

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
29 November 2007 (29.11.2007)

PCT

(10) International Publication Number
WO 2007/136646 A2

(51) International Patent Classification:

C07K 14/43 (2006.01)

(21) International Application Number:

PCT/US2007/011717

(22) International Filing Date: 16 May 2007 (16.05.2007)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/801,119 17 May 2006 (17.05.2006) US

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

(72) Inventors; and

(75) Inventors/Applicants (for US only): **DAMUDE, Howard, Glenn** [CA/US]; 4 Kenwick Road, Hockessin, Delaware 19707 (US). **POLLAK, Dana, M., Walters** [US/US]; 123C East Baker Street, Media, Pennsylvania 19063 (US). **XUE, Zhixiong** [CN/US]; 111 Harvey Lane, Chadds Ford, Pennsylvania 19317 (US). **ZHU, Quinn, Qun** [US/US]; 544 Revere Road, West Chester, Pennsylvania 19382 (US).

(74) Agent: **FELTHAM, S., Neil**; E. I. du Pont de Nemours and Company, Legal Patent Records Center, 4417 Lancaster Pike, Wilmington, Delaware 19805 (US).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, 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, 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, HU, IE, IS, IT, LT, LU, LV, MC, MT, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

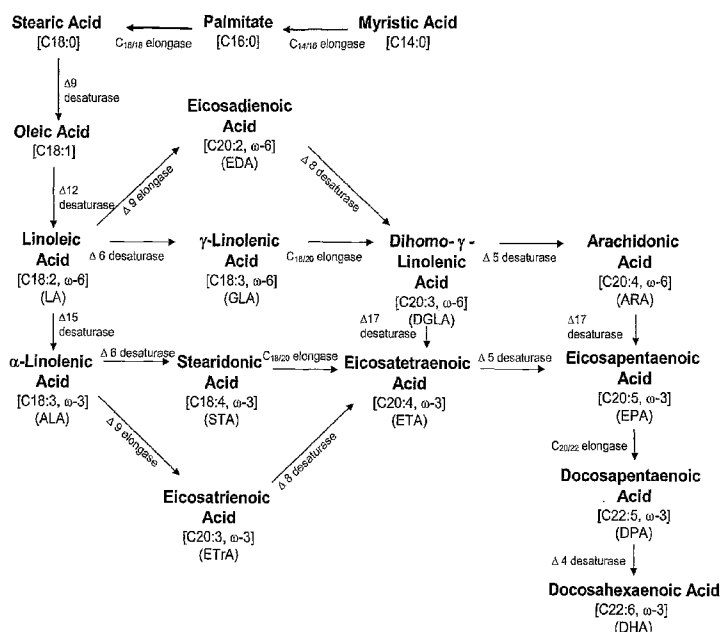
Published:

— without international search report and to be republished upon receipt of that report

— with sequence listing part of description published separately in electronic form and available upon request from the International Bureau

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: DELTA5 DESATURASE AND ITS USE IN MAKING POLYUNSATURATED FATTY ACIDS



(57) Abstract: The present invention relates to a $\Delta 5$ desaturase, which has the ability to convert dihomo- γ -linolenic acid (DGLA; 20:3 ω -6) to arachidonic acid (ARA; 20:4 ω -6) and/or eicosatetraenoic acid (ETA; 20:4 ω -3) to eicosapentaenoic acid (EPA; 20:5 ω -3). Isolated nucleic acid fragments and recombinant constructs comprising such fragments encoding $\Delta 5$ desaturase along with a method of making long chain polyunsaturated fatty acids (PUFAs) using this $\Delta 5$ desaturase in oleaginous yeast are disclosed.

WO 2007/136646 A2

TITLE

Δ 5 DESATURASE AND ITS USE IN MAKING POLYUNSATURATED FATTY ACIDS

This application claims the benefit of United States Provisional
5 Patent Application 60/801,119, filed May 17, 2006.

FIELD OF THE INVENTION

This invention is in the field of biotechnology. More specifically, this
invention pertains to the identification of nucleic acid fragments encoding a
 Δ 5 fatty acid desaturase enzyme and the use of this desaturase in making
10 long chain polyunsaturated fatty acids (PUFAs).

BACKGROUND OF THE INVENTION

The importance of PUFAs is undisputed. For example, certain
PUFAs are important biological components of healthy cells and are
recognized as: "essential" fatty acids that cannot be synthesized *de novo*
15 in mammals and instead must be obtained either in the diet or derived by
further desaturation and elongation of linoleic acid (LA; 18:2 ω -6) or α -
linolenic acid (ALA; 18:3 ω -3); constituents of plasma membranes of cells,
where they may be found in such forms as phospholipids or
triacylglycerols; necessary for proper development (particularly in the
20 developing infant brain) and for tissue formation and repair; and,
precursors to several biologically active eicosanoids of importance in
mammals (e.g., prostacyclins, eicosanoids, leukotrienes, prostaglandins).
Additionally, a high intake of long-chain ω -3 PUFAs produces
cardiovascular protective effects (Dyerberg, J. et al., *Amer. J. Clin. Nutr.*,
25 28:958-966 (1975); Dyerberg, J. et al., *Lancet*, 2(8081):117-119 (July 15,
1978); Shimokawa, H., *World Rev. Nutr. Diet*, 88:100-108 (2001); von
Schacky, C. and Dyerberg, J., *World Rev. Nutr. Diet*, 88:90-99 (2001)).
And, numerous other studies document wide-ranging health benefits
conferred by administration of ω -3 and/or ω -6 PUFAs against a variety of
30 symptoms and diseases (e.g., asthma, psoriasis, eczema, diabetes,
cancer).

A variety of different hosts including plants, algae, fungi 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 LA and ALA 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 arachidonic acid (ARA; 20:4 ω -6), eicosapentaenoic acid (EPA; 20:5 ω -3) and docosahexaenoic acid (DHA; 22:6 ω -3) may all require expression of a Δ 5 desaturase.

Most Δ 5 desaturase enzymes identified so far have the primary ability to convert dihomo- γ -linolenic acid (DGLA; 20:3 ω -6) to ARA, with secondary activity in converting eicosatetraenoic acid (ETA; 20:4 ω -3) to EPA (where DHA is subsequently synthesized from EPA following reaction with an additional $C_{20/22}$ elongase and a Δ 4 desaturase). The Δ 5 desaturase has a role in both the Δ 6 desaturase/ Δ 6 elongase pathway (which is predominantly found in algae, mosses, fungi, nematodes and humans and which is characterized by the production of γ -linoleic acid (GLA; 18:3 ω -6) and/or stearidonic acid (STA; 18:4 ω -3)) and the Δ 9 elongase/ Δ 8 desaturase pathway (which operates in some organisms, such as euglenoid species and which is characterized by the production of eicosadienoic acid (EDA; 20:2 ω -6) and/or eicosatrienoic acid (ETrA; 20:3 ω -3)) (Figure 1).

Based on the role Δ 5 desaturase enzymes play in the synthesis of e.g., ARA, EPA and DHA, there has been considerable effort to identify and characterize these enzymes from various sources. As such, numerous Δ 5 desaturases have been disclosed in both the open literature (e.g., GenBank Accession No. AF199596, No. AF226273, No. AF320509, No. AB072976, No. AF489588, No. AJ510244, No. AF419297, No. AF07879, No. AF067654 and No. AB022097) and the patent literature (e.g., U.S. Patents 5,972,664 and 6,075,183), see also, commonly owned, co-pending 60/801172, disclosing amino acid and nucleic acid sequences for a Δ 5 desaturase enzyme from *Euglena gracilis*, and No.

60/915733, disclosing amino acid and nucleic acid sequences for a $\Delta 5$ desaturase enzyme from *Euglena anabaena*.

Despite the disclosures cited above, it was desirable for the Applicants herein to identify and isolate additional genes encoding $\Delta 5$ desaturases that would be suitable for heterologous expression in a
5 variety of host organisms for use in the production of ω -3/ ω -6 fatty acids.

Applicants have solved the stated problem by isolating a gene encoding $\Delta 5$ desaturase from *Peridinium* sp. CCMP626.

SUMMARY OF THE INVENTION

10 The present invention relates to new genetic constructs encoding polypeptides having $\Delta 5$ desaturase activity, and their use in algae, bacteria, yeast, and fungi for the production of PUFAs.

Accordingly the invention provides an isolated nucleic acid molecule selected from the group consisting of:

- 15 (a) an isolated nucleotide sequence encoding a $\Delta 5$ desaturase enzyme as set forth in SEQ ID NO:2;
- (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
20 SSC, 0.1% SDS; and,
- (c) an isolated nucleotide sequence that is completely complementary to (a) or (b).

In a preferred embodiment the nucleic acid molecule encoding the $\Delta 5$ desaturase of the invention are codon optimized for expression in the
25 appropriate host.

In a preferred embodiment the invention provides an isolated nucleic acid molecule comprising a first nucleotide sequence encoding a $\Delta 5$ desaturase enzyme of at least 463 amino acids that has at least 95% identity based on BLASTP algorithms when compared to a polypeptide
30 having the sequence as set forth in SEQ ID NO:2;

or a second nucleotide sequence comprising the complement of the first nucleotide sequence.

The invention additionally provides genetic chimera comprising the nucleic acids encoding the $\Delta 5$ desaturase enzymes described herein as well as transformed host comprising the same.

In a preferred embodiment the invention provides a method for the production of arachidonic acid comprising:

- a) providing a host cell comprising:
 - (i) an isolated nucleotide molecule encoding a $\Delta 5$ desaturase polypeptide having at least 95% identity when compared to a polypeptide having the amino acid sequence as set forth in SEQ ID NO:2, based on BLASTP algorithms; and,
 - (ii) a source of dihomo- γ -linoleic acid;
- b) growing the host cell of step (a) under conditions wherein the nucleic acid fragment encoding the $\Delta 5$ desaturase polypeptide is expressed and the dihomo- γ -linoleic acid is converted to arachidonic acid; and,
- c) optionally recovering the arachidonic acid of step (b).

Alternatively the invention provides methods for the production of eicosapentaenoic acid comprising:

- a) providing a host cell comprising:
 - (i) an isolated nucleotide molecule encoding a $\Delta 5$ desaturase polypeptide having at least 95% identity when compared to a polypeptide having the amino acid sequence as set forth in SEQ ID NO:2, based on BLASTP algorithms; and,
 - (ii) a source of eicosatetraenoic acid;
- b) growing the host cell of step (a) under conditions wherein the nucleic acid fragment encoding the $\Delta 5$ desaturase polypeptide is expressed and the eicosatetraenoic acid is converted to eicosapentaenoic acid; and,
- c) optionally recovering the eicosapentaenoic acid of step (b).

In another embodiment the invention provides a method of obtaining a nucleic acid fragment encoding a $\Delta 5$ desaturase enzyme comprising:

- a.) probing a genomic library with:
 - 5 i.) an isolated nucleic acid fragment encoding an amino acid sequence as set forth in SEQ ID NO:2; or,
 - ii.) an isolated nucleic acid fragment that is complementary to (i);
- b.) identifying a DNA clone that hybridizes with the nucleic acid
10 fragment of step (a); and,
- c.) sequencing the genomic fragment that comprises the clone identified in step (b);

wherein the sequenced genomic fragment encodes a $\Delta 5$ desaturase enzyme.

- 15 Similarly the invention provides a method of obtaining a nucleic acid fragment encoding a $\Delta 5$ desaturase enzyme comprising:

- a.) synthesizing at least one oligonucleotide primer corresponding to a portion of the sequence selected from the group consisting of SEQ ID NOs:1 and 3; and,
- 20 b.) amplifying an insert present in a cloning vector using the oligonucleotide primer of step (a);

wherein the amplified insert encodes a portion of an amino acid sequence encoding a $\Delta 5$ desaturase enzyme.

BRIEF DESCRIPTION OF THE DRAWINGS AND SEQUENCE LISTINGS

25

Figure 1 illustrates the ω -3/ ω -6 fatty acid biosynthetic pathway.

Figure 2 shows a chromatogram of the lipid profile of a *Peridinium* sp. CCMP626 cell extract as described in Example 1.

Figure 3 shows a portion of an alignment between and among $\Delta 5$
30 desaturase proteins and $\Delta 8$ desaturase proteins using a Clustal W analysis (MegAlign™ program of DNASTAR software).

Figure 4 graphically represents the relationship between SEQ ID NOs:1, 2, 4, 5, 6, 8 and 10, each of which relates to the *Peridinium* sp. CCMP626 $\Delta 5$ desaturase.

Figure 5A illustrates the cloning strategy utilized for amplification of the *Peridinium* sp. CCMP626 $\Delta 5$ desaturase gene (RD5). Figure 5B is a plasmid map of pZUF17, while Figure 5C is a plasmid map of pDMW368.

Figure 6 provides plasmid maps for the following: (A) pKUNF12T6E; (B) pRD5S; and, (C) pZURD5S.

Figures 7A and 7B show a comparison of the DNA sequence of the *Peridinium* sp. CCMP626 $\Delta 5$ desaturase gene (designated as "RD5"; SEQ ID NO:1) and the synthetic gene (designated as "RD5S"; SEQ ID NO:3) codon-optimized for expression in *Yarrowia lipolytica*.

Figure 8A and 8B show a Clustal V alignment (with default parameters) of a *Pavlova lutheri* $\Delta 8$ desaturase (SEQ ID NO:18), a *Pavlova salina* $\Delta 8$ desaturase (SEQ ID NO:66), a *Euglena gracilis* $\Delta 8$ desaturase (SEQ ID NO:16) and two different *Rhizopus stolonifer* $\Delta 6$ fatty acid desaturases (SEQ ID NOs:53 and 65).

Figure 9 provides a plasmid map for pY98.

Figure 10A provides the fatty acid profiles for *Yarrowia lipolytica* expressing pY98 (SEQ ID NO:76; comprising a *Mortierella alpina* $\Delta 5$ desaturase gene designated as "MaD5") or pDMW368 (SEQ ID NO:23; comprising the *Peridinium* sp. CCMP626 $\Delta 5$ desaturase gene designated as "RD5") and fed various substrates. Figure 10B provides a comparison of the ω -3 and ω -6 substrate specificity of MaD5 versus RD5.

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

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.

- 5 SEQ ID NOs:1-26, 50, 51, 53-56, 63-72 and 75-76 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.
<i>Peridinium</i> sp. CCMP626 $\Delta 5$ desaturase ("RD5")	1 (1392 bp)	2 (463 AA)
Synthetic $\Delta 5$ desaturase, derived from <i>Peridinium</i> sp. CCMP626, codon-optimized for expression in <i>Yarrowia lipolytica</i> ("RD5S")	3 (1392 bp)	2 (463 AA)
<i>Peridinium</i> sp. CCMP626—fragment of pT-12-D5	4 (563 bp)	5 (187 AA)
<i>Peridinium</i> sp. CCMP626-- fragment of pT-RD5-5'C2	6 (693 bp)	--
<i>Peridinium</i> sp. CCMP626--5' sequence relative to SEQ ID NO:4	7 (511 bp)	--
<i>Peridinium</i> sp. CCMP626--fragment of pT-RD5-5'2 nd	8 (358 bp)	--
<i>Peridinium</i> sp. CCMP626--5' sequence relative to SEQ ID NO:6	9 (161 bp)	--
<i>Peridinium</i> sp. CCMP626-- fragment of pT-RD5-3'	10 (299 bp)	--
<i>Peridinium</i> sp. CCMP626—3' sequence relative to SEQ ID NO:4	11 (247 bp)	--
<i>Pythium irregulare</i> $\Delta 5$ desaturase (GenBank Accession No. AAL13311)	--	12 (456 AA)
<i>Phytophthora megasperma</i> $\Delta 5$ desaturase (GenBank Accession No. CAD53323)	--	13 (477 AA)
<i>Phaeodactylum tricornutum</i> $\Delta 5$ desaturase (GenBank Accession No. AAL92562)	--	14 (469 AA)
<i>Dictyostelium discoideum</i> $\Delta 5$ desaturase (GenBank Accession No. XP_640331)	--	15 (467 AA)
<i>Euglena gracilis</i> $\Delta 8$ desaturase (PCT Publications No. WO 2006/012325 and No. WO 2006/012326)	--	16 (421 AA)
<i>Pavlova lutheri</i> (CCMP459) $\Delta 8$ desaturase	17 (1269 bp)	18 (423 AA)
Conserved Region 1	--	19

		(7 AA)
Conserved Region 2	--	20 (7 AA)
<i>Thalassiosira pseudonana</i> $\Delta 8$ sphingolipid desaturase (GenBank Accession No. AAX14502)	--	21 (476 AA)
Plasmid pZUF17	22 (8165 bp)	--
Plasmid pDMW368	23 (8480 bp)	--
Plasmid pKUNF12T6E	24 (12,649 bp)	--
Synthetic C _{18/20} elongase gene derived from <i>Thraustochytrium aureum</i> (U.S. Patent 6,677,145), codon-optimized for expression in <i>Yarrowia lipolytica</i> ("EL2S")	25 (819 bp)	26 (272 AA)
Plasmid pRD5S	50 (4112 bp)	--
Plasmid pZURD5S	51 (8480 bp)	--
<i>Rhizopus stolonifer</i> $\Delta 6$ fatty acid desaturase (NCBI Accession No. AAX22052)	--	53 (459 AA)
<i>Pavlova lutheri</i> $\Delta 8$ desaturase--portion of cDNA insert from clone eps1c.pk002.f22 (5' end of cDNA insert)	54 (695 bp)	--
<i>Pavlova lutheri</i> $\Delta 8$ desaturase--fully sequenced EST eps1c.pk002.f22: fis (full insert sequence)	55 (1106 bp)	--
<i>Pavlova lutheri</i> $\Delta 8$ desaturase--translation of nucleotides 1-864 of fully sequenced EST eps1c.pk002.f22: fis (full insert sequence; SEQ ID NO:55)	--	56 (287 AA)
<i>Pavlova lutheri</i> $\Delta 8$ desaturase--full 5' end sequence from genome walking	63 (1294 bp)	--
<i>Pavlova lutheri</i> $\Delta 8$ desaturase--assembled sequence	64 (1927 bp)	--
<i>Rhizopus stolonifer</i> $\Delta 6$ fatty acid desaturase (NCBI Accession No. ABB96724)	--	65 (459 AA)
<i>Pavlova salina</i> $\Delta 8$ desaturase	--	66 (427 AA)
<i>Mortierella alpina</i> $\Delta 5$ desaturase	67 (1338 bp)	68 (446 AA)
Plasmid pY5-22	69 (6473 bp)	--
Plasmid pY5-22GPD	70 (6970 bp)	--

<i>Yarrowia lipolytica</i> glyceraldehyde-3-phosphate dehydrogenase promoter (GPD)	71 (968 bp)	--
Plasmid pYZDE2-S	72 (8630 bp)	--
Plasmid pKR136	75 (6339 bp)	--
Plasmid pY98	76 (8319 bp)	--

SEQ ID NOs:27-29 correspond to AP primer, Smart IV oligonucleotide primer and CDSIII 5' primer, respectively, used for *Peridinium* sp. CCMP626 cDNA synthesis.

- 5 SEQ ID NOs:30-33 correspond to degenerate oligonucleotide primers 5-1A, 5-1B, 5-1C and 5-1D, respectively, that encode Conserved Region 1.

- SEQ ID NOs:34-37 correspond to degenerate oligonucleotide primers 5-4AR, 5-4BR, 5-4CR and 5-4DR, respectively, that encode
10 Conserved Region 2.

 SEQ ID NOs:38-42 correspond to primers ODMW520, ODMW521, DNR CDS 5', ODMW541 and ODMW542, respectively, used for 5' RACE.

 SEQ ID NOs:43-45 correspond to primers ODMW523, AUAP and ODMW524, respectively, used for 3' RACE.

- 15 SEQ ID NOs:46-49 correspond to primers YL807, YL810, YL808 and YL809, respectively, used for amplification of the full length cDNA of RD5.

 SEQ ID NO:52 corresponds to primer T7, used for sequencing the *Pavlova lutheri* (CCMP459) cDNA library.

- 20 SEQ ID NOs:57 and 58 correspond to primers SeqE and SeqW, respectively, used for sequencing *Pavlova lutheri* (CCMP459) clones.

 SEQ ID NOs:59 and 60 correspond to the universal primer AP1 and primer GSP PvDES, respectively, used for amplification of genomic *Pavlova lutheri* (CCMP459) DNA.

- 25 SEQ ID NOs:61 and 62 correspond to primers M13-28Rev and PavDES seq, respectively, used for sequencing *Pavlova lutheri* (CCMP459) genomic inserts.

