

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
23 September 2010 (23.09.2010)

PCT

(10) International Publication Number
WO 2010/108114 A2

(51) International Patent Classification:

C12N 15/52 (2006.01) *C12N 9/00* (2006.01)
C12N 15/63 (2006.01) *A61K 38/43* (2006.01)
C12N 5/10 (2006.01) *A61K 48/00* (2006.01)
C12N 1/00 (2006.01) *A61P 3/00* (2006.01)
C12P 7/64 (2006.01)

(21) International Application Number:

PCT/US2010/028009

(22) International Filing Date:

19 March 2010 (19.03.2010)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

61/161,742 19 March 2009 (19.03.2009) US
61/296,460 19 January 2010 (19.01.2010) US

(71) Applicant (*for all designated States except US*):
MARTEK BIOSCIENCES CORPORATION
[US/US]; 6480 Dobbin Road, Columbia, Maryland 21045
(US).

(72) Inventors; and

(71) Applicants : **APT, Kirk, E.** [US/US]; 3874 Paul Mill
Road, Ellicott City, Maryland 21042 (US). **RICHTER,**
Leslie [US/US]; 1085 West Midway Blvd., Broomfield,
Colorado 80020 (US). **SIMPSON, David** [US/US]; 2145
Mapleton Ave., Boulder, Colorado 80304 (US).
ZIRKLE, Ross [US/US]; 2108 Huntington Terrace, Mt.
Airy, Maryland 21771 (US).

(74) Agents: **KIM, Judith, U.** et al.; Sterne, Kessler, Gold-
stein & Fox, P.L.L.C., 1100 New York Avenue, N.W.,
Washington, District of Columbia 20005-3934 (US).

(81) Designated States (*unless otherwise indicated, for every
kind of national protection available*): AE, AG, AL, AM,
AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ,
CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO,
DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT,
HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP,
KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD,
ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI,
NO, NZ, OM, PE, PG, PH, PL, PT, RO, RS, RU, SC, SD,
SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR,
TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States (*unless otherwise indicated, for every
kind of regional protection available*): ARIPO (BW, GH,
GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM,
ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ,
TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE,
ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV,
MC, MK, MT, NL, NO, PL, PT, RO, SE, SI, SK, SM,
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 (Rule 48.2(g))*
- *with sequence listing part of description (Rule 5.2(a))*

(54) Title: POLYUNSATURATED FATTY ACID SYNTHASE NUCLEIC ACID MOLECULES AND POLYPEPTIDES,
COMPOSITIONS, AND METHODS OF MAKING AND USES THEREOF

(57) Abstract: The present invention is directed to isolated nucleic acid molecules and polypeptides of thraustochytrid polyunsaturated fatty acid (PUFA) synthases involved in the production of PUFAs, including PUFAs enriched in docosahexaenoic acid (DHA), eicosapentaenoic acid (EPA), or a combination thereof. The present invention is directed to vectors and host cells comprising the nucleic acid molecules, polypeptides encoded by the nucleic acid molecules, compositions comprising the nucleic acid molecules or polypeptides, and methods of making and uses thereof.



WO 2010/108114 A2

POLYUNSATURATED FATTY ACID SYNTHASE NUCLEIC ACID MOLECULES AND POLYPEPTIDES, COMPOSITIONS, AND METHODS OF MAKING AND USES THEREOF

BACKGROUND OF THE INVENTION

Field of the Invention

[0001] The present invention is directed to isolated nucleic acid molecules and polypeptides of polyunsaturated fatty acid (PUFA) synthases involved in the production of PUFAs, including PUFAs enriched in docosahexaenoic acid (DHA), eicosapentaenoic acid (EPA), or a combination thereof. The present invention is directed to vectors and host cells comprising the nucleic acid molecules, polypeptides encoded by the nucleic acid molecules, compositions comprising the nucleic acid molecules or polypeptides, and methods of making and uses thereof.

Background of the Invention

[0002] Thraustochytrids are microorganisms of the order Thraustochytriales, including members of the genus *Thraustochytrium* and the genus *Schizochytrium*, and have been recognized as an alternative source of PUFAs. See, e.g., U.S. Patent No. 5,130,242. It has recently been shown that polyketide synthase (PKS)-like systems in marine bacteria and thraustochytrids are capable of synthesizing polyunsaturated fatty acids (PUFAs) from acetyl-CoA and malonyl-CoA. These PKS synthase-like systems are also referred to herein as PUFA synthase systems. PUFA synthase systems in the marine bacteria *Shewanella* and *Vibrio marinus* are described in U.S. Patent No. 6,140,486. A PUFA synthase system in a thraustochytrid of the genus *Schizochytrium* is described in U.S. Patent No. 6,566,583. PUFA synthase systems in thraustochytrids of the genus *Schizochytrium* and the genus *Thraustochytrium* are also described in U.S. Patent No. 7,247,461. U.S. Patent No. 7,211,418 describes a PUFA synthase system in a thraustochytrid of the genus *Thraustochytrium* and the production of eicosapentaenoic acid (C20:5, omega-3) (EPA) and other PUFAs using the system. U.S. Patent No. 7,217,856 describes PUFA synthase systems in *Shewanella olleyana* and *Shewanella japonica*. WO 2005/097982 describes a PUFA synthase system in strain SAM2179. U.S.

Patent Nos. 7,208,590 and 7,368,552 describe PUFA synthase genes and proteins from *Thraustochytrium aureum*.

[0003] PKS systems have been traditionally described in the literature as falling into one of three basic types, typically referred to as Type I (modular or iterative), Type II, and Type III. The Type I modular PKS system has also been referred to as a "modular" PKS system, and the Type I iterative PKS system has also been referred to as a "Type I" PKS system. The Type II system is characterized by separable proteins, each of which carries out a distinct enzymatic reaction. The enzymes work in concert to produce the end product and each individual enzyme of the system typically participates several times in the production of the end product. This type of system operates in a manner analogous to the fatty acid synthase (FAS) systems found in plants and bacteria. Type I iterative PKS systems are similar to the Type II system in that the enzymes are used in an iterative fashion to produce the end product. The Type I iterative system differs from the Type II system in that enzymatic activities, instead of being associated with separable proteins, occur as domains of larger proteins. This system is analogous to the Type I FAS systems found in animals and fungi.

[0004] In contrast to the Type II systems, each enzyme domain in the Type I modular PKS systems is used only once in the production of the end product. The domains are found in very large proteins and the product of each reaction is passed on to another domain in the PKS protein.

[0005] Type III systems have been more recently discovered and belong to the plant chalcone synthase family of condensing enzymes. Type III PKSs are distinct from Type I and Type II PKS systems and utilize free CoA substrates in iterative condensation reactions to usually produce a heterocyclic end product.

[0006] In the conventional or standard pathway for PUFA synthesis, medium chain-length saturated fatty acids (products of a fatty acid synthase (FAS) system) are modified by a series of elongation and desaturation reactions. The substrates for the elongation reaction are fatty acyl-CoA (the fatty acid chain to be elongated) and malonyl-CoA (the source of the two carbons added during each elongation reaction). The product of the elongase reaction is a fatty acyl-CoA that has two additional carbons in the linear chain. The desaturases create cis double bonds in the preexisting fatty acid chain by extraction of two hydrogens in an oxygen-dependant reaction. The substrates for the desaturases are

either acyl-CoA (in some animals) or the fatty acid that is esterified to the glycerol backbone of a phospholipid (e.g., phosphatidylcholine).

[0007] Fatty acids are classified based on the length and saturation characteristics of the carbon chain. Fatty acids are termed short chain, medium chain, or long chain fatty acids based on the number of carbons present in the chain, are termed saturated fatty acids when no double bonds are present between the carbon atoms, and are termed unsaturated fatty acids when double bonds are present. Unsaturated long chain fatty acids are monounsaturated when only one double bond is present and are polyunsaturated when more than one double bond is present.

[0008] PUFAs are classified based on the position of the first double bond from the methyl end of the fatty acid: omega-3 (n-3) fatty acids contain a first double bond at the third carbon, while omega-6 (n-6) fatty acids contain a first double bond at the sixth carbon. For example, docosahexaenoic acid ("DHA") is an omega-3 PUFA with a chain length of 22 carbons and 6 double bonds, often designated as "22:6 n-3." Other omega-3 PUFAs include eicosapentaenoic acid ("EPA"), designated as "20:5 n-3," and omega-3 docosapentaenoic acid ("DPA n-3"), designated as "22:5 n-3." DHA and EPA have been termed "essential" fatty acids. Omega-6 PUFAs include arachidonic acid ("ARA"), designated as "20:4 n-6," and omega-6 docosapentaenoic acid ("DPA n-6"), designated as "22:5 n-6."

[0009] Omega-3 fatty acids are biologically important molecules that affect cellular physiology due to their presence in cell membranes, regulate production and gene expression of biologically active compounds, and serve as biosynthetic substrates. Roche, H. M., *Proc. Nutr. Soc.* 58: 397-401 (1999). DHA, for example, accounts for approximately 15%-20% of lipids in the human cerebral cortex, and 30%-60% of lipids in the retina, is concentrated in the testes and sperm, and is an important component of breast milk. Bergé, J.P., and Barnathan, G. *Adv. Biochem. Eng. Biotechnol.* 96:49-125 (2005). DHA accounts for up to 97% of the omega-3 fatty acids in the brain and up to 93% of the omega-3 fatty acids in the retina. Moreover, DHA is essential for both fetal and infant development, as well as maintenance of cognitive functions in adults. *Id.* Because omega-3 fatty acids are not synthesized *de novo* in the human body, these fatty acids must be derived from nutritional sources.

- [0010] Flaxseed oil and fish oils are considered good dietary sources of omega-3 fatty acids. Flaxseed oil contains no EPA, DHA, DPA, or ARA but rather contains linolenic acid (C18:3 n-3), a building block enabling the body to manufacture EPA. There is evidence, however, that the rate of metabolic conversion can be slow and variable, particularly among those with impaired health. Fish oils vary considerably in the type and level of fatty acid composition depending on the particular species and their diets. For example, fish raised by aquaculture tend to have a lower level of omega-3 fatty acids than those in the wild. Furthermore, fish oils carry the risk of containing environmental contaminants and can be associated with stability problems and a fishy odor or taste.
- [0011] Oils produced from thraustochytrids often have simpler polyunsaturated fatty acid profiles than corresponding fish or microalgal oils. Lewis, T.E., *Mar. Biotechnol.* 1: 580-587 (1999). Strains of thraustochytrid species have been reported to produce omega-3 fatty acids as a high percentage of the total fatty acids produced by the organisms. U.S. Patent No. 5,130,242; Huang, J. *et al.*, *J. Am. Oil. Chem. Soc.* 78: 605-610 (2001); Huang, J. *et al.*, *Mar. Biotechnol.* 5: 450-457 (2003). However, isolated thraustochytrids vary in the identity and amounts of PUFAs produced, such that some previously described strains can have undesirable PUFA profiles.
- [0012] Efforts have been made to produce PUFAs in oil-seed crop plants by modification of the endogenously-produced fatty acids. Genetic modification of these plants with various individual genes for fatty acid elongases and desaturases has produced leaves or seeds containing measurable levels of PUFAs such as EPA, but also containing significant levels of mixed shorter-chain and less unsaturated PUFAs (Qi *et al.*, *Nature Biotech.* 22:739 (2004); PCT Publ. No. WO 04/071467; Abbadi *et al.*, *Plant Cell* 16:1 (2004)); Napier and Sayanova, *Proc. Nutrition Society* 64:387-393 (2005); Robert *et al.*, *Functional Plant Biology* 32:473-479 (2005); and U.S. Appl. Publ. No. 2004/0172682).
- [0013] As such, a continuing need exists for the isolation of nucleic acid molecules and polypeptides associated with desirable PUFA profiles and methods to produce desirable PUFA profiles through use of such nucleic acid molecules and polypeptides.

BRIEF SUMMARY OF THE INVENTION

- [0014] The present invention is directed to an isolated nucleic acid molecule selected from the group consisting of: (a) a nucleic acid molecule comprising a polynucleotide

sequence at least 80% identical to SEQ ID NO:1, wherein the polynucleotide sequence encodes a polypeptide comprising PUFA synthase activity selected from the group consisting of beta-ketoacyl-ACP synthase (KS) activity, malonyl-CoA:ACP acyltransferase (MAT) activity, acyl carrier protein (ACP) activity, ketoreductase (KR) activity, beta-hydroxyacyl-ACP dehydrase (DH) activity, and combinations thereof; (b) a nucleic acid molecule comprising a polynucleotide sequence at least 80% identical to SEQ ID NO:7, wherein the polynucleotide sequence encodes a polypeptide comprising KS activity; (c) a nucleic acid molecule comprising a polynucleotide sequence at least 80% identical to SEQ ID NO:9, wherein the polynucleotide sequence encodes a polypeptide comprising MAT activity; (d) a nucleic acid molecule comprising a polynucleotide sequence at least 80% identical to any one of SEQ ID NOs:13, 15, 17, 19, 21, or 23, wherein the polynucleotide sequence encodes a polypeptide comprising ACP activity; (e) a nucleic acid molecule comprising a polynucleotide sequence at least 80% identical to SEQ ID NO:11, wherein the polynucleotide sequence encodes a polypeptide comprising ACP activity; (f) a nucleic acid molecule comprising a polynucleotide sequence at least 80% identical to SEQ ID NO:25, wherein the polynucleotide sequence encodes a polypeptide comprising KR activity; and (g) a nucleic acid molecule comprising a polynucleotide sequence at least 80% identical to SEQ ID NO:27, wherein the polynucleotide sequence encodes a polypeptide comprising DH activity. In some embodiments, the polynucleotide sequences are at least 90% identical or at least 95% identical to SEQ ID NOs:1, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, and 27, respectively. In some embodiments, the nucleic acid molecules comprise the polynucleotide sequences of SEQ ID NOs:1, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, and 27, respectively.

[0015] The present invention is directed to an isolated nucleic acid molecule selected from the group consisting of: (a) a nucleic acid molecule comprising a polynucleotide sequence encoding a polypeptide, wherein the polypeptide comprises an amino acid sequence at least 80% identical to SEQ ID NO:2, and wherein the polypeptide comprises a PUFA synthase activity selected from the group consisting of KS activity, MAT activity, ACP activity, KR activity, DH activity, and combinations thereof; (b) a nucleic acid molecule comprising a polynucleotide sequence encoding a polypeptide, wherein the polypeptide comprises an amino acid sequence at least 80% identical to SEQ ID NO:8, and wherein the polypeptide comprises KS activity; (c) a nucleic acid molecule

comprising a polynucleotide sequence encoding a polypeptide, wherein the polypeptide comprises an amino acid sequence at least 80% identical to SEQ ID NO:10, and wherein the polypeptide comprises MAT activity; (d) a nucleic acid molecule comprising a polynucleotide sequence encoding a polypeptide, wherein the polypeptide comprises an amino acid sequence at least 80% identical to any one of SEQ ID NOs:14, 16, 18, 20, 22, or 24, and wherein the polypeptide comprises ACP activity; (e) a nucleic acid molecule comprising a polynucleotide sequence encoding a polypeptide, wherein the polypeptide comprises an amino acid sequence at least 80% identical to SEQ ID NO:12, and wherein the polypeptide comprises ACP activity; (f) a nucleic acid molecule comprising a polynucleotide sequence encoding a polypeptide, wherein the polypeptide comprises an amino acid sequence at least 80% identical to SEQ ID NO:26, and wherein the polypeptide comprises KR activity; and (g) a nucleic acid molecule comprising a polynucleotide sequence encoding a polypeptide, wherein the polypeptide comprises an amino acid sequence at least 80% identical to SEQ ID NO:28, and wherein the polypeptide comprises DH activity. In some embodiments, the amino acid sequences are at least 90% identical or at least 95% identical to SEQ ID NOs:2, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, and 28, respectively. In some embodiments, the polypeptides comprise the amino acid sequences of SEQ ID NOs:2, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, and 28, respectively.

[0016] The present invention is directed to an isolated nucleic acid molecule selected from the group consisting of: (a) a nucleic acid molecule comprising a polynucleotide sequence at least 80% identical to SEQ ID NO:3, wherein the polynucleotide sequence encodes a polypeptide comprising PUFA synthase activity selected from the group consisting of KS activity, chain length factor (CLF) activity, acyltransferase (AT) activity, enoyl-ACP reductase (ER) activity, and combinations thereof; (b) a nucleic acid molecule comprising a polynucleotide sequence at least 80% identical to SEQ ID NO:29, wherein the polynucleotide sequence encodes a polypeptide comprising KS activity; (c) a nucleic acid molecule comprising a polynucleotide sequence at least 80% identical to SEQ ID NO:31, wherein the polynucleotide sequence encodes a polypeptide comprising CLF activity; (d) a nucleic acid molecule comprising a polynucleotide sequence at least 80% identical to SEQ ID NO:33, wherein the polynucleotide sequence encodes a polypeptide comprising AT activity; and (e) a nucleic acid molecule comprising a

polynucleotide sequence at least 80% identical to SEQ ID NO:35, wherein the polynucleotide sequence encodes a polypeptide comprising ER activity. In some embodiments, the polynucleotide sequences are at least 90% identical or at least 95% identical to SEQ ID NOs:3, 29, 31, 33, and 35, respectively. In some embodiments, the nucleic acid molecules comprise the polynucleotide sequences of SEQ ID NOs:3, 29, 31, 33, and 35, respectively.

[0017] The present invention is directed to an isolated nucleic acid molecule selected from the group consisting of: (a) a nucleic acid molecule comprising a polynucleotide sequence encoding a polypeptide, wherein the polypeptide comprises an amino acid sequence at least 80% identical to SEQ ID NO:4, and wherein the polypeptide comprises a PUFA synthase activity selected from the group consisting of KS activity, CLF activity, AT activity, ER activity, and combinations thereof; (b) a nucleic acid molecule comprising a polynucleotide sequence encoding a polypeptide, wherein the polypeptide comprises an amino acid sequence at least 80% identical to SEQ ID NO:30, and wherein the polypeptide comprises KS activity; (c) a nucleic acid molecule comprising a polynucleotide sequence encoding a polypeptide, wherein the polypeptide comprises an amino acid sequence at least 80% identical to SEQ ID NO:32, and wherein the polypeptide comprises CLF activity; (d) a nucleic acid molecule comprising a polynucleotide sequence encoding a polypeptide, wherein the polypeptide comprises an amino acid sequence at least 80% identical to SEQ ID NO:34, and wherein the polypeptide comprises AT activity; and (e) a nucleic acid molecule comprising a polynucleotide sequence encoding a polypeptide, wherein the polypeptide comprises an amino acid sequence at least 80% identical to SEQ ID NO:36, and wherein the polypeptide comprises ER activity. In some embodiments, the amino acid sequences are at least 90% identical or at least 95% identical to SEQ ID NOs:4, 30, 32, 34, and 36, respectively. In some embodiments, the polypeptides comprise the amino acid sequence of SEQ ID NOs:4, 30, 32, 34, and 36, respectively.

[0018] The present invention is directed to an isolated nucleic acid molecule selected from the group consisting of: (a) a nucleic acid molecule comprising a polynucleotide sequence at least 80% identical to SEQ ID NO:5, wherein the polynucleotide sequence encodes a polypeptide comprising PUFA synthase activity selected from the group consisting of DH activity, ER activity, and combinations thereof; (b) a nucleic acid

molecule comprising a polynucleotide sequence at least 80% identical to SEQ ID NO:37, wherein the polynucleotide sequence encodes a polypeptide comprising DH activity; (c) a nucleic acid molecule comprising a polynucleotide sequence at least 80% identical to SEQ ID NO:39, wherein the polynucleotide sequence encodes a polypeptide comprising DH activity; and (d) a nucleic acid molecule comprising a polynucleotide sequence at least 80% identical to SEQ ID NO:41, wherein the polynucleotide sequence encodes a polypeptide comprising ER activity. In some embodiments, the polynucleotide sequences are at least 90% identical or at least 95% identical to SEQ ID NOs:5, 37, 39, and 41, respectively. In some embodiments, the nucleic acid molecules comprise the polynucleotide sequences of SEQ ID NOs:5, 37, 39, and 41, respectively.

[0019] The present invention is directed to an isolated nucleic acid molecule selected from the group consisting of: (a) a nucleic acid molecule comprising a polynucleotide sequence encoding a polypeptide, wherein the polypeptide comprises an amino acid sequence at least 80% identical to SEQ ID NO:6, wherein the polypeptide comprises PUFA synthase activity selected from the group consisting of DH activity, ER activity, and combinations thereof; (b) a nucleic acid molecule comprising a polynucleotide sequence encoding a polypeptide, wherein the polypeptide comprises an amino acid sequence at least 80% identical to SEQ ID NO:38, and wherein the polypeptide comprises DH activity; (c) a nucleic acid molecule comprising a polynucleotide sequence encoding a polypeptide, wherein the polypeptide comprises an amino acid sequence at least 80% identical to SEQ ID NO:40, and wherein the polypeptide comprises DH activity; and (d) a nucleic acid molecule comprising a polynucleotide sequence encoding a polypeptide, wherein the polypeptide comprises an amino acid sequence at least 80% identical to SEQ ID NO:42, and wherein the polypeptide comprises ER activity. In some embodiments, the amino acid sequences are at least 90% identical or at least 95% identical to SEQ ID NOs:6, 38, 40, and 42, respectively. In some embodiments, the polypeptides comprise the amino acid sequences of SEQ ID NOs:6, 38, 40, and 42, respectively.

[0020] The present invention is directed to an isolated nucleic acid molecule selected from the group consisting of: (a) a nucleic acid molecule comprising a polynucleotide sequence at least 80% identical to SEQ ID NO:68 or SEQ ID NO:120, wherein the polynucleotide sequence encodes a polypeptide comprising PUFA synthase activity selected from the group consisting of KS activity, MAT activity, ACP activity, KR

activity, DH activity, and combinations thereof; (b) a nucleic acid molecule comprising a polynucleotide sequence at least 80% identical to SEQ ID NO:74, wherein the polynucleotide sequence encodes a polypeptide comprising KS activity; (c) a nucleic acid molecule comprising a polynucleotide sequence at least 80% identical to SEQ ID NO:76, wherein the polynucleotide sequence encodes a polypeptide comprising MAT activity; (d) a nucleic acid molecule comprising a polynucleotide sequence at least 80% identical to any one of SEQ ID NOs: 80, 82, 84, 86, 88, 90, 92, 94, 96, or 98, wherein the polynucleotide sequence encodes a polypeptide comprising ACP activity; (e) a nucleic acid molecule comprising a polynucleotide sequence at least 80% identical to SEQ ID NO:78, wherein the polynucleotide sequence encodes a polypeptide comprising ACP activity; (f) a nucleic acid molecule comprising a polynucleotide sequence at least 80% identical to SEQ ID NO:100, wherein the polynucleotide sequence encodes a polypeptide comprising KR activity; and (g) a nucleic acid molecule comprising a polynucleotide sequence at least 80% identical to SEQ ID NO:118, wherein the polynucleotide sequence encodes a polypeptide comprising DH activity. In some embodiments, the polynucleotide sequences are at least 90% identical or at least 95% identical to SEQ ID NOs:68, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96, 98, 100, 118, and 120, respectively. In some embodiments, the nucleic acid molecules comprise the polynucleotide sequences of SEQ ID NOs:68, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96, 98, 100, 118, and 120, respectively.

[0021] The present invention is directed to an isolated nucleic acid molecule selected from the group consisting of: (a) a nucleic acid molecule comprising a polynucleotide sequence encoding a polypeptide, wherein the polypeptide comprises an amino acid sequence at least 80% identical to SEQ ID NO:69, and wherein the polypeptide comprises a PUFA synthase activity selected from the group consisting of KS activity, MAT activity, ACP activity, KR activity, DH activity, and combinations thereof; (b) a nucleic acid molecule comprising a polynucleotide sequence encoding a polypeptide, wherein the polypeptide comprises an amino acid sequence at least 80% identical to SEQ ID NO:75, and wherein the polypeptide comprises KS activity; (c) a nucleic acid molecule comprising a polynucleotide sequence encoding a polypeptide, wherein the polypeptide comprises an amino acid sequence at least 80% identical to SEQ ID NO:77, and wherein the polypeptide comprises MAT activity; (d) a nucleic acid molecule comprising a

polynucleotide sequence encoding a polypeptide, wherein the polypeptide comprises an amino acid sequence at least 80% identical to any one of SEQ ID NOs:81, 83, 85, 87, 89, 91, 93, 95, 97, or 99, and wherein the polypeptide comprises ACP activity; (e) a nucleic acid molecule comprising a polynucleotide sequence encoding a polypeptide, wherein the polypeptide comprises an amino acid sequence at least 80% identical to SEQ ID NO:79, and wherein the polypeptide comprises ACP activity; (f) a nucleic acid molecule comprising a polynucleotide sequence encoding a polypeptide, wherein the polypeptide comprises an amino acid sequence at least 80% identical to SEQ ID NO:101, and wherein the polypeptide comprises KR activity; and (g) a nucleic acid molecule comprising a polynucleotide sequence encoding a polypeptide, wherein the polypeptide comprises an amino acid sequence at least 80% identical to SEQ ID NO:119, and wherein the polypeptide comprises DH activity. In some embodiments, the amino acid sequences are at least 90% identical or at least 95% identical to SEQ ID NOs:69, 75, 77, 79, 81, 83, 85, 87, 89, 91, 93, 95, 97, 99, 101, and 119, respectively. In some embodiments, the polypeptides comprise the amino acid sequences of SEQ ID NOs:69, 75, 77, 79, 81, 83, 85, 87, 89, 91, 93, 95, 97, 99, 101, and 119, respectively.

[0022] The present invention is directed to an isolated nucleic acid molecule selected from the group consisting of: (a) a nucleic acid molecule comprising a polynucleotide sequence at least 80% identical to SEQ ID NO:70 or SEQ ID NO:121, wherein the polynucleotide sequence encodes a polypeptide comprising PUFA synthase activity selected from the group consisting of KS activity, chain length factor (CLF) activity, acyltransferase (AT) activity, enoyl-ACP reductase (ER) activity, and combinations thereof; (b) a nucleic acid molecule comprising a polynucleotide sequence at least 80% identical to SEQ ID NO:102, wherein the polynucleotide sequence encodes a polypeptide comprising KS activity; (c) a nucleic acid molecule comprising a polynucleotide sequence at least 80% identical to SEQ ID NO:104, wherein the polynucleotide sequence encodes a polypeptide comprising CLF activity; (d) a nucleic acid molecule comprising a polynucleotide sequence at least 80% identical to SEQ ID NO:106, wherein the polynucleotide sequence encodes a polypeptide comprising AT activity; and (e) a nucleic acid molecule comprising a polynucleotide sequence at least 80% identical to SEQ ID NO:108, wherein the polynucleotide sequence encodes a polypeptide comprising ER activity. In some embodiments, the polynucleotide sequences are at least 90% identical

or at least 95% identical to SEQ ID NOs:70, 102, 104, 106, 108, and 121, respectively. In some embodiments, the nucleic acid molecules comprise the polynucleotide sequences of SEQ ID NOs:70, 102, 104, 106, 108, and 121, respectively.

[0023] The present invention is directed to an isolated nucleic acid molecule selected from the group consisting of: (a) a nucleic acid molecule comprising a polynucleotide sequence encoding a polypeptide, wherein the polypeptide comprises an amino acid sequence at least 80% identical to SEQ ID NO:71, and wherein the polypeptide comprises a PUFA synthase activity selected from the group consisting of KS activity, CLF activity, AT activity, ER activity, and combinations thereof; (b) a nucleic acid molecule comprising a polynucleotide sequence encoding a polypeptide, wherein the polypeptide comprises an amino acid sequence at least 80% identical to SEQ ID NO:103, and wherein the polypeptide comprises KS activity; (c) a nucleic acid molecule comprising a polynucleotide sequence encoding a polypeptide, wherein the polypeptide comprises an amino acid sequence at least 80% identical to SEQ ID NO:105, and wherein the polypeptide comprises CLF activity; (d) a nucleic acid molecule comprising a polynucleotide sequence encoding a polypeptide, wherein the polypeptide comprises an amino acid sequence at least 80% identical to SEQ ID NO:107, and wherein the polypeptide comprises AT activity; and (e) a nucleic acid molecule comprising a polynucleotide sequence encoding a polypeptide, wherein the polypeptide comprises an amino acid sequence at least 80% identical to SEQ ID NO:109, and wherein the polypeptide comprises ER activity. In some embodiments, the amino acid sequences are at least 90% identical or at least 95% identical to SEQ ID NOs:71, 103, 105, 107, and 109, respectively. In some embodiments, the polypeptides comprise the amino acid sequence of SEQ ID NOs:71, 103, 105, 107, and 109, respectively.

[0024] The present invention is directed to an isolated nucleic acid molecule selected from the group consisting of: (a) a nucleic acid molecule comprising a polynucleotide sequence at least 80% identical to SEQ ID NO:72 or SEQ ID NO:122, wherein the polynucleotide sequence encodes a polypeptide comprising PUFA synthase activity selected from the group consisting of DH activity, ER activity, and combinations thereof; (b) a nucleic acid molecule comprising a polynucleotide sequence at least 80% identical to SEQ ID NO:110, wherein the polynucleotide sequence encodes a polypeptide comprising DH activity; (c) a nucleic acid molecule comprising a polynucleotide

sequence at least 80% identical to SEQ ID NO:112, wherein the polynucleotide sequence encodes a polypeptide comprising DH activity; and (d) a nucleic acid molecule comprising a polynucleotide sequence at least 80% identical to SEQ ID NO:114, wherein the polynucleotide sequence encodes a polypeptide comprising ER activity. In some embodiments, the polynucleotide sequences are at least 90% identical or at least 95% identical to SEQ ID NOs:72, 110, 112, 114, and 122, respectively. In some embodiments, the nucleic acid molecules comprise the polynucleotide sequences of SEQ ID NOs:72, 110, 112, 114, and 122, respectively.

[0025] The present invention is directed to an isolated nucleic acid molecule selected from the group consisting of: (a) a nucleic acid molecule comprising a polynucleotide sequence encoding a polypeptide, wherein the polypeptide comprises an amino acid sequence at least 80% identical to SEQ ID NO:73, wherein the polypeptide comprises PUFA synthase activity selected from the group consisting of DH activity, ER activity, and combinations thereof; (b) a nucleic acid molecule comprising a polynucleotide sequence encoding a polypeptide, wherein the polypeptide comprises an amino acid sequence at least 80% identical to SEQ ID NO:111, and wherein the polypeptide comprises DH activity; (c) a nucleic acid molecule comprising a polynucleotide sequence encoding a polypeptide, wherein the polypeptide comprises an amino acid sequence at least 80% identical to SEQ ID NO:113, and wherein the polypeptide comprises DH activity; and (d) a nucleic acid molecule comprising a polynucleotide sequence encoding a polypeptide, wherein the polypeptide comprises an amino acid sequence at least 80% identical to SEQ ID NO:115, and wherein the polypeptide comprises ER activity. In some embodiments, the amino acid sequences are at least 90% identical or at least 95% identical to SEQ ID NOs:73, 111, 113, and 115, respectively. In some embodiments, the polypeptides comprise the amino acid sequences of SEQ ID NOs:73, 111, 113, and 115, respectively.

