acid fragments not only have the above homologies but typically encode a polypeptide having at least 50 amino acids, preferably at least 100 amino acids, more preferably at least 150 amino acids, still more preferably at least 200 amino acids, and most preferably at least 250 amino acids.

The term “motif” means a set of amino acids conserved at specific positions along an aligned sequence of evolutionarily related proteins. While amino acids at other positions can vary between homologous proteins, amino acids that are highly conserved at specific positions indicate amino acids that are essential in the structure, the stability, or the activity of a protein. Because they are identified by their high degree of conservation in aligned sequences of a family of protein homologues, they can be used as identifiers, or “signatures”, to determine

if a protein with a newly determined sequence belongs to a previously identified protein family.

The term "sequence analysis software" refers to any computer algorithm or software program that is useful for the analysis of nucleotide or amino acid sequences. "Sequence analysis software" may be commercially available or independently developed. Typical sequence analysis software include, but is not limited to: 1) the GCG suite of programs (Wisconsin Package Version 9.0, Genetics Computer Group (GCG), Madison, WI); 2) BLASTP, BLASTN, BLASTX (Altschul et al., *J. Mol. Biol.,* 215:403-410 (1990)); 3) DNASTAR (DNASTAR, Inc. Madison, WI); 4) Sequencher (Gene Codes Corporation, Ann Arbor, MI); and, 5) the FASTA program incorporating the Smith-Waterman algorithm (W. R. Pearson, *Comput. Methods Genome Res.,* [Proc. Int. Symp.] (1994), Meeting Date 1992, 111-20. Editor(s): Suhai, Sandor. Plenum: New York, NY). Within this description, whenever sequence analysis software is used for analysis, the analytical results are based on the "default values" of the program referenced, unless otherwise specified. As used herein "default values" will mean any set of values or parameters that originally load with the software when first initialized.

Standard recombinant DNA and molecular cloning techniques used herein are well known in the art and are described by Sambrook, J., Fritsch, E. F. and Maniatis, T., Molecular Cloning: A Laboratory Manual, 2nd ed., Cold Spring Harbor Laboratory: Cold Spring Harbor, NY (1989) (hereinafter "Maniatis"); by Silhavy, T. J., Bennan, M. L. and Enquist, L. W., Experiments with Gene Fusions, Cold Spring Harbor Laboratory: Cold Spring Harbor, NY (1984); and by Ausubel, F. M. et al., Current Protocols in Molecular Biology, published by Greene Publishing Assoc. and Wiley-Interscience, Hoboken, NJ (1987).

An Overview: Microbial Biosynthesis Of Fatty Acids And Triacylglycerols

[0016] In general, lipid accumulation in oleaginous microorganisms is triggered in response to the overall carbon to nitrogen ratio present in the growth medium. This process, leading to the *de novo* synthesis of free

palmitate (16:0) in oleaginous microorganisms, is described in detail in U.S. Pat. No. 7,238,482. Palmitate is the precursor of longer-chain saturated and unsaturated fatty acid derivates, which are formed through the action of elongases and desaturases (FIG. 1).

[0017] TAGs, the primary storage unit for fatty acids, are formed by a series of reactions that involve: 1) esterification of one molecule of acyl-CoA to glycerol-3-phosphate via an acyltransferase to produce lysophosphatidic acid; 2) esterification of a second molecule of acyl-CoA via an acyltransferase to yield 1,2-diacylglycerol phosphate, commonly identified as phosphatidic acid; 3) removal of a phosphate by phosphatidic acid phosphatase to yield 1,2-diacylglycerol ["DAG"]; and, 4) addition of a third fatty acid by the action of an acyltransferase to form TAG. A wide spectrum of fatty acids can be incorporated into TAGs, including saturated and unsaturated fatty acids and short-chain and long-chain fatty acids.

Biosynthesis Of Omega Fatty Acids

[0018] The metabolic process wherein oleic acid is converted to ω-3/ω-6 fatty acids involves elongation of the carbon chain through the addition of carbon atoms and desaturation of the molecule through the addition of double bonds. This requires a series of special desaturation and elongation enzymes present in the endoplasmic reticulum membrane. However, as seen in FIG. 1 and as described below, there are often multiple alternate pathways for production of a specific ω-3/ω-6 fatty acid.

[0019] Specifically, FIG. 1 depicts the pathways described below. All pathways require the initial conversion of oleic acid to linoleic acid ["LA"], the first of the ω-6 fatty acids, by a Δ12 desaturase. Then, using the "Δ9 elongase/ Δ8 desaturase pathway" and LA as substrate, long-chain ω-6 and ω-3 fatty acids are formed as follows: 1) LA is converted to eicosadienoic acid ["EDA"] by a Δ9 elongase; 2) EDA is converted to dihomo-γ-linolenic acid ["DGLA"] by a Δ8 desaturase; 3) DGLA is converted to arachidonic acid ["ARA"] by a Δ5 desaturase; 4) ARA is

converted to docosatetraenoic acid ["DTA"] by a $C_{20/22}$ elongase; and, 5) DTA is converted to docosapentaenoic acid ["DPAn-6"] by a Δ4 desaturase. Alternatively, the "Δ9 elongase/ Δ8 desaturase pathway" can use α-linolenic acid ["ALA"] as substrate to produce long-chain ω-3 fatty acids as follows: 1) LA is converted to ALA, the first of the ω-3 fatty acids, by a Δ15 desaturase; 2) ALA is converted to eicosatrienoic acid ["ETrA"] by a Δ9 elongase; 3) ETrA is converted to eicosatetraenoic acid ["ETA"] by a Δ8 desaturase; 4) ETA is converted to eicosapentaenoic acid ["EPA"] by a Δ5 desaturase; 5) EPA is converted to docosapentaenoic acid ["DPA"] by a $C_{20/22}$ elongase; and, 6) DPA is converted to docosahexaenoic acid ["DHA"] by a Δ4 desaturase. Optionally, ω-6 fatty acids may be converted to ω-3 fatty acids. For example, ETA and EPA are produced from DGLA and ARA, respectively, by Δ17 desaturase activity.

[0020] Alternate pathways for the biosynthesis of ω-3/ω-6 fatty acids utilize a Δ6 desaturase and $C_{18/20}$ elongase, that is, the "Δ6 desaturase/ Δ6 elongase pathway". More specifically, LA and ALA may be converted to γ-linolenic acid ["GLA"] and stearidonic acid ["STA"], respectively, by a Δ6 desaturase; then, a $C_{18/20}$ elongase converts GLA to DGLA and/or STA to ETA. Downstream PUFAs are subsequently formed as described above.

[0021] It is contemplated that the particular functionalities required to be expressed in a specific host organism for production of ω-3/ω-6 fatty acids will depend on the host cell (and its native PUFA profile and/or desaturase/elongase profile), the availability of substrate, and the desired end product(s). One skilled in the art will be able to identify various candidate genes encoding each of the enzymes desired for ω-3/ω-6 fatty acid biosynthesis. Useful desaturase and elongase sequences may be derived from any source, e.g., isolated from a natural source such as from bacteria, algae, fungi, oomycete, yeast, stramenopiles, plants, animals, etc., produced via a semi-synthetic route or synthesized *de novo*. Although the particular source of the desaturase and elongase genes

introduced into the host is not critical, considerations for choosing a specific polypeptide having desaturase or elongase activity include: 1) the substrate specificity and activity of the polypeptide; 2) whether the polypeptide or a component thereof is a rate-limiting enzyme; 3) whether the desaturase or elongase is essential for synthesis of a desired PUFA; 4) co-factors required by the polypeptide; and/or, 5) whether the polypeptide was modified after its production, such as by a kinase or a prenyltransferase. The expressed polypeptide preferably has parameters compatible with the biochemical environment of its location in the host cell. See U.S. Pat. No. 7,238,482.

[0022] It may also be useful to consider the conversion efficiency of each particular desaturase and/or elongase. More specifically, since each enzyme rarely functions with 100% efficiency to convert substrate to product, the final lipid profile of un-purified oils produced in a host cell is typically a mixture of various PUFAs consisting of the desired ω-3/ω-6 fatty acid, as well as various upstream intermediary PUFAs. Thus, the conversion efficiency of each enzyme is also a variable to consider when optimizing biosynthesis of a desired fatty acid.

[0023] With each of the considerations above in mind, candidate genes having the appropriate desaturase and elongase activities (e.g., Δ6 desaturases, $C_{18/20}$ elongases, Δ5 desaturases, Δ17 desaturases, Δ15 desaturases, Δ9 desaturases, Δ12 desaturases, $C_{14/16}$ elongases, $C_{16/18}$ elongases, Δ9 elongases, Δ8 desaturases, Δ4 desaturases and $C_{20/22}$ elongases) can be identified according to publicly available literature such as e.g. GenBank, the patent literature, and experimental analysis of organisms having the ability to produce PUFAs. These genes will be suitable for introduction into a specific host organism, to enable or enhance the organism's synthesis of PUFAs.

Sequence Identification Of A Novel *Eutreptiella cf_gymnastica* CCMP1594 Δ4 Desaturase

[0024] The present disclosure relates to a nucleotide sequence (SEQ ID NO:1) isolated from *Eutreptiella cf_gymnastica* CCMP1594, encoding a Δ4 desaturase (SEQ ID NO:2). This sequence is designated herein as "E1594D4".

[0025] Comparison of the E1594D4 nucleotide base and deduced amino acid sequences to public databases reveals that the most similar known sequences are about 68% identical to the amino acid sequence of E1594D4 reported herein over a length of 514 amino acids using the Clustal W method of alignment (described by Higgins and Sharp, *CABIOS*, 5:151-153 (1989); Higgins, D.G. et al., *Comput. Appl. Biosci.,* 8:189-191(1992); found in the MegAlign™ v6.1 program of the LASERGENE bioinformatics computing suite (DNASTAR Inc.)). More preferred amino acid fragments are at least about 70%-80% identical to the sequences herein, where those sequences that are at least about 80%-90% identical are particularly suitable and those sequences that are at least about 90%-95% identical are most preferred. Similarly, preferred E1594D4 encoding nucleic acid sequences corresponding to the ORF are those encoding active proteins and which are at least about 70%-80% identical to the nucleic acid sequences of E1594D4 reported herein, where those sequences that are at least about 80%-90% identical are particularly suitable and those sequences that are at least about 90%-95% identical are most preferred.

[0026] In alternate embodiments, the E1594D4 desaturase sequence can be codon-optimized for expression in a particular host organism. As is well known in the art, this can be a useful means to further optimize the expression of the enzyme in the alternate host, since use of host-preferred codons can substantially enhance the expression of the foreign gene encoding the polypeptide. In general, host-preferred codons can be determined within a particular host species of interest by examining codon

usage in proteins, preferably those expressed in the largest amount, and determining which codons are used with highest frequency. Then, the coding sequence for a polypeptide of interest having e.g., desaturase activity can be synthesized in whole or in part using the codons preferred in the host species.

[0027] Thus, E1594D4 was codon-optimized for expression in *Yarrowia lipolytica*. This was possible based on previous determination of the *Y. lipolytica* codon usage profile, identification of those codons that were preferred, and determination of the consensus sequence around the 'ATG' initiation codon (see U.S. Pat. No. 7,238,482 and U.S. Pat. No. 7,125,672). The codon-optimized synthetic gene, designated herein as "E1594D4S", had one additional alanine amino acid inserted between amino acid residues 1 and 2 of the wildtype E1594D4; thus, the total length of E1594D4S is 1548 nucleotides (SEQ ID NO:3), while the encoded protein set forth as SEQ ID NO:4 is 515 amino acids in length.

[0028] One skilled in the art would be able to use the teachings herein to create various other codon-optimized Δ4 desaturase proteins suitable for optimal expression in alternate hosts (i.e., other than *Yarrowia lipolytica*), based on the wildtype E1594D4 sequence. Accordingly, the disclosure herein relates to any codon-optimized Δ4 desaturase protein that is derived from the wildtype E1594D4, that is, encoded by SEQ ID NO:2. This includes, but is not limited to, the nucleotide sequence set forth in SEQ ID NO:3, which encodes a synthetic Δ4 desaturase protein (i.e., E1594D4S as set forth in SEQ ID NO:4) that was codon-optimized for expression in *Yarrowia lipolytica.*

Identification And Isolation Of Homologs

[0029] Any of the instant desaturase sequences (i.e., E1594D4, E1594D4S) or portions thereof may be used to search for Δ4 desaturase homologs in the same or other bacterial, algal, fungal, oomycete, yeast, stramenopiles, euglenoid, plant or animal species using sequence

analysis software. In general, such computer software matches similar sequences by assigning degrees of homology to various substitutions, deletions, and other modifications.

[0030] Alternatively, any of the instant desaturase sequences or portions thereof may also be employed as hybridization reagents for the identification of Δ4 homologs. The basic components of a nucleic acid hybridization test include a probe, a sample suspected of containing the gene or gene fragment of interest and a specific hybridization method. Probes of the present invention are typically single-stranded nucleic acid sequences that are complementary to the nucleic acid sequences to be detected. Probes are "hybridizable" to the nucleic acid sequence to be detected. Although the probe length can vary from 5 bases to tens of thousands of bases, typically a probe length of about 15 bases to about 30 bases is suitable. Only part of the probe molecule need be complementary to the nucleic acid sequence to be detected. In addition, the complementarity between the probe and the target sequence need not be perfect. Hybridization does occur between imperfectly complementary molecules with the result that a certain fraction of the bases in the hybridized region are not paired with the proper complementary base.

[0031] Hybridization methods are well defined. Typically the probe and sample must be mixed under conditions that will permit nucleic acid hybridization. This involves contacting the probe and sample in the presence of an inorganic or organic salt under the proper concentration and temperature conditions. The probe and sample nucleic acids must be in contact for a long enough time that any possible hybridization between the probe and sample nucleic acid may occur. The concentration of probe or target in the mixture will determine the time necessary for hybridization to occur. The higher the probe or target concentration, the shorter the hybridization incubation time needed. Optionally, a chaotropic agent may be added (e.g., guanidinium chloride, guanidinium thiocyanate, sodium thiocyanate, lithium tetrachloroacetate, sodium perchlorate, rubidium

tetrachloroacetate, potassium iodide, cesium trifluoroacetate). If desired, one can add formamide to the hybridization mixture, typically 30-50% (v/v).

[0032] Various hybridization solutions can be employed. Typically, these comprise from about 20 to 60% volume, preferably 30%, of a polar organic solvent. A common hybridization solution employs about 30-50% v/v formamide, about 0.15 to 1 M sodium chloride, about 0.05 to 0.1 M buffers (e.g., sodium citrate, Tris-HCl, PIPES or HEPES (pH range about 6-9)), about 0.05 to 0.2% detergent (e.g., sodium dodecylsulfate), or between 0.5-20 mM EDTA, FICOLL (Pharmacia Inc.) (about 300-500 kdal), polyvinylpyrrolidone (about 250-500 kdal), and serum albumin. Also included in the typical hybridization solution will be unlabeled carrier nucleic acids from about 0.1 to 5 mg/mL, fragmented nucleic DNA (e.g., calf thymus or salmon sperm DNA, or yeast RNA), and optionally from about 0.5 to 2% wt/vol glycine. Other additives may also be included, such as volume exclusion agents that include a variety of polar water-soluble or swellable agents (e.g., polyethylene glycol), anionic polymers (e.g., polyacrylate or polymethylacrylate) and anionic saccharidic polymers (e.g., dextran sulfate).

[0033] Nucleic acid hybridization is adaptable to a variety of assay formats. One of the most suitable is the sandwich assay format. The sandwich assay is particularly adaptable to hybridization under non-denaturing conditions. A primary component of a sandwich-type assay is a solid support. The solid support has adsorbed to it or covalently coupled to it immobilized nucleic acid probe that is unlabeled and complementary to one portion of the sequence.

[0034] In additional embodiments, any of the Δ4 desaturase nucleic acid fragments described herein (or any homologs identified thereof) may be used to isolate genes encoding homologous proteins from the same or other bacterial, algal, fungal, oomycete, yeast, stramenopiles, euglenoid,

plant or animal species. Isolation of homologous genes using sequence-dependent protocols is well known in the art. Examples of sequence-dependent protocols include, but are not limited to: 1) methods of nucleic acid hybridization; 2) methods of DNA and RNA amplification, as exemplified by various uses of nucleic acid amplification technologies such as polymerase chain reaction ["PCR"] (U.S. Pat. No.4,683,202); ligase chain reaction ["LCR"] (Tabor, S. et al., *Proc. Natl. Acad. Sci. U.S.A.,* 82:1074 (1985)); or strand displacement amplification ["SDA"] (Walker, et al., *Proc. Natl. Acad. Sci. U.S.A.*, 89:392 (1992)); and, 3) methods of library construction and screening by complementation.

[0035] For example, genes encoding similar proteins or polypeptides to the Δ4 desaturases described herein could be isolated directly by using all or a portion of the nucleic acid fragments as DNA hybridization probes to screen libraries from any desired organism using well known methods, wherein those organisms producing DPAn-6 or DHA would be preferred. Specific oligonucleotide probes based upon the nucleic acid sequences can be designed and synthesized by methods known in the art (Maniatis, *supra*). Moreover, the entire sequences can be used directly to synthesize DNA probes by methods known to the skilled artisan, such as random primers DNA labeling, nick translation or end-labeling techniques, or RNA probes using available *in vitro* transcription systems. In addition, specific primers can be designed and used to amplify a part of or the full length of the Δ4 desaturase sequences. The resulting amplification products can be labeled directly during amplification reactions or labeled after amplification reactions, and used as probes to isolate full-length DNA fragments under conditions of appropriate stringency.

[0036] Typically, in PCR-type amplification techniques, the primers have different sequences and are not complementary to each other. Depending on the desired test conditions, the sequences of the primers should be designed to provide for both efficient and faithful replication of the target nucleic acid. Methods of PCR primer design are common and well known

(Thein and Wallace, "The use of oligonucleotides as specific hybridization probes in the Diagnosis of Genetic Disorders", in *Human Genetic Diseases: A Practical Approach*, K. E. Davis Ed., (1986) pp 33-50, IRL: Herndon, VA; and Rychlik, W., In Methods in Molecular Biology, White, B. A. Ed., (1993) Vol. 15, pp 31-39, PCR Protocols: Current Methods and Applications. Humania: Totowa, NJ).

[0037] Generally two short segments of the Δ4 desaturase sequences may be used in PCR protocols to amplify longer nucleic acid fragments encoding homologous genes from DNA or RNA. PCR may also be performed on a library of cloned nucleic acid fragments wherein the sequence of one primer is derived from the disclosed nucleic acid fragments. The sequence of the other primer takes advantage of the presence of the polyadenylic acid tracts to the 3' end of the mRNA precursor encoding eukaryotic genes.

[0038] Alternatively, the second primer sequence may be based upon sequences derived from the cloning vector. For example, the skilled artisan can follow the RACE protocol (Frohman et al., *Proc. Natl Acad. Sci. U.S.A.,* 85:8998 (1988)) to generate cDNAs by using PCR to amplify copies of the region between a single point in the transcript and the 3' or 5' end. Primers oriented in the 3' and 5' directions can be designed from the disclosed sequences. Using commercially available 3' RACE or 5' RACE systems (e.g., Gibco/BRL, Gaithersburg, MD), specific 3' or 5' cDNA fragments can be isolated (Ohara et al., *Proc. Natl Acad. Sci. U.S.A.,* 86:5673 (1989); Loh et al., *Science,* 243:217 (1989)).

[0039] Alternately, any of the Δ4 desaturase nucleic acid fragments described herein (or any homologs identified thereof) may be used for creation of new and improved fatty acid desaturases. As is well known in the art, *in vitro* mutagenesis and selection, chemical mutagenesis, "gene shuffling" methods or other means can be employed to obtain mutations of naturally occurring desaturase genes. Furthermore, improved fatty acids

may be synthesized by domain swapping, wherein a functional domain from any of the Δ4 desaturase nucleic acid fragments described herein is exchanged with a functional domain in an alternate desaturase gene to thereby result in a novel protein.

Methods For Production Of Various ω-3 And/Or ω-6 Fatty Acids

[0040] It is expected that introduction of chimeric genes encoding the Δ4 desaturases described herein (i.e., E1594D4, E1594D4S or other mutant enzymes, codon-optimized enzymes or homologs thereof), under the control of the appropriate promoters, will result in increased production of DPAn-6 and/or DHA in the transformed host organism, respectively. As such, described herein is a method for the direct production of PUFAs comprising exposing a fatty acid substrate (i.e., DTA or DPA) to the desaturase enzymes described herein (e.g., E1594D4, E1594D4S), such that the substrate is converted to the desired fatty acid product (i.e., DPAn-6 or DHA, respectively).

[0041] More specifically, a method for the production of DHA in a host cell is provided herein, wherein the host cell comprises:

(i) an isolated nucleotide molecule encoding a Δ4 desaturase polypeptide having at least 68% identity when compared to a polypeptide having an amino acid sequence as set forth in SEQ ID NO:2, based on the Clustal W method of alignment; and,

(ii) a source of DPA;

wherein the host cell is grown under conditions such that the Δ4 desaturase is expressed and the DPA is converted to DHA, and wherein the DHA is optionally recovered.

[0042] The person of skill in the art will recognize that the broad substrate range of the Δ4 desaturase may additionally allow for the use of the enzyme for the conversion DTA to DPAn-6. Accordingly, described herein

is also a method for the production of DPAn-6, wherein the host cell comprises:

(i) an isolated nucleotide molecule encoding a Δ4 desaturase polypeptide having at least 68% identity when compared to a polypeptide having an amino acid sequence as set forth in SEQ ID NO:2, based on the Clustal W method of alignment; and,

(ii) a source of DTA;

wherein the host cell is grown under conditions such that the Δ4 desaturase is expressed and the DTA is converted to DPAn-6, and wherein the DPAn-6 is optionally recovered.

[0043] The source of the DTA or DPA used as substrate in either of the methods above may be produced by the host either naturally or transgenically, or the substrate may be provided exogenously. In particular, it is contemplated that the Δ4 desaturases described herein (e.g., E1594D4, E1594D4S or other mutant enzymes, codon-optimized enzymes or homologs thereof) may be expressed in conjunction with additional genes encoding enzymes of the PUFA biosynthetic pathway, such as Δ6 desaturases, $C_{18/20}$ elongases, Δ17 desaturases, Δ15 desaturases, Δ9 desaturases, Δ12 desaturases, $C_{14/16}$ elongases, $C_{16/18}$ elongases, Δ9 elongases, Δ8 desaturases, Δ5 desaturases and/or $C_{20/22}$ elongases, to result in production of DPAn-6 and/or DHA. The particular genes included within a particular expression cassette will depend on the host cell (and its PUFA profile and/or desaturase/elongase profile), the availability of substrate and the desired end product(s).

[0044] In alternative embodiments, it may be useful to disrupt a host organism's native Δ4 desaturase, based on the complete sequences described herein, the complement of those complete sequences, substantial portions of those sequences, codon-optimized desaturases derived therefrom and those sequences that are substantially homologous thereto.

Expression Systems, Cassettes And Vectors

[0045] The genes and gene products described herein may be expressed in heterologous host cells. Expression in recombinant hosts may be useful for the production of various PUFA pathway intermediates, or for the modulation of PUFA pathways already existing in the host for the synthesis of new products heretofore not possible using the host.

[0046] Expression systems and expression vectors containing regulatory sequences that direct high level expression of foreign proteins are well known. Any of these could be used to construct chimeric genes for production of any of the gene products of the instant sequences. These chimeric genes could then be introduced into appropriate host cells via transformation to provide high-level expression of the encoded enzymes.

[0047] Vectors (e.g., constructs, plasmids) and DNA expression cassettes useful for the transformation of suitable host cells are well known. The specific choice of sequences present in the construct is dependent upon the desired expression products (*supra*), the nature of the host cell, and the proposed means of separating transformed cells versus non-transformed cells. Typically, however, the vector contains at least one expression cassette, a selectable marker and sequences allowing autonomous replication or chromosomal integration. Suitable expression cassettes comprise a region 5' of the gene that controls transcriptional initiation, i.e., a promoter, the gene coding sequence, and a region 3' of the DNA fragment that controls transcriptional termination, i.e., a terminator. It is most preferred when both control regions are derived from genes from the transformed host cell, although it is to be understood that such control regions need not be derived from the genes native to the specific species chosen as a production host.

[0048] Transcriptional control regions or promoters useful for driving expression of Δ4 desaturase ORFs in the desired host cell are numerous and well known. These control regions may comprise a promoter,

enhancer, silencer, intron sequences, 3' UTR and/or 5' UTR regions, and protein and/or RNA stabilizing elements. Such elements may vary in their strength and specificity. Virtually any promoter (i.e., native, synthetic, or chimeric) capable of directing expression of these genes in the selected host cell is suitable, although transcriptional and translational regions from the host species are particularly useful. Expression in a host cell can occur in an induced or constitutive fashion. Induced expression occurs by inducing the activity of a regulatable promoter operably linked to the gene of interest, while constitutive expression occurs by the use of a constitutive promoter operably linked to the gene of interest.

[0049] When the host cell is, for example, yeast, transcriptional and translational regions functional in yeast cells are provided, particularly from the host species. See Int'l. App. Pub. No. WO 2006/052870 for preferred transcriptional initiation regulatory regions for use in *Yarrowia lipolytica*. Any of a number of regulatory sequences may be used, depending upon whether constitutive or induced transcription is desired, the efficiency of the promoter in expressing the ORF of interest, the ease of construction, etc.

[0050] 3' non-coding sequences encoding transcription termination signals, i.e., a "termination region", must be provided in a recombinant construct and may be derived from the 3' region of the gene from which the initiation region was obtained or from a different gene. A large number of termination regions are known and function satisfactorily in a variety of hosts when utilized in both the same and different genera and species from which they were derived. The termination region usually is selected more as a matter of convenience rather than because of any particular property. Termination control regions may also be derived from various genes native to the preferred hosts. The 3'-region can also be synthetic, as one of skill in the art can utilize available information to design and synthesize a 3'-region sequence that functions as a transcription

terminator. A termination region may be unnecessary, but is highly preferred.

[0051] Merely inserting a gene, such as a desaturase, into a cloning vector does not ensure its expression at the desired rate, concentration, amount, etc. In response to the need for a high expression rate, many specialized expression vectors have been created by manipulating a number of different genetic elements that control transcription, RNA stability, translation, protein stability and location, oxygen limitation, and secretion from the host cell. Some of the manipulated features include: the nature of the relevant transcriptional promoter and terminator sequences, the number of copies of the cloned gene (wherein additional copies may be cloned within a single expression construct and/or additional copies may be introduced into the host cell by increasing the plasmid copy number or by multiple integration of the cloned gene into the genome), whether the gene is plasmid-borne or integrated into the genome of the host cell, the final cellular location of the synthesized foreign protein, the efficiency of translation and correct folding of the protein in the host organism, the intrinsic stability of the mRNA and protein of the cloned gene within the host cell, and the codon usage within the cloned gene such that its frequency approaches the frequency of preferred codon usage of the host cell. Each of these may be used in the methods and host cells described herein to further optimize expression of the Δ4 desaturases.

Transformation Of Host Cells

[0052] After a recombinant construct is created, e.g., comprising a chimeric gene comprising a promoter, ORF and terminator, it is placed in a plasmid vector capable of autonomous replication in the host cell or it is directly integrated into the genome of the host cell. Integration of expression cassettes can occur randomly within the host genome or can be targeted through the use of constructs containing regions of homology with the host genome sufficient to target recombination with the host locus. Where constructs are targeted to an endogenous locus, all or some

of the transcriptional and translational regulatory regions can be provided by the endogenous locus.

[0053] When two or more genes are expressed from separate replicating vectors, each vector may have a different means of selection and should lack homology to the other construct(s) to maintain stable expression and prevent reassortment of elements among constructs. Judicious choice of regulatory regions, selection means and method of propagation of the introduced construct(s) can be experimentally determined so that all introduced genes are expressed at the necessary levels to provide for synthesis of the desired products.

[0054] Constructs comprising the gene(s) of interest may be introduced into a host cell by any standard technique. These techniques include transformation, e.g., lithium acetate transformation (*Methods in Enzymology*, 194:186-187 (1991)), protoplast fusion, biolistic impact, electroporation, microinjection, vacuum filtration or any other method that introduces the gene(s) of interest into the host cell.

[0055] For convenience, a host cell that has been manipulated by any method to take up a DNA sequence, for example, in an expression cassette, is referred to herein as "transformed" or "recombinant". The transformed host will have at least one copy of the expression cassette and may have two or more, depending upon whether the expression cassette is integrated into the genome amplified, or is present on an extrachromosomal element having multiple copy numbers. The transformed host cell can be identified by various selection techniques, as described in U.S. Pat. No. 7,238,482 and U.S. Pat. No. 7,259,255.

[0056] Following transformation, substrates suitable for Δ4 desaturases (and, optionally other PUFA enzymes that are co-expressed within the host cell) may be produced by the host either naturally or transgenically, or they may be provided exogenously.