SEQ ID NOs:73 and 74 are primers GPDsense and GPDantisense, respectively, used for amplifying the GPD promoter.

DETAILED DESCRIPTION OF THE INVENTION

All patents, patent applications, and publications cited herein are
5 incorporated by reference in their entirety. This specifically includes the
following Applicants' Assignee's co-pending applications: U.S. Patent
7,125,672, U.S. Patent 7,189,559, U.S. Patent 7,192,762, U.S. Patent
7,198,937, U.S. Patent 7,202,356, U.S. Patent Applications No. 10/840579
and No. 10/840325 (filed May 6, 2004), U.S. Patent Application No.
10 10/869630 (filed June 16, 2004), U.S. Patent Application No. 10/882760
(filed July 1, 2004), U.S. Patent Applications No. 10/985254 and No.
10/985691 (filed November 10, 2004), U.S. Patent Application No.
11/024544 (filed December 29, 2004), U.S. Patent Application No.
11/166993 (filed June 24, 2005), U.S. Patent Application No. 11/183664
15 (filed July 18, 2005), U.S. Patent Application No. 11/185301 (filed July 20,
2005), U.S. Patent Application No. 11/190750 (filed July 27, 2005), U.S.
Patent Application No. 11/198975 (filed August 8, 2005), U.S. Patent
Application No. 11/225354 (filed September 13, 2005), U.S. Patent
Application No. 11/253882 (filed October 19, 2005), U.S. Patent
20 Applications No. 11/264784 and No. 11/264737 (filed November 1, 2005),
U.S. Patent Application No. 11/265761 (filed November 2, 2005), U.S.
Patent Application No. 60/795810 (filed April 28, 2006), U.S. Patent
Application No. 60/793575 (filed April 20, 2006), U.S. Patent Application
No. 60/796637 (filed May 2, 2006), U.S. Patent Applications No.
25 60/801172 and No. 60/801119 (filed May 17, 2006), U.S. Patent
Application No. 60/853563 (filed October 23, 2006), U.S. Patent
Application No. 60/855177 (filed October 30, 2006), U.S. Patent
Applications No. 11/601563 and No. 11/601564 (filed November 16,
2006), U.S. Patent Application No. 11/635258 (filed December 7, 2006),
30 U.S. Patent Application No. 11/613420 (filed December 20, 2006), U.S.
Patent Application No. 60/909790 (filed April 3, 2007), U.S. Patent
Application No. 60/911925 (filed April 16, 2007), U.S. Patent Application

No. 60/910831 (filed April 10, 2007) and U.S. Patent Application No. 60/915733, (filed May 3, 2007).

In accordance with the subject invention, Applicants identify a novel *Peridinium* sp. CCMP626 $\Delta 5$ 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.

PUFAs, or derivatives thereof, made by the methodology disclosed herein can be used as dietary substitutes, or supplements, particularly infant formulas, for patients undergoing intravenous feeding or for preventing or treating malnutrition. Alternatively, the purified PUFAs (or derivatives thereof) may be incorporated into cooking oils, fats or margarines formulated so that in normal use the recipient would receive the desired amount for dietary supplementation. The 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 (human or veterinary).

Supplementation of humans or animals with PUFAs produced by recombinant means can result in increased levels of the added PUFAs, as well as their metabolic progeny. For example, treatment with EPA can result not only in increased levels of EPA, but also downstream products of EPA such as eicosanoids (i.e., prostaglandins, leukotrienes, thromboxanes). Complex regulatory mechanisms can make it desirable to combine various PUFAs, or add different conjugates of PUFAs, in order to prevent, control or overcome such mechanisms to achieve the desired levels of specific PUFAs in an individual.

Definitions

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

"Open reading frame" is abbreviated ORF.

"Polymerase chain reaction" is abbreviated PCR.

"American Type Culture Collection" is abbreviated ATCC.

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

"Triacylglycerols" are abbreviated TAGs.

The term "invention" or "present invention" as used herein is not meant to be limiting to any one specific embodiment of the invention but
5 applies generally to any and all embodiments of the invention as described in the claims and specification.

The term "fatty acids" refers to long chain aliphatic acids (alkanoic acids) of varying chain lengths, from about C₁₂ to C₂₂ (although both longer and shorter chain-length acids are known). The predominant chain
10 lengths are between C₁₆ and C₂₂. 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",
15 "monounsaturated fatty acids" versus "polyunsaturated fatty acids" (or "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. Patent Publication No. 2005/0136519.

Nomenclature used to describe PUFAs in the present disclosure is
20 shown below 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
25 names of ω -3 and ω -6 fatty acids and their precursors, the abbreviations that will be used throughout the specification and each compounds' chemical name.

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	--	<i>cis</i> -9-octadecenoic	18:1
Linoleic	LA	<i>cis</i> -9, 12-octadecadienoic	18:2 ω -6
γ -Linoleic	GLA	<i>cis</i> -6, 9, 12-octadecatrienoic	18:3 ω -6
Eicosadienoic	EDA	<i>cis</i> -11, 14-eicosadienoic	20:2 ω -6
Dihomo- γ -Linoleic	DGLA	<i>cis</i> -8, 11, 14-eicosatrienoic	20:3 ω -6
Arachidonic	ARA	<i>cis</i> -5, 8, 11, 14-eicosatetraenoic	20:4 ω -6
α -Linolenic	ALA	<i>cis</i> -9, 12, 15-octadecatrienoic	18:3 ω -3
Stearidonic	STA	<i>cis</i> -6, 9, 12, 15-octadecatetraenoic	18:4 ω -3
Eicosatrienoic	ETra or ERA	<i>cis</i> -11, 14, 17-eicosatrienoic	20:3 ω -3
Sciadonic	SCI	<i>cis</i> -5,11,14-eicosatrienoic	20:3b ω -6
Juniperonic	JUP	<i>cis</i> -5,11,14,17-eicosatetraenoic	20:4b ω -3
Eicosa-tetraenoic	ETA	<i>cis</i> -8, 11, 14, 17-eicosatetraenoic	20:4 ω -3
Eicosa-pentaenoic	EPA	<i>cis</i> -5, 8, 11, 14, 17-eicosapentaenoic	20:5 ω -3
Docosa-pentaenoic	DPA	<i>cis</i> -7, 10, 13, 16, 19-docosapentaenoic	22:5 ω -3
Docosa-hexaenoic	DHA	<i>cis</i> -4, 7, 10, 13, 16, 19-docosahexaenoic	22:6 ω -3

The terms "triacylglycerol", "oil" and "TAGs" refer to neutral lipids composed of three fatty acyl residues esterified to a glycerol molecule (and such terms will be used interchangeably throughout the present disclosure herein). Such oils can contain long chain PUFAs, as well as shorter saturated and unsaturated fatty acids and longer chain saturated fatty acids. Thus, "oil biosynthesis" generically refers to the synthesis of TAGs in the cell.

“Percent (%) PUFAs in the total lipid and oil fractions” refers to the percent of PUFAs relative to the total fatty acids in those fractions. The term “total lipid fraction” or “lipid fraction” both refer to the sum of all lipids (i.e., neutral and polar) within an oleaginous organism, thus including
5 those lipids that are located in the phosphatidylcholine (PC) fraction, phosphatidyletanolamine (PE) fraction and triacylglycerol (TAG or oil) fraction. However, the terms “lipid” and “oil” will be used interchangeably throughout the specification.

A metabolic pathway, or biosynthetic pathway, in a biochemical
10 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 (then called a flux generating step). Many of these pathways are elaborate, and involve a step by step modification of
15 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 LA, EDA, GLA, DGLA, ARA, ALA, STA, ETrA, ETA, EPA, DPA and DHA. This process is well described in the
20 literature (e.g., see PCT Publication No. WO 2006/052870). Briefly, this process involves elongation of the carbon chain through the addition of carbon atoms and desaturation of the molecule through the addition of double bonds, via a series of special desaturation and elongation enzymes (i.e., “PUFA biosynthetic pathway enzymes”) present in the endoplasmic
25 reticulim 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 $\Delta 4$ desaturase, a $\Delta 5$ desaturase, a $\Delta 6$ desaturase, a $\Delta 12$ desaturase, a $\Delta 15$ desaturase, a $\Delta 17$ desaturase, a $\Delta 9$ desaturase, a $\Delta 8$ desaturase, a $\Delta 9$
30 elongase, a $C_{14/16}$ elongase, a $C_{16/18}$ elongase, a $C_{18/20}$ elongase and/or a $C_{20/22}$ elongase.

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 Figure 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 where one portion will generate ω -3 fatty acids and the other portion, only ω -6 fatty acids. That portion that only generates ω -3 fatty acids will be referred to herein as the ω -3 fatty acid biosynthetic pathway, whereas that portion that generates only ω -6 fatty acids will be referred to herein as the ω -6 fatty acid biosynthetic pathway.

The term "functional" as used herein in context with 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 the genes listed in the above paragraph are required, as a number of fatty acid products will only require the expression of a subset of the genes of this pathway.

The term " Δ 6 desaturase/ Δ 6 elongase pathway" will refer to a PUFA biosynthetic pathway that minimally includes at least one Δ 6 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, EPA, DPA and DHA may also be synthesized.

The term " Δ 9 elongase/ Δ 8 desaturase pathway" will refer to a PUFA biosynthetic pathway that minimally includes at least one Δ 9 elongase and at least one Δ 8 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, 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 $\Delta 9$ elongase/ $\Delta 8$ 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 "by-product fatty acid" refers to any fatty acid produced in a fatty acid metabolic pathway that is not the intended fatty acid product of the pathway nor an "intermediate fatty acid" of the pathway. For instance, when EPA is produced using the $\Delta 9$ elongase/ $\Delta 8$ desaturase pathway, sciadonic acid (SCI) and juniperonic acid (JUP) also can be produced by the action of a $\Delta 5$ desaturase on either EDA or ETrA, respectively. They are considered to be "by-product fatty acids" since neither can be further converted to EPA by the action of other metabolic pathway enzymes.

The term "desaturase" refers to a polypeptide that can desaturate, i.e., introduce a double bond, in one or more fatty acids to produce 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 $\Delta 5$ desaturases that catalyze the conversion of DGLA to ARA and/or ETA to EPA. Other desaturases include: 1.) $\Delta 17$ 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 ARA to EPA and/or DGLA to ETA; 2.) $\Delta 6$ desaturases that catalyze the conversion of LA to GLA and/or ALA to STA; 3.) $\Delta 12$ desaturases that catalyze the conversion of oleic acid to LA; 4.) $\Delta 15$ desaturases that catalyze the conversion of LA to ALA and/or GLA to STA; 5.) $\Delta 4$ desaturases that catalyze the conversion of DPA to DHA; 6.) $\Delta 8$ desaturases that catalyze the conversion of EDA to DGLA and/or ETrA

to ETA; and, 7.) $\Delta 9$ desaturases that catalyze the conversion of palmitate to palmitoleic acid (16:1) and/or stearate to oleic acid. In the art, $\Delta 15$ and $\Delta 17$ 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). In some embodiments, it is most 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.

For the purposes herein, the term "RD5" refers to a $\Delta 5$ desaturase enzyme (SEQ ID NO:2) isolated from *Peridinium* sp. CCMP626, encoded by SEQ ID NO:1 herein. Similarly, the term "RD5S" refers to a synthetic $\Delta 5$ desaturase derived from *Peridinium* sp. CCMP626 that is codon-optimized for expression in *Yarrowia lipolytica* (i.e., SEQ ID NOs:3 and 2).

The terms "conversion efficiency" and "percent substrate conversion" refer to the efficiency by which a particular enzyme (e.g., a desaturase) can convert substrate to product. The conversion efficiency is measured according to the following formula: $[\text{product}]/[\text{substrate} + \text{product}] * 100$, 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 U.S. Patent Publication No. 2005/0132442 and PCT Publication No. WO 2005/047480. Examples of reactions catalyzed by elongase systems are the conversion of GLA to DGLA, 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 will utilize a C_{14} substrate (e.g., myristic acid), a $C_{16/18}$ elongase will utilize a C_{16} substrate

(e.g., palmitate), a C_{18/20} elongase (also known as a $\Delta 6$ elongase as the terms can be used interchangeably) will utilize a C₁₈ substrate (e.g., GLA, STA) and a C_{20/22} elongase will utilize a C₂₀ substrate (e.g., EPA). In like manner, a $\Delta 9$ 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 (e.g., thereby acting as both a C_{16/18} elongase and a C_{18/20} elongase).

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. 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*, *Rhodospiridium*, *Cryptococcus*, *Trichosporon* and *Lipomyces*.

As used herein, the terms "isolated nucleic acid fragment" or "isolated nucleic acid molecule" will be 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), particularly Chapter 11 and Table 11.1 therein (entirely incorporated herein by reference). 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 instant specification teaches the complete amino acid and nucleotide sequence encoding encoding particular algal 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 instant invention comprises the complete sequences as reported in the accompanying Sequence Listing, as well as substantial portions of those sequences as defined above.

5 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, the invention herein also includes 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.

15 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 of the present invention 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.

25 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 (e.g., 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 to any of the nucleic acid sequences disclosed herein.

30 "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, the instant invention relates to any nucleic acid fragment that encodes all or a

substantial portion of the amino acid sequence encoding the instant algal polypeptide as set forth in SEQ ID NO:2. 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
5 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.

"Chemically synthesized", as related to a sequence of DNA, means that the component nucleotides were assembled *in vitro*. Manual chemical
10 synthesis of DNA may be accomplished using well-established procedures or, automated chemical synthesis can be performed using one of a number of commercially available machines. "Synthetic genes" can be assembled from oligonucleotide building blocks that are chemically synthesized using procedures known to those skilled in the art. These
15 building blocks are ligated and annealed 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
20 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.

"Gene" refers to a nucleic acid fragment that expresses a specific protein, and that may refer to the coding region alone or may include
25 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,
30 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, and which influence the transcription, RNA processing or stability, or translation of the associated coding sequence. Regulatory sequences may include promoters, translation leader sequences, 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 that 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. Patent 5,107,065; PCT Publication 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

the nucleic acid fragments of the invention. Expression may also refer to translation of mRNA into a polypeptide.

5 “Mature” protein refers to a post-translationally processed polypeptide, i.e., one from which any pre- or propeptides present in the primary translation product have been removed. “Precursor” protein refers to the primary product of translation of mRNA, i.e., with pre- and propeptides still present. Pre- and propeptides may be (but are not limited to) intracellular localization signals.

10 “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.

15 The terms “plasmid”, “vector” and “cassette” 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
20 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 a promoter fragment and DNA sequence for a selected gene product along with appropriate 3' untranslated sequence into a cell.

25 “Expression cassette” refers to a specific vector containing a foreign gene and having elements in addition to the foreign gene that allow for enhanced expression of that gene in a foreign host.

30 The term “percent identity”, as known in the art, is a relationship between two or more polypeptide sequences or two or more polynucleotide sequences, as determined by comparing the sequences. In the art, “identity” also means the degree of sequence relatedness between polypeptide or polynucleotide sequences, as the case may be, as determined by the match between strings of such sequences. “Identity”

and "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.) Humana: 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 identity are designed to give the best match between the sequences tested. Methods to determine identity and 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" corresponding to the alignment method labeled Clustal V (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.). For multiple alignments, 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. After alignment of the sequences using the Clustal V program, it is possible to obtain a "percent identity" by viewing the "sequence distances" table in the same program. Additionally the "Clustal W method of alignment" is available and corresponds to the alignment method labeled Clustal W (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™ v6.1 program of the LASERGENE bioinformatics computing suite (DNASTAR Inc.). Default parameters for multiple alignment correspond to GAP PENALTY=10, GAP LENGTH PENALTY=0.2, Delay Divergen Seqs(%)=30, DNA Transition Weight=0.5, Protein Weight Matrix=Gonnet Series, DNA Weight Matrix=IUB. After alignment of the sequences using the Clustal W program, it is possible to obtain a "percent identity" by viewing the "sequence distances" table in the same program.

It is well understood by one skilled in the art that many levels of sequence identity are useful in identifying polypeptides, from other species, wherein such polypeptides have the same or similar function or activity. Suitable nucleic acid fragments (isolated polynucleotides of the present invention) 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 67% to 100% may be useful in describing the present invention, such as 68%, 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 "conserved domain" or "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 will 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 the context of this application it will be understood that where sequence analysis software is used for analysis, that the results of the analysis will be 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. With regard to the BLASTP algorithm used herein default parameters will include the Robinson and Robinson amino acid frequencies (Robinson A.B., Robinson L.R., *Proc. Natl Acad. Sci. U.S.A.*, 88:8880-8884 (1991)), the BLOSUM62 scoring matrix and the gap cost $\Delta(g) = 11 + g$.

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., Bennis, 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).

10 An Overview: Microbial Biosynthesis Of Fatty Acids and Triacylglycerols

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 PCT Publication No. WO 2004/101757. Palmitate is the precursor of longer-chain saturated and unsaturated fatty acid derivatives, which are formed through the action of elongases and desaturases (Figure 1).

20 TAGs (the primary storage unit for fatty acids) are formed by a series of reactions that involve: 1.) the esterification of one molecule of acyl-CoA to glycerol-3-phosphate via an acyltransferase to produce lysophosphatidic acid; 2.) the 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.) 25 the 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

30 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 Figure 1 and as described below, there are often multiple alternate pathways for production of a specific ω -3/ ω -6 fatty acid.

Specifically, all pathways require the initial conversion of oleic acid to LA, the first of the ω -6 fatty acids, by a Δ 12 desaturase. Then, using the "Δ6 desaturase/Δ6 elongase pathway", ω -6 fatty acids are formed as follows: (1) LA is converted to GLA by a Δ6 desaturase; (2) GLA is converted to DGLA by a C_{18/20} elongase; and, (3) DGLA is converted to ARA by a Δ5 desaturase. Alternatively, the "Δ6 desaturase/Δ6 elongase pathway" can be utilized for formation of 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 STA by a Δ6 desaturase; (3) STA is converted to ETA by a C_{18/20} elongase; (4) ETA is converted to EPA by a Δ5 desaturase; (5) EPA is converted to DPA by a C_{20/22} elongase; and, (6) DPA is converted to 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.

Alternate pathways for the biosynthesis of ω -3/ ω -6 fatty acids utilize a Δ9 elongase and Δ8 desaturase. More specifically, LA and ALA may be converted to EDA and ETrA, respectively, by a Δ9 elongase; then, a Δ8 desaturase converts EDA to DGLA and/or ETrA to ETA.

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 (from bacteria, algae, fungi, 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 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; and/or, 4.) co-factors required by the polypeptide. The expressed polypeptide preferably has parameters compatible with the biochemical environment of its location in the host cell (see PCT Publication No. WO 2004/101757 for additional details).

In additional embodiments, it will 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 will typically be a mixture of various PUFAs consisting of the desired ω -3/ ω -6 fatty acid, as well as various upstream intermediary PUFAs. Thus, each enzyme's conversion efficiency is also a variable to consider, when optimizing biosynthesis of a desired fatty acid.

With each of the considerations above in mind, candidate genes having the appropriate desaturase and elongase activities (e.g., $\Delta 6$ desaturases, $C_{18/20}$ elongases, $\Delta 5$ desaturases, $\Delta 17$ desaturases, $\Delta 15$ desaturases, $\Delta 9$ desaturases, $\Delta 12$ desaturases, $C_{14/16}$ elongases, $C_{16/18}$ elongases, $\Delta 9$ elongases, $\Delta 8$ desaturases, $\Delta 4$ desaturases and $C_{20/22}$ elongases) can be identified according to publicly available literature (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 *Peridinium* sp. CCMP626 $\Delta 5$ Desaturase

In the present invention, a nucleotide sequence (SEQ ID NO:1) has been isolated from *Peridinium* sp. CCMP626 encoding a $\Delta 5$ desaturase (SEQ ID NO:2), designated herein as "RD5".

Comparison of the RD5 nucleotide base and deduced amino acid sequences to public databases reveals that the most similar known

sequences are about 67% identical to the amino acid sequence of RD5 reported herein over a length of 463 amino acids using a BLASTP search algorithm. 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 RD5 encoding nucleic acid sequences corresponding to the instant ORF are those encoding active proteins and which are at least about 70%-80% identical to the nucleic acid sequences of RD5 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.

In alternate embodiments, the instant RD5 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.