[0026] The present invention is directed to an isolated nucleic acid molecule comprising a polynucleotide sequence encoding a polypeptide comprising PUFA synthase activity selected from the group consisting of KS activity, MAT activity, ACP activity, KR activity, CLF activity, AT activity, ER activity, DH activity, and combinations thereof, wherein the polynucleotide hybridizes under stringent conditions to the complement of any of the polynucleotide sequences described above.

- [0027] The present invention is directed to an isolated nucleic acid molecule comprising a polynucleotide sequence that is fully complementary to any of the polynucleotide sequences described above.
- [0028] The present invention is directed to a recombinant nucleic acid molecule comprising any of the nucleic acid molecules described above or combinations thereof and a transcription control sequence. In some embodiments, the recombinant nucleic acid molecule is a recombinant vector.
- [0029] The present invention is directed to a host cell that expresses any of the nucleic acid molecules described above, any of the recombinant nucleic acid molecules described above, and combinations thereof. In some embodiments, the host cell is selected from the group consisting of a plant cell, a microbial cell, and an animal cell. In some embodiments, the microbial cell is a bacterium. In some embodiments, the bacterium is *E. coli*. In some embodiments, the bacterium is a marine bacterium. In some embodiments, the microbial cell is a thraustochytrid. In some embodiments, the thraustochytrid is a *Schizochytrium*. In some embodiments, the thraustochytrid is a *Thraustochytrium*. In some embodiments, the thraustochytrid is an *Ulkenia*.
- [0030] The present invention is directed to a method to produce at least one PUFA, comprising: expressing a PUFA synthase gene in a host cell under conditions effective to produce PUFA, wherein the PUFA synthase gene comprises any of the isolated nucleic acid molecules described above, any of the recombinant nucleic acid molecules described above, or combinations thereof, and wherein at least one PUFA is produced. In one aspect of this embodiment, the host cell is selected from the group consisting of a plant cell, an isolated animal cell, and a microbial cell. In another aspect of this embodiment, the at least one PUFA comprises docosahexaenoic acid (DHA) or eicosapentaenoic acid (EPA).
- [0031] The present invention is directed to a method to produce lipids enriched for DHA, EPA, or a combination thereof, comprising: expressing a PUFA synthase gene in a host cell under conditions effective to produce lipids, wherein the PUFA synthase gene comprises any of the isolated nucleic acid molecules described above, any of the recombinant nucleic acid molecules described above, or combinations thereof in the host cell, and wherein lipids enriched with DHA, EPA, or a combination thereof are produced. The present invention is directed to a method for making a recombinant vector

comprising inserting any one of the isolated nucleic acid molecules described above into a vector.

[0032] The present invention is directed to a method of making a recombinant host cell comprising introducing a recombinant vector as described above into a host cell. In some embodiments, the host cell is selected from the group consisting of a plant cell, an isolated animal cell, and a microbial cell.

[0033] The present invention is directed to an isolated polypeptide encoded by any of the polynucleotide sequences described above.

[0034] The present invention is directed to an isolated polypeptide selected from the group consisting of: (a) a polypeptide comprising an amino acid sequence at least 80% identical to SEQ ID NO:2, wherein the polypeptide comprises a PUFA synthase activity selected from the group consisting of KS activity, MAT activity, ACP activity, KR activity, DH activity, and combinations thereof; (b) a polypeptide comprising an amino acid sequence at least 80% identical to SEQ ID NO:8, wherein the polypeptide comprises KS activity; (c) a polypeptide comprising an amino acid sequence at least 80% identical to SEQ ID NO:10, wherein the polypeptide comprises MAT activity; (d) a polypeptide comprising an amino acid sequence at least 80% identical to any one of SEQ ID NOs:14, 16, 18, 20, 22, or 24, wherein the polypeptide comprises ACP activity; (e) a polypeptide comprising an amino acid sequence at least 80% identical to SEQ ID NO:12, wherein the polypeptide comprises ACP activity; (f) a polypeptide comprising an amino acid sequence at least 80% identical to SEQ ID NO:26, wherein the polypeptide comprises KR activity; and (g) a polypeptide comprising an amino acid sequence at least 80% identical to SEQ ID NO:28, wherein the polypeptide comprises DH activity. In some embodiments, the amino acid sequences are at least 90% identical or at least 95% identical to SEQ ID NOs:2, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, and 28, respectively. In some embodiments, the polypeptides comprise the amino acid sequences of SEQ ID NOs:2, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, and 28, respectively.

[0035] The present invention is directed to an isolated polypeptide selected from the group consisting of: (a) a polypeptide comprising an amino acid sequence at least 80% identical to SEQ ID NO:4, wherein the polypeptide comprises a PUFA synthase activity selected from the group consisting of KS activity, CLF activity, AT activity, ER activity, and combinations thereof; (b) a polypeptide comprising an amino acid sequence at least

80% identical to SEQ ID NO:30, wherein the polypeptide comprises KS activity; (c) a polypeptide comprising an amino acid sequence at least 80% identical to SEQ ID NO:32, wherein the polypeptide comprises CLF activity; (d) a polypeptide comprising an amino acid sequence at least 80% identical to SEQ ID NO:34, wherein the polypeptide comprises AT activity; and (e) a polypeptide comprising an amino acid sequence at least 80% identical to SEQ ID NO:36, wherein the polypeptide comprises ER activity. In some embodiments, the amino acid sequences are at least 90% identical or at least 95% identical to SEQ ID NOs:4, 30, 32, 34, and 36, respectively. In some embodiments, the polypeptides comprise the amino acid sequence of SEQ ID NOs:4, 30, 32, 34, and 36, respectively.

[0036] The present invention is directed to an isolated polypeptide selected from the group consisting of: (a) a polypeptide comprising an amino acid sequence at least 80% identical to SEQ ID NO:6, wherein the polypeptide comprises a PUFA synthase activity selected from the group consisting of DH activity, ER activity, and combinations thereof; (b) a polypeptide comprising an amino acid sequence at least 80% identical to SEQ ID NO:38, wherein the polypeptide comprises DH activity; (c) a polypeptide comprising an amino acid sequence at least 80% identical to SEQ ID NO:40, wherein the polypeptide comprises DH activity; and (d) a polypeptide comprising an amino acid sequence at least 80% identical to SEQ ID NO:42, wherein the polypeptide comprises ER activity. In some embodiments, the amino acid sequences are at least 90% identical or at least 95% identical to SEQ ID NOs:6, 38, 40, and 42, respectively. In some embodiments, the polypeptides comprise the amino acid sequences of SEQ ID NOs:6, 38, 40, and 42, respectively.

[0037] The present invention is directed to an isolated polypeptide selected from the group consisting of: (a) a polypeptide comprising an amino acid sequence at least 80% identical to SEQ ID NO:69, wherein the polypeptide comprises a PUFA synthase activity selected from the group consisting of KS activity, MAT activity, ACP activity, KR activity, DH activity, and combinations thereof; (b) a polypeptide comprising an amino acid sequence at least 80% identical to SEQ ID NO:75, wherein the polypeptide comprises KS activity; (c) a polypeptide comprising an amino acid sequence at least 80% identical to SEQ ID NO:77, wherein the polypeptide comprises MAT activity; (d) a polypeptide comprising an amino acid sequence at least 80% identical to any one of SEQ

ID NOs:81, 83, 85, 87, 89, 91, 93, 95, 97, or 99, wherein the polypeptide comprises ACP activity; (e) a polypeptide comprising an amino acid sequence at least 80% identical to SEQ ID NO:79, wherein the polypeptide comprises ACP activity; (f) a polypeptide comprising an amino acid sequence at least 80% identical to SEQ ID NO:101, wherein the polypeptide comprises KR activity; and (g) a polypeptide comprising an amino acid sequence at least 80% identical to SEQ ID NO:119, wherein the polypeptide comprises DH activity. In some embodiments, the amino acid sequences are at least 90% identical or at least 95% identical to SEQ ID NOs:69, 75, 77, 79, 81, 83, 85, 87, 89, 91, 93, 95, 97, 99, 101, and 119, respectively. In some embodiments, the polypeptides comprise the amino acid sequences of SEQ ID NOs:69, 75, 77, 79, 81, 83, 85, 87, 89, 91, 93, 95, 97, 99, 101, and 119, respectively.

[0038] The present invention is directed to an isolated polypeptide selected from the group consisting of: (a) a polypeptide comprising an amino acid sequence at least 80% identical to SEQ ID NO:71, wherein the polypeptide comprises a PUFA synthase activity selected from the group consisting of KS activity, CLF activity, AT activity, ER activity, and combinations thereof; (b) a polypeptide comprising an amino acid sequence at least 80% identical to SEQ ID NO:103, wherein the polypeptide comprises KS activity; (c) a polypeptide comprising an amino acid sequence at least 80% identical to SEQ ID NO:105, wherein the polypeptide comprises CLF activity; (d) a polypeptide comprising an amino acid sequence at least 80% identical to SEQ ID NO:107, wherein the polypeptide comprises AT activity; and (e) a polypeptide comprising an amino acid sequence at least 80% identical to SEQ ID NO:109, wherein the polypeptide comprises ER activity. In some embodiments, the amino acid sequences are at least 90% identical or at least 95% identical to SEQ ID NOs:71, 103, 105, 107, and 109, respectively. In some embodiments, the polypeptides comprise the amino acid sequence of SEQ ID NOs:71, 103, 105, 107, and 109, respectively.

[0039] The present invention is directed to an isolated polypeptide selected from the group consisting of: (a) a polypeptide comprising an amino acid sequence at least 80% identical to SEQ ID NO:73, wherein the polypeptide comprises a PUFA synthase activity selected from the group consisting of DH activity, ER activity, and combinations thereof; (b) a polypeptide comprising an amino acid sequence at least 80% identical to SEQ ID NO:111, wherein the polypeptide comprises DH activity; (c) a polypeptide comprising an

amino acid sequence at least 80% identical to SEQ ID NO:113, wherein the polypeptide comprises DH activity; and (d) a polypeptide comprising an amino acid sequence at least 80% identical to SEQ ID NO:115, wherein the polypeptide comprises ER activity. In some embodiments, the amino acid sequences are at least 90% identical or at least 95% identical to SEQ ID NOs:73, 111, 113, and 115, respectively. In some embodiments, the polypeptides comprise the amino acid sequences of SEQ ID NOs:73, 111, 113, and 115, respectively.

[0040] In some embodiments, any of the isolated polypeptides of the invention can be a fusion polypeptide.

[0041] The present invention is directed to a composition comprising any of the polypeptides described above and a biologically acceptable carrier.

[0042] The present invention is directed to a method of increasing production of DHA, EPA, or a combination thereof in an organism having PUFA synthase activity, comprising: expressing any of the isolated nucleic acid molecules described above, any of the recombinant nucleic acid molecules described above, or combinations thereof in the organism under conditions effective to produce DHA, EPA, or a combination thereof, wherein the PUFA synthase activity replaces an inactive or deleted activity, introduces a new activity, or enhances an existing activity in the organism, and wherein production of DHA, EPA, or a combination thereof in the organism is increased.

[0043] The present invention is directed to a method of isolating lipids from a host cell, comprising: (a) expressing a PUFA synthase gene in the host cell under conditions effective to produce lipids, wherein the PUFA synthase gene comprises any of the isolated nucleic acid molecules described above, any of the recombinant nucleic acid molecules described above, or combinations thereof in the host cell, and (b) isolating lipids from the host cell. In some embodiments, the host cell is selected from the group consisting of a plant cell, an isolated animal cell, and a microbial cell. In some embodiments, the lipids comprise DHA, EPA, or a combination thereof.

BRIEF DESCRIPTION OF THE DRAWINGS/FIGURES

[0044] FIG. 1 shows the gene architecture of the *Schizochytrium* sp. ATCC PTA-9695 PUFA synthases of the invention.

- [0045] FIG. 2 shows the gene architecture of the *Thraustochytrium* sp. ATCC PTA-10212 PUFA synthases of the invention.
- [0046] FIG. 3 shows the domain architecture of the *Schizochytrium* sp. ATCC PTA-9695 and *Thraustochytrium* sp. ATCC PTA-10212 PUFA synthases of the invention and synthases from *Schizochytrium* sp. ATCC 20888, *Thraustochytrium* sp. ATCC 20892, *Thraustochytrium aureum*, and SAM2179.
- [0047] FIG. 4 shows an alignment of a *Schizochytrium* sp. ATCC PTA-9695 Pfa1p amino acid sequence (SEQ ID NO:2) and a *Thraustochytrium* sp. ATCC PTA-10212 Pfa1p amino acid sequence (SEQ ID NO:69) of the invention with the OrfA sequences from *Schizochytrium* sp. ATCC 20888 (SEQ ID NO:54) and *Thraustochytrium* sp. ATCC 20892 (SEQ ID NO:56) and the ORF A sequence from *Thraustochytrium aureum* (SEQ ID NO:55).
- [0048] FIG. 5 shows an alignment of a *Schizochytrium* sp. ATCC PTA-9695 Pfa2p amino acid sequence (SEQ ID NO:4) and a *Thraustochytrium* sp. ATCC PTA-10212 Pfa2p amino acid sequence (SEQ ID NO:71) of the invention with the OrfB sequences from *Schizochytrium* sp. ATCC 20888 (SEQ ID NO:57) and *Thraustochytrium* sp. ATCC 20892 (SEQ ID NO:58) and the ORF B sequence from *Thraustochytrium aureum* (SEQ ID NO:59).
- [0049] FIG. 6 shows an alignment of a *Schizochytrium* sp. ATCC PTA-9695 Pfa3p amino acid sequence (SEQ ID NO:6) and a *Thraustochytrium* sp. ATCC PTA-10212 Pfa3p amino acid sequence (SEQ ID NO:73) of the invention with the OrfC sequences from *Schizochytrium* sp. ATCC 20888 (SEQ ID NO:61) and *Thraustochytrium* sp. ATCC 20892 (SEQ ID NO:60).
- [0050] FIG. 7 shows the *Schizochytrium* sp. ATCC PTA-9695 *PFA1* polynucleotide sequence (SEQ ID NO:1).
- [0051] FIG. 8 shows the *Schizochytrium* sp. ATCC PTA-9695 Pfa1p amino acid sequence (SEQ ID NO:2).
- [0052] FIG. 9 shows the *Schizochytrium* sp. ATCC PTA-9695 *PFA2* polynucleotide sequence (SEQ ID NO:3).
- [0053] FIG. 10 shows the *Schizochytrium* sp. ATCC PTA-9695 Pfa2p amino acid sequence (SEQ ID NO:4).

- [0054] FIG. 11 shows the *Schizochytrium* sp. ATCC PTA-9695 *PFA3* polynucleotide sequence (SEQ ID NO:5).
- [0055] FIG. 12 shows the *Schizochytrium* sp. ATCC PTA-9695 *Pfa3p* amino acid sequence (SEQ ID NO:6).
- [0056] FIG. 13 shows the *Thraustochytrium* sp. ATCC PTA-10212 *PFA1* polynucleotide sequence (SEQ ID NO:68).
- [0057] FIG. 14 shows a *Thraustochytrium* sp. ATCC PTA-10212 *PFA1* polynucleotide sequence (SEQ ID NO:120) that has been codon-optimized for expression in *Schizochytrium*.
- [0058] FIG. 15 shows the *Thraustochytrium* sp. ATCC PTA-10212 *Pfa1p* amino acid sequence (SEQ ID NO:69).
- [0059] FIG. 16 shows the *Thraustochytrium* sp. ATCC PTA-10212 *PFA2* polynucleotide sequence (SEQ ID NO:70).
- [0060] FIG. 17 shows a *Thraustochytrium* sp. ATCC PTA-10212 *PFA2* polynucleotide sequence (SEQ ID NO:121) that has been codon-optimized for expression in *Schizochytrium*.
- [0061] FIG. 18 shows the *Thraustochytrium* sp. ATCC PTA-10212 *Pfa2p* amino acid sequence (SEQ ID NO:71).
- [0062] FIG. 19 shows the *Thraustochytrium* sp. ATCC PTA-10212 *PFA3* polynucleotide sequence (SEQ ID NO:72).
- [0063] FIG. 20 shows a *Thraustochytrium* sp. ATCC PTA-10212 *PFA3* polynucleotide sequence (SEQ ID NO:122) that has been codon-optimized for expression in *Schizochytrium*.
- [0064] FIG. 21 shows the *Thraustochytrium* sp. ATCC PTA-10212 *Pfa3p* amino acid sequence (SEQ ID NO:73).
- [0065] FIG. 22 shows a codon usage table for *Schizochytrium*.

DETAILED DESCRIPTION OF THE INVENTION

- [0066] The present invention is directed to isolated nucleic acid molecules and polypeptides of polyunsaturated fatty acid (PUFA) synthases involved in the production of PUFAs, including PUFAs enriched in docosahexaenoic acid (DHA), eicosapentaenoic acid (EPA), or a combination thereof. The present invention is directed to vectors and

host cells comprising the nucleic acid molecules, polypeptides encoded by the nucleic acid molecules, compositions comprising the nucleic acid molecules or polypeptides, and methods of making and uses thereof.

PUFA Synthases

[0067] As used herein, the term "PUFA synthase" refers to an enzyme that is involved in the production of polyunsaturated fatty acids. *See, e.g., Metz et al., Science 293:290-293 (2001).*

[0068] The present invention is directed in part to three PUFA synthase subunits termed Pfa1p (SEQ ID NO:2 or SEQ ID NO:69), Pfa2p (SEQ ID NO:4 or SEQ ID NO:71), and Pfa3p (SEQ ID NO:6 or SEQ ID NO:73), as well as the genes that encode the subunits termed *PFA1* (SEQ ID NO:1, SEQ ID NO:68, or SEQ ID NO:120), *PFA2* (SEQ ID NO:3, SEQ ID NO:70, or SEQ ID NO:121), and *PFA3* (SEQ ID NO:5, SEQ ID NO:72, or SEQ ID NO:122). *See, FIGs. 1-3 and 7-21.* PUFA synthases in other thraustochytrids have also been designated as ORF 1, ORF 2, and ORF 3, respectively, or as OrfA, OrfB, and OrfC, respectively. *See, e.g., Schizochytrium sp. (ATCC 20888) and Thraustochytrium sp. (ATCC 20892) in U.S. Pat. Nos. 7,247,461 and 7,256,022, referring to orfA, orfB, and orfC genes and corresponding OrfA, OrfB, and OrfC proteins, and Thraustochytrium aureum (ATCC 34304) in U.S. Pat. No. 7,368,552, referring to ORF A, ORF B, and ORF C genes and proteins. See also, strain SAM2179 in WO/2005/097982, referring to ORF 1, ORF 2, and ORF 3 genes and proteins.*

Nucleic Acid Molecules

[0069] The present invention is directed to isolated nucleic acid molecules comprising polynucleotide sequences for PUFA synthase genes and domains derived from an isolated microorganism that is the subject of co-pending U.S. Appl. No. 12/407,687, filed on March 19, 2009, incorporated herein by reference in its entirety. The microorganism was deposited under the Budapest Treaty at the American Type Culture Collection, Patent Depository, 10801 University Boulevard, Manassas, VA 20110-2209, on January 7, 2009, and given ATCC Accession No. PTA-9695, and is also referred to as *Schizochytrium sp.* ATCC PTA-9695. When expressed, these genes produce unique fatty acid profiles, characterized in part by high levels of omega-3 fatty acids, in particular high levels of DHA.

[0070] The present invention is directed to isolated nucleic acid molecules comprising polynucleotide sequences for PUFA synthase genes and domains derived from an isolated microorganism that is the subject of co-pending U.S. Appl. No. 61/296,456, filed on January 19, 2010, incorporated herein by reference in its entirety. The microorganism was deposited under the Budapest Treaty at the American Type Culture Collection, Patent Depository, 10801 University Boulevard, Manassas, VA 20110-2209, on July 14, 2009, and given ATCC Accession No. PTA-10212, and is also referred to as *Thraustochytrium* sp. ATCC PTA-10212. When expressed, these genes produce unique fatty acid profiles, characterized in part by high levels of omega-3 fatty acids, in particular high levels of DHA, EPA, or a combination thereof.

[0071] As used herein, a "polynucleotide" can comprise a conventional phosphodiester bond or a non-conventional bond (e.g., an amide bond, such as found in peptide nucleic acids (PNA)). A polynucleotide can contain the nucleotide sequence of the full length cDNA sequence, including the untranslated 5' and 3' sequences, the coding sequences, as well as fragments, epitopes, domains, and variants of the nucleic acid sequence. The polynucleotide can be composed of any polyribonucleotide or polydeoxyribonucleotide, which can be unmodified RNA or DNA or modified RNA or DNA. For example, polynucleotides can be composed of single- and double-stranded DNA, DNA that is a mixture of single- and double-stranded regions, single- and double-stranded RNA, and RNA that is mixture of single- and double-stranded regions, hybrid molecules comprising DNA and RNA that can be single-stranded or, more typically, double-stranded or a mixture of single- and double-stranded regions. In addition, the polynucleotides can be composed of triple-stranded regions comprising RNA or DNA or both RNA and DNA. Polynucleotides can contain ribonucleosides (adenosine, guanosine, uridine, or cytidine; "RNA molecules") or deoxyribonucleosides (deoxyadenosine, deoxyguanosine, deoxythymidine, or deoxycytidine; "DNA molecules"), or any phosphoester analogs thereof, such as phosphorothioates and thioesters. Polynucleotides can also contain one or more modified bases or DNA or RNA backbones modified for stability or for other reasons. "Modified" bases include, for example, tritylated bases and unusual bases such as inosine. A variety of modifications can be made to DNA and RNA; thus, "polynucleotide" embraces chemically, enzymatically, or metabolically modified forms. The term nucleic acid molecule refers only to the primary and secondary structure of the

molecule, and does not limit it to any particular tertiary forms. Thus, this term includes double-stranded DNA found, inter alia, in linear or circular DNA molecules (e.g., restriction fragments), plasmids, and chromosomes. In discussing the structure of particular double-stranded DNA molecules, sequences can be described herein according to the normal convention of giving only the sequence in the 5' to 3' direction along the non-transcribed strand of DNA (i.e., the strand having a sequence homologous to the mRNA).

[0072] The terms "isolated" nucleic acid molecule refers to a nucleic acid molecule, DNA or RNA, which has been removed from its native environment. Further examples of isolated nucleic acid molecules include nucleic acid molecules comprising recombinant polynucleotides maintained in heterologous host cells or purified (partially or substantially) polynucleotides in solution. Isolated RNA molecules include *in vivo* or *in vitro* RNA transcripts of polynucleotides of the present invention. Isolated nucleic acid molecules according to the present invention further include such molecules produced synthetically. In addition, a nucleic acid molecule or polynucleotide can include a regulatory element such as a promoter, ribosome binding site, or a transcription terminator.

[0073] A "gene" refers to an assembly of nucleotides that encode a polypeptide, and includes cDNA and genomic DNA nucleic acids. "Gene" also refers to a nucleic acid fragment that expresses a specific protein, including intervening sequences (introns) between individual coding segments (exons), as well as 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.

[0074] The present invention is directed to isolated nucleic acid molecules comprising polynucleotide sequences at least 80% identical to the polynucleotide sequences of *Schizochytrium sp.* ATCC PTA-9695 *PFA1* (SEQ ID NO:1), *Schizochytrium sp.* ATCC PTA-9695 *PFA2* (SEQ ID NO:3), *Schizochytrium sp.* ATCC PTA-9695 *PFA3* (SEQ ID NO:5), *Thraustochytrium sp.* ATCC PTA-10212 *PFA1* (SEQ ID NO:68 or SEQ ID NO:120), *Thraustochytrium sp.* ATCC PTA-10212 *PFA2* (SEQ ID NO:70 or SEQ ID NO:121), *Thraustochytrium sp.* ATCC PTA-10212 *PFA3* (SEQ ID NO:72 or SEQ ID NO:122), and combinations thereof, wherein the polynucleotides encode polypeptides comprising one or more PUFA synthase activities.

- [0075] The PUFA synthase activities are associated with one or more domains in each synthase polypeptide, wherein the domains can be identified by their conserved structural or functional motifs based on their homology to known motifs and can also be identified based upon their specific biochemical activities. *See, e.g.,* U.S. Patent No. 7,217,856, incorporated by reference herein in its entirety. Examples of PUFA synthase domains include: the beta-ketoacyl-ACP synthase (KS) domain, malonyl-CoA:ACP acyltransferase (MAT) domain, acyl carrier protein (ACP) domains, ketoreductase (KR) domain, and beta-hydroxyacyl-ACP dehydrase (DH) domain in Pfa1p; the KS domain, chain length factor (CLF) domain, acyltransferase (AT) domain, and enoyl-ACP reductase (ER) domain in Pfa2p; and the DH domains and the ER domain in Pfa3p.
- [0076] A polypeptide or domain of a polypeptide having beta-ketoacyl-ACP synthase (KS) biological activity (function) has been previously shown to be capable of carrying out the initial step of the fatty acid elongation reaction cycle. The term "beta-ketoacyl-ACP synthase" has been used interchangeably with the terms "3-keto acyl-ACP synthase," "beta-ketoacyl-ACP synthase," and "keto-acyl ACP synthase." In other systems, it has been shown that the acyl group for elongation is linked to a cysteine residue at the active site of KS by a thioester bond, and the acyl-KS undergoes condensation with malonyl-ACP to form -ketoacyl-ACP, CO₂, and unbound ("free") KS. In such systems, KS has been shown to possess greater substrate specificity than other polypeptides of the reaction cycle. Polypeptides (or domains of polypeptides) can be readily identified as belonging to the KS family by homology to known KS sequences.
- [0077] A polypeptide or a domain of a polypeptide having malonyl-CoA:ACP acyltransferase (MAT) activity has been previously shown to be capable of transferring the malonyl moiety from malonyl-CoA to ACP. The term "malonyl-CoA:ACP acyltransferase" has been used interchangeably with "malonyl acyltransferase." In addition to the active site motif (GxSxG), MATs have been shown to possess an extended motif (R and Q amino acids in key positions). Polypeptides (or domains of polypeptides) can be readily identified as belonging to the MAT family by their homology to known MAT sequences and by their extended motif structure.
- [0078] A polypeptide or a domain of a polypeptide having acyl carrier protein (ACP) activity has been previously shown to be capable of functioning as a carrier for growing fatty acyl chains via a thioester linkage to a covalently bound co-factor. ACPs are

typically about 80 to about 100 amino acids long and have been shown to be converted from inactive apo-forms to functional holo-forms by transfer of the phosphopantetheinyl moiety of CoA to a highly conserved serine residue of the ACP. It has also been shown that acyl groups are attached to ACPs by a thioester linkage at the free terminus of the phosphopantetheinyl moiety. The presence of variations of an active site motif (LGIDS*) has also been recognized as a signature of ACPs. The functionality of the active site serine (S*) has been demonstrated in a bacterial PUFA synthase (Jiang *et al.*, *J. Am. Chem. Soc.* 130:6336-7 (2008)). Polypeptides (or domains of polypeptides) can be readily identified as belonging to the ACP family by labeling with radioactive pantetheine and by sequence homology to known ACPs.

[0079] A polypeptide or a domain of a polypeptide having dehydrase or dehydratase (DH) activity has been previously shown to be capable of catalyzing a dehydration reaction. Reference to DH activity typically refers to FabA-like beta-hydroxyacyl-ACP dehydrase biological activity. FabA-like beta-hydroxyacyl-ACP dehydrase biological activity removes HOH from a beta-ketoacyl-ACP and initially produces a trans double bond in the carbon chain. The term "FabA-like beta-hydroxyacyl-ACP dehydrase" has been used interchangeably with the terms "FabA-like beta-hydroxy acyl-ACP dehydrase," "beta-hydroxyacyl-ACP dehydrase," and "dehydrase." The DH domains of PUFA synthase systems have previously been demonstrated as showing homology to bacterial DH enzymes associated with FAS systems (rather than to the DH domains of other PKS systems). *See*, e.g., U.S. Patent No. 7,217,856, incorporated by reference herein in its entirety. A subset of bacterial DHs, the FabA-like DHs, possesses cis-trans isomerase activity (Heath *et al.*, *J. Biol. Chem.*, 271, 27795 (1996)). Based on homology to the FabA-like DH proteins, one or all of the PUFA synthase system DH domains can be responsible for insertion of cis double bonds in the PUFA synthase products. A polypeptide or domain can also have non-FabA-like DH activity, or non-FabA-like beta-hydroxyacyl-ACP dehydrase (DH) activity. More specifically, a conserved active site motif of about 13 amino acids in length has been previously identified in PUFA synthase DH domains: LxxHxxxGxxxxP (the L position can also be an I in the motif). *See*, e.g., U.S. Patent No. 7,217,856, and Donadio S, Katz L., *Gene* 111(1):51-60 (1992), each of which is incorporated by reference herein in its entirety. This conserved motif is found in

a similar region of all known PUFA synthase sequences and could be responsible for a non-FabA like dehydration.

[0080] A polypeptide or a domain of a polypeptide having beta-ketoacyl-ACP reductase (KR) activity has been previously shown to be capable of catalyzing the pyridine-nucleotide-dependent reduction of 3-ketoacyl forms of ACP. The term "beta-ketoacyl-ACP reductase" has been used interchangeably with the terms "ketoreductase," "3-ketoacyl-ACP reductase," and "keto-acyl ACP reductase." It has been determined in other systems that KR function involves the first reductive step in the de novo fatty acid biosynthesis elongation cycle. Polypeptides (or domains of polypeptides) can be readily identified as belonging to the KR family by sequence homology to known PUFA synthase KRs.

[0081] A polypeptide or a domain of a polypeptide having chain length factor (CLF) activity has been previously defined as having one or more of the following activities or characteristics: (1) it can determine the number of elongation cycles and hence chain length of the end product, (2) it has homology to KS, but lacks the KS active site cysteine, (3) it can heterodimerize with KS, (4) it can provide the initial acyl group to be elongated, or (5) it can decarboxylate malonate (as malonyl-ACP), thus forming an acetate group that can be transferred to the KS active site and that can act as the 'priming' molecule that undergoes the initial elongation (condensation) reaction. A CLF domain is found in all currently identified PUFA synthase systems and in each case is found as part of a multidomain protein. Polypeptides (or domains of polypeptides) can be readily identified as belonging to the CLF family by sequence homology to known PUFA synthase CLFs.

[0082] A polypeptide or a domain of a polypeptide having acyltransferase (AT) activity has been previously defined as having one or more of the following activities or characteristics: (1) it can transfer the fatty acyl group from the ACP domain(s) to water (i.e., a thioesterase), releasing the fatty acyl group as a free fatty acid, (2) it can transfer a fatty acyl group to an acceptor such as CoA, (3) it can transfer the acyl group among the various ACP domains, or (4) it can transfer the fatty acyl group to a lipophilic acceptor molecule (e.g. to lysophosphadidic acid). Polypeptides (or domains of polypeptides) can be readily identified as belonging to the AT family by sequence homology to known PUFA synthase ATs.