Metabolic Engineering Of ω-3 And/Or ω-6 Fatty Acid Biosynthesis

[0057] Knowledge of the sequences of the present Δ4 desaturases will be useful for manipulating ω-3 and/or ω-6 fatty acid biosynthesis in various host cells. This may require metabolic engineering directly within the PUFA biosynthetic pathway or additional manipulation of pathways that contribute carbon to the PUFA biosynthetic pathway.

[0058] Techniques useful for up-regulating desirable biochemical pathways and down-regulating undesirable biochemical pathways are well known in the art. For example, biochemical pathways competing with the ω-3 and/or ω-6 fatty acid biosynthetic pathways for energy or carbon, or native PUFA biosynthetic pathway enzymes that interfere with production of a particular PUFA end-product, may be eliminated by gene disruption or down-regulated by other means, such as antisense mRNA and zinc-finger targeting technologies.

[0059] The following discuss altering the PUFA biosynthetic pathway as a means to increase GLA, ARA, EPA or DHA, respectively, and desirable manipulations in the TAG biosynthetic pathway and in the TAG degradation pathway: Int'l. App. Pub. No. WO 2006/033723, Int'l. App. Pub. No. WO 2006/055322 [U.S. Pat. Appl. Pub. No. 2006-0094092-A1], Int'l. App. Pub. No. WO 2006/052870 [U.S. Pat. Appl. Pub. No. 2006-0115881-A1] and Int'l. App. Pub. No. WO 2006/052871 [U.S. Pat. Appl. Pub. No. 2006-0110806-A1], respectively.

Preferred Hosts For Recombinant Expression Of Δ4 Desaturases

[0060] A variety of eukaryotic organisms are suitable as host, to thereby yield a transformant host organism comprising a Δ4 desaturase as described herein. These may include hosts that grow on a variety of feedstocks, including simple or complex carbohydrates, fatty acids, organic acids, oils, glycerols and alcohols, and/or hydrocarbons over a wide range of temperature and pH values. Based on the needs of the

Applicants' Assignee, the genes described herein were initially isolated for expression in an oleaginous yeast (and in particular *Yarrowia lipolytica*); however, it is contemplated that because transcription, translation and the protein biosynthetic apparatus are highly conserved, any bacteria, yeast, algae, stramenopile, oomycete, euglenoid and/or fungus will be a suitable host for expression of the present nucleic acid fragments.

[0061] Preferred hosts are oleaginous organisms, such as oleaginous yeast. These oleaginous organisms are naturally capable of oil synthesis and accumulation, wherein the total oil content can comprise greater than about 25% of the cellular dry weight, more preferably greater than about 30% of the cellular dry weight, and most preferably greater than about 40% of the cellular dry weight. Various algae, moss, fungi, yeast and stramenopiles are naturally classified as oleaginous. Genera typically identified as oleaginous yeast include, but are not limited to: *Yarrowia, Candida, Rhodotorula, Rhodosporidium*, *Cryptococcus, Trichosporon* and *Lipomyces*. More specifically, illustrative oil-synthesizing yeast include: *Rhodosporidium toruloides, Lipomyces starkeyii, L. lipoferus*, *Candida revkaufi, C. pulcherrima, C. tropicalis, C. utilis, Trichosporon pullans, T. cutaneum, Rhodotorula glutinus, R. graminis* and *Yarrowia lipolytica* (formerly classified as *Candida lipolytica*).

[0062] Most preferred is the oleaginous yeast *Yarrowia lipolytica;* and, in a further embodiment, most preferred are the *Y. lipolytica* strains designated as ATCC #76982, ATCC #20362, ATCC #8862, ATCC #18944 and/or LGAM S(7)1 (Papanikolaou S., and Aggelis G., *Bioresour. Technol.*, 82(1):43-9 (2002)).

[0063] Specific teachings relating to transformation of *Yarrowia lipolytica* include U.S. Pat. No. 4,880,741, U.S. Pat. No. 5,071,764 and Chen, D. C. et al. (*Appl. Microbiol. Biotechnol.*, 48(2):232-235 (1997)). Specific teachings applicable for engineering ARA, EPA and DHA in *Y. lipolytica* are provided in Int'l. App. Pub. No.WO 2006/055322, Int'l. App. Pub. No.

WO 2006/052870 and Int'l. App. Pub. No. WO 2006/052871, respectively. Detailed means for the synthesis and transformation of expression vectors comprising Δ4 desaturases in oleaginous yeast (i.e., *Yarrowia lipolytica*) are provided in Int'l. App. Pub. No. WO 2006/052871.

[0064] The preferred method of expressing genes in *Yarrowia lipolytica* is by integration of linear DNA into the genome of the host. Integration into multiple locations within the genome can be particularly useful when high level expression of genes are desired, such as in the *Ura3* locus (GenBank Accession No. AJ306421), the *Leu2* gene locus (GenBank Accession No. AF260230), the *Lys5* gene locus (GenBank Accession No. M34929), the *Aco2* gene locus (GenBank Accession No. AJ001300), the *Pox3* gene locus (Pox3: GenBank Accession No. XP_503244; or, Aco3: GenBank Accession No. AJ001301), the Δ12 desaturase gene locus (Int'l. App. Pub. No. WO 2004/104167), the *Lip1* gene locus (GenBank Accession No. Z50020), the *Lip2* gene locus (GenBank Accession No. AJ012632), the *SCP2* gene locus (GenBank Accession No. AJ431362), the *Pex3* gene locus (GenBank Accession No. CAG78565), the *Pex16* gene locus (GenBank Accession No. CAG79622) and/or the *Pex10* gene locus (GenBank Accession No. CAG81606).

[0065] Preferred selection methods for use in *Yarrowia lipolytica* are resistance to kanamycin, hygromycin and the amino glycoside G418, as well as ability to grow on media lacking uracil, leucine, lysine, tryptophan or histidine. 5-fluoroorotic acid [5-fluorouracil-6-carboxylic acid monohydrate; "5-FOA"] may also be used for selection of yeast *Ura*$^-$ mutants. This compound is toxic to yeast cells that possess a functioning URA3 gene encoding orotidine 5'-monophosphate decarboxylase [OMP decarboxylase]; thus, based on this toxicity, 5-FOA is especially useful for the selection and identification of *Ura*$^-$ mutant yeast strains (Bartel, P.L. and Fields, S., Yeast 2-Hybrid System, Oxford University: New York, v. 7, pp 109-147, 1997; see also Int'l. App. Pub. No. WO 2006/052870 for 5-FOA use in *Yarrowia*).

[0066] Other microbial hosts include oleaginous bacteria, algae, euglenoids, stramenopiles and other fungi; and, within this broad group of microbial hosts, of particular interest are microorganisms that naturally produce ω-3/ω-6 fatty acids. For example, ARA, EPA and/or DHA is produced via *Cyclotella* sp., *Nitzschia* sp., *Pythium, Thraustochytrium* sp., *Schizochytrium* sp. and *Mortierella.* Thus, for example, transformation of *Mortierella alpine,* which is commercially used for production of ARA, with any of the present Δ4 desaturase genes under the control of inducible or regulated promoters (in addition to a Δ17 desaturase and a $C_{20/22}$ elongase) could yield a transformant organism capable of synthesizing DHA. The method of transformation of *M. alpina* is described by Mackenzie et al. (*Appl. Environ. Microbiol.,* 66:4655 (2000)). Similarly, methods for transformation of Thraustochytriales microorganisms (e.g., *Thraustochytrium*, *Schizochytrium*) are disclosed in U.S. Pat. No. 7,001,772.

[0067] In alternate embodiments, hosts may be plants or other animals. For example, using oilseed plants that can be readily engineered for PUFA production include: soybean (*Glycine* and *Soja sp.*), corn (*Zea mays*), flax (*Linum sp.*), rapeseed (*Brassica sp.*), primrose, canola, maize, cotton, safflower (*Carthamus sp.*) and sunflower (*Helianthus sp.*). See, for example, U.S. Pat. Appl. Pub. No. 2007-0237876 A1.

[0068] Regardless of the selected host or expression construct, multiple transformants must be screened to obtain a strain displaying the desired expression level, regulation and pattern. Such screening may be accomplished by Southern analysis of DNA blots (Southern, *J. Mol. Biol.,* 98:503 (1975)), Northern analysis of mRNA expression (Kroczek, *J. Chromatogr. Biomed. Appl.*, 618(1-2):133-145 (1993)), Western and/or Elisa analyses of protein expression, phenotypic analysis or GC analysis of the PUFA products.

Fermentation Processes For Omega Fatty Acid Production

[0069] The transformed host cell is grown under conditions that optimize expression of chimeric desaturase genes and produce the greatest and most economical yield of desired PUFAs. In general, media conditions may be optimized by modifying the type and amount of carbon source, the type and amount of nitrogen source, the carbon-to-nitrogen ratio, the amount of different mineral ions, the oxygen level, growth temperature, pH, length of the biomass production phase, length of the oil accumulation phase and the time and method of cell harvest. Oleaginous yeast of interest, such as *Yarrowia lipolytica,* are generally grown in complex media such as yeast extract-peptone-dextrose broth (YPD) or a defined minimal media that lacks a component necessary for growth and thereby forces selection of the desired expression cassettes (e.g., Yeast Nitrogen Base (DIFCO Laboratories, Detroit, MI)).

[0070] Fermentation media for the methods and host cells described herein must contain a suitable carbon source such as are taught in U.S. Pat. No. 7,238,482. Suitable sources of carbon encompass a wide variety of sources, with sugars, glycerol and/or fatty acids being preferred. Most preferred is glucose and/or fatty acids containing between 10-22 carbons.

[0071] Nitrogen may be supplied from an inorganic (e.g., $(NH_4)_2SO_4$) or organic (e.g., urea or glutamate) source. In addition to appropriate carbon and nitrogen sources, the fermentation media must also contain suitable minerals, salts, cofactors, buffers, vitamins and other components known to those skilled in the art suitable for the growth of the oleaginous host and promotion of the enzymatic pathways of PUFA production. Particular attention is given to several metal ions, such as Fe^{+2}, Cu^{+2}, Mn^{+2}, Co^{+2}, Zn^{+2}, Mg^{+2}, that promote synthesis of lipids and PUFAs (Nakahara, T. et al., *Ind. Appl. Single Cell Oils,* D. J. Kyle and R. Colin, eds. pp 61-97 (1992)).

[0072] Preferred growth media for the methods and host cells described herein are common commercially prepared media, such as Yeast Nitrogen Base (DIFCO Laboratories, Detroit, MI). Other defined or synthetic growth media may also be used and the appropriate medium for growth of the transformant host cells is well known in microbiology or fermentation science. A suitable pH range for the fermentation is typically between about pH 4.0 to pH 8.0, wherein pH 5.5 to pH 7.5 is preferred as the range for the initial growth conditions. The fermentation may be conducted under aerobic or anaerobic conditions, wherein microaerobic conditions are preferred.

[0073] Typically, accumulation of increased amounts of PUFAs and TAGs in oleaginous yeast cells requires a two-stage process, since the metabolic state must be "balanced" between growth and synthesis/storage of fats. Thus, most preferably, a two-stage fermentation process is necessary for the production of oils comprising PUFAs in oleaginous yeast. This approach is described in U.S. Pat. No. 7,238,482, as are various suitable fermentation process designs (i.e., batch, fed-batch and continuous) and considerations during growth.

Purification And Processing Of PUFAs

[0074] Fatty acids, including PUFAs, may be found in the host microorganisms as free fatty acids or in esterified forms such as acylglycerols, phospholipids, sulfolipids or glycolipids. These fatty acids may be extracted from the host cells through a variety of means well-known in the art. One review of extraction techniques, quality analysis and acceptability standards for yeast lipids is that of Z. Jacobs (*Critical Reviews in Biotechnology,* 12(5/6):463-491 (1992)). A brief review of downstream processing is also available by A. Singh and O. Ward (*Adv. Appl. Microbiol.*, 45:271-312 (1997)).

[0075] In general, means for the purification of fatty acids (including PUFAs) may include extraction (e.g., U.S. Pat. No. 6,797,303 and U.S.

Pat. No. 5,648,564) with organic solvents, sonication, supercritical fluid extraction (e.g., using carbon dioxide), saponification and physical means such as presses, or combinations thereof. See U.S. Pat. No. 7,238,482.

PUFAs For Use In Foodstuffs, Health Food Products, Pharmaceuticals And Animal Feeds

[0076] The market place contains many food and feed products, incorporating ω-3 and/or ω-6 fatty acids, particularly ALA, GLA, ARA, EPA, DPA and DHA. It is contemplated that oleaginous yeast biomass comprising long-chain PUFAs, partially purified biomass comprising PUFAs, purified oil comprising PUFAs, and/or purified PUFAs made by the methods and host cells described herein impart the health benefits, upon ingestion of foods or feed improved by their addition. These oils can be added to food analogs, drinks, meat products, cereal products, baked foods, snack foods and dairy products, to name a few. See U.S. Pat. Appl. Pub. No. 2006-0094092.

[0077] These compositions may impart health benefits by being added to medical foods including medical nutritionals, dietary supplements, infant formula and pharmaceuticals. The skilled artisan will appreciate the amount of the oils to be added to food, feed, dietary supplements, nutriceuticals, pharmaceuticals, and other ingestible products as to impart health benefits. Health benefits from ingestion of these oils are described in the art, known to the skilled artisan and continuously investigated. Such an amount is referred to herein as an "effective" amount and depends on, among other things, the nature of the ingested products containing these oils and the physical conditions they are intended to address.

EXAMPLES

[0078] The present invention is further described in the following Examples, which illustrate reductions to practice of the invention but do not completely define all of its possible variations.

GENERAL METHODS

[0079] Standard recombinant DNA and molecular cloning techniques used in the Examples are well known in the art and are described by: 1) Sambrook, J., Fritsch, E. F. and Maniatis, T., *Molecular Cloning: A Laboratory Manual*; Cold Spring Harbor Laboratory: Cold Spring Harbor, NY (1989) (Maniatis); 2) T. J. Silhavy, M. L. Bennan, and L. W. Enquist, Experiments with Gene Fusions; Cold Spring Harbor Laboratory: Cold Spring Harbor, NY (1984); and, 3) Ausubel, F. M. et al., Current Protocols in Molecular Biology, published by Greene Publishing Assoc. and Wiley-Interscience, Hoboken, NJ (1987).

[0080] Materials and methods suitable for the maintenance and growth of microbial cultures are well known in the art. Techniques suitable for use in the following examples may be found as set out in Manual of Methods for General Bacteriology (Phillipp Gerhardt, R. G. E. Murray, Ralph N. Costilow, Eugene W. Nester, Willis A. Wood, Noel R. Krieg and G. Briggs Phillips, Eds), American Society for Microbiology: Washington, D.C. (1994)); or by Thomas D. Brock in Biotechnology: A Textbook of Industrial Microbiology, 2nd ed., Sinauer Associates: Sunderland, MA (1989). All reagents, restriction enzymes and materials used for the growth and maintenance of microbial cells were obtained from Aldrich Chemicals (Milwaukee, WI), DIFCO Laboratories (Detroit, MI), GIBCO/BRL (Gaithersburg, MD), or Sigma Chemical Company (St. Louis, MO), unless otherwise specified. *E. coli* strains were typically grown at 37 °C on Luria Bertani (LB) plates.

[0081] General molecular cloning was performed according to standard methods (Sambrook et al., *supra*). DNA sequence was generated on an ABI Automatic sequencer using dye terminator technology (U.S. Pat. No. 5,366,860; EP 272,007) using a combination of vector and insert-specific primers. Sequence editing was performed in Sequencher (Gene Codes Corporation, Ann Arbor, MI). All sequences represent coverage at least two times in both directions. Comparisons of genetic sequences were

accomplished using DNASTAR software (DNASTAR Inc., Madison, WI) unless otherwise specified.

[0082] The meaning of abbreviations is as follows: "sec" means second(s), "min" means minute(s), "h" means hour(s), "d" means day(s), "µL" means microliter(s), "mL" means milliliter(s), "L" means liter(s), "µM" means micromolar, "mM" means millimolar, "M" means molar, "mmol" means millimole(s), "µmole" mean micromole(s), "g" means gram(s), "µg" means microgram(s), "ng" means nanogram(s), "U" means unit(s), "bp" means base pair(s) and "kB" means kilobase(s).

Nomenclature For Expression Cassettes

[0083] The structure of an expression cassette is represented by a simple notation system of "X::Y::Z", wherein X describes the promoter fragment, Y describes the gene fragment, and Z describes the terminator fragment, which are all operably linked to one another.

Transformation And Cultivation Of *Yarrowia lipolytica*

[0084] *Yarrowia lipolytica* strain ATCC #20362 was purchased from the American Type Culture Collection (Rockville, MD). *Yarrowia lipolytica* strains were routinely grown at 28-30 °C in several media, according to the recipes shown below. Agar plates were prepared as required by addition of 20 g/L agar to each liquid media, according to standard methodology.

YPD agar medium (per liter): 10 g of yeast extract [Difco]; 20 g of Bacto peptone [Difco]; and 20 g of glucose.

Basic Minimal Media (MM) (per liter): 20 g glucose; 1.7 g yeast nitrogen base without amino acids; 1.0 g proline; and pH 6.1 (not adjusted).

High Glucose Media (HGM) (per liter): 80 glucose; 2.58 g KH_2PO_4; 5.36 g K_2HPO_4; pH 7.5 (do not need to adjust).

[0085] Transformation of *Y. lipolytica* was performed according to the method of Chen, D. C. et al. (*Appl. Microbiol.Biotechnol.*, 48(2):232-235 (1997)), unless otherwise noted. Briefly, *Yarrowia* was streaked onto a YPD plate and grown at 30 °C for approximately 18 hr. Several large loopfuls of cells were scraped from the plate and resuspended in 1 mL of transformation buffer containing: 2.25 mL of 50% PEG, average MW 3350; 0.125 mL of 2 M Li acetate, pH 6.0; and 0.125 mL of 2 M DTT. Then, approximately 500 ng of linearized plasmid DNA was incubated in 100 μl of resuspended cells, and maintained at 39 °C for 1 hr with vortex mixing at 15 min intervals. The cells were plated onto selection media plates and maintained at 30 °C for 2 to 3 days.

Isolation Of *Yarrowia lipolytica* Strain Y4184U

[0086] *Yarrowia lipolytica* strain Y4184, producing EPA relative to the total lipids via expression of a Δ9 elongase/Δ8 desaturase pathway, was generated as described in Example 7 of Int'l. App. Pub. No. WO 2008/073367. Briefly, as diagrammed in FIG. 5, strain Y4184 was derived from *Y. lipolytica* ATCC #20362 via construction of strain Y2224 (a FOA resistant mutant from an autonomous mutation of the *Ura3* gene of wildtype *Yarrowia* strain ATCC #20362), strain Y4001 (producing 17% EDA with a *Leu-* phenotype), strain Y4001U1 (*Leu-* and *Ura-*), strain Y4036 (producing 18% DGLA with a *Leu-* phenotype), strain Y4036U (*Leu-* and *Ura-*), strain Y4069 (producing 12% ARA with a *Ura-* phenotype), strain Y4084 (producing 14% EPA), strain Y4084U1 (*Ura-*), strain Y4127 (producing 18% EPA and deposited with the American Type Culture Collection on November 29, 2007, under accession number ATCC PTA-8802), strain Y4127U2 (*Ura-*), strain Y4158 (producing 25% EPA), strain Y4158U1 (producing *Ura-*) and strain Y4184 (producing 30.7% EPA relative to the total TFAs).

[0087] The final genotype of strain Y4184 with respect to wildtype *Yarrowia lipolytica* ATCC #20362 was *unknown 1-, unknown 2-, unknown 3-, unknown 4-, unknown 5-, unknown 6-,* YAT1::ME3S::Pex16,

EXP1::ME3S::Pex20 (2 copies), GPAT::EgD9e::Lip2, FBAINm::EgD9eS::Lip2, EXP1::EgD9eS::Lip1, FBA::EgD9eS::Pex20, YAT1::EgD9eS::Lip2, GPD::EgD9eS::Lip2, GPDIN::EgD8M::Lip1, YAT1::EgD8M::Aco, EXP1::EgD8M::Pex16, FBAINm::EgD8M::Pex20, FBAIN::EgD8M::Lip1 (2 copies), GPM/FBAIN::FmD12S::Oct, EXP1::FmD12S::Aco, YAT1::FmD12::Oct, GPD::FmD12::Pex20, EXP1::EgD5S::Pex20, YAT1::EgD5S::Aco, YAT1::RD5S::Oct, FBAIN::EgD5::Aco, FBAINm::PaD17::Aco, EXP1::PaD17::Pex16, YAT1::PaD17S::Lip1, YAT1::YlCPT1::Aco, GPD::YlCPT1::Aco (wherein FmD12 is a *Fusarium moniliforme* Δ12 desaturase gene [Int'l. App. Pub. No. WO 2005/047485]; FmD12S is a codon-optimized Δ12 desaturase gene, derived from *Fusarium moniliforme* [Int'l. App. Pub. No. WO 2005/047485]; ME3S is a codon-optimized $C_{16/18}$ elongase gene, derived from *Mortierella alpina* [Int'l. App. Pub. No. WO 2007/046817]; EgD9e is a *Euglena gracilis* Δ9 elongase gene [Int'l. App. Pub. No. WO 2007/061742]; EgD9eS is a codon-optimized Δ9 elongase gene, derived from *Euglena gracilis* [Int'l. App. Pub. No. WO 2007/061742]; EgD8M is a synthetic mutant Δ8 desaturase [Int'l. App. Pub. No. WO 2008/073271], derived from *Euglena gracilis* [U.S. Pat. No. 7,256,033]; EgD5 is a *Euglena gracilis* Δ5 desaturase [U.S. Pat. App. Pub. US 2007-0292924-A1]; EgD5S is a codon-optimized Δ5 desaturase gene, derived from *Euglena gracilis* [U.S. Pat. App. Pub. No. 2007-0292924]; RD5S is a codon-optimized Δ5 desaturase, derived from *Peridinium sp.* CCMP626 [U.S. Pat. App. Pub. No. 2007-0271632]; PaD17 is a *Pythium aphanidermatum* Δ17 desaturase [Int'l. App. Pub. No. WO 2008/054565]; PaD17S is a codon-optimized Δ17 desaturase, derived from *Pythium aphanidermatum* [Int'l. App. Pub. No. WO 2008/054565]; and, YlCPT1 is a *Yarrowia lipolytica* diacylglycerol cholinephosphotransferase gene [Int'l. App. Pub. No. WO 2006/052870]).

[0088] Finally, in order to disrupt the Ura3 gene in strain Y4184, construct pZKUE3S (described in Table 22 of Int'l. App. Pub. No. WO 2008/073367 used to integrate a EXP1::ME3S::Pex20 chimeric gene into the *Ura3* gene of strain Y4184 to result in strains Y4184U1 (11.2% EPA of total lipids),

Y4184U2 (10.6% EPA of total lipids) and Y4184U4 (15.5% EPA of total lipids), respectively (collectively, Y4184U).

Fatty Acid Analysis Of *Yarrowia lipolytica*

[0089] For fatty acid analysis, cells were collected by centrifugation and lipids were extracted as described in Bligh, E. G. & Dyer, W. J. (*Can. J. Biochem. Physiol.*, 37:911-917 (1959)). Fatty acid methyl esters ["FAMEs"] were prepared by transesterification of the lipid extract with sodium methoxide (Roughan, G., and Nishida I., *Arch Biochem Biophys.,* 276(1):38-46 (1990)) and subsequently analyzed with a Hewlett-Packard 6890 GC fitted with a 30-m X 0.25 mm (i.d.) HP-INNOWAX (Hewlett-Packard) column. The oven temperature was from 170 °C (25 min hold) to 185 °C at 3.5 °C/min.

[0090] For direct base transesterification, *Yarrowia* culture (3 mL) was harvested, washed once in distilled water, and dried under vacuum in a Speed-Vac for 5-10 min. Sodium methoxide (100 μl of 1 %) was added to the sample, and then the sample was vortexed and rocked for 20 min. After adding 3 drops of 1 M NaCl and 400 μl hexane, the sample was vortexed and spun. The upper layer was removed and analyzed by GC as described above.

EXAMPLE 1

Eutreptiella cf_gymnastica CCMP1594 Lipid Profile, Total RNA Isolation And Genomic DNA Isolation

[0091] *Eutreptiella cf_gymnastica* CCMP1594 cells (1 liter of culture) were purchased from the Provasoli-Guillard National Center for Culture of Marine Phytoplankton (CCMP) (Bigelow Laboratory for Ocean Sciences, West Boothbay Harbor, Maine). Cells from 50 mL culture were resuspended in 600 μl of sodium methoxide dissolved in methanol. The sample was shaken for 20 min, and 50 μl of 1 M NaCl was added. After mixing, 600 μl of heptane was added. The sample was vortexed and

centrifuged in an Eppendorf microfuge for 1 min. The upper layer was carefully separated from the lower layer and placed in a glass vial for GC analysis. The results of the analysis are shown below in Table 3. Fatty acids are identified as 16:0 (palmitate), 16:1 (palmitoleic acid), 18:0, 18:1 (oleic acid), 18:2, GLA, ALA, DGLA, ARA, EPA, DPA and DHA; and the composition of each is presented as the weight percent of total fatty acids ["TFAs"].

Table 3

Lipid Profile Of *Eutreptiella cf_gymnastica* CCMP1594 Cells

16:0	16:1	18:0	18:1	18:2	GLA	ALA	DGLA	ARA	EPA	DPA	DHA
18.5	2.5	10	27.5	5	0	10.2	0.1	0.3	5.3	4.7	10.7

[0092] Based on the presence of EPA, DPA and DHA, it was concluded that the *Eutreptiella cf_gymnastica* CCMP1594 had a functional Δ4 desaturase capable of converting DPA (22:5, ω-3) to DHA (22:6, ω-3).

[0093] Total RNA and genomic DNA were isolated from *Eutreptiella cf_gymnastica* CCMP1594 using the trizol reagent (Invitrogen, Carlsbad, CA), according to the manufacturer's protocol. Cell pellet from the 1 L culture (~0.25 mL in volume) was resuspended in 0.75 mL of trizol reagent, mixed with 0.5 mL of 0.5 mm glass beads, and homogenized in a Biospec mini beadbeater (Bartlesville, OK) at the highest setting for 3 min. The mixture was centrifuged in an Eppendorf centrifuge for 30 sec at 14,000 rpm to remove debri and glass beads. The supernatant was extracted with 150 µL of 24:1 chloroform:isoamy alcohol. The upper aqueous phase was used for RNA isolation, while the lower organic phase was used for DNA isolation.

[0094] For RNA isolation, the aqueous phase was mixed with 0.375 mL of isopropyl alcohol and allowed to incubate at room temperature for 5 min. Precipitated RNA was collected by centrifugation at 8,000 rpm and 4 °C

for 5 min. The pellet was washed once with 0.7 mL of 80% ethanol and air dried. Thus, 720 µg of total RNA were obtained.

[0095] For genomic DNA isolation, the lower organic phase was mixed with 75 µL of ethanol and incubated at room temperature for 5 min. The sample was then centrifuged at 5,000 rpm for 2 min in an Eppendorf centrifuge. The pellet was washed with 0.75 mL of 0.1 M sodium citrate in 10% ethanol twice. Each time, the sample was incubated for 15 min at room temperature in the wash solution, followed by centrifugation at 5,000 rpm for 5 min at 4 °C in an Eppendorf centrifuge. The pellet was air dried and re-dissolved in 300 µL of 8 mM NaOH. The pH of the sample was adjusted to 7.5 with 1 M HEPES. The DNA sample was then further purified with the Qiagen PCR purification kit according to the manufacturer's protocol. In this way, 45 µg genomic DNA were obtained from *Eutreptiella cf_gymnastica* CCMP1594.

EXAMPLE 2

Eutreptiella cf_gymnastica CCMP1594 cDNA Synthesis

[0096] cDNA was synthesized directly from the *Eutreptiella cf_gymnastica* CCMP1594 mRNA as follows. Total RNA (2.4 µg) from *Eutreptiella cf_gymnastica* CCMP1594 was used as template to synthesize double stranded cDNA. The Creator™ SMART™ cDNA Library Construction Kit from BD Bioscience Clontech (Palo Alto, CA) was used. One (1) µL of the total RNA sample was mixed with 1 µL of SMART IV oligonucleotide (SEQ ID NO:10), 1 µL CDSIII/3' PCR primer (SEQ ID NO:11) and 2 µL of water. The mixture was heated to 75 °C for 5 min and then cooled on ice for 5 min. To the sample were added 2 µL of 5X first strand buffer, 1 µL 20 mM DTT, 1 µL of dNTP mix (10 mM each of dATP, dCTP, dGTP and dTTP) and 1 µL of PowerScript reverse transcriptase. The sample was incubated at 42 °C for 1 h. The resulting first strand cDNA was then used as template for amplification. The reaction mixture contained 2 µL of the above first strand cDNA sample, 80 µL of water, 10 µL of 10X Advantage 2 PCR buffer, 2 µL 50X dNTP mix (10 mM each of dATP, dCTP, dGTP

and dTTP), 2 µL of 5' CDSIII PCR primer (SEQ ID NO:12), 2 µL CDSIII/3' PCR primer (SEQ ID NO:11) and 2 µL 50X Advantage 2 polymerase mix. PCR amplification was performed using the following conditions: 95 °C for 1 min, followed by 20 cycles of 95 °C for 10 sec and 68 °C for 6 min. Amplification product was purified with a Qiagen PCR purification kit according to the manufacturer's protocol. Purified product was eluted with 50 µL of water.