In one preferred embodiment of the invention herein, RD5 was codon-optimized for expression in *Yarrowia lipolytica*. This was possible by first determining the *Y. lipolytica* codon usage profile (see PCT Publication No. WO 04/101757 and U.S. Patent 7,125,672) and identifying those codons that were preferred. Then, for further optimization of gene expression in *Y. lipolytica*, the consensus sequence around the 'ATG' initiation codon was determined. This optimization resulted in modification of 247 bp of the 1392 bp coding region (17.7%) and optimization of 229 codons of the total 463 codons (49.4%). None of the modifications in the

codon-optimized gene ("RD5S"; SEQ ID NO:3) changed the amino acid sequence of the encoded protein (SEQ ID NO:2). As described in Example 11, the codon-optimized gene was 8.9% more efficient desaturating DGLA to ARA than the wildtype gene, when expressed in *Y. lipolytica*.

One skilled in the art would be able to use the teachings herein to create various other codon-optimized $\Delta 5$ desaturase proteins suitable for optimal expression in alternate hosts (i.e., other than *Yarrowia lipolytica*), based on the wildtype RD5 sequence. Accordingly, the instant invention relates to any codon-optimized $\Delta 5$ desaturase protein that is derived from the wildtype RD5 (i.e., 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 $\Delta 5$ desaturase protein (i.e., RD5S) that was codon-optimized for expression in *Yarrowia lipolytica*.

Identification And Isolation Of Homologs

Any of the instant desaturase sequences (i.e., RD5, RD5S) or portions thereof may be used to search for $\Delta 5$ desaturase homologs in the same or other bacterial, algal, fungal, or plant 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.

Alternatively, any of the instant desaturase sequences or portions thereof may also be employed as hybridization reagents for the identification of $\Delta 5$ 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.

5 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
10 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
15 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).

20 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)),
25 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.,
30 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).

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.

In additional embodiments, any of the $\Delta 5$ 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, or plant 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 [e.g., polymerase chain reaction (PCR), Mullis et al., U.S. Patent 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.

For example, genes encoding similar proteins or polypeptides to the $\Delta 5$ desaturases described herein could be isolated directly by using all or a portion of the instant nucleic acid fragments as DNA hybridization probes to screen libraries from e.g., any desired yeast or fungus using methodology well known to those skilled in the art (wherein those organisms producing ARA [or derivatives thereof] would be preferred). Specific oligonucleotide probes based upon the instant 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 (e.g., 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 full-length of) the instant sequences. The resulting amplification products can be labeled directly during amplification reactions or labeled after
5 amplification reactions, and used as probes to isolate full-length DNA fragments under conditions of appropriate stringency.

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
10 designed to provide for both efficient and faithful replication of the target nucleic acid. Methods of PCR primer design are common and well known in the art (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,
15 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. Humana: Totowa, NJ).

Generally two short segments of the instant sequences may be used in PCR protocols to amplify longer nucleic acid fragments encoding
20 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 instant nucleic acid fragments, and 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
25 eukaryotic genes.

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
30 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 instant sequences. Using commercially available 3' RACE or 5' RACE systems (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)).

In other embodiments, any of the $\Delta 5$ 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. Alternatively, improved fatty acids may be synthesized by domain swapping, wherein a functional domain from any of the $\Delta 5$ desaturase nucleic acid fragments described herein are 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

It is expected that introduction of chimeric genes encoding the $\Delta 5$ desaturases described herein (i.e., RD5, RD5S or other mutant enzymes, codon-optimized enzymes or homologs thereof), under the control of the appropriate promoters will result in increased production of ARA and/or EPA in the transformed host organism, respectively. As such, the present invention encompasses a method for the direct production of PUFAs comprising exposing a fatty acid substrate (i.e., DGLA or ETA) to the desaturase enzymes described herein (e.g., RD5, RD5S), such that the substrate is converted to the desired fatty acid product (i.e., ARA or EPA, respectively).

More specifically, it is an object of the present invention to provide a method for the production of ARA in a host cell (e.g., oleaginous yeast), wherein the host cell comprises:

- (i) an isolated nucleotide molecule encoding a $\Delta 5$ desaturase polypeptide having at least 95% identity when compared to a polypeptide having an amino acid sequence as set forth in SEQ ID NO:2, based on BLASTP algorithms; and,
- (ii) a source of dihomo- γ -linoleic acid;

wherein the host cell is grown under conditions such that the $\Delta 5$ desaturase is expressed and the DGLA is converted to ARA, and wherein the ARA is optionally recovered.

The person of skill in the art will recognize that the broad substrate
5 range of the $\Delta 5$ desaturase may additionally allow for the use of the enzyme for the conversion of ETA to EPA. Accordingly the invention provides a method for the production of EPA, wherein the host cell comprises:

- 10 (i) an isolated nucleotide molecule encoding a $\Delta 5$ desaturase polypeptide having at least 95% identity when compared to a polypeptide having an amino acid sequence as set forth in SEQ ID NO:2, based on BLASTP algorithms; and,
- (ii) a source of ETA;

wherein the host cell is grown under conditions such that the $\Delta 5$
15 desaturase is expressed and the ETA is converted to EPA, and wherein the EPA is optionally recovered.

Alternatively, each $\Delta 5$ desaturase gene and its corresponding enzyme product described herein can be used indirectly for the production of ω -3 fatty acids (see U.S. Patent Publication No. 2005/0136519).
20 Indirect production of ω -3/ ω -6 PUFAs occurs wherein the fatty acid substrate is converted indirectly into the desired fatty acid product, via means of an intermediate step(s) or pathway intermediate(s). Thus, it is contemplated that the $\Delta 5$ desaturases described herein (e.g., RD5, RD5S or other mutant enzymes, codon-optimized enzymes or homologs thereof)
25 may be expressed in conjunction with additional genes encoding enzymes of the PUFA biosynthetic pathway (e.g., $\Delta 6$ desaturases, $C_{18/20}$ elongases, $\Delta 17$ desaturases, $\Delta 15$ desaturases, $\Delta 9$ desaturases, $\Delta 12$ desaturases, $C_{14/16}$ elongases, $C_{16/18}$ elongases, $\Delta 9$ elongases, $\Delta 8$ desaturases, $\Delta 4$ desaturases, $C_{20/22}$ elongases) to result in higher levels of production of
30 longer-chain ω -3 fatty acids (e.g., EPA, DPA and 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).

In alternative embodiments, it may be useful to disrupt a host organism's native $\Delta 5$ desaturase, based on the complete sequences described herein, the complement of those complete sequences,
5 substantial portions of those sequences, codon-optimized desaturases derived therefrom and those sequences that are substantially homologous thereto.

Expression Systems, Cassettes And Vectors

10 The genes and gene products of the instant sequences 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
15 possible using the host.

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

Vectors or DNA cassettes useful for the transformation of suitable host cells are well known in the art. The specific choice of sequences
25 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 or cassette contains sequences directing transcription and translation of the relevant gene(s), a selectable marker and
30 sequences allowing autonomous replication or chromosomal integration. Suitable vectors comprise a region 5' of the gene that controls transcriptional initiation (e.g., a promoter) 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.

5 Initiation control regions or promoters which are useful to drive expression of the instant $\Delta 5$ desaturase ORFs in the desired host cell are numerous and familiar to those skilled in the art. Virtually any promoter capable of directing expression of these genes in the selected host cell is suitable for the present invention. Expression in a host cell can be
10 accomplished in a transient or stable fashion. Transient expression can be accomplished by inducing the activity of a regulatable promoter operably linked to the gene of interest. Stable expression can be achieved by the use of a constitutive promoter operably linked to the gene of interest. As an example, when the host cell is yeast, transcriptional and
15 translational regions functional in yeast cells are provided, particularly from the host species (e.g., see U.S. Patent Application No. 11/265761, corresponding to PCT Publication No. WO 2006/052870 for preferred transcriptional initiation regulatory regions for use in *Yarrowia lipolytica*). Any one of a number of regulatory sequences can be used, depending
20 upon whether constitutive or induced transcription is desired, the efficiency of the promoter in expressing the ORF of interest, the ease of construction and the like.

 The termination region can be derived from the 3' region of the gene from which the initiation region was obtained or from a different
25 gene. A large number of termination regions are known and function satisfactorily in a variety of hosts (when utilized both in the same and different genera and species from where they were derived). The termination region usually is selected more as a matter of convenience rather than because of any particular property. Termination control
30 regions may also be derived from various genes native to the preferred hosts. Optionally, a termination site may be unnecessary; however, it is most preferred if included.

As one of skill in the art is aware, merely inserting a gene into a cloning vector does not ensure that it will be successfully expressed at the level needed. 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 aspects of transcription, translation, protein stability, oxygen limitation, and secretion from the host cell. More specifically, some of the molecular features that have been manipulated to control gene expression include: 1.) the nature of the relevant transcriptional promoter and terminator sequences; 2.) the number of copies of the cloned gene and whether the gene is plasmid-borne or integrated into the genome of the host cell; 3.) the final cellular location of the synthesized foreign protein; 4.) the efficiency of translation and correct folding of the protein in the host organism; 5.) the intrinsic stability of the mRNA and protein of the cloned gene within the host cell; and, 6.) 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 types of modifications are encompassed in the present invention, as means to further optimize expression of the $\Delta 5$ desaturases described herein.

20 Transformation Of Host Cells

Once the DNA encoding a polypeptide suitable for expression in an appropriate host cell has been obtained (e.g., 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.

Where two or more genes are expressed from separate replicating vectors, it is desirable that each vector has 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
5 levels to provide for synthesis of the desired products.

Constructs comprising the gene 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,
10 electroporation, microinjection, or any other method that introduces the gene of interest into the host cell.

For convenience, a host cell that has been manipulated by any method to take up a DNA sequence (e.g., an expression cassette) will be referred to as "transformed" or "recombinant" herein. The transformed
15 host will have at least one copy of the expression construct and may have two or more, depending upon whether the gene 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. Patent Publications No.
20 2005/0136519 and No. 2005/0014270, corresponding to PCT Publications No. WO 2004/101757 and No. WO 2005/003310, respectively.

Following transformation, substrates suitable for the instant $\Delta 5$ desaturases (and, optionally other PUFA enzymes that are co-expressed within the host cell) may be produced by the host either naturally or
25 transgenically, or they may be provided exogenously.

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

Knowledge of the sequences of the present $\Delta 5$ 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
30 PUFA biosynthetic pathway or additional manipulation of pathways that contribute carbon to the PUFA biosynthetic pathway. Methods useful for up-regulating desirable biochemical pathways and down-regulating undesirable biochemical pathways are well known to those skilled 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 (e.g., antisense mRNA).

Detailed discussion of manipulations within the PUFA biosynthetic pathway as a means to increase ARA, EPA or DHA (and associated techniques thereof) are presented in PCT Publication No. WO 2006/055322 [U.S. Patent Publication No. 2006-0094092-A1], PCT Publication No. WO 2006/052870 [U.S. Patent Publication No. 2006-0115881-A1] and PCT Publication No. WO 2006/052871 [U.S. Patent Publication No. 2006-0110806-A1], respectively, as are desirable manipulations in the TAG biosynthetic pathway and the TAG degradation pathway (and associated techniques thereof).

Preferred Hosts For Recombinant Expression Of $\Delta 5$ Desaturases

Host cells for expression of the instant genes and nucleic acid fragments may include hosts that grow on a variety of feedstocks, including simple or complex carbohydrates, fatty acids, organic acids, oils 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 in the instant invention 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, and/or fungus will be a suitable host for expression of the present nucleic acid fragments.

Preferred hosts are oleaginous organisms, such as oleaginous yeast. These oleaginous organisms are naturally capable of oil synthesis and accumulation, wherein the oil 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. Genera typically identified as oleaginous yeast include, but are not limited to: *Yarrowia*, *Candida*, *Rhodotorula*,

Rhodospiridium, *Cryptococcus*, *Trichosporon* and *Lipomyces*. More specifically, illustrative oil-synthesizing yeast include: *Rhodospiridium toruloides*, *Lipomyces starkeyii*, *L. lipoferus*, *Candida revkaui*, *C. pulcherrima*, *C. tropicalis*, *C. utilis*, *Trichosporon pullans*, *T. cutaneum*,
5 *Rhodotorula glutinus*, *R. graminis* and *Yarrowia lipolytica* (formerly classified as *Candida lipolytica*).

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
10 LGAM S(7)1 (Papanikolaou S., and Aggelis G., *Bioresour. Technol.*, 82(1):43-9 (2002)).

Specific teachings applicable for transformation of oleaginous yeasts (i.e., *Yarrowia lipolytica*) include U.S. Patent 4,880,741, U.S. Patent 5,071,764 (incorporated herein by reference) and Chen, D. C. et al. (*Appl. Microbiol. Biotechnol.*, 48(2):232-235 (1997)). Specific teachings
15 applicable for engineering ARA, EPA and DHA in *Y. lipolytica* are provided in U.S. Patent Applications No. 11/264784, No. 11/265761 and No. 11/264737, respectively and incorporated herein by reference. Detailed means for the synthesis and transformation of expression vectors
20 comprising $\Delta 5$ desaturases in oleaginous yeast (i.e., *Yarrowia lipolytica*) are provided in PCT Publications No. WO 2004/101757 and No. WO 2006/052870. The preferred method of expressing genes in this yeast is by integration of linear DNA into the genome of the host; and, integration into multiple locations within the genome can be particularly useful when
25 high level expression of genes are desired [e.g., 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*:
30 GenBank Accession No. AJ001301), the $\Delta 12$ desaturase gene locus (PCT Publication No. WO 2004/104167), the *Lip1* gene locus (GenBank Accession No. Z50020) and/or the *Lip2* gene locus (GenBank Accession No. AJ012632)].

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. In alternate embodiments, 5-fluoroorotic acid (5-fluorouracil-6-carboxylic acid monohydrate; "5-FOA") is used for selection of yeast *Ura⁻* mutants. The 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).

Other preferred microbial hosts include oleaginous bacteria, algae, and other fungi; and, within this broad group of microbial hosts, of particular interest are microorganisms that synthesize ω -3/ ω -6 fatty acids. Thus, for example, transformation of *Mortierella alpina* (which is commercially used for production of ARA) with any of the present Δ 5 desaturase genes under the control of inducible or regulated promoters could yield a transformant organism capable of synthesizing increased quantities of ARA. 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 are disclosed in U.S. Patent 7,001,772.

No matter what particular host is selected for expression of the Δ 5 desaturases described herein, multiple transformants must be screened in order to obtain a strain displaying the desired expression level 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 (Kroczeck, *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

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 that may be optimized include the type and amount of carbon source, the type and amount of nitrogen source, the carbon-to-nitrogen ratio, the 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. *Yarrowia lipolytica* are generally grown in complex media (e.g., 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)).

Fermentation media in the present invention must contain a suitable carbon source. Suitable carbon sources are taught in PCT Publication No. WO 2004/101757. Although it is contemplated that the source of carbon utilized in the present invention may encompass a wide variety of carbon-containing sources, preferred carbon sources are sugars, glycerol, and/or fatty acids. Most preferred is glucose and/or fatty acids containing between 10-22 carbons.

Nitrogen may be supplied from an inorganic (e.g., $(\text{NH}_4)_2\text{SO}_4$) 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 necessary for PUFA production. Particular attention is given to several metal ions (e.g., 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)).

Preferred growth media in the present invention are common commercially prepared media, such as Yeast Nitrogen Base (DIFCO Laboratories). Other defined or synthetic growth media may also be used and the appropriate medium for growth of the transformant host cells will be known by one skilled in the art of 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.

5 Typically, accumulation of high levels of PUFAs 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 PUFAs in *Yarrowia lipolytica*. This approach is described in
10 PCT Publication No. WO 2004/101757, as are various suitable fermentation process designs (i.e., batch, fed-batch and continuous) and considerations during growth.

Purification And Processing Of PUFA Oils

PUFAs may be found in the host microorganisms as free fatty acids
15 or in esterified forms such as acylglycerols, phospholipids, sulfolipids or glycolipids, and 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
20 review of downstream processing is also available by A. Singh and O. Ward (*Adv. Appl. Microbiol.*, 45:271-312 (1997)).

In general, means for the purification of PUFAs may include extraction with organic solvents, sonication, supercritical fluid extraction (e.g., using carbon dioxide), saponification and physical means such as
25 presses, or combinations thereof. One is referred to the teachings of PCT Publication No. WO 2004/101757 for additional details.

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

The market place currently supports a large variety of food and feed
30 products, incorporating ω -3 and/or ω -6 fatty acids (particularly ARA, EPA and DHA). It is contemplated that the oils of the invention comprising long-chain PUFAs will function in food and feed products to impart the health benefits of current formulations. More specifically, oils of the

invention containing ω -3 and/or ω -6 fatty acids will be suitable for use in a variety of food and feed products including, but not limited to food analogs, meat products, cereal products, baked foods, snack foods and dairy products (see U.S. Patent Publication No. 2006/0094092 for details).

5 Additionally the present oils may be used in formulations to impart health benefits in medical foods including medical nutritionals, dietary supplements, infant formula as well as pharmaceutical products. One of skill in the art of food processing and food formulation will understand how the amount and composition of the present oils may be added to the food
10 or feed product. Such an amount will be referred to herein as an "effective" amount and will depend on the food or feed product, the diet that the product is intended to supplement or the medical condition that the medical food or medical nutritional is intended to correct or treat.

EXAMPLES

15 The present invention is further defined in the following Examples. It should be understood that these Examples, while indicating preferred embodiments of the invention, are given by way of illustration only. From the above discussion and these Examples, one skilled in the art can ascertain the essential characteristics of this invention, and without
20 departing from the spirit and scope thereof, can make various changes and modifications of the invention to adapt it to various usages and conditions.

GENERAL METHODS

Standard recombinant DNA and molecular cloning techniques used
25 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. Bannan, and L. W. Enquist, *Experiments with Gene Fusions*; Cold Spring Harbor Laboratory: Cold
30 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).

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.

- 5 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
- 10 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* (XL1-Blue) competent cells were purchased from the Stratagene Company (San Diego, CA). *E. coli* strains were
- 15 typically grown at 37 °C on Luria Bertani (LB) plates.

- 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. Patent 5,366,860; EP 272,007) using a combination of vector and insert-specific
- 20 primers. Sequence editing was performed in Sequencher (Gene Codes Corporation, Ann Arbor, MI). All sequences represent coverage at least two times in both directions. Unless otherwise indicated herein comparisons of genetic sequences were accomplished using DNASTAR software (DNASTAR Inc., Madison, WI).

- 25 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"
- 30 means microgram(s), "ng" means nanogram(s), "U" means unit(s), "bp" means base pair(s) and "kB" means kilobase(s).

Transformation And Cultivation Of *Yarrowia lipolytica*

Yarrowia lipolytica strain ATCC #20362 was purchased from the American Type Culture Collection (Rockville, MD). *Y. lipolytica* strains were usually grown at 28 °C on YPD agar (1% yeast extract, 2%
5 bactopeptone, 2% glucose, 2% agar).

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
10 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; 0.125 mL of 2 M DTT; and 50 µg sheared salmon sperm DNA. Then, approximately 500 ng of linearized plasmid DNA was incubated in 100 µl of resuspended cells, and
15 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.

For selection of transformants, minimal medium ("MM") was generally used; the composition of MM is as follows: 0.17% yeast nitrogen
20 base (DIFCO Laboratories, Detroit, MI) without ammonium sulfate or amino acids, 2% glucose, 0.1% proline, pH 6.1). Supplements of uracil were added as appropriate to a final concentration of 0.01% (thereby producing "MMU" selection media, prepared with 20 g/L agar).

Alternatively, transformants were selected on 5-fluoroorotic acid
25 ("FOA"; also 5-fluorouracil-6-carboxylic acid monohydrate) selection media, comprising: 0.17% yeast nitrogen base (DIFCO Laboratories) without ammonium sulfate or amino acids, 2% glucose, 0.1% proline, 75 mg/L uracil, 75 mg/L uridine, 900 mg/L FOA (Zymo Research Corp., Orange, CA) and 20 g/L agar.

30 Fatty Acid Analysis Of *Yarrowia lipolytica*

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 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.

- 5 The oven temperature was from 170 °C (25 min hold) to 185 °C at 3.5 °C/min.

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
10 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

15 *Peridinium* sp. CCMP626 Lipid Profile And RNA Isolation

Peridinium sp. CCMP626 (red algae) was purchased from The Provasoli-Guillard National Center for Culture of Marine Phytoplankton (Boothbay Harbor, Maine). About 200 mg cells were dried for 5 min under vacuum, resuspended in 100 µL of trimethylsulfonium hydroxide (TMSH)
20 and incubated at room temperature for 15 min with shaking. After this, 0.5 mL of hexane was added and the vials were incubated for 15 min at room temperature with shaking. Fatty acid methyl esters (5 µL injected from hexane layer) were separated and quantified using a Hewlett-Packard 6890 Gas Chromatograph fitted with an Omegawax 320 fused silica
25 capillary column (Catalog #24152, Supelco Inc., Bellefonte, PA). The oven temperature was programmed to hold at 220 °C for 2.7 min, increase to 240 °C at 20 °C/min and then hold for an additional 2.3 min. Carrier gas was supplied by a Whatman hydrogen generator. Retention times were compared to those for methyl esters of standards commercially available
30 (Catalog #U-99-A, Nu-Chek Prep, Inc., Elysian, MN) and the resulting chromatogram is shown in Figure 2.