[0083] A polypeptide or a domain of a polypeptide having enoyl-ACP reductase (ER) biological activity has been previously shown to be capable of reducing the trans-double bond (introduced by the DH activity) in the fatty acyl-ACP, resulting in saturation of the associated carbons. The ER domain in PUFA synthase systems has previously been shown to have homology to a family of ER enzymes (Heath *et al.*, *Nature* 406: 145-146 (2000), incorporated by reference herein in its entirety), and an ER homolog has been shown to function as an enoyl-ACP reductase *in vitro* (Bumpus *et al.* *J. Am. Chem. Soc.*, 130: 11614-11616 (2008), incorporated by reference herein in its entirety). The term "enoyl-ACP reductase" has been used interchangeably with "enoyl reductase," "enoyl ACP-reductase," and "enoyl acyl-ACP reductase." Polypeptides (or domains of polypeptides) can be readily identified as belonging to the ER family by sequence homology to known PUFA synthase ERs.

[0084] In some embodiments, the present invention is directed to nucleic acid molecules comprising a polynucleotide sequence at least 80% identical to a polynucleotide sequence within *PFAI* (SEQ ID NO:1, SEQ ID NO:68, or SEQ ID NO:120) that encodes one or more PUFA synthase domains. In some embodiments, the nucleic acid molecule comprises a polynucleotide sequence at least 80% identical to a polynucleotide sequence within *PFAI* (SEQ ID NO:1, SEQ ID NO:68, or SEQ ID NO:120) that encodes one or more PUFA synthase domains such as a KS domain (SEQ ID NO:7 or SEQ ID NO:74), a MAT domain (SEQ ID NO:9 or SEQ ID NO:76), an ACP domain (such as any one of SEQ ID NOs:13, 15, 17, 19, 21, 23, 80, 82, 84, 86, 88, 90, 92, 94, 96, or 98), a combination of two or more ACP domains, such as two, three, four, five, six, seven, eight, nine, or ten ACP domains, including tandem domains (SEQ ID NO:11 or SEQ ID NO:78, and portions thereof), a KR domain (SEQ ID NO:25 or SEQ ID NO:100), a DH domain (SEQ ID NO:27 or SEQ ID NO:118), and combinations thereof. In some embodiments, the nucleic acid molecule comprises two or more polynucleotide sequences, wherein each of the at least two or more polynucleotide sequences is 80% identical to a polynucleotide sequence within *PFAI* (SEQ ID NO:1, SEQ ID NO:68, or SEQ ID NO:120) that encodes one or more PUFA synthase domains. In some embodiments, the at least two or more polynucleotide sequences are 80% identical to the same polynucleotide sequence within SEQ ID NO:1, SEQ ID NO:68, or SEQ ID NO:120 that encodes one or more PUFA synthase domains. In some embodiments, the at least

two or more polynucleotide sequences are 80% identical to different polynucleotide sequences within SEQ ID NO:1, SEQ ID NO:68, or SEQ ID NO:120 that each encode one or more PUFA synthase domains. In some embodiments, the at least two or more polynucleotide sequences are 80% identical to different polynucleotide sequences within SEQ ID NO:1, SEQ ID NO:68, or SEQ ID NO:120, wherein the at least two or more polynucleotide sequences are located in the same order or a different order in the nucleic acid molecule as compared to the order of the corresponding sequences within SEQ ID NO:1, SEQ ID NO:68, or SEQ ID NO:120. In some embodiments, each of the at least two or more polynucleotide sequences are 80% identical to a polynucleotide sequence within *PFAI* (SEQ ID NO:1, SEQ ID NO:68, or SEQ ID NO:120) that encodes one or more PUFA synthase domains such as a KS domain (SEQ ID NO:7 or SEQ ID NO:74), a MAT domain (SEQ ID NO:9 or SEQ ID NO:76), an ACP domain (such as any one of SEQ ID NOs:13, 15, 17, 19, 21, 23, 80, 82, 84, 86, 88, 90, 92, 94, 96, or 98), a combination of two, three, four, five, six, seven, eight, nine, or ten ACP domains, including tandem domains (SEQ ID NO:11 or SEQ ID NO:78, and portions thereof), a KR domain (SEQ ID NO:25 or SEQ ID NO:100), a DH domain (SEQ ID NO:27 or SEQ ID NO:118), and combinations thereof. In some embodiments, the nucleic acid molecule comprises one or more polynucleotide sequences within *PFAI* (SEQ ID NO:1, SEQ ID NO:68, or SEQ ID NO:120) that encodes one or more PUFA synthase domains, including one or more copies of any individual domain in combination with one or more copies of any other individual domain.

[0085] In some embodiments, the present invention is directed to nucleic acid molecules comprising a polynucleotide sequence at least 80% identical to a polynucleotide sequence within *PFA2* (SEQ ID NO:3, SEQ ID NO:70, or SEQ ID NO:121) that encodes one or more PUFA synthase domains. In some embodiments, the nucleic acid molecule comprises a polynucleotide sequence at least 80% identical to the polynucleotide sequence within *PFA2* (SEQ ID NO:3, SEQ ID NO:70, or SEQ ID NO:121) that encodes one or more PUFA synthase domains such as a KS domain (SEQ ID NO:29 or SEQ ID NO:102), a CLF domain (SEQ ID NO:31 or SEQ ID NO:104), an AT domain (SEQ ID NO:33 or SEQ ID NO:106), an ER domain (SEQ ID NO:35 or SEQ ID NO:108), and combinations thereof. In some embodiments, the nucleic acid molecule comprises two or more polynucleotide sequences, wherein each of the at least two or more polynucleotide

sequences is 80% identical to a polynucleotide sequence within *PFA2* (SEQ ID NO:3, SEQ ID NO:70, or SEQ ID NO:121) that encodes one or more PUFA synthase domains. In some embodiments, the at least two or more polynucleotide sequences are 80% identical to the same polynucleotide sequence within SEQ ID NO:3, SEQ ID NO:70, or SEQ ID NO:121 that encodes one or more PUFA synthase domains. In some embodiments, the at least two or more polynucleotide sequences are 80% identical to different polynucleotide sequences within SEQ ID NO:3, SEQ ID NO:70, or SEQ ID NO:121 that each encode one or more PUFA synthase domains. In some embodiments, the at least two or more polynucleotide sequences are 80% identical to different polynucleotide sequences within SEQ ID NO:3, SEQ ID NO:70, or SEQ ID NO:121, wherein the at least two or more polynucleotide sequences are located in the same order or a different order in the nucleic acid molecule as compared to the order of the corresponding sequences within SEQ ID NO:3, SEQ ID NO:70, or SEQ ID NO:121. In some embodiments, each of the at least two or more polynucleotide sequences are 80% identical to a polynucleotide sequence within *PFA2* (SEQ ID NO:3, SEQ ID NO:70, or SEQ ID NO:121) that encodes one or more PUFA synthase domains such as a KS domain (SEQ ID NO:29 or SEQ ID NO:102), a CLF domain (SEQ ID NO:31 or SEQ ID NO:104), an AT domain (SEQ ID NO:33 or SEQ ID NO:106), an ER domain (SEQ ID NO:35 or SEQ ID NO:108), and combinations thereof. In some embodiments, the nucleic acid molecule comprises one or more polynucleotide sequences within *PFA2* (SEQ ID NO:3, SEQ ID NO:70, or SEQ ID NO:121) that encodes one or more PUFA synthase domains, including one or more copies of any individual domain in combination with one or more copies of any other individual domain.

[0086] In some embodiments, the present invention is directed to nucleic acid molecules comprising a polynucleotide sequence at least 80% identical to a polynucleotide sequence within *PFA3* (SEQ ID NO:5, SEQ ID NO:72, or SEQ ID NO:122) that encodes one or more PUFA synthase domains. In some embodiments, the nucleic acid molecule comprises a polynucleotide sequence at least 80% identical to a polynucleotide sequence within *PFA3* (SEQ ID NO:5, SEQ ID NO:72, or SEQ ID NO:122) that encodes one or more PUFA synthase domains such as a DH domain (such as SEQ ID NO:37, SEQ ID NO:39, SEQ ID NO:110, or SEQ ID NO:112), an ER domain (SEQ ID NO:41 or SEQ ID NO:114), and combinations thereof. In some embodiments, the nucleic acid molecule

comprises two or more polynucleotide sequences, wherein each of the at least two or more polynucleotide sequences is 80% identical to a polynucleotide sequence within *PFA3* (SEQ ID NO:5, SEQ ID NO:72, or SEQ ID NO:122) that encodes one or more PUFA synthase domains. In some embodiments, the at least two or more polynucleotide sequences are 80% identical to the same polynucleotide sequence within SEQ ID NO:5, SEQ ID NO:72, or SEQ ID NO:122 that encodes one or more PUFA synthase domains. In some embodiments, the at least two or more polynucleotide sequences are 80% identical to different polynucleotide sequences within SEQ ID NO:5, SEQ ID NO:72, or SEQ ID NO:122 that each encode one or more PUFA synthase domains. In some embodiments, the at least two or more polynucleotide sequences are 80% identical to different polynucleotide sequences within SEQ ID NO:5, SEQ ID NO:72, or SEQ ID NO:122, wherein the at least two or more polynucleotide sequences are located in the same order or a different order in the nucleic acid molecule as compared to the order of the corresponding sequences within SEQ ID NO:5, SEQ ID NO:72, or SEQ ID NO:122. In some embodiments, each of the at least two or more polynucleotide sequences is 80% identical to a polynucleotide sequence within *PFA3* (SEQ ID NO:5, SEQ ID NO:72, or SEQ ID NO:122) that encodes one or more PUFA synthase domains such as a DH domain (such as SEQ ID NO:37, SEQ ID NO:39, SEQ ID NO:110, or SEQ ID NO:112), an ER domain (SEQ ID NO:41 or SEQ ID NO:114), and combinations thereof. In some embodiments, the nucleic acid molecule comprises one or more polynucleotide sequences within *PFA3* (SEQ ID NO:5, SEQ ID NO:72, or SEQ ID NO:122) that encodes one or more PUFA synthase domains, including one or more copies of any individual domain in combination with one or more copies of any other individual domain.

[0087] In some embodiments, the present invention is directed to a nucleic acid molecule comprising a polynucleotide sequence at least 80% identical to SEQ ID NO:1, SEQ ID NO:68, or SEQ ID NO:120, wherein the polynucleotide sequence encodes a polypeptide comprising PUFA synthase activity selected from the group consisting of KS activity, MAT activity, ACP activity, KR activity, DH activity, and combinations thereof.

[0088] In some embodiments, the present invention is directed to a nucleic acid molecule comprising a polynucleotide sequence at least 80% identical to SEQ ID NO:7 or SEQ ID NO:74, wherein the polynucleotide sequence encodes a polypeptide comprising KS activity.

- [0089] In some embodiments, the present invention is directed to a nucleic acid molecule comprising a polynucleotide sequence at least 80% identical to SEQ ID NO:9 or SEQ ID NO:76, wherein the polynucleotide sequence encodes a polypeptide comprising MAT activity.
- [0090] In some embodiments, the present invention is directed to a nucleic acid molecule comprising a polynucleotide sequence at least 80% identical to any one of SEQ ID NOs:13, 15, 17, 19, 21, 23, 80, 82, 84, 86, 88, 90, 92, 94, 96, or 98, wherein the polynucleotide sequence encodes a polypeptide comprising ACP activity.
- [0091] In some embodiments, the present invention is directed to a nucleic acid molecule comprising a polynucleotide sequence at least 80% identical to SEQ ID NO:11 or SEQ ID NO:78, wherein the polynucleotide sequence encodes a polypeptide comprising ACP activity.
- [0092] In some embodiments, the nucleic acid molecule comprises a polynucleotide sequence at least 80% identical to a polynucleotide sequence within SEQ ID NO:11 that encodes one, two, three, four, five, or six ACP domains, wherein the polynucleotide sequence encodes a polypeptide comprising ACP activity associated with one or more ACP domains. SEQ ID NOs:13, 15, 17, 19, 21, and 23 are representative polynucleotides sequence that each encode a single ACP domain within SEQ ID NO:11.
- [0093] In some embodiments, the nucleic acid molecule comprises a polynucleotide sequence at least 80% identical to a polynucleotide sequence within SEQ ID NO:78 that encodes one, two, three, four, five, six, seven, eight, nine, or ten ACP domains, wherein the polynucleotide sequence encodes a polypeptide comprising ACP activity associated with one or more ACP domains. SEQ ID NOs:80, 82, 84, 86, 88, 90, 92, 94, 96, and 98 are representative polynucleotides sequence that each encode a single ACP domain within SEQ ID NO:78.
- [0094] In some embodiments, the present invention is directed to a nucleic acid molecule comprising a polynucleotide sequence at least 80% identical to SEQ ID NO:25 or SEQ ID NO:100, wherein the polynucleotide sequence encodes a polypeptide comprising KR activity.
- [0095] In some embodiments, the present invention is directed to a nucleic acid molecule comprising a polynucleotide sequence at least 80% identical to SEQ ID NO:27 or SEQ ID

NO:118, wherein the polynucleotide sequence encodes a polypeptide comprising DH activity.

[0096] In some embodiments, the present invention is directed to a nucleic acid molecule comprising a polynucleotide sequence at least 80% identical to SEQ ID NO:3, SEQ ID NO:70, or SEQ ID NO:121, wherein the polynucleotide sequence encodes a polypeptide comprising PUFA synthase activity selected from the group consisting of KS activity, CLF activity, AT activity, ER activity, and combinations thereof.

[0097] In some embodiments, the present invention is directed to a nucleic acid molecule comprising a polynucleotide sequence at least 80% identical to SEQ ID NO:29 or SEQ ID NO:102, wherein the polynucleotide sequence encodes a polypeptide comprising KS activity.

[0098] In some embodiments, the present invention is directed to a nucleic acid molecule comprising a polynucleotide sequence at least 80% identical to SEQ ID NO:31 or SEQ ID NO:104, wherein the polynucleotide sequence encodes a polypeptide comprising CLF activity.

[0099] In some embodiments, the present invention is directed to a nucleic acid molecule comprising a polynucleotide sequence at least 80% identical to SEQ ID NO:33 or SEQ ID NO:106, wherein the polynucleotide sequence encodes a polypeptide comprising AT activity.

[0100] In some embodiments, the present invention is directed to a nucleic acid molecule comprising a polynucleotide sequence at least 80% identical to SEQ ID NO:35 or SEQ ID NO:108, wherein the polynucleotide sequence encodes a polypeptide comprising ER activity.

[0101] In some embodiments, the present invention is directed to a nucleic acid molecule comprising a polynucleotide sequence at least 80% identical to SEQ ID NO:5, SEQ ID NO:72, or SEQ ID NO:122, wherein the polynucleotide sequence encodes a polypeptide comprising PUFA synthase activity selected from the group consisting of DH activity, ER activity, and combinations thereof.

[0102] In some embodiments, the present invention is directed to a nucleic acid molecule comprising a polynucleotide sequence at least 80% identical to SEQ ID NO:37, wherein the polynucleotide sequence encodes a polypeptide comprising DH activity.

- [0103] In some embodiments, the present invention is directed to a nucleic acid molecule comprising a polynucleotide sequence at least 80% identical to SEQ ID NO:39, wherein the polynucleotide sequence encodes a polypeptide comprising DH activity.
- [0104] In some embodiments, the present invention is directed to a nucleic acid molecule comprising a polynucleotide sequence at least 80% identical to SEQ ID NO:110, wherein the polynucleotide sequence encodes a polypeptide comprising DH activity.
- [0105] In some embodiments, the present invention is directed to a nucleic acid molecule comprising a polynucleotide sequence at least 80% identical to SEQ ID NO:112, wherein the polynucleotide sequence encodes a polypeptide comprising DH activity.
- [0106] In some embodiments, the present invention is directed to a nucleic acid molecule comprising a polynucleotide sequence at least 80% identical to SEQ ID NO:41 or SEQ ID NO:114, wherein the polynucleotide sequence encodes a polypeptide comprising ER activity.
- [0107] The present invention is directed to isolated nucleic acid molecules comprising polynucleotide sequences encoding polypeptides, wherein the polypeptides comprise amino acid sequences that are at least 80% identical to the amino acid sequences of Pfa1p (SEQ ID NO:2 or SEQ ID NO:69), Pfa2p (SEQ ID NO:4 or SEQ ID NO:71), or Pfa3p (SEQ ID NO:6 or SEQ ID NO:73), wherein the polynucleotides encode polypeptides comprising one or more PUFA synthase activities.
- [0108] The present invention is directed to nucleic acid molecules comprising a polynucleotide sequence encoding a polypeptide, wherein the polypeptide comprises an amino acid sequence at least 80% identical to the amino acid sequences of one or more PUFA synthase domains of the PUFA synthases of the invention.
- [0109] In some embodiments, the present invention is directed to nucleic acid molecules comprising a polynucleotide sequence encoding a polypeptide, wherein the polypeptide comprises an amino acid sequence at least 80% identical to an amino acid sequence within Pfa1p (SEQ ID NO:2 or SEQ ID NO:69) comprising one or more PUFA synthase domains. In some embodiments, the polypeptide comprises an amino acid sequence at least 80% identical to an amino acid sequence within Pfa1p (SEQ ID NO:2 or SEQ ID NO:69) comprising one or more PUFA synthase domains such as a KS domain (SEQ ID NO:8 or SEQ ID NO:75), a MAT domain (SEQ ID NO:10 or SEQ ID NO:77), an ACP domain (such as any one of SEQ ID NOs:14, 16, 18, 20, 22, 24, 81, 83, 85, 87, 89, 91, 93,

95, 97, or 99), a combination of two or more ACP domains, such as two, three, four, five, six, seven, eight, nine, or ten ACP domains, including tandem domains (SEQ ID NO:12 or SEQ ID NO:79, and portions thereof), a KR domain (SEQ ID NO:26 or SEQ ID NO:101), a DH domain (SEQ ID NO:28 or SEQ ID NO:119), and combinations thereof. In some embodiments, the polypeptide comprises two or more amino acid sequences, wherein each of the at least two or more amino acid sequences is 80% identical to an amino acid sequence within Pfa1p (SEQ ID NO:2 or SEQ ID NO:69) comprising one or more PUFA synthase domains. In some embodiments, the at least two or more amino acid sequences are 80% identical to the same amino acid sequence within Pfa1p (SEQ ID NO:2 or SEQ ID NO:69) comprising one or more PUFA synthase domains. In some embodiments, the at least two or more amino acid sequences are 80% identical to different amino acid sequences within Pfa1p (SEQ ID NO:2 or SEQ ID NO:69) that each comprise one or more PUFA synthase domains. In some embodiments, the at least two or more amino acid sequences are 80% identical to different amino acid sequences within Pfa1p (SEQ ID NO:2 or SEQ ID NO:69), wherein the at least two or more amino acid sequences are located in the same order or a different order in the polypeptide as compared to the order of the corresponding domains within Pfa1p (SEQ ID NO:2 or SEQ ID NO:69). In some embodiments, the at least two or more amino acid sequences are 80% identical to an amino acid sequence within Pfa1p (SEQ ID NO:2 or SEQ ID NO:69) comprising one or more PUFA synthase domains such as a KS domain (SEQ ID NO:8 or SEQ ID NO:75), a MAT domain (SEQ ID NO:10 or SEQ ID NO:77), an ACP domain (such as any one of SEQ ID NOs:14, 16, 18, 20, 22, 24, 81, 83, 85, 87, 89, 91, 93, 95, 97, or 99), a combination of two, three, four, five, six, seven, eight, nine, or ten ACP domains, including tandem domains (SEQ ID NO:12 or SEQ ID NO:79, and portions thereof), a KR domain (SEQ ID NO:26 or SEQ ID NO:101), a DH domain (SEQ ID NO:28 or SEQ ID NO:119), and combinations thereof. In some embodiments, the polypeptide comprises one or more amino acid sequences within Pfa1p (SEQ ID NO:2 or SEQ ID NO:69) comprising one or more PUFA synthase domains, including one or more copies of any individual domain in combination with one or more copies of any other individual domain.

[0110] In some embodiments, the present invention is directed to nucleic acid molecules comprising a polynucleotide sequence encoding a polypeptide, wherein the polypeptide

comprises an amino acid sequence at least 80% identical to an amino acid sequence within Pfa2p (SEQ ID NO:4 or SEQ ID NO:71) comprising one or more PUFA synthase domains. In some embodiments, the polypeptide comprises an amino acid sequence at least 80% identical to an amino acid sequence within Pfa2p (SEQ ID NO:4 or SEQ ID NO:71) comprising one or more PUFA synthase domains such as a KS domain (SEQ ID NO:30 or SEQ ID NO:103), a CLF domain (SEQ ID NO:32 or SEQ ID NO:105), an AT domain (SEQ ID NO:34 or SEQ ID NO:107), an ER domain (SEQ ID NO:36 or SEQ ID NO:109), and combinations thereof. In some embodiments, the polypeptide comprises two or more amino acid sequences, wherein each of the at least two or more amino acid sequences is 80% identical to an amino acid sequence within Pfa2p (SEQ ID NO:4 or SEQ ID NO:71) comprising one or more PUFA synthase domains. In some embodiments, the at least two or more amino acid sequences are 80% identical to the same amino acid sequence within Pfa2p (SEQ ID NO:4 or SEQ ID NO:71). In some embodiments, the at least two or more amino acid sequences are 80% identical to different amino acid sequences within Pfa2p (SEQ ID NO:4 or SEQ ID NO:71) that each comprise one or more PUFA synthase domains. In some embodiments, the at least two or more amino acid sequences are 80% identical to different amino acid sequences within Pfa2p (SEQ ID NO:4 or SEQ ID NO:71), wherein the at least two or more amino acid sequences are located in the same order or a different order in the polypeptide as compared to the order of the corresponding domains within Pfa2p (SEQ ID NO:4 or SEQ ID NO:71). In some embodiments, the at least two or more amino acid sequences are 80% identical to an amino acid sequence within Pfa2p (SEQ ID NO:4 or SEQ ID NO:71) comprising one or more PUFA synthase domains such as a KS domain (SEQ ID NO:30 or SEQ ID NO:103), a CLF domain (SEQ ID NO:32 or SEQ ID NO:105), an AT domain (SEQ ID NO:34 or SEQ ID NO:107), an ER domain (SEQ ID NO:36 or SEQ ID NO:109), and combinations thereof. In some embodiments, the polypeptide comprises one or more amino acid sequences within Pfa2p (SEQ ID NO:4 or SEQ ID NO:71) comprising one or more PUFA synthase domains, including one or more copies of any individual domain in combination with one or more copies of any other individual domain.

[0111] In some embodiments, the present invention is directed to nucleic acid molecules comprising a polynucleotide sequence encoding a polypeptide, wherein the polypeptide

comprises an amino acid sequence at least 80% identical to an amino acid sequence within Pfa3p (SEQ ID NO:6 or SEQ ID NO:73) comprising one or more PUFA synthase domains. In some embodiments, the polypeptide comprises an amino acid sequence at least 80% identical to an amino acid sequence within Pfa3p (SEQ ID NO:6 or SEQ ID NO:73) comprising one or more PUFA synthase domains such as a DH domain (such as SEQ ID NO:38, SEQ ID NO:40, SEQ ID NO:111, or SEQ ID NO:113), an ER domain (SEQ ID NO:42 or SEQ ID NO:115), and combinations thereof. In some embodiments, the polypeptide comprises two or more amino acid sequences, wherein each of the at least two or more amino acid sequences is 80% identical to an amino acid sequence within Pfa3p (SEQ ID NO:6 or SEQ ID NO:73) comprising one or more PUFA synthase domains. In some embodiments, the at least two or more amino acid sequences are 80% identical to the same amino acid sequence within Pfa3p (SEQ ID NO:6 or SEQ ID NO:73) comprising one or more PUFA synthase domains. In some embodiments, the at least two or more amino acid sequences are 80% identical to different amino acid sequences within Pfa3p (SEQ ID NO:6 or SEQ ID NO:73) that each comprise one or more PUFA synthase domains. In some embodiments, the at least two or more amino acid sequences are 80% identical to different amino acid sequences within Pfa3p (SEQ ID NO:6 or SEQ ID NO:73), wherein the at least two or more amino acid sequences are located in the same order or a different order in the polypeptide as compared to the order of the corresponding domains within Pfa3p (SEQ ID NO:6 or SEQ ID NO:73). In some embodiments, the at least two or more amino acid sequences are 80% identical to an amino acid sequence within Pfa3p (SEQ ID NO:6 or SEQ ID NO:73) comprising one or more PUFA synthase domains such as a DH domain (such as SEQ ID NO:38, SEQ ID NO:40, SEQ ID NO:111, or SEQ ID NO:113), an ER domain (SEQ ID NO:42 or SEQ ID NO:115), and combinations thereof. In some embodiments, the polypeptide comprises one or more amino acid sequences within Pfa3p (SEQ ID NO:6 or SEQ ID NO:73) comprising one or more PUFA synthase domains, including one or more copies of any individual domain in combination with one or more copies of any other individual domain.

[0112] In some embodiments, the present invention is directed to a nucleic acid molecule comprising a polynucleotide sequence encoding a polypeptide, wherein the polypeptide comprises an amino acid sequence at least 80% identical to SEQ ID NO:2 or SEQ ID

NO:69, and wherein the polypeptide comprises a PUFA synthase activity selected from the group consisting of KS activity, MAT activity, ACP activity, KR activity, DH activity, and combinations thereof.

[0113] In some embodiments, the present invention is directed to a nucleic acid molecule comprising a polynucleotide sequence encoding a polypeptide, wherein the polypeptide comprises an amino acid sequence at least 80% identical to SEQ ID NO:8 or SEQ ID NO:75, and wherein the polypeptide comprises KS activity.

[0114] In some embodiments, the present invention is directed to a nucleic acid molecule comprising a polynucleotide sequence encoding a polypeptide, wherein the polypeptide comprises an amino acid sequence at least 80% identical to SEQ ID NO:10 or SEQ ID NO:77, and wherein the polypeptide comprises MAT activity.

[0115] In some embodiments, the present invention is directed to a nucleic acid molecule comprising a polynucleotide sequence encoding a polypeptide, wherein the polypeptide comprises an amino acid sequence at least 80% identical to any one of SEQ ID NOs:14, 16, 18, 20, 22, 24, 81, 83, 85, 87, 89, 91, 93, 95, 97, or 99, and wherein the polypeptide comprises ACP activity.

[0116] In some embodiments, the present invention is directed to a nucleic acid molecule comprising a polynucleotide sequence encoding a polypeptide, wherein the polypeptide comprises an amino acid sequence at least 80% identical to SEQ ID NO:12 or SEQ ID NO:79, and wherein the polypeptide comprises ACP activity.

[0117] In some embodiments, the present invention is directed to nucleic acid molecules comprising a polynucleotide sequence encoding a polypeptide, wherein the polypeptide comprises an amino acid sequence at least 80% identical to an amino acid sequence within SEQ ID NO:12, wherein the polypeptide comprises ACP activity. In some embodiments, the amino acid sequence is at least 80% identical to an amino acid sequence within SEQ ID NO:12 comprising one, two, three, four, five, or six ACP domains, wherein the polypeptide comprises ACP activity associated with one or more ACP domains. SEQ ID NOs:14, 16, 18, 20, 22 and 24 are representative amino acid sequences, each comprising a single ACP domain within SEQ ID NO: 12.

[0118] In some embodiments, the present invention is directed to nucleic acid molecules comprising a polynucleotide sequence encoding a polypeptide, wherein the polypeptide comprises an amino acid sequence at least 80% identical to an amino acid sequence

within SEQ ID NO:79, wherein the polypeptide comprises ACP activity. In some embodiments, the amino acid sequence is at least 80% identical to an amino acid sequence within SEQ ID NO:79 comprising one, two, three, four, five, six, seven, eight, nine, or ten ACP domains, wherein the polypeptide comprises ACP activity associated with one or more ACP domains. SEQ ID NOs:81, 83, 85, 87, 89, 91, 93, 95, 97, and 99 are representative amino acid sequences, each comprising a single ACP domain within SEQ ID NO:79.

- [0119] In some embodiments, the present invention is directed to a nucleic acid molecule comprising a polynucleotide sequence encoding a polypeptide, wherein the polypeptide comprises an amino acid sequence at least 80% identical to SEQ ID NO:26 or SEQ ID NO:101, and wherein the polypeptide comprises KR activity.
- [0120] In some embodiments, the present invention is directed to a nucleic acid molecule comprising a polynucleotide sequence encoding a polypeptide, wherein the polypeptide comprises an amino acid sequence at least 80% identical to SEQ ID NO:28 or SEQ ID NO:119, and wherein the polypeptide comprises DH activity.
- [0121] In some embodiments, the present invention is directed to nucleic acid molecules comprising a polynucleotide sequence encoding a polypeptide, wherein the polypeptide comprises an amino acid sequence at least 80% identical to SEQ ID NO:4 or SEQ ID NO:71, wherein the polypeptide comprises a PUFA synthase activity selected from the group consisting of KS activity, CLF activity, AT activity, ER activity, and combinations thereof.
- [0122] In some embodiments, the present invention is directed to a nucleic acid molecule comprising a polynucleotide sequence encoding a polypeptide, wherein the polypeptide comprises an amino acid sequence at least 80% identical to SEQ ID NO:30 or SEQ ID NO:103, and wherein the polypeptide comprises KS activity.
- [0123] In some embodiments, the present invention is directed to a nucleic acid molecule comprising a polynucleotide sequence encoding a polypeptide, wherein the polypeptide comprises an amino acid sequence at least 80% identical to SEQ ID NO:32 or SEQ ID NO:105, and wherein the polypeptide comprises CLF activity.
- [0124] In some embodiments, the present invention is directed to a nucleic acid molecule comprising a polynucleotide sequence encoding a polypeptide, wherein the polypeptide

comprises an amino acid sequence at least 80% identical to SEQ ID NO:34 or SEQ ID NO:107, and wherein the polypeptide comprises AT activity.

[0125] In some embodiments, the present invention is directed to a nucleic acid molecule comprising a polynucleotide sequence encoding a polypeptide, wherein the polypeptide comprises an amino acid sequence at least 80% identical to SEQ ID NO:36 or SEQ ID NO:109, and wherein the polypeptide comprises ER activity.

[0126] In some embodiments, the present invention is directed to a nucleic acid molecule comprising a polynucleotide sequence encoding a polypeptide, wherein the polypeptide comprises an amino acid sequence at least 80% identical to SEQ ID NO:6 or SEQ ID NO:73, wherein the polypeptide comprises a PUFA synthase activity selected from the group consisting of DH activity, ER activity, and combinations thereof.

[0127] In some embodiments, the present invention is directed to a nucleic acid molecule comprising a polynucleotide sequence encoding a polypeptide, wherein the polypeptide comprises an amino acid sequence at least 80% identical to SEQ ID NO:38, and wherein the polypeptide comprises DH activity.