EXAMPLE 3

Isolation Of A Portion Of The Coding Region Of The *Eutreptiella cf_gymnastica* CCMP1594 Δ4 Desaturase Gene

[0097] The present Example describes the identification of a portion of the *Eutreptiella cf_gymnastica* CCMP1594 gene encoding Δ4 desaturase (designated herein as "E1594D4" (SEQ ID NOs:1 and 2)), by use of primers derived from conserved regions of other known Δ4 desaturase sequences.

[0098] The *Euglena gracilis* Δ4 fatty acid desaturase (SEQ ID NO:13; GenBank Accession No. AY278558; Meyer et al., *Biochemistry,* 42(32):9779-9788 (2003)), *Thalassiosira pseudonana* Δ4 fatty acid desaturase (SEQ ID NO:37; GenBank Accession No. AAX14506; Tonon et al., *FEBS J.,* 272 (13):3401-3412 (2005)), *Thraustochytrium* sp. FJN-10 Δ4 fatty acid desaturase (SEQ ID NO:38; GenBank Accession No. AAZ43257), and *Pavlova lutheri* (SEQ ID NO:42; GenBank Accession No. AAQ98793; Tonon et al., *FEBS Lett.,* 553(3):440-444 (2003)) were aligned as shown in FIG. 2, using the method of Clustal W (slow, accurate, Gonnet option; Thompson et al., *Nucleic Acids Res.*, 22:4673-4680 (1994)) of the MegAlign™ program of DNASTAR software. Based on this alignment, degenerate primers were designed as shown in Table 4 (location of primers with respect to SEQ ID NOs:13, 37, 38 and 42 are shown within the boxed regions of FIG. 2).

Table 4

Degenerate Oligonucleotides Used To Amplify The Δ4 Desaturase Gene From *Eutreptiella cf_gymnastica* CCMP1594

Primer	Nucleotide Sequence	Amino Acid Sequence
D4-F1	TTCCTNGCNAARCAYCCNGG (SEQ ID NO:15)	FLAKHPG (SEQ ID NO:18)
D4-F2	TTTCTNGCNAARCAYCCNGG (SEQ ID NO:16)	FLAKHPG (SEQ ID NO:18)
D4-F3	TTYTTRGCNAARCAYCCNGG (SEQ ID NO:17)	FLAKHPG (SEQ ID NO:18)
D4-F4	ATHCARCAYGAYGGNAAYCA (SEQ ID NO:19)	IQHDGNH (SEQ ID NO:20)
D4-F5	CAYGAYGGNAAYCAYGGNGC (SEQ ID NO:21)	HDGNHGA (SEQ ID NO:22)
D4-F6	GGNCAYCAYAGYTTYACNAA (SEQ ID NO:23)	GHHQYTN (SEQ ID NO:26)
D4-F7	GGYCAYCAYTCNTTYACNAA (SEQ ID NO:24)	GHHQYTN (SEQ ID NO:26)
D4-F8	GGRCAYCAYTCNTTYACNAA (SEQ ID NO:25)	GHHQYTN (SEQ ID NO:26)
D4-R1	AANAGRTGRTGYTCDATYTG (SEQ ID NO:27)	QIEHHLF (SEQ ID NO:29)
D4-R2	AAYAARTGRTGYTCDATYTG (SEQ ID NO:28)	QIEHHLF (SEQ ID NO:29)

[Note: The nucleic acid degeneracy code used for SEQ ID NOs:15-29 was as follows: R= A/G; Y=C/T; H= A/C/T; D=A/G/T; and N=A/C/T/G.]

[0099] A total of 16 different PCR amplification reactions were performed, using all possible combinations of the 8 forward and 2 reverse primers. Each reaction mixture contained 1 μl of 1:10 diluted *Eutreptiella cf_gymnastica* CCMP1594 cDNA (from Example 2), 5 μl each of the forward and reverse primers (20 μM), 14 μl water and 25 μl of TaKaRa ExTaq 2X premix (TaKaRa Bio, Mountain View, CA). The thermocycler conditions were set for 94 °C for 1 min, then 30 cycles of 94 °C for 20 sec, 55 °C for 20 sec and 72 °C for 1 min, followed by a final extension at 72 °C for 7 min. PCR products were analyzed by electrophoresis on standard agarose gels, and putative Δ4 desaturase fragments were detected as shown below in Table 5.

Table 5

Detected Putative Δ4 Desaturase Fragments

Product	Forward Primer	Reverse Primer
~800 bp fragment	D4-F3 or D4-F4	D4-R1
~800 bp fragment	D4-F3	D4-R2
~700 bp fragment	D4-F6, D4-F7 or D4-F8	D4-R1 or D4-R2

[0100] Each of the fragments described above in Table 5 were purified with a Qiagen PCR purification kit (Valencia, CA), cloned into pCR2.1-TOPO (Invitrogen) and sequenced.

[0101] Identity of the *Eutreptiella cf_gymnastica* CCMP1594 sequences were determined by conducting BLAST (Basic Local Alignment Search Tool; Altschul, S. F., et al., *J. Mol. Biol.*, 215:403-410 (1993)) searches for similarity to sequences contained in the BLAST "nr" database (comprising all non-redundant GenBank CDS translations, sequences derived from the 3-dimensional structure Brookhaven Protein Data Bank, the SWISS-PROT protein sequence database, EMBL and DDBJ databases). The sequences were analyzed for similarity to all publicly available DNA sequences contained in the "nr" database using the BLASTN algorithm provided by the National Center for Biotechnology Information ["NCBI"].

[0102] BLAST sequence analysis showed that the fragments generated by primer pairs D4-F4/D4-R1 and D4-F7/D4-R1 were from a single gene that showed extensive homology to the known Δ4 desaturases from other organisms. The sequences were assembled into a 847 bp contig (SEQ ID NO:5), which was assumed to encode a portion of a putative Δ4 desaturase from *Eutreptiella cf_gymnastica* CCMP1594.

EXAMPLE 4

Isolation Of The Full-Length Δ4 Desaturase From *Eutreptiella cf_gymnastica* CCMP1594

[0103] Primers were designed to isolate the 5' and 3' ends of the putative Δ4 desaturase gene from cDNA samples of *Eutreptiella cf_gymnastica* CCMP1594, based on the partial 847 bp sequence set forth in SEQ ID NO:5 and described in Example 3.

Isolation Of The Δ4 Desaturase 5' Coding Region

[0104] The 5' region of the putative Δ4 desaturase from *Eutreptiella cf_gymnastica* CCMP1594 was isolated by nested PCR amplification of cDNA ends. Based on the partial sequence of the putative Δ4 desaturase gene, primer 1594D4-5-1 (SEQ ID NO:30) was used in combination with 5' CDSIII PCR primer (SEQ ID NO:12) from the BD-Clontech Creator™ Smart™ cDNA library kit for the first round of amplification. The reaction mixture contained 1 μl of each primer (10 μM), 1 μl of *Eutreptiella cf_gymnastica* CCMP1594 cDNA (~50 ng), 22 μl water and 25 μl TaKaRa ExTaq 2X premix. The thermocycler conditions were set for 94 °C for 60 sec, then 30 cycles at 94 °C for 20 sec, 55 °C for 20 sec and 72 °C for 30 sec, followed by a final extension at 72 °C for 5 min.

[0105] The second round of PCR amplification used 1 μl of diluted product from the first round PCR reaction as template, wherein the PCR product was diluted 1:50 in water. Amplification was conducted as described above, with the exception that 1 μl each of primers 1594D4-5-2 (SEQ ID NO:31) and DNR CDS 5-2 (SEQ ID NO:32) were used (stock solution of 10 μM for each primer).

[0106] A 359 bp DNA fragment from the second round PCR product was cloned into pCR2.1-TOPO (Invitrogen) and sequenced. The fragment (SEQ ID NO:6) was designated "E1594D4-5'-A", as analysis showed that this fragment overlapped partly with the original E1594D4 partial fragment (SEQ ID NO:5) and extended further upstream. However, there was no translation initiation codon in the extended 359 bp fragment of SEQ ID NO:6. Based on sequence comparison with known Δ4 desaturases, approximately 400 bp was assumed to be missing from the 5'-end.

[0107] The methodology utilized above to obtain fragment E1594D4-5'-A was repeated to obtain an additional 5' region of the E1594D4 gene, utilizing identical PCR conditions to those described above. However,

primer 1594D4-5-4 (SEQ ID NO:33) replaced primer 1594D4-5-1 (SEQ ID NO:30) in the first round of amplification. Following a 1:50 dilution of the first round product, a second round PCR was conducted using primer 1594D4-5-5 (SEQ ID NO:34) instead of primer 1594D4-5-2 (SEQ ID NO:31).

[0108] A ~400 bp DNA fragment in the second round PCR product was cloned into pCR2.1-TOPO and sequenced. Sequence analysis showed that this fragment contained the 5' end of the E1594D4 gene, including the start codon and 9 bp of the 5' untranslated region. The fragment was designated "1594D4-5'-B" (SEQ ID NO:7).

Isolation Of The Δ4 Desaturase 3' Coding Region

[0109] The 3' region of the putative Δ4 desaturase was also isolated by nested PCR amplification. In the first round, the reaction mixture contained 1 μl each of primer 1594D4-3-1 (SEQ ID NO:35, 10 μM) and primer CDSIII/3' PCR primer (SEQ ID NO:11, 10 μM), 1 μl of *Eutreptiella cf_gymnastica* CCMP1594 cDNA (~50 ng), 22 μl water and 25 μl TaKaRa ExTaq 2X premix. The thermocycler conditions were set for 94 °C for 60 sec, then 30 cycles at 94 °C for 20 sec, 55 °C for 20 sec and 72 °C for 30 sec, followed by a final extension at 72 °C for 7 min. For the second round PCR, the reaction mixture contained 1 μl each of primer 1594D4-3-2 (SEQ ID NO:36, 10 μM) and CDSIII/3' PCR primer (SEQ ID NO:11, 10 μM), 1 μl of 1:50 diluted first round PCR product, 22 μl water and 25 μl TaKaRa ExTaq 2X premix. PCR conditions were otherwise identical to that used for the first round PCR.

[0110] A ~900 bp DNA fragment was generated by the second round of PCR. This fragment was cloned into pCR2.1-TOPO and sequenced. Sequence analysis showed that this fragment (designated as "1594D4-3" and set forth as SEQ ID NO:8) included the 3' region of the E1594D4 gene.

Assembly And Analysis Of The Complete Δ4 Desaturase Coding Sequence

[0111] The cDNA sequence of the entire putative *Eutreptiella cf_gymnastica* CCMP1594 Δ4 desaturase (E1594D4) gene was determined by assembly of the E1594D4 partial fragment (SEQ ID NO:5), the 1594D4-5'-A fragment (SEQ ID NO:6), the 1594D4-5'-B fragment (SEQ ID NO:7) and the 1594D4-3' fragment (SEQ ID NO:8). The 2070 bp cDNA sequence, including 9 bp of the 5' untranslated region and 516 bp of 3' untranslated region, was designated "E1594D4-cDNA" (SEQ ID NO:9). The E1594D4 CDS was 1345 bp in length (SEQ ID NO:1) and encoded a polypeptide of 514 amino acids (SEQ ID NO:2).

[0112] The E1594D4 sequence (i.e., SEQ ID NO:2) was compared for similarity to all publicly available protein sequences contained in the "nr" database, using the BLASTX algorithm (Gish, W. and States, D. J., *Nature Genetics,* 3:266-272 (1993)) provided by the NCBI. The results of the BLASTX comparison summarizing the sequence to which SEQ ID NO:2 has the most similarity are reported according to the % identity, % similarity and Expectation value. "% Identity" is defined as the percentage of amino acids that are identical between the two proteins. "% Similarity" is defined as the percentage of amino acids that are identical or conserved between the two proteins. "Expectation value" estimates the statistical significance of the match, specifying the number of matches, with a given score, that are expected in a search of a database of this size absolutely by chance.

[0113] Thus, SEQ ID NO:2 was found to share 65% identity and 76% similarity with the amino acid sequence of the Δ4 fatty acid desaturase from *Thalassiosira pseuduonana* (SEQ ID NO:37; GenBank Accession No. AAX14506), with an Expectation value of 0.0. Additionally, the full length E1594D4 gene shared identity and similarity with other Δ4 fatty acid desaturases. More specifically, pairwise comparison between and among

Δ4 desaturase proteins from *Eutreptiella cf_gymnastica* CCMP1594 (SEQ ID NO:2), *Thalassiosira pseduonana* (SEQ ID NO:37, *supra*), *Euglena gracilis* (SEQ ID NO:13; GenBank Accession No. AY278558) and *Thraustochytrium* sp. FJN-10 (SEQ ID NO:38; GenBank Accession No. AAZ43257) using a Clustal W analysis (MegAlign™ program of DNASTAR software) resulted in the percent similarities shown below in Table 6.

Table 6

Percent Similarities Between And Among Various Δ4 Desaturases

	SEQ ID NO:2	SEQ ID NO:37	SEQ ID NO:13	SEQ ID NO:38
Eutreptiella cf_gymnastica CCMP1594 (SEQ ID NO:2)	100	68	40	59
Thalassiosira pseuduonana (SEQ ID NO:37)	--	100	41	56
Euglena gracilis (SEQ ID NO:13)	--	--	100	42
Thraustochytrium sp. FJN-10 (SEQ ID NO:38)	--	--	--	100

EXAMPLE 5

Synthesis Of A Codon-Optimized Δ4 Desaturase Gene ["E1594D4S"] For *Yarrowia lipolytica*

[0114] The codon usage of the Δ4 desaturase gene of *Eutreptiella cf_gymnastica* CCMP1594 (SEQ ID NOs:1 and 2; "E1594D4") was optimized for expression in *Yarrowia lipolytica,* in a manner similar to that described in Int'l. App. Pub. No. WO 2004/101753 and U.S. Pat. No. 7,125,672. Specifically, a codon-optimized Δ4 desaturase gene (designated "E1594D4S", SEQ ID NOs:3 and 4) was designed based on the coding sequence of the Δ4 desaturase gene of E1594D4, according to the *Yarrowia* codon usage pattern (Int'l. App. Pub. No. WO 2004/101753), the consensus sequence around the 'ATG' translation initiation codon, and the general rules of RNA stability (Guhaniyogi, G. and J. Brewer, *Gene,* 265(1-2):11-23 (2001)). A total of 200 bp of the 1545 bp coding region were modified (12.9%; FIG. 3) and 191 codons were optimized (37.1%). The GC content was reduced from 56.1% within the wild type gene (i.e.,

E1594D4) to 54.6% within the synthetic gene (i.e., E1594D4S). A *Nco*I site and *Not*I sites were incorporated around the translation initiation codon and after the stop codon of E1594D4S, respectively. In order to add a *Nco*I site around the translation initiation codon, E1594D4S had one additional alanine amino acid inserted between amino acid residues 1 and 2 of the wildtype E1594D4; thus, the total length of E1594D4S is 515 amino acids (SEQ ID NO:4). The designed E1594D4S gene (SEQ ID NO:3; labeled as "1594D4S" in FIG. 4A) was synthesized by GenScript Corporation (Piscataway, NJ) and cloned into pUC57 (GenBank Accession No. Y14837) to generate p1594D4S (FIG. 4A; SEQ ID NO:39).

EXAMPLE 6

Generation Of Construct pZKL4-220ESC4, Comprising E1594D4S

[0115] The present Example describes the construction of plasmid pZKL4-220ESC4. This plasmid was constructed to integrate two chimeric $C_{20/22}$ elongase genes and one chimeric E1594D4S gene into the lipase 4-like locus (GenBank Accession No. XM_503825) of *Yarrowia lipolytica.* This was designed to integrate the chimeric genes into the genome of *Yarrowia lipolytica* and then permit study of the function of the codon-optimized Δ4 desaturase derived from *Eutreptiella cf_gymnastica* CCMP1594 in *Yarrowia lipolytica.*

[0116] Plasmid pZKL4-220ESC4 (FIG. 4B) contained the following components:

Table 7

Components Of Plasmid pZKL4-220ESC4 (SEQ ID NO:40)

RE Sites And Nucleotides Within SEQ ID NO:40	Description Of Fragment And Chimeric Gene Components
Asc I/*Bsi*W I (6742-5990)	745 bp 5' portion of the *Yarrowia* Lipase 4-like gene (labeled as "Lip4L-5'" in Figure; GenBank Accession No. XM_503825)
*Pac*I / *Sph*I (10238-9450)	782 bp 3' portion of *Yarrowia* Lipase 4-like gene (labeled as "Lip4L-3'" in Figure; GenBank Accession No. XM_503825)

Swa I/*BsiW* I (3847-5990)	FBAINm::EaC20ES::Pex20, comprising: • FBAINm: *Yarrowia lipolytica* FBAINm promoter (U.S. Pat. No. 7,202,356); • EaC20ES: codon-optimized C20 elongase gene (SEQ ID NO:46), derived from *Euglena anabaena* (U.S. Pat. Appl. Pub. No. 2008/0254191-A1); • Pex20: Pex20 terminator sequence from *Yarrowia Pex20* gene (GenBank Accession No. AF054613)
Pme I/*Swa* I (1868-3847)	YAT1::EgC20ES::Lip1, comprising: • YAT1: *Yarrowia lipolytica* YAT1 promoter (U.S. Pat. Appl. Pub. No. 2006/0094102-A1); • EgC20ES: codon-optimized C20 elongase gene (SEQ ID NO:44), derived from *Euglena gracilis* (U.S. Pat. Appl. Pub. No. 2008/0254191-A1); • Lip1: Lip1 terminator sequence from *Yarrowia Lip1* gene (GenBank Accession No. Z50020)
Cla I/*Pme* I (12339-1868)	EXP1::E1594D4S::Oct, comprising: • EXP1: *Yarrowia lipolytica* export protein (EXP1) promoter (Int'l. App. Pub. No. WO 2006/052870); • E1594D4S: codon-optimized Δ4 desaturase (SEQ ID NO:3), derived from *Eutreptiella cf_gymnastica* CCMP1594 (labeled as "D4S-1594" in Figure); • OCT: OCT terminator sequence of *Yarrowia OCT* gene (GenBank Accession No. X69988)
Sal I/*EcoR* I (11889-10270)	• *Yarrowia* Ura3 gene (GenBank Accession No. AJ306421)

EXAMPLE 7

Expression Of The Codon-Optimized Δ4 Desaturase ("E1594D4S") In *Yarrowia lipolytica* Strain Y4184U4

[0117] The pZKL4-220ESC4 plasmid comprising E1594D4S (Example 6) was digested with *Asc*I/*Sph*I, and then used for transformation of strain Y4184U4 (General Methods), using standard transformation procedures. The transformants were selected on MM plates. After 4 days growth at 30 °C, 3 transformants grown on the MM plates were picked and re-streaked onto fresh MM plates. Once grown, these strains and the control strain were individually inoculated into 3 mL liquid MM at 30 °C and shaken at 250 rpm/min for 2 days. The cells were collected by centrifugation, resuspended in High Glucose Media ["HGM"] and then shaken at 250 rpm/min for 5 days. The cells were collected by centrifugation, lipids were extracted, and fatty acid methyl esters ["FAMEs"] were prepared by trans-

esterification, and subsequently analyzed with a Hewlett-Packard 6890 GC.

[0118] Results are shown below in Table 8. Specifically, fatty acids are identified as 16:0 (palmitate), 16:1, 18:0 (stearic acid), 18:1 (oleic acid), LA, ALA, EDA, DGLA, ARA, ETrA, ETA, EPA, DPA and DHA; and, fatty acid compositions were expressed as the weight percent (wt. %) of total fatty acids ["TFAs"].

Table 8

Fatty Acid Composition In Transformants Expressing E1594D4S (SEQ ID NO:3)

	Fatty acid composition (wt. % of TFAs)													
	16:0	16:1	18:0	18:1	LA	ALA	EDA	DGLA	ARA	ETrA	ETA	EPA	DPA	DHA
Y4184-Control														
1	3.6	1.4	1.9	8.3	29.0	6.3	4.7	1.1	0.4	2.1	1.1	30.4	0.5	0.0
2	3.8	1.7	1.7	8.4	31.1	6.2	4.3	1.1	0.4	1.9	1.2	28.8	0.3	0.0
3	4.5	1.8	1.9	8.1	32.0	5.9	4.8	1.2	0.4	1.9	1.2	28.0	0.2	0.0
Average	4.0	1.6	1.8	8.3	30.7	6.2	4.6	1.1	0.4	2.0	1.2	29.1	0.3	0.0
Vector pZKL4-220EC4-1 in Y4184U4														
1	4.5	1.3	1.5	6.3	29.4	5.1	3.3	1.2	0.5	1.4	1.0	22.1	9.8	2.0
2	4.3	1.3	1.9	7.0	29.5	5.7	3.2	1.1	0.5	1.4	1.0	21.2	10.4	1.9
3	4.3	1.6	1.4	6.2	29.1	4.0	3.3	1.3	0.8	1.4	0.8	22.0	10.3	2.0
Average	4.4	1.4	1.6	6.5	29.3	4.9	3.2	1.2	0.6	1.4	0.9	21.7	10.2	2.0

[0119] The GC analyses of Table 8 showed that there were about 2% DHA and 10.2% DPA of total lipids produced in all three transformants, but not in the control Y4184 strain. Conversion efficiency of the substrate, DPA, to DHA in the three transformant strains expressing E1594D4S was determined to be about 16%. The conversion efficiency was measured according to the following formula: ([product]/[substrate + product])*100, where 'product' includes the immediate product and all products in the pathway derived from it. Thus, this experimental data demonstrated that the synthetic Δ4 desaturase derived from *Eutreptiella cf_gymnastica* CCMP1594 and codon-optimized for expression in *Yarrowia lipolytica* (E1594D4S, as set forth in SEQ ID NO:3) was active to convert the substrate, DPA, to DHA.

The Claims defining the invention are as follows:

1. An isolated nucleic acid molecule selected from the group consisting of:
   (a) an isolated nucleotide sequence encoding a delta-4 desaturase enzyme selected from the group consisting of SEQ ID NO:2 and SEQ ID NO:4;
   (b) an isolated nucleotide sequence that hybridizes with (a) under the following hybridization conditions: 0.1X SSC, 0.1% SDS, 65 °C and washed with 2X SSC, 0.1% SDS followed by 0.1X SSC, 0.1% SDS;
   (c) an isolated nucleotide sequence that is completely complementary to (a) or (b); and
   (d) an isolated nucleotide sequence encoding a delta-4 desaturase enzyme of at least 514 amino acids that has at least 68% identity based on the Clustal W method of alignment when compared to the amino acid sequence set forth in SEQ ID NO:2.

2. The isolated nucleic acid molecule of Claim 1, wherein the nucleic acid molecule is the isolated nucleotide sequence of (a).

3. The isolated nucleic acid molecule of Claim 1, wherein at least 191 codons are codon-optimized for expression in *Yarrowia*.

4. The isolated nucleic acid molecule of Claim 1 selected from the group consisting of SEQ ID NO:1 and SEQ ID NO:3.

5. A chimeric gene comprising the isolated nucleic acid molecule of any one of Claims 1 to 4 operably linked to at least one regulatory sequence.

6. A host cell comprising the chimeric gene of Claim 5.

7. The host cell of Claim 6, wherein the host cell is selected from the group consisting of yeast, algae, bacteria, euglenoids, stramenopiles, fungi and plants.

8. The host cell of Claim 7, wherein the host cell is a yeast.

9. The host cell of Claim 8, wherein the yeast is an oleaginous yeast.

10. The host cell of Claim 9, wherein the oleaginous yeast is selected from the group consisting of *Yarrowia, Candida, Rhodotorula, Rhodosporidium*, *Cryptococcus, Trichosporon* and *Lipomyces.*

11. The host cell of claim 10, wherein the host cell is a *Yarrowia* sp.

12. The host cell of claim 7, wherein the host cell is a plant cell, wherein said plant is selected from the group consisting of soybean, corn, flax, rapeseed, primrose, canola, maize, cotton, safflower and sunflower.

13. A method of making a polyunsaturated fatty acid selected from the group consisting of docosahexaenoic acid and all-*cis*-4,7,10,13,16-docosapentaenoic acid (22:5, omega-6), comprising:

a) providing a host cell comprising:

(i) an isolated nucleotide molecule encoding a delta-4 desaturase polypeptide having at least 68% amino acid identity when compared to a polypeptide having the amino acid sequence as set forth in SEQ ID NO:2, based on the Clustal W method of alignment; and,

(ii) a source fatty acid selected from the group consisting of all-*cis*-7,10,13,16,19-docosapentaenoic acid (22:5, omega-3) and docosatetraenoic acid;

b) growing the host cell of step (a) under conditions to express the nucleotide molecule encoding the delta-4 desaturase polypeptide and to convert the source fatty acid to a polyunsaturated fatty acid selected from the group consisting of docosahexaenoic acid and all-cis-4,7,10,13,16-docosapentaenoic acid (22:5, omega-6),
such that when all-*cis*-7,10,13,16,19-docosapentaenoic acid (22:5, omega-3) is the source fatty acid, then docosahexaenoic acid is the

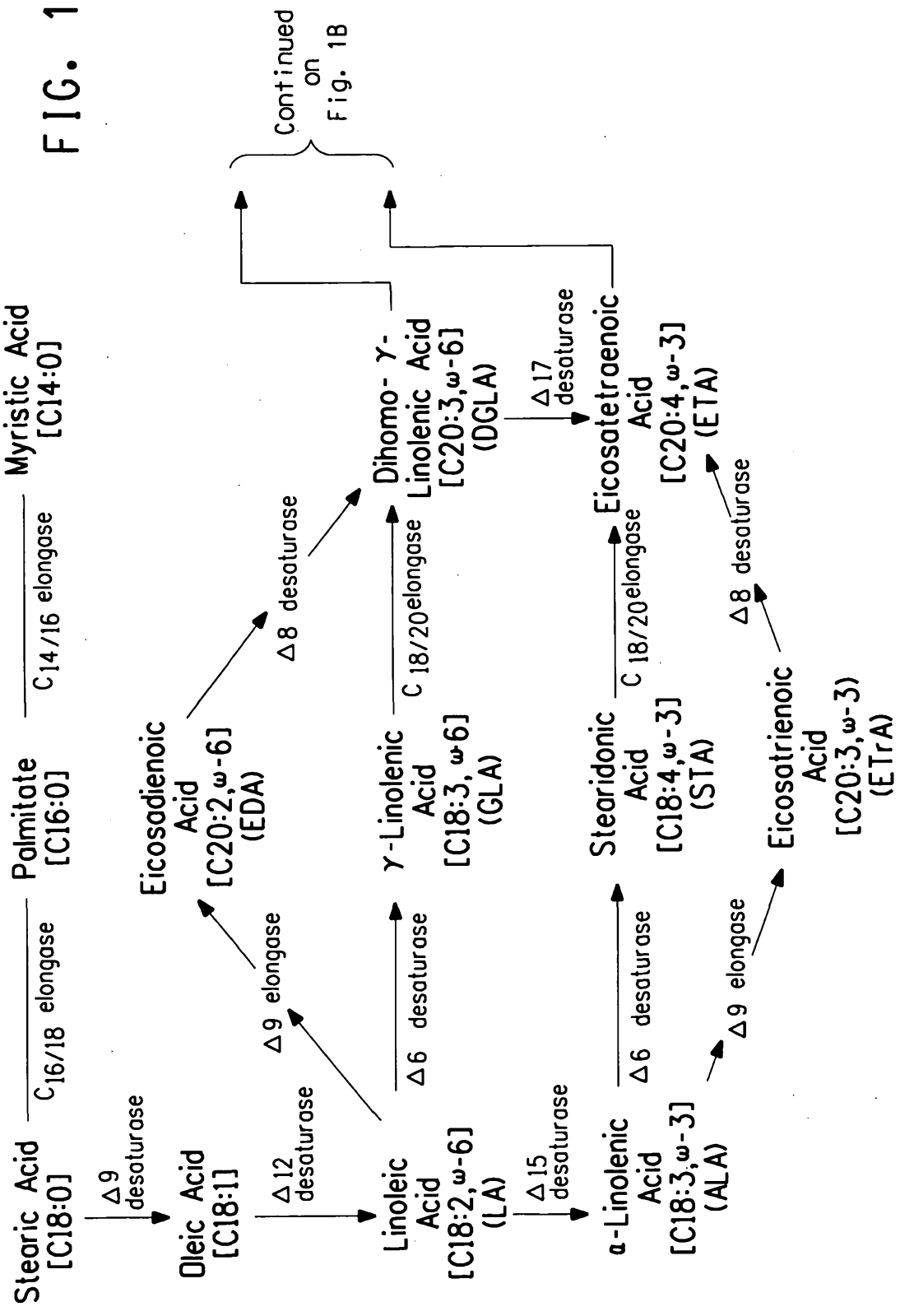
polyunsaturated fatty acid produced; and when docosatetraenoic acid is the source fatty acid, then all-*cis*-4,7,10,13,16-docosapentaenoic acid (22:5, omega-6) is the polyunsaturated fatty acid produced; and,

c) optionally recovering the polyunsaturated fatty acid produced in step (b).