The GC profile demonstrated that *Peridinium* sp. CCMP626 cells produced large amounts of both ARA and EPA, thus suggesting the

presence of a $\Delta 5$ desaturase with high efficiency. Since only a trace amount of GLA was detected, it was hypothesized that *Peridinium* sp. CCMP626 utilizes a $\Delta 9$ elongase/ $\Delta 8$ desaturase pathway for PUFA production.

5 Total RNA was extracted from CCMP626 cells. Specifically, cells from 1 L of culture were collected by centrifugation, quick frozen in liquid N₂ and stored at -80 °C. The cell pellet was resuspended in 1 mL of Trizol reagent (Invitrogen, Carlsbad, CA), mixed with 0.6 mL of glass beads (0.5 mm), and the mixture was homogenized at the highest setting on a
10 Biospec Mini Beadbeater (Bartlesville, OK) for 3 min. Total RNA was then isolated, according to Invitrogen's protocol for Trizol. In this way, total RNA (34 µg) was obtained from 1 L culture. The total RNA sample was used for preparation of cDNA.

EXAMPLE 2

15 *Peridinium* sp. CCMP626 cDNA Synthesis

cDNA was synthesized directly from the *Peridinium* sp. CCMP626 total RNA as follows. Specifically, the total RNA was primed with adapter primer AP (SEQ ID NO:27) from Invitrogen's 3'-RACE kit (Carlsbad, CA), in the presence of the Smart IV oligonucleotide (SEQ ID
20 NO:28) from the BD-Clontech Creator™ Smart™ cDNA library kit (Mississauga, ON, Canada). The reverse transcription was done with Superscript II reverse transcriptase from the 3'-RACE kit according to the protocol of the Creator™ Smart™ cDNA library kit.

The 1st strand cDNA synthesis mixture was used as template for
25 PCR amplification, using primer AP (SEQ ID NO:27) as the 3' primer and CDSIII 5' primer (SEQ ID NO:29) as the 5' primer (supplied with the BD-Clontech Creator™ Smart™ cDNA library kit). Amplification was carried out with Clontech Advantage cDNA polymerase mix at 94 °C for 30 sec, followed by 20 cycles of 94 °C for 10 sec and 68 °C for 6 min. A final
30 extension at 68 °C for 7 min was performed.

EXAMPLE 3

Isolation Of A Portion Of The Coding Region Of The *Peridinium* sp.

CCMP626 $\Delta 5$ Desaturase Gene

The present Example describes the identification of a portion of the
5 *Peridinium* sp. CCMP626 gene encoding $\Delta 5$ desaturase (designated
herein as "RD5" [i.e., red algae D5] and corresponding to SEQ ID NOs:1
and 2), by use of primers derived from conserved regions of other known
 $\Delta 5$ and $\Delta 8$ desaturase sequences.

Various considerations were made when evaluating which
10 desaturases might enable design of degenerate primers suitable to isolate
the *Peridinium* sp. CCMP626 $\Delta 5$ desaturase. Specifically, the Applicants
knew that only $\Delta 5$, $\Delta 6$ and $\Delta 8$ desaturase sequences comprise a
conserved 'HPGG' motif at their N-terminus (wherein the 'HPGG' domain
is part of the well-known cytochrome B5 domain); in contrast, $\Delta 9$
15 desaturases possess a 'HPGG' motif of the cytochrome B5 domain at their
C-terminus, while both $\Delta 17$ and $\Delta 12$ desaturases lack the cytochrome B5
domain. Based on the GC results described in Figure 2, it was assumed
that a $\Delta 9$ elongase/ $\Delta 8$ desaturase pathway operated in *Peridinium* sp.
CCMP626; thus, among the desaturases sharing the N-terminal
20 conserved 'HPGG' motif, only $\Delta 5$ and $\Delta 8$ desaturases were expected
within the organism. Finally, although only a few $\Delta 8$ desaturase
sequences are known, numerous $\Delta 5$ desaturase are publicly available.
The Applicants selected those $\Delta 5$ desaturase sequences that possessed
lower homology to "traditional" $\Delta 5$ desaturase genes and that also shared
25 high homology to one another.

Based on the above, the four $\Delta 5$ desaturases and two $\Delta 8$
desaturases shown below in Table 3 were aligned, 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
30 software.

Table 3
 Δ 5 And Δ 8 Desaturases Aligned To Identify Regions Of Conserved
Amino Acids

Desaturase	Organism	Reference	SEQ ID NO:
Δ 5	<i>Pythium irregulare</i>	GenBank Accession No. AAL13311	12
Δ 5	<i>Phytophthora megasperma</i>	GenBank Accession No. CAD53323	13
Δ 5	<i>Phaeodactylum tricornutum</i>	GenBank Accession No. AAL92562	14
Δ 5	<i>Dictyostelium discoideum</i>	GenBank Accession No. XP_640331	15
Δ 8	<i>Euglena gracilis</i>	PCT Publications No. WO 2006/012325 and No. WO 2006/012326	16
Δ 8	<i>Pavlova lutheri</i>	Example 12 (<i>infra</i>)	18

5 Figure 3 shows a portion of the resulting alignment, containing several stretches of conserved amino acid sequence among the 6 different organisms. Based on this alignment, two sets of degenerate oligonucleotides were designed to amplify a portion of the coding region of the Δ 5 desaturase gene from *Peridinium* sp. CCMP626, corresponding to

10 the regions of Figure 3 that are labeled as "Conserved Region 1" and "Conserved Region 2". Specifically, the conserved amino acid sequence GHH(I/V)YTN (SEQ ID NO:19) was designed to correspond to Conserved Region 1, while the conserved amino acid sequence NFQ(V/A)(S/N)HV (SEQ ID NO:20) was designed to correspond to Conserved Region 2. In

15 order to reduce the degeneracy of the oligonucleotides, 4 sets of oligonucleotides (i.e., 5-1A, 5-1B, 5-1C and 5-1D) were designed to encode Conserved Region 1; and, 4 sets of oligonucleotides (i.e., 5-4AR, 5-4BR, 5-4CR and 5-4DR) were designed to encode the anti-sense strand of Conserved Region 2.

20

Table 4

Degenerate Oligonucleotides Used To Amplify The $\Delta 5$ Desaturase Gene
From *Peridinium* sp. CCMP626

<u>Oligonucleotide</u> <u>Name</u>	<u>Sequence</u>	<u>SEQ ID NO</u>
5-1A	GGHCAYCAYRTBTAYACAAA	SEQ ID NO:30
5-1B	GGHCAYCAYRTBTAYACCAA	SEQ ID NO:31
5-1C	GGHCAYCAYRTBTAYACGAA	SEQ ID NO:32
5-1D	GGHCAYCAYRTBTAYACTAA	SEQ ID NO:33
5-4AR	ACRTGRYTNACYTGRAAGTT	SEQ ID NO:34
5-4BR	ACRTGRYTNACYTGRAAATT	SEQ ID NO:35
5-4CR	ACRTGNGANACYTGRAAGTT	SEQ ID NO:36
5-4DR	ACRTGNGANACYTGRAAATT	SEQ ID NO:37

[Note: The nucleic acid degeneracy code used for SEQ ID NOs:30 to 37 was as follows: R= A/G; Y=C/T; B=G/T/C; and H=A/C/T.]

5

Based on the full-length sequences of the $\Delta 5$ sequences of Figure 3, it was hypothesized that the *Peridinium* sp. CCMP626 $\Delta 5$ gene fragment amplified as described above would be about 630 bp in length (lacking about 210 amino acids at its N-terminal and 70 amino acids at its C-terminal).

10

A total of sixteen different PCR amplifications were conducted, as all combinations of the primers were tested (i.e., primer 5-1A was used with each of 5-4AR, 5-4BR, 5-4CR and 5-4DR, individually; similarly, primer 5-1B was used with each 5-4AR, 5-4BR, 5-4CR and 5-4DR; etc.). The PCR amplifications were carried out in a 50 μ l total volume comprising: PCR buffer (containing 10 mM KCl, 10 mM $(\text{NH}_4)_2\text{SO}_4$, 20 mM Tris-HCl (pH 8.75), 2 mM MgSO_4 , 0.1% Triton X-100), 100 μ g/mL BSA (final concentration), 200 μ M each deoxyribonucleotide triphosphate, 10 pmole of each primer, 10 ng cDNA of *Peridinium* sp. CCMP626 and 1 μ l of *Taq* DNA polymerase (Epicentre Technologies, Madison, WI). The thermocycler conditions were set for 35 cycles at 95 °C for 1 min, 56 °C for 30 sec and 72 °C for 1 min, followed by a final extension at 72 °C for 10 min.

15

20

25

The PCR products were purified using a Qiagen PCR purification kit (Valencia, CA). One fragment of the approximate expected size was then further purified following gel electrophoresis in 1% (w/v) agarose and then

cloned into the pGEM-T-easy vector (Promega, Madison, WI). The ligated DNA was used to transform cells of *E. coli* DH10B and transformants were selected on LB (1% bacto-tryptone, 0.5% bacto-yeast extract and 1% NaCl) agar containing ampicillin (100 µg/mL). Analysis of the plasmid DNA from a group of 12 transformants confirmed the presence of the insert with the expected size (plasmids were designated as "pT-12-D1", "pT-12-D2", "pT-12-D3", etc. to "pT-12-D12").

Sequence analyses showed that pT-12-D5 contained a 563 bp fragment (SEQ ID NO:4), which encoded 187 amino acids (SEQ ID NO:5) (including amino acids that corresponded to Conserved Region 1 and 2). Identity of the *Peridinium* sp. CCMP626 sequence was 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 sequence was 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). SEQ ID NO:4 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:4 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. Thus, the translated amino acid sequence of SEQ ID NO:4 (i.e., SEQ ID NO:5) had 64% identity and

78% similarity with the amino acid sequence of the $\Delta 8$ -sphingolipid desaturase of *Thalassiosira pseudonana* (GenBank Accession No. AAX14502; SEQ ID NO:21), with an Expectation value of 2E-64; additionally, the partial fragment of SEQ ID NO:4 had 65% identity and
5 78% similarity with the $\Delta 5$ fatty acid desaturase of *Phaeodactylum tricornutum* (GenBank Accession No. AAL92562; SEQ ID NO:14), with an Expectation value of 7E-62.

EXAMPLE 4

Isolation Of The 5' Coding Region Of The *Peridinium* sp. CCMP626

10 $\Delta 5$ Desaturase Gene

To isolate the N-terminal portion of the putative $\Delta 5$ desaturase identified in Example 3, a modified 5' RACE technique based on RACE protocols from two different companies (i.e., Invitrogen and BD-Clontech) was utilized.

15 Briefly, the double-stranded cDNA of *Peridinium* sp. CCMP626 (Example 2) was used as the template in a 5' RACE experiment, comprising two separate rounds of PCR amplification. In the first round of PCR amplification, the oligonucleotide primers consisted of a gene specific oligonucleotide (i.e., ODMW520; SEQ ID NO:38) and the generic
20 oligonucleotide CDSIII 5' primer (SEQ ID NO:29) from the BD-Clontech Creator™ Smart™ cDNA library kit. The PCR amplifications were carried out in a 50 μ l total volume, comprising: 25 μ l of LA Taq™ pre-mix (TaKaRa Bio Inc., Otsu, Shiga, 520-2193, Japan), 10 pmole of each primer and 1 μ l of Taq DNA polymerase (Epicentre Technologies,
25 Madison, WI). The thermocycler conditions were set for 35 cycles at 95 °C for 1 min, 56 °C for 30 sec and 72 °C for 1 min, followed by a final extension at 72 °C for 10 min.

The second round of PCR amplification used 1 μ l of the product from the first round PCR reaction as template. Primers consisted of a
30 gene specific oligonucleotide (i.e., ODMW521; SEQ ID NO:39) and the generic oligonucleotide DNR CDS 5' (SEQ ID NO:40), supplied with BD-

Clontech's Creator™ Smart™ cDNA library kit. Amplification was conducted as described above.

The products of the second round PCR reaction were electrophoresed in 1% (w/v) agarose. Products between 400 bp and 800
5 bp were then purified from the gel and cloned into the pGEM-T-easy vector (Promega, Madison, WI). The ligated DNA was used to transform *E. coli* DH10β and transformants were selected on LB agar containing ampicillin (100 µg/mL).

Analysis of the plasmid DNA from one transformant comprising the
10 5' region of the putative Δ5 desaturase gene confirmed the presence of the expected plasmid, designated pT-RD5-5'C2. Sequence analyses showed that pT-RD5-5'C2 contained a fragment of 693 bp (SEQ ID NO:6), which over-lapped with 182 bp from the 5' end of the 563 bp fragment of pT-12-D5 (Example 3; SEQ ID NO:4) and additionally provided 511 bp of
15 5' upstream sequence (SEQ ID NO:7) (Figure 4). The sequence of pT-RD5-5'C2 also corrected the sequence corresponding to Conserved Region 1, resulting from use of a degenerate oligonucleotide for initial PCR amplification of the 563 bp fragment in pT-12-D5 (Example 3). However, there was no translation initiation codon in the extended 693 bp
20 fragment of SEQ ID NO:6.

A second round of the modified 5' RACE was carried out as described above, except that oligonucleotides ODMW541 (SEQ ID NO:41) and ODMW542 (SEQ ID NO:42) were used as gene-specific primers. Products between 200 bp and 400 bp were then purified from a gel and
25 cloned into the pGEM-T-easy vector (Promega, Madison, WI). The ligated DNA was transformed into *E. coli* DH10β and transformants were selected on LB agar containing ampicillin (100 µg/mL).

Analysis of the plasmid DNA from one transformant comprising the
30 5' region of the putative Δ5 desaturase gene confirmed the presence of the expected plasmid, designated pT-RD5-5'2nd. Sequence analyses showed that pT-RD5-5'2nd contained a fragment of 358 bp (SEQ ID NO:8), which over-lapped with 197 bp of the 5' end of the DNA fragment in pT-RD5-5'C2 described above and additionally provided 161 bp of 5'

upstream sequence (SEQ ID NO:9). One-hundred sixteen (116) bp of the 5' extended fragment encoded the N-terminal portion of the putative $\Delta 5$ desaturase gene, including the translation initiation codon, thus providing the complete 5' sequence of the gene.

EXAMPLE 5

Isolation Of The 3' Coding Region Of The *Peridinium* sp. CCMP626 $\Delta 5$ Desaturase Gene

To isolate the C-terminal portion of the putative $\Delta 5$ desaturase identified in Example 3, a 3' RACE technique was utilized. The methodology was described above in Example 4; however, the primers used on both the first and second round of PCR amplification were as shown below in Table 5.

Table 5

Oligonucleotide Primers Used For 3' RACE

PCR Amplification	Gene Specific Oligonucleotide	Generic Oligonucleotide
1 st Round	ODMW523 (SEQ ID NO:43)	AUAP (SEQ ID NO:44)
2 nd Round	ODMW524 (SEQ ID NO:45)	AUAP (SEQ ID NO:44)

* Primer AUAP was supplied in Invitrogen's 3'-RACE kit (Carlsbad, CA).

Following isolation and purification of products (i.e., 200-500 bp), the fragments were cloned into the pGEM-T-easy vector (Promega) and transformed into *E. coli* DH10B, as in Example 4.

Analysis of the plasmid DNA from one transformant comprising the 3' region of the $\Delta 5$ desaturase gene confirmed the presence of the expected plasmid, designated pT-RD5-3'. Sequence analyses showed that pT-RD5-3' contained a fragment of 299 bp (SEQ ID NO:10), which over-lapped with 52 bp from the 3' end of the 563 bp fragment of pT-12-D5 (Example 3, SEQ ID NO:4) and provided 247 bp of additional 3' downstream sequence (SEQ ID NO:11). The first 202 bp of the extended 247 bp fragment included within pT-RD5-3' encoded the C-terminal coding region (including the translation stop codon) of the putative $\Delta 5$ desaturase gene. The sequence of pT-RD5-3' also corrected the sequence

corresponding to Conserved Region 2, resulting from use of a degenerate oligonucleotide for initial PCR amplification of the 563 bp fragment in pT-12-D5 (Example 3).

After 2 rounds of 5' RACE and one round of 3' RACE, the DNA
 5 sequence of the entire putative *Peridinium* sp. CCMP626 $\Delta 5$ desaturase (RD5) coding region was determined. As shown in Figure 4, the RD5 CDS was 1392 bp in length (SEQ ID NO:1) and encoded a polypeptide with 463 amino acids (SEQ ID NO:2), based on alignment of SEQ ID NOs:4, 6, 8 and 10. The results of BLASTP sequence analysis algorithms
 10 using the full length RD5 gene as the query sequence showed that it shared 67% identity and 81% similarity with the $\Delta 5$ fatty acid desaturase of *Phaeodactylum tricornutum* (GenBank Accession No. AAL92562; SEQ ID NO:14), with an Expectation value of 0.0. Additionally, the full length RD5 gene shared 64% identity and 76% similarity with the $\Delta 8$ -sphingolipid
 15 desaturase of *Thalassiosira pseudonana* (GenBank Accession No. AAX14502; SEQ ID NO:21), with an Expectation value of 2E-178.

EXAMPLE 6

Generation Of Construct pDMW368, Comprising RD5

The present Example describes the generation of pDMW368,
 20 comprising a chimeric FBAIN::RD5::Pex20 gene (Figure 5C). This was designed to integrate the chimeric gene into the genome of *Yarrowia lipolytica* and then study the function of the *Peridinium* sp. CCMP626 $\Delta 5$ desaturase in *Yarrowia lipolytica*.

Based on the full length cDNA of RD5 (SEQ ID NO:1),
 25 oligonucleotides YL807 and YL810 (SEQ ID NOs:46 and 47, respectively) were used as primers to amplify the first portion of RD5 (Figure 5A). Primer YL807 contained a *Nco*I site and primer YL810 contained a *Bam*H1 site. Then, primers YL808 and YL809 (SEQ ID NOs:48 and 49, respectively) were used as primers to amplify the second portion of RD5.
 30 Primer YL809 contained a *Bam*H1 site, while primer YL808 contained a *Not*I site. The PCR reactions, using primer pairs YL807/YL810 or YL808/YL809, with *Peridinium* sp. CCMP626 cDNA (Example 2) as template, were individually carried out in a 50 μ l total volume comprising:

PCR buffer (containing 10 mM KCl, 10 mM (NH₄)₂SO₄, 20 mM Tris-HCl (pH 8.75), 2 mM MgSO₄, 0.1% Triton X-100), 100 µg/mL BSA (final concentration), 200 µM each deoxyribonucleotide triphosphate, 10 pmole of each primer and 1 µl of *Pfu* DNA polymerase (Stratagene, San Diego, CA). The thermocycler conditions were set for 35 cycles at 95 °C for 1 min, 56 °C for 30 sec and 72 °C for 1 min, followed by a final extension at 72 °C for 10 min.