[0128] In some embodiments, the present invention is directed to a nucleic acid molecule comprising a polynucleotide sequence encoding a polypeptide, wherein the polypeptide comprises an amino acid sequence at least 80% identical to SEQ ID NO:40, and wherein the polypeptide comprises DH activity.

[0129] In some embodiments, the present invention is directed to a nucleic acid molecule comprising a polynucleotide sequence encoding a polypeptide, wherein the polypeptide comprises an amino acid sequence at least 80% identical to SEQ ID NO:111, and wherein the polypeptide comprises DH activity.

[0130] In some embodiments, the present invention is directed to a nucleic acid molecule comprising a polynucleotide sequence encoding a polypeptide, wherein the polypeptide comprises an amino acid sequence at least 80% identical to SEQ ID NO:113, and wherein the polypeptide comprises DH activity.

[0131] In some embodiments, the present invention is directed to a nucleic acid molecule comprising a polynucleotide sequence encoding a polypeptide, wherein the polypeptide comprises an amino acid sequence at least 80% identical to SEQ ID NO:42 or SEQ ID NO:115, and wherein the polypeptide comprises ER activity.

[0132] In some embodiments, the nucleic acid molecules comprise polynucleotide sequences at least about 80%, 85%, or 90% identical to the polynucleotide sequences reported herein, or at least about 95%, 96%, 97%, 98%, 99%, or 100% identical to the polynucleotide sequences reported herein. The term "percent identity," as known in the art, is a relationship between two or more amino acid 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 amino acid or polynucleotide sequences, as the case may be, as determined by the match between strings of such sequences.

[0133] By a nucleic acid molecule having a polynucleotide sequence at least, for example, 95% "identical" to a reference polynucleotide sequence of the present invention, it is intended that the polynucleotide sequence of the nucleic acid molecule is identical to the reference sequence except that the polynucleotide sequence can include up to five nucleotide differences per each 100 nucleotides of the reference polynucleotide sequence. In other words, to obtain a nucleic acid molecule having a polynucleotide sequence at least 95% identical to a reference polynucleotide sequence, up to 5% of the nucleotides in the reference sequence can be deleted or substituted with another nucleotide, or a number of nucleotides up to 5% of the total nucleotides in the reference sequence can be inserted into the reference sequence.

[0134] As a practical matter, whether any particular polynucleotide sequence or amino acid sequence is at least 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% identical to a polynucleotide sequence or amino acid sequence of the present invention can be determined conventionally using known computer programs. A method for determining the best overall match between a query sequence (a sequence of the present invention) and a subject sequence can be determined using the alignment of sequences and calculation of identity scores. The alignments were done using the computer program AlignX, which is a component of the Vector NTI Suite 10.0 package from Invitrogen (www.invitrogen.com). The alignments were performed using a ClustalW alignment (Thompson, J.D., *et al. Nucl. Acids Res.* 22: 4673-4680 (1994)) for both amino acid and polynucleotide sequence alignments. The default scoring matrices Blosum62mt2 and swgapdnamt were used for amino acid and polynucleotide sequence alignments, respectively. For amino acid sequences, the default gap opening penalty is 10 and the gap

extension penalty 0.1. For polynucleotide sequences, the default gap opening penalty is 15 and the gap extension penalty is 6.66.

[0135] The present invention is directed to an isolated nucleic acid molecule comprising a polynucleotide sequence encoding a polypeptide comprising PUFA synthase activity selected from the group consisting of KS activity, MAT activity, ACP activity, KR activity, CLF activity, AT activity, ER activity, DH activity, and combinations thereof, wherein the polynucleotide hybridizes under stringent conditions to the complement of any of the polynucleotide sequences described above.

[0136] A nucleic acid molecule is "hybridizable" to another nucleic acid molecule, such as a cDNA, genomic DNA, or RNA, when a single stranded form of the nucleic acid molecule can anneal to the other nucleic acid molecule under the appropriate conditions of temperature and solution ionic strength. Hybridization and washing conditions are well known and exemplified. *See*, e.g., Sambrook J. and Russell D. 2001. Molecular cloning: A laboratory manual, 3rd edition. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York. 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 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. For more stringent conditions, washes are performed at 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 are increased to 60°C. Another set of highly stringent conditions uses two final washes in 0.1X SSC, 0.1% SDS at 65°C. An additional set of highly stringent conditions are defined by hybridization at 0.1X SSC, 0.1% SDS, 65°C and washed with 2X SSC, 0.1% SDS followed by 0.1X SSC, 0.1% SDS.

[0137] The present invention is directed to an isolated nucleic acid molecule comprising a polynucleotide sequence that is fully complementary to any of the polynucleotide sequences described above. The term "complementary" is used to describe the relationship between nucleotide bases that are capable to hybridizing to one another. For

example, with respect to DNA, adenosine is complementary to thymine and cytosine is complementary to guanine.

[0138] In certain embodiments, the polynucleotide or nucleic acid is DNA. In the case of DNA, a nucleic acid molecule comprising a polynucleotide sequence which encodes a polypeptide can normally include a promoter and/or other transcription or translation control elements operably associated with one or more coding regions. An operable association is when a coding region for a gene product, e.g., a polypeptide, is associated with one or more regulatory sequences in such a way as to place expression of the gene product under the influence or control of the regulatory sequence(s). Two DNA fragments (such as a polypeptide coding region and a promoter associated therewith) are "operably associated" if induction of promoter function results in the transcription of mRNA encoding the desired gene product and if the nature of the linkage between the two DNA fragments does not interfere with the ability of the expression regulatory sequences to direct the expression of the gene product or interfere with the ability of the DNA template to be transcribed. Thus, a promoter region would be operably associated with a polynucleotide sequence encoding a polypeptide if the promoter was capable of effecting transcription of that polynucleotide sequence. The promoter can be a cell-specific promoter that directs substantial transcription of the DNA only in predetermined cells. In general, a coding region is located 3' to a promoter. Promoters can 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 can 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 which 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 can have identical promoter activity. A promoter is generally bounded at its 3' terminus by the transcription initiation site and extends upstream (5' direction) to include the minimum number of bases or elements necessary to initiate transcription at levels detectable above background. Within the promoter will be found a transcription initiation

site (conveniently defined for example, by mapping with nuclease S1), as well as protein binding domains (consensus sequences) responsible for the binding of RNA polymerase.

[0139] Suitable regulatory regions include nucleic acid regions located upstream (5' non-coding sequences), within, or downstream (3' non-coding sequences) of a coding region, and which influence the transcription, RNA processing or stability, or translation of the associated coding region. Regulatory regions can include promoters, translation leader sequences, RNA processing sites, effector binding sites, and stem-loop structures. Other transcription control elements, besides a promoter, for example enhancers, operators, repressors, and transcription termination signals, can be operably associated with the polynucleotide to direct cell-specific transcription. The boundaries of the coding region are determined by a start codon at the 5' (amino) terminus and a translation stop codon at the 3' (carboxyl) terminus. A coding region can include, but is not limited to, prokaryotic regions, cDNA from mRNA, genomic DNA molecules, synthetic DNA molecules, or RNA molecules. If the coding region is intended for expression in a eukaryotic cell, a polyadenylation signal and transcription termination sequence will usually be located 3' to the coding region.

[0140] In certain aspects of the invention, polynucleotide sequences having at least 20 bases, at least 30 bases, or at least 50 bases and that hybridize to a polynucleotide sequence of the present invention can be employed as PCR primers. Typically, in PCR-type amplification techniques, the primers have different sequences and are not complementary to each other. Depending on the desired test conditions, the sequences of the primers should be designed to provide for both efficient and faithful replication of the target nucleic acid. Methods of PCR primer design are common and well known in the art. Generally two short segments of the instant sequences can be used in polymerase chain reaction (PCR) protocols to amplify longer nucleic acid fragments encoding homologous genes from DNA or RNA. The polymerase chain reaction can 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 microbial genes. Alternatively, the second primer sequence can be based upon sequences derived from the cloning vector.

[0141] 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 amplification reactions, and used as probes to isolate full length DNA fragments under conditions of appropriate stringency.

[0142] Therefore, the nucleic acid molecules of the present invention can be used to isolate genes encoding homologous proteins from the same or other species or bacterial 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, methods of nucleic acid hybridization, and methods of DNA and RNA amplification as exemplified by various uses of nucleic acid amplification technologies (e.g., polymerase chain reaction, Mullis *et al.*, U.S. Pat. No. 4,683,202; ligase chain reaction (LCR) (Tabor, S. *et al.*, *Proc. Acad. Sci. USA* 82: 1074 (1985)); or strand displacement amplification (SDA; Walker, *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* 89: 392 (1992)).

[0143] In some embodiments, the isolated nucleic acid molecules of the present invention are used to isolate homologous nucleic acid molecules from other organisms in order to identify PUFA synthases that produce similar or improved PUFA profiles. In some embodiments, the isolated nucleic acid molecules of the present invention are used to isolate homologous nucleic acid molecules from other organisms that are involved in producing high amounts of DHA.

[0144] The nucleic acid molecules of the present invention also comprise polynucleotide sequences encoding a PUFA synthase gene, a domain of a PUFA synthase gene, or a fragment of the PUFA synthase gene fused in frame to a marker sequence which allows for detection of the polypeptide of the present invention. Marker sequences include auxotrophic or dominant markers known to one of ordinary skill in the art such as *ZEO* (zeocin), *NEO* (G418), hygromycin, arsenite, *HPH*, *NAT*, and the like.

[0145] The present invention also encompasses variants of the PUFA synthase gene. Variants can contain alterations in the coding regions, non-coding regions, or both. Examples are polynucleotide sequence variants containing alterations which produce silent substitutions, additions, or deletions, but do not alter the properties or activities of the encoded polypeptide. In certain embodiments, polynucleotide sequence variants are produced by silent substitutions due to the degeneracy of the genetic code. In further

embodiments, polynucleotide sequence variants can be produced for a variety of reasons, e.g., to optimize codon expression for a particular host (e.g., changing codons in the thraustochytrid mRNA to those preferred by other organisms such as *E. coli* or *Saccharomyces cerevisiae*).

[0146] Also provided in the present invention are allelic variants, orthologs, and/or species homologs. Procedures known in the art can be used to obtain full-length genes, allelic variants, splice variants, full-length coding portions, orthologs, and/or species homologs of the genes described herein using information from the sequences disclosed herein. For example, allelic variants and/or species homologs can be isolated and identified by making suitable probes or primers from the sequences provided herein and screening a suitable nucleic acid source for allelic variants and/or the desired homologue.

[0147] The present invention is directed to a recombinant nucleic acid molecule comprising any of the nucleic acid molecules described above or combinations thereof and a transcription control sequence. In some embodiments, the recombinant nucleic acid molecule is a recombinant vector.

[0148] The present invention is directed to a method for making a recombinant vector comprising inserting one or more isolated nucleic acid molecules as described herein into a vector.

[0149] The vectors of this invention can be, for example, a cloning vector or an expression vector. The vector can be, for example, in the form of a plasmid, a viral particle, a phage, etc.

[0150] The polynucleotide sequences of the invention can be included in any one of a variety of expression vectors for expressing a polypeptide. Such vectors include chromosomal, nonchromosomal, and synthetic DNA or RNA sequences, e.g., derivatives of SV40; bacterial plasmids; and yeast plasmids. However, any other appropriate vector known to one of ordinary skill in the art can be used.

[0151] The appropriate DNA sequence can be inserted into the vector by a variety of procedures. In general, the DNA sequence is inserted into an appropriate restriction endonuclease site(s) by procedures known in the art. Such procedures and others are deemed to be within the scope of those skilled in the art.

[0152] The present invention also includes recombinant constructs comprising one or more of the polynucleotide sequences described above. The constructs comprise a vector,

such as a plasmid or viral vector, into which one or more sequences of the invention has been inserted, in a forward or reverse orientation. In one aspect of this embodiment, the construct further comprises regulatory sequences, including, for example, a promoter, operably associated to the sequence. Large numbers of suitable vectors and promoters are known to those of skill in the art, and are commercially available.

Polypeptides

[0153] The present invention is directed to isolated polypeptides comprising amino acid sequences for PUFA synthase proteins and domains derived from the isolated microorganisms deposited as ATCC Accession Nos. PTA-9695 and PTA-10212.

[0154] As used herein, the term "polypeptide" is intended to encompass a singular "polypeptide" as well as plural "polypeptides" and refers to a molecule composed of monomers (amino acids) linearly linked by amide bonds (also known as peptide bonds). The term "polypeptide" refers to any chain or chains of two or more amino acids and does not refer to a specific length of the product. Thus, peptides, dipeptides, tripeptides, oligopeptides, "protein," "amino acid chain," or any other term used to refer to a chain or chains of two or more amino acids are included within the definition of "polypeptide," and the term "polypeptide" can be used instead of or interchangeably with any of these terms. The term "polypeptide" is also intended to refer to the products of post-expression modifications of the polypeptide, including without limitation glycosylation, acetylation, phosphorylation, amidation, derivatization by known protecting/blocking groups, proteolytic cleavage, or modification by non-naturally occurring amino acids.

[0155] Polypeptides as described herein can include fragment, variant, or derivative molecules thereof without limitation. The terms "fragment," "variant," "derivative" and "analog" when referring to a polypeptide include any polypeptide which retains at least some biological activity. Polypeptide fragments can include proteolytic fragments, deletion fragments, and fragments which more easily reach the site of action when delivered to an animal. Polypeptide fragments further include any portion of the polypeptide which comprises an antigenic or immunogenic epitope of the native polypeptide, including linear as well as three-dimensional epitopes. Polypeptide fragments can comprise variant regions, including fragments as described above, and also polypeptides with altered amino acid sequences due to amino acid substitutions, deletions, or insertions. Variants can occur naturally, such as an allelic variant. By an

"allelic variant" is intended alternate forms of a gene occupying a given locus on a chromosome of an organism. Non-naturally occurring variants can be produced using art-known mutagenesis techniques. Polypeptide fragments of the invention can comprise conservative or non-conservative amino acid substitutions, deletions, or additions. Variant polypeptides can also be referred to herein as "polypeptide analogs." Polypeptide fragments of the present invention can also include derivative molecules. As used herein a "derivative" of a polypeptide or a polypeptide fragment refers to a subject polypeptide having one or more residues chemically derivatized by reaction of a functional side group. Also included as "derivatives" are those peptides which contain one or more naturally occurring amino acid derivatives of the twenty standard amino acids. For example, 4-hydroxyproline can be substituted for proline; 5-hydroxylysine can be substituted for lysine; 3-methylhistidine can be substituted for histidine; homoserine can be substituted for serine; and ornithine can be substituted for lysine.

[0156] Polypeptides of the invention can be encoded by any of the nucleic acid molecules of the invention.

[0157] The present invention is directed to isolated polypeptides comprising amino acid sequences that are at least 80% identical to the amino acid sequences of Pfa1p (SEQ ID NO:2 or SEQ ID NO:69), Pfa2p (SEQ ID NO:4 or SEQ ID NO:71), Pfa3p (SEQ ID NO:6 or SEQ ID NO:73), and combinations thereof, wherein the polypeptides comprise one or more PUFA synthase activities.

[0158] The present invention is directed to polypeptides comprising amino acid sequences that are at least 80% identical to the amino acid sequences of one or more PUFA synthase domains of the PUFA synthases of the invention.

[0159] In some embodiments, the present invention is directed to polypeptides comprising amino acid sequences that are at least 80% identical to an amino acid sequence within Pfa1p (SEQ ID NO:2 or SEQ ID NO:69) comprising one or more PUFA synthase domains. In some embodiments, the polypeptide comprises an amino acid sequence at least 80% identical to an amino acid sequence within Pfa1p (SEQ ID NO:2 or SEQ ID NO:69) comprising one or more PUFA synthase domains such as a KS domain (SEQ ID NO:8 or SEQ ID NO:75), a MAT domain (SEQ ID NO:10 or SEQ ID NO:77), an ACP domain (such as any one of SEQ ID NOs:14, 16, 18, 20, 22, 24, 81, 83, 85, 87, 89, 91, 93, 95, 97, or 99), a combination of two or more ACP domains such as two, three,

four, five, six, seven, eight, nine, or ten ACP domains, including tandem domains (SEQ ID NO:12 or SEQ ID NO:79, and portions thereof), a KR domain (SEQ ID NO:26 or SEQ ID NO:101), a DH domain (SEQ ID NO:28 or SEQ ID NO:119), and combinations thereof. In some embodiments, the polypeptide comprises two or more amino acid sequences, wherein each of the at least two or more amino acid sequences is 80% identical to an amino acid sequence within Pfa1p (SEQ ID NO:2 or SEQ ID NO:69) comprising one or more PUFA synthase domains. In some embodiments, the at least two or more amino acid sequences are 80% identical to the same amino acid sequence within Pfa1p (SEQ ID NO:2 or SEQ ID NO:69) comprising one or more PUFA synthase domains. In some embodiments, the at least two or more amino acid sequences are 80% identical to different amino acid sequences within Pfa1p (SEQ ID NO:2 or SEQ ID NO:69) that each comprise one or more PUFA synthase domains. In some embodiments, the at least two or more amino acid sequences are 80% identical to different amino acid sequences within Pfa1p (SEQ ID NO:2 or SEQ ID NO:69), wherein the at least two or more amino acid sequences are located in the same order or a different order in the polypeptide as compared to the order of the corresponding domains within Pfa1p (SEQ ID NO:2 or SEQ ID NO:69). In some embodiments, the at least two or more amino acid sequences are 80% identical to an amino acid sequence within Pfa1p (SEQ ID NO:2 or SEQ ID NO:69) comprising one or more PUFA synthase domains such as a KS domain (SEQ ID NO:8 or SEQ ID NO:75), a MAT domain (SEQ ID NO:10 or SEQ ID NO:77), an ACP domain (such as any one of SEQ ID NOs:14, 16, 18, 20, 22, 24, 81, 83, 85, 87, 89, 91, 93, 95, 97, or 99), a combination of two, three, four, five, six, seven, eight, nine, or ten ACP domains, including tandem domains (SEQ ID NO:12 or SEQ ID NO:79, and portions thereof), a KR domain (SEQ ID NO:26 or SEQ ID NO:101), a DH domain (SEQ ID NO:28 or SEQ ID NO:119), and combinations thereof. In some embodiments, the polypeptide comprises one or more amino acid sequences within Pfa1p (SEQ ID NO:2 or SEQ ID NO:69) comprising one or more PUFA synthase domains, including one or more copies of any individual domain in combination with one or more copies of any other individual domain.

[0160] In some embodiments, the present invention is directed to polypeptides comprising amino acid sequences that are at least 80% identical to an amino acid sequence within Pfa2p (SEQ ID NO:4 or SEQ ID NO:71) comprising one or more PUFA

synthase domains. In some embodiments, the polypeptide comprises an amino acid sequence at least 80% identical to an amino acid sequence within Pfa2p (SEQ ID NO:4 or SEQ ID NO:71) comprising one or more PUFA synthase domains such as a KS domain (SEQ ID NO:30 or SEQ ID NO:103), a CLF domain (SEQ ID NO:32 or SEQ ID NO:105), an AT domain (SEQ ID NO:34 or SEQ ID NO:107), an ER domain (SEQ ID NO:36 or SEQ ID NO:109), and combinations thereof. In some embodiments, the polypeptide comprises two or more amino acid sequences, wherein each of the at least two or more amino acid sequences is 80% identical to an amino acid sequence within Pfa2p (SEQ ID NO:4 or SEQ ID NO:71) comprising one or more PUFA synthase domains. In some embodiments, the at least two or more amino acid sequences are 80% identical to the same amino acid sequence within Pfa2p (SEQ ID NO:4 or SEQ ID NO:71). In some embodiments, the at least two or more amino acid sequences are 80% identical to different amino acid sequences within Pfa2p (SEQ ID NO:4 or SEQ ID NO:71) that each comprise one or more PUFA synthase domains. In some embodiments, the at least two or more amino acid sequences are 80% identical to different amino acid sequences within Pfa2p (SEQ ID NO:4 or SEQ ID NO:71), wherein the at least two or more amino acid sequences are located in the same order or a different order in the polypeptide as compared to the order of the corresponding domains within Pfa2p (SEQ ID NO:4 or SEQ ID NO:71). In some embodiments, the at least two or more amino acid sequences are 80% identical to an amino acid sequence within Pfa2p (SEQ ID NO:4 or SEQ ID NO:71) comprising one or more PUFA synthase domains such as a KS domain (SEQ ID NO:30 or SEQ ID NO:103), a CLF domain (SEQ ID NO:32 or SEQ ID NO:105), an AT domain (SEQ ID NO:24 or SEQ ID NO:107), an ER domain (SEQ ID NO:36 or SEQ ID NO:109), and combinations thereof. In some embodiments, the polypeptide comprises one or more amino acid sequences sequence within Pfa2p (SEQ ID NO:4 or SEQ ID NO:71) comprising one or more PUFA synthase domains, including one or more copies of any individual domain in combination with one or more copies of any other individual domain.

[0161] In some embodiments, the present invention is directed to polypeptides comprising amino acid sequences that are at least 80% identical to an amino acid sequence within Pfa3p (SEQ ID NO:6 or SEQ ID NO:73) comprising one or more PUFA synthase domains. In some embodiments, the polypeptide comprises an amino acid

sequence at least 80% identical to an amino acid sequence within Pfa3p (SEQ ID NO:6 or SEQ ID NO:73) comprising one or more PUFA synthase domains such as a *DH* domain (such as SEQ ID NO:38, SEQ ID NO:40, SEQ ID NO:111, or SEQ ID NO:113), an *ER* domain (SEQ ID NO:42 or SEQ ID NO:115), and combinations thereof. In some embodiments, the polypeptide comprises two or more amino acid sequences, wherein each of the at least two or more amino acid sequences is 80% identical to an amino acid sequence within Pfa3p (SEQ ID NO:6 or SEQ ID NO:73) comprising one or more PUFA synthase domains. In some embodiments, the at least two or more amino acid sequences are 80% identical to the same amino acid sequence within Pfa3p (SEQ ID NO:6 or SEQ ID NO:73) comprising one or more PUFA synthase domains. In some embodiments, the at least two or more amino acid sequences are 80% identical to different amino acid sequences within Pfa3p (SEQ ID NO:6 or SEQ ID NO:73) that each comprise one or more PUFA synthase domains. In some embodiments, the at least two or more amino acid sequences are 80% identical to different amino acid sequences within Pfa3p (SEQ ID NO:6 or SEQ ID NO:73), wherein the at least two or more amino acid sequences are located in the same order or a different order in the polypeptide as compared to the order of the corresponding domains within Pfa3p (SEQ ID NO:6 or SEQ ID NO:73). In some embodiments, the at least two or more amino acid sequences are 80% identical to an amino acid sequence within Pfa3p (SEQ ID NO:6 or SEQ ID NO:73) comprising one or more PUFA synthase domains such as a *DH* domain (such as SEQ ID NO:38, SEQ ID NO:40, SEQ ID NO:111, or SEQ ID NO:113), an *ER* domain (SEQ ID NO:42 or SEQ ID NO:115), and combinations thereof. In some embodiments, the polypeptide comprises one or more amino acid sequences within Pfa3p (SEQ ID NO:6 or SEQ ID NO:73) comprising one or more PUFA synthase domains, including one or more copies of any individual domain in combination with one or more copies of any other individual domain.

[0162] In some embodiments, the present invention is directed to a polypeptide comprising an amino acid sequence at least 80% identical to SEQ ID NO:2 or SEQ ID NO:69, wherein the polypeptide comprises a PUFA synthase activity selected from the group consisting of KS activity, MAT activity, ACP activity, KR activity, DH activity, and combinations thereof.

- [0163] In some embodiments, the present invention is directed to a polypeptide comprising an amino acid sequence at least 80% identical to SEQ ID NO:8 or SEQ ID NO:75, wherein the polypeptide comprises KS activity.
- [0164] In some embodiments, the present invention is directed to a polypeptide comprising an amino acid sequence at least 80% identical to SEQ ID NO:10 or SEQ ID NO:77, wherein the polypeptide comprises MAT activity.
- [0165] In some embodiments, the present invention is directed to a polypeptide comprising an amino acid sequence at least 80% identical to such as any one of SEQ ID NOs:14, 16, 18, 20, 22, 24, 81, 83, 85, 87, 89, 91, 93, 95, 97, or 99, wherein the polypeptide comprises ACP activity.
- [0166] In some embodiments, the present invention is directed to a polypeptide comprising an amino acid sequence at least 80% identical to SEQ ID NO:12 or SEQ ID NO:79, wherein the polypeptide comprises ACP activity.
- [0167] In some embodiments, the present invention is directed to a polypeptide comprising an amino acid sequence at least 80% identical to an amino acid sequence within SEQ ID NO:12, wherein the polypeptide comprises ACP activity. In some embodiments, the amino acid sequence is at least 80% identical to an amino acid sequence within SEQ ID NO:12 comprising one, two, three, four, five, or six ACP domains, wherein the polypeptide comprises ACP activity associated with one or more ACP domains. SEQ ID NOs:14, 16, 18, 20, 22, and 24 are representative amino acid sequences comprising a single ACP domain within SEQ ID NO: 12.
- [0168] In some embodiments, the present invention is directed to a polypeptide comprising an amino acid sequence at least 80% identical to an amino acid sequence within SEQ ID NO:79, wherein the polypeptide comprises ACP activity. In some embodiments, the amino acid sequence is at least 80% identical to an amino acid sequence within SEQ ID NO:79 comprising one, two, three, four, five, six, seven, eight, nine, or ten ACP domains, wherein the polypeptide comprises ACP activity associated with one or more ACP domains. SEQ ID NOs:81, 83, 85, 87, 89, 91, 93, 95, 97, and 99 are representative amino acid sequences comprising a single ACP domain within SEQ ID NO:79.

- [0169] In some embodiments, the present invention is directed to a polypeptide comprising an amino acid sequence at least 80% identical to SEQ ID NO:26 or SEQ ID NO:101, wherein the polypeptide comprises KR activity.
- [0170] In some embodiments, the present invention is directed to a polypeptide comprising an amino acid sequence at least 80% identical to SEQ ID NO:28 or SEQ ID NO:119, wherein the polypeptide comprises DH activity.
- [0171] In some embodiments, the present invention is directed to a polypeptide comprising an amino acid sequence at least 80% identical to SEQ ID NO:4 or SEQ ID NO:71, wherein the polypeptide comprises a PUFA synthase activity selected from the group consisting of KS activity, CLF activity, AT activity, ER activity, and combinations thereof.
- [0172] In some embodiments, the present invention is directed to a polypeptide comprising an amino acid sequence at least 80% identical to SEQ ID NO:30 or SEQ ID NO:103, wherein the polypeptide comprises KS activity.
- [0173] In some embodiments, the present invention is directed to a polypeptide comprising an amino acid sequence at least 80% identical to SEQ ID NO:32 or SEQ ID NO:105, wherein the polypeptide comprises CLF activity.
- [0174] In some embodiments, the present invention is directed to a polypeptide comprising an amino acid sequence at least 80% identical to SEQ ID NO:34 or SEQ ID NO:107, wherein the polypeptide comprises AT activity.
- [0175] In some embodiments, the present invention is directed to a polypeptide comprising an amino acid sequence at least 80% identical to SEQ ID NO:36 or SEQ ID NO:109, wherein the polypeptide comprises ER activity.
- [0176] In some embodiments, the present invention is directed to a polypeptide comprising an amino acid sequence at least 80% identical to SEQ ID NO:6 or SEQ ID NO:73, wherein the polypeptide comprises a PUFA synthase activity selected from the group consisting of DH activity, ER activity, and combinations thereof.
- [0177] In some embodiments, the present invention is directed to a polypeptide comprising an amino acid sequence at least 80% identical to SEQ ID NO:38, wherein the polypeptide comprises DH activity.

- [0178] In some embodiments, the present invention is directed to a polypeptide comprising an amino acid sequence at least 80% identical to SEQ ID NO:40, wherein the polypeptide comprises DH activity.
- [0179] In some embodiments, the present invention is directed to a polypeptide comprising an amino acid sequence at least 80% identical to SEQ ID NO:111, wherein the polypeptide comprises DH activity.
- [0180] In some embodiments, the present invention is directed to a polypeptide comprising an amino acid sequence at least 80% identical to SEQ ID NO:113, wherein the polypeptide comprises DH activity.
- [0181] In some embodiments, the present invention is directed to a polypeptide comprising an amino acid sequence at least 80% identical to SEQ ID NO:42 or SEQ ID NO:115, wherein the polypeptide comprises ER activity.
- [0182] In some embodiments, the polypeptides comprise amino acid sequences at least about 80%, 85%, or 90% identical to the amino acid sequences reported herein, or at least about 95%, 96%, 97%, 98%, 99%, or 100% identical to the amino acid sequences reported herein.
- [0183] By a polypeptide having an amino acid sequence at least, for example, 95% "identical" to a query amino acid sequence of the present invention, it is intended that the amino acid sequence of the subject polypeptide is identical to the query sequence except that the subject polypeptide sequence can include up to five amino acid alterations per each 100 amino acids of the query amino acid sequence. In other words, to obtain a polypeptide having an amino acid sequence at least 95% identical to a query amino acid sequence, up to 5% of the amino acid residues in the subject sequence can be inserted, deleted, (indels) or substituted with another amino acid. These alterations of the reference sequence can occur at the amino or carboxy terminal positions of the reference amino acid sequence or anywhere between those terminal positions, interspersed either individually among residues in the reference sequence or in one or more contiguous groups within the reference sequence.
- [0184] As a practical matter, whether any particular polypeptide having an amino acid sequence that is at least 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% identical to, for instance, the amino acid sequence of the present invention can be determined conventionally using known computer programs. As discussed above, a method for

determining the best overall match between a query sequence (a sequence of the present invention) and a subject sequence can be determined using the alignment of sequences and calculation of identity scores. The alignments were done using the computer program AlignX, which is a component of the Vector NTI Suite 10.0 package from Invitrogen (www.invitrogen.com). The alignments were performed using a ClustalW alignment (J. Thompson *et al.*, *Nucleic Acids Res.* 22(22):4673-4680 (1994). The default scoring matrix Blosum62mt2 was used. The default gap opening penalty is 10 and the gap extension penalty 0.1.

[0185] In further aspects of the invention, nucleic acid molecules having polynucleotide sequences at least 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% identical to the polynucleotide sequences disclosed herein, encode a polypeptide having one or more PUFA synthase activities. Polypeptides having one or more PUFA synthase activities exhibit one or more activities similar to, but not necessarily identical to, one or more activities of a PUFA synthase of the present invention.

[0186] Of course, due to the degeneracy of the genetic code, one of ordinary skill in the art will immediately recognize that a large portion of the nucleic acid molecules having a polynucleotide sequence at least 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identical to the polynucleotide sequences described herein will encode polypeptides "having PUFA synthase functional activity." In fact, since degenerate variants of any of these polynucleotide sequences all encode the same polypeptide, in many instances, it can be predicted by the skilled artisan based on knowledge of conservative substitutions as well as conserved functional domains, which polypeptides will exhibit activity. In certain aspects of the invention, the polypeptides and polynucleotides of the present invention are provided in an isolated form, e.g., purified to homogeneity. Alternatively, the polypeptides and polynucleotides of the invention can be synthetically produced by conventional synthesizers.