14. The method of Claim 13, wherein the isolated nucleotide molecule encodes a delta-4 desaturase polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:2 and SEQ ID NO:4.

15. The method of Claim 14, wherein:
the isolated nucleotide molecule has a nucleic acid sequence selected from the group consisting of SEQ ID NO:1 and SEQ ID NO:3; and the host cell is selected from the group consisting of: algae; bacteria; yeast; oleaginous yeast preferably selected from the group consisting of *Yarrowia, Candida, Rhodotorula, Rhodosporidium*, *Cryptococcus, Trichosporon* and *Lipomyces;* stramenopiles; euglenoids, fungi; plant cells and animal cells.

FIG. 1A


Stearic Acid [C18:0]
Δ9 desaturase
Oleic Acid [C18:1]
Δ12 desaturase
Linoleic Acid [C18:2,ω-6] (LA)
Δ15 desaturase
α-Linolenic Acid [C18:3,ω-3] (ALA)
C16/18 elongase
Palmitate [C16:0]
C14/16 elongase
Myristic Acid [C14:0]
Δ9 elongase
Eicosadienoic Acid [C20:2,ω-6] (EDA)
Δ6 desaturase
γ-Linolenic Acid [C18:3, ω6] (GLA)
Δ6 desaturase
Stearidonic Acid [C18:4,ω-3] (STA)
Δ9 elongase
Eicosatrienoic Acid [C20:3,ω-3) (ETrA)
Δ8 desaturase
C18/20elongase
Dihomo- γ-Linolenic Acid [C20:3,ω-6] (DGLA)
Δ8 desaturase
C18/20elongase
Eicosatetraenoic Acid [C20:4,ω-3] (ETA)
Δ17 desaturase
Continued on Fig. 1B

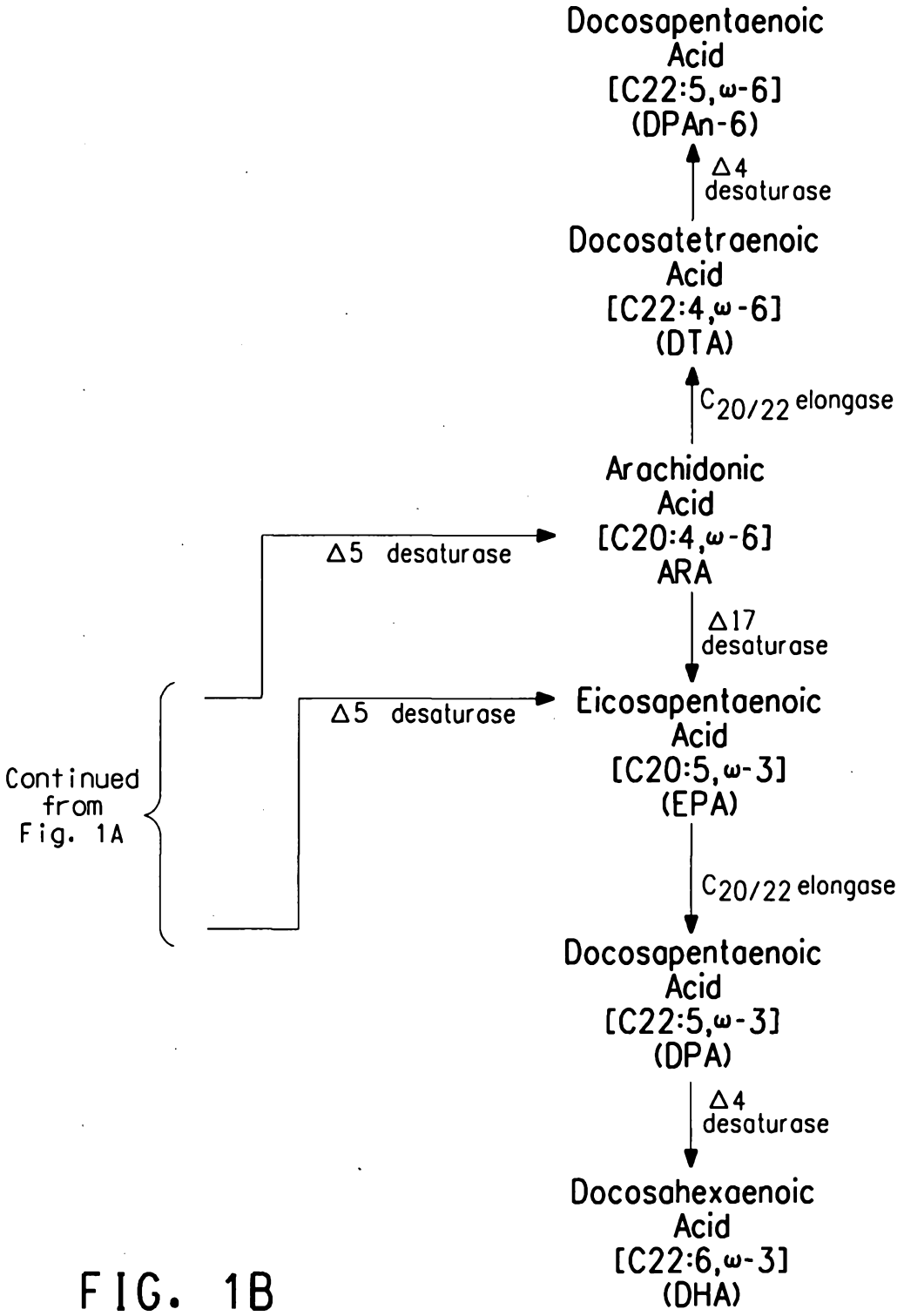

Docosapentaenoic Acid [C22:5,ω-6] (DPAn-6)
Δ4 desaturase
Docosatetraenoic Acid [C22:4,ω-6] (DTA)
C20/22 elongase
Arachidonic Acid [C20:4,ω-6] ARA
Δ5 desaturase
Δ17 desaturase
Δ5 desaturase
Eicosapentaenoic Acid [C20:5,ω-3] (EPA)
Continued from Fig. 1A
C20/22 elongase
Docosapentaenoic Acid [C22:5,ω-3] (DPA)
Δ4 desaturase
Docosahexaenoic Acid [C22:6,ω-3] (DHA)

FIG. 1B

```
                          1                                                  50
SEQ ID NO:13      (1)  MLVLFGNFYVKQYSQKNGKPENGATPENGAKPQPCENGTVEKRENDTANV
SEQ ID NO:37      (1)  -------------MGNGNLPASTAQLKSTSKPQ-----------------
SEQ ID NO:38      (1)  ----------------------------------M-----------------
SEQ ID NO:42      (1)  ------------------------MPPSAASEG-----------------

                          51                                                 100
SEQ ID NO:13     (51)  RPTRPAGPPPATYYDSLAVSGQGKERLFTTDEVRRHILPTDGWLTCHEG-
SEQ ID NO:37     (21)  --------------------QQHEHRTISKSELAQHNTPKSAWCAVHSTP
SEQ ID NO:38      (2)  --------------------TVGYDGEIPFEQVRAHNKPDDAWCAIHG--
SEQ ID NO:42     (10)  --------------------GVAELRAAEVASYTRKAVDERPDLTIVG--

                          101                                                150
SEQ ID NO:13    (100)  ---------------VYDVTDFLAKHPGGG-VITLGLGRDCTILIESYHP
SEQ ID NO:37     (51)  ATDPSHSNNKQHAHLVLDITDFASRHPGGD-LILLASGKDASVLFETYHP
SEQ ID NO:38     (30)  --------------HVYDVTKFASVHPGGD-IILLAAGKDATVLYETYHV
SEQ ID NO:42     (38)  -------------DAVYDAKAFRDEHPGGAHFVSLFGGRDATEAFMEYHR

                          151                                                200
SEQ ID NO:13    (134)  AGRPDKVMEKYRIGTLQDP----------------KTFYAWGESDFYPEL
SEQ ID NO:37    (100)  RGVPTSLIQKLQIGVMEE-------------EAFRDSFYSWTDSDFYTVL
SEQ ID NO:38     (65)  RGVSDAVLRKYRIGKLPDGQGGANEKEKRTLSGLSSASYYTWNSDFYRVM
SEQ ID NO:42     (75)  RAWPKARMSKFFVGSLDAS-----------------EKP-TQADSAYLRL

                          201                                                250
SEQ ID NO:13    (168)  KRRALARLKEAGQARRGG--LGVKALLVLTLFFVSWYMWVA------HKS
SEQ ID NO:37    (137)  KRRVVERLEERGLDRRGSKEIWIKALFLLVGFWYCLYKMYTTSDIDQYGI
SEQ ID NO:38    (115)  RERVVARLKERGKARRGGYELWIKALLLLVGFWSSLCWMCT---LDPSFG
SEQ ID NO:42    (107)  CAEVNALLPK-GSGGFAPPSYWLKAAALVVAAVSIEGYMLLR------GK
```

FIG. 2A

```
                          251                                                300
SEQ ID NO:13   (210) FLWA-AVWGFAGSHVGLSIQHDGNHGAFSRNTLVNRLAGWGMDLIGASST
SEQ ID NO:37   (187) AIAYSIGMGTFAAFIGTCIQHDGNHGAFAQNKLLNKLAGWTLDMIGASAF
SEQ ID NO:38   (162) AILAAMSLGVFAAFVGTCIQHDGNHGAFAQSRWVNKVAGWTLDMIGASGM
SEQ ID NO:42   (150) TLLLSVFLGLVFAWIGLNIQHDANHGALSRHSVINYCLGYAQDWIGGNMV

                          301                                                350
SEQ ID NO:13   (259) VWEYQHVIGHHQYTNLVS-------------DTLFSLPENDPDVFSSYPL
SEQ ID NO:37   (237) TWELQHMLGHHPYTNVLDGVEEERKERGEDVALEEKDQESDPDVFSSFPL
SEQ ID NO:38   (212) TWEFQHALGHHPYTNLIEEENGLQKVSGKKMDTKLADQESDPDVFSTYPM
SEQ ID NO:42   (200) LWLQEHVVMHHLHTNDVD---------------------ADPDQKAHG-V

                          351                                                400
SEQ ID NO:13   (296) MRMHPDTAWQPHHRFQHLFAFPLFALMTISKVLTSDFAVCLSMKKGSIDC
SEQ ID NO:37   (287) MRMHPHHTTSWYHKYQHLYAPPLFALMTLAKVFQQDFEVATSGRLYHIDA
SEQ ID NO:38   (262) MRLHPWHQKRWYHRFQHIYGPFIFGFMTINKVVTQDVGVVFRKRLFQIDA
SEQ ID NO:42   (228) LRLKPTDGWMPWHALQQLYILPGEAMYAFKLLFLDALELLAWRWEG--EK

                          401                                                450
SEQ ID NO:13   (346) SSRLVPLEGQLLFWGAKLANFLLQIVLPCYLHGTAMGLALFSVAHLVSGE
SEQ ID NO:37   (337) NVRYGSVWNVMRFWAMKVITMGYMMGLPIYFHGVLRGVGLFVIGHLACGE
SEQ ID NO:38   (312) ECRYASPMYVARFWIMKALTVLYMVALPCYMQGPWHGLKLFAIAHFTCGE
SEQ ID NO:42   (276) -ISPLARALFAPAVACKLGFWARFVALPLWLQPTVHTALCICATVCTGSF
```

FIG. 2B

```
                        451                                                 500
SEQ ID NO:13   (396) YLAICFIINHISESCEFMNTSFQTAAR------------R---TEMLQAA
SEQ ID NO:37   (387) LLATMFIVNHVIEGVSYGTKDLVGGASHGDEKKIVKPTTVLGDTPMEKTR
SEQ ID NO:38   (362) VLATMFIVNHVIEGVSYASKDAVKGT-------MAPPKTMHGVTPMNNTR
SEQ ID NO:42   (325) YLAFFFFISHNFDGVGSVGPKGSLPR------------------------

                        501                                                 550
SEQ ID NO:13   (431) HQAAEAKK----------VKPTPPPNDWAVTQVQCCVNWRSGGVLANHLS
SEQ ID NO:37   (437) EEALKSNSNNNKKKGEKNSVPSVPFNDWAAVQCQTSVNWSPGSWFWNHFS
SEQ ID NO:38   (405) KEVEAEAS----KSG--AVVKSVPLDDWAAVQCQTSANWSVGSWFWNHFS
SEQ ID NO:42   (351) -------------------------SATFVQRQVETSSNVGGYWLGVLN

                        551                                                 600
SEQ ID NO:13   (471) GGLNHQIEHHLFPSISHANYPTIAPVVKEVCEEYGLPYKNYVTFWDAVCG
SEQ ID NO:37   (487) GGLSHQIEHHLFPSICHTNYCHIQDVVESTCAEYGVPYQSESNLFVAYGK
SEQ ID NO:38   (449) GGLNHQIEHHLFPGLSHETYYHIQDVVQSTCAEYGVPYQHEPSLWTAYWK
SEQ ID NO:42   (375) GGLNFQIEHHLFPRLHHSYYAQIAPVVRTHIEKLGFKYRHFPTVGSNLSS

                        601          621
SEQ ID NO:13   (521) MVQHLRLMGAPPVPTNGDKKS
SEQ ID NO:37   (537) MISHLKFLGKAKCE-------
SEQ ID NO:38   (499) MLEHLRRLGNEETHESWQRAA
SEQ ID NO:42   (425) MLQHMGKMGTRPGAEKGGKAE
```

FIG. 2C

```
                                    1                                                  50
 E1594D4 (SEQ ID NO:1)        (1)  ATG---CAGTCAACCAAGGCGGCCGACACCGCCGCTACCGACAAGAGTCT
E1594D4S (SEQ ID NO:3)        (1)  ATGGCTCAGTCCACCAAGGCTGCCGACACTGCTGCCACCGACAAGTCTCT

                                   51                                                 100
 E1594D4 (SEQ ID NO:1)       (48)  CGACAAGAACCGCCTCATCTCTCGGGATGAGCTTCGTTCTCACAATGTCC
E1594D4S (SEQ ID NO:3)       (51)  CGACAAGAACCGACTCATCTCCCGAGACGAGCTGCGGTCTCACAACGTTC

                                  101                                                 150
 E1594D4 (SEQ ID NO:1)       (98)  CCCAGGATGCGTGGGCTGCTGTCCACGGGAGGGTCATCAACATCACGGAG
E1594D4S (SEQ ID NO:3)      (101)  CCCAGGATGCCTGGGCTGCCGTCCACGGCAGAGTCATCAACATTACCGAG

                                  151                                                 200
 E1594D4 (SEQ ID NO:1)      (148)  TTCGCCCGACGTCATCCTGGCGGCGACATCATCCTCCTTGCCGCAGGGAA
E1594D4S (SEQ ID NO:3)      (151)  TTCGCCCGACGGCATCCTGGTGGCGACATCATTCTGCTTGCCGCAGGAAA

                                  201                                                 250
 E1594D4 (SEQ ID NO:1)      (198)  GGATGCCACAGTCCTCTTCGAGACCTACCATCCCCGCGGTGTCCCCACCT
E1594D4S (SEQ ID NO:3)      (201)  GGATGCCACCGTGCTCTTCGAGACTTACCATCCTCGAGGTGTTCCCACCT

                                  251                                                 300
 E1594D4 (SEQ ID NO:1)      (248)  CCATCCTCGACAAGCTCCAGGTGGGAAAGATGAAGGACGGGGAGCTGCCC
E1594D4S (SEQ ID NO:3)      (251)  CGATCCTCGACAAGCTGCAGGTCGGCAAGATGAAGGACGGAGAACTTCCC

                                  301                                                 350
 E1594D4 (SEQ ID NO:1)      (298)  TCCTCCTTCTACTCGTGGGATTCTGACTTTTACAAGACCCTGCGCGCCCG
E1594D4S (SEQ ID NO:3)      (301)  TCCTCGTTCTACTCGTGGGATTCCGACTTTTACAAGACCCTGCGAGCTCG
```

FIG. 3A

```
                                    351                                                 400
 E1594D4 (SEQ ID NO:1)    (348) CGTTGTTGAGAGGTTGGACAAGCTCAACCTGCCGCGAAGGGGAGGGTATG
E1594D4S (SEQ ID NO:3)    (351) AGTGGTCGAGCGATTGGACAAGCTCAACCTGCCTCGAAGAGGTGGCTACG

                                    401                                                 450
 E1594D4 (SEQ ID NO:1)    (398) AGATCTGGGTCAAGGCAGTATTCCTCCTGGCAGGATTCTGGTTCAGCCTC
E1594D4S (SEQ ID NO:3)    (401) AGATTTGGGTCAAGGCAGTATTCCTCCTGGCTGGATTCTGGTTCAGCCTC

                                    451                                                 500
 E1594D4 (SEQ ID NO:1)    (448) TACAAGATGTCTGTGAACGAGACCTACTGGGCCGCATCGCTCTGGTCCGT
E1594D4S (SEQ ID NO:3)    (451) TACAAGATGTCCGTCAACGAGACCTACTGGGCTGCCTCGCTGTGGTCCGT

                                    501                                                 550
 E1594D4 (SEQ ID NO:1)    (498) GTCCATGGGAGTGTTCGCCGCCTTCATCGGCACTTGCATCCAGCACGATG
E1594D4S (SEQ ID NO:3)    (501) GTCTATGGGAGTCTTTGCTGCCTTCATCGGCACTTGCATTCAACACGATG

                                    551                                                 600
 E1594D4 (SEQ ID NO:1)    (548) GAAACCATGGCGCCTTCTCGACCAGCCCGGCTCTGAACAAGGTGGCGGGC
E1594D4S (SEQ ID NO:3)    (551) GAAACCACGGTGCCTTCTCGACCAGCCCTGCTCTCAACAAGGTTGCAGGC

                                    601                                                 650
 E1594D4 (SEQ ID NO:1)    (598) TGGACTCTGGACATGATTGGGGCGTCAGGTTTCACGTGGGAAATCCAACA
E1594D4S (SEQ ID NO:3)    (601) TGGACTCTGGACATGATCGGTGCTTCTGGCTTTACATGGGAGATTCAGCA

                                    651                                                 700
 E1594D4 (SEQ ID NO:1)    (648) TATGCTCGGCCATCATCCCTACACCAACGTTCTTGACGTGGACGAAGAAA
E1594D4S (SEQ ID NO:3)    (651) TATGCTCGGACACCATCCCTACACCAACGTCCTGGACGTGGACGAAGAGA
```

FIG. 3B

```
                                    701                                 750
 E1594D4  (SEQ ID NO:1)    (698) AGAGGAAGGAAGCTGGCGACGACTGCCCGATGGAAGACAAGGACCAGGAG
E1594D4S  (SEQ ID NO:3)    (701) AGCGAAAGGAAGCTGGCGACGATTGTCCTATGGAGGACAAGGATCAGGAG

                                    751                                 800
 E1594D4  (SEQ ID NO:1)    (748) TCCGACCCAGATGTCTTCTCCTCCTTCCCTCTCATGCGCATGCACCCATA
E1594D4S  (SEQ ID NO:3)    (751) TCCGACCCAGATGTCTTCTCTTCGTTTCCTCTCATGCGAATGCACCCCTA

                                    801                                 850
 E1594D4  (SEQ ID NO:1)    (798) CCACAAGGCTGAGTGGTACCATCGCTATCAGCACCTGTACGCGCCCGTTC
E1594D4S  (SEQ ID NO:3)    (801) CCACAAGGCCGAGTGGTACCACCGATATCAGCACCTGTACGCACCCGTTC

                                    851                                 900
 E1594D4  (SEQ ID NO:1)    (848) TCTTCGCGTTCATGACGCTAGCCAAGGTGTTCCAGCAGGATATCGAGGTC
E1594D4S  (SEQ ID NO:3)    (851) TCTTTGCTTTCATGACTCTTGCCAAGGTGTTCCAACAGGACATCGAAGTC

                                    901                                 950
 E1594D4  (SEQ ID NO:1)    (898) GCCACCACCCAGAGATTGTACCATATCGATGCCAAGTGCCGATACAATTC
E1594D4S  (SEQ ID NO:3)    (901) GCTACCACTCAGCGACTGTACCACATCGACGCCAAGTGCCGATACAATTC

                                    951                                1000
 E1594D4  (SEQ ID NO:1)    (948) TATTCTGAATGTCTTGCGCTTTTGGTCGATGAAGGTGCTTTCGATCGGAT
E1594D4S  (SEQ ID NO:3)    (951) CATTCTCAATGTCCTTCGGTTTTGGTCGATGAAGGTGCTCTCCATCGGCT

                                   1001                                1050
 E1594D4  (SEQ ID NO:1)    (998) ATATGCTGGCTGTGCCCTGCTACTTCCACGGCATTCTTGGTGGCCTTGGC
E1594D4S  (SEQ ID NO:3)   (1001) ACATGCTGGCTGTTCCCTGCTACTTCCACGGAATCCTTGGTGGCCTTGGA
```

FIG. 3C

```
                                        1051                                                1100
 E1594D4 (SEQ ID NO:1)   (1048) CTTTTCCTTATCGGCCACTTTGCCTGCGGTGAGCTTCTGGCGACCATGTT
E1594D4S (SEQ ID NO:3)   (1051) CTGTTTCTCATCGGCCACTTTGCCTGTGGAGAGCTTCTGGCAACCATGTT

                                        1101                                                1150
 E1594D4 (SEQ ID NO:1)   (1098) CATTGTCAATCACGTCATTGAGGGAGTCTCCTTTGGCAAGAAGGGTGAAT
E1594D4S (SEQ ID NO:3)   (1101) CATTGTCAATCACGTCATCGAGGGTGTGTCCTTTGGCAAAAAGGGAGAAT

                                        1151                                                1200
 E1594D4 (SEQ ID NO:1)   (1148) CGCTGGGACTTTCCAAGGACGTGGAGTTCAAGCCCACCACCGTTTCGGGC
E1594D4S (SEQ ID NO:3)   (1151) CTCTCGGTCTGTCCAAGGACGTGGAGTTCAAGCCTACAACCGTTTCTGGA

                                        1201                                                1250
 E1594D4 (SEQ ID NO:1)   (1198) CGCACGCCCATGGAACAGACCCGTGCCGAAGCCAAGAAGGCGGCCAACGG
E1594D4S (SEQ ID NO:3)   (1201) CGAACTCCAATGGAGCAGACCCGTGCCGAGGCCAAAAAGGCTGCCAATGG

                                        1251                                                1300
 E1594D4 (SEQ ID NO:1)   (1248) CGGAAACGTGAAGGATGTCCCCTACAACGACTGGGCGGCCGTTCAATGCC
E1594D4S (SEQ ID NO:3)   (1251) AGGCAACGTCAAGGATGTTCCCTACAACGACTGGGCTGCCGTTCAGTGTC

                                        1301                                                1350
 E1594D4 (SEQ ID NO:1)   (1298) AAACGAGTGTGAACTGGAGTCCTGGATCGTGGTTCTGGAATCACTTCAGC
E1594D4S (SEQ ID NO:3)   (1301) AAACGAGCGTCAACTGGTCTCCTGGATCGTGGTTCTGGAATCACTTCTCC

                                        1351                                                1400
 E1594D4 (SEQ ID NO:1)   (1348) GGCGGTCTATCGCATCAGATTGAGCACCATCTTTTCCCTAGCATTTGCCA
E1594D4S (SEQ ID NO:3)   (1351) GGTGGCCTCTCCCACCAGATCGAGCACCATCTGTTTCCCAGCATTTGTCA
```

FIG. 3D

```
                                    1401                                                1450
 E1594D4 (SEQ ID NO:1)   (1398) CACCAATTACGCTCATATCCAAGACGTTGTCCAAAAGACTTGCGAGGAGT
E1594D4S (SEQ ID NO:3)   (1401) CACCAACTACGCTCACATCCAGGACGTTGTCCAGAAGACTTGCGAAGAGT

                                    1451                                                1500
 E1594D4 (SEQ ID NO:1)   (1448) ASGGCGTTCCTTACCAAAGCGAGCCCTCTTTGTAYTCCGCCTATGGCAAG
E1594D4S (SEQ ID NO:3)   (1451) ACGGTGTTCCTTACCAGTCCGAACCCTCTTTGTTCTCCGCCTATGGCAAG

                                    1501                                        1548
 E1594D4 (SEQ ID NO:1)   (1498) ATGTTGAGCCATCTCAAGTACCTCGGAAACGAGAAGAAGGTGGCTTAG
E1594D4S (SEQ ID NO:3)   (1501) ATGCTGTCTCATCTCAAGTACCTCGGAAACGAGAAAAAGGTCGCTTAA
```

FIG. 3E

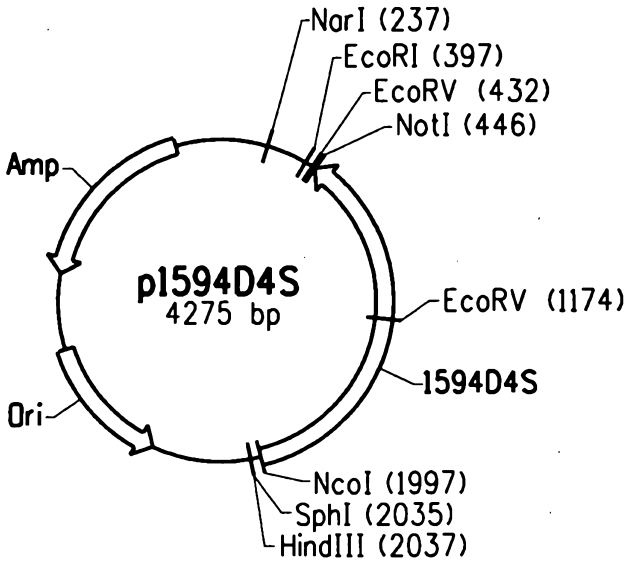

NarI (237)
EcoRI (397)
EcoRV (432)
NotI (446)
Amp
p1594D4S
4275 bp
EcoRV (1174)
1594D4S
Ori
NcoI (1997)
SphI (2035)
HindIII (2037)

FIG. 4A

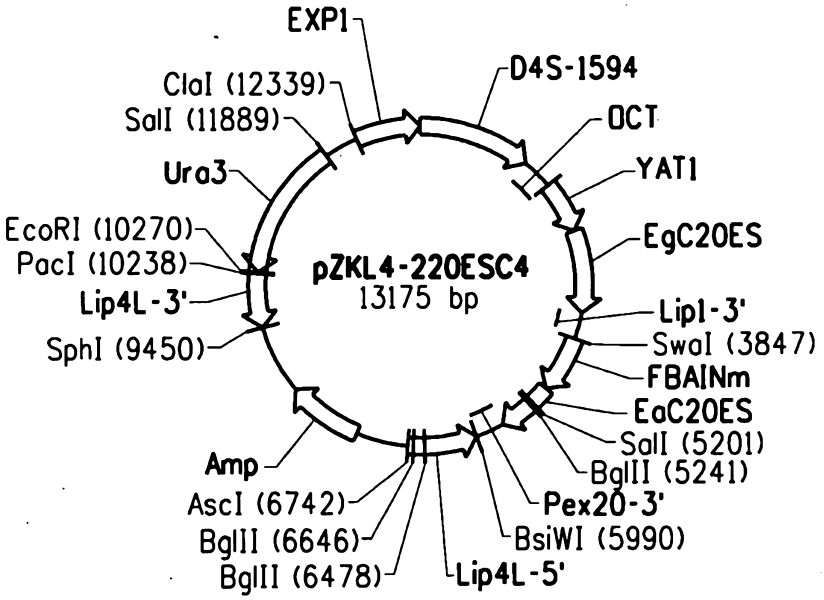

EXP1
D4S-1594
ClaI (12339)
SalI (11889)
OCT
Ura3
YAT1
EcoRI (10270)
PacI (10238)
pZKL4-220ESC4
13175 bp
EgC20ES
Lip4L-3'
Lip1-3'
SphI (9450)
SwaI (3847)
FBAINm
EaC20ES
SalI (5201)
BglII (5241)
Amp
Pex20-3'
AscI (6742)
BglII (6646)
BsiWI (5990)
BglII (6478)
Lip4L-5'