The individual PCR products were purified using a Qiagen PCR purification kit. The PCR products from the reaction amplified with primers YL807/YL810 were digested with *Nco*I and *Bam*H1, while the PCR products from the reaction amplified with primers YL808/YL809 were digested with *Bam*H1 and *Not*I. The *Nco*I/*Bam*H1- and the *Bam*H1/*Not*I-digested DNA fragments were purified following gel electrophoresis in 1% (w/v) agarose, and then directionally ligated with *Nco*I/*Not*I-digested pZUF17 (Figure 5B; SEQ ID NO:22; comprising a synthetic Δ17 desaturase gene ["D17st"] derived from *Saprolegnia diclina* (U.S. Patent Publication No. 2003/0196217 A1), codon-optimized for expression in *Yarrowia lipolytica* (PCT Publication No. WO 2004/101757)). The product of this ligation was pDMW368 (Figure 5C; SEQ ID NO:23, which thereby contained the following components:

Table 6
Components Of Plasmid pDMW368

RE Sites And Nucleotides Within SEQ ID NO:23	Description Of Fragment And Chimeric Gene Components
<i>Eco</i> R I/ <i>Bsi</i> W I (6063-318)	FBAIN::RD5::Pex20, comprising: • FBAIN: <i>Yarrowia lipolytica</i> FBAIN promoter (PCT Publication No. WO 2005/049805; U.S. Patent 7,202,356) • RD5: <i>Peridinium</i> sp.CCMP626 Δ5 desaturase (SEQ ID NO:1 described herein) • Pex20: Pex20 terminator sequence of <i>Yarrowia</i> Pex20 gene (GenBank Accession No. AF054613)
1354-474	ColE1 plasmid origin of replication
2284-1424	ampicillin-resistance gene (Amp ^R) for selection in <i>E. coli</i>

3183-4487	<i>Yarrowia</i> autonomous replication sequence (ARS18; GenBank Accession No. A17608)
6020-4533	<i>Yarrowia Ura 3</i> gene (GenBank Accession No. AJ306421)

The term "FBAIN promoter" or "FBAIN promoter region" refers to the 5' upstream untranslated region in front of the 'ATG' translation initiation codon of the *Yarrowia lipolytica* fructose-bisphosphate aldolase enzyme (E.C. 4.1.2.13) encoded by the *fba1* gene and that is necessary for expression, plus a portion of 5' coding region that has an intron of the *fba1* gene.

EXAMPLE 7

Generation Of *Yarrowia lipolytica* Strain M4 To Produce About 8% DGLA Of Total Lipids

The present Example describes the construction of strain M4, derived from *Yarrowia lipolytica* ATCC #20362, capable of producing 8% DGLA relative to the total lipids. This strain was engineered to express the $\Delta 6$ desaturase/ $\Delta 6$ elongase pathway, via introduction of construct pKUNF12T6E (Figure 6A; SEQ ID NO:24). This construct was generated to integrate four chimeric genes (comprising a $\Delta 12$ desaturase, a $\Delta 6$ desaturase and two $C_{18/20}$ elongases) into the *Ura3* loci of wild type *Yarrowia* strain ATCC #20362, to thereby enable production of DGLA. Thus, pKUNF12T6E contained the following components:

Table 7

Description of Plasmid pKUNF12T6E (SEQ ID NO:24)

RE Sites And Nucleotides Within SEQ ID NO:24	Description Of Fragment And Chimeric Gene Components
<i>AscI/BsWI</i> (9420-8629)	784 bp 5' portion of <i>Yarrowia Ura3</i> gene (GenBank Accession No. AJ306421)
<i>SphI/PacI</i> (12128-1)	516 bp 3' portion of <i>Yarrowia Ura3</i> gene (GenBank Accession No. AJ306421)

Swal/BsWI (6380-8629)	FBAIN::EL1S::Pex20, comprising: • FBAIN: <i>Yarrowia lipolytica</i> FBAIN promoter (PCT Publication No. WO 2005/049805; U.S. Patent 7,202,356; labeled as "Fba1+intron" in Figure) • EL1S: codon-optimized elongase 1 gene (PCT Publication No. WO 2004/101753), derived from <i>Mortierella alpina</i> (GenBank Accession No. AX464731) • Pex20: Pex20 terminator sequence from <i>Yarrowia Pex20</i> gene (GenBank Accession No. AF054613)
BglII/Swal (4221-6380)	TEF::Δ6S::Lip1, comprising: • TEF: <i>Yarrowia lipolytica</i> TEF promoter (GenBank Accession No. AF054508) • Δ6S: codon-optimized Δ6 desaturase gene (PCT Publication No. WO 2004/101753), derived from <i>Mortierella alpina</i> (GenBank Accession No. AF465281) • Lip1: Lip1 terminator sequence from <i>Yarrowia Lip1</i> gene (GenBank Accession No. Z50020)
PmeI/ClaI (4207-1459)	FBA::F.Δ12::Lip2, comprising: • FBA: <i>Yarrowia lipolytica</i> FBA promoter (PCT Publication No. WO 2005/049805; U.S. Patent 7,202,356; labeled as "FBA1" in Figure) • F.Δ12: <i>Fusarium moniliforme</i> Δ12 desaturase gene (PCT Publication No. WO 2005/047485) • Lip2: Lip2 terminator sequence from <i>Yarrowia Lip2</i> gene (GenBank Accession No. AJ012632)
ClaI/PacI (1459-1)	TEF::EL2Syn::XPR2, comprising: • TEF: <i>Yarrowia lipolytica</i> TEF promoter (GenBank Accession No. AF054508) • EL2Syn: codon-optimized elongase gene (SEQ ID NO:25), derived from <i>Thraustochytrium aureum</i> (U.S. Patent 6,677,145) • XPR2: ~100 bp of the 3' region of the <i>Yarrowia Xpr</i> gene (GenBank Accession No. M17741)

Plasmid pKUNF12T6E was digested with *Ascl*/*SphI*, and then used for transformation of wild type *Y. lipolytica* ATCC #20362 according to the General Methods. The transformant cells were plated onto FOA selection media plates and maintained at 30 °C for 2 to 3 days. The FOA resistant colonies were picked and streaked onto MM and MMU selection plates. The colonies that could grow on MMU plates but not on MM plates were selected as *Ura*⁻ strains. Single colonies of *Ura*⁻ strains were then inoculated into liquid MMU at 30 °C and shaken at 250 rpm/min for 2 days. The cells were collected by centrifugation, lipids were extracted, and fatty

acid methyl esters were prepared by trans-esterification, and subsequently analyzed with a Hewlett-Packard 6890 GC.

GC analyses showed the presence of DGLA in the transformants containing the 4 chimeric genes of pKUNF12T6E, but not in the wild type
5 *Yarrowia* control strain. Most of the selected 32 *Ura*⁻ strains produced about 6% DGLA of total lipids. There were 2 strains (i.e., strains M4 and 13-8) that produced about 8% DGLA of total lipids.

EXAMPLE 8

Functional Analysis Of RD5 Gene In *Yarrowia lipolytica* Strain M4

10 Plasmid pDMW368 (Example 6; comprising a chimeric FBAIN::RD5::Pex20 gene was transformed into strain M4 (Example 7), as described in the General Methods. The transformants were selected on MM plates. After 2 days grown at 30 °C, 3 transformants grown on the MM plates were picked and re-streaked onto fresh MM plates. Once
15 grown, these strains 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, lipids were extracted, and fatty acid methyl esters were prepared by trans-esterification, and subsequently analyzed with a Hewlett-Packard 6890 GC.

20 GC analyses showed that there were about 4.2% DGLA and 2.2% ARA of total lipids produced in all three transformants, wherein the conversion efficiency of DGLA to ARA in these three strains was determined to be about 35% (average). The conversion efficiency was measured according to the following formula: ($\frac{[\text{product}]}{[\text{substrate} + \text{product}]}$)*100, where 'product' includes the immediate product and all
25 products in the pathway derived from it. Thus, this experimental data demonstrated that the cloned *Peridinium* sp. CCMP626 $\Delta 5$ desaturase, described herein as SEQ ID NOs:1 and 2, efficiently desaturated DGLA to ARA.

EXAMPLE 9

Synthesis Of A Codon-Optimized $\Delta 5$ Desaturase Gene ("RD5S") For Expression In *Yarrowia lipolytica*

The codon usage of the $\Delta 5$ desaturase gene of *Peridinium* sp. CCMP626 (SEQ ID NOs:1 and 2; RD5) was optimized for expression in *Yarrowia lipolytica*, in a manner similar to that described in PCT Publication No. WO 2004/101753 and U.S. Patent 7,125,672. Specifically, a codon-optimized $\Delta 5$ desaturase gene (designated "RD5S", SEQ ID NO:3) was designed based on the coding sequence of the $\Delta 5$ desaturase gene of RD5 (SEQ ID NO:2), according to the *Yarrowia* codon usage pattern (PCT Publication 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)). In addition to modification of the translation initiation site, 247 bp of the 1392 bp coding region was modified (17.7%, Figures 7A and 7B) and 229 codons were optimized (49.4%). The GC content was increased from 49.3% within the wild type gene (i.e., RD5) to 54.2% within the synthetic gene (i.e., RD5S). A *Nco*I site and *Not*I sites were incorporated around the translation initiation codon and after the stop codon of RD5S (SEQ ID NO:3), respectively. None of the modifications in the codon-optimized gene changed the amino acid sequence of the encoded protein (SEQ ID NO:2). The designed RD5S gene was synthesized by GenScript Corporation (Piscataway, NJ) and cloned into pUC57 (GenBank Accession No. Y14837) to generate pRD5S (SEQ ID NO:50; Figure 6B) (RD5S labeled as "RD5S(626)" in Figure).

EXAMPLE 10

Generation Of Construct pZURD5S, Comprising RD5S

The present Example describes the construction of plasmid pZURD5S comprising a chimeric FBAIN::RD5S::Pex20 gene. Plasmid pZURD5S (SEQ ID NO:51; Figure 6C) was constructed by replacing the *Nco* I/*Not* I fragment of pZUF17 (Figure 5B; SEQ ID NO:22) with the *Nco* I/*Not* I RD5S fragment from pRD5S (SEQ ID NO:50; Figure 6B). The

product of this ligation was pZURD5S (Figure 6C; SEQ ID NO:51), which thereby contained the following components:

Table 8

Components Of Plasmid pZURD5S

RE Sites And Nucleotides Within SEQ ID NO:51	Description Of Fragment And Chimeric Gene Components
<i>EcoR</i> I/BsW I (7458-1713)	FBAIN::RD5S::Pex20, comprising: • FBAIN: <i>Yarrowia lipolytica</i> FBAIN promoter (PCT Publication No. WO 2005/049805; U.S. Patent 7,202,356) • RD5S: codon-optimized $\Delta 5$ desaturase (SEQ ID NO:3, described herein as RD5S), derived from <i>Peridinium</i> sp. CCMP626 (labeled as "RD5S(626)" in Figure) • Pex20: Pex20 terminator sequence of <i>Yarrowia</i> Pex20 gene (GenBank Accession No. AF054613)
2749-1869	ColE1 plasmid origin of replication
3679-2819	ampicillin-resistance gene (Amp ^R) for selection in <i>E. coli</i>
4578-5882	<i>Yarrowia</i> autonomous replication sequence (ARS18; GenBank Accession No. A17608)
7415-5928	<i>Yarrowia</i> Ura 3 gene (GenBank Accession No. AJ306421)

EXAMPLE 11

Expression Of The Codon-Optimized $\Delta 5$ Desaturase ("RD5S") In *Yarrowia lipolytica* Strain M4

Plasmid pZURD5S (Example 10; comprising a chimeric FBAIN::RD5S::Pex20 gene) was used for transformation into strain M4 (Example 7), as described in the General Methods. Eight (8) transformants grown on the MM plates were picked and re-streaked onto fresh MM plates. Once grown, these strains 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, lipids were extracted, and fatty acid methyl esters were prepared by trans-esterification, and subsequently analyzed with a Hewlett-Packard 6890 GC.

The GC results showed that there were about 5.4% DGLA and 3.3% ARA of total lipids produced in all 8 transformants. The conversion efficiency whereby RD5S converted DGLA to ARA in these 8 strains was

at an average rate of about 38%. 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,
5 this experimental data demonstrated that the codon-optimized $\Delta 5$ desaturase gene (RD5S, as set forth in SEQ ID NO:3) derived from *Peridinium* sp. CCMP626 was about 8.9% more efficient converting DGLA to ARA than the wild type RD5 (Example 8).

EXAMPLE 12

10 Isolation Of A *Pavlova lutheri* (CCMP459) $\Delta 8$ Desaturase

The present example describes the isolation of the *Pavlova lutheri* (CCMP459) $\Delta 8$ desaturase utilized in Example 3 and in Figure 3 (also described in U.S. Patent Application No. 11/737,772 incorporated herein by reference). This required: synthesis of *Pavlova lutheri* (CCMP459)
15 cDNA; library construction and sequencing; identification of $\Delta 8$ desaturase homologs; and, cloning of a full-length $\Delta 8$ desaturase from genomic DNA.
Pavlova lutheri (CCMP459) cDNA Synthesis, Library Construction And Sequencing

A cDNA library of *Pavlova lutheri* (CCMP459) was synthesized as
20 described in PCT Publication No. WO 2004/071467 (published August 26, 2004). Briefly, frozen pellets of Pav459 were obtained from the Provasoli-Guillard National Center for Culture of Marine Phytoplankton (CCMP, West Boothbay Harbor, ME). These pellets were crushed in liquid nitrogen and total RNA was extracted from Pav459 by using the Qiagen
25 RNeasy® Maxi Kit (Qiagen, Valencia, CA), per the manufacturer's instructions. From this total RNA, mRNA was isolated using oligo dT cellulose resin, which was then used for the construction of a cDNA library using the pSport1 vector (Invitrogen, Carlsbad, CA). The cDNA thus produced was directionally cloned (5' *Sall*/3' *NotI* into pSport1 vector. The
30 Pav459 library contained approximately 6.1×10^5 clones per mL, each with an average insert size of approximately 1200 bp. The *Pavlova lutheri* library was named eps1c.

For sequencing, clones first were recovered from archived glycerol cultures grown/frozen in 384-well freezing media plates, and inoculated with an automatic QPix[®] colony picker (Genetix) in 96-well deep-well plates containing LB + 100 mg/mL ampicillin. After growing 20 hrs at 37 °C, cells were pelleted by centrifugation and stored at -20 °C. Plasmids then were isolated on an Eppendorf 5Prime robot, using a modified 96-well format alkaline lysis miniprep method (Eppendorf PerfectPrep[®]). Briefly, a filter and vacuum manifold was used to facilitate removal of cellular debris after acetate precipitation. Plasmid DNA was then bound on a second filter plate directly from the filtrate, washed, dried and eluted.

Plasmids were end-sequenced in 384-well plates, using vector-primed T7 primer (SEQ ID NO:52) and the ABI BigDye version 3 Prism sequencing kit. For the sequencing reaction, 100-200 ng of template and 6.4 pmol of primer were used, and the following reaction conditions were repeated 25 times: 96 °C for 10 sec, 50 °C for 5 sec and 60 °C for 4 min. After ethanol-based cleanup, cycle sequencing reaction products were resolved and detected on Perkin-Elmer ABI 3700 automated sequencers.

Identification Of $\Delta 8$ Desaturase Enzyme Homologs From *Pavlova lutheri* cDNA Library eps1c

cDNA clones encoding *Pavlova lutheri* $\Delta 8$ desaturase homologs (hereby called $\Delta 8$ desaturases) were identified by conducting BLAST searches for similarity to sequences contained in the BLAST "nr" database (as described in Example 3). The P-value (probability) of observing a match of a cDNA sequence to a sequence contained in the searched databases merely by chance as calculated by BLAST are reported herein as "pLog" values, which represent the negative of the logarithm of the reported P-value. Accordingly, the greater the pLog value, the greater the likelihood that the cDNA sequence and the BLAST "hit" represent homologous proteins.

The BLASTX search using the nucleotide sequence from clone eps1c.pk002.f22 revealed similarity of the protein encoded by the cDNA to the $\Delta 6$ desaturase from *Rhizopus stolonifer* (SEQ ID NO:53) (NCBI Accession No. AAX22052 (GI 60499699), locus AAX22052, CDS

AY795076; Lu et al., unpublished). The sequence of a portion of the cDNA insert from clone eps1c.pk002.f22 is shown in SEQ ID NO:54 (5' end of cDNA insert). Subsequently, the full insert sequence (eps1c.pk002.f22:fis) was obtained and is shown in SEQ ID NO:55.

5 Sequence for the deduced amino acid sequence (from nucleotide 1 of SEQ ID NO:55 to the first stop codon at nucleotide 864 of SEQ ID NO:55) is shown in SEQ ID NO:56. Full insert sequencing was carried out using a modified transposition protocol. Clones identified for full insert sequencing were recovered from archived glycerol stocks as single colonies, and
10 plasmid DNA was isolated via alkaline lysis. Plasmid templates were transposed via the Template Generation System (TGS II) transposition kit (Finnzymes Oy, Espoo, Finland), following the manufacturer's protocol. The transposed DNA was transformed into EH10B electro-competent cells (Edge BioSystems, Gaithersburg, MD) via electroporation. Multiple
15 transformants were randomly selected from each transposition reaction, plasmid DNA was prepared, and templates were sequenced as above (ABI BigDye v3.1) outward from the transposition event site, utilizing unique primers SeqE (SEQ ID NO:57) and SeqW (SEQ ID NO:58).

Sequence data was collected (ABI Prism Collections software) and
20 assembled using the Phrap sequence assembly program (P. Green, University of Washington, Seattle). Assemblies were viewed by the Consed sequence editor (D. Gordon, University of Washington, Seattle) for final editing.

The amino acid sequence set forth in SEQ ID NO:56 was evaluated
25 by BLASTP, yielding a pLog value of 19.52 (E value of $3e-20$) versus the $\Delta 6$ desaturase from *Mortierella alpina* (NCBI Accession No. BAC82361 (GI 34221934), locus BAC82361, CDS AB070557; Sakuradani and Shimizu, *Biosci. Biotechnol. Biochem.*, 67:704-711 (2003)). Based on the results from the BLASTP comparison to the *Mortierella alpina* and other
30 fatty acid desaturases, the *Pavlova lutheri* $\Delta 8$ desaturase was not full length and was lacking sequence at the 5' end.

Cloning A Full-Length $\Delta 8$ Desaturase From *Pavlova lutheri* Genomic DNA

Genomic DNA was isolated from *Pavlova lutheri* (CCMP459) using the Qiagen DNeasy® Plant Maxi Prep Kit according to the manufacturer's protocol. Using 1 maxi column per 1 gm of frozen cell pellet, a total of 122
5 μg of genomic DNA was isolated from 4 gm of *Pavlova lutheri* culture. The final concentration of genomic DNA was 22.8 ng/ μL . GenomeWalker libraries were synthesized using the Universal GenomeWalker™ kit (BD Biosciences Clontech, Palo Alto, CA) following the manufacturer's protocol (Prot # PT3042-1, version PRO3300). Briefly, four restriction
10 digests were set up as per the protocol using 300 ng of genomic DNA per reaction. After phenol clean up, pellets were dissolved in 4 μL of water and adapters were ligated as per the protocol.

For the primary PCR, the Advantage®-GC Genomic PCR kit (BD Biosciences Clontech) was used following the manufacturer's protocol
15 (Prot # PT3090-1, version PR1X433). For each restriction digest, 1 μL of library was combined with 22.8 μL of PCR grade water, 10 μL of 5X GC Genomic PCR Reaction Buffer, 2.2 μL of 25 mM $\text{Mg}(\text{CH}_3\text{CO}_2)_2$, 10 μL of GC-Melt (5 M), 1 μL of 50X dNTP mix (10 mM each), 1 μL of Advantage-GC Genomic Pol. Mix (50 X), 1 μL of Universal GenomeWalker™ primer
20 AP1 (10 μM , SEQ ID NO:59) and 1 μL of GSP PvDES (10 μM , SEQ ID NO:60). After denaturation at 95 °C, the following reaction conditions were repeated 35 times: 94 °C for 30 sec, 68 °C for 6 min. After these reaction conditions, an additional extension at 68 °C was carried out for 6 min followed by cooling to 15 °C until removed.

25 The primary PCR reaction for each library was analyzed by agarose gel electrophoresis and DNA bands with molecular weights around 6 kB, 3.5 kB, 2.5 kB and 1.2 kB were observed. DNA bands for each library were purified using the Zymoclean™ Gel DNA Recovery Kit (Zymo Research, Orange, CA) following the manufacturer's protocol. The
30 resulting DNA was cloned into the pGEM®-T Easy Vector (Promega) following the manufacturer's protocol and inserts were sequenced using the T7 (SEQ ID NO:52) and M13-28Rev (SEQ ID NO:61) primers as described above. Additional sequence was then obtained using a gene-

specific sequencing primer PavDES seq (SEQ ID NO:62) that was derived from the newly acquired sequence data. The full 5' end sequence obtained by genome walking is shown in SEQ ID NO:63. The sequence of the overlapping regions of the genomic sequence (SEQ ID NO:63) and the fully sequenced EST eps1c.pk002.f22:fis (SEQ ID NO:55) were aligned using Sequencher™ (Version 4.2, Gene Codes Corporation, Ann Arbor, MI) using the Large Gap assembly algorithm. Interestingly, the comparison showed that the EST that was originally sequenced (SEQ ID NO:55) was lacking 459 bp when compared to the genomic sequence (SEQ ID NO:63). This missing sequence in the EST appeared to be a deletion rather than an intron as no clear intron splice sites were identified in the genomic DNA at the 5' end of the gene. The genomic sequence for the 5' end (SEQ ID NO:63) was combined with the 3' end of the EST sequence (SEQ ID NO:55) to yield SEQ ID NO:64. Using EditSeq™ 6.1 sequence analysis software (DNASTAR Inc., Madison, WI), an ORF was identified (SEQ ID NO:17). The amino acid sequence coded for by SEQ ID NO:17 is shown in SEQ ID NO:18.