[0187] As known in the art "similarity" between two polypeptides is determined by comparing the amino acid sequence and conserved amino acid substitutes thereto of the polypeptide to the sequence of a second polypeptide.

[0188] In some embodiments, a polypeptide of the invention is a fusion polypeptide.

[0189] As used herein, "fusion polypeptide" means a polypeptide comprising a first polypeptide linearly connected, via peptide bonds, to a second polypeptide. The first

polypeptide and the second polypeptide can be identical or different, and they can be directly connected, or connected via a peptide linker. As used herein, the terms "linked," "fused," or "fusion" are used interchangeably. These terms refer to the joining together of two more elements or components by any means including chemical conjugation or recombinant means. An "in-frame fusion" refers to the joining of two or more open reading frames to form a continuous longer open reading frame, in a manner that maintains the correct reading frame of the original open reading frames. Thus, the resulting recombinant fusion protein is a single protein containing two or more segments that correspond to polypeptides encoded by the original open reading frames (which segments are not normally so joined in nature). Although the reading frame is thus made continuous throughout the fused segments, the segments can be physically or spatially separated by, for example, in-frame linker sequence. A "linker" sequence is a series of one or more amino acids separating two polypeptide coding regions in a fusion protein.

[0190] The invention is directed to a composition comprising one or more polypeptides of the invention and a biologically acceptable carrier.

[0191] In some embodiments, the composition includes a biologically acceptable "excipient," wherein the excipient is a component, or mixture of components, that is used in a composition of the present invention to give desirable characteristics to the composition, and also include carriers. "Biologically acceptable" means a compound, material, composition, salt, and/or dosage form which is, within the scope of sound medical judgment, suitable for contact with the tissues of living cells without excessive toxicity, irritation, inflammatory response, or other problematic complications over the desired duration of contact commensurate with a reasonable benefit/risk ratio. Various excipients can be used. In some embodiments, the excipient can be, but is not limited to, an alkaline agent, a stabilizer, an antioxidant, an adhesion agent, a separating agent, a coating agent, an exterior phase component, a controlled-release component, a solvent, a surfactant, a humectant, a buffering agent, a filler, an emollient, or combinations thereof. Excipients in addition to those discussed herein can include excipients listed in, though not limited to, *Remington: The Science and Practice of Pharmacy*, 21st ed. (2005). Inclusion of an excipient in a particular classification herein (e.g., "solvent") is intended to illustrate rather than limit the role of the excipient. A particular excipient can fall within multiple classifications.

[0192] The present invention further relates to a fragment, variant, derivative, or analog of any of the polypeptide disclosed herein.

[0193] The polypeptide of the present invention can be a recombinant polypeptide, a natural polypeptide, or a synthetic polypeptide.

Host Cells

[0194] The present invention is directed to a host cell that expresses any of the nucleic acid molecules and recombinant nucleic acid molecules described above as well as combinations thereof.

[0195] The term "expression" as used herein refers to a process by which a gene produces a biochemical, for example, a RNA or polypeptide. The process includes any manifestation of the functional presence of the gene within the cell including, without limitation, gene knockdown as well as both transient expression and stable expression. It includes, without limitation, transcription of the gene into messenger RNA (mRNA), transfer RNA (tRNA), small hairpin RNA (shRNA), small interfering RNA (siRNA), or any other RNA product, and the translation of such mRNA into polypeptide(s). If the final desired product is biochemical, expression includes the creation of that biochemical and any precursors.

[0196] To produce one or more desired polyunsaturated fatty acids, a host cell can be genetically modified to introduce a PUFA synthase system of the present invention into the host cell.

[0197] When genetically modifying organisms to express a PUFA synthase system according to the present invention, some host organisms can endogenously express accessory proteins that are required in conjunction with a PUFA synthase system in order to produce PUFAs. However, it may be necessary to transform some organisms with nucleic acid molecules encoding one or more accessory protein(s) in order to enable or to enhance production of PUFAs by the organism, even if the organism endogenously produces a homologous accessory protein. Some heterologous accessory proteins can operate more effectively or efficiently with the transformed PUFA synthase proteins than the host cells' endogenous accessory protein(s).

[0198] Accessory proteins are defined herein as proteins that are not considered to be part of the core PUFA synthase system (i.e., not part of the PUFA synthase enzyme complex itself) but which may be necessary for PUFA production or efficient PUFA production

using the core PUFA synthase enzyme complex of the present invention. For example, in order to produce PUFAs, a PUFA synthase system must work with an accessory protein that transfers a 4'-phosphopantetheinyl moiety from coenzyme A to the acyl carrier protein (ACP) domain(s). Therefore, a PUFA synthase system can be considered to include at least one 4'-phosphopantetheinyl transferase (PPTase) domain, or such a domain can be considered to be an accessory domain or protein to the PUFA synthase system. Structural and functional characteristics of PPTases have been described in detail, e.g., in U.S. Appl. Publ. Nos. 2002/0194641; 2004/0235127; and 2005/0100995.

[0199] A domain or protein having 4'-phosphopantetheinyl transferase (PPTase) biological activity (function) is characterized as the enzyme that transfers a 4'-phosphopantetheinyl moiety from Coenzyme A to the acyl carrier protein (ACP). This transfer to an invariant serine residue of the ACP activates the inactive apo-form to the holo-form. In both polyketide and fatty acid synthesis, the phosphopantetheine group forms thioesters with the growing acyl chains. The PPTases are a family of enzymes that have been well characterized in fatty acid synthesis, polyketide synthesis, and non-ribosomal peptide synthesis. The sequences of many PPTases are known, crystal structures have been determined (e.g., Reuter K., *et al.*, *EMBO J.* 18(23):6823-31 (1999)), and mutational analysis has identified amino acid residues important for activity (Mofid M.R., *et al.*, *Biochemistry* 43(14):4128-36 (2004)).

[0200] One heterologous PPTase which has been previously demonstrated to recognize *Schizochytrium* ACP domains as substrates is the Het I protein of *Nostoc sp.* PCC 7120 (formerly called *Anabaena sp.* PCC 7120). Het I is present in a cluster of genes in *Nostoc* known to be responsible for the synthesis of long chain hydroxy-fatty acids that are a component of a glyco-lipid layer present in heterocysts of that organism (Black and Wolk, *J. Bacteriol.* 176: 2282-2292 (1994); Campbell *et al.*, *Arch. Microbiol.* 167: 251-258 (1997)). Het I is likely to activate the ACP domains of a protein, Hgl E, present in that cluster. Sequences and constructs containing Het I have been described in, e.g., U.S. Appl. Publ. No. 2007/0244192, incorporated by reference herein in its entirety.

[0201] Another heterologous PPTase which has been demonstrated previously to recognize the *Schizochytrium* ACP domains is Sfp, derived from *Bacillus subtilis*. Sfp has been well characterized and is widely used due to its ability to recognize a broad range of substrates. Based on published sequence information (Nakana, *et al.*, *Molecular*

and *General Genetics* 232: 313-321 (1992)), an expression vector was previously produced for Sfp by cloning the coding region, along with defined up- and downstream flanking DNA sequences, into a pACYC-184 cloning vector. This construct encodes a functional PPTase as demonstrated by its ability to be co-expressed with *Schizochytrium* Orfs in *E. coli* which, under appropriate conditions, resulted in the accumulation of DHA in those cells (see, U.S. Appl. Publ. No. 2004/0235127, incorporated by reference herein in its entirety).

[0202] Host cells can include microbial cells; animal cells; plant cells; and insect cells. Representative examples of appropriate hosts include bacterial cells; thermophilic or mesophilic bacteria; marine bacteria; thraustochytrids; fungal cells, such as yeast; plant cells; insect cells; and isolated animal cells. Host cells can be either untransfected cells or cells that are already transfected with at least one other recombinant nucleic acid molecule. Host cells can also include transgenic cells that have been engineered to express a PUFA synthase. The selection of an appropriate host is deemed to be within the scope of those skilled in the art from the teachings herein.

[0203] Host cells include any microorganism of the order *Thraustochytriales*, such as microorganisms from a genus including, but not limited to: *Thraustochytrium*, *Labyrinthuloides*, *Japonochytrium*, and *Schizochytrium*. Species within these genera include, but are not limited to: any *Schizochytrium* species, including *Schizochytrium aggregatum*, *Schizochytrium limacinum*, *Schizochytrium minutum*; any *Thraustochytrium* species (including former *Ulkenia* species such as *U. visurgensis*, *U. amoeboida*, *U. sarkariana*, *U. profunda*, *U. radiata*, *U. minuta* and *Ulkenia* sp. BP-5601), and including *Thraustochytrium striatum*, *Thraustochytrium aureum*, *Thraustochytrium roseum*; and any *Japonochytrium* species. Strains of *Thraustochytriales* include, but are not limited to: *Schizochytrium* sp. (S31) (ATCC 20888); *Schizochytrium* sp. (S8) (ATCC 20889); *Schizochytrium* sp. (LC-RM) (ATCC 18915); *Schizochytrium* sp. (SR21); *Schizochytrium aggregatum* (Goldstein et Belsky) (ATCC 28209); *Schizochytrium limacinum* (Honda et Yokochi) (IFO 32693); *Thraustochytrium* sp. (23B) (ATCC 20891); *Thraustochytrium striatum* (Schneider) (ATCC 24473); *Thraustochytrium aureum* (Goldstein) (ATCC 34304); *Thraustochytrium roseum* (Goldstein) (ATCC 28210); and *Japonochytrium* sp. (L1) (ATCC 28207). Other examples of suitable host microorganisms for genetic modification include, but are not limited to, yeast including *Saccharomyces cerevisiae*,

Saccharomyces carlsbergensis, or other yeast such as *Candida*, *Kluyveromyces*, or other fungi, for example, filamentous fungi such as *Aspergillus*, *Neurospora*, *Penicillium*, etc. Bacterial cells also can be used as hosts. This includes *Escherichia coli*, which can be useful in fermentation processes. Alternatively, a host such as a *Lactobacillus* species or *Bacillus* species can be used as a host.

[0204] Plant host cells include, but are not limited to, any higher plants, including both dicotyledonous and monocotyledonous plants, and consumable plants, including crop plants and plants used for their oils. Such plants can include, for example: canola, soybeans, rapeseed, linseed, corn, safflowers, sunflowers, and tobacco. Other plants include those plants that are known to produce compounds used as pharmaceutical agents, flavoring agents, nutraceutical agents, functional food ingredients, cosmetically active agents, or plants that are genetically engineered to produce these compounds/agents. Thus, any plant species or plant cell can be selected. Examples of plants and plant cells, and plants grown or derived therefrom, include, but are not limited to, plants and plant cells obtainable from canola (*Brassica rapa* L.); canola cultivars NQC02CNX12 (ATCC PTA-6011), NQC02CNX21 (ATCC PTA-6644), and NQC02CNX25 (ATCC PTA-6012) as well as cultivars, breeding cultivars, and plant parts derived from canola cultivars NQC02CNX12, NQC02CNX21, and NQC02CNX25 (see, U.S. Patent Nos. 7,355,100, 7,456,340, and 7,348,473, respectively); soybean (*Glycine max*); rapeseed (*Brassica* spp.); linseed/flax (*Linum usitatissimum*); maize (corn) (*Zea mays*); safflower (*Carthamus tinctorius*); sunflower (*Helianthus annuus*); tobacco (*Nicotiana tabacum*); *Arabidopsis thaliana*, Brazil nut (*Betholettia excelsa*); castor bean (*Ricinus communis*); coconut (*Cocus nucifera*); coriander (*Coriandrum sativum*); cotton (*Gossypium* spp.); groundnut (*Arachis hypogaea*); jojoba (*Simmondsia chinensis*); mustard (*Brassica* spp. and *Sinapis alba*); oil palm (*Elaeis guineensis*); olive (*Olea europaea*); rice (*Oryza sativa*); squash (*Cucurbita maxima*); barley (*Hordeum vulgare*); wheat (*Triticum aestivum*); and duckweed (*Lemnaceae* sp.). Plant lines from these and other plants can be produced, selected, or optimized for a desirable trait such as or associated with, but not limited to, seed yield, lodging resistance, emergence, disease resistance or tolerance, maturity, late season plant intactness, plant height, shattering resistance, ease of plant transformation, oil content, or oil profile. Plant lines can be selected through plant breeding such as pedigree breeding, recurrent selection breeding, intercross and backcross breeding, as

well as methods such as marker assisted breeding and tilling. *See*, e.g., U.S. Patent No. 7,348,473.

[0205] Animal cells include any isolated animal cells.

[0206] The present invention is directed to a host cell that expresses one or more nucleic acid molecules or recombinant nucleic acid molecules, including vectors, of the invention.

[0207] The present invention is directed to a method for making a recombinant host cell comprising introducing a recombinant vector into a host cell.

[0208] Host cells can be genetically engineered (transduced or transformed or transfected) with the vectors of this invention that can be, for example, a cloning vector or an expression vector. The vector can be, for example, in the form of a plasmid, a viral particle, a phage, etc. The vector containing a polynucleotide sequence as described herein, as well as an appropriate promoter or control sequence, can be employed to transform an appropriate host to permit expression of the polypeptide encoded by the polynucleotide sequence. The genetic modification of host cells can also include the optimization of genes for preferred or optimal host codon usage.

[0209] The engineered host cells can be cultured in conventional nutrient media modified as appropriate for activating promoters, selecting transformants, or amplifying the genes of the present invention. The culture conditions, such as temperature, pH, and the like, are those previously used with the host cell selected for expression, and will be apparent to the ordinarily skilled artisan.

[0210] In some embodiments, the present invention is directed to genetically modifying a plant or part of a plant to express a PUFA synthase system described herein, which includes at least the core PUFA synthase enzyme complex. A "part of a plant" or "plant part" as defined herein includes any part of a plant, such as, but not limited to, seeds (immature or mature), oils, pollen, embryos, flowers, fruits, shoots, leaves, roots, stems, explants, etc. In some embodiments, the genetically modified plant or part of a plant produces one or more PUFAs, such as EPA, DHA, DPA (n-3 or n-6), ARA, GLA, SDA, other PUFAs, and combinations thereof. Plants are not known to endogenously contain a PUFA synthase system; therefore, the PUFA synthase systems of the present invention can be used to engineer plants with unique fatty acid production capabilities. In a further embodiment, the plant or part of a plant is further genetically modified to express at least

one PUFA synthase accessory protein, (e.g., a PPTase). In some embodiments, the plant is an oil seed plant, wherein the oil seeds, and/or the oil in the oil seeds, contain PUFAs produced by the PUFA synthase system. In some embodiments, the genetically modified plants, parts of plants, oil seeds, and/or oils in the oil seeds contain a detectable amount of at least one PUFA that is the product of the PUFA synthase system. In further embodiments, such plants, parts of plants, oil seeds, and/or oils in the oil seeds can be substantially free of intermediate or side products that are not the primary PUFA products of the introduced PUFA synthase system and that are not naturally produced by the endogenous FAS system in the wild-type plants. While wild-type plants produce some short or medium chain PUFAs, such as 18 carbon PUFAs via the FAS system, new or additional PUFAs will be produced in the plant, parts of plants, oil seeds, and/or oils in the oil seeds as a result of genetic modification with a PUFA synthase system described herein.

[0211] Genetic modification of a plant can be accomplished using classical strain development and/or molecular genetic techniques. See, U.S. Appl. Publ. No. 2007/0244192. Methods for producing a transgenic plant, wherein a recombinant nucleic acid molecule encoding a desired amino acid sequence is incorporated into the genome of the plant, are known in the art. For example, viral vectors can be used to produce transgenic plants, such as by transformation of a monocotyledonous plant with a viral vector using the methods described in U.S. Pat. Nos. 5,569,597; 5,589,367; and 5,316,931. Methods for the genetic engineering or modification of plants by transformation are also well known in the art, including biological and physical transformation protocols. See, e.g., B.L. Miki *et al.*, *Procedures for Introducing Foreign DNA into Plants*, in *METHODS IN PLANT MOLECULAR BIOLOGY AND BIOTECHNOLOGY* 67-88 (Glick, B. R. and Thompson, J. E. eds., CRC Press, Inc., Boca Raton, 1993). In addition, vectors and *in vitro* culture methods for plant cell or tissue transformation and regeneration of plants are available. See, e.g., M. Y. Gruber *et al.*, *Vectors for Plant Transformation*, in *METHODS IN PLANT MOLECULAR BIOLOGY AND BIOTECHNOLOGY* 89-119 (Glick, B. R. and Thompson, J. E. eds., CRC Press, Inc., Boca Raton, 1993).

[0212] A widely utilized method for introducing an expression vector into plants is based on the natural transformation system of *Agrobacterium*. See, e.g., Horsch *et al.*, *Science* 227:1229 (1985) and U.S. Patent No. 6,051,757. *A. tumefaciens* and *A. rhizogenes* are

plant pathogenic soil bacteria which genetically transform plant cells. The Ti and Ri plasmids of *A. tumefaciens* and *A. rhizogenes*, respectively, carry genes responsible for genetic transformation of the plant. See, e.g., Kado, C. I., *Crit. Rev. Plant. Sci.* 10:1 (1991). Descriptions of Agrobacterium vector systems and methods for Agrobacterium-mediated gene transfer are provided by numerous references, including Gruber et al., supra; Miki et al., supra; Moloney et al., *Plant Cell Reports* 8:238 (1989); U.S. Pat. Nos. 5,177,010; 5,104,310; 5,149,645; 5,469,976; 5,464,763; 4,940,838; 4,693,976; 5,591,616; 5,231,019; 5,463,174; 4,762,785; 5,004,863; and 5,159,135; and European Patent Appl. Nos. 0131624, 120516, 159418, 176112, 116718, 290799, 320500, 604662, 627752, 0267159, and 0292435.

[0213] Other methods of plant transformation include microprojectile-mediated transformation, wherein DNA is carried on the surface of microprojectiles. The expression vector is introduced into plant tissues with a biolistic device that accelerates the microprojectiles to speeds sufficient to penetrate plant cell walls and membranes. See, e.g., Sanford et al., *Part. Sci. Technol.* 5:27 (1987), Sanford, J. C., *Trends Biotech.* 6:299 (1988), Sanford, J. C., *Physiol. Plant* 79:206 (1990), Klein et al., *Biotechnology* 10:268 (1992), and U.S. Patent Nos. 5,015,580 and 5,322,783. Techniques for accelerating genetic material coated onto microparticles directed into cells is also described, e.g., in U.S. Patent Nos. 4,945,050 and 5,141,141. Another method for physical delivery of DNA to plants is sonication of target cells. See, e.g., Zhang et al., *Bio/Technology* 9:996 (1991). Alternatively, liposome or spheroplast fusion have been used to introduce expression vectors into plants. See, e.g., Deshayes et al., *EMBO J.*, 4:2731 (1985), Christou et al., *Proc Natl. Acad. Sci. USA* 84:3962 (1987). Direct uptake of DNA into protoplasts using CaCl_2 precipitation, DNA injection, polyvinyl alcohol or poly-L-ornithine have also been reported. See, e.g., Hain et al., *Mol. Gen. Genet.* 199:161 (1985) and Draper et al., *Plant Cell Physiol.* 23:451 (1982). Electroporation of protoplasts and whole cells and tissues has also been described. See, e.g., Donn et al., in Abstracts of VIIth International Congress on Plant Cell and Tissue Culture IAPTC, A2-38, p. 53 (1990); D'Halluin et al., *Plant Cell* 4:1495-1505 (1992); Spencer et al., *Plant Mol. Biol.* 24:51-61 (1994); International Appl. Publ. Nos. WO 87/06614, WO 92/09696, and WO 93/21335; and U.S. Patent Nos. 5,472,869 and 5,384,253. Other transformation

technology includes whiskers technology, see, e.g., U.S. Patent Nos. 5,302,523 and 5,464,765.

[0214] Chloroplasts or plastids can also be directly transformed. As such, recombinant plants can be produced in which only the chloroplast or plastid DNA has been modified with any of the nucleic acid molecules and recombinant nucleic acid molecules described above as well as combinations thereof. Promoters which function in chloroplasts and plastids are known in the art. See, e.g., Hanley-Bowden *et al.*, *Trends in Biochemical Sciences* 12:67-70 (1987). Methods and compositions for obtaining cells containing chloroplasts into which heterologous DNA has been inserted have been described, e.g., in U.S. Patent Nos. 5,693,507 and 5,451,513.

[0215] Any other methods which provide for efficient transformation can also be employed.

[0216] Vectors suitable for use in plant transformation are known in the art. See, e.g., U.S. Patent Nos. 6,495,738; 7,271,315; 7,348,473; 7,355,100; 7,456,340; and references disclosed therein.

[0217] Expression vectors can include at least one genetic marker, operably linked to a regulatory element (a promoter, for example) that allows transformed cells containing the marker to be either recovered by negative selection, i.e., inhibiting growth of cells that do not contain the selectable marker gene, or by positive selection, i.e., screening for the product encoded by the genetic marker. Many commonly used selectable marker genes for plant transformation are well known in the transformation arts, and include, for example, genes that code for enzymes that metabolically detoxify a selective chemical agent which can be an antibiotic or an herbicide, or genes that encode an altered target which is insensitive to the inhibitor. Selectable markers suitable for use in plant transformation include, but are not limited to, the aminoglycoside phosphotransferase gene of transposon Tn5 (Aph II) which encodes resistance to the antibiotics kanamycin, neomycin, and G418, as well as those genes which encode for resistance or tolerance to glyphosate, hygromycin, methotrexate, phosphinothricin (bialophos), imidazolinones, sulfonylureas and triazolopyrimidine herbicides, such as chlorsulfuron, bromoxynil, dalapon, and the like. One commonly used selectable marker gene for plant transformation is the neomycin phosphotransferase II (nptII) gene under the control of plant regulatory signals which confers resistance to kanamycin. See, e.g., Fraley *et al.*,

Proc. Natl. Acad. Sci. U.S.A. 80: 4803 (1983). Another commonly used selectable marker gene is the hygromycin phosphotransferase gene which confers resistance to the antibiotic hygromycin. See, e.g., Vanden Elzen *et al.*, *Plant Mol. Biol.* 5:299 (1985). Additional selectable marker genes of bacterial origin that confer resistance to antibiotics include gentamycin acetyl transferase, streptomycin phosphotransferase, aminoglycoside-3'-adenyl transferase, and the bleomycin resistance determinant. See, e.g., Hayford *et al.*, *Plant Physiol.* 86:1216 (1988), Jones *et al.*, *Mol. Gen. Genet.* 210: 86 (1987), Svab *et al.*, *Plant Mol. Biol.* 14:197 (1990), Hille *et al.*, *Plant Mol. Biol.* 7:171 (1986). Other selectable marker genes confer resistance to herbicides such as glyphosate, glufosinate, or bromoxynil. See, e.g., Comai *et al.*, *Nature* 317:741-744 (1985), Gordon-Kamm *et al.*, *Plant Cell* 2:603-618 (1990) and Stalker *et al.*, *Science* 242:419-423 (1988). Other selectable marker genes for plant transformation are not of bacterial origin. These genes include, for example, mouse dihydrofolate reductase, plant 5-enolpyruvylshikimate-3-phosphate synthase and plant acetolactate synthase. See, e.g., Eichholtz *et al.*, *Somatic Cell Mol. Genet.* 13:67 (1987), Shah *et al.*, *Science* 233:478 (1986), Charest *et al.*, *Plant Cell Rep.* 8:643 (1990).

[0218] A reporter gene can be used with or without a selectable marker. Reporter genes are genes which are typically not present in the recipient organism or tissue and typically encode for proteins resulting in some phenotypic change or enzymatic property. See, e.g., K. Weising *et al.*, *Ann. Rev. Genetics* 22: 421 (1988). Reporter genes include, but are not limited to beta-glucuronidase (GUS), beta-galactosidase, chloramphenicol acetyltransferase, green fluorescent protein, and luciferase genes. See, e.g., Jefferson, R. A., *Plant Mol. Biol. Rep.* 5:387 (1987), Teeri *et al.*, *EMBO J.* 8:343 (1989), Koncz *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* 84:131 (1987), DeBlock *et al.*, *EMBO J.* 3:1681 (1984), and Chalfie *et al.*, *Science* 263:802 (1994). An assay for detecting reporter gene expression can be performed at a suitable time after the gene has been introduced into recipient cells. One such assay entails the use of the gene encoding beta-glucuronidase (GUS) of the uidA locus of *E. coli* as described by Jefferson *et al.*, *Biochem. Soc. Trans.* 15: 17-19 (1987).

[0219] Promoter regulatory elements from a variety of sources can be used efficiently in plant cells to express foreign genes. For example, promoter regulatory elements of bacterial origin, such as the octopine synthase promoter, the nopaline synthase promoter, the mannopine synthase promoter, as well as promoters of viral origin, such as the

cauliflower mosaic virus (35S and 19S), 35T (which is a re-engineered 35S promoter, *see* International Appl. Publ. No. WO 97/13402) can be used. Plant promoter regulatory elements include but are not limited to ribulose-1,6-bisphosphate (RUBP) carboxylase small subunit (ssu), beta-conglycinin promoter, beta-phaseolin promoter, ADH promoter, heat-shock promoters, and tissue specific promoters. Matrix attachment regions, scaffold attachment regions, introns, enhancers, and polyadenylation sequences can also be used to improve transcription efficiency or DNA integration. Such elements can be included to obtain optimal performance of the transformed DNA in the plant. Typical elements include, but are not limited to, Adh-intron 1, Adh-intron 6, the alfalfa mosaic virus coat protein leader sequence, the maize streak virus coat protein leader sequence, as well as others available to a skilled artisan. Constitutive promoter regulatory elements can also be used to direct continuous gene expression. Constitutive promoters include, but are not limited to, promoters from plant viruses such as the 35S promoter from CaMV (Odell *et al.*, *Nature* 313:810-812 (1985)), and promoters from such genes as rice actin (McElroy *et al.*, *Plant Cell* 2:163-171 (1990)), ubiquitin (Christensen *et al.*, *Plant Mol. Biol.* 12:619-632 (1989) and Christensen *et al.*, *Plant Mol. Biol.* 18:675-689 (1992)), pEMU (Last *et al.*, *Theor. Appl. Genet.* 81:581-588 (1991)), MAS (Velten *et al.*, *EMBO J.* 3:2723-2730 (1984)), maize H3 histone (Lepetit *et al.*, *Mol. Gen. Genetics* 231:276-285 (1992) and Atanassova *et al.*, *Plant Journal* 2(3): 291-300 (1992)), and the ALS promoter, Xba1/NcoI fragment 5' to the *Brassica napus* ALS3 structural gene (or a nucleotide sequence similar to the Xba1/NcoI fragment) (International Appl. Publ. No. WO 96/30530). Tissue-specific promoter regulatory elements can also be used for gene expression in specific cell or tissue types, such as leaves or seeds (e.g., zein, oleosin, napin, ACP, globulin, and the like). Tissue-specific or tissue-preferred promoters include, but are not limited to, a root-preferred promoter, such as from the phaseolin gene (Murai *et al.*, *Science* 23:476-482 (1983) and Sengupta-Gopalan *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* 82:3320-3324 (1985)); a leaf-specific and light-induced promoter such as from cab or rubisco (Simpson *et al.*, *EMBO J.* 4(11):2723-2729 (1985) and Timko *et al.*, *Nature* 318:579-582 (1985)); an anther-specific promoter such as from LAT52 (Twell *et al.*, *Mol. Gen. Genetics* 217:240-245 (1989)); a pollen-specific promoter such as from Zm13 (Guerrero *et al.*, *Mol. Gen. Genetics* 244:161-168 (1993)); or a microspore-preferred promoter such as from apg (Twell *et al.*, *Sex. Plant Reprod.* 6:217-224 (1993)).

Promoter regulatory elements can also be active during a certain stage of a plants' development as well as plant tissues and organs, including, but not limited to, pollen-specific, embryo specific, corn silk specific, cotton fiber specific, root specific, and seed endosperm specific promoter regulatory elements. An inducible promoter regulatory element can be used, which is responsible for expression of genes in response to a specific signal, such as: physical stimulus (heat shock genes); light (RUBP carboxylase); hormone (Em); metabolites; chemicals; and stress. Inducible promoters include, but are not limited to, a promoter from the ACEI system which responds to copper (Mett *et al.*, *PNAS* 90:4567-4571 (1993)); from the In2 gene from maize which responds to benzenesulfonamide herbicide safeners (Hershey *et al.*, *Mol. Gen. Genetics* 227:229-237 (1991) and Gatz *et al.*, *Mol. Gen. Genetics* 243:32-38 (1994)), from the Tet repressor from Tn10 (Gatz *et al.*, *Mol. Gen. Genetics* 227:229-237 (1991)); and from a steroid hormone gene, the transcriptional activity of which is induced by a glucocorticosteroid hormone (Sчена *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* 88:0421 (1991)).

[0220] Signal sequences can also be used to direct a polypeptide to either an intracellular organelle or subcellular compartment or for secretion to the apoplast. *See*, e.g., Becker *et al.*, *Plant Mol. Biol.* 20:49 (1992), Knox, C., *et al.*, *Plant Mol. Biol.* 9:3-17 (1987), Lerner *et al.*, *Plant Physiol.* 91:124-129 (1989), Fontes *et al.*, *Plant Cell* 3:483-496 (1991), Matsuoka *et al.*, *Proc. Natl. Acad. Sci.* 88:834 (1991), Gould *et al.*, *J. Cell. Biol.* 108:1657 (1989), Creissen *et al.*, *Plant J.* 2:129 (1991), Kalderon, *et al.*, *Cell* 39:499-509 (1984), and Steifel *et al.*, *Plant Cell* 2:785-793 (1990). Such targeting sequences provide for the desired expressed protein to be transferred to the cell structure in which it most effectively functions or to areas of the cell in which cellular processes necessary for desired phenotypic functions are concentrated.