FIG. 4B

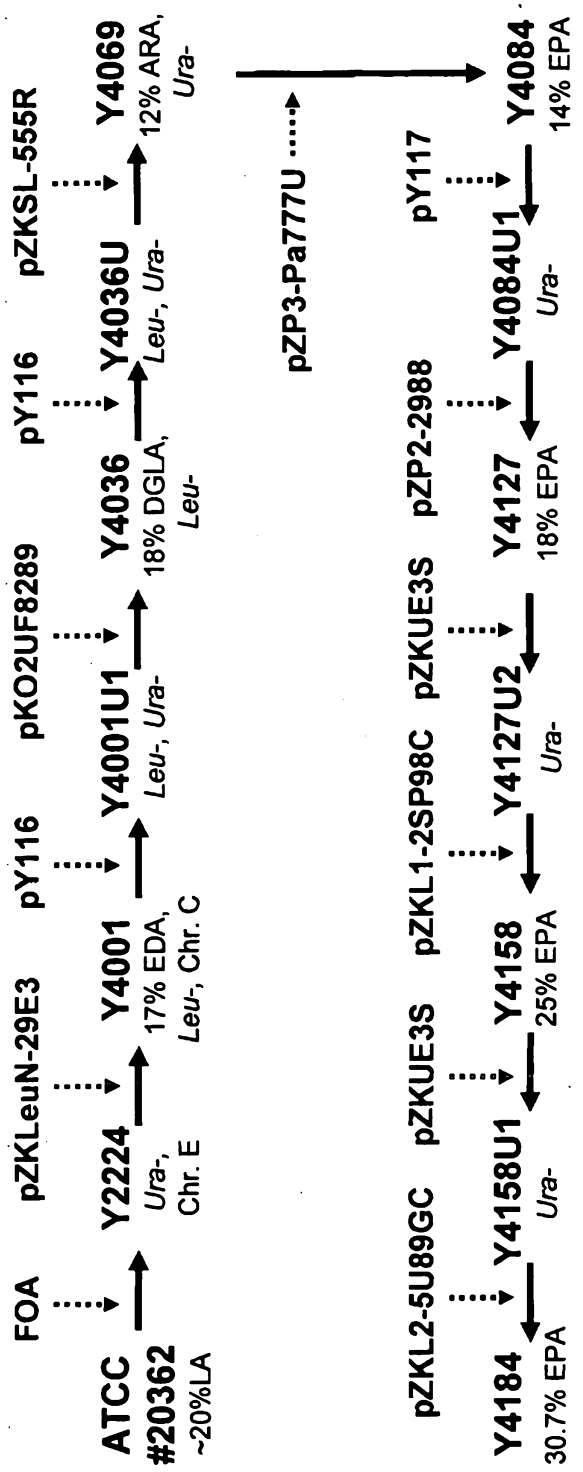

FOA
ATCC #20362 ~20%LA
pZKLeuN-29E3
Y2224 Ura-, Chr. E
pY116
Y4001 17% EDA, Leu-, Chr. C
pKO2UF8289
Y4001U1 Leu-, Ura-
pY116
Y4036 18% DGLA, Leu-
pZKSL-555R
Y4036U Leu-, Ura-
Y4069 12% ARA, Ura-
pZP3-Pa777U
pZKL2-5U89GC
Y4184 30.7% EPA
pZKUE3S
Y4158U1 Ura-
pZKL1-2SP98C
Y4158 25% EPA
pZKUE3S
Y4127U2 Ura-
pZP2-2988
Y4127 18% EPA
pY117
Y4084U1 Ura-
Y4084 14% EPA
pZKUE3S
Y4184U Ura-

FIG. 5

SEQUENCE LISTING

<110> E. I. du Pont de Nemours & Co., Inc.
Zhu, Quinn
Xue, Zhixiong

<120> DELTA-4 DESATURASE AND ITS USE IN MAKING POLYUNSATURATED FATTY ACIDS

<130> CL3798JPPCT

<140> JP 2011-503123
<141> 2009-04-01

<150> US 61/041,716
<151> 2008-04-02

<160> 47

<170> PatentIn version 3.5

<210> 1
<211> 1545
<212> DNA
<213> Eutreptiella cf_gymnastica CCMP1594

<220>
<221> misc_feature
<222> (1)..(1545)
<223> delta-4 desaturase ("E1594D4")

<400> 1
atgcagtcaa ccaaggcggc cgacaccgcc gctaccgaca agagtctcga caagaaccgc 60

ctcatctctc gggatgagct tcgttctcac aatgtccccc aggatgcgtg ggctgctgtc 120

cacgggaggg tcatcaacat cacggagttc gcccgacgtc atcctggcgg cgacatcatc 180

ctccttgccg cagggaagga tgccacagtc ctcttcgaga cctaccatcc ccgcggtgtc 240

cccacctcca tcctcgacaa gctccaggtg ggaaagatga aggacgggga gctgccctcc 300

tccttctact cgtgggattc tgacttttac aagaccctgc gcgcccgcgt tgttgagagg 360

ttggacaagc tcaacctgcc gcgaagggga gggtatgaga tctgggtcaa ggcagtattc 420

ctcctggcag gattctggtt cagcctctac aagatgtctg tgaacgagac ctactgggcc 480

gcatcgctct ggtccgtgtc catgggagtg ttcgccgcct tcatcggcac ttgcatccag 540

cacgatggaa accatggcgc cttctcgacc agcccggctc tgaacaaggt ggcgggctgg 600

actctggaca tgattggggc gtcaggtttc acgtgggaaa tccaacatat gctcggccat 660

catccctaca ccaacgttct tgacgtggac gaagaaaaga ggaaggaagc tggcgacgac 720

tgcccgatgg aagacaagga ccaggagtcc gacccagatg tcttctcctc cttccctctc 780

atgcgcatgc acccatacca caaggctgag tggtaccatc gctatcagca cctgtacgcg 840

cccgttctct tcgcgttcat gacgctagcc aaggtgttcc agcaggatat cgaggtcgcc 900

accacccaga gattgtacca tatcgatgcc aagtgccgat acaattctat tctgaatgtc 960

ttgcgctttt ggtcgatgaa ggtgctttcg atcggatata tgctggctgt gccctgctac 1020

ttccacggca ttcttggtgg ccttggcctt ttccttatcg gccactttgc ctgcggtgag 1080

cttctggcga ccatgttcat tgtcaatcac gtcattgagg gagtctcctt tggcaagaag 1140

ggtgaatcgc tgggactttc caaggacgtg gagttcaagc ccaccaccgt ttcgggccgc 1200

acgcccatgg aacagacccg tgccgaagcc aagaaggcgg ccaacggcgg aaacgtgaag 1260

gatgtcccct acaacgactg ggcggccgtt caatgccaaa cgagtgtgaa ctggagtcct 1320

ggatcgtggt tctggaatca cttcagcggc ggtctatcgc atcagattga gcaccatctt 1380

ttccctagca tttgccacac caattacgct catatccaag acgttgtcca aaagacttgc 1440

gaggagtasg gcgttcctta ccaaagcgag ccctctttgt aytccgccta tggcaagatg 1500

ttgagccatc tcaagtacct cggaaacgag aagaaggtgg cttag 1545

<210> 2
<211> 514
<212> PRT
<213> Eutreptiella cf_gymnastica CCMP1594

<220>
<221> MISC_FEATURE
<222> (1)..(514)
<223> delta-4 desaturase ("E1594D4")

<220>
<221> MISC_FEATURE
<222> (483)..(483)
<223> Xaa = any amino acid

<400> 2

Met Gln Ser Thr Lys Ala Ala Asp Thr Ala Ala Thr Asp Lys Ser Leu
1 5 10 15

Asp Lys Asn Arg Leu Ile Ser Arg Asp Glu Leu Arg Ser His Asn Val
20 25 30

Pro Gln Asp Ala Trp Ala Ala Val His Gly Arg Val Ile Asn Ile Thr
35 40 45

Glu Phe Ala Arg Arg His Pro Gly Gly Asp Ile Ile Leu Leu Ala Ala
50 55 60

Gly Lys Asp Ala Thr Val Leu Phe Glu Thr Tyr His Pro Arg Gly Val
65 70 75 80

Pro Thr Ser Ile Leu Asp Lys Leu Gln Val Gly Lys Met Lys Asp Gly
85 90 95

Glu Leu Pro Ser Ser Phe Tyr Ser Trp Asp Ser Asp Phe Tyr Lys Thr
100 105 110

```
Leu Arg Ala Arg Val Val Glu Arg Leu Asp Lys Leu Asn Leu Pro Arg
        115                 120                 125

Arg Gly Gly Tyr Glu Ile Trp Val Lys Ala Val Phe Leu Leu Ala Gly
    130                 135                 140

Phe Trp Phe Ser Leu Tyr Lys Met Ser Val Asn Glu Thr Tyr Trp Ala
145                 150                 155                 160

Ala Ser Leu Trp Ser Val Ser Met Gly Val Phe Ala Ala Phe Ile Gly
                165                 170                 175

Thr Cys Ile Gln His Asp Gly Asn His Gly Ala Phe Ser Thr Ser Pro
            180                 185                 190

Ala Leu Asn Lys Val Ala Gly Trp Thr Leu Asp Met Ile Gly Ala Ser
        195                 200                 205

Gly Phe Thr Trp Glu Ile Gln His Met Leu Gly His His Pro Tyr Thr
    210                 215                 220

Asn Val Leu Asp Val Asp Glu Glu Lys Arg Lys Glu Ala Gly Asp Asp
225                 230                 235                 240

Cys Pro Met Glu Asp Lys Asp Gln Glu Ser Asp Pro Asp Val Phe Ser
                245                 250                 255

Ser Phe Pro Leu Met Arg Met His Pro Tyr His Lys Ala Glu Trp Tyr
            260                 265                 270

His Arg Tyr Gln His Leu Tyr Ala Pro Val Leu Phe Ala Phe Met Thr
        275                 280                 285

Leu Ala Lys Val Phe Gln Gln Asp Ile Glu Val Ala Thr Thr Gln Arg
    290                 295                 300

Leu Tyr His Ile Asp Ala Lys Cys Arg Tyr Asn Ser Ile Leu Asn Val
305                 310                 315                 320

Leu Arg Phe Trp Ser Met Lys Val Leu Ser Ile Gly Tyr Met Leu Ala
                325                 330                 335

Val Pro Cys Tyr Phe His Gly Ile Leu Gly Gly Leu Gly Leu Phe Leu
            340                 345                 350

Ile Gly His Phe Ala Cys Gly Glu Leu Leu Ala Thr Met Phe Ile Val
        355                 360                 365

Asn His Val Ile Glu Gly Val Ser Phe Gly Lys Lys Gly Glu Ser Leu
    370                 375                 380

Gly Leu Ser Lys Asp Val Glu Phe Lys Pro Thr Thr Val Ser Gly Arg
385                 390                 395                 400

Thr Pro Met Glu Gln Thr Arg Ala Glu Ala Lys Lys Ala Ala Asn Gly
                405                 410                 415

Gly Asn Val Lys Asp Val Pro Tyr Asn Asp Trp Ala Ala Val Gln Cys
            420                 425                 430

Gln Thr Ser Val Asn Trp Ser Pro Gly Ser Trp Phe Trp Asn His Phe
        435                 440                 445
```

```
Ser Gly Gly Leu Ser His Gln Ile Glu His His Leu Phe Pro Ser Ile
    450                 455                 460

Cys His Thr Asn Tyr Ala His Ile Gln Asp Val Val Gln Lys Thr Cys
465                 470                 475                 480

Glu Glu Xaa Gly Val Pro Tyr Gln Ser Glu Pro Ser Leu Tyr Ser Ala
                485                 490                 495

Tyr Gly Lys Met Leu Ser His Leu Lys Tyr Leu Gly Asn Glu Lys Lys
            500                 505                 510

Val Ala
```

<210> 3
<211> 1548
<212> DNA
<213> Eutreptiella cf_gymnastica CCMP1594

<220>
<221> CDS
<222> (1)..(1548)
<223> synthetic delta-4 desaturase (codon-optimized for Yarrowia lipolytica) ("E1594D4S")

<400> 3

```
atg gct cag tcc acc aag gct gcc gac act gct gcc acc gac aag tct          48
Met Ala Gln Ser Thr Lys Ala Ala Asp Thr Ala Ala Thr Asp Lys Ser
1               5                   10                  15

ctc gac aag aac cga ctc atc tcc cga gac gag ctg cgg tct cac aac          96
Leu Asp Lys Asn Arg Leu Ile Ser Arg Asp Glu Leu Arg Ser His Asn
            20                  25                  30

gtt ccc cag gat gcc tgg gct gcc gtc cac ggc aga gtc atc aac att         144
Val Pro Gln Asp Ala Trp Ala Ala Val His Gly Arg Val Ile Asn Ile
        35                  40                  45

acc gag ttc gcc cga cgg cat cct ggt ggc gac atc att ctg ctt gcc         192
Thr Glu Phe Ala Arg Arg His Pro Gly Gly Asp Ile Ile Leu Leu Ala
    50                  55                  60

gca gga aag gat gcc acc gtg ctc ttc gag act tac cat cct cga ggt         240
Ala Gly Lys Asp Ala Thr Val Leu Phe Glu Thr Tyr His Pro Arg Gly
65                  70                  75                  80

gtt ccc acc tcg atc ctc gac aag ctg cag gtc ggc aag atg aag gac         288
Val Pro Thr Ser Ile Leu Asp Lys Leu Gln Val Gly Lys Met Lys Asp
                85                  90                  95

gga gaa ctt ccc tcc tcg ttc tac tcg tgg gat tcc gac ttt tac aag         336
Gly Glu Leu Pro Ser Ser Phe Tyr Ser Trp Asp Ser Asp Phe Tyr Lys
            100                 105                 110

acc ctg cga gct cga gtg gtc gag cga ttg gac aag ctc aac ctg cct         384
Thr Leu Arg Ala Arg Val Val Glu Arg Leu Asp Lys Leu Asn Leu Pro
        115                 120                 125

cga aga ggt ggc tac gag att tgg gtc aag gca gta ttc ctc ctg gct         432
```

```
Arg Arg Gly Gly Tyr Glu Ile Trp Val Lys Ala Val Phe Leu Leu Ala
    130                 135                 140

gga ttc tgg ttc agc ctc tac aag atg tcc gtc aac gag acc tac tgg  480
Gly Phe Trp Phe Ser Leu Tyr Lys Met Ser Val Asn Glu Thr Tyr Trp
145                 150                 155                 160

gct gcc tcg ctg tgg tcc gtg tct atg gga gtc ttt gct gcc ttc atc  528
Ala Ala Ser Leu Trp Ser Val Ser Met Gly Val Phe Ala Ala Phe Ile
                165                 170                 175

ggc act tgc att caa cac gat gga aac cac ggt gcc ttc tcg acc agc  576
Gly Thr Cys Ile Gln His Asp Gly Asn His Gly Ala Phe Ser Thr Ser
            180                 185                 190

cct gct ctc aac aag gtt gca ggc tgg act ctg gac atg atc ggt gct  624
Pro Ala Leu Asn Lys Val Ala Gly Trp Thr Leu Asp Met Ile Gly Ala
        195                 200                 205

tct ggc ttt aca tgg gag att cag cat atg ctc gga cac cat ccc tac  672
Ser Gly Phe Thr Trp Glu Ile Gln His Met Leu Gly His His Pro Tyr
    210                 215                 220

acc aac gtc ctg gac gtg gac gaa gag aag cga aag gaa gct ggc gac  720
Thr Asn Val Leu Asp Val Asp Glu Glu Lys Arg Lys Glu Ala Gly Asp
225                 230                 235                 240

gat tgt cct atg gag gac aag gat cag gag tcc gac cca gat gtc ttc  768
Asp Cys Pro Met Glu Asp Lys Asp Gln Glu Ser Asp Pro Asp Val Phe
                245                 250                 255

tct tcg ttt cct ctc atg cga atg cac ccc tac cac aag gcc gag tgg  816
Ser Ser Phe Pro Leu Met Arg Met His Pro Tyr His Lys Ala Glu Trp
            260                 265                 270

tac cac cga tat cag cac ctg tac gca ccc gtt ctc ttt gct ttc atg  864
Tyr His Arg Tyr Gln His Leu Tyr Ala Pro Val Leu Phe Ala Phe Met
        275                 280                 285

act ctt gcc aag gtg ttc caa cag gac atc gaa gtc gct acc act cag  912
Thr Leu Ala Lys Val Phe Gln Gln Asp Ile Glu Val Ala Thr Thr Gln
    290                 295                 300

cga ctg tac cac atc gac gcc aag tgc cga tac aat tcc att ctc aat  960
Arg Leu Tyr His Ile Asp Ala Lys Cys Arg Tyr Asn Ser Ile Leu Asn
305                 310                 315                 320

gtc ctt cgg ttt tgg tcg atg aag gtg ctc tcc atc ggc tac atg ctg 1008
Val Leu Arg Phe Trp Ser Met Lys Val Leu Ser Ile Gly Tyr Met Leu
                325                 330                 335

gct gtt ccc tgc tac ttc cac gga atc ctt ggt ggc ctt gga ctg ttt 1056
Ala Val Pro Cys Tyr Phe His Gly Ile Leu Gly Gly Leu Gly Leu Phe
            340                 345                 350

ctc atc ggc cac ttt gcc tgt gga gag ctt ctg gca acc atg ttc att 1104
Leu Ile Gly His Phe Ala Cys Gly Glu Leu Leu Ala Thr Met Phe Ile
        355                 360                 365

gtc aat cac gtc atc gag ggt gtg tcc ttt ggc aaa aag gga gaa tct 1152
Val Asn His Val Ile Glu Gly Val Ser Phe Gly Lys Lys Gly Glu Ser
    370                 375                 380
```

```
ctc ggt ctg tcc aag gac gtg gag ttc aag cct aca acc gtt tct gga       1200
Leu Gly Leu Ser Lys Asp Val Glu Phe Lys Pro Thr Thr Val Ser Gly
385                 390                 395                 400

cga act cca atg gag cag acc cgt gcc gag gcc aaa aag gct gcc aat       1248
Arg Thr Pro Met Glu Gln Thr Arg Ala Glu Ala Lys Lys Ala Ala Asn
                405                 410                 415

gga ggc aac gtc aag gat gtt ccc tac aac gac tgg gct gcc gtt cag       1296
Gly Gly Asn Val Lys Asp Val Pro Tyr Asn Asp Trp Ala Ala Val Gln
            420                 425                 430

tgt caa acg agc gtc aac tgg tct cct gga tcg tgg ttc tgg aat cac       1344
Cys Gln Thr Ser Val Asn Trp Ser Pro Gly Ser Trp Phe Trp Asn His
        435                 440                 445

ttc tcc ggt ggc ctc tcc cac cag atc gag cac cat ctg ttt ccc agc       1392
Phe Ser Gly Gly Leu Ser His Gln Ile Glu His His Leu Phe Pro Ser
    450                 455                 460

att tgt cac acc aac tac gct cac atc cag gac gtt gtc cag aag act       1440
Ile Cys His Thr Asn Tyr Ala His Ile Gln Asp Val Val Gln Lys Thr
465                 470                 475                 480

tgc gaa gag tac ggt gtt cct tac cag tcc gaa ccc tct ttg ttc tcc       1488
Cys Glu Glu Tyr Gly Val Pro Tyr Gln Ser Glu Pro Ser Leu Phe Ser
                485                 490                 495

gcc tat ggc aag atg ctg tct cat ctc aag tac ctc gga aac gag aaa       1536
Ala Tyr Gly Lys Met Leu Ser His Leu Lys Tyr Leu Gly Asn Glu Lys
            500                 505                 510

aag gtc gct taa                                                      1548
Lys Val Ala
        515
```

<210> 4
<211> 515
<212> PRT
<213> Eutreptiella cf_gymnastica CCMP1594

<400> 4

```
Met Ala Gln Ser Thr Lys Ala Ala Asp Thr Ala Ala Thr Asp Lys Ser
1               5                   10                  15

Leu Asp Lys Asn Arg Leu Ile Ser Arg Asp Glu Leu Arg Ser His Asn
            20                  25                  30

Val Pro Gln Asp Ala Trp Ala Ala Val His Gly Arg Val Ile Asn Ile
        35                  40                  45

Thr Glu Phe Ala Arg Arg His Pro Gly Gly Asp Ile Ile Leu Leu Ala
    50                  55                  60

Ala Gly Lys Asp Ala Thr Val Leu Phe Glu Thr Tyr His Pro Arg Gly
65                  70                  75                  80

Val Pro Thr Ser Ile Leu Asp Lys Leu Gln Val Gly Lys Met Lys Asp
                85                  90                  95
```

```
Gly Glu Leu Pro Ser Ser Phe Tyr Ser Trp Asp Ser Asp Phe Tyr Lys
            100                 105                 110

Thr Leu Arg Ala Arg Val Val Glu Arg Leu Asp Lys Leu Asn Leu Pro
        115                 120                 125

Arg Arg Gly Gly Tyr Glu Ile Trp Val Lys Ala Val Phe Leu Leu Ala
    130                 135                 140

Gly Phe Trp Phe Ser Leu Tyr Lys Met Ser Val Asn Glu Thr Tyr Trp
145                 150                 155                 160

Ala Ala Ser Leu Trp Ser Val Ser Met Gly Val Phe Ala Ala Phe Ile
                165                 170                 175

Gly Thr Cys Ile Gln His Asp Gly Asn His Gly Ala Phe Ser Thr Ser
            180                 185                 190

Pro Ala Leu Asn Lys Val Ala Gly Trp Thr Leu Asp Met Ile Gly Ala
        195                 200                 205

Ser Gly Phe Thr Trp Glu Ile Gln His Met Leu Gly His His Pro Tyr
    210                 215                 220

Thr Asn Val Leu Asp Val Asp Glu Glu Lys Arg Lys Glu Ala Gly Asp
225                 230                 235                 240

Asp Cys Pro Met Glu Asp Lys Asp Gln Glu Ser Asp Pro Asp Val Phe
                245                 250                 255

Ser Ser Phe Pro Leu Met Arg Met His Pro Tyr His Lys Ala Glu Trp
            260                 265                 270

Tyr His Arg Tyr Gln His Leu Tyr Ala Pro Val Leu Phe Ala Phe Met
        275                 280                 285

Thr Leu Ala Lys Val Phe Gln Gln Asp Ile Glu Val Ala Thr Thr Gln
    290                 295                 300

Arg Leu Tyr His Ile Asp Ala Lys Cys Arg Tyr Asn Ser Ile Leu Asn
305                 310                 315                 320

Val Leu Arg Phe Trp Ser Met Lys Val Leu Ser Ile Gly Tyr Met Leu
                325                 330                 335

Ala Val Pro Cys Tyr Phe His Gly Ile Leu Gly Gly Leu Gly Leu Phe
            340                 345                 350

Leu Ile Gly His Phe Ala Cys Gly Glu Leu Leu Ala Thr Met Phe Ile
        355                 360                 365

Val Asn His Val Ile Glu Gly Val Ser Phe Gly Lys Lys Gly Glu Ser
    370                 375                 380

Leu Gly Leu Ser Lys Asp Val Glu Phe Lys Pro Thr Thr Val Ser Gly
385                 390                 395                 400

Arg Thr Pro Met Glu Gln Thr Arg Ala Glu Ala Lys Lys Ala Ala Asn
                405                 410                 415

Gly Gly Asn Val Lys Asp Val Pro Tyr Asn Asp Trp Ala Ala Val Gln
```

```
            420                 425                 430

Cys Gln Thr Ser Val Asn Trp Ser Pro Gly Ser Trp Phe Trp Asn His
        435                 440                 445

Phe Ser Gly Gly Leu Ser His Gln Ile Glu His His Leu Phe Pro Ser
    450                 455                 460

Ile Cys His Thr Asn Tyr Ala His Ile Gln Asp Val Val Gln Lys Thr
465                 470                 475                 480

Cys Glu Glu Tyr Gly Val Pro Tyr Gln Ser Glu Pro Ser Leu Phe Ser
                485                 490                 495

Ala Tyr Gly Lys Met Leu Ser His Leu Lys Tyr Leu Gly Asn Glu Lys
            500                 505                 510

Lys Val Ala
        515
```

<210> 5
<211> 847
<212> DNA
<213> Eutreptiella cf_gymnastica CCMP1594

<220>
<221> misc_feature
<222> (781)..(781)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (789)..(789)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (831)..(831)
<223> n is a, c, g, or t

<400> 5

```
attcaacacg acgggaacca tggcgccttc tcgaccagcc cggctctgaa caaggtggcg       60

ggctggactc tggacatgat tggggcgtca ggtttcacgt gggaaatcca acatatgctc      120

ggccatcatc cctacaccaa cgttcttgac gtggacgaag aaaagaggaa ggaagctggc      180

gacgactgcc cgatggaaga caaggaccag gagtccgacc cagatgtctt ctcctccttc      240

cctctcatgc gcatgcaccc ataccacaag gctgagtggt accatcgcta tcagcacctg      300

tacgcgcccg ttctcttcgc gttcatgacg ctagccaagg tgttccagca ggatatcgag      360

gtcgccacca cccagagatt gtaccatatc gatgccaagt gccgatacaa ttctattctg      420

aatgtcttgc gcttttggtc gatgaaggtg ctttcgatcg gatatatgct ggctgtgccc      480

tgctacttcc acggcattct tggtggcctt ggccttttcc ttatcggcca ctttgcctgc      540

ggtgagcttc tggcgaccat gttcattgtc aatcacgtca ttgagggagt ctcctttggc      600
```