The amino acid sequence set forth in SEQ ID NO:18 was evaluated by BLASTP, yielding a pLog value of 35.10 (E value of 8e-36) versus the $\Delta 6$ desaturase from *Rhizopus stolonifer* (SEQ ID NO:65) (NCBI Accession No. ABB96724 (GI 83027409), locus ABB96724, CDS DQ291156; Zhang et al., unpublished). Furthermore, the *Pavlova lutheri* $\Delta 8$ desaturase is 78.0% identical to the *Pavlova salina* $\Delta 8$ desaturase sequence (SEQ ID NO:66) disclosed in PCT Publication No. WO 2005/103253 (published April 22, 2005) using the Jotun Hein method. Sequence percent identity calculations performed by the Jotun Hein method (Hein, J. J., *Meth. Enz.*, 183:626-645 (1990)) were done using the MegAlign™ v6.1 program of the LASERGENE bioinformatics computing suite (DNASTAR Inc., Madison, WI) with the default parameters for pairwise alignment (KTUPLE=2). The *Pavlova lutheri* $\Delta 8$ desaturase is 76.4% identical to the *Pavlova salina* $\Delta 8$ desaturase sequence using the Clustal V method. Sequence percent identity calculations performed by the Clustal V method (Higgins, D.G. and Sharp, P.M., *Comput. Appl. Biosci.*, 5:151-153 (1989);

Higgins et al., *Comput. Appl. Biosci.*, 8:189-191 (1992)) were done using the MegAlign™ v6.1 program of the LASERGENE bioinformatics computing suite (DNASTAR Inc.) with the default parameters for pairwise alignment (KTUPLE=1, GAP PENALTY=3, WINDOW=5, DIAGONALS
 5 SAVED=5 and GAP LENGTH PENALTY=10). BLAST scores and probabilities indicate that the fragment of SEQ ID NO:17 encodes an entire *Pavlova lutheri* $\Delta 8$ desaturase.

Figure 8A and 8B show a Clustal V alignment (with default parameters) of SEQ ID NO:18 (the amino acid sequence of the *Pavlova*
 10 *lutheri* $\Delta 8$ desaturase), SEQ ID NO:66 (the amino acid sequence of *Pavlova salina* $\Delta 8$ desaturase sequence, *supra*), SEQ ID NO:16 (the amino acid sequence of *Euglena gracilis* $\Delta 8$ desaturase sequence disclosed as SEQ ID NO:2 in PCT Publication No. WO 2006/012325; published February 2, 2006), SEQ ID NO:65 (the amino acid sequence for
 15 the *Rhizopus stolonifer* $\Delta 6$ fatty acid desaturase (NCBI Accession No. ABB96724, *supra*)) and SEQ ID NO:53 (the amino acid sequence for the *Rhizopus stolonifer* $\Delta 6$ fatty acid desaturase (NCBI Accession No. AAX22052, *supra*)). The results of the Clustal V alignment show that SEQ ID NO:18 is 76.4%, 22.6%, 22.2% and 22.2% identical to SEQ ID NO:66,
 20 SEQ ID NO:16, SEQ ID NO:65 and SEQ ID NO:53, respectively.

EXAMPLE 13

Comparing The Substrate Specificity Of The *Mortierella alpina* $\Delta 5$ Desaturase (MaD5) With The *Peridinium* sp. CCMP626 $\Delta 5$ Desaturase (RD5) In *Yarrowia lipolytica*

25 The present Example describes comparison of the substrate specificity of a *Mortierella alpina* $\Delta 5$ desaturase (MaD5; SEQ ID NOs:67 and 68), which is described in U.S. Patent 6,075,183 and PCT Publication Nos. WO 2004/071467 and No. WO 2005/047479) to that of RD5 (SEQ ID NO:2) in *Yarrowia lipolytica*.

30 This work included the following steps: (1) construction of *Yarrowia* expression vector pY98 comprising MaD5; (2) transformation of pY98 and pDMW368 into *Yarrowia* strain Y2224; and, 3.) comparison of lipid profiles

within transformant organisms comprising pY98 or pDMW368 after feeding fatty acid substrates.

Construction Of *Yarrowia* Expression Vector pY98, Comprising MaD5

Plasmid pY5-22 (SEQ ID NO:69) is a shuttle plasmid that can
5 replicate both in *E. coli* and *Yarrowia lipolytica*, containing the following: a
Yarrowia autonomous replication sequence (ARS18; GenBank Accession
No. M91600); a ColE1 plasmid origin of replication; an ampicillin-
resistance gene (Amp^R) for selection in *E. coli*; a *Yarrowia* URA3 gene
(GenBank Accession No. AJ306421) for selection in *Yarrowia*; and, a
10 chimeric TEF::*NcoI*/*NotI*::XPR cassette, wherein "XPR" was ~100 bp of the
3' region of the *Yarrowia Xpr* gene (GenBank Accession No. M17741).
Although the construction of plasmid pY5-22 is not described herein in
detail, it was derived from pY5 (previously described in PCT Publication
No. WO 2004/101757).

15 Plasmid pY5-22GPD (SEQ ID NO:70) was created from pY5-22
(SEQ ID NO:69), by replacing the TEF promoter with the *Yarrowia*
lipolytica GPD promoter (SEQ ID NO:71) using techniques well known to
one skilled in the art. The *Yarrowia* "GPD promoter" refers to the 5'
upstream untranslated region in front of the 'ATG' translation initiation
20 codon of a protein encoded by the *Yarrowia lipolytica* glyceraldehyde-3-
phosphate dehydrogenase (GPD) gene and that is necessary for
expression (PCT Publication No. WO 2005/003310). More specifically,
the *Yarrowia lipolytica* GPD promoter was amplified from plasmid
pYZDE2-S (SEQ ID NO:72; which was previously described in U.S. Patent
25 Application No. 60/795810 (the contents of which are hereby incorporated
by reference)) using oligonucleotides GPDsense (SEQ ID NO:73) and
GPDantisense (SEQ ID NO:74). The resulting DNA fragment was
digested with *Sall*/*NotI* and cloned into the *Sall*/*NotI* fragment of pY5-22
(SEQ ID NO:69), thus replacing the TEF promoter and *NcoI*/*NotI* site with
30 the GPD promoter and a unique *NotI* site, and thereby producing pY5-
22GPD (SEQ ID NO:70).

The *Mortierella alpina* $\Delta 5$ desaturase gene (SEQ ID NO:67) was
released from pKR136 (SEQ ID NO:75; which was previously described in

PCT Publication No. WO 2004/071467 (the contents of which are hereby incorporated by reference)) by digestion with *NotI* and cloned into the *NotI* site of pY5-22GPD to produce pY98 (SEQ ID NO:76; Figure 9).

Transformation Of pY98 (Comprising MaD5) And pDMW368 (Comprising RD5) Into *Yarrowia* Strain Y2224 And Comparison Of Lipid Profiles

Strain Y2224 was isolated in the following manner: *Yarrowia lipolytica* ATCC #20362 cells from a YPD agar plate (1% yeast extract, 2% bactopectone, 2% glucose, 2% agar) were streaked onto a MM plate (75 mg/L each of uracil and uridine, 6.7 g/L YNB with ammonia sulfate, without amino acid, and 20 g/L glucose) containing 250 mg/L 5-FOA (Zymo Research). Plates were incubated at 28 °C and four of the resulting colonies were patched separately onto MM plates containing 200 mg/mL 5-FOA and MM plates lacking uracil and uridine to confirm uracil *Ura3* auxotrophy.

Strain Y2224 was transformed with pY98 (SEQ ID NO:76, Figure 9) and pDMW368 (SEQ ID NO:23; Figure 5C; Example 6) as described in the General Methods.

Single colonies of transformant *Yarrowia lipolytica* containing pY98 (SEQ ID NO:76) or pDMW368 (SEQ ID NO:23) were grown in 3 mL MM lacking uracil supplemented with 0.2% tergitol at 30 °C for 1 day. After this, 0.1 mL was transferred to 3 mL of the same medium supplemented with either EDA, ETrA, DGLA, ETA or no fatty acid. These were incubated for 16 h at 30 °C, 250 rpm and then pellets were obtained by centrifugation. Cells were washed once with water, pelleted by centrifugation and air dried. Pellets were transesterified (Roughan, G. and Nishida, I., *Arch. Biochem. Biophys.*, 276(1):38-46 (1990)) with 500 µL of 1% sodium methoxide for 30 min at 50 °C after which 500 µL of 1M NaCl and 100 µL of heptane were added. After thorough mixing and centrifugation, fatty acid methyl esters (FAMES) were analyzed by GC.

FAMES (5 µL injected from hexane layer) were separated and quantified using a Hewlett-Packard 6890 Gas Chromatograph fitted with an Omegawax 320 fused silica capillary column (Catalog No. 24152, Supelco Inc.). The oven temperature was programmed to hold at 220 °C

for 2.6 min, increase to 240 °C at 20 °C/min and then hold for an additional 2.4 min. Carrier gas was supplied by a Whatman hydrogen generator. Retention times were compared to those for methyl esters of standards commercially available (Nu-Chek Prep, Inc.).

5 The fatty acid profiles for *Yarrowia lipolytica* expressing pY98 (SEQ ID NO:76) or pDMW368 (SEQ ID NO:23) and fed various substrates are shown in Figure 10A. In Figure 10A shading indicates the substrates fed and products produced; fatty acids are identified as 16:0 (palmitate), 16:1, 18:0 (stearic acid), 18:1 (oleic acid), LA, GLA, ALA, STA, EDA, SCI
10 (sciadonic acid or *cis*-5,11,14-eicosatrienoic acid; 20:3 ω -6), DGLA, ARA, ETrA, JUP (juniperonic acid or *cis*-5,11,14,17-eicosatrienoic acid; 20:4 ω -3), ETA and EPA. Fatty acid compositions were expressed as the weight percent (wt. %) of total fatty acids.

 Percent Δ 5 desaturation ("% delta-5 desat") of RD5 and MaD5 for
15 each substrate is shown in Figure 10B and was calculated by dividing the wt. % for product (either SCI, JUP, ARA or EPA) by the sum of the wt. % for the substrate and product (either EDA and SCI, ETrA and JUP, DGLA and ARA, or ETA and EPA, respectively) and multiplying by 100 to express as a %, depending on which substrate was fed.

20 The activities of MaD5 and RD5 are compared using the ratio of the percent Δ 5 desaturation ("Ratio Desat R/Ma") in Figure 10B and are calculated by dividing the percent Δ 5 desaturation for RD5 on a particular substrate by the percent Δ 5 desaturation for MaD5 on the same substrate.

 The substrate specificity of RD5 and MaD5 for the correct ω -6 fatty
25 acid substrate (i.e., DGLA) versus the by-product fatty acid (i.e., SCI) or the correct ω -3 fatty acid substrate (i.e., ETA) versus the by-product fatty acid (i.e., JUP) is also shown in Figure 10B. Specifically, the substrate specificity ("Ratio Prod/By-Prod") for ω -6 substrates was calculated by dividing the percent Δ 5 desaturation (% delta-5 desat) for DGLA by the
30 percent Δ 5 desaturation (% delta-5 desat) for EDA and is shown on the same lines as the results for DGLA. The substrate specificity ("Ratio Prod/By-Prod") for ω -3 substrates was calculated by dividing the percent Δ 5 desaturation (% delta-5 desat) for ETA by the percent Δ 5 desaturation

(% Δ -5 desat) for ETrA and is shown on the same lines as the results for ETA. Furthermore, the ratio of substrate specificity ("Ratio Prod/By-Prod R/Ma") for ω -6 substrates was determined by dividing the substrate specificity for RD5 on the ω -6 substrates (i.e., DGLA/EDA) by that for MaD5. The ratio of substrate specificity ("Ratio Prod/By-Prod R/Ma") for ω -3 substrates was calculated by dividing the substrate specificity for RD5 on the ω -3 substrates (i.e., ETA/ETrA) by that for MaD5.

The preference of RD5 and MaD5 for ω -6 or ω -3 substrates is compared using the ratio of the percent Δ 5 desaturation ("Ratio n-6/n-3") in Figure 10B and is calculated by dividing the percent Δ 5 desaturation for RD5 and MaD5 on a particular ω -6 substrate (either DGLA or EDA) by the percent Δ 5 desaturation on the corresponding ω -3 substrate (either ETA or ETrA, respectively).

From the results in Figure 10B, it is clear that RD5 is approximately 3.0- to 9.7-fold more active in *Yarrowia* than MaD5 when DGLA, EDA, ETrA and ETA are used as substrates. The substrate specificity of RD5 compared to MaD5 (RD5/MaD5) for the correct ω -6 substrate (i.e., DGLA versus EDA) is approximately 0.4 and for the ω -3 substrate (i.e., ETA versus ETrA) is approximately 0.6. RD5 also has an approximate 1.4-fold preference for ω -6 substrates (i.e., EDA and DGLA) over the ω -3 substrates (i.e., ETrA and ETA), respectively.

CLAIMS

What is claimed is:

- 5 1. An isolated nucleic acid molecule selected from the group consisting of:
 - (a) an isolated nucleotide sequence encoding a $\Delta 5$ desaturase enzyme as set forth in SEQ ID NO:2;
 - 10 (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).
- 15 2. The isolated nucleic acid molecule of Claim 1, wherein at least 229 codons are codon-optimized for expression in *Yarrowia*.
3. The isolated nucleic acid molecule of Claim 1 selected from the group of SEQ ID NOs:1 and 3.
4. An isolated nucleic acid fragment of Claim 1 isolated from
20 *Peridinium* sp. CCMP626.
5. An isolated nucleic acid molecule comprising a first nucleotide sequence encoding a $\Delta 5$ desaturase enzyme of at least 463 amino acids that has at least 95% identity based on BLASTP algorithms when compared to a polypeptide having the sequence as set forth in SEQ ID
25 NO:2;
 - or a second nucleotide sequence comprising the complement of the first nucleotide sequence.
6. A polypeptide encoding a $\Delta 5$ desaturase enzyme as set forth in SEQ ID NO:2.
- 30 7. A chimeric gene comprising the isolated nucleic acid molecule of any one of Claims 1-5 operably linked to at least one regulatory sequence.
8. A transformed *Yarrowia* sp. comprising the isolated nucleic acid molecule of Claim 1.

9. A method for the production of arachidonic acid comprising:

a) providing a host cell comprising:

(i) an isolated nucleotide molecule encoding a $\Delta 5$ desaturase polypeptide having at least 95% identity when compared to a polypeptide having the amino acid sequence as set forth in SEQ ID NO:2, based on BLASTP algorithms; and,

(ii) a source of dihomo- γ -linoleic acid;

b) growing the host cell of step (a) under conditions wherein the nucleic acid fragment encoding the $\Delta 5$ desaturase polypeptide is expressed and the dihomo- γ -linoleic acid is converted to arachidonic acid; and,

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

10. A method for the production of eicosapentaenoic acid

comprising:

a) providing a host cell comprising:

(i) an isolated nucleotide molecule encoding a $\Delta 5$ desaturase polypeptide having at least 95% identity when compared to a polypeptide having the amino acid sequence as set forth in SEQ ID NO:2, based on BLASTP algorithms; and,

(ii) a source of eicosatetraenoic acid;

b) growing the host cell of step (a) under conditions wherein the nucleic acid fragment encoding the $\Delta 5$ desaturase polypeptide is expressed and the eicosatetraenoic acid is converted to eicosapentaenoic acid; and,

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

11. A method according to either of Claims 9 or 10 wherein the isolated nucleic acid molecule encoding the $\Delta 5$ desaturase polypeptide is set forth in SEQ ID NO:2, wherein at least 229 codons are codon-optimized for expression in *Yarrowia*.

12. A method according to either of Claims 9 or 10 wherein the isolated nucleic acid molecule encodes a $\Delta 5$ desaturase polypeptide having the amino acid sequence as set forth in SEQ ID NO:2.

13. A method according to either of Claims 9 or 10 wherein:

- 5 a.) the isolated nucleic acid molecule has the nucleic acid sequence selected from the group consisting of SEQ ID NOs:1 and SEQ ID NO:3; and,
b.) the host cell is *Yarrowia lipolytica*.

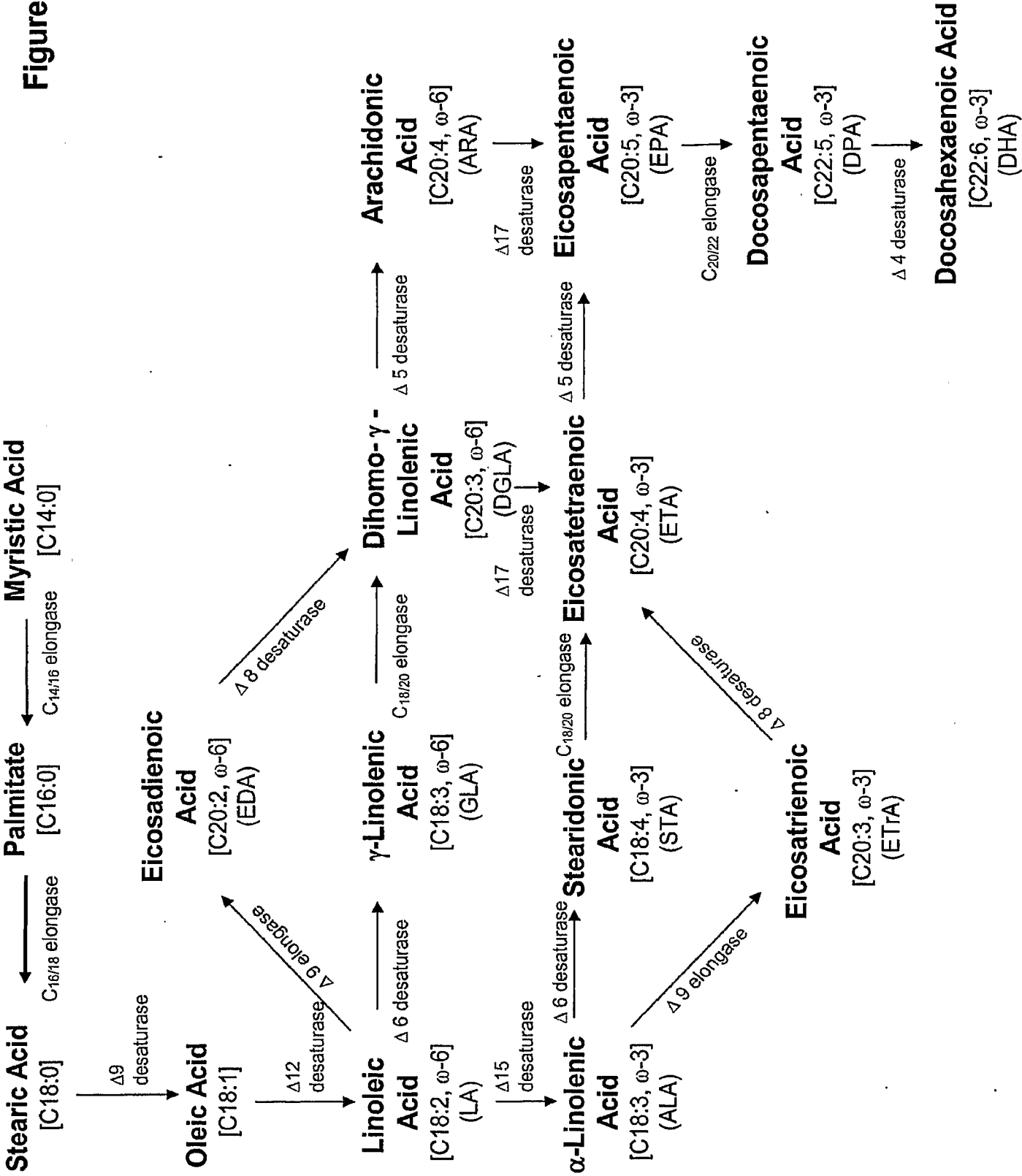
10 14. A method according to either of Claims 9 or 10, wherein the host cell is selected from the group consisting of: algae, bacteria, yeast, and fungi.

15 15. A method according to Claim 14 wherein the host cell is a fungus selected from the group consisting of: *Thraustochytrium* sp., *Schizochytrium* sp. and *Mortierella* sp.