[0221] In some embodiments, signal sequences are used to direct proteins of the invention to a subcellular compartment, for example, to the plastid or chloroplast. Gene products, including heterologous gene products, can be targeted to the plastid or chloroplast by fusing the gene product to a signal sequence which is cleaved during chloroplast import yielding the mature protein. *See*, e.g., Comai *et al.*, *J. Biol. Chem.* 263: 15104-15109 (1988) and van den Broeck *et al.*, *Nature* 313: 358-363 (1985). DNA encoding for appropriate signal sequences can be isolated from cDNAs encoding the RUBISCO protein, the CAB protein, the EPSP synthase enzyme, the GS2 protein, or

from any naturally occurring chloroplast targeted protein that contains a signal sequence (also termed a chloroplast transit peptide (CTP)) that directs the targeted protein to the chloroplast. Such chloroplast targeted proteins are well known in the art. The chloroplast targeted proteins are synthesized as larger precursor proteins that contain an amino-terminal CTP, which directs the precursor to the chloroplast import machinery. CTPs are generally cleaved by specific endoproteases located within the chloroplast organelle, thus releasing the targeted mature protein, including active proteins such as enzymes, from the precursor into the chloroplast milieu. Examples of sequences encoding peptides suitable for targeting a gene or gene product to the chloroplast or plastid of the plant cell include the petunia EPSPS CTP, the *Arabidopsis* EPSPS CTP2 and intron, and other sequences known in the art. Specific examples of CTPs include, but are not limited to, the *Arabidopsis thaliana* ribulose biphosphate carboxylase small subunit *at*s1A transit peptide, an *Arabidopsis thaliana* EPSPS transit peptide, and a *Zea mays* ribulose biphosphate carboxylase small subunit transit peptide. An optimized transit peptide is described, e.g., by Van den Broeck *et al.*, *Nature* 313:358-363 (1985). Prokaryotic and eukaryotic signal sequences are disclosed, e.g., by Michaelis *et al.*, *Ann. Rev. Microbiol.* 36: 425 (1982). Additional examples of transit peptides that can be used in the invention include chloroplast transit peptides described in Von Heijne *et al.*, *Plant Mol. Biol. Rep.* 9:104-126(1991); Mazur *et al.*, *Plant Physiol.* 85: 1110 (1987); Vorst *et al.*, *Gene* 65: 59 (1988); Chen & Jagendorf, *J. Biol. Chem.* 268: 2363-2367 (1993); a transit peptide from the *rbcS* gene from *Nicotiana glauca* (Poulsen *et al.*, *Mol. Gen. Genet.* 205: 193-200 (1986)); and a transit peptide derived from *Brassica napus* acyl-ACP thioesterase (Loader *et al.*, *Plant Mol. Biol.* 23: 769-778 (1993); Loader *et al.*, *Plant Physiol.* 110:336-336 (1995).

[0222] Genetically modified plants of the invention can be further modified to delete or inactivate an endogenous fatty acid synthase, to reduce endogenous competition with the exogenous PUFA synthase system for malonyl CoA, to increase the level of malonyl CoA in the organism, and combinations thereof. See, e.g., U.S. Appl. Publ. No. 2007/0245431.

[0223] A genetically modified plant can be cultured in a fermentation medium or grown in a suitable medium such as soil. A suitable growth medium for higher plants includes any growth medium for plants, such as, but not limited to, soil, sand, any other particulate media that support root growth (e.g. vermiculite, perlite, etc.) or hydroponic culture as

well as suitable light, water, and nutritional supplements which optimize the growth of the higher plant. PUFAs can be recovered from the genetically modified plants through purification processes which extract the compounds from the plant. PUFAs can be recovered by harvesting the plant as well as by harvesting the oil from the plant (e.g., from the oil seeds). The plant can also be consumed in its natural state or further processed into consumable products. In some embodiments, the present invention is directed to a genetically modified plant, wherein the plant produces at least one PUFA as a result of the genetic modification, and wherein the total fatty acid profile in the plant, or the part of the plant that accumulates PUFAs, comprises a detectable amount of the PUFA produced as a result of genetic modification of the plant. In some embodiments, the plant is an oil seed plant. In some embodiments, the oil seed plant produces PUFAs in its mature seeds or contains the PUFAs in the oil of its seeds.

[0224] Various mammalian cell culture systems can also be employed to express recombinant protein. Expression vectors will comprise an origin of replication, a suitable promoter and enhancer, and also any necessary ribosome binding sites, polyadenylation site, splice donor and acceptor sites, transcriptional termination sequences, and 5' flanking nontranscribed sequences.

Methods Involving Heterologous Expression

[0225] The present invention is directed to a method to produce at least one PUFA comprising expressing a PUFA synthase system in a host cell under conditions effective to produce PUFA, wherein the PUFA synthase system comprises any of the isolated nucleic acid molecules and recombinant nucleic acid molecules described herein as well as combinations thereof, wherein at least one PUFA is produced. In some embodiments, the at least one PUFA includes DHA, EPA, or a combination thereof. In some embodiments, the host cell is a plant cell, an isolated animal cell, or a microbial cell. In some embodiments the host cell is a thraustochytrid.

[0226] The present invention is directed to a method to produce lipids enriched for DHA, EPA, or a combination thereof, comprising expressing a PUFA synthase gene in a host cell under conditions effective to produce lipids, wherein the PUFA synthase gene comprises any of the isolated nucleic acid molecules and recombinant nucleic acid molecules described herein as well as combinations thereof in the host cell, wherein lipids enriched with DHA, EPA, or a combination thereof are produced.

[0227] The invention is directed to a method of isolating lipids from a host cell, comprising expressing a PUFA synthase gene in the host cell under conditions effective to produce lipids, and isolating lipids from the host cell, wherein the PUFA synthase system in the host cell comprises any of the isolated nucleic acid molecules and recombinant nucleic acid molecules described herein as well as combinations thereof.

[0228] In some embodiments, one or more lipid fractions containing PUFAs are isolated from the host cells. In some embodiments, the one or more fractions isolated from the host cell includes the total fatty acid fraction, the sterol esters fraction, the triglyceride fraction, the free fatty acid fraction, the sterol fraction, the diglycerol fraction, the phospholipid fraction, or combination thereof. In some embodiments, PUFAs are isolated from the host cells, wherein the PUFAs are enriched for omega-3 fatty acids omega-6 fatty acids, or combinations thereof based on the composition of the PUFA synthase system introduced into a host cell. In some embodiments, the PUFAs are enriched for DHA, EPA, DPA n-6, ARA, or combinations thereof based on the composition of the PUFA synthase system introduced into a host cell. In some embodiments, the PUFAs are enriched for DHA, EPA, or a combination thereof. In some embodiments, the PUFA profile of PUFAs isolated from a host cell include high concentrations of DHA and lower concentrations of EPA, ARA, DPA n-6, or combinations thereof. In some embodiments, the PUFA profile of PUFAs isolated from a host cell include high concentrations of DHA and EPA, and lower concentrations of ARA, DPA n-6, or combinations thereof. In some embodiments, the PUFA profile of PUFAs isolated from a host cell include high concentrations of EPA and lower concentrations of DHA, ARA, DPA n-6, or combinations thereof.

[0229] The invention is directed to a method of replacing an inactive or deleted PUFA synthase activity, introducing a new PUFA synthase activity, or enhancing an existing PUFA synthase activity in an organism having PUFA synthase activity, comprising expressing any of the isolated nucleic acid molecules and recombinant nucleic acid molecules described herein as well as combinations thereof in the organism under conditions effective to express the PUFA synthase activity. In some embodiments, the nucleic acid molecule comprises one or more *PFA1*, *PFA2*, or *PFA3* PUFA synthase polynucleotide sequences described herein that encode one or more PUFA synthase domains. In some embodiments, the PUFA profiles of the organisms are altered by the

introduction of the one or more nucleic acid molecules of the invention. In some embodiments, the altered PUFA profiles include an increase in omega-3 fatty acids and a decrease in omega-6 fatty acids. In some embodiments, the altered PUFA profiles include an increase in omega-6 fatty acids and a decrease in omega-3 fatty acids. In some embodiments, both omega-3 and omega-6 fatty acids are increased. In some embodiments, the amount of DHA is increased while the amounts of one or more of EPA, ARA, DPA n-6, or combinations thereof are maintained or decrease. In some embodiments, the amounts of EPA and DHA are increased while the amounts of ARA, DPA n-6, or a combination thereof are maintained or decrease. In some embodiments, the amount of EPA is increased while the amounts of one or more of EPA, ARA, DPA n-6, or combinations thereof are maintained or decrease. In some embodiments, the nucleic acid molecule comprises the polynucleotide sequence of *PFA3* or one or more domains therein. In some embodiments, the nucleic acid molecule comprises the polynucleotide sequence of *PFA3* or one or more domains therein and the amount of omega-3 fatty acids in the organism is increased while the amount of omega-6 fatty acids is decreased. In some embodiments, the nucleic acid molecule comprises the polynucleotide sequence of *PFA2* or one or more domains therein and the amount of DHA in the organism is increased while the amount of EPA is decreased.

[0230] The invention is directed to methods of increasing production of DHA, EPA, or a combination thereof in an organism having PUFA synthase activity, comprising expressing any of the isolated nucleic acid molecules and recombinant nucleic acid molecules described herein as well as combinations thereof in the organism under conditions effective to produce DHA, EPA, or a combination thereof, wherein the PUFA synthase activity replaces an inactive or deleted activity, introduces a new activity, or enhances an existing activity in the organism, and wherein production of DHA, EPA, or a combination thereof in the organism is increased.

[0231] Having generally described this invention, a further understanding can be obtained by reference to the examples provided herein. These examples are for purposes of illustration only and are not intended to be limiting.

EXAMPLE 1

- [0232] Degenerate primers for the KS and DH PUFA synthase domains were designed in order to isolate the corresponding sequences from the isolated microorganism deposited under ATCC Accession No. PTA-9695, also known as *Schizochytrium sp.* ATCC PTA-9695.
- [0233] Degenerate primers for the KS region of *Schizochytrium sp.* ATCC PTA-9695 *PFA1* (i.e., the region containing the KS domain) were designed based on the published *PFA1* (previously termed *orfA* or ORF 1) sequences for *Shewanella japonica*, *Schizochytrium sp.* ATCC 20888, *Thraustochytrium aureum* (ATCC 34304), and *Thraustochytrium sp.* 23B ATCC 20892:
- prDS173 (forward): GATCTACTGCAAGCGCGGNGGNTTYAT (SEQ ID NO:62), and
- prDS174 (reverse): GGCGCAGGCGGCRTCNACNAC (SEQ ID NO:63).
- [0234] Degenerate primers for the DH region of *Schizochytrium sp.* ATCC PTA-9695 *PFA3* (previously termed *orfC* or ORF 3) were designed based on the published sequences for *Moritella marina*; *Schizochytrium sp.* ATCC 20888; *Shewanella sp.* SCRC-2738; *Photobacter profundum*; and *Thraustochytrium sp.* 23B ATCC 20892:
- JGM190 (forward): CAYTGGTAYTTYCCNTGYCAYTT (SEQ ID NO:64); and
- BLR242 (reverse): CCNGGCATNACNGGRTC (SEQ ID NO:65).
- [0235] The PCR conditions with chromosomal DNA template were as follows: 0.2 μ M dNTPs, 0.1 μ M each primer, 8% DMSO, 200 ng chromosomal DNA, 2.5 U Herculase® II fusion polymerase (Stratagene), and 1X Herculase® buffer (Stratagene) in a 50 μ L total volume. The PCR Protocol included the following steps: (1) 98°C for 3 minutes; (2) 98°C for 30 seconds; (3) 50°C for 30 seconds; (4) 72°C for 2 minutes; (5) repeat steps 2-4 for 40 cycles; (6) 72°C for 5 minutes; and (7) hold at 6°C.
- [0236] For both primer pairs, PCR yielded distinct DNA products with the expected sizes using chromosomal templates from *Schizochytrium sp.* ATCC Accession No. PTA-9695. The respective PCR products were cloned into the vector pJET1.2/blunt (Fermentas) according to the manufacturer's instructions, and the insert sequence was determined using supplied standard primers.
- [0237] The DNA sequences obtained from the PCR products were compared with known sequences available from the NCBI GenBank in a standard BLASTx search (BLASTx

parameters: Low complexity filter on; Matrix: BLOSUM62; Gap cost; Existence 11, Extension 1. Stephen F. Altschul, Thomas L. Madden, Alejandro A. Schäffer, Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman (1997), "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs", *Nucleic Acids Res.* 25:3389-3402.).

[0238] At the amino acid level, the sequences with the highest level of homology to deduced amino acid sequence derived from the cloned DNA containing the KS fragment from *Schizochytrium sp.* ATCC PTA-9695 were: *Schizochytrium sp.* ATCC 20888 "polyunsaturated fatty acid synthase subunit A" (Identity = 87%; positives = 92%); *Shewanella oneidensis* MR-1 "multi-domain beta-ketoacyl synthase" (Identity = 49%; positives = 64%); and *Shewanella sp. MR-4* "beta-ketoacyl synthase" (Identity = 49%; positives = 64%).

[0239] At the amino acid level, the sequences with the highest level of homology to the deduced amino acid sequence derived from the cloned DNA containing the DH fragment from *Schizochytrium sp.* ATCC PTA-9695 were: *Schizochytrium sp.* ATCC 20888 "polyunsaturated fatty acid synthase subunit C" (Identity = 61%; positives = 71%); *Shewanella pealeana* ATCC 700345 "Beta-hydroxyacyl-(acyl-carrier-protein) dehydratase FabA/FabZ" (Identity = 35%; positives = 50%); and *Shewanella sediminis* HAW-EB3 "omega-3 polyunsaturated fatty acid synthase PfaC" (Identity = 34%; positives = 50%).

EXAMPLE 2

[0240] PUFA synthase genes were identified from *Schizochytrium sp.* ATCC PTA-9695.

[0241] Genomic DNA was prepared from the microorganism by standard procedures. See, e.g., Sambrook J. and Russell D. 2001. Molecular cloning: A laboratory manual, 3rd edition. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York. Briefly: (1) 500 µL of cells were pelleted from mid-log culture. The cells were Re-spun, and all traces of liquid were removed from the cell pellet with a small-bore tip; (2) pellets were resuspended with 200µL lysis buffer (20mM Tris pH 8.0, 125 µg/mL Proteinase K, 50mM NaCl, 10mM EDTA pH 8.0, 0.5% SDS); (3) cells were lysed at 50°C for 1 hour; (4) the lysis mixture was pipetted into phase-lock gel (PLG -Eppendorf) 2mL tubes; (5) equal volume of P:C:I was added and allowed to mix for 1.5 hours; (6) the tubes were

centrifuged at 12k x g for 5 minutes; (7) the aqueous phase was removed from above the gel within the PLG tube and an equal volume of chloroform was added to the aqueous phase, and mixed for 30 minutes; (8) the tubes were centrifuged at 14k for approximately 5 minutes; (9) the top layer (aqueous phase) was pipetted away from the chloroform, and placed in a new tube; (10) 0.1 volume of 3M NaOAc was added and mixed (inverted several times); (11) 2 volumes of 100% EtOH were added and mixed (inverted several times) with genomic DNA precipitant forming at this stage; (12) the tubes were spun at 4°C in a microcentrifuge at 14k for approximately 15 minutes; (13) the liquid was gently poured off with genomic DNA remaining at the bottom of the tube; (14) the pellet was washed with 0.5mL 70% EtOH; (15) the tubes were spun at 4°C in a microcentrifuge at 14k for approximately 5 minutes; (16) the EtOH was gently poured off and the genomic DNA pellet was dried; and (17) a suitable volume of H₂O and RNase was added directly to the genomic DNA pellet.

[0242] The isolated genomic DNA was used to generate recombinant libraries consisting of large fragments (approximately 40 kB) according to the manufacturer's instructions in the cosmid pWEB-TNC™ (Epicentre). The cosmid libraries were screened by standard colony hybridization procedures using ³²P radioactively labeled probes (Sambrook J. and Russell D. 2001. *Molecular cloning: A laboratory manual*, 3rd edition. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York). The probes contained DNA homologous to published PUFA synthase sequences from other organisms as described in Example 1. These probes were generated by a DNA restriction digest of the cloned fragments from respective clones from pJET1.2/blunt described above and labeled by standard methods. In all cases, strong hybridization of the individual probes to certain cosmids indicated clones containing DNA homologous to PUFA synthase genes.

[0243] Cosmid clone pDS115 demonstrated strong hybridization of probe to the KS region and was selected for DNA sequencing of the *Schizochytrium sp.* ATCC PTA-9695 *PFA1* gene. Cosmid clone pDS115, containing the *Schizochytrium sp.* ATCC PTA-9695 *PFA1* and *PFA2* genes, was deposited under the Budapest Treaty, at the American Type Culture Collection, Patent Depository, 10801 University Boulevard, Manassas, VA 20110-2209, on January 27, 2009, and given ATCC Accession No. PTA-9737. Sequencing primers to the DNA sequence of the KS region determined in Example 1 were designed using standard methods. To determine the DNA sequence of

Schizochytrium sp. ATCC PTA-9695 *PFA1*, successive rounds of DNA sequencing, involving subsequent sequencing primer design by standard methods, was carried out in order to "walk" the cosmid clone.

[0244] In previously published thraustochytrid PUFA synthase systems, the PUFA synthase genes *PFA1* and *PFA2* have been clustered together and arranged as to be divergently transcribed. This is also the case for *PFA1* and *PFA2* from *Schizochytrium sp.* ATCC PTA-9695. Through the "walking" of DNA sequence from cosmid clone pDS115, the conceptual start of *PFA2* was found to be 493 nucleotides from the start of *PFA1* and divergently transcribed. Each nucleotide base pair of the *Schizochytrium sp.* ATCC PTA-9695 *PFA1* and *PFA2* PUFA synthase genes were covered by at least two separate DNA sequencing reactions with high-quality with at least a minimum aggregated Phred score of 40 (confidence level of 99.99%).

[0245] Cosmid clone pBS4 demonstrated strong hybridization of probe to the DH region and was selected for DNA sequencing of the *Schizochytrium sp.* ATCC PTA-9695 *PFA3* gene. Cosmid clone pBS4, containing the *Schizochytrium sp.* ATCC PTA-9695 *PFA3* gene, was deposited under the Budapest Treaty, at the American Type Culture Collection, Patent Depository, 10801 University Boulevard, Manassas, VA 20110-2209, on January 27, 2009, and given ATCC Accession No. PTA-9736. Sequencing primers were designed using standard methods to the DH region DNA sequence determined in Example 1. To determine the DNA sequence of the *Schizochytrium sp.* ATCC PTA-9695 *PFA3* gene, successive rounds of DNA sequencing, involving subsequent sequencing primer design by standard methods, was carried out in order to "walk" the cosmid clone. Each nucleotide base pair of the *Schizochytrium sp.* ATCC PTA-9695 *PFA3* gene was covered by at least two separate DNA sequencing reactions of high-quality with at least a minimum aggregated Phred score of 40 (confidence level of 99.99%).

[0246] Table 1 shows identities for the *Schizochytrium sp.* ATCC PTA-9695 *PFA1* (SEQ ID NO:1), *PFA2* (SEQ ID NO:3), and *PFA3* (SEQ ID NO:5) polynucleotide sequences as compared to previously published sequences. Identities were determined by the scoring matrix "swgapdnamt" within the "AlignX" program of the VectorNTI program, a standard for DNA alignment.

Table 1: Percent Identity to *PFA1*, *PFA2*, and *PFA3* Polynucleotide Sequences

Source of Published <i>PFA1</i> , <i>PFA2</i> , and <i>PFA3</i> Sequences	% Identity of published <i>PFA1</i> (<i>orfA</i>) to <i>PFA1</i> (SEQ ID NO:1)	% Identity of published <i>PFA2</i> (<i>orfB</i>) to <i>PFA2</i> (SEQ ID NO:3)	% Identity of published <i>PFA3</i> (<i>orfC</i>) to <i>PFA3</i> (SEQ ID NO:5)
<i>Schizochytrium</i> sp. ATCC 20888	70	66	75
<i>Thraustochytrium aureum</i> ATCC 34304	65	62	not published
<i>Thraustochytrium</i> sp. 23B ATCC 20892	56	55	67

[0247] Table 2 shows identities for the *Schizochytrium* sp. ATCC PTA-9695 Pfa1p (SEQ ID NO:2), Pfa2p (SEQ ID NO:4), and Pfa3p (SEQ ID NO:6) amino acid sequences as compared to previously published PUFA synthase amino acid sequences. Identities were determined through use of the scoring matrix "blosum62mt2" within the "AlignX" program of the VectorNTI program, a standard for protein alignment.

Table 2: Percent Identity to Pfa1p, Pfa2p, and Pfa3p Amino Acid Sequences

Source of Published Pfa1p, Pfa2p, and Pfa3p Sequences	% Identity of published Pfa1p (OrfA) to Pfa1p (SEQ ID NO:2)	% Identity of published Pfa2p (OrfB) to Pfa2p (SEQ ID NO:4)	% Identity of published Pfa3p (OrfC) to Pfa3p (SEQ ID NO:6)
<i>Schizochytrium</i> sp. ATCC 20888	60	53	70
<i>Thraustochytrium aureum</i> ATCC 34304	60	54	not published
<i>Thraustochytrium</i> sp. 23B ATCC 20892	52	52	70

EXAMPLE 3

[0248] Domain analysis was performed to annotate the sequence coordinates for the PUFA synthase domains and active sites of *Schizochytrium* sp. ATCC PTA-9695 *PFA1*, *PFA2*, and *PFA3*, respectively. Domains were identified based on homology to known PUFA synthase, fatty acid synthase, and polyketide synthase domains.

[0249] Table 3 shows the domains and active sites associated with *Schizochytrium sp.* ATCC PTA-9695 *PFA1*.

Table 3: *Schizochytrium sp.* ATCC PTA-9695 *PFA1* Domain Analysis

Domain	DNA position	AA position	Sites	DNA position	AA position
KS	7-1401 of SEQ ID NO:1	3-467 of SEQ ID NO:2	Active - DXAC*	607-609 of SEQ ID NO:1	C203 of SEQ ID NO:2
	(SEQ ID NO:7)	(SEQ ID NO:8)	(SEQ ID NO:43)		
			End - GFGG	1363-1374 of SEQ ID NO:1	455-458 of SEQ ID NO:2
			(SEQ ID NO:44)	(SEQ ID NO:45)	
MAT	1798-2700 of SEQ ID NO:1	600-900 of SEQ ID NO:2	Active GHS*LG	2095-2097 of SEQ ID NO:1	S699 of SEQ ID NO:2
	(SEQ ID NO:9)	(SEQ ID NO:10)	(SEQ ID NO:46)		
ACP	3298-5400 of SEQ ID NO:1	1100-1800 of SEQ ID NO:2	ACP1 domain	3325-3600 of SEQ ID NO:1	1109-1200 of SEQ ID NO:2
	(SEQ ID NO:11)	(SEQ ID NO:12)		(SEQ ID NO:13)	(SEQ ID NO:14)
			ACP1 Active LGIDS*	3454-3456 of SEQ ID NO:1	S1152 of SEQ ID NO:2
			(SEQ ID NO:47)		
			ACP2 domain	3667-3942 of SEQ ID NO:1	1223-1314 of SEQ ID NO:2
				(SEQ ID NO:15)	(SEQ ID NO:16)
			ACP2 Active LGIDS*	3796-3798 of SEQ ID NO:1	S1266 of SEQ ID NO:2
			(SEQ ID NO:47)		
			ACP3 domain	4015-4290 of SEQ ID NO:1	1339-1430 of SEQ ID NO:2
				(SEQ ID NO:17)	(SEQ ID NO:18)
			ACP3 Active LGIDS*	4144-4146 of SEQ ID NO:1	S1382 of SEQ ID NO:2
			(SEQ ID NO:47)		
			ACP4 domain	4363-4638 of SEQ ID NO:1	1455-1546 of SEQ ID NO:2
				(SEQ ID NO:19)	(SEQ ID NO:20)
			ACP4 Active LGIDS*	4492-4494 of SEQ ID NO:1	S1498 of SEQ ID NO:2
			(SEQ ID NO:47)		

Domain	DNA position	AA position	Sites	DNA position	AA position
			ACP5 domain	4711-4986 of SEQ ID NO:1 (SEQ ID NO:21)	1571-1662 of SEQ ID NO:2 (SEQ ID NO:22)
			ACP5 Active LGIDS* (SEQ ID NO:47)	4840-4842 of SEQ ID NO:1	S1614 of SEQ ID NO:2
			ACP6 domain	5053-5328 of SEQ ID NO:1 (SEQ ID NO:23)	1685-1776 of SEQ ID NO:2 (SEQ ID NO:24)
			ACP6 Active LGIDS* (SEQ ID NO:47)	5182-5184 of SEQ ID NO:1	S1728 of SEQ ID NO:2
KR	5623-7800 of SEQ ID NO:1 (SEQ ID NO:25)	1875-2600 of SEQ ID NO:2 (SEQ ID NO:26)	“core region”	5998-6900 of SEQ ID NO:1 (SEQ ID NO:48)	2000-2300 of SEQ ID NO:2 (SEQ ID NO:49)
DH Motif	7027-7065 of SEQ ID NO:1 (SEQ ID NO:27)	2343-2355 of SEQ ID NO:2 (SEQ ID NO:28)	LxxHxxxGxxxxP (SEQ ID NO:50)	7027-7065 of SEQ ID NO:1 (SEQ ID NO:27)	2343-2355 of SEQ ID NO:2 (SEQ ID NO:28)

[0250] The first domain in *Schizochytrium sp.* ATCC PTA-9695 Pfa1 is a KS domain. The nucleotide sequence containing the sequence encoding the *Schizochytrium sp.* ATCC PTA-9695 Pfa1 KS domain is represented herein as SEQ ID NO:7, corresponding to positions 7-1401 of SEQ ID NO:1. The amino acid sequence containing the *Schizochytrium sp.* ATCC PTA-9695 Pfa1 KS domain is represented herein as SEQ ID NO:8, corresponding to positions 3-467 of SEQ ID NO:2. The KS domain contains an active site motif: DXAC* (SEQ ID NO:43), with an *acyl binding cite corresponding to C203 of SEQ ID NO:2. Also, a characteristic motif is present at the end of the KS domain: GFGG (SEQ ID NO:44), corresponding to positions 455-458 of SEQ ID NO:2 and positions 453-456 of SEQ ID NO:8.

[0251] The second domain in *Schizochytrium sp.* ATCC PTA-9695 Pfa1 is a MAT domain. The nucleotide sequence containing the sequence encoding the *Schizochytrium sp.* ATCC PTA-9695 Pfa1 MAT domain is represented herein as SEQ ID NO:9,

corresponding to positions 1798-2700 of SEQ ID NO:1. The amino acid sequence containing the *Schizochytrium* sp. ATCC PTA-9695 Pfa1 MAT domain is represented herein as SEQ ID NO:10, corresponding to positions 600-900 of SEQ ID NO:2. The MAT domain contains an active site motif: GHS*XG (SEQ ID NO:46), with an *acyl binding cite corresponding to S699 of SEQ ID NO:2.

[0252] The third through eighth domains of *Schizochytrium* sp. ATCC PTA-9695 Pfa1 are six tandem ACP domains, also referred to herein as ACP1, ACP2, ACP3, ACP4, ACP5, and ACP6. The nucleotide sequence containing the first ACP domain, ACP1, is represented herein as SEQ ID NO:13 and is contained within the nucleotide sequence spanning from about position 3325 to about position 3600 of SEQ ID NO:1. The amino acid sequence containing ACP1, represented herein as SEQ ID NO:14, is contained within the amino acid sequence spanning from about position 1109 to about position 1200 of SEQ ID NO:2. The nucleotide sequence containing ACP2, represented herein as SEQ ID NO:15, is contained within the nucleotide sequence spanning from about position 3667 to about position 3942 of SEQ ID NO:1. The amino acid sequence containing ACP2, represented herein as SEQ ID NO:16, is contained within the amino acid sequence spanning from about position 1223 to about position 1314 of SEQ ID NO:2. The nucleotide sequence containing ACP3, represented herein as SEQ ID NO:17, is contained within the nucleotide sequence spanning from about position 4015 to about position 4290 of SEQ ID NO:1. The amino acid sequence containing ACP3, represented herein as SEQ ID NO:18, is contained within the amino acid sequence spanning from about position 1339 to about position 1430 of SEQ ID NO:2. The nucleotide sequence containing ACP4, represented herein as SEQ ID NO:19, is contained within the nucleotide sequence spanning from about position 4363 to about position 4638 of SEQ ID NO:1. The amino acid sequence containing ACP4, represented herein as SEQ ID NO:20, is contained within the amino acid sequence spanning from about position 1455 to about position 1546 of SEQ ID NO:2. The nucleotide sequence containing ACP5, represented herein as SEQ ID NO:21, is contained within the nucleotide sequence spanning from about position 4711 to about position 4986 of SEQ ID NO:1. The amino acid sequence containing ACP5, represented herein as SEQ ID NO:22, is contained within the amino acid sequence spanning from about position 1571 to about position 1662 of SEQ ID NO:2. The nucleotide sequence containing ACP6, represented herein as SEQ ID NO:23, is contained

within the nucleotide sequence spanning from about position 5053 to about position 5328 of SEQ ID NO:1. The amino acid sequence containing ACP6, represented herein as SEQ ID NO:24, is contained within the amino acid sequence spanning from about position 1685 to about position 1776 of SEQ ID NO:2. All six ACP domains together span a region of *Schizochytrium sp.* ATCC PTA-9695 Pfa1 of from about position 3298 to about position 5400 of SEQ ID NO:1, corresponding to amino acid positions of about 1100 to about 1800 of SEQ ID NO:2. The nucleotide sequence for the entire ACP region containing all six domains is represented herein as SEQ ID NO:11; while the amino acid sequence for the entire ACP region containing all six domains is represented herein as SEQ ID NO:12. The repeat interval for the six ACP domains within SEQ ID NO:11 was found to be approximately every 342 nucleotides (the actual number of amino acids measured between adjacent active site serines ranges from 114 to 116 amino acids). Each of the six ACP domains contains a pantetheine binding motif LGIDS* (SEQ ID NO:47) wherein S* is the pantetheine binding site serine (S). The pantetheine binding site serine (S) is located near the center of each ACP domain sequence. The locations of the active site serine residues (i.e., the pantetheine binding site) for each of the six ACPD domains, with respect to the amino acid sequence of SEQ ID NO:2 are: ACP1 = S1152, ACP2 = S1266, ACP3 = S1382, ACP4 = S1498, ACP5 = S1614, and ACP6 = S1728.

[0253] The ninth domain in *Schizochytrium sp.* ATCC PTA-9695 Pfa1 is a KR domain. The nucleotide sequence containing the sequence encoding the *Schizochytrium sp.* ATCC PTA-9695 Pfa1 KR domain is represented herein as SEQ ID NO:25, corresponding to positions 5623-7800 of SEQ ID NO:1. The amino acid sequence containing the *Schizochytrium sp.* ATCC PTA-9695 Pfa1 KR domain is represented herein as SEQ ID NO:26, corresponding to positions 1875-2600 of SEQ ID NO:2. Within the KR domain is a core region (contained within the nucleotide sequence of SEQ ID NO:48, and the amino acid sequence of SEQ ID NO:49) with homology to short chain aldehyde-dehydrogenases (KR is a member of this family). This core region spans from about position 5998 to about 6900 of SEQ ID NO:1, which corresponds to amino acid positions 2000-2300 of SEQ ID NO:2.