aagaagggtg aatcgctggg actttccaag gacgtggagt tcaagcccac caccgtttcg 660

ggccgcacgc ccatggaaca gacccgtgcc gaagccaaga aggcggccaa cggcggaaac 720

gtgaaggatg tcccctacaa cgactgggcg gccgttcaat gccaaacgag tgtgaactgg 780

ngtcctggnt cgtggttctg gaatcacttc agcggcggtc tatcgcatca nattgagcac 840

cacctgt 847

<210> 6
<211> 359
<212> DNA
<213> Eutreptiella cf_gymnastica CCMP1594

<400> 6
ctactcgtgg gattctgact tttacaagac cctgcgcgcc cgcgttgttg agaggttgga 60

caagctcaac ctgccgcgaa ggggagggta tgagatctgg gtcaaggcag tattcctcct 120

ggcaggattc tggttcagcc tctacaagat gtctgtgaac gagacctact gggccgcatc 180

gctctggtcc gtgtccatgg gagtgttcgc cgccttcatc ggcacttgca tccagcacga 240

tggaaaccat ggcgccttct cgaccagccc ggctctgaac aaggtggcgg gctggactct 300

ggacatgatt ggggcgtcag gtttcacgtg ggaaatccaa catatgctcg gccatcatc 359

<210> 7
<211> 395
<212> DNA
<213> Eutreptiella cf_gymnastica CCMP1594

<400> 7
ggaggagaga tgcagtcaac caaggcggcc gacaccgccg ctaccgacaa gagtctcgac 60

aagaaccgcc tcatctctcg ggatgagctt cgttctcaca atgtccccca ggatgcgtgg 120

gctgctgtcc acgggagggt catcaacatc acggagttcg cccgacgtca tcctggcggc 180

gacatcatcc tccttgccgc agggaaggat gccacagtcc tcttcgagac ctaccatccc 240

cgcggtgtcc ccacctccat cctcgacgag ctccaggtgg gaaagatgaa ggacggggag 300

ctgccctcct ccttctactc gtgggattct gacttttaca agaccctgcg cgcccgcgtt 360

gttgagaggt tggacaagct caacctgccg cgaag 395

<210> 8
<211> 873
<212> DNA
<213> Eutreptiella cf_gymnastica CCMP1594

<400> 8
gtttcgggcc gcacgcccat ggaacagacc cgtgccgaag ccaagaaggc ggccaacggc 60

ggaaacgtga aggatgtccc ctacaacgac tgggsggccg ttcaatgcca aacgagtgtg 120

aactggagtc ctggatcgtg gttctggaat cacttcagcg gcggtctatc gcatcagatt 180

gagcaccatc ttttccctag catttgccac accaattacg ctcatatcca agacgttgtc 240

caaaagactt gcgaggagta sggcgttcct taccaaagcg agccctcttt gtaytccgcc 300

tatggcaaga tgttgagcca tctcaagtac ctcggaaacg agaagaaggt ggcttaggca 360

ttggcgaact gaaaataaat tgctattgat ttttaaaaga ttttagcgag gaaattttcg 420

accaaataca acgcgtgttc ctyttgggcg gtcctgattc ggcacactgt gttttgcagg 480

atcatgctgc ctcacagggt ggggtcccat ctggtggttg tgtgaggtgc tgccggctgc 540

gtgctggaac acacgcatgc tgtcctatgg ttgggccggt gaggggtgac ggtcgcgaat 600

atggtggtga tggcggcggc ggtgggggcc ctggttgcgt caagcggcac aaaactacag 660

agttatacga cgatgtacac tatgcccctc tcctagggcc acccttgcct acaaggtgca 720

ttaactggct aatggtactc cagccaatga tctataccct tgcatggttg ttattgctgc 780

gccccmccgg csccgcatct ggcgttgcgc tttcctgcac cccagtgcaa cctctggcgt 840

ctctatattt ttcaaaaaaa aaaaaaaaaa aaa 873

<210> 9
<211> 2070
<212> DNA
<213> Eutreptiella cf_gymnastica CCMP1594

<400> 9
ggaggagaga tgcagtcaac caaggcggcc gacaccgccg ctaccgacaa gagtctcgac 60

aagaaccgcc tcatctctcg ggatgagctt cgttctcaca atgtccccca ggatgcgtgg 120

gctgctgtcc acgggagggt catcaacatc acggagttcg cccgacgtca tcctggcggc 180

gacatcatcc tccttgccgc agggaaggat gccacagtcc tcttcgagac ctaccatccc 240

cgcggtgtcc ccacctccat cctcgacaag ctccaggtgg gaaagatgaa ggacggggag 300

ctgccctcct ccttctactc gtgggattct gacttttaca agaccctgcg cgcccgcgtt 360

gttgagaggt tggacaagct caacctgccg cgaaggggag ggtatgagat ctgggtcaag 420

gcagtattcc tcctggcagg attctggttc agcctctaca agatgtctgt gaacgagacc 480

tactgggccg catcgctctg gtccgtgtcc atgggagtgt tcgccgcctt catcggcact 540

tgcatccagc acgatggaaa ccatggcgcc ttctcgacca gcccggctct gaacaaggtg 600

gcgggctgga ctctggacat gattggggcg tcaggtttca cgtgggaaat ccaacatatg 660

ctcggccatc atccctacac caacgttctt gacgtggacg aagaaaagag gaaggaagct 720

ggcgacgact gcccgatgga agacaaggac caggagtccg acccagatgt cttctcctcc 780

ttccctctca tgcgcatgca cccataccac aaggctgagt ggtaccatcg ctatcagcac 840

ctgtacgcgc ccgttctctt cgcgttcatg acgctagcca aggtgttcca gcaggatatc 900

gaggtcgcca ccacccagag attgtaccat atcgatgcca agtgccgata caattctatt 960

ctgaatgtct tgcgcttttg gtcgatgaag gtgctttcga tcggatatat gctggctgtg 1020

ccctgctact tccacggcat tcttggtggc cttggccttt tccttatcgg ccactttgcc 1080

tgcggtgagc ttctggcgac catgttcatt gtcaatcacg tcattgaggg agtctccttt 1140

ggcaagaagg gtgaatcgct gggactttcc aaggacgtgg agttcaagcc caccaccgtt 1200

tcgggccgca cgcccatgga acagacccgt gccgaagcca agaaggcggc caacggcgga 1260

aacgtgaagg atgtccccta caacgactgg gcggccgttc aatgccaaac gagtgtgaac 1320

tggagtcctg gatcgtggtt ctggaatcac ttcagcggcg gtctatcgca tcagattgag 1380

caccatcttt tccctagcat ttgccacacc aattacgctc atatccaaga cgttgtccaa 1440

aagacttgcg aggagtasgg cgttccttac caaagcgagc cctctttgta ytccgcctat 1500

ggcaagatgt tgagccatct caagtacctc ggaaacgaga agaaggtggc ttaggcattg 1560

gcgaactgaa aataaattgc tattgatttt taaaagattt tagcgaggaa attttcgacc 1620

aaatacaacg cgtgttccty ttgggcggtc ctgattcggc acactgtgtt ttgcaggatc 1680

atgctgcctc acagggtggg gtcccatctg gtggttgtgt gaggtgctgc cggctgcgtg 1740

ctggaacaca cgcatgctgt cctatggttg ggccggtgag gggtgacggt cgcgaatatg 1800

gtggtgatgg cggcggcggt gggggccctg gttgcgtcaa gcggcacaaa actacagagt 1860

tatacgacga tgtacactat gcccctctcc tagggccacc cttgcctaca aggtgcatta 1920

actggctaat ggtactccag ccaatgatct ataccccttgc atggttgtta ttgctgcgcc 1980

ccmccggcsc cgcatctggc gttgcgcttt cctgcacccc agtgcaacct ctggcgtctc 2040

tatatttttc aaaaaaaaaa aaaaaaaaaa 2070

<210> 10
<211> 39
<212> DNA
<213> Artificial Sequence

<220>
<223> Smart IV oligonucleotide primer

<400> 10
aagcagtggt atcaacgcag agtggccatt acggccggg 39

<210> 11
<211> 59
<212> DNA
<213> Artificial Sequence

<220>
<223> CDSIII/3'PCR primer

<220>

<221> misc_feature
<222> (28)..(57)
<223> thymidine (dT); see BD Biosciences Clontech's SMART cDNA technology

<220>
<221> misc_feature
<222> (59)..(59)
<223> n is a, c, g, or t

<400> 11
attctagagg ccgaggcggc cgacatgttt ttttttttt ttttttttt ttttttvn 59

<210> 12
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> 5' CDSIII PCR primer

<400> 12
aagcagtggt atcaacgcag agt 23

<210> 13
<211> 541
<212> PRT
<213> Euglena gracilis

<220>
<221> MISC_FEATURE
<222> (1)..(541)
<223> delta-4 desaturase; GenBank Accession No. AY278558

<400> 13

Met Leu Val Leu Phe Gly Asn Phe Tyr Val Lys Gln Tyr Ser Gln Lys
1 5 10 15

Asn Gly Lys Pro Glu Asn Gly Ala Thr Pro Glu Asn Gly Ala Lys Pro
20 25 30

Gln Pro Cys Glu Asn Gly Thr Val Glu Lys Arg Glu Asn Asp Thr Ala
35 40 45

Asn Val Arg Pro Thr Arg Pro Ala Gly Pro Pro Pro Ala Thr Tyr Tyr
50 55 60

Asp Ser Leu Ala Val Ser Gly Gln Gly Lys Glu Arg Leu Phe Thr Thr
65 70 75 80

Asp Glu Val Arg Arg His Ile Leu Pro Thr Asp Gly Trp Leu Thr Cys
85 90 95

His Glu Gly Val Tyr Asp Val Thr Asp Phe Leu Ala Lys His Pro Gly
100 105 110

Gly Gly Val Ile Thr Leu Gly Leu Gly Arg Asp Cys Thr Ile Leu Ile
115 120 125

```
Glu Ser Tyr His Pro Ala Gly Arg Pro Asp Lys Val Met Glu Lys Tyr
    130                 135                 140

Arg Ile Gly Thr Leu Gln Asp Pro Lys Thr Phe Tyr Ala Trp Gly Glu
145                 150                 155                 160

Ser Asp Phe Tyr Pro Glu Leu Lys Arg Arg Ala Leu Ala Arg Leu Lys
                165                 170                 175

Glu Ala Gly Gln Ala Arg Arg Gly Gly Leu Gly Val Lys Ala Leu Leu
            180                 185                 190

Val Leu Thr Leu Phe Phe Val Ser Trp Tyr Met Trp Val Ala His Lys
        195                 200                 205

Ser Phe Leu Trp Ala Ala Val Trp Gly Phe Ala Gly Ser His Val Gly
    210                 215                 220

Leu Ser Ile Gln His Asp Gly Asn His Gly Ala Phe Ser Arg Asn Thr
225                 230                 235                 240

Leu Val Asn Arg Leu Ala Gly Trp Gly Met Asp Leu Ile Gly Ala Ser
                245                 250                 255

Ser Thr Val Trp Glu Tyr Gln His Val Ile Gly His His Gln Tyr Thr
            260                 265                 270

Asn Leu Val Ser Asp Thr Leu Phe Ser Leu Pro Glu Asn Asp Pro Asp
        275                 280                 285

Val Phe Ser Ser Tyr Pro Leu Met Arg Met His Pro Asp Thr Ala Trp
    290                 295                 300

Gln Pro His His Arg Phe Gln His Leu Phe Ala Phe Pro Leu Phe Ala
305                 310                 315                 320

Leu Met Thr Ile Ser Lys Val Leu Thr Ser Asp Phe Ala Val Cys Leu
                325                 330                 335

Ser Met Lys Lys Gly Ser Ile Asp Cys Ser Ser Arg Leu Val Pro Leu
            340                 345                 350

Glu Gly Gln Leu Leu Phe Trp Gly Ala Lys Leu Ala Asn Phe Leu Leu
        355                 360                 365

Gln Ile Val Leu Pro Cys Tyr Leu His Gly Thr Ala Met Gly Leu Ala
    370                 375                 380

Leu Phe Ser Val Ala His Leu Val Ser Gly Glu Tyr Leu Ala Ile Cys
385                 390                 395                 400

Phe Ile Ile Asn His Ile Ser Glu Ser Cys Glu Phe Met Asn Thr Ser
                405                 410                 415

Phe Gln Thr Ala Ala Arg Arg Thr Glu Met Leu Gln Ala Ala His Gln
            420                 425                 430

Ala Ala Glu Ala Lys Lys Val Lys Pro Thr Pro Pro Pro Asn Asp Trp
        435                 440                 445

Ala Val Thr Gln Val Gln Cys Cys Val Asn Trp Arg Ser Gly Gly Val
```

```
    450                 455                 460

Leu Ala Asn His Leu Ser Gly Gly Leu Asn His Gln Ile Glu His His
465                 470                 475                 480

Leu Phe Pro Ser Ile Ser His Ala Asn Tyr Pro Thr Ile Ala Pro Val
                485                 490                 495

Val Lys Glu Val Cys Glu Glu Tyr Gly Leu Pro Tyr Lys Asn Tyr Val
            500                 505                 510

Thr Phe Trp Asp Ala Val Cys Gly Met Val Gln His Leu Arg Leu Met
        515                 520                 525

Gly Ala Pro Pro Val Pro Thr Asn Gly Asp Lys Lys Ser
    530                 535                 540


<210>  14
<211>  515
<212>  PRT
<213>  Thraustochytrium aureum

<220>
<221>  MISC_FEATURE
<222>  (1)..(515)
<223>  delta 4-desaturase; GenBank Accession No. AAN75707

<400>  14

Met Thr Val Gly Phe Asp Glu Thr Val Thr Met Asp Thr Val Arg Asn
1               5                   10                  15

His Asn Met Pro Asp Asp Ala Trp Cys Ala Ile His Gly Thr Val Tyr
            20                  25                  30

Asp Ile Thr Lys Phe Ser Lys Val His Pro Gly Gly Asp Ile Ile Met
        35                  40                  45

Leu Ala Ala Gly Lys Glu Ala Thr Ile Leu Phe Glu Thr Tyr His Ile
    50                  55                  60

Lys Gly Val Pro Asp Ala Val Leu Arg Lys Tyr Lys Val Gly Lys Leu
65                  70                  75                  80


Pro Gln Gly Lys Lys Gly Glu Thr Ser His Met Pro Thr Gly Leu Asp
                85                  90                  95

Ser Ala Ser Tyr Tyr Ser Trp Asp Ser Glu Phe Tyr Arg Val Leu Arg
            100                 105                 110

Glu Arg Val Ala Lys Lys Leu Ala Glu Pro Gly Leu Met Gln Arg Ala
        115                 120                 125

Arg Met Glu Leu Trp Ala Lys Ala Ile Phe Leu Leu Ala Gly Phe Trp
    130                 135                 140

Gly Ser Leu Tyr Ala Met Cys Val Leu Asp Pro His Gly Gly Ala Met
145                 150                 155                 160

Val Ala Ala Val Thr Leu Gly Val Phe Ala Ala Phe Val Gly Thr Cys
```

```
                165                 170                 175

Ile Gln His Asp Gly Ser His Gly Ala Phe Ser Lys Ser Arg Phe Met
            180                 185                 190

Asn Lys Ala Ala Gly Trp Thr Leu Asp Met Ile Gly Ala Ser Ala Met
        195                 200                 205

Thr Trp Glu Met Gln His Val Leu Gly His His Pro Tyr Thr Asn Leu
    210                 215                 220

Ile Glu Met Glu Asn Gly Leu Ala Lys Val Lys Gly Ala Asp Val Asp
225                 230                 235                 240

Pro Lys Lys Val Asp Gln Glu Ser Asp Pro Asp Val Phe Ser Thr Tyr
                245                 250                 255

Pro Met Leu Arg Leu His Pro Trp His Arg Gln Arg Phe Tyr His Lys
            260                 265                 270

Phe Gln His Leu Tyr Ala Pro Leu Ile Phe Gly Phe Met Thr Ile Asn
        275                 280                 285

Lys Val Ile Ser Gln Asp Val Gly Val Val Leu Arg Lys Arg Leu Phe
    290                 295                 300

Gln Ile Asp Ala Asn Cys Arg Tyr Gly Ser Pro Trp Asn Val Ala Arg
305                 310                 315                 320

Phe Trp Ile Met Lys Leu Leu Thr Thr Leu Tyr Met Val Ala Leu Pro
                325                 330                 335

Met Tyr Met Gln Gly Pro Ala Gln Gly Leu Lys Leu Phe Phe Met Ala
            340                 345                 350

His Phe Thr Cys Gly Glu Val Leu Ala Thr Met Phe Ile Val Asn His
        355                 360                 365

Ile Ile Glu Gly Val Ser Tyr Ala Ser Lys Asp Ala Val Lys Gly Val
    370                 375                 380

Met Ala Pro Pro Arg Thr Val His Gly Val Thr Pro Met Gln Val Thr
385                 390                 395                 400

Gln Lys Ala Leu Ser Ala Ala Glu Ser Thr Lys Ser Asp Ala Asp Lys
                405                 410                 415

Thr Thr Met Ile Pro Leu Asn Asp Trp Ala Ala Val Gln Cys Gln Thr
            420                 425                 430

Ser Val Asn Trp Ala Val Gly Ser Trp Phe Trp Asn His Phe Ser Gly
        435                 440                 445

Gly Leu Asn His Gln Ile Glu His His Cys Phe Pro Gln Asn Pro His
    450                 455                 460

Thr Val Asn Val Tyr Ile Ser Gly Ile Val Lys Glu Thr Cys Glu Glu
465                 470                 475                 480

Tyr Gly Val Pro Tyr Gln Ala Glu Ile Ser Leu Phe Ser Ala Tyr Phe
                485                 490                 495
```

Lys Met Leu Ser His Leu Arg Thr Leu Gly Asn Glu Asp Leu Thr Ala
500 505 510

Trp Ser Thr
515

<210> 15
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer D4-F1

<220>
<221> misc_feature
<222> (6)..(6)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (9)..(9)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (18)..(18)
<223> n is a, c, g, or t

<400> 15
ttcctngcna arcayccngg 20

<210> 16
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer D4-F2

<220>
<221> misc_feature
<222> (6)..(6)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (9)..(9)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (18)..(18)
<223> n is a, c, g, or t

<400> 16
tttctngcna arcayccngg 20

<210> 17

<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer D4-F3

<220>
<221> misc_feature
<222> (9)..(9)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (18)..(18)
<223> n is a, c, g, or t

<400> 17
ttyttrgcna arcayccngg 20

<210> 18
<211> 7
<212> PRT
<213> Artificial Sequence

<220>
<223> Translation of primers D4-F1, D4-F2 and D4-F3

<400> 18

Phe Leu Ala Lys His Pro Gly
1 5

<210> 19
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer D4-F4

<220>
<221> misc_feature
<222> (15)..(15)
<223> n is a, c, g, or t

<400> 19
athcarcayg ayggnaayca 20

<210> 20
<211> 7
<212> PRT
<213> Artificial Sequence

<220>
<223> Translation of primer D4-F4

<400> 20

Ile Gln His Asp Gly Asn His
1 5

<210> 21
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer D4-F5

<220>
<221> misc_feature
<222> (9)..(9)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (18)..(18)
<223> n is a, c, g, or t

<400> 21
caygayggna aycayggngc 20

<210> 22
<211> 7
<212> PRT
<213> Artificial Sequence

<220>
<223> Translation of primer D4-F5

<400> 22

His Asp Gly Asn His Gly Ala
1 5

<210> 23
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer D4-F6

<220>
<221> misc_feature
<222> (3)..(3)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (18)..(18)
<223> n is a, c, g, or t

<400> 23
ggncaycaya gyttyacnaa 20

<210> 24
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer D4-F7

<220>
<221> misc_feature
<222> (12)..(12)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (18)..(18)
<223> n is a, c, g, or t

<400> 24
ggycaycayt cnttyacnaa 20

<210> 25
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer D4-F8

<220>
<221> misc_feature
<222> (12)..(12)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (18)..(18)
<223> n is a, c, g, or t

<400> 25
ggrcaycayt cnttyacnaa 20

<210> 26
<211> 7
<212> PRT
<213> Artificial Sequence

<220>
<223> Translation of primers D4-F6, D4-F7 and D4-F8

<400> 26

Gly His His Gln Tyr Thr Asn
1 5

<210> 27
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer D4-R1

<220>
<221> misc_feature
<222> (3)..(3)
<223> n is a, c, g, or t

<400> 27
aanagrtgrt gytcdatytg 20

<210> 28
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer D4-R2

<400> 28
aayaartgrt gytcdatytg 20

<210> 29
<211> 7
<212> PRT
<213> Artificial Sequence

<220>
<223> Translation of primers D4-R1 and D4-R2

<400> 29

Gln Ile Glu His His Leu Phe
1 5

<210> 30
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer 1594D4-5-1

<400> 30
gtccacgtca agaacgttgg tg 22

<210> 31
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer 1594D4-5-2

<400> 31
gatgatggcc gagcatatgt tg 22

<210> 32
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer DNR CDS 5-2

<400> 32
caacgcagag tggccattac gg 22

<210> 33
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer 1594D4-5-4

<400> 33
gaatactgcc ttgacccaga tc 22

<210> 34
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer 1594D4-5-5

<400> 34
cttcgcggca ggttgagctt gtc 23

<210> 35
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer 1594D4-3-1

<400> 35
acgtggagtt caagcccacc ac 22

<210> 36
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer 1594D4-3-2

<400> 36
gtttcgggcc gcacgcccat g 21

<210> 37

<211> 550
<212> PRT
<213> Thalassiosira pseudonana

<220>
<221> MISC_FEATURE
<222> (1)..(550)
<223> delta-4 desaturase; GenBank Accession No. AAX14506

<400> 37

Met Gly Asn Gly Asn Leu Pro Ala Ser Thr Ala Gln Leu Lys Ser Thr
1 5 10 15

Ser Lys Pro Gln Gln Gln His Glu His Arg Thr Ile Ser Lys Ser Glu
20 25 30

Leu Ala Gln His Asn Thr Pro Lys Ser Ala Trp Cys Ala Val His Ser
35 40 45

Thr Pro Ala Thr Asp Pro Ser His Ser Asn Asn Lys Gln His Ala His
50 55 60

Leu Val Leu Asp Ile Thr Asp Phe Ala Ser Arg His Pro Gly Gly Asp
65 70 75 80

Leu Ile Leu Leu Ala Ser Gly Lys Asp Ala Ser Val Leu Phe Glu Thr
85 90 95

Tyr His Pro Arg Gly Val Pro Thr Ser Leu Ile Gln Lys Leu Gln Ile
100 105 110

Gly Val Met Glu Glu Glu Ala Phe Arg Asp Ser Phe Tyr Ser Trp Thr
115 120 125

Asp Ser Asp Phe Tyr Thr Val Leu Lys Arg Arg Val Val Glu Arg Leu
130 135 140

Glu Glu Arg Gly Leu Asp Arg Arg Gly Ser Lys Glu Ile Trp Ile Lys
145 150 155 160

Ala Leu Phe Leu Leu Val Gly Phe Trp Tyr Cys Leu Tyr Lys Met Tyr
165 170 175

Thr Thr Ser Asp Ile Asp Gln Tyr Gly Ile Ala Ile Ala Tyr Ser Ile
180 185 190

Gly Met Gly Thr Phe Ala Ala Phe Ile Gly Thr Cys Ile Gln His Asp
195 200 205

Gly Asn His Gly Ala Phe Ala Gln Asn Lys Leu Leu Asn Lys Leu Ala
210 215 220

Gly Trp Thr Leu Asp Met Ile Gly Ala Ser Ala Phe Thr Trp Glu Leu
225 230 235 240

Gln His Met Leu Gly His His Pro Tyr Thr Asn Val Leu Asp Gly Val
245 250 255

Glu Glu Glu Arg Lys Glu Arg Gly Glu Asp Val Ala Leu Glu Glu Lys
260 265 270

```
Asp Gln Glu Ser Asp Pro Asp Val Phe Ser Ser Phe Pro Leu Met Arg
        275                 280                 285

Met His Pro His His Thr Thr Ser Trp Tyr His Lys Tyr Gln His Leu
    290                 295                 300

Tyr Ala Pro Pro Leu Phe Ala Leu Met Thr Leu Ala Lys Val Phe Gln
305                 310                 315                 320

Gln Asp Phe Glu Val Ala Thr Ser Gly Arg Leu Tyr His Ile Asp Ala
                325                 330                 335

Asn Val Arg Tyr Gly Ser Val Trp Asn Val Met Arg Phe Trp Ala Met
            340                 345                 350

Lys Val Ile Thr Met Gly Tyr Met Met Gly Leu Pro Ile Tyr Phe His
        355                 360                 365

Gly Val Leu Arg Gly Val Gly Leu Phe Val Ile Gly His Leu Ala Cys
    370                 375                 380

Gly Glu Leu Leu Ala Thr Met Phe Ile Val Asn His Val Ile Glu Gly
385                 390                 395                 400

Val Ser Tyr Gly Thr Lys Asp Leu Val Gly Gly Ala Ser His Gly Asp
                405                 410                 415

Glu Lys Lys Ile Val Lys Pro Thr Thr Val Leu Gly Asp Thr Pro Met
            420                 425                 430

Glu Lys Thr Arg Glu Glu Ala Leu Lys Ser Asn Ser Asn Asn Asn Lys
        435                 440                 445

Lys Lys Gly Glu Lys Asn Ser Val Pro Ser Val Pro Phe Asn Asp Trp
    450                 455                 460

Ala Ala Val Gln Cys Gln Thr Ser Val Asn Trp Ser Pro Gly Ser Trp
465                 470                 475                 480

Phe Trp Asn His Phe Ser Gly Gly Leu Ser His Gln Ile Glu His His
                485                 490                 495

Leu Phe Pro Ser Ile Cys His Thr Asn Tyr Cys His Ile Gln Asp Val
            500                 505                 510

Val Glu Ser Thr Cys Ala Glu Tyr Gly Val Pro Tyr Gln Ser Glu Ser
        515                 520                 525

Asn Leu Phe Val Ala Tyr Gly Lys Met Ile Ser His Leu Lys Phe Leu
    530                 535                 540

Gly Lys Ala Lys Cys Glu
545                 550
```

<210> 38
<211> 519
<212> PRT
<213> Thraustochytrium sp. FJN-10

<220>
<221> MISC_FEATURE

<222> (1)..(519)
<223> delta-4 desaturase; GenBank Accession No. AAZ43257

<400> 38

```
Met Thr Val Gly Tyr Asp Gly Glu Ile Pro Phe Glu Gln Val Arg Ala
1               5                  10                  15

His Asn Lys Pro Asp Asp Ala Trp Cys Ala Ile His Gly His Val Tyr
            20                  25                  30

Asp Val Thr Lys Phe Ala Ser Val His Pro Gly Gly Asp Ile Ile Leu
        35                  40                  45

Leu Ala Ala Gly Lys Asp Ala Thr Val Leu Tyr Glu Thr Tyr His Val
    50                  55                  60

Arg Gly Val Ser Asp Ala Val Leu Arg Lys Tyr Arg Ile Gly Lys Leu
65                  70                  75                  80

Pro Asp Gly Gln Gly Gly Ala Asn Glu Lys Glu Lys Arg Thr Leu Ser
                85                  90                  95

Gly Leu Ser Ser Ala Ser Tyr Tyr Thr Trp Asn Ser Asp Phe Tyr Arg
            100                 105                 110

Val Met Arg Glu Arg Val Val Ala Arg Leu Lys Glu Arg Gly Lys Ala
        115                 120                 125

Arg Arg Gly Gly Tyr Glu Leu Trp Ile Lys Ala Leu Leu Leu Leu Val
    130                 135                 140

Gly Phe Trp Ser Ser Leu Cys Trp Met Cys Thr Leu Asp Pro Ser Phe
145                 150                 155                 160

Gly Ala Ile Leu Ala Ala Met Ser Leu Gly Val Phe Ala Ala Phe Val
                165                 170                 175

Gly Thr Cys Ile Gln His Asp Gly Asn His Gly Ala Phe Ala Gln Ser
            180                 185                 190

Arg Trp Val Asn Lys Val Ala Gly Trp Thr Leu Asp Met Ile Gly Ala
        195                 200                 205

Ser Gly Met Thr Trp Glu Phe Gln His Ala Leu Gly His His Pro Tyr
    210                 215                 220

Thr Asn Leu Ile Glu Glu Glu Asn Gly Leu Gln Lys Val Ser Gly Lys
225                 230                 235                 240

Lys Met Asp Thr Lys Leu Ala Asp Gln Glu Ser Asp Pro Asp Val Phe
                245                 250                 255

Ser Thr Tyr Pro Met Met Arg Leu His Pro Trp His Gln Lys Arg Trp
            260                 265                 270

Tyr His Arg Phe Gln His Ile Tyr Gly Pro Phe Ile Phe Gly Phe Met
        275                 280                 285

Thr Ile Asn Lys Val Val Thr Gln Asp Val Gly Val Val Phe Arg Lys
    290                 295                 300
```

```
Arg Leu Phe Gln Ile Asp Ala Glu Cys Arg Tyr Ala Ser Pro Met Tyr
305                 310                 315                 320

Val Ala Arg Phe Trp Ile Met Lys Ala Leu Thr Val Leu Tyr Met Val
                325                 330                 335


Ala Leu Pro Cys Tyr Met Gln Gly Pro Trp His Gly Leu Lys Leu Phe
            340                 345                 350

Ala Ile Ala His Phe Thr Cys Gly Glu Val Leu Ala Thr Met Phe Ile
        355                 360                 365

Val Asn His Val Ile Glu Gly Val Ser Tyr Ala Ser Lys Asp Ala Val
    370                 375                 380

Lys Gly Thr Met Ala Pro Pro Lys Thr Met His Gly Val Thr Pro Met
385                 390                 395                 400

Asn Asn Thr Arg Lys Glu Val Glu Ala Glu Ala Ser Lys Ser Gly Ala
                405                 410                 415

Val Val Lys Ser Val Pro Leu Asp Asp Trp Ala Ala Val Gln Cys Gln
            420                 425                 430

Thr Ser Ala Asn Trp Ser Val Gly Ser Trp Phe Trp Asn His Phe Ser
        435                 440                 445

Gly Gly Leu Asn His Gln Ile Glu His His Leu Phe Pro Gly Leu Ser
    450                 455                 460

His Glu Thr Tyr Tyr His Ile Gln Asp Val Val Gln Ser Thr Cys Ala
465                 470                 475                 480

Glu Tyr Gly Val Pro Tyr Gln His Glu Pro Ser Leu Trp Thr Ala Tyr
                485                 490                 495

Trp Lys Met Leu Glu His Leu Arg Arg Leu Gly Asn Glu Glu Thr His
            500                 505                 510

Glu Ser Trp Gln Arg Ala Ala
        515


<210>  39
<211>  4275
<212>  DNA
<213>  Artificial Sequence

<220>
<223>  Plasmid p1594D4S

<400>  39
tcgcgcgttt cggtgatgac ggtgaaaacc tctgacacat gcagctcccg gagacggtca       60

cagcttgtct gtaagcggat gccgggagca gacaagcccg tcagggcgcg tcagcgggtg      120

ttggcgggtg tcggggctgg cttaactatg cggcatcaga gcagattgta ctgagagtgc      180

accatatgcg gtgtgaaata ccgcacagat gcgtaaggag aaaataccgc atcaggcgcc      240

attcgccatt caggctgcgc aactgttggg aagggcgatc ggtgcgggcc tcttcgctat      300
```