16 16. A method according to Claim 14 wherein the yeast is an oleaginous yeast.

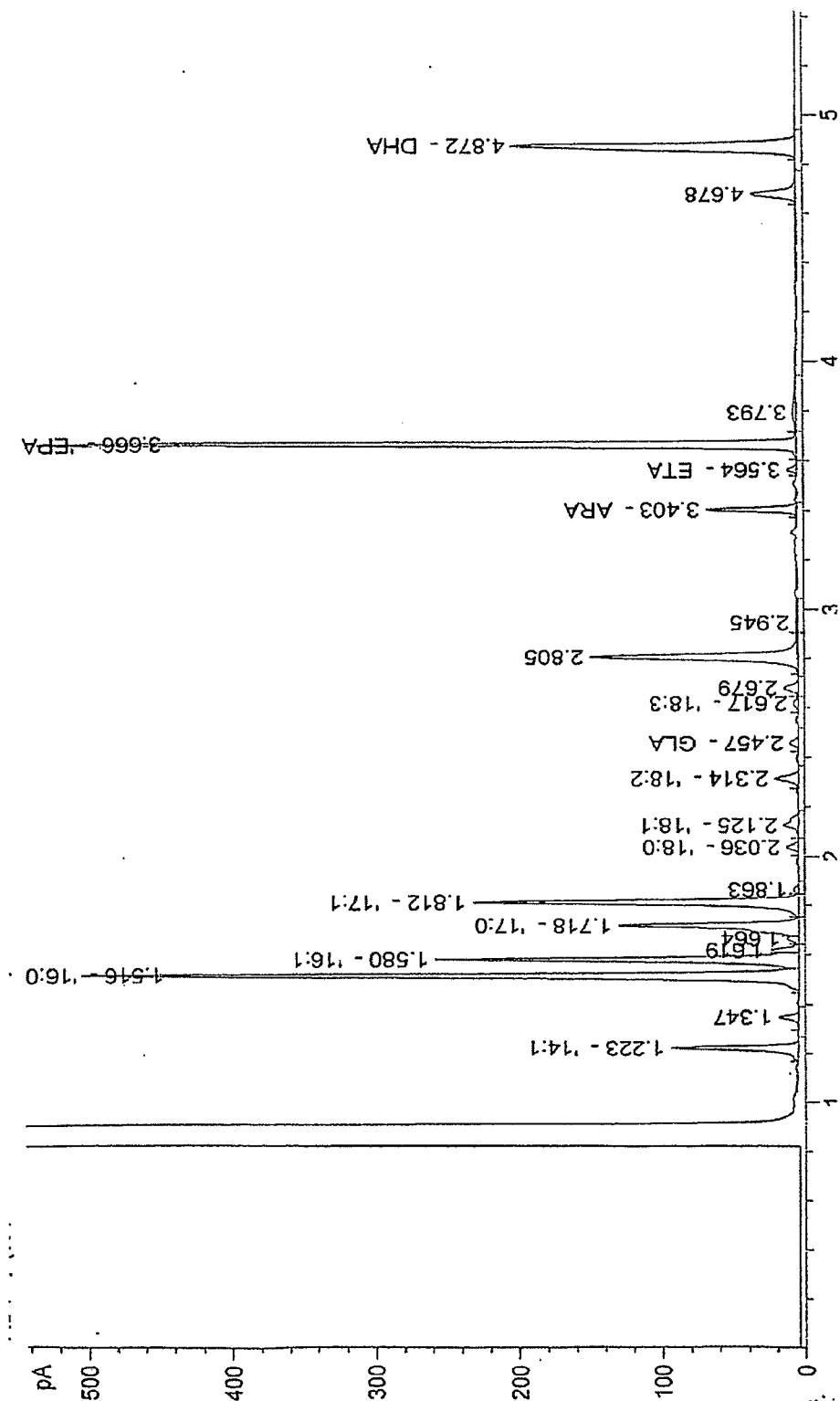
17. A method according to Claim 16 wherein the oleaginous yeast is selected from the group consisting of: *Yarrowia*, *Candida*, *Rhodotorula*, *Rhodospiridium*, *Cryptococcus*, *Trichosporon* and *Lipomyces*.

Figure 1



2/12

Figure 2



205	W	Q	E	Q	H	W	T	-	H	H	A	Y	T	N	H	A	E	M	D	P	D	S	F	G	A	E	P	M	L	L	F	N	D	-	-	Y	P	L	D	H	SEQ ID NO:14
225	W	L	N	Q	H	V	V	G	H	H	I	Y	T	N	V	A	G	A	D	P	D	-	L	P	V	D	F	K	S	D	V	R	-	-	R	I	V	Y	R	SEQ ID NO:13	
204	W	L	N	Q	H	V	V	G	H	H	I	Y	T	N	V	A	G	A	D	P	D	-	L	P	V	D	F	E	S	D	V	R	-	-	R	I	V	H	R	SEQ ID NO:12	
207	W	C	H	Q	H	V	I	G	H	H	L	Y	T	N	V	R	N	A	D	P	D	-	L	G	Q	G	-	E	I	D	F	R	-	-	V	T	P	Y	SEQ ID NO:15		
178	W	K	D	R	H	N	A	-	H	H	S	A	T	N	V	Q	G	H	D	P	D	-	I	D	N	L	P	L	A	W	S	E	D	D	V	T	R	A	S	SEQ ID NO:16	
205	W	R	A	R	H	N	T	-	H	H	V	C	T	N	E	D	G	S	D	P	D	-	I	K	T	A	P	L	L	I	Y	V	R	E	N	P	S	I	A	K	SEQ ID NO:18

SEQ ID NO:14	SEQ ID NO:13	SEQ ID NO:12	SEQ ID NO:15	SEQ ID NO:16	SEQ ID NO:18
242	P A R T W L H R F Q	A F F Y M P V L A G Y W L S A V F N P Q I L D - L Q Q R G A	242	242	243
261	Q V L L P I Y K Y Q	- H L Y L P P L Y G V L G - - - L K F R V Q D D V F E T F V T	261	240	243
240	Q V L L P I Y K F Q	- H I Y L P P L Y G V L G - - - L K F R I Q D D V F E T F V S	240	242	243
242	Q A R S W Y H K Y Q	- H I Y A P I L Y G V Y A - - - L K Y R I Q D D - H E I F T K	242	216	243
216	P I S R K L I Q F Q	- Q Y Y F L V I C I L L R - - - F I W C F Q S V L T V R S L	216	243	243
243	R L N - F F Q R W Q	- Q Y Y V P T M A I L D - - - L Y W R L E S I A Y V A V R	243		

281	L	S	V	G	I	R	L	D	N	A	F	I	H	S	R	R	K	Y	A	V	F	W	R	A	V	Y	I	A	V	N	V	I	A	P	F	Y	T	N	S	G	SEQ	ID	NO:14
297	L	T	N	G	P	L	R	V	N	P	L	S	V	G	D	W	A	E	M	I	L	S	K	A	F	W	V	F	Y	R	I	Y	L	P	L	A	V	L	Q	V	SEQ	ID	NO:13
276	L	T	N	G	P	V	R	V	N	P	H	P	V	S	D	W	V	Q	M	I	F	A	K	A	F	W	T	F	Y	R	I	Y	I	P	L	V	W	L	K	I	SEQ	ID	NO:12
277	K	S	N	G	A	I	R	Y	S	P	I	S	T	I	D	T	A	I	F	I	L	G	K	L	V	F	I	I	S	R	F	I	L	P	L	I	Y	N	H	S	SEQ	ID	NO:15
252	K	D	-	-	-	-	R	D	N	Q	F	Y	R	S	Q	Y	K	K	E	A	I	G	L	A	L	H	W	T	L	K	T	L	F	H	L	F	F	M	P	S	SEQ	ID	NO:16
278	-	-	-	-	-	-	-	-	-	-	-	L	P	K	M	W	M	Q	A	A	A	L	A	A	H	Y	A	L	-	-	L	C	W	V	F	A	A	H	L	SEQ	ID	NO:18	

	321	337	316	317	288	304	SEQ ID NO:14	SEQ ID NO:13	SEQ ID NO:12	SEQ ID NO:15	SEQ ID NO:16	SEQ ID NO:18
-	L E W S	W R V F	G N I M	L M G V A E S L A L A V L	F S L S H N F E S A D R D P							
D P A R F W G V F	-	-	-	F L A E F S T G W Y L A F N F	Q V S H V S T A C E Y P G							
T P S T F W G V F	-	-	-	F L A E F T T G W Y L A F N F	Q V S H V S T E C E Y P C							
- F S S H L I C F F	-	-	-	L I S E L V L G W Y L A I S	F Q V S H V V E D L Q F M A							
I L T S L L V F F	-	-	-	V S E L V G G F G I A I V	V F M N H Y P L E K I G D S							
N L I P L M M V A	-	-	-	R - - - G F A T G I V	V F F A T H Y G E D I L D R E							

Figure 4

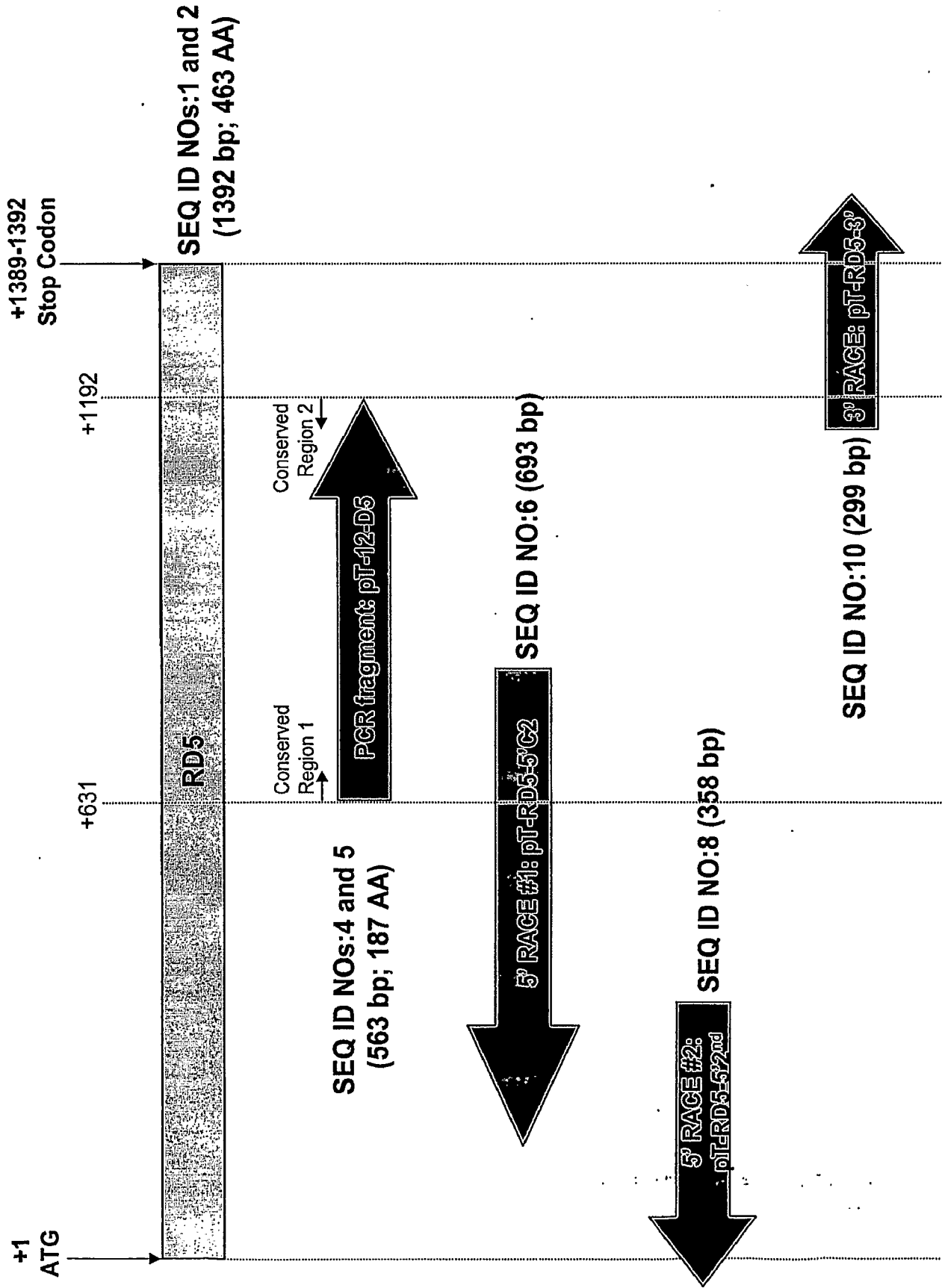


Figure 5

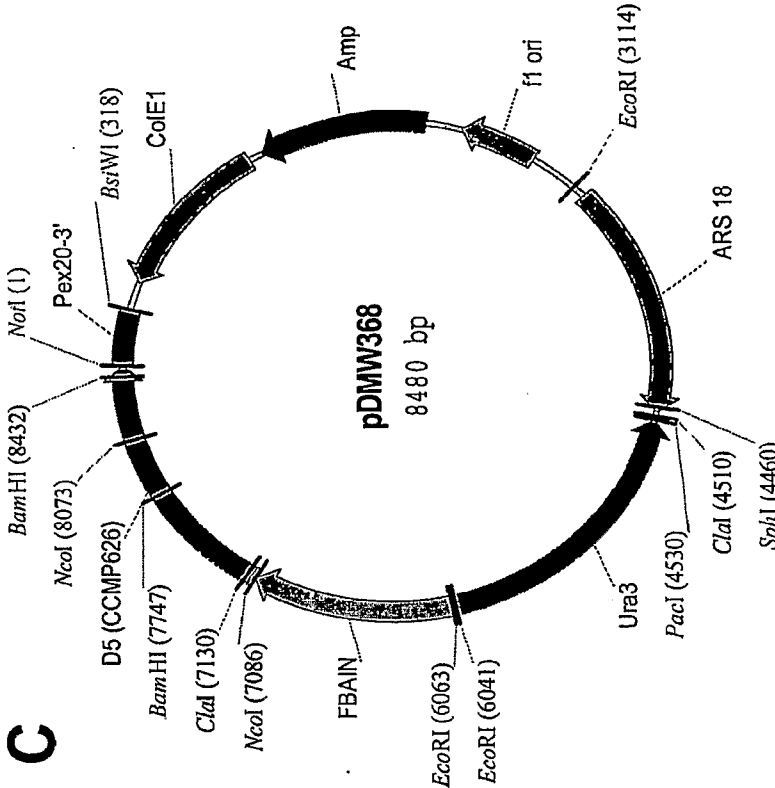
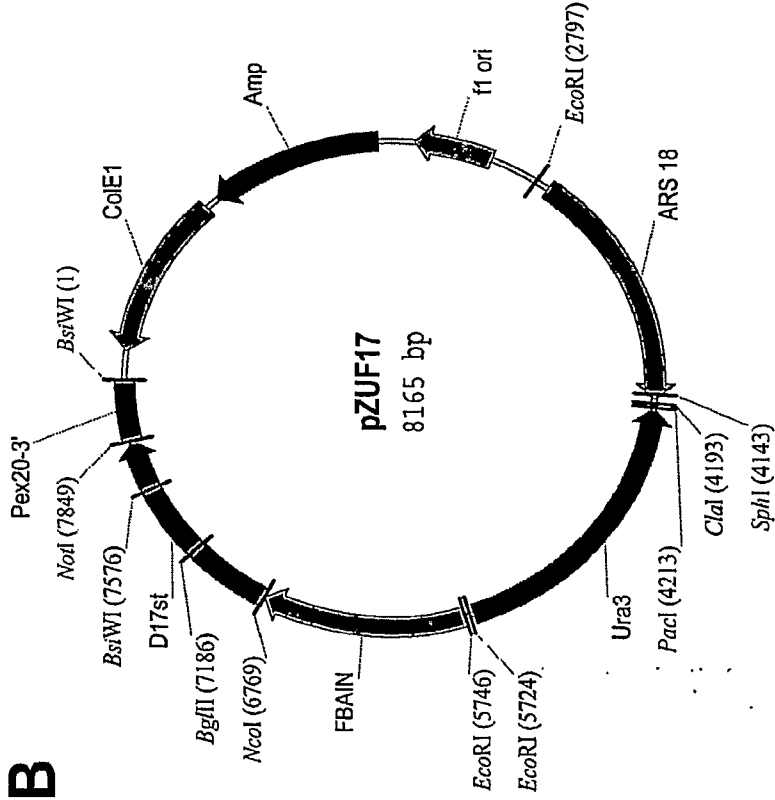
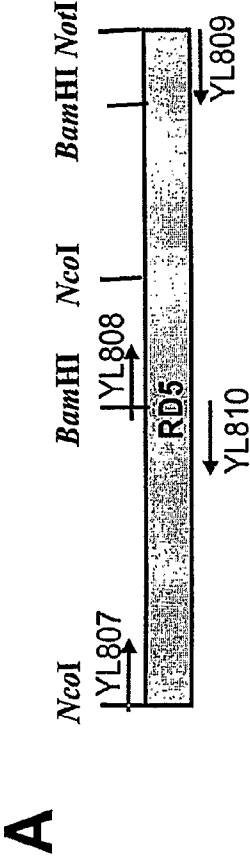