[0254] The tenth domain in *Schizochytrium sp.* ATCC PTA-9695 Pfa1 is a DH domain. The nucleotide sequence containing the sequence encoding the *Schizochytrium sp.* ATCC PTA-9695 Pfa1 DH domain is represented herein as SEQ ID NO:27, corresponding to

positions 7027-7065 of SEQ ID NO:1. The amino acid sequence containing the *Schizochytrium sp.* ATCC PTA-9695 Pfa1 DH domain is represented herein as SEQ ID NO:28, corresponding to positions 2343-2355 of SEQ ID NO:2. The DH domain contains a conserved active site motif (*See*, Donadio, S. and Katz., L., *Gene* 111(1): 51-60 (1992)): LxxHxxxGxxxxP (SEQ ID NO:50).

[0255] Table 4 shows the domains and active sites associated with *Schizochytrium sp.* ATCC PTA-9695 PFA2.

Table 4: *Schizochytrium sp.* ATCC PTA-9695 PFA2 Domain Analysis

Domain	DNA positions	AA positions	Sites	DNA positions	AA positions
KS	10-1350 of SEQ ID NO:3 (SEQ ID NO:29)	4-450 of SEQ ID NO:4 (SEQ ID NO:30)	DXAC* (SEQ ID NO:43)	571-573 of SEQ ID NO:3	C191 of SEQ ID NO:4
			End - GFGG (SEQ ID NO:44)	1312-1323 of SEQ ID NO:3 (SEQ ID NO:51)	438-441 of SEQ ID NO:4
CLF	1408-2700 of SEQ ID NO:3 (SEQ ID NO:31)	470-900 of SEQ ID NO:4 (SEQ ID NO:32)			
AT	2998-4200 of SEQ ID NO:3 (SEQ ID NO:33)	1000-1400 of SEQ ID NO:4 (SEQ ID NO:34)	GxS*xG (SEQ ID NO:52)	3421-3423 of SEQ ID NO:3	S1141 of SEQ ID NO:4
ER	4498-5700 of SEQ ID NO:3 (SEQ ID NO:35)	1500-1900 of SEQ ID NO:4 (SEQ ID NO:36)			

[0256] The first domain in *Schizochytrium sp.* ATCC PTA-9695 Pfa2 is a KS domain. The nucleotide sequence containing the sequence encoding the *Schizochytrium sp.* ATCC PTA-9695 Pfa2 KS domain is represented herein as SEQ ID NO:29, corresponding to positions 10-1350 of SEQ ID NO:3. The amino acid sequence containing the *Schizochytrium sp.* ATCC PTA-9695 Pfa2 KS domain is represented herein as SEQ ID NO:30, corresponding to positions 4-450 of SEQ ID NO:4. The KS domain contains an active site motif: DXAC* (SEQ ID NO:43), with an *acyl binding cite corresponding to C191 of SEQ ID NO:4. Also, a characteristic motif is present at the end of the KS

domain: GFGG (SEQ ID NO:44), corresponding to positions 438-441 of SEQ ID NO:4 and positions 435-438 of SEQ ID NO:30.

[0257] The third domain in *Schizochytrium sp.* ATCC PTA-9695 Pfa2 is a CLF domain. The nucleotide sequence containing the sequence encoding the *Schizochytrium sp.* ATCC PTA-9695 Pfa2 CLF domain is represented herein as SEQ ID NO:31, corresponding to positions 1408-2700 of SEQ ID NO :3. The amino acid sequence containing the *Schizochytrium sp.* ATCC PTA-9695 Pfa2 CLF domain is represented herein as SEQ ID NO:32, corresponding to positions 470-900 of SEQ ID NO:4.

[0258] The third domain in *Schizochytrium sp.* ATCC PTA-9695 Pfa2 is an AT domain. The nucleotide sequence containing the sequence encoding the *Schizochytrium sp.* ATCC PTA-9695 Pfa2 AT domain is represented herein as SEQ ID NO:33, corresponding to positions 2998-4200 of SEQ ID NO:3. The amino acid sequence containing the *Schizochytrium sp.* ATCC PTA-9695 Pfa2 AT domain is represented herein as SEQ ID NO:34, corresponding to positions 1000-1400 of SEQ ID NO:4. The AT domain contains an active site motif of GxS*xG (SEQ ID NO:52) that is characteristic of acyltransferase (AT) proteins, with an active site serine residue corresponding to S1141 of SEQ ID NO:4.

[0259] The fourth domain of *Schizochytrium sp.* ATCC PTA-9695 Pfa2 is an ER domain. The nucleotide sequence containing the sequence encoding the *Schizochytrium sp.* ATCC PTA-9695 Pfa2 ER domain is represented herein as SEQ ID NO:35, corresponding to positions 4498-5700 of SEQ ID NO:3. The amino acid sequence containing the Pfa2 ER domain is represented herein as SEQ ID NO:36, corresponding to positions 1500-1900 of SEQ ID NO:4.

[0260] Table 5 shows the domains and active sites associated with *Schizochytrium sp.* ATCC PTA-9695 *PFA3*.

Table 5: *Schizochytrium sp.* ATCC PTA-9695 PFA3 Domain Analysis

Domain	DNA positions	AA positions	Sites	DNA positions	AA positions
DH1	1-1350 of SEQ ID NO:5 (SEQ ID NO:37)	1-450 of SEQ ID NO:6 (SEQ ID NO:38)	FxxH*F (SEQ ID NO:53)	931-933 of SEQ ID NO:5	H310 of SEQ ID NO:6
DH2	1501-2700 of SEQ ID NO:5 (SEQ ID NO:39)	501-900 of SEQ ID NO:6 (SEQ ID NO:40)	FxxH*F (SEQ ID NO:53)	2401-2403 of SEQ ID NO:5	H801 of SEQ ID NO:6
ER	2848-4200 of SEQ ID NO:5 (SEQ ID NO:41)	950-1400 of SEQ ID NO:6 (SEQ ID NO:42)			

[0261] The first and second domains of *Schizochytrium sp.* ATCC PTA-9695 Pfa3 are DH domains, referred to herein as DH1 and DH2, respectively. The nucleotide sequence containing the sequence encoding the *Schizochytrium sp.* ATCC PTA-9695 Pfa3 DH1 domain is represented herein as SEQ ID NO:37, corresponding to positions 1-1350 of SEQ ID NO:5. The amino acid sequence containing the *Schizochytrium sp.* ATCC PTA-9695 Pfa3 DH1 domain is represented herein as SEQ ID NO:38, corresponding to positions 1-450 of SEQ ID NO:6. The nucleotide sequence containing the sequence encoding the *Schizochytrium sp.* ATCC PTA-9695 Pfa3 DH2 domain is represented herein as SEQ ID NO:39, corresponding to positions 1501-2700 of SEQ ID NO:5. The amino acid sequence containing the *Schizochytrium sp.* ATCC PTA-9695 Pfa3 DH2 domain is represented herein as SEQ ID NO:40, corresponding to positions 501-900 of SEQ ID NO:6. The DH domains contain an active site motif: FxxH*F (SEQ ID NO:53). The nucleotide sequence containing the active site motif in DH1 corresponds to positions 931-933 of SEQ ID NO:5, while the nucleotide sequence containing the active site motif in DH2 corresponds to positions 2401-2403 of SEQ ID NO:5. The active site H* in the motif FxxH*F is based on data from Leesong *et al.*, *Structure* 4:253-64 (1996) and Kimber *et al.* *J Biol Chem.* 279:52593-602 (2004), with the active site H* in DH1 corresponding to H310 of SEQ ID NO:6 and the active site H* in DH2 corresponding to H801 of SEQ ID NO:6.

[0262] The third domain of *Schizochytrium sp.* ATCC PTA-9695 Pfa3 is an ER domain. The nucleotide sequence containing the sequence encoding the *Schizochytrium sp.* ATCC PTA-9695 Pfa3 ER domain is represented herein as SEQ ID NO:41, corresponding to

positions 2848-4200 of SEQ ID NO:5. The amino acid sequence containing the *Schizochytrium sp.* ATCC PTA-9695 Pfa3 ER domain is represented herein as SEQ ID NO:42, corresponding to positions 950-1400 of SEQ ID NO:6.

EXAMPLE 4

[0263] Degenerate primers for the KS, ER, and DH PUFA synthase domains were designed in order to isolate the corresponding sequences from the isolated microorganism deposited under ATCC Accession No. PTA-10212, also known as *Thraustochytrium sp.* ATCC PTA-10212.

[0264] Degenerate primers for the KS region of *Thraustochytrium sp.* ATCC PTA-10212 *PFA1* (i.e., the region containing the KS domain) were designed based on the published *PFA1* (previously termed *orfA* or ORF 1) sequences for *Schizochytrium sp.* ATCC 20888, *Thraustochytrium aureum* (ATCC 34304), and *Thraustochytrium sp.* 23B ATCC 20892:

prDS233 (forward): TGATATGGGAGGAATGAATTGTGTNGTNGAYGC

(SEQ ID NO:123)

prDS235 (reverse): TTCCATAACAAAATGATAATTAGCTCCNCCRAANCC

(SEQ ID NO:124).

[0265] Degenerate primers for the ER region of *Thraustochytrium sp.* ATCC PTA-10212 *PFA2* (i.e., the region containing the ER domain) were designed based on the published *PFA2* (previously termed *orfB* or ORF 2) sequences for *Shewanella japonica*, *Schizochytrium sp.* ATCC 20888, *Thraustochytrium aureum* (ATCC 34304), and *Thraustochytrium sp.* 23B ATCC 20892:

prDS183 (forward): GGCGGCCACACCGAYAAYMGNCC (SEQ ID NO:125)

prDS184 (reverse): CGGGGCCGCACCANAYYTGRTA (SEQ ID NO:126).

[0266] Degenerate primers for the ER region of *Thraustochytrium sp.* ATCC PTA-10212 *PFA3* (i.e., the region containing the ER domain) were designed based on the published *PFA3* (previously termed *orfC* or ORF 3) sequences for *Shewanella japonica*, *Schizochytrium sp.* ATCC 20888, *Thraustochytrium aureum* (ATCC 34304), and *Thraustochytrium sp.* 23B ATCC 20892:

prDS181 (forward): TCCTTCGGNGCNGSNGG (SEQ ID NO:127)

prDS184 (reverse): CGGGGCCGCACCANAYYTGRTA (SEQ ID NO:126).

- [0267] Degenerate primers JGM190 (forward, SEQ ID NO:64) and BLR242 (reverse, SEQ ID NO:65), as described above, were used to amplify the DH region of *PFA3* from *Thraustochytrium sp.* ATCC PTA-10212.
- [0268] The PCR conditions with chromosomal DNA template were as follows: 0.2 μ M dNTPs, 0.1 μ M each primer, 6% DMSO, 200 ng chromosomal DNA, 2.5 U Herculase® II fusion polymerase (Stratagene), and 1X Herculase® buffer (Stratagene) in a 50 μ L total volume. The PCR Protocol included the following steps: (1) 98°C for 3 minutes; (2) 98°C for 30 seconds; (3) 54°C for 45 seconds; (4) 72°C for 1 minutes; (5) repeat steps 2-4 for 40 cycles; (6) 72°C for 5 minutes; and (7) hold at 6°C.
- [0269] For all primer pairs, PCR yielded distinct DNA products with the expected sizes using chromosomal templates from *Thraustochytrium sp.* ATCC PTA-10212. The respective PCR products were cloned into the vector pJET1.2/blunt (Fermentas) according to the manufacturer's instructions, and the insert sequence was determined using supplied standard primers.
- [0270] The DNA sequences obtained from the PCR products were compared with known sequences available from the NCBI GenBank as described in Example 1.
- [0271] At the amino acid level, the sequences with the highest level of homology to deduced amino acid sequence derived from the cloned DNA containing the KS fragment from *PFA1* from *Thraustochytrium sp.* ATCC PTA-10212 were: *Schizochytrium sp.* ATCC 20888 "polyunsaturated fatty acid synthase subunit A" (Identity = 80%; positives = 90%); *Shewanella benthica* KT99 "omega-3 polyunsaturated fatty acid synthase PfaA" (Identity = 51%; positives = 67%); *Shewanella loihica* PV-4 "beta-ketoacyl synthase" (Identity = 50%; positives = 67%); *Shewanella woodyi* ATCC 51908 "polyketide-type polyunsaturated fatty acid synthase PfaA" (Identity = 51%; positives = 66%).
- [0272] At the amino acid level, the sequences with the highest level of homology to deduced amino acid sequence derived from the cloned DNA containing the ER fragment from *PFA2* from *Thraustochytrium sp.* ATCC PTA-10212 were: *Schizochytrium sp.* ATCC 20888 "polyunsaturated fatty acid synthase subunit B" (Identity = 70%; positives = 85%); *Schizochytrium sp.* ATCC 20888 "polyunsaturated fatty acid synthase subunit C" (Identity = 66%; positives = 83%); *Nodularia spumigena* CCY9414 "2-nitropropane dioxygenase" (Identity = 57%; positives = 74%); *Moritella sp.* PE36 "polyunsaturated fatty acid synthase PfaD" (Identity = 57%; positives = 71%).

- [0273] At the amino acid level, the sequences with the highest level of homology to deduced amino acid sequence derived from the cloned DNA containing the ER fragment from *PFA3* from *Thraustochytrium* sp. ATCC PTA-10212 were: *Schizochytrium* sp. ATCC 20888 "polyunsaturated fatty acid synthase subunit C" (Identity = 80%; positives = 90%); *Schizochytrium* sp. ATCC 20888 "polyunsaturated fatty acid synthase subunit B" (Identity = 78%; positives = 89%); *Moritella* sp. PE36 "polyunsaturated fatty acid synthase PfaD" (Identity = 56%; positives = 71%); *Shewanella amazonensis* SB2B "omega-3 polyunsaturated fatty acid synthase PfaD" (Identity = 55%; positives = 73%).
- [0274] At the amino acid level, the sequences with the highest level of homology to deduced amino acid sequence derived from the cloned DNA containing the DH fragment from *PFA3* from *Thraustochytrium* sp. ATCC PTA-10212 were: *Schizochytrium* sp. ATCC 20888 "polyunsaturated fatty acid synthase subunit C" (Identity = 63%; positives = 76%); *Shewanella pealeana* ATCC 700345 "Beta-hydroxyacyl-(acyl-carrier-protein) dehydratase FabA/FabZ" (Identity = 35%; positives = 53%); *Shewanella piezotolerans* WP3 "Multi-domain beta-ketoacyl synthase" (Identity = 36%; positives = 52%); *Shewanella benthica* KT99 "omega-3 polyunsaturated fatty acid synthase PfaC" (Identity = 35%; positives = 51%).

EXAMPLE 5

- [0275] PUFA synthase genes were identified from *Thraustochytrium* sp. ATCC PTA-10212.
- [0276] From a -80°C cryovial, 1 mL of cells were thawed at room temperature and added to 50 mL of liquid HSFM media (below) in a 250 mL non-baffled flask. The flask was incubated at 23°C for 3 days. Cells were collected and utilized for standard Bacterial Artificial Chromosome (BAC) library construction (Lucigen Corporation, Middleton, WI USA).

Table 6: HSFM Media

<u>Ingredient</u>	<u>concentration</u>	<u>ranges</u>
Na ₂ SO ₄	g/L 31.0	0-50, 15-45, or 25-35
NaCl	g/L 0.625	0-25, 0.1-10, or 0.5-5
KCl	g/L 1.0	0-5, 0.25-3, or 0.5-2
MgSO ₄ ·7H ₂ O	g/L 5.0	0-10, 2-8, or 3-6
(NH ₄) ₂ SO ₄	g/L 0.44	0-10, 0.25-5, or 0.05-3

-85-

MSG·1H ₂ O	g/L	6.0	0-10, 4-8, or 5-7
CaCl ₂	g/L	0.29	0.1-5, 0.15-3, or 0.2-1
T 154 (yeast extract)	g/L	6.0	0-20, 0.1-10, or 1-7
KH ₂ PO ₄	g/L	0.8	0.1-10, 0.5-5, or 0.6-1.8

Post autoclave (Metals)

Citric acid	mg/L	3.5	0.1-5000, 10-3000, or 3-2500
FeSO ₄ ·7H ₂ O	mg/L	10.30	0.1-100, 1-50, or 5-25
MnCl ₂ ·4H ₂ O	mg/L	3.10	0.1-100, 1-50, or 2-25
ZnSO ₄ ·7H ₂ O	mg/L	3.10	0.01-100, 1-50, or 2-25
CoCl ₂ ·6H ₂ O	mg/L	0.04	0-1, 0.001-0.1, or 0.01-0.1
Na ₂ MoO ₄ ·2H ₂ O	mg/L	0.04	0.001-1, 0.005-0.5, or 0.01-0.1
CuSO ₄ ·5H ₂ O	mg/L	2.07	0.1-100, 0.5-50, or 1-25
NiSO ₄ ·6H ₂ O	mg/L	2.07	0.1-100, 0.5-50, or 1-25

Post autoclave (Vitamins)

Thiamine	mg/L	9.75	0.1-100, 1-50, or 5-25
Vitamin B12	mg/L	0.16	0.01-100, 0.05-5, or 0.1-1
Ca ^{1/2} -pantothenate	mg/L	2.06	0.1-100, 0.1-50, or 1-10
Biotin	mg/L	3.21	0.1-100, 0.1-50, or 1-10

Post autoclave (Carbon)

Glycerol	g/L	30.0	5-150, 10-100, or 20-50
----------	-----	------	-------------------------

Nitrogen Feed:

<u>Ingredient</u>	<u>Concentration</u>	
MSG·1H ₂ O	g/L	17
		0-150, 10-100, or 15-50

Typical cultivation conditions would include the following:

pH	about 6.5 – about 9.5, about 6.5 – about 8.0, or about 6.8 – about 7.8;
temperature:	about 15 – about 30 degrees Celsius, about 18 – about 28 degrees Celsius, or about 21 to about 23 degrees Celsius;
dissolved oxygen:	about 0.1 – about 100% saturation, about 5 – about 50% saturation, or about 10 – about 30% saturation; and/or
glycerol controlled @:	about 5 – about 50 g/L, about 10 – about 40 g/L, or about 15 – about 35 g/L.

[0277] The recombinant BAC libraries, consisting of large fragments (average of approximately 120 kB) were handled according to the manufacturer's instructions in the BAC vector pSMART® (Lucigen Corporation). The BAC libraries were screened by standard colony hybridization procedures using ³²P radioactively labeled probes (Sambrook J. and Russell D. 2001. Molecular cloning: A laboratory manual, 3rd edition.

Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York). The probes contained DNA homologous to published PUFA synthase sequences from other organisms as described in Example 4. These probes were generated by a DNA restriction digest of the cloned fragments from respective clones from pJET1.2/blunt described above and labeled by standard methods. In all cases, strong hybridization of the individual probes to certain BACs indicated clones containing DNA homologous to PUFA synthase genes.

[0278] BAC clone pLR130 (also known as LuMaBAC 2M23) demonstrated strong hybridization of probe to both the KS region and ER region, indicating that it contained the *PFA1* and *PFA2* genes, and was selected for DNA sequencing of the *Thraustochytrium* sp. ATCC PTA-10212 *PFA1* and *PFA2* genes. The BAC was sequenced by standard procedures (Eurofins MWG Operon, Huntsville, AL). BAC clone pLR130, containing the *PFA1* and *PFA2* genes, was deposited under the Budapest Treaty, at the American Type Culture Collection, Patent Depository, 10801 University Boulevard, Manassas, VA 20110-2209, on December 1, 2009, and given ATCC Accession No. PTA-10511.

[0279] In previously published thraustochytrid PUFA synthase systems, the PUFA synthase genes *PFA1* and *PFA2* have been clustered together and arranged as to be divergently transcribed. This is also the case for *PFA1* and *PFA2* from *Thraustochytrium* sp. ATCC PTA-10212. The conceptual start of *PFA2* was found to be 693 nucleotides from the start of *PFA1* and divergently transcribed.

[0280] BAC clone pDS127 (also known as LuMaBAC 9K17) demonstrated strong hybridization of probe to both the DH region and ER region of *PFA3* and was selected for DNA sequencing of the *PFA3* gene. BAC clone pDS127, containing the *PFA3* gene, was deposited under the Budapest Treaty, at the American Type Culture Collection, Patent Depository, 10801 University Boulevard, Manassas, VA 20110-2209, on December 1, 2009, and given ATCC Accession No. PTA-10510. Sequencing primers were designed using standard methods to the DH region and ER region and the DNA sequence determined in Example 4. To determine the DNA sequence of the *Thraustochytrium* sp. ATCC PTA-10212 *PFA3* gene, successive rounds of DNA sequencing, involving subsequent sequencing primer design by standard methods, was carried out in order to "walk" the BAC clone. Each nucleotide base pair of the *PFA3* gene was covered by at

least two separate DNA sequencing reactions of high-quality with at least a minimum aggregated Phred score of 40 (confidence level of 99.99%).

[0281] Table 7 shows identities for the *Thraustochytrium* sp. ATCC PTA-10212 *PFA1* (SEQ ID NO:68), *PFA2* (SEQ ID NO:70), and *PFA3* (SEQ ID NO:72) polynucleotide sequences as compared to previously published sequences and the sequences from *Schizochytrium* sp. PTA-9695. Identities were determined by the scoring matrix "swgapdnamt" within the "AlignX" program of the VectorNTI program, a standard for DNA alignment.

Table 7: Percent Identity to *PFA1*, *PFA2*, and *PFA3* Polynucleotide Sequences

Source of Comparison <i>PFA1</i> , <i>PFA2</i> , and <i>PFA3</i> Sequences	% Identity of Comparison <i>PFA1</i> (<i>orfA</i>) to <i>PFA1</i>	% Identity of Comparison <i>PFA2</i> (<i>orfB</i>) to <i>PFA2</i>	% Identity of Comparison <i>PFA3</i> (<i>orfC</i>) to <i>PFA3</i>
<i>Schizochytrium</i> sp. ATCC 20888	55	54	59
<i>Thraustochytrium aureum</i> ATCC 34304	55	53	not published
<i>Thraustochytrium</i> sp. 23B ATCC 20892	55	57	62
<i>Schizochytrium</i> sp. PTA-9695	55	52	59

[0282] Table 8 shows identities for the *Thraustochytrium* sp. ATCC PTA-10212 Pfa1p (SEQ ID NO:69), Pfa2p (SEQ ID NO:71), and Pfa3p (SEQ ID NO:73) amino acid sequences as compared to previously published PUFA synthase amino acid sequences and the sequences from *Schizochytrium* sp. PTA-9695. Identities were determined through use of the scoring matrix "blosum62mt2" within the "AlignX" program of the VectorNTI program, a standard for protein alignment.

Table 8: Percent Identity to Pfa1p, Pfa2p, and Pfa3p Amino Acid Sequences

Source of Comparison Pfa1p, Pfa2p, and Pfa3p Sequences	% Identity of Comparison Pfa1p (<i>OrfA</i>) to Pfa1p	% Identity of Comparison Pfa2p (<i>OrfB</i>) to Pfa2p	% Identity of Comparison Pfa3p (<i>OrfC</i>) to Pfa3p
<i>Schizochytrium</i> sp. ATCC 20888	62	57	69
<i>Thraustochytrium aureum</i> ATCC 34304	58	54	not published
<i>Thraustochytrium</i> sp. 23B ATCC 20892	54	54	71
<i>Schizochytrium</i> sp. PTA-9695	59	53	73

EXAMPLE 6

[0283] Domain analysis was performed to annotate the sequence coordinates for the PUFA synthase domains and active sites of *Thraustochytrium sp.* ATCC PTA-10212 *PFA1*, *PFA2*, and *PFA3*, respectively. Domains were identified based on homology to known PUFA synthase, fatty acid synthase, and polyketide synthase domains.

[0284] Table 9 shows the domains and active sites associated with *Thraustochytrium sp.* ATCC PTA-10212 *PFA1*.

Table 9: *Thraustochytrium sp.* ATCC PTA-10212 *PFA1* Domain Analysis

Domain	DNA position	AA position	Sites	DNA position	AA position
KS	13-1362 of SEQ ID NO:68	5-545 of SEQ ID NO:69	Active - DXAC*	601-612 of SEQ ID NO:68	C204 of SEQ ID NO:69
	(SEQ ID NO:74)	(SEQ ID NO:75)	(SEQ ID NO:43)		
			End - GFGG	1351-1362 of SEQ ID NO:68	451-454 of SEQ ID NO:69
			(SEQ ID NO:44)	(SEQ ID NO:45)	
MAT	1783-2703 of SEQ ID NO:68	595-901 of SEQ ID NO:69	Active GHS*LG	2083-2085 of SEQ ID NO:68	S695 of SEQ ID NO:69
	(SEQ ID NO:76)	(SEQ ID NO:77)	(SEQ ID NO:46)	(SEQ ID NO:116)	
ACP	3208-6510 of SEQ ID NO:68	1070-2170 of SEQ ID NO:69	ACP1 domain	3280-3534 of SEQ ID NO:68	1094-1178 of SEQ ID NO:69
	(SEQ ID NO:78)	(SEQ ID NO:79)		(SEQ ID NO:80)	(SEQ ID NO:81)
			ACP1 Active LGIDS*	3403-3405 of SEQ ID NO:68	S1135 of SEQ ID NO:69
			(SEQ ID NO:47)		
			ACP2 domain	3607-3861 of SEQ ID NO:68	1203-1287 of SEQ ID NO:69
				(SEQ ID NO:82)	(SEQ ID NO:83)
			ACP2 Active LGIDS*	3730-3732 of SEQ ID NO:68	S1244 of SEQ ID NO:69
			(SEQ ID NO:47)		
			ACP3 domain	3934-4185 of SEQ ID NO:68	1312-1396 of SEQ ID NO:69
				(SEQ ID NO:84)	(SEQ ID NO:85)

Domain	DNA position	AA position	Sites	DNA position	AA position
			ACP3 Active LGIDS* (SEQ ID NO:47)	4057-4059 of SEQ ID NO:68	S1353 of SEQ ID NO:69
			ACP4 domain (SEQ ID NO:86)	4261-4515 of SEQ ID NO:68 (SEQ ID NO:86)	1421-1505 of SEQ ID NO:69 (SEQ ID NO:87)
			ACP4 Active LGIDS* (SEQ ID NO:47)	4384-4386 of SEQ ID NO:68	S1462 of SEQ ID NO:69
			ACP5 domain (SEQ ID NO:88)	4589-4842 of SEQ ID NO:68 (SEQ ID NO:88)	1530-1614 of SEQ ID NO:69 (SEQ ID NO:89)
			ACP5 Active LGIDS* (SEQ ID NO:47)	4711-4713 of SEQ ID NO:68	S1571 of SEQ ID NO:69
			ACP6 domain (SEQ ID NO:90)	4915-5169 of SEQ ID NO:68 (SEQ ID NO:90)	1639-1723 of SEQ ID NO:69 (SEQ ID NO:91)
			ACP6 Active LGIDS* (SEQ ID NO:47)	5038-5040 of SEQ ID NO:68	S1680 of SEQ ID NO:69
			ACP7 domain (SEQ ID NO:92)	5242-5496 of SEQ ID NO:68 (SEQ ID NO:92)	1748-1832 of SEQ ID NO:69 (SEQ ID NO:93)
			ACP7 Active LGIDS* (SEQ ID NO:47)	5365-5367 of SEQ ID NO:68	S1789 of SEQ ID NO:69
			ACP8 domain (SEQ ID NO:94)	5569-5823 of SEQ ID NO:68 (SEQ ID NO:94)	1857-1941 of SEQ ID NO:69 (SEQ ID NO:95)
			ACP8 Active LGIDS* (SEQ ID NO:47)	5692-5694 of SEQ ID NO:68	S1898 of SEQ ID NO:69
			ACP9 domain (SEQ ID NO:96)	5896-6150 of SEQ ID NO:68 (SEQ ID NO:96)	1966-2050 of SEQ ID NO:69 (SEQ ID NO:97)

Domain	DNA position	AA position	Sites	DNA position	AA position
			ACP9 Active LGIDS* (SEQ ID NO:47)	6019-6021 of SEQ ID NO:68	S2007 of SEQ ID NO:69
			ACP10 domain (SEQ ID NO:98)	6199-6453 of SEQ ID NO:68 (SEQ ID NO:98)	2067-2151 of SEQ ID NO:69 (SEQ ID NO:99)
			ACP10 Active LGIDS* (SEQ ID NO:47)	6322-6324 of SEQ ID NO:68	S2108 of SEQ ID NO:69
KR	6808-8958 of SEQ ID NO:68 (SEQ ID NO:100)	2270-2986 of SEQ ID NO:69 (SEQ ID NO:101)	"core region"	7198-8100 of SEQ ID NO:68 (SEQ ID NO:116)	2400-2600 of SEQ ID NO:69 (SEQ ID NO:117)
DH Motif	8203-8241 of SEQ ID NO:68 (SEQ ID NO:118)	2735-2747 of SEQ ID NO:69 (SEQ ID NO:119)	LxxHxxxGxxxxP (SEQ ID NO:50)	8203-8241 of SEQ ID NO:68 (SEQ ID NO:118)	2735-2747 of SEQ ID NO:69 (SEQ ID NO:119)

[0285] The first domain in *Thraustochytrium sp.* ATCC PTA-10212 Pfa1 is a KS domain. The nucleotide sequence containing the sequence encoding the *Thraustochytrium sp.* ATCC PTA-10212 Pfa1 KS domain is represented herein as SEQ ID NO:74, corresponding to positions 13-1362 of SEQ ID NO:68. The amino acid sequence containing the *Thraustochytrium sp.* ATCC PTA-10212 Pfa1 KS domain is represented herein as SEQ ID NO:75, corresponding to positions 5-454 of SEQ ID NO:69. The KS domain contains an active site motif: DXAC* (SEQ ID NO:43), with an *acyl binding cite corresponding to C204 of SEQ ID NO:69. Also, a characteristic motif is present at the end of the KS domain: GFGG (SEQ ID NO:44), corresponding to positions 451-454 of SEQ ID NO:69 and positions 447-450 of SEQ ID NO:75.

[0286] The second domain in *Thraustochytrium sp.* ATCC PTA-10212 Pfa1 is a MAT domain. The nucleotide sequence containing the sequence encoding the *Thraustochytrium sp.* ATCC PTA-10212 Pfa1 MAT domain is represented herein as SEQ ID NO:76, corresponding to positions 1783-2703 of SEQ ID NO:68. The amino acid sequence containing the *Thraustochytrium sp.* ATCC PTA-10212 Pfa1 MAT domain is represented herein as SEQ ID NO:77, corresponding to positions 595-901 of SEQ ID

NO:69. The MAT domain contains an active site motif: GHS*XG (SEQ ID NO:46), with an *acyl binding cite corresponding to S695 of SEQ ID NO:69.