tacgccagct ggcgaaaggg ggatgtgctg caaggcgatt aagttgggta acgccagggt 360

tttcccagtc acgacgttgt aaaacgacgg ccagtgaatt cgagctcggt acctcgcgaa 420

tgcatctaga tatcggatcc cgggcggccg cttaagcgac ctttttctcg tttccgaggt 480

acttgagatg agacagcatc ttgccatagg cggagaacaa agagggttcg gactggtaag 540

gaacaccgta ctcttcgcaa gtcttctgga caacgtcctg gatgtgagcg tagttggtgt 600

gacaaatgct gggaaacaga tggtgctcga tctggtggga gaggccaccg gagaagtgat 660

tccagaacca cgatccagga gaccagttga cgctcgtttg acactgaacg gcagcccagt 720

cgttgtaggg aacatccttg acgttgcctc cattggcagc ctttttggcc tcggcacggg 780

tctgctccat tggagttcgt ccagaaacgg ttgtaggctt gaactccacg tccttggaca 840

gaccgagaga ttctcccttt ttgccaaagg acacaccctc gatgacgtga ttgacaatga 900

acatggttgc cagaagctct ccacaggcaa agtggccgat gagaaacagt ccaaggccac 960

caaggattcc gtggaagtag cagggaacag ccagcatgta gccgatggag agcaccttca 1020

tcgaccaaaa ccgaaggaca ttgagaatgg aattgtatcg gcacttggcg tcgatgtggt 1080

acagtcgctg agtggtagcg acttcgatgt cctgttggaa caccttggca agagtcatga 1140

aagcaaagag aacgggtgcg tacaggtgct gatatcggtg gtaccactcg gccttgtggt 1200

aggggtgcat tcgcatgaga ggaaacgaag agaagacatc tgggtcggac tcctgatcct 1260

tgtcctccat aggacaatcg tcgccagctt cctttcgctt ctcttcgtcc acgtccagga 1320

cgttggtgta gggatggtgt ccgagcatat gctgaatctc ccatgtaaag ccagaagcac 1380

cgatcatgtc cagagtccag cctgcaacct tgttgagagc agggctggtc gagaaggcac 1440

cgtggtttcc atcgtgttga atgcaagtgc cgatgaaggc agcaaagact cccatagaca 1500

cggaccacag cgaggcagcc cagtaggtct cgttgacgga catcttgtag aggctgaacc 1560

agaatccagc caggaggaat actgccttga cccaaatctc gtagccacct cttcgaggca 1620

ggttgagctt gtccaatcgc tcgaccactc gagctcgcag ggtcttgtaa aagtcggaat 1680

cccacgagta gaacgaggag ggaagttctc cgtccttcat cttgccgacc tgcagcttgt 1740

cgaggatcga ggtgggaaca cctcgaggat ggtaagtctc gaagagcacg gtggcatcct 1800

ttcctgcggc aagcagaatg atgtcgccac caggatgccg tcgggcgaac tcggtaatgt 1860

tgatgactct gccgtggacg gcagcccagg catcctgggg aacgttgtga gaccgcagct 1920

cgtctcggga gatgagtcgg ttcttgtcga gagacttgtc ggtggcagca gtgtcggcag 1980

ccttggtgga ctgagccatg gattgggggc ccgtcgactg cagaggcctg catgcaagct 2040

tggcgtaatc atggtcatag ctgtttcctg tgtgaaattg ttatccgctc acaattccac 2100

acaacatacg agccggaagc ataaagtgta aagcctgggg tgcctaatga gtgagctaac 2160

tcacattaat tgcgttgcgc tcactgcccg ctttccagtc gggaaacctg tcgtgccagc 2220

tgcattaatg aatcggccaa cgcgcgggga gaggcggttt gcgtattggg cgctcttccg 2280

cttcctcgct cactgactcg ctgcgctcgg tcgttcggct gcggcgagcg gtatcagctc 2340

actcaaaggc ggtaatacgg ttatccacag aatcagggga taacgcagga aagaacatgt 2400

gagcaaaagg ccagcaaaag gccaggaacc gtaaaaaggc cgcgttgctg gcgtttttcc 2460

ataggctccg ccccccctgac gagcatcaca aaaatcgacg ctcaagtcag aggtggcgaa 2520

acccgacagg actataaaga taccaggcgt ttcccccctgg aagctccctc gtgcgctctc 2580

ctgttccgac cctgccgctt accggatacc tgtccgcctt tctcccttcg ggaagcgtgg 2640

cgctttctca tagctcacgc tgtaggtatc tcagttcggt gtaggtcgtt cgctccaagc 2700

tgggctgtgt gcacgaaccc cccgttcagc ccgaccgctg cgccttatcc ggtaactatc 2760

gtcttgagtc caacccggta agacacgact tatcgccact ggcagcagcc actggtaaca 2820

ggattagcag agcgaggtat gtaggcggtg ctacagagtt cttgaagtgg tggcctaact 2880

acggctacac tagaagaaca gtatttggta tctgcgctct gctgaagcca gttaccttcg 2940

gaaaaagagt tggtagctct tgatccggca aacaaaccac cgctggtagc ggtggttttt 3000

ttgtttgcaa gcagcagatt acgcgcagaa aaaaaggatc tcaagaagat cctttgatct 3060

tttctacggg gtctgacgct cagtggaacg aaaactcacg ttaagggatt ttggtcatga 3120

gattatcaaa aaggatcttc acctagatcc ttttaaatta aaaatgaagt tttaaatcaa 3180

tctaaagtat atatgagtaa acttggtctg acagttacca atgcttaatc agtgaggcac 3240

ctatctcagc gatctgtcta tttcgttcat ccatagttgc ctgactcccc gtcgtgtaga 3300

taactacgat acgggagggc ttaccatctg gccccagtgc tgcaatgata ccgcgagacc 3360

cacgctcacc ggctccagat ttatcagcaa taaaccagcc agccggaagg gccgagcgca 3420

gaagtggtcc tgcaacttta tccgcctcca tccagtctat taattgttgc cgggaagcta 3480

gagtaagtag ttcgccagtt aatagtttgc gcaacgttgt tgccattgct acaggcatcg 3540

tggtgtcacg ctcgtcgttt ggtatggctt cattcagctc cggttcccaa cgatcaaggc 3600

gagttacatg atcccccatg ttgtgcaaaa aagcggttag ctccttcggt cctccgatcg 3660

ttgtcagaag taagttggcc gcagtgttat cactcatggt tatggcagca ctgcataatt 3720

ctcttactgt catgccatcc gtaagatgct tttctgtgac tggtgagtac tcaaccaagt 3780

cattctgaga atagtgtatg cggcgaccga gttgctcttg cccggcgtca atacgggata 3840

ataccgcgcc acatagcaga actttaaaag tgctcatcat tggaaaacgt tcttcggggc 3900

gaaaactctc aaggatctta ccgctgttga gatccagttc gatgtaaccc actcgtgcac 3960

ccaactgatc ttcagcatct tttactttca ccagcgtttc tgggtgagca aaaacaggaa 4020

ggcaaaatgc cgcaaaaaag ggaataaggg cgacacggaa atgttgaata ctcatactct 4080

tccttttca atattattga agcatttatc agggttattg tctcatgagc ggatacatat 4140

ttgaatgtat ttagaaaaat aaacaaatag gggttccgcg cacatttccc cgaaaagtgc 4200

cacctgacgt ctaagaaacc attattatca tgacattaac ctataaaaat aggcgtatca 4260

cgaggccctt tcgtc 4275

<210> 40
<211> 13175
<212> DNA
<213> Artificial Sequence

<220>
<223> Plasmid pZKL4-220ESC4

<400> 40
catggctcag tccaccaagg ctgccgacac tgctgccacc gacaagtctc tcgacaagaa 60

ccgactcatc tcccgagacg agctgcggtc tcacaacgtt ccccaggatg cctgggctgc 120

cgtccacggc agagtcatca acattaccga gttcgcccga cggcatcctg gtggcgacat 180

cattctgctt gccgcaggaa aggatgccac cgtgctcttc gagacttacc atcctcgagg 240

tgttcccacc tcgatcctcg acaagctgca ggtcggcaag atgaaggacg gagaacttcc 300

ctcctcgttc tactcgtggg attccgactt ttacaagacc ctgcgagctc gagtggtcga 360

gcgattggac aagctcaacc tgcctcgaag aggtggctac gagatttggg tcaaggcagt 420

attcctcctg gctggattct ggttcagcct ctacaagatg tccgtcaacg agacctactg 480

ggctgcctcg ctgtggtccg tgtctatggg agtctttgct gccttcatcg gcacttgcat 540

tcaacacgat ggaaaccacg gtgccttctc gaccagccct gctctcaaca aggttgcagg 600

ctggactctg gacatgatcg gtgcttctgg ctttacatgg gagattcagc atatgctcgg 660

acaccatccc tacaccaacg tcctggacgt ggacgaagag aagcgaaagg aagctggcga 720

cgattgtcct atggaggaca aggatcagga gtccgaccca gatgtcttct cttcgtttcc 780

tctcatgcga atgcacccct accacaaggc cgagtggtac caccgatatc agcacctgta 840

cgcacccgtt ctctttgctt tcatgactct tgccaaggtg ttccaacagg acatcgaagt 900

cgctaccact cagcgactgt accacatcga cgccaagtgc cgatacaatt ccattctcaa 960

tgtccttcgg ttttggtcga tgaaggtgct ctccatcggc tacatgctgg ctgttccctg 1020

ctacttccac ggaatccttg gtggccttgg actgtttctc atcggccact ttgcctgtgg 1080

agagcttctg gcaaccatgt tcattgtcaa tcacgtcatc gagggtgtgt cctttggcaa 1140

aaagggagaa tctctcggtc tgtccaagga cgtggagttc aagcctacaa ccgtttctgg 1200

acgaactcca atggagcaga cccgtgccga ggccaaaaag gctgccaatg gaggcaacgt 1260

caaggatgtt ccctacaacg actgggctgc cgttcagtgt caaacgagcg tcaactggtc 1320

tcctggatcg tggttctgga atcacttctc cggtggcctc tcccaccaga tcgagcacca 1380

tctgtttccc agcatttgtc acaccaacta cgctcacatc caggacgttg tccagaagac 1440

ttgcgaagag tacggtgttc cttaccagtc cgaaccctct ttgttctccg cctatggcaa 1500

gatgctgtct catctcaagt acctcggaaa cgagaaaaag gtcgcttaag cggccgcatg 1560

tacatacaag attatttata gaaatgaatc gcgatcgaac aaagagtacg agtgtacgag 1620

taggggatga tgataaaagt ggaagaagtt ccgcatcttt ggatttatca acgtgtagga 1680

cgatacttcc tgtaaaaatg caatgtcttt accataggtt ctgctgtaga tgttattaac 1740

taccattaac atgtctactt gtacagttgc agaccagttg gagtatagaa tggtacactt 1800

accaaaaagt gttgatggtt gtaactacga tatataaaac tgttgacggg atctgcgtac 1860

actgtttaaa cagagtgtga aagactcact atggtccggg cttatctcga ccaatagcca 1920

aagtctggag tttctgagag aaaaaggcaa gatacgtatg taacaaagcg acgcatggta 1980

caataatacc ggaggcatgt atcatagaga gttagtggtt cgatgatggc actggtgcct 2040

ggtatgactt tatacggctg actacatatt tgtcctcaga catacaatta cagtcaagca 2100

cttacccttg gacatctgta ggtacccccc ggccaagacg atctcagcgt gtcgtatgtc 2160

ggattggcgt agctccctcg ctcgtcaatt ggctcccatc tactttcttc tgcttggcta 2220

cacccagcat gtctgctatg gctcgttttc gtgccttatc tatcctccca gtattaccaa 2280

ctctaaatga catgatgtga ttgggtctac actttcatat cagagataag gagtagcaca 2340

gttgcataaa aagcccaact ctaatcagct tcttcctttc ttgtaattag tacaaaggtg 2400

attagcgaaa tctggaagct tagttggccc taaaaaaatc aaaaaaagca aaaaacgaaa 2460

aacgaaaaac cacagttttg agaacaggga ggtaacgaag gatcgtatat atatatatat 2520

atatatatac ccacggatcc cgagaccggc ctttgattct tccctacaac caaccattct 2580

caccacccta attcacaacc atggctgact ctcccgtcat caacctctcc accatgtgga 2640

agcctctgtc gctcatggcc ttggatcttg ctgttctggg acacgtctgg aagcaggcac 2700

aacaggaggg ctccatctcg gcttacgccg actctgtgtg gactcccctc atcatgtccg 2760

gtctgtacct ctccatgatc ttcgtgggat gtcgatggat gaagaaccga gagcccttcg 2820

aaatcaagac ctacatgttt gcctacaacc tgtaccagac cctcatgaac ctttgcattg 2880

tgctgggctt cctctaccag gtccacgcta ccggtatgcg attctgggga tctggcgtgg 2940

accgatcgcc caagggtctg ggaattggct ttttcatcta tgcccattac cacaacaagt 3000

acgtcgagta cttcgacaca ctcttcatgg tgctgcggaa aaagaacaac cagatttcct 3060

ttcttcacgt ctaccatcac gctctgctca cctgggcttg gtttgccgtg gtctacttcg 3120

ctcctggagg tgacggctgg tttggagcct gctacaattc ctccattcat gtcctgatgt 3180

actcttacta tctgcttgcc accttcggca tctcctgtcc ctggaaaaag atcctcaccc 3240

agctgcaaat ggttcagttc tgcttttgct tcacccactc gatctacgtg tggatttgcg 3300

gttccgaaat ctaccctcga cccttgactg ctctccagtc cttcgtgatg gtcaacatgc 3360

tggttctctt tggcaacttc tacgtcaagc agtattctca gaagaatgga aagcccgaga 3420

acggtgccac tcctgagaac ggtgccaagc ctcagccctg cgagaacggc accgtcgaga 3480

agcgagagaa cgacactgcc aacgttcgat aagcggccgc atgagaagat aaatatataa 3540

atacattgag atattaaatg cgctagatta gagagcctca tactgctcgg agagaagcca 3600

agacgagtac tcaaagggga ttacaccatc catatccaca gacacaagct ggggaaaggt 3660

tctatataca ctttccggaa taccgtagtt tccgatgtta tcaatggggg cagccaggat 3720

ttcaggcact tcggtgtctc ggggtgaaat ggcgttcttg gcctccatca agtcgtacca 3780

tgtcttcatt tgcctgtcaa agtaaaacag aagcagatga agaatgaact tgaagtgaag 3840

gaatttaaat tgccccggag aagacggcca ggccgcctag atgacaaatt caacaactca 3900

cagctgactt tctgccattg ccactagggg ggggcctttt tatatggcca agccaagctc 3960

tccacgtcgg ttgggctgca cccaacaata aatgggtagg gttgcaccaa caaagggatg 4020

ggatgggggg tagaagatac gaggataacg gggctcaatg gcacaaataa gaacgaatac 4080

tgccattaag actcgtgatc cagcgactga caccattgca tcatctaagg gcctcaaaac 4140

tacctcggaa ctgctgcgct gatctggaca ccacagaggt tccgagcact ttaggttgca 4200

ccaaatgtcc caccaggtgc aggcagaaaa cgctggaaca gcgtgtacag tttgtcttaa 4260

caaaaagtga gggcgctgag gtcgagcagg gtggtgtgac ttgttatagc ctttagagct 4320

gcgaaagcgc gtatggattt ggctcatcag gccagattga gggtctgtgg acacatgtca 4380

tgttagtgta cttcaatcgc cccctggata tagccccgac aataggccgt ggcctcattt 4440

ttttgccttc cgcacatttc cattgctcgg tacccacacc ttgcttctcc tgcacttgcc 4500

aaccttaata ctggtttaca ttgaccaaca tcttacaagc ggggggcttg tctagggtat 4560

atataaacag tggctctccc aatcggttgc cagtctcttt tttccttttct ttccccacag 4620

attcgaaatc taaactacac atcacagaac tccgagccgt gagtatccac gacaagatca 4680

gtgtcgagac gacgcgtttt gtgtaatgac acaatccgaa agtcgctagc aacacacact 4740

ctctacacaa actaacccag ctctggtacc atggccgagg gcaagtccga cggtcccgtc 4800

gttaccctcc agtccatgtg gaagcccctg gctctcatgg ccatcgacgt cggcatcctg 4860

gtcaacgtgc gacggaaggc cttcaccgag ttcgacggac actcgaacgt cttcgccgat 4920

cccgtgtaca ttccctttgt catgaacctg ttctacctca ccatgatctt tgctggctgc 4980

cgatggatga agactcgaga acccttcgag atcaagtcct acatgtttgc ctacaacgct 5040

taccagacaa tgatgaactt tctcattgtg gtcggcttca tgtatgaggt tcactccacc 5100

ggtatgcgat actggggatc cagaatcgac acttctacca agggcttggg actgggtttc 5160

ctcatctatg cccattacca caacaagtac gtggagtacg tcgacaccct gttcatgatt 5220

ctgcggaaga aaaacaatca gatctcgttc cttcacgttt accaccattc cctgctcact 5280

tgggcatggt gggctgtggt ctactgggct cctggcggag atgcctggtt cggtgcctgt 5340

tacaactcct tcatccacgt tctcatgtac tcctactatc tgtttgccac cttcggcatt 5400

cgatgtccct ggaaaaagat gctcacccag ttgcaaatgg tccagttctg cttttgcttc 5460

gctcatgcca tgtacgttgg atggcttggt cacgaggtgt accctcgatg gctcactgct 5520

ctgcaggcct ttgtgatgct caacatgctg gtcctctttg gcaacttcta catgaagtct 5580

tactccaagg cgagcaagct cgaaccagcc tctcccgtgt cgcctgcctc tcttgctcag 5640

aagcccttcg agaacgccaa ggtcaagtaa gcggccgcaa gtgtggatgg ggaagtgagt 5700

gcccggttct gtgtgcacaa ttggcaatcc aagatggatg gattcaacac agggatatag 5760

cgagctacgt ggtggtgcga ggatatagca acggatattt atgtttgaca cttgagaatg 5820

tacgatacaa gcactgtcca agtacaatac taaacatact gtacatactc atactcgtac 5880

ccgggcaacg gtttcacttg agtgcagtgg ctagtgctct tactcgtaca gtgtgcaata 5940

ctgcgtatca tagtctttga tgtatatcgt attcattcat gttagttgcg tacgtaggga 6000

tcaggtgctt aggaagctcg accaaccacg gagactgttg aaactggatg tcggtaacag 6060

catctggaat gctgaatgtt cctcgaataa caacatattt ctccttgttg aggtgatcat 6120

aagctatgta tccggtgatt gaagtggaat agaagtctcc tccgaagact gagtccaacg 6180

tcatgttcgg gaaataccga caactctctc cacatgtaaa atcagttcgt agaggagtga 6240

ctggcgcatt gacacagtag gcgatgtttg caatccgaga aaacttggcc gtaaagttgt 6300

acagctcctg ggaggcttga actcgagttt ttgaaagtgt cgctggtggc tcgccgaaga 6360

gggaggcata gaggtacgca accacttgcc cgagcgtgag gttcatgatg ccaatagtga 6420

atgtcattta tcaccgtact gcgcagtatt tatatagggc tcatcggtcc atgtatagat 6480

ctgtccactt atgacacccc catgtctcat taatgtgtaa aggtggagac gggtggagta 6540

caggtacaga gttggaggaa atcaggatag tggggttaag acatgctccg agtccaaatt 6600

tcaactctcc attgtcacaa gacctctggt ttcagagtta ttacagatct aggcctgttt 6660

caaggtgagg ggacctcatc tggatcggca cgacgatcgt caccttacag aggacgtctg 6720

tcgcagggaa aggtgatgtg gcgcgccagc tgcattaatg aatcggccaa cgcgcgggga 6780

gaggcggttt gcgtattggg cgctcttccg cttcctcgct cactgactcg ctgcgctcgg 6840

tcgttcggct gcggcgagcg gtatcagctc actcaaaggc ggtaatacgg ttatccacag 6900

aatcagggga taacgcagga aagaacatgt gagcaaaagg ccagcaaaag gccaggaacc 6960

gtaaaaaggc cgcgttgctg gcgtttttcc ataggctccg ccccccctgac gagcatcaca 7020

aaaatcgacg ctcaagtcag aggtggcgaa acccgacagg actataaaga taccaggcgt 7080

ttcccccctgg aagctccctc gtgcgctctc ctgttccgac cctgccgctt accggatacc 7140

tgtccgcctt tctcccttcg ggaagcgtgg cgctttctca tagctcacgc tgtaggtatc 7200

tcagttcggt gtaggtcgtt cgctccaagc tgggctgtgt gcacgaaccc cccgttcagc 7260

ccgaccgctg cgccttatcc ggtaactatc gtcttgagtc caacccggta agacacgact 7320

tatcgccact ggcagcagcc actggtaaca ggattagcag agcgaggtat gtaggcggtg 7380

ctacagagtt cttgaagtgg tggcctaact acggctacac tagaagaaca gtatttggta 7440

tctgcgctct gctgaagcca gttaccttcg gaaaaagagt tggtagctct tgatccggca 7500

aacaaaccac cgctggtagc ggtggttttt ttgtttgcaa gcagcagatt acgcgcagaa 7560

aaaaaggatc tcaagaagat cctttgatct tttctacggg gtctgacgct cagtggaacg 7620

aaaactcacg ttaagggatt ttggtcatga gattatcaaa aaggatcttc acctagatcc 7680

ttttaaatta aaaatgaagt tttaaatcaa tctaaagtat atatgagtaa acttggtctg 7740

acagttacca atgcttaatc agtgaggcac ctatctcagc gatctgtcta tttcgttcat 7800

ccatagttgc ctgactcccc gtcgtgtaga taactacgat acgggagggc ttaccatctg 7860

gccccagtgc tgcaatgata ccgcgagacc cacgctcacc ggctccagat ttatcagcaa 7920

taaaccagcc agccggaagg gccgagcgca gaagtggtcc tgcaacttta tccgcctcca 7980

tccagtctat taattgttgc cgggaagcta gagtaagtag ttcgccagtt aatagtttgc 8040

gcaacgttgt tgccattgct acaggcatcg tggtgtcacg ctcgtcgttt ggtatggctt 8100

cattcagctc cggttcccaa cgatcaaggc gagttacatg atcccccatg ttgtgcaaaa 8160

aagcggttag ctccttcggt cctccgatcg ttgtcagaag taagttggcc gcagtgttat 8220

cactcatggt tatggcagca ctgcataatt ctcttactgt catgccatcc gtaagatgct 8280

tttctgtgac tggtgagtac tcaaccaagt cattctgaga atagtgtatg cggcgaccga 8340

gttgctcttg cccggcgtca atacgggata ataccgcgcc acatagcaga actttaaaag 8400

tgctcatcat tggaaaacgt tcttcggggc gaaaactctc aaggatctta ccgctgttga 8460

gatccagttc gatgtaaccc actcgtgcac ccaactgatc ttcagcatct tttactttca 8520

ccagcgtttc tgggtgagca aaaacaggaa ggcaaaatgc cgcaaaaaag ggaataaggg 8580

cgacacggaa atgttgaata ctcatactct tccttttttca atattattga agcatttatc 8640

agggttattg tctcatgagc ggatacatat ttgaatgtat ttagaaaaat aaacaaatag 8700

gggttccgcg cacatttccc cgaaaagtgc cacctgatgc ggtgtgaaat accgcacaga 8760

tgcgtaagga gaaaataccg catcaggaaa ttgtaagcgt taatattttg ttaaaattcg 8820

cgttaaattt ttgttaaatc agctcatttt ttaaccaata ggccgaaatc ggcaaaatcc 8880

cttataaatc aaaagaatag accgagatag ggttgagtgt tgttccagtt tggaacaaga 8940

gtccactatt aaagaacgtg gactccaacg tcaaagggcg aaaaaccgtc tatcagggcg 9000

atggcccact acgtgaacca tcaccctaat caagtttttt ggggtcgagg tgccgtaaag 9060

cactaaatcg gaaccctaaa gggagccccc gatttagagc ttgacgggga aagccggcga 9120

acgtggcgag aaaggaaggg aagaaagcga aaggagcggg cgctagggcg ctggcaagtg 9180

tagcggtcac gctgcgcgta accaccacac ccgccgcgct taatgcgccg ctacagggcg 9240

cgtccattcg ccattcaggc tgcgcaactg ttgggaaggg cgatcggtgc gggcctcttc 9300

gctattacgc cagctggcga aaggggatg tgctgcaagg cgattaagtt gggtaacgcc 9360

agggttttcc cagtcacgac gttgtaaaac gacggccagt gaattgtaat acgactcact 9420

atagggcgaa ttgggcccga cgtcgcatgc caaggcgtat attagttggt gggaaccagt 9480

gtacgaccgg gtcctgtata accaggttca gtggcatact tgtaggagtt gttcccgtgg 9540

tatttgggca tggctaagac atttcgccga ccaatgttaa gtgcacaata gccgatgtag 9600

tagatgtaag ccagatggtt ttggagcagg tcgattcgag accacagatt gaaagtgcct 9660

cgatggcagg cctcgttttc tcctccttca cagacagaca ctttgtcgag tgcagggtag 9720

acgctttctt ggtcaatgta cacttctcca atggcgtgtg tgtaattgga ccaatctggc 9780

agtccaacga agatatcgtt ccagtgggtg agcctgtagt ttcgtcgctt gtcgtttact 9840

tcaaggcctg tgtcattaaa ccacagctgg ttgatgtagt ttgcaaactc tgagtttcct 9900

actctcggct ggccaaagtt gatcatggta ggatcatgtc caagtaattt gaagtgagtt 9960

gcgaaaagaa gagcttgagc agcgcccagc gagtgaccag taacatacat tttgtagtca 10020

gtgtggttgg tgaggaactt ttcaaactga ggagcagcat tgaccatagt ttcgttgaag 10080

gccttggcga acccatcatg gatcatgcag cctttgcact cactagtttt gatgggaata 10140

agacgggggt cttcaaccac cagagcccgc tctttgaggt tgtcaagacc tttgttctcc 10200

acttccaagt ctggtcggac tgcccatctc tgttaattaa gttgcgacac atgtcttgat 10260

agtatcttga attctctctc ttgagctttt ccataacaag ttcttctgcc tccaggaagt 10320

ccatgggtgg tttgatcatg gttttggtgt agtggtagtg cagtggtggt attgtgactg 10380

gggatgtagt tgagaataag tcatacacaa gtcagctttc ttcgagcctc atataagtat 10440

aagtagttca acgtattagc actgtaccca gcatctccgt atcgagaaac acaacaacat 10500

gccccattgg acagatcatg cggatacaca ggttgtgcag tatcatacat actcgatcag 10560

acaggtcgtc tgaccatcat acaagctgaa caagcgctcc atacttgcac gctctctata 10620

tacacagtta aattacatat ccatagtcta acctctaaca gttaatcttc tggtaagcct 10680

cccagccagc cttctggtat cgcttggcct cctcaatagg atctcggttc tggccgtaca 10740

gacctcggcc gacaattatg atatccgttc cggtagacat gacatcctca acagttcggt 10800

actgctgtcc gagagcgtct cccttgtcgt caagacccac cccgggggtc agaataagcc 10860

agtcctcaga gtcgccctta ggtcggttct gggcaatgaa gccaaccaca aactcggggt 10920

cggatcgggc aagctcaatg gtctgcttgg agtactcgcc agtggccaga gagcccttgc 10980

aagacagctc ggccagcatg agcagacctc tggccagctt ctcgttggga gaggggacta 11040

ggaactcctt gtactgggag ttctcgtagt cagagacgtc ctccttcttc tgttcagaga 11100

cagtttcctc ggcaccagct cgcaggccag caatgattcc ggttccgggt acaccgtggg 11160

cgttggtgat atcggaccac tcggcgattc ggtgacaccg gtactggtgc ttgacagtgt 11220

tgccaatatc tgcgaacttt ctgtcctcga acaggaagaa accgtgctta agagcaagtt 11280

ccttgagggg gagcacagtg ccggcgtagg tgaagtcgtc aatgatgtcg atatgggttt 11340

tgatcatgca cacataaggt ccgaccttat cggcaagctc aatgagctcc ttggtggtgg 11400

taacatccag agaagcacac aggttggttt tcttggctgc cacgagcttg agcactcgag 11460

cggcaaaggc ggacttgtgg acgttagctc gagcttcgta ggagggcatt ttggtggtga 11520

agaggagact gaaataaatt tagtctgcag aacttttat cggaacctta tctggggcag 11580

tgaagtatat gttatggtaa tagttacgag ttagttgaac ttatagatag actggactat 11640

acggctatcg gtccaaatta gaaagaacgt caatggctct ctgggcgtcg cctttgccga 11700

caaaaatgtg atcatgatga aagccagcaa tgacgttgca gctgatattg ttgtcggcca 11760

accgcgccga aaacgcagct gtcagaccca cagcctccaa cgaagaatgt atcgtcaaag 11820

tgatccaagc acactcatag ttggagtcgt actccaaagg cggcaatgac gagtcagaca 11880

gatactcgtc gacctttcc ttgggaacca ccaccgtcag cccttctgac tcacgtattg 11940

tagccaccga cacaggcaac agtccgtgga tagcagaata tgtcttgtcg gtccatttct 12000

caccaacttt aggcgtcaag tgaatgttgc agaagaagta tgtgccttca ttgagaatcg 12060

gtgttgctga tttcaataaa gtcttgagat cagtttggcc agtcatgttg tgggggtaa 12120

ttggattgag ttatcgccta cagtctgtac aggtatactc gctgcccact ttatactttt 12180

tgattccgct gcacttgaag caatgtcgtt taccaaaagt gagaatgctc cacagaacac 12240

accccagggt atggttgagc aaaaaataaa cactccgata cggggaatcg aaccccggtc 12300

tccacggttc tcaagaagta ttcttgatga gagcgtatcg atggttaatg ctgctgtgtg 12360

ctgtgtgtgt gtgttgtttg gcgctcattg ttgcgttatg cagcgtacac cacaatattg 12420

gaagcttatt agcctttcta ttttttcgtt tgcaaggctt aacaacattg ctgtggagag 12480

ggatggggat atggaggccg ctggagggag tcggagaggc gttttggagc ggcttggcct 12540

ggcgcccagc tcgcgaaacg cacctaggac cctttggcac gccgaaatgt gccacttttc 12600

agtctagtaa cgccttacct acgtcattcc atgcgtgcat gtttgcgcct tttttccctt 12660

gcccttgatc gccacacagt acagtgcact gtacagtgga ggttttgggg gggtcttaga 12720

tgggagctaa aagcggccta gcggtacact agtgggattg tatggagtgg catggagcct 12780

aggtggagcc tgacaggacg cacgaccggc tagcccgtga cagacgatgg gtggctcctg 12840

ttgtccaccg cgtacaaatg tttgggccaa agtcttgtca gccttgcttg cgaacctaat 12900

tcccaatttt gtcacttcgc acccccattg atcgagccct aacccctgcc catcaggcaa 12960

tccaattaag ctcgcattgt ctgccttgtt tagtttggct cctgcccgtt tcggcgtcca 13020

cttgcacaaa cacaaacaag cattatatat aaggctcgtc tctccctccc aaccacactc 13080

actttttgc ccgtcttccc ttgctaacac aaaagtcaag aacacaaaca accaccccaa 13140

ccccccttaca cacaagacat atctacagca atggc 13175

<210> 41
<211> 509
<212> PRT
<213> Schizochytrium aggregatum

<220>
<221> MISC_FEATURE
<222> (1)..(509)
<223> delta 4-desaturase

<300>
<302> DELTA 4-DESATURASE GENES AND USES THEREOF
<310> WO 2002/090493
<311> 2002-05-02
<312> 2002-11-14
<313> (1)..(509)