Figure 6

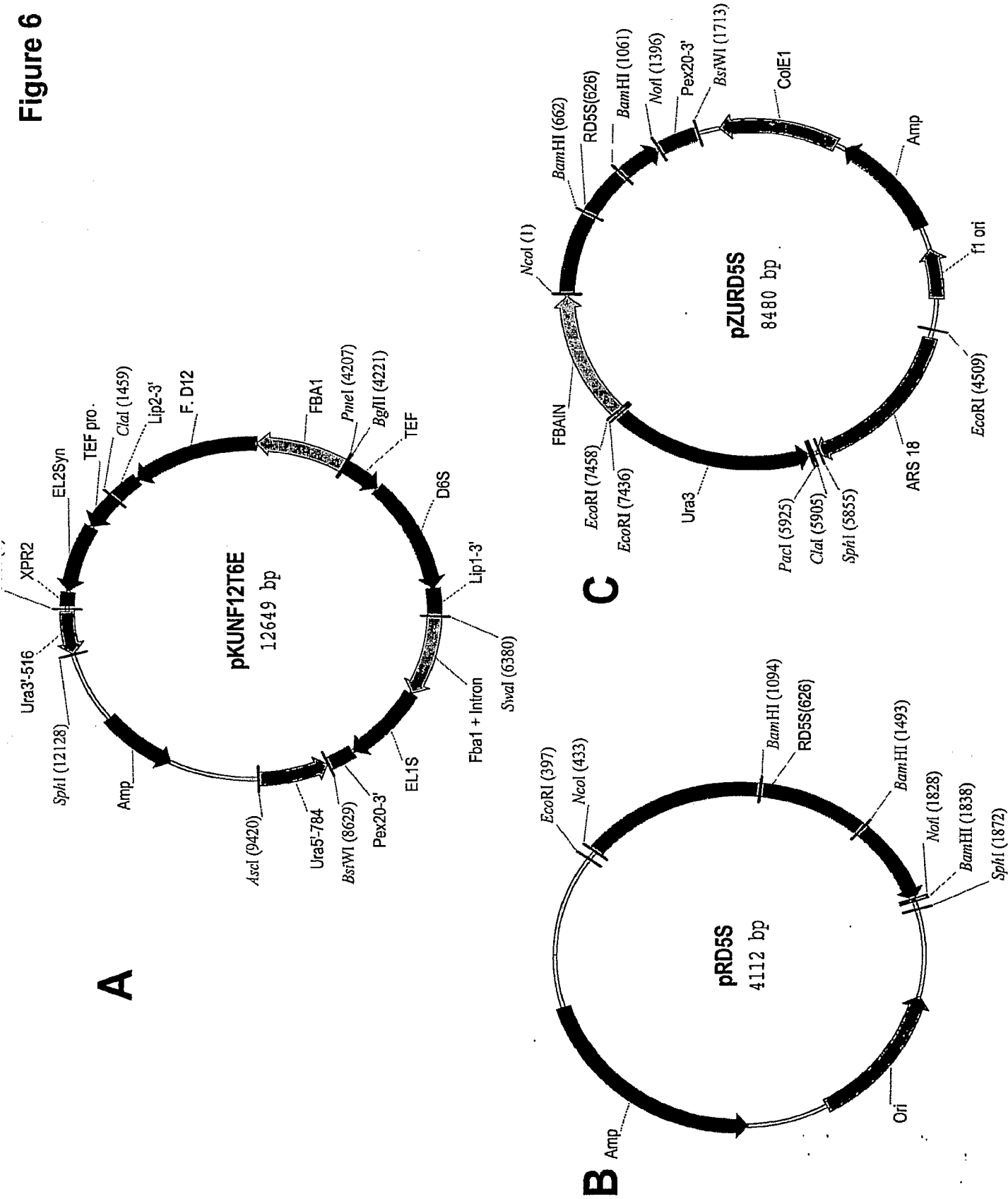


Figure 7A

7/12

1	ATGGCTTCCAGATGCGGACAAGTTTGA	GACAGCGCAAGGCGCAATCGATTTCAA	GACACGCGCT	SEQ	ID NO:1	
1	ATGGCTTCCGACGCGCAAGCTTGGCA	CAAGGCTCAGTCCATCCAGGACAC	TGCC	SEQ	ID NO:3	
61	GATTCGCAAGCTACCGA	ACTCAAAGATTGGCA	CCCTGAAGGCTTTC	AGGCAAGAAATC	SEQ	ID NO:1
61	GATTCAGGCTACCGAGCTCAAGATTGGCA	CCCTGAAGGCTTTC	CAAGGC	ACCGAGATC	SEQ	ID NO:3
121	GTCAATGATGGAGACATTTACGATAT	AAAGACTTTTGATCACCCCGGTGGTGA	ATCCATC	SEQ	ID NO:1	
121	GTCAATGATGGGACATCTACGACATCA	CAAGACTTCGATCACCCCTGGAGGC	GAATCCATC	SEQ	ID NO:3	
181	ATGACTTTTGGGGA	AAACGATGTCAACGCCACGTACA	AGATGATCACCCCTACCACTCT	SEQ	ID NO:1	
181	ATGACTTTGGTGGCAACGACGTTACT	GTCCACCTACAAGATGATTCATCCCTACCACTCG	SEQ	ID NO:3		
241	AAGCACCATTTGGAGAAAGATGA	AAAGAAAGTGGGACGAGTTCGGGACTA	CACCTCGGAAATAC	SEQ	ID NO:1	
241	AAGCATCATCTGGAGAAAGATGA	AAAGTTCGGTCGAGTCCCGACTACACCTCCGAGTAC	SEQ	ID NO:3		
301	AAGTTTGATACTCCCTTTGAGCGTGAA	ATCAAGCAAGAGGCTTTCGCTACAT	TGTCGACGACGA	SEQ	ID NO:1	
301	AAGTTTGATACTCCCTTTGACGAGAGATCA	AAACAGGAGGCTTTCAGAGATTTGTCGACGA	AGA	SEQ	ID NO:3	
361	GGC	CGGAGTTTGGAAACACCTGGATACTTCTTC	CGGGCTTTCGCTACATTTGGACTTTTC	SEQ	ID NO:1	
361	GGTCGAGAGTTTGGAAACACCTGGCTACTTCTTTC	CGAGCTTTCGCTACATCGGTCCTTC	SEQ	ID NO:3		
421	TTT	TACTTGCAGTATTTGTGGGTCA	CGACTCCACTACCTTTGCTTCTCTAT	SEQ	ID NO:1	
421	TTT	TACTTGCAGTATCTTGGGTTACTACTCCCTAC	CACTTTGCTTCTCTATCTCTAC	SEQ	ID NO:3	
481	GGTGT	TTTCGCAAGCTTTCAATGGTTTGA	ACGTACAACATGATGCCAACCA	CGGAGCTGCC	SEQ	ID NO:1
481	GGTGT	TTTCAGGCTTTCAATGGCTTGA	ACGTCCAGCACGCCAACCA	CGGAGCTGCC	SEQ	ID NO:3
541	TCCA	AGAAGCCCTTGGATCAATAA	CTTGCTAGGATTGGGCTGACTTTAT	TCGGAGGTTCC	SEQ	ID NO:1
541	TCCA	AGAAGCCCTTGGATCAACAATTGCTCGGC	CTGGGTGCTGACTTTAT	TCGGAGGCTCC	SEQ	ID NO:3
601	AA	TGGTTGTGGATGAACAGCACTGGACGCAC	CAACAATACACCAACCA	CATGAGAAAG	SEQ	ID NO:1
601	AAGTGGCT	CTGGATGAACAGCACTGGACCCAT	CACACTTACACCAACCA	TCAGAGAAAG	SEQ	ID NO:3
661	GATCCCGATGCCCTTGGGCGCTGA	ACCAATGTTGTTGTTCAATGATTA	TATCCCTTGGGTCA	C	SEQ	ID NO:1
661	GATCCCGACGCCCTTGGGTGCA	AGACCTATGCTCTTCAACGACTAT	TATCCCTTGGGTCA	C	SEQ	ID NO:3

Figure 7B

8/12

721 CCAAGCGTACTTTGATTCAACACTTCCAGGCCCTTCTATTACCTTTTGGTCTTGGCCGGA SEQ ID NO:1
721 CCAAGCGAACCCTCAATTCATCACTTCCAAGCCTTCTACTATCTGTGTTGTCCTTGCTGGC SEQ ID NO:3

781 TACTGGGTCTCTTCGGTCTTCAACCCCTCAAAATTTTGGACTTGCAACACCCGGGTGCTCAA SEQ ID NO:1
781 TACTGGGTGCTCTCGGTGTTCAACCCCTCAGATCCCTGGACCTCCAGCACCGAGGTGCCAG SEQ ID NO:3

841 GCGGTTGGAATGAATAATGGAGAAACGATTACATTGCCAAAGCCGAAAGTATGCCATCTTC SEQ ID NO:1
841 GCTGTGGCATGAAGATGGAGAACGACTACATTGCCAAGTCTCGAAAGTACGCTATCTTC SEQ ID NO:3

901 TTGCGTCTCTTGATATATTACACCAACATTTGTGCTCCGATCCAAACCAAGGCTTCTCG SEQ ID NO:1
901 CTGCGACTCTCTGTACATCTACACCAACATTTGTGCTCCCATCCAGAACCAAGGCTTTTCG SEQ ID NO:3

961 TTGACCGGTGGTGGCCACATTTTGACCATGGGCGTCCGCTTCCAGTTTGACTTTGGCGACT SEQ ID NO:1
961 CTCACCGTCTGTGCTCACATTTTACTATGGGTGTCGCTCCAGCCTTGACCTCGCTACT SEQ ID NO:3

1021 CTTTGTGCTTGTGCGACAAATTTTGAAACGCGGATCGCGATCCCACTTACGAGGCCCGC SEQ ID NO:1
1021 CTGTTGGCCCTCTCCACAACTTCGAGAACGCGAGATCGGGATCCCACTTACGAGGCTCGA SEQ ID NO:3

1081 AAGGAGGAGAGCCCTGTTTGTGTTCAAGTCGCAAGTCCGAAACCTCGTCAACTTACGGA SEQ ID NO:1
1081 AAGGAGGCGAGCCCTGTTCTGTTGTTCAAGTCGAGGTGGAAACCTCTCTACTTACGGT SEQ ID NO:3

1141 GGTTCATCTCGGGTTGCTTGACGGGCGGACTCAACTTCCAAAGTGGAACAACACTTGTTT SEQ ID NO:1
1141 GGCTTCATTTCCGGTTGCCCTTACAGCCGGACTCAACTTTCAGGTGAGCATCACTGTTT SEQ ID NO:3

1201 CCTCGTATGAGTTCCGGCCTGGTACCCCTACATTGCCCTACTGTTGAGAGGTTTGCAAA SEQ ID NO:1
1201 CCTCGAATGTCCTCTGCCCTGGTACCCCTACATCGCTCCTACCGTTCCGAGAGGTCTGCAAA SEQ ID NO:3

1261 AAGCACGGAAGTCAAGTACGCACTACTATCCCTGGGTCTGGCAAAACCTTGATTTCAACTGTC SEQ ID NO:1
1261 AAGCACGGGCTCAAGTACGCCCTACTATCCCTGGGTGTTGGCAGAACCTCATCTCGACCTGTC SEQ ID NO:3

1321 AAGTATCTGCATCAAAAGCGGAACCTGGATCCAACTGGAAGAATGGCGCCAAACCCCTACTCG SEQ ID NO:1
1321 AAGTACTGTCATCAGTCCGGAACTGGCTCGAACTGGAAGAAGACGGTGCCAAATCCCTACTCT SEQ ID NO:3

1381 GGAAAAATTGTAA SEQ ID NO:1
1381 GGCAAGCTGTAA SEQ ID NO:3

Figure 8A

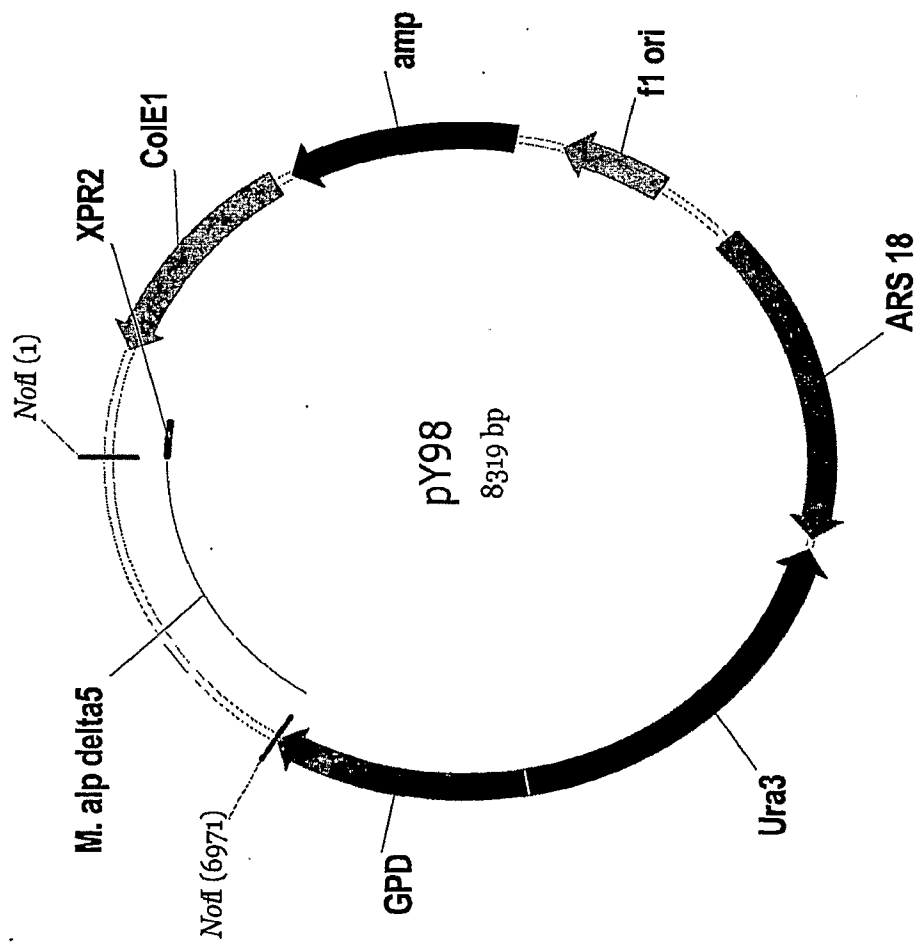
1	MKGGDGGGAQAVSGTDASL	10	20	30	40	SEQ ID NO:18 (Pavlova lutheri).pro
1	MGRGDSGGQAHPAAELAVPSDRAEVSNA	10	20	30	40	SEQ ID NO:66 (Pavlova salina).pro
1	MK--S-KRQALP	10	20	30	40	SEQ ID NO:16 (Euglena gracilis).pro
1	MS--TLDQRSIFTIKELESISQRIHDGDEE	10	20	30	40	SEQ ID NO:65 (Rhizopus stolonifer).pro
1	MS--TLDQRSIFTIKELESISQRIHDGDEE	10	20	30	40	SEQ ID NO:53 (Rhizopus stolonifer).pro
37	KRV-DVTKFQKAHPGGSKVFRI	50	60	70	80	SEQ ID NO:18 (Pavlova lutheri).pro
41	KRV-DVTKFQRTHPGGSKVFRI	50	60	70	80	SEQ ID NO:66 (Pavlova salina).pro
14	GTTYDVSAWVNHPGGAEI	50	60	70	80	SEQ ID NO:16 (Euglena gracilis).pro
37	KKVVDTVTEFIEDHPGGAQVLL	50	60	70	80	SEQ ID NO:65 (Rhizopus stolonifer).pro
37	KKVVDTVTEFIEDHPGGAQVLL	50	60	70	80	SEQ ID NO:53 (Rhizopus stolonifer).pro
76	IKMMEGMLKKSEDA	90	100	110	120	SEQ ID NO:18 (Pavlova lutheri).pro
80	IKMMEGMLKKSEDA	90	100	110	120	SEQ ID NO:66 (Pavlova salina).pro
54	FDKLR-MPKINPSSELPQA	90	100	110	120	SEQ ID NO:16 (Euglena gracilis).pro
77	YEVLNNYFVGVQETVVT	90	100	110	120	SEQ ID NO:65 (Rhizopus stolonifer).pro
77	YEVLNNYFVGVQETVVT	90	100	110	120	SEQ ID NO:53 (Rhizopus stolonifer).pro
116	GLYDPCPLDELFLKLT	130	140	150	160	SEQ ID NO:18 (Pavlova lutheri).pro
120	GYDPCPLDELFLKLT	130	140	150	160	SEQ ID NO:66 (Pavlova salina).pro
93	GMFDASPLWYSYKIST	130	140	150	160	SEQ ID NO:16 (Euglena gracilis).pro
117	GYFHSKLFYAYKVLST	130	140	150	160	SEQ ID NO:65 (Rhizopus stolonifer).pro
117	GYFHSKLFYAYKVLST	130	140	150	160	SEQ ID NO:53 (Rhizopus stolonifer).pro
151	AGALSMGFGFYLDGWL	170	180	190	200	SEQ ID NO:18 (Pavlova lutheri).pro
156	CGALMVSFGWYLDGWL	170	180	190	200	SEQ ID NO:66 (Pavlova salina).pro
128	IGAVLLGMHYQQMGWL	170	180	190	200	SEQ ID NO:16 (Euglena gracilis).pro
157	ASAITVGI	170	180	190	200	SEQ ID NO:65 (Rhizopus stolonifer).pro
157	ASAITVGI	170	180	190	200	SEQ ID NO:53 (Rhizopus stolonifer).pro
191	GYALG-FLQGYDVA	210	220	230	240	SEQ ID NO:18 (Pavlova lutheri).pro
196	GYFLG-FVQGYAVE	210	220	230	240	SEQ ID NO:66 (Pavlova salina).pro
163	GLVFGNGLQGFST	210	220	230	240	SEQ ID NO:16 (Euglena gracilis).pro
192	VVFLGNFCQGFSL	210	220	230	240	SEQ ID NO:65 (Rhizopus stolonifer).pro
192	VVFLGNFCQGFSL	210	220	230	240	SEQ ID NO:53 (Rhizopus stolonifer).pro

Figure 8B

10/12

230	LLIY	250	260	270	280	SEQ ID NO:18 (Pavlova lutheri).pro
235	LLIY	250	260	270	280	SEQ ID NO:66 (Pavlova salina).pro
203	LAWSE	250	260	270	280	SEQ ID NO:16 (Euglena gracilis).pro
232	VLLWDEYASAA	250	260	270	280	SEQ ID NO:65 (Rhizopus stolonifer).pro
232	VLLWDEYASAA	250	260	270	280	SEQ ID NO:53 (Rhizopus stolonifer).pro
286	VPTMAI	290	300	310	320	SEQ ID NO:18 (Pavlova lutheri).pro
261	VPVMAI	290	300	310	320	SEQ ID NO:66 (Pavlova salina).pro
230	LVI	290	300	310	320	SEQ ID NO:16 (Euglena gracilis).pro
272	FIL	290	300	310	320	SEQ ID NO:65 (Rhizopus stolonifer).pro
272	FIL	290	300	310	320	SEQ ID NO:53 (Rhizopus stolonifer).pro
287	---	330	340	350	360	SEQ ID NO:18 (Pavlova lutheri).pro
292	---	330	340	350	360	SEQ ID NO:66 (Pavlova salina).pro
286	EAI	330	340	350	360	SEQ ID NO:16 (Euglena gracilis).pro
305	ERV	330	340	350	360	SEQ ID NO:65 (Rhizopus stolonifer).pro
305	ERV	330	340	350	360	SEQ ID NO:53 (Rhizopus stolonifer).pro
316	ATGI	370	380	390	400	SEQ ID NO:18 (Pavlova lutheri).pro
321	GTGI	370	380	390	400	SEQ ID NO:66 (Pavlova salina).pro
305	GIAI	370	380	390	400	SEQ ID NO:16 (Euglena gracilis).pro
344	TLAL	370	380	390	400	SEQ ID NO:65 (Rhizopus stolonifer).pro
344	TLAL	370	380	390	400	SEQ ID NO:53 (Rhizopus stolonifer).pro
356	WLVN	410	420	430	440	SEQ ID NO:18 (Pavlova lutheri).pro
361	WLVN	410	420	430	440	SEQ ID NO:66 (Pavlova salina).pro
345	IITD	410	420	430	440	SEQ ID NO:16 (Euglena gracilis).pro
384	PLGD	410	420	430	440	SEQ ID NO:65 (Rhizopus stolonifer).pro
384	PLGD	410	420	430	440	SEQ ID NO:53 (Rhizopus stolonifer).pro
396	HGLE	450	460	470		SEQ ID NO:18 (Pavlova lutheri).pro
401	HGLE	450	460	470		SEQ ID NO:66 (Pavlova salina).pro
385	HNLP	450	460	470		SEQ ID NO:16 (Euglena gracilis).pro
424	YDIN	450	460	470		SEQ ID NO:65 (Rhizopus stolonifer).pro
424	YDIN	450	460	470		SEQ ID NO:53 (Rhizopus stolonifer).pro

Figure 9



12/12

Figure 10

A

Fatty Acid	Clone	Gene	Fatty acid composition (wt.%)															
			16:0	16:1	18:0	18:1	LA	GLA	ALA	STA	EDA	SCI	DGLA	ARA	ETra	JUP	ETA	EPA
None	pY98	MaD5	9.5	9.5	0.7	37.4	42.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	pDMW368	RD5	10.0	9.3	0.8	39.4	40.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
EDA	pY98	MaD5	9.8	8.2	0.8	39.7	34.1	0.0	0.0	0.0	7.1	0.2	0.0	0.0	0.0	0.0	0.0	0.0
	pDMW368	RD5	8.9	8.2	1.1	41.2	32.7	0.0	0.0	0.0	5.4	2.6	0.0	0.0	0.0	0.0	0.0	0.0
DGLA	pY98	MaD5	12.0	6.8	1.4	38.8	19.3	0.6	0.0	0.0	0.0	0.0	18.2	3.0	0.0	0.0	0.0	0.0
	pDMW368	RD5	12.1	7.0	1.2	40.3	18.3	0.6	0.0	0.0	0.0	0.0	10.2	10.3	0.0	0.0	0.0	0.0
ETra	pY98	MaD5	8.9	7.2	1.0	41.2	17.2	0.0	12.1	0.0	0.0	0.0	0.0	0.0	11.7	0.6	0.0	0.0
	pDMW368	RD5	8.1	7.3	0.9	42.3	16.1	0.0	12.1	0.0	0.0	0.0	0.0	0.0	10.1	3.1	0.0	0.0
ETA	pY98	MaD5	11.9	6.5	0.8	38.5	12.2	0.0	0.0	0.4	0.0	0.0	0.0	0.0	0.0	0.0	26.1	3.6
	pDMW368	RD5	11.9	6.4	0.8	40.0	11.1	0.0	0.0	0.4	0.0	0.0	0.0	0.0	0.0	0.0	18.8	10.5

B

Fatty Acid	Clone	Gene	% delta-5		Ratio Desat		Ratio Prod/By-Prod		Ratio n-6/n-3
			desat	R/Ma	Desat	R/Ma	Prod/By-Prod	R/Ma	
None	pY98	MaD5							
	pDMW368	RD5							
EDA	pY98	MaD5	3.4		9.7				0.68
	pDMW368	RD5	32.6						1.40
DGLA	pY98	MaD5	14.0		3.6		4.17	0.37	1.17
	pDMW368	RD5	50.3				1.54		1.40
ETra	pY98	MaD5	4.9		4.7				
	pDMW368	RD5	23.2						
ETA	pY98	MaD5	12.0		3.0		2.44	0.63	
	pDMW368	RD5	35.9				1.55		