<400> 41

Met Thr Val Gly Gly Asp Glu Val Tyr Ser Met Ala Gln Val Arg Asp
1 5 10 15

His Asn Thr Pro Asp Asp Ala Trp Cys Ala Ile His Gly Glu Val Tyr
20 25 30

Glu Leu Thr Lys Phe Ala Arg Thr His Pro Gly Gly Asp Ile Ile Leu
35 40 45

Leu Ala Ala Gly Lys Glu Ala Thr Ile Leu Phe Glu Thr Tyr His Val
50 55 60

Arg Pro Ile Ser Asp Ala Val Leu Arg Lys Tyr Arg Ile Gly Lys Leu
65 70 75 80

```
Ala Ala Ala Gly Lys Asp Glu Pro Ala Asn Asp Ser Thr Tyr Tyr Ser
                85                  90                  95

Trp Asp Ser Asp Phe Tyr Lys Val Leu Arg Gln Arg Val Val Ala Arg
            100                 105                 110

Leu Glu Glu Arg Lys Ile Ala Arg Arg Gly Gly Pro Glu Ile Trp Ile
        115                 120                 125

Lys Ala Ala Ile Leu Val Ser Gly Phe Trp Ser Met Leu Tyr Leu Met
    130                 135                 140

Cys Thr Leu Asp Pro Asn Arg Gly Ala Ile Leu Ala Ala Ile Ala Leu
145                 150                 155                 160

Gly Ile Val Ala Ala Phe Val Gly Thr Cys Ile Gln His Asp Gly Asn
                165                 170                 175


His Gly Ala Phe Ala Phe Ser Pro Phe Met Asn Lys Leu Ser Gly Trp
            180                 185                 190

Thr Leu Asp Met Ile Gly Ala Ser Ala Met Thr Trp Glu Met Gln His
        195                 200                 205

Val Leu Gly His His Pro Tyr Thr Asn Leu Ile Glu Met Glu Asn Gly
    210                 215                 220

Thr Gln Lys Val Thr His Ala Asp Val Asp Pro Lys Lys Ala Asp Gln
225                 230                 235                 240

Glu Ser Asp Pro Asp Val Phe Ser Thr Tyr Pro Met Leu Arg Leu His
                245                 250                 255

Pro Trp His Arg Lys Arg Phe Tyr His Arg Phe Gln His Leu Tyr Ala
            260                 265                 270

Pro Leu Leu Phe Gly Phe Met Thr Ile Asn Lys Val Ile Thr Gln Asp
        275                 280                 285

Val Gly Val Val Leu Ser Lys Arg Leu Phe Gln Ile Asp Ala Asn Cys
    290                 295                 300

Arg Tyr Ala Ser Lys Ser Tyr Val Ala Arg Phe Trp Ile Met Lys Leu
305                 310                 315                 320

Leu Thr Val Leu Tyr Met Val Ala Leu Pro Val Tyr Thr Gln Gly Leu
                325                 330                 335

Val Asp Gly Leu Lys Leu Phe Phe Ile Ala His Phe Ser Cys Gly Glu
            340                 345                 350

Leu Leu Ala Thr Met Phe Ile Val Asn His Ile Ile Glu Gly Val Ser
        355                 360                 365

Tyr Ala Ser Lys Asp Ser Val Lys Gly Thr Met Ala Pro Pro Arg Thr
    370                 375                 380

Val His Gly Val Thr Pro Met His Asp Thr Arg Asp Ala Leu Gly Lys
385                 390                 395                 400

Glu Lys Ala Ala Thr Lys His Val Pro Leu Asn Asp Trp Ala Ala Val
```

```
                405                 410                 415

Gln Cys Gln Thr Ser Val Asn Trp Ser Ile Gly Ser Trp Phe Trp Asn
            420                 425                 430

His Phe Ser Gly Gly Leu Asn His Gln Ile Glu His His Leu Phe Pro
        435                 440                 445

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

Thr Cys Ala Glu Tyr Gly Val Pro Tyr Gln Ser Glu Gln Ser Leu Phe
465                 470                 475                 480

Ser Ala Tyr Phe Lys Met Leu Ser His Leu Arg Ala Leu Gly Asn Glu
                485                 490                 495

Pro Met Pro Ser Trp Glu Lys Asp His Pro Lys Ser Lys
            500                 505
```

<210> 42
<211> 445
<212> PRT
<213> Pavlova lutheri

<220>
<221> MISC_FEATURE
<222> (1)..(445)
<223> delta-4 desaturase; GenBank Accession No. AAQ98793

<400> 42

```
Met Pro Pro Ser Ala Ala Ser Glu Gly Gly Val Ala Glu Leu Arg Ala
1               5                   10                  15

Ala Glu Val Ala Ser Tyr Thr Arg Lys Ala Val Asp Glu Arg Pro Asp
            20                  25                  30

Leu Thr Ile Val Gly Asp Ala Val Tyr Asp Ala Lys Ala Phe Arg Asp
        35                  40                  45

Glu His Pro Gly Gly Ala His Phe Val Ser Leu Phe Gly Gly Arg Asp
    50                  55                  60

Ala Thr Glu Ala Phe Met Glu Tyr His Arg Arg Ala Trp Pro Lys Ala
65                  70                  75                  80

Arg Met Ser Lys Phe Phe Val Gly Ser Leu Asp Ala Ser Glu Lys Pro
                85                  90                  95

Thr Gln Ala Asp Ser Ala Tyr Leu Arg Leu Cys Ala Glu Val Asn Ala
            100                 105                 110

Leu Leu Pro Lys Gly Ser Gly Gly Phe Ala Pro Pro Ser Tyr Trp Leu
        115                 120                 125

Lys Ala Ala Ala Leu Val Val Ala Ala Val Ser Ile Glu Gly Tyr Met
    130                 135                 140

Leu Leu Arg Gly Lys Thr Leu Leu Leu Ser Val Phe Leu Gly Leu Val
145                 150                 155                 160
```

```
Phe Ala Trp Ile Gly Leu Asn Ile Gln His Asp Ala Asn His Gly Ala
                165                 170                 175

Leu Ser Arg His Ser Val Ile Asn Tyr Cys Leu Gly Tyr Ala Gln Asp
            180                 185                 190

Trp Ile Gly Gly Asn Met Val Leu Trp Leu Gln Glu His Val Val Met
        195                 200                 205

His His Leu His Thr Asn Asp Val Asp Ala Asp Pro Asp Gln Lys Ala
    210                 215                 220

His Gly Val Leu Arg Leu Lys Pro Thr Asp Gly Trp Met Pro Trp His
225                 230                 235                 240

Ala Leu Gln Gln Leu Tyr Ile Leu Pro Gly Glu Ala Met Tyr Ala Phe
                245                 250                 255

Lys Leu Leu Phe Leu Asp Ala Leu Glu Leu Leu Ala Trp Arg Trp Glu
            260                 265                 270

Gly Glu Lys Ile Ser Pro Leu Ala Arg Ala Leu Phe Ala Pro Ala Val
        275                 280                 285

Ala Cys Lys Leu Gly Phe Trp Ala Arg Phe Val Ala Leu Pro Leu Trp
    290                 295                 300

Leu Gln Pro Thr Val His Thr Ala Leu Cys Ile Cys Ala Thr Val Cys
305                 310                 315                 320

Thr Gly Ser Phe Tyr Leu Ala Phe Phe Phe Phe Ile Ser His Asn Phe
                325                 330                 335

Asp Gly Val Gly Ser Val Gly Pro Lys Gly Ser Leu Pro Arg Ser Ala
            340                 345                 350

Thr Phe Val Gln Arg Gln Val Glu Thr Ser Ser Asn Val Gly Gly Tyr
        355                 360                 365


Trp Leu Gly Val Leu Asn Gly Gly Leu Asn Phe Gln Ile Glu His His
    370                 375                 380

Leu Phe Pro Arg Leu His His Ser Tyr Tyr Ala Gln Ile Ala Pro Val
385                 390                 395                 400

Val Arg Thr His Ile Glu Lys Leu Gly Phe Lys Tyr Arg His Phe Pro
                405                 410                 415

Thr Val Gly Ser Asn Leu Ser Ser Met Leu Gln His Met Gly Lys Met
            420                 425                 430

Gly Thr Arg Pro Gly Ala Glu Lys Gly Gly Lys Ala Glu
        435                 440                 445
```

<210> 43
<211> 433
<212> PRT
<213> Isochrysis galbana

<220>
<221> MISC_FEATURE
<222> (1)..(433)
<223> delta-4 desaturase; GenBank Accession No. AAV33631

<400> 43

```
Met Cys Asn Ala Ala Gln Val Glu Thr Gln Ala Leu Arg Ala Lys Glu
1               5                  10                  15

Ala Ala Lys Pro Thr Trp Thr Lys Ile His Gly Arg Thr Val Asp Val
            20                  25                  30

Glu Thr Phe Arg His Pro Gly Gly Asn Ile Leu Asp Leu Phe Leu Gly
        35                  40                  45

Met Asp Ala Thr Thr Ala Phe Glu Thr Phe His Gly His His Lys Gly
    50                  55                  60

Ala Trp Lys Met Leu Lys Thr Leu Pro Glu Lys Glu Val Ala Ala Ala
65                  70                  75                  80

Asp Ile Pro Ala Gln Lys Glu Glu His Val Ala Glu Met Thr Arg Leu
                85                  90                  95

Met Ala Ser Trp Arg Glu Arg Gly Leu Phe Lys Pro Arg Pro Val Ala
            100                 105                 110

Ser Ser Ile Tyr Gly Leu Cys Val Ile Phe Ala Ile Ala Ala Ser Val
        115                 120                 125

Ala Cys Ala Pro Tyr Ala Pro Val Leu Ala Gly Ile Ala Val Gly Thr
    130                 135                 140

Cys Trp Ala Gln Cys Gly Phe Leu Gln His Met Gly Gly His Arg Glu
145                 150                 155                 160

Trp Gly Arg Thr Trp Ser Phe Ala Phe Gln His Leu Phe Glu Gly Leu
                165                 170                 175

Leu Lys Gly Gly Ser Ala Ser Trp Trp Arg Asn Arg His Asn Lys His
            180                 185                 190

His Ala Lys Thr Asn Val Leu Gly Glu Asp Gly Asp Leu Arg Thr Thr
        195                 200                 205

Pro Phe Phe Ala Trp Asp Pro Thr Leu Ala Lys Lys Val Pro Asp Trp
    210                 215                 220

Ser Leu Arg Thr Gln Ala Phe Thr Phe Leu Pro Ala Leu Gly Ala Tyr
225                 230                 235                 240

Val Phe Val Phe Ala Phe Thr Val Arg Lys Tyr Ser Val Val Lys Arg
                245                 250                 255

Leu Trp His Glu Val Ala Leu Met Val Ala His Tyr Ala Leu Phe Ser
            260                 265                 270

Trp Ala Leu Ser Ala Ala Gly Ala Ser Leu Ser Ser Gly Leu Thr Phe
        275                 280                 285

Tyr Cys Thr Gly Tyr Ala Trp Gln Gly Ile Tyr Leu Gly Phe Phe Phe
```

```
    290                 295                 300

Gly Leu Ser His Phe Ala Val Glu Arg Val Pro Ser Thr Ala Thr Trp
305                 310                 315                 320

Leu Glu Ser Thr Met Met Gly Thr Val Asp Trp Gly Gly Ser Ser Ala
                325                 330                 335

Phe Cys Gly Tyr Leu Ser Gly Phe Leu Asn Ile Gln Ile Glu His His
            340                 345                 350

Met Ala Pro Gln Met Pro Met Glu Asn Leu Arg Gln Ile Arg Ala Asp
        355                 360                 365

Cys Lys Ala Ala Ala His Lys Phe Gly Leu Pro Tyr Arg Glu Leu Thr
    370                 375                 380

Phe Val Ala Ala Thr Lys Leu Met Met Ser Gly Leu Tyr Arg Thr Gly
385                 390                 395                 400

Lys Asp Glu Leu Lys Leu Arg Ala Asp Arg Arg Lys Phe Thr Arg Ala
                405                 410                 415

Gln Ala Tyr Met Gly Ala Ala Ser Ala Leu Val Asp Thr Leu Lys Ala
            420                 425                 430

Asp
```

<210> 44
<211> 912
<212> DNA
<213> Euglena gracilis

<220>
<221> CDS
<222> (1)..(912)
<223> synthetic C20 elongase (codon-optimized for Yarrowia lipolytica) ("EgC20ES")

<300>
<302> MULTIZYMES AND THEIR USE IN MAKING POLYUNSATURATED FATTY ACIDS
<310> US-2008-0254191-A1
<311> 2008-04-03
<312> 2008-10-16
<313> (1)..(912)

<400> 44

```
atg gct gac tct ccc gtc atc aac ctc tcc acc atg tgg aag cct ctg         48
Met Ala Asp Ser Pro Val Ile Asn Leu Ser Thr Met Trp Lys Pro Leu
1               5                   10                  15

tcg ctc atg gcc ttg gat ctt gct gtt ctg gga cac gtc tgg aag cag         96
Ser Leu Met Ala Leu Asp Leu Ala Val Leu Gly His Val Trp Lys Gln
            20                  25                  30

gca caa cag gag ggc tcc atc tcg gct tac gcc gac tct gtg tgg act        144
Ala Gln Gln Glu Gly Ser Ile Ser Ala Tyr Ala Asp Ser Val Trp Thr
        35                  40                  45

ccc ctc atc atg tcc ggt ctg tac ctc tcc atg atc ttc gtg gga tgt        192
```

```
Pro Leu Ile Met Ser Gly Leu Tyr Leu Ser Met Ile Phe Val Gly Cys
    50                  55                  60

cga tgg atg aag aac cga gag ccc ttc gaa atc aag acc tac atg ttt        240
Arg Trp Met Lys Asn Arg Glu Pro Phe Glu Ile Lys Thr Tyr Met Phe
65                  70                  75                  80

gcc tac aac ctg tac cag acc ctc atg aac ctt tgc att gtg ctg ggc        288
Ala Tyr Asn Leu Tyr Gln Thr Leu Met Asn Leu Cys Ile Val Leu Gly
                85                  90                  95

ttc ctc tac cag gtc cac gct acc ggt atg cga ttc tgg gga tct ggc        336
Phe Leu Tyr Gln Val His Ala Thr Gly Met Arg Phe Trp Gly Ser Gly
            100                 105                 110

gtg gac cga tcg ccc aag ggt ctg gga att ggc ttt ttc atc tat gcc        384
Val Asp Arg Ser Pro Lys Gly Leu Gly Ile Gly Phe Phe Ile Tyr Ala
        115                 120                 125

cat tac cac aac aag tac gtc gag tac ttc gac aca ctc ttc atg gtg        432
His Tyr His Asn Lys Tyr Val Glu Tyr Phe Asp Thr Leu Phe Met Val
    130                 135                 140

ctg cgg aaa aag aac aac cag att tcc ttt ctt cac gtc tac cat cac        480
Leu Arg Lys Lys Asn Asn Gln Ile Ser Phe Leu His Val Tyr His His
145                 150                 155                 160

gct ctg ctc acc tgg gct tgg ttt gcc gtg gtc tac ttc gct cct gga        528
Ala Leu Leu Thr Trp Ala Trp Phe Ala Val Val Tyr Phe Ala Pro Gly
                165                 170                 175

ggt gac ggc tgg ttt gga gcc tgc tac aat tcc tcc att cat gtc ctg        576
Gly Asp Gly Trp Phe Gly Ala Cys Tyr Asn Ser Ser Ile His Val Leu
            180                 185                 190

atg tac tct tac tat ctg ctt gcc acc ttc ggc atc tcc tgt ccc tgg        624
Met Tyr Ser Tyr Tyr Leu Leu Ala Thr Phe Gly Ile Ser Cys Pro Trp
        195                 200                 205

aaa aag atc ctc acc cag ctg caa atg gtt cag ttc tgc ttt tgc ttc        672
Lys Lys Ile Leu Thr Gln Leu Gln Met Val Gln Phe Cys Phe Cys Phe
    210                 215                 220

acc cac tcg atc tac gtg tgg att tgc ggt tcc gaa atc tac cct cga        720
Thr His Ser Ile Tyr Val Trp Ile Cys Gly Ser Glu Ile Tyr Pro Arg
225                 230                 235                 240

ccc ttg act gct ctc cag tcc ttc gtg atg gtc aac atg ctg gtt ctc        768
Pro Leu Thr Ala Leu Gln Ser Phe Val Met Val Asn Met Leu Val Leu
                245                 250                 255

ttt ggc aac ttc tac gtc aag cag tat tct cag aag aat gga aag ccc        816
Phe Gly Asn Phe Tyr Val Lys Gln Tyr Ser Gln Lys Asn Gly Lys Pro
            260                 265                 270

gag aac ggt gcc act cct gag aac ggt gcc aag cct cag ccc tgc gag        864
Glu Asn Gly Ala Thr Pro Glu Asn Gly Ala Lys Pro Gln Pro Cys Glu
        275                 280                 285

aac ggc acc gtc gag aag cga gag aac gac act gcc aac gtt cga taa        912
Asn Gly Thr Val Glu Lys Arg Glu Asn Asp Thr Ala Asn Val Arg
    290                 295                 300
```

<210> 45
<211> 303
<212> PRT
<213> Euglena gracilis

<400> 45

Met Ala Asp Ser Pro Val Ile Asn Leu Ser Thr Met Trp Lys Pro Leu
1 5 10 15

Ser Leu Met Ala Leu Asp Leu Ala Val Leu Gly His Val Trp Lys Gln
20 25 30

Ala Gln Gln Glu Gly Ser Ile Ser Ala Tyr Ala Asp Ser Val Trp Thr
35 40 45

Pro Leu Ile Met Ser Gly Leu Tyr Leu Ser Met Ile Phe Val Gly Cys
50 55 60

Arg Trp Met Lys Asn Arg Glu Pro Phe Glu Ile Lys Thr Tyr Met Phe
65 70 75 80

Ala Tyr Asn Leu Tyr Gln Thr Leu Met Asn Leu Cys Ile Val Leu Gly
85 90 95

Phe Leu Tyr Gln Val His Ala Thr Gly Met Arg Phe Trp Gly Ser Gly
100 105 110

Val Asp Arg Ser Pro Lys Gly Leu Gly Ile Gly Phe Phe Ile Tyr Ala
115 120 125

His Tyr His Asn Lys Tyr Val Glu Tyr Phe Asp Thr Leu Phe Met Val
130 135 140

Leu Arg Lys Lys Asn Asn Gln Ile Ser Phe Leu His Val Tyr His His
145 150 155 160

Ala Leu Leu Thr Trp Ala Trp Phe Ala Val Val Tyr Phe Ala Pro Gly
165 170 175

Gly Asp Gly Trp Phe Gly Ala Cys Tyr Asn Ser Ser Ile His Val Leu
180 185 190

Met Tyr Ser Tyr Tyr Leu Leu Ala Thr Phe Gly Ile Ser Cys Pro Trp
195 200 205

Lys Lys Ile Leu Thr Gln Leu Gln Met Val Gln Phe Cys Phe Cys Phe
210 215 220

Thr His Ser Ile Tyr Val Trp Ile Cys Gly Ser Glu Ile Tyr Pro Arg
225 230 235 240

Pro Leu Thr Ala Leu Gln Ser Phe Val Met Val Asn Met Leu Val Leu
245 250 255

Phe Gly Asn Phe Tyr Val Lys Gln Tyr Ser Gln Lys Asn Gly Lys Pro
260 265 270

Glu Asn Gly Ala Thr Pro Glu Asn Gly Ala Lys Pro Gln Pro Cys Glu
275 280 285

```
Asn Gly Thr Val Glu Lys Arg Glu Asn Asp Thr Ala Asn Val Arg
    290                 295                 300


<210>  46
<211>  900
<212>  DNA
<213>  Euglena anabaena

<220>
<221>  CDS
<222>  (1)..(900)
<223>  synthetic C20 elongase (codon-optimized for Yarrowia lipolytica)
       ("EaC20ES")

<300>
<302>  MULTIZYMES AND THEIR USE IN MAKING POLYUNSATURATED FATTY ACIDS
<310>  US-2008-0254191-A1
<311>  2008-04-03
<312>  2008-10-16
<313>  (1)..(900)

<400>  46
atg gcc gag ggc aag tcc gac ggt ccc gtc gtt acc ctc cag tcc atg         48
Met Ala Glu Gly Lys Ser Asp Gly Pro Val Val Thr Leu Gln Ser Met
1               5                   10                  15

tgg aag ccc ctg gct ctc atg gcc atc gac gtc ggc atc ctg gtc aac         96
Trp Lys Pro Leu Ala Leu Met Ala Ile Asp Val Gly Ile Leu Val Asn
            20                  25                  30

gtg cga cgg aag gcc ttc acc gag ttc gac gga cac tcg aac gtc ttc        144
Val Arg Arg Lys Ala Phe Thr Glu Phe Asp Gly His Ser Asn Val Phe
        35                  40                  45

gcc gat ccc gtg tac att ccc ttt gtc atg aac ctg ttc tac ctc acc        192
Ala Asp Pro Val Tyr Ile Pro Phe Val Met Asn Leu Phe Tyr Leu Thr
    50                  55                  60

atg atc ttt gct ggc tgc cga tgg atg aag act cga gaa ccc ttc gag        240
Met Ile Phe Ala Gly Cys Arg Trp Met Lys Thr Arg Glu Pro Phe Glu
65                  70                  75                  80

atc aag tcc tac atg ttt gcc tac aac gct tac cag aca atg atg aac        288
Ile Lys Ser Tyr Met Phe Ala Tyr Asn Ala Tyr Gln Thr Met Met Asn
                85                  90                  95

ttt ctc att gtg gtc ggc ttc atg tat gag gtt cac tcc acc ggt atg        336
Phe Leu Ile Val Val Gly Phe Met Tyr Glu Val His Ser Thr Gly Met
            100                 105                 110

cga tac tgg gga tcc aga atc gac act tct acc aag ggc ttg gga ctg        384
Arg Tyr Trp Gly Ser Arg Ile Asp Thr Ser Thr Lys Gly Leu Gly Leu
        115                 120                 125

ggt ttc ctc atc tat gcc cat tac cac aac aag tac gtg gag tac gtc        432
Gly Phe Leu Ile Tyr Ala His Tyr His Asn Lys Tyr Val Glu Tyr Val
    130                 135                 140

gac acc ctg ttc atg att ctg cgg aag aaa aac aat cag atc tcg ttc        480
Asp Thr Leu Phe Met Ile Leu Arg Lys Lys Asn Asn Gln Ile Ser Phe
```

```
145                 150                 155                 160

ctt cac gtt tac cac cat tcc ctg ctc act tgg gca tgg tgg gct gtg        528
Leu His Val Tyr His His Ser Leu Leu Thr Trp Ala Trp Trp Ala Val
                165                 170                 175

gtc tac tgg gct cct ggc gga gat gcc tgg ttc ggt gcc tgt tac aac        576
Val Tyr Trp Ala Pro Gly Gly Asp Ala Trp Phe Gly Ala Cys Tyr Asn
            180                 185                 190

tcc ttc atc cac gtt ctc atg tac tcc tac tat ctg ttt gcc acc ttc        624
Ser Phe Ile His Val Leu Met Tyr Ser Tyr Tyr Leu Phe Ala Thr Phe
        195                 200                 205

ggc att cga tgt ccc tgg aaa aag atg ctc acc cag ttg caa atg gtc        672
Gly Ile Arg Cys Pro Trp Lys Lys Met Leu Thr Gln Leu Gln Met Val
    210                 215                 220

cag ttc tgc ttt tgc ttc gct cat gcc atg tac gtt gga tgg ctt ggt        720
Gln Phe Cys Phe Cys Phe Ala His Ala Met Tyr Val Gly Trp Leu Gly
225                 230                 235                 240

cac gag gtg tac cct cga tgg ctc act gct ctg cag gcc ttt gtg atg        768
His Glu Val Tyr Pro Arg Trp Leu Thr Ala Leu Gln Ala Phe Val Met
                245                 250                 255

ctc aac atg ctg gtc ctc ttt ggc aac ttc tac atg aag tct tac tcc        816
Leu Asn Met Leu Val Leu Phe Gly Asn Phe Tyr Met Lys Ser Tyr Ser
            260                 265                 270

aag gcg agc aag ctc gaa cca gcc tct ccc gtg tcg cct gcc tct ctt        864
Lys Ala Ser Lys Leu Glu Pro Ala Ser Pro Val Ser Pro Ala Ser Leu
        275                 280                 285

gct cag aag ccc ttc gag aac gcc aag gtc aag taa                       900
Ala Gln Lys Pro Phe Glu Asn Ala Lys Val Lys
    290                 295
```

<210> 47
<211> 299
<212> PRT
<213> Euglena anabaena

<400> 47

```
Met Ala Glu Gly Lys Ser Asp Gly Pro Val Val Thr Leu Gln Ser Met
1               5                   10                  15

Trp Lys Pro Leu Ala Leu Met Ala Ile Asp Val Gly Ile Leu Val Asn
            20                  25                  30

Val Arg Arg Lys Ala Phe Thr Glu Phe Asp Gly His Ser Asn Val Phe
        35                  40                  45

Ala Asp Pro Val Tyr Ile Pro Phe Val Met Asn Leu Phe Tyr Leu Thr
    50                  55                  60

Met Ile Phe Ala Gly Cys Arg Trp Met Lys Thr Arg Glu Pro Phe Glu
65                  70                  75                  80

Ile Lys Ser Tyr Met Phe Ala Tyr Asn Ala Tyr Gln Thr Met Met Asn
```

```
                85                  90                  95

Phe Leu Ile Val Val Gly Phe Met Tyr Glu Val His Ser Thr Gly Met
            100                 105                 110

Arg Tyr Trp Gly Ser Arg Ile Asp Thr Ser Thr Lys Gly Leu Gly Leu
        115                 120                 125

Gly Phe Leu Ile Tyr Ala His Tyr His Asn Lys Tyr Val Glu Tyr Val
    130                 135                 140

Asp Thr Leu Phe Met Ile Leu Arg Lys Lys Asn Asn Gln Ile Ser Phe
145                 150                 155                 160

Leu His Val Tyr His His Ser Leu Leu Thr Trp Ala Trp Trp Ala Val
                165                 170                 175

Val Tyr Trp Ala Pro Gly Gly Asp Ala Trp Phe Gly Ala Cys Tyr Asn
            180                 185                 190

Ser Phe Ile His Val Leu Met Tyr Ser Tyr Tyr Leu Phe Ala Thr Phe
        195                 200                 205

Gly Ile Arg Cys Pro Trp Lys Lys Met Leu Thr Gln Leu Gln Met Val
    210                 215                 220

Gln Phe Cys Phe Cys Phe Ala His Ala Met Tyr Val Gly Trp Leu Gly
225                 230                 235                 240

His Glu Val Tyr Pro Arg Trp Leu Thr Ala Leu Gln Ala Phe Val Met
                245                 250                 255

Leu Asn Met Leu Val Leu Phe Gly Asn Phe Tyr Met Lys Ser Tyr Ser
            260                 265                 270

Lys Ala Ser Lys Leu Glu Pro Ala Ser Pro Val Ser Pro Ala Ser Leu
        275                 280                 285

Ala Gln Lys Pro Phe Glu Asn Ala Lys Val Lys
    290                 295
```