[0287] The third through twelfth domains of *Thraustochytrium sp.* ATCC PTA-10212 Pfa1p are ten tandem ACP domains, also referred to herein as ACP1, ACP2, ACP3, ACP4, ACP5, ACP6, ACP7, ACP8, ACP9, and ACP10. The nucleotide sequence containing the first ACP domain, ACP1, is represented herein as SEQ ID NO:80 and is contained within the nucleotide sequence spanning from about position 3280 to about position 3534 of SEQ ID NO:68. The amino acid sequence containing ACP1, represented herein as SEQ ID NO:81, is contained within the amino acid sequence spanning from about position 1094 to about position 1178 of SEQ ID NO:69. The nucleotide sequence containing ACP2, represented herein as SEQ ID NO:82, is contained within the nucleotide sequence spanning from about position 3607 to about position 3861 of SEQ ID NO:68. The amino acid sequence containing ACP2, represented herein as SEQ ID NO:83, is contained within the amino acid sequence spanning from about position 1203 to about position 1287 of SEQ ID NO:69. The nucleotide sequence containing ACP3, represented herein as SEQ ID NO:84, is contained within the nucleotide sequence spanning from about position 3934 to about position 4185 of SEQ ID NO:68. The amino acid sequence containing ACP3, represented herein as SEQ ID NO:85, is contained within the amino acid sequence spanning from about position 1312 to about position 1396 of SEQ ID NO:69. The nucleotide sequence containing ACP4, represented herein as SEQ ID NO:86, is contained within the nucleotide sequence spanning from about position 4261 to about position 4515 of SEQ ID NO:68. The amino acid sequence containing ACP4, represented herein as SEQ ID NO:87, is contained within the amino acid sequence spanning from about position 1421 to about position 1505 of SEQ ID NO:69. The nucleotide sequence containing ACP5, represented herein as SEQ ID NO:88, is contained within the nucleotide sequence spanning from about position 4589 to about position 4842 of SEQ ID NO:68. The amino acid sequence containing ACP5, represented herein as SEQ ID NO:89, is contained within the amino acid sequence spanning from about position 1530 to about position 1614 of SEQ ID NO:69. The nucleotide sequence containing ACP6, represented herein as SEQ ID NO:90, is contained within the nucleotide sequence spanning from about position 4915 to about position 5169 of SEQ ID NO:68. The amino acid sequence containing ACP6, represented herein as

SEQ ID NO:91, is contained within the amino acid sequence spanning from about position 1639 to about position 1723 of SEQ ID NO:69. The nucleotide sequence containing ACP7, represented herein as SEQ ID NO:92, is contained within the nucleotide sequence spanning from about position 5242 to about position 5496 of SEQ ID NO:68. The amino acid sequence containing ACP7, represented herein as SEQ ID NO:93, is contained within the amino acid sequence spanning from about position 1748 to about position 1832 of SEQ ID NO:69. The nucleotide sequence containing ACP8, represented herein as SEQ ID NO:94, is contained within the nucleotide sequence spanning from about position 5569 to about position 5832 of SEQ ID NO:68. The amino acid sequence containing ACP8, represented herein as SEQ ID NO:95, is contained within the amino acid sequence spanning from about position 1857 to about position 1941 of SEQ ID NO:69. The nucleotide sequence containing ACP9, represented herein as SEQ ID NO:96, is contained within the nucleotide sequence spanning from about position 5896 to about position 6150 of SEQ ID NO:68. The amino acid sequence containing ACP9, represented herein as SEQ ID NO:97, is contained within the amino acid sequence spanning from about position 1966 to about position 2050 of SEQ ID NO:69. The nucleotide sequence containing ACP10, represented herein as SEQ ID NO:98, is contained within the nucleotide sequence spanning from about position 6199 to about position 6453 of SEQ ID NO:68. The amino acid sequence containing ACP10, represented herein as SEQ ID NO:99, is contained within the amino acid sequence spanning from about position 2067 to about position 2151 of SEQ ID NO:69. All ten ACP domains together span a region of *Thraustochytrium* sp. ATCC PTA-10212 Pfa1 of from about position 3208 to about position 6510 of SEQ ID NO:68, corresponding to amino acid positions of about 1070 to about 2170 of SEQ ID NO:69. The nucleotide sequence for the entire ACP region containing all 10 domains is represented herein as SEQ ID NO:78; while the amino acid sequence for the entire ACP region containing all six domains is represented herein as SEQ ID NO:79. The repeat interval for the 10 ACP domains within SEQ ID NO:78 was found to be approximately every 327 nucleotides (the actual number of amino acids measured between adjacent active site serines ranges from 101 to 109 amino acids). Each of the ten ACP domains contains a pantetheine binding motif LGIDS* (SEQ ID NO:47) wherein S* is the pantetheine binding site serine (S). The pantetheine binding site serine (S) is located near the center of each ACP domain

sequence. The locations of the active site serine residues (i.e., the pantetheine binding site) for each of the six ACPD domains, with respect to the amino acid sequence of SEQ ID NO:69 are: ACP1 = S1135, ACP2 = S1244, ACP3 = S1353, ACP4 = S1462, ACP5 = S1571, ACP6 = S1680, ACP7 = S1789, ACP8 = S1898, ACP9 = S2007, and ACP10 = S2108.

[0288] The thirteenth domain in *Thraustochytrium sp.* ATCC PTA-10212 Pfa1 is a KR domain. The nucleotide sequence containing the sequence encoding the Pfa1 KR domain is represented herein as SEQ ID NO:100, corresponding to positions 6808-8958 of SEQ ID NO:68. The amino acid sequence containing the Pfa1 KR domain is represented herein as SEQ ID NO:101, corresponding to positions 2270-2986 of SEQ ID NO:69. Within the KR domain is a core region (contained within the nucleotide sequence of SEQ ID NO:116, and the amino acid sequence of SEQ ID NO:117) with homology to short chain aldehyde-dehydrogenases (KR is a member of this family). This core region spans from about position 5998 to about 6900 of SEQ ID NO:68, which corresponds to amino acid positions 2000-2300 of SEQ ID NO:69.

[0289] The fourteenth domain in *Thraustochytrium sp.* ATCC PTA-10212 Pfa1 is a DH domain. The nucleotide sequence containing the sequence encoding the Pfa1 DH domain is represented herein as SEQ ID NO:118, corresponding to positions 7027-7065 of SEQ ID NO:68. The amino acid sequence containing the Pfa1 DH domain is represented herein as SEQ ID NO:119, corresponding to positions 2343-2355 of SEQ ID NO:69. The DH domain contains a conserved active site motif (see, Donadio, S. and Katz., L., Gene 111(1): 51-60 (1992)): LxxHxxxGxxxxP (SEQ ID NO:50).

[0290] Table 10 shows the domains and active sites associated with *Thraustochytrium sp.* ATCC PTA-10212 PFA2.

Table 10: *Thraustochytrium sp.* ATCC PTA-10212 PFA2 Domain Analysis

Domain	DNA positions	AA positions	Sites	DNA positions	AA positions
KS	10-1320 of SEQ ID NO:70 (SEQ ID NO:102)	4-440 of SEQ ID NO:71 (SEQ ID NO:103)	DXAC* (SEQ ID NO:43)	571-573 of SEQ ID NO:70	C191 of SEQ ID NO:71
			End - GFGG (SEQ ID NO:44)	1267-1278 of SEQ ID NO:70	423-426 of SEQ ID NO:71

CLF	1378-2700 of SEQ ID NO:70 (SEQ ID NO:104)	460-900 of SEQ ID NO:71 (SEQ ID NO:105)			
AT	2848-4200 of SEQ ID NO:70 (SEQ ID NO:106)	950-1400 of SEQ ID NO:71 (SEQ ID NO:107)	GxS*xG (SEQ ID NO:52)	3361-3363 of SEQ ID NO:70	S1121 of SEQ ID NO:71
ER	4498-5700 of SEQ ID NO:70 (SEQ ID NO:108)	1500-1900 of SEQ ID NO:71 (SEQ ID NO:109)			

[0291] The first domain in *Thraustochytrium sp.* ATCC PTA-10212 Pfa2 is a KS domain.

The nucleotide sequence containing the sequence encoding the *Thraustochytrium sp.* ATCC PTA-10212 Pfa2 KS domain is represented herein as SEQ ID NO:102, corresponding to positions 10-1320 of SEQ ID NO:70. The amino acid sequence containing the *Thraustochytrium sp.* ATCC PTA-10212 Pfa2 KS domain is represented herein as SEQ ID NO:103, corresponding to positions 4-440 of SEQ ID NO:71. The KS domain contains an active site motif: DXAC* (SEQ ID NO:43), with an *acyl binding cite corresponding to C191 of SEQ ID NO:71. Also, a characteristic motif is present at the end of the KS domain: GFGG (SEQ ID NO:44), corresponding to positions 423-426 of SEQ ID NO:71 and positions 1267-1278 of SEQ ID NO:70.

[0292] The second domain in *Thraustochytrium sp.* ATCC PTA-10212 Pfa2 is a CLF domain. The nucleotide sequence containing the sequence encoding the *Thraustochytrium sp.* ATCC PTA-10212 Pfa2 CLF domain is represented herein as SEQ ID NO:104, corresponding to positions 1378-2700 of SEQ ID NO:70. The amino acid sequence containing the *Thraustochytrium sp.* ATCC PTA-10212 Pfa2 CLF domain is represented herein as SEQ ID NO:105, corresponding to positions 460-900 of SEQ ID NO:71.

[0293] The third domain in *Thraustochytrium sp.* ATCC PTA-10212 Pfa2 is an AT domain. The nucleotide sequence containing the sequence encoding the *Thraustochytrium sp.* ATCC PTA-10212 Pfa2 AT domain is represented herein as SEQ ID NO:106, corresponding to positions 2848-4200 of SEQ ID NO:70. The amino acid sequence containing the *Thraustochytrium sp.* ATCC PTA-10212 Pfa2 AT domain is

represented herein as SEQ ID NO:107, corresponding to positions 950-1400 of SEQ ID NO:71. The AT domain contains an active site motif of GxS*xG (SEQ ID NO:50) that is characteristic of acyltransferase (AT) proteins, with an active site serine residue corresponding to S1121 of SEQ ID NO:71.

[0294] The fourth domain of *Thraustochytrium sp.* ATCC PTA-10212 Pfa2 is an ER domain. The nucleotide sequence containing the sequence encoding the *Thraustochytrium sp.* ATCC PTA-10212 Pfa2 ER domain is represented herein as SEQ ID NO:108, corresponding to positions 4498-5700 of SEQ ID NO:70. The amino acid sequence containing the *Thraustochytrium sp.* ATCC PTA-10212 Pfa2 ER domain is represented herein as SEQ ID NO:109, corresponding to positions 1500-1900 of SEQ ID NO:71.

[0295] Table 11 shows the domains and active sites associated with *Thraustochytrium sp.* ATCC PTA-10212 PFA3.

Table 11: *Thraustochytrium sp.* ATCC PTA-10212 PFA3 Domain Analysis

Domain	DNA positions	AA positions	Sites	DNA positions	AA positions
DH1	1-1350 of SEQ ID NO:72 (SEQ ID NO:110)	1-450 of SEQ ID NO:73 (SEQ ID NO:111)	FxxH*F (SEQ ID NO:53)	934-936 of SEQ ID NO:72	H312 of SEQ ID NO:73
DH2	1501-2700 of SEQ ID NO:72 (SEQ ID NO:112)	501-900 of SEQ ID NO:73 (SEQ ID NO:113)	FxxH*F (SEQ ID NO:53)	2401-2403 of SEQ ID NO:72	H801 of SEQ ID NO:73
ER	2848-4212 of SEQ ID NO:72 (SEQ ID NO:114)	950-1404 of SEQ ID NO:73 (SEQ ID NO:115)			

[0296] The first and second domains of *Thraustochytrium sp.* ATCC PTA-10212 Pfa3 are DH domains, referred to herein as DH1 and DH2, respectively. The nucleotide sequence containing the sequence encoding the *Thraustochytrium sp.* ATCC PTA-10212 Pfa3 DH1 domain is represented herein as SEQ ID NO:110, corresponding to positions 1-1350 of SEQ ID NO:72. The amino acid sequence containing the *Thraustochytrium sp.* ATCC PTA-10212 Pfa3 DH1 domain is represented herein as SEQ ID NO:111, corresponding to positions 1-450 of SEQ ID NO:73. The nucleotide sequence containing the sequence

encoding the *Thraustochytrium sp.* ATCC PTA-10212 Pfa3 DH2 domain is represented herein as SEQ ID NO:112, corresponding to positions 1501-2700 of SEQ ID NO:72. The amino acid sequence containing the *Thraustochytrium sp.* ATCC PTA-10212 Pfa3 DH2 domain is represented herein as SEQ ID NO:113, corresponding to positions 501-900 of SEQ ID NO:73. The DH domains contain an active site motif: FxxH*F (SEQ ID NO:53). The nucleotide sequence containing the active site motif in DH1 corresponds to positions 934-936 of SEQ ID NO:72, while the nucleotide sequence containing the active site motif in DH2 corresponds to positions 2401-2403 of SEQ ID NO:72. The active site H* in the motif FxxH*F is based on data from Leesong et al., Structure 4:253-64 (1996) and Kimber et al. J Biol Chem. 279:52593-602 (2004), with the active site H* in DH1 corresponding to H312 of SEQ ID NO:73 and the active site H* in DH2 corresponding to H801 of SEQ ID NO:73.

[0297] The third domain of *Thraustochytrium sp.* ATCC PTA-10212 Pfa3 is an ER domain. The nucleotide sequence containing the sequence encoding the *Thraustochytrium sp.* ATCC PTA-10212 Pfa3 ER domain is represented herein as SEQ ID NO:114, corresponding to positions 2848-4200 of SEQ ID NO:72. The amino acid sequence containing the *Thraustochytrium sp.* ATCC PTA-10212 Pfa3 ER domain is represented herein as SEQ ID NO:115, corresponding to positions 950-1400 of SEQ ID NO:73.

EXAMPLE 7

[0298] The inactivation of native PUFA synthase genes in *Schizochytrium sp.* ATCC 20888, to generate PUFA auxotrophs, and the replacement of such inactivated genes with exogenously introduced homologous genes to restore PUFA synthesis has been previously demonstrated and described. See, e.g., U.S. Patent No. 7,217,856, incorporated by reference herein in its entirety. The three PUFA synthase genes from *Schizochytrium sp.* ATCC 20888 have been previously termed *orfA*, *orfB*, and *orfC*, corresponding to the *PFA1*, *PFA2*, and *PFA3* nomenclature used herein, respectively. *Id.*

[0299] The native *orfA* gene in *Schizochytrium sp.* ATCC 20888 was replaced by homologous recombination following transformation with a vector containing the ZeocinTM resistance marker surrounded by sequences from the *orfA* flanking region. A

mutant strain was generated lacking a functional *orfA* gene. The mutant strain was auxotrophic and required PUFA supplementation for growth.

[0300] *Schizochytrium* sp. ATCC PTA-9695 *PFA1* (SEQ ID NO:1) was cloned into expression vector pREZ37 to generate pREZ345. The expression vector contained approximately 2 kb of DNA from the flanking region of the native *orfA* gene locus from *Schizochytrium* sp. ATCC 20888. The *Schizochytrium* sp. ATCC 20888 mutant lacking functional *orfA* was transformed via electroporation with enzyme pretreatment (see below) with pREZ345 containing *PFA1*. Based on homologous regions flanking the ZeocinTM resistance marker in the mutant and flanking the *PFA1* gene in pREZ345, double-crossover recombination occurred such that *PFA1* was inserted into the native *orfA* locus. Recombination with *Schizochytrium* sp. ATCC PTA-9695 *PFA1* (SEQ ID NO:1) restored PUFA production in the *Schizochytrium* sp. ATCC 20888 mutant lacking *orfA*. In brief, cells were grown in M2B liquid media (see following paragraph) at 30°C with 200 rpm shaking for 3 days. Cells were harvested and the fatty acids were converted to methyl-esters using standard techniques. Fatty acid profiles were determined using gas chromatography with flame ionization detection (GC-FID) as fatty acid methyl esters (FAME). The native *Schizochytrium* sp. ATCC 20888 strain containing a functional *orfA* gene produced DHA and DPA n-6 in a ratio of 2.3:1. The recombinant strain containing *Schizochytrium* sp. ATCC PTA-9695 *PFA1* (SEQ ID NO:1) in place of the inactivated *orfA* gene also produced DHA and DPA n-6 in a ratio of 2.4:1. The EPA content of the recombinant strain was 2.7% of fatty acid methyl-esters (FAME), the DPA n-3 content was 0.7%, the DPA n-6 content was 8.8%, and the DHA content was 21.2%.

[0301] *M2B medium* - 10 g/L glucose, 0.8 g/L (NH₄)₂SO₄, 5 g/L Na₂SO₄, 2 g/L MgSO₄•7H₂O, 0.5 g/L KH₂PO₄, 0.5 g/L KCl, 0.1 g/L CaCl₂•2H₂O, 0.1 M MES (pH 6.0) 0.1% PB26 metals, and 0.1% PB26 Vitamins (v/v). PB26 vitamins consisted of 50 mg/mL vitamin B12, 100 µg/mL thiamine, and 100 µg/mL Ca-pantothenate. PB26 metals were adjusted to pH 4.5 and consisted of 3 g/L FeSO₄•7H₂O, 1 g/L MnCl₂•4H₂O, 800 mg/mL ZnSO₄•7H₂O, 20 mg/mL CoCl₂•6H₂O, 10 mg/mL Na₂MoO₄•2H₂O, 600 mg/mL CuSO₄•5H₂O, and 800 mg/mL NiSO₄•6H₂O. PB26 stock solutions were filter-sterilized separately and added to the broth after autoclaving. Glucose, KH₂PO₄, and CaCl₂•2H₂O were each autoclaved separately from the remainder of the broth ingredients

before mixing to prevent salt precipitation and carbohydrate caramelizing. All medium ingredients were purchased from Sigma Chemical (St. Louis, MO).

[0302] *Electroporation with enzyme pretreatment* - Cells were grown in 50 mL of M50-20 media (see U.S. Publ. No. 2008/0022422) on a shaker at 200 rpm for 2 days at 30°C. The cells were diluted at 1:100 into M2B media and grown overnight (16-24 h), attempting to reach mid-log phase growth (OD₆₀₀ of 1.5-2.5). The cells were centrifuged in a 50 mL conical tube for 5 min at about 3000 x g. The supernatant was removed and the cells were resuspended in 1 M mannitol, pH 5.5, in a suitable volume to reach a final concentration of 2 OD₆₀₀ units. 5 mL of cells were aliquoted into a 25 mL shaker flask and amended with 10 mM CaCl₂ (1.0 M stock, filter sterilized) and 0.25 mg/mL Protease XIV (10 mg/mL stock, filter sterilized; Sigma-Aldrich, St. Louis, MO). Flasks were incubated on a shaker at 30°C and about 100 rpm for 4 h. Cells were monitored under the microscope to determine the degree of protoplasting, with single cells desired. The cells were centrifuged for 5 min at about 2500 x g in round-bottom tubes (i.e., 14 mL Falcon™ tubes, BD Biosciences, San Jose, CA). The supernatant was removed and the cells were gently resuspended with 5 mL of ice cold 10% glycerol. The cells were re-centrifuged for 5 min at about 2500 x g in round-bottom tubes. The supernatant was removed and the cells were gently resuspended with 500 µL of ice cold 10% glycerol, using wide-bore pipette tips. 90 µL of cells were aliquoted into a prechilled electro-cuvette (Gene Pulser® cuvette - 0.1 cm gap or 0.2 cm gap, Bio-Rad, Hercules, CA). One µg to 5 µg of DNA (in less than or equal to a 10 µL volume) was added to the cuvette, mixed gently with a pipette tip, and placed on ice for 5 min. Cells were electroporated at 200 ohms (resistance), 25 µF (capacitance), and either 250V (for 0.1 cm gap) or 500V (0.2 cm gap). 0.5 mL of M50-20 media was added immediately to the cuvette. The cells were then transferred to 4.5 mL of M50-20 media in a 25 mL shaker flask and incubated for 2-3 h at 30°C and about 100 rpm on a shaker. The cells were centrifuged for 5 min at about 2500 x g in round bottom tubes. The supernatant was removed and the cell pellet was resuspended in 0.5 mL of M50-20 media. Cells were plated onto an appropriate number (2 to 5) of M2B plates with appropriate selection and incubated at 30°C.

- [0303] The *Schizochytrium* sp. ATCC 20888 mutant lacking functional *orfA* is also transformed with pREZ345 containing *PFA1*, such that *PFA1* is randomly integrated in the mutant and restores PUFA production.

EXAMPLE 8

- [0304] *Thraustochytrium* sp. ATCC PTA-10212 *PFA1* (SEQ ID NO:68) was re-synthesized (DNA2.0) and codon-optimized for expression in *Schizochytrium* (SEQ ID NO:120) and was cloned into an expression vector to generate pLR95. Codon-optimization occurred using the *Schizochytrium* codon usage table in FIG. 22. The expression vector contained approximately 2 kb of DNA from the flanking region of the native *orfA* gene locus from *Schizochytrium* sp. ATCC 20888.
- [0305] The *Schizochytrium* sp. ATCC 20888 mutant lacking functional *orfA* from Example 7 was transformed via electroporation with enzyme pretreatment (See Example 7) with pLR95 containing codon-optimized *Thraustochytrium* sp. ATCC PTA-10212 *PFA1* (SEQ ID NO:120). Based on homologous regions flanking the ZeocinTM resistance marker in the mutant and flanking the *PFA1* gene in pLR95, double-crossover recombination occurred such that codon-optimized *Thraustochytrium* sp. ATCC PTA-10212 *PFA1* was inserted into the native *orfA* locus. Recombination with codon-optimized *Thraustochytrium* sp. ATCC PTA-10212 *PFA1* (SEQ ID NO:120) restored PUFA production in the *Schizochytrium* sp. ATCC 20888 mutant lacking *orfA*. Cells were grown and analyzed for FAMES as described in Example 7. The native *Schizochytrium* sp. ATCC 20888 strain containing a functional *orfA* gene produced DHA and EPA in a ratio of 25:1. The recombinant strain containing codon-optimized *Thraustochytrium* sp. ATCC PTA-10212 *PFA1* (SEQ ID NO:120) in place of the inactivated *orfA* gene produced DHA and EPA in a ratio of 5.4:1, further demonstrating that the PUFA profile of *Schizochytrium* can be altered by the nucleic acid molecules described herein. The EPA content of the recombinant strain was 4.4% of FAME, the DPA n-3 content was 2.3%, the DPA n-6 content was 4.9%, and the DHA content was 24.0%.

- [0306] The *Schizochytrium* sp. ATCC 20888 mutant lacking functional *orfA* is also transformed with pLR95 containing *PFA1*, such that *PFA1* is randomly integrated in the mutant and restores PUFA production.

EXAMPLE 9

- [0307] The native *orfB* gene in *Schizochytrium* sp. ATCC 20888 was replaced by homologous recombination following transformation via electroporation with enzyme pretreatment (See Example 7) with a vector containing the ZeocinTM resistance marker surrounded by sequences from the *orfB* flanking region. A mutant strain was generated lacking a functional *orfB* gene. The mutant strain was auxotrophic and required PUFA supplementation for growth.
- [0308] *Schizochytrium* sp. ATCC PTA-9695 *PFA2* (SEQ ID NO:3) was cloned into expression vector pDS04 to generate pREZ331. The expression vector contained approximately 2 kb of DNA from the flanking region of the native *orfB* gene locus from *Schizochytrium* sp. ATCC 20888.
- [0309] The *Schizochytrium* sp. ATCC 20888 mutant lacking functional *orfB* was transformed with pREZ331 containing *PFA2*. Based on random integration in the mutant, PUFA production was restored by *Schizochytrium* sp. ATCC PTA-9695 *PFA2* (SEQ ID NO:3). Cells were grown and analyzed for FAMES as described in Example 7. The native *Schizochytrium* sp. ATCC 20888 strain containing a functional *orfB* gene produced DHA and DPA n-6 in a ratio of 2.3:1. The recombinant strain containing *Schizochytrium* sp. ATCC PTA-9695 *PFA2* (SEQ ID NO:3) as a replacement of the inactivated *orfB* gene produced DHA and DPA n-6 in a ratio of 3.5:1. The EPA content of the recombinant strain was 0.8% of FAME, the DPA n-3 content was 0.1%, the DPA n-6 content was 7.1%, and the DHA content was 25.1%.
- [0310] The *Schizochytrium* sp. ATCC 20888 mutant lacking functional *orfB* is also transformed with pREZ331 containing *PFA2*, such that *PFA2* is inserted into the native *orfB* locus and restores PUFA production.

EXAMPLE 10

[0311] *Thraustochytrium* sp. ATCC PTA-10212 *PFA2* (SEQ ID NO:70) was re-synthesized (DNA2.0) and codon-optimized for expression in *Schizochytrium* (SEQ ID NO:121) and was cloned into an expression vector to generate pLR85. Codon-optimization occurred using the *Schizochytrium* codon usage table in FIG. 22. The expression vector contained approximately 2 kb of DNA from the flanking region of the native *orfB* gene locus from *Schizochytrium* sp. ATCC 20888.

[0312] Replacement of *orf* genes was also studied in a daughter strain of *Schizochytrium* sp. ATCC 20888 having improved DHA productivity. The native *orfB* gene in the daughter strain was replaced by homologous recombination following transformation via electroporation with enzyme pretreatment (See Example 7) with a vector containing the ZeocinTM resistance marker surrounded by sequences from the *orfB* flanking region. A mutant strain was generated lacking a functional *orfB* gene. The mutant strain was auxotrophic and required PUFA supplementation for growth. The mutant strain was transformed via electroporation with enzyme pretreatment (see Example 8) with pLR85 containing codon-optimized *Thraustochytrium* sp. ATCC PTA-10212 *PFA2* (SEQ ID NO:121). Based on homologous regions flanking the ZeocinTM resistance marker in the mutant and flanking the *PFA2* gene in pLR85, double-crossover recombination occurred such that codon-optimized *Thraustochytrium* sp. ATCC PTA-10212 *PFA2* (SEQ ID NO:121) was inserted into the native *orfB* locus of the mutant strain. Recombination with codon-optimized *Thraustochytrium* sp. ATCC PTA-10212 *PFA2* (SEQ ID NO:121) restored PUFA production in the daughter strain mutant lacking *orfB*. Cells were grown and analyzed for FAMES as described in Example 7. The EPA content of the recombinant strain was 1.0% of FAME, the DPA n-3 content was 0.3%, the DPA n-6 content was 7.0%, and the DHA content was 31.0%.

[0313] In an experiment to be performed, the *Schizochytrium* sp. ATCC 20888 mutant lacking functional *orfB* from Example 9 is transformed via electroporation with enzyme pretreatment (see Example 8) with pLR85 containing codon-optimized *Thraustochytrium* sp. ATCC PTA-10212 *PFA2* (SEQ ID NO:121). Based on homologous regions flanking the ZeocinTM resistance marker in the mutant and flanking the *PFA2* gene in pLR85, double-crossover recombination occurs such that codon-optimized *Thraustochytrium* sp. ATCC PTA-10212 *PFA2* (SEQ ID NO:121) is inserted into the native *orfB* locus.

Recombination with codon-optimized *Thraustochytrium* sp. ATCC PTA-10212 *PFA2* (SEQ ID NO:121) restores PUFA production in the *Schizochytrium* sp. ATCC 20888 mutant lacking *orfB*.

- [0314] The *Schizochytrium* sp. ATCC 20888 and daughter strain mutants lacking functional *orfB* are also transformed with pLR85 containing *PFA2*, such that *PFA2* is randomly integrated in the mutants and restores PUFA production in each of the mutants.

EXAMPLE 11

- [0315] A plasmid containing a paromomycin resistance marker cassette functional in *Schizochytrium* was developed for *Schizochytrium* sp. ATCC 20888 by replacement of the bleomycin/ZeocinTM resistance gene (*ble*) coding region in pMON50000/pTUBZEO11-2 (US Patent 7,001,772 B2) with that of neomycin phosphotransferase II (*npt*), originally from bacterial transposon Tn5. In pMON50000, the *ble* resistance gene is driven by the *Schizochytrium* α -tubulin promoter and is followed by the SV40 transcription terminator. The *ble* region in pMON50000 encompasses a *Nco*I restriction site at the ATG start codon and a *Pml*I restriction site immediately following the TGA stop signal. PCR was used to amplify the *npt* coding region present in pCaMVnpt (Shimizu *et al.*, *Plant J.* 26(4):375 (2001)) such that the product included a *Bsp*HI restriction site (underlined below, primer CAX055) at the start ATG (bold) and a *Pml*I restriction site (underlined below, primer CAX056) immediately following the stop signal (bold – reverse complement):

CAX055 (forward): GTCATGATTGAACAAGATGGATTGCAC (SEQ ID NO:66)

CAX056 (reverse): CCACGTGTCAGAAGAAGAACTCGTCAAGAA (SEQ ID NO:67).

- [0316] PCR was carried out with the TaqMaster polymerase kit (5Prime), products were cloned into pCR4-TOPO (Invitrogen), and resulting plasmids were transformed into *E. coli* TOP10 (Invitrogen). DNA sequence analysis using vector primers identified multiple clones containing the desired 805bp structure (i.e., the sequences match those of the source template plus the engineered restriction sites). The modified *npt* coding region was isolated by digestion with *Bsp*HI plus *Pml*I restriction enzymes, and the purified DNA fragment was ligated with a pMON50000 vector fragment generated by digestion

with *NcoI* plus *PmlI* enzymes. Restriction enzymes *BspHI* and *NcoI* leave compatible overlapping ends, and *PmlI* leaves blunt ends. The resulting plasmid, pTS-NPT, contains the *npt* neomycin/paromomycin resistance gene in the identical context as that of the original *ble* gene in pMON50000.

- [0317] Particle bombardment of *Schizochytrium* (US Patent 7,001,772 B2) was used to evaluate the function of the novel paromomycin resistance cassette in pTS-NPT. Selection for paromomycin (PAR) resistance was carried out on agar plates containing 50 µg/mL paromomycin sulfate (Sigma). Paromomycin-resistant *Schizochytrium* transformants were found at frequencies similar to those for Zeocin™-resistance from pMON50000. The " α -tubulin promoter/*npt*/SV40 terminator" cassette can be freed from pTS-NPT with various restriction enzymes for subsequent use in other development efforts.

EXAMPLE 12

- [0318] The native *orfC* gene in *Schizochytrium sp.* ATCC 20888 was replaced by homologous recombination following transformation with a vector containing the paromomycin resistance marker surrounded by sequences from the *orfC* flanking region. A mutant strain was generated lacking a functional *orfC* gene. The mutant strain was auxotrophic and required PUFA supplementation for growth.
- [0319] *Schizochytrium sp.* ATCC PTA-9695 *PFA3* (SEQ ID NO:5) was cloned into expression vector pREZ22 to generate pREZ324. The expression vector contained approximately 2 kb of DNA from the flanking region of the native *orfC* gene locus from *Schizochytrium sp.* ATCC 20888.
- [0320] The *Schizochytrium sp.* ATCC 20888 mutant lacking functional *orfC* was transformed with pREZ324 containing *Schizochytrium sp.* ATCC PTA-9695 *PFA3*. Based on homologous regions flanking the paromomycin resistance marker in the mutant and flanking the *Schizochytrium sp.* ATCC PTA-9695 *PFA3* gene in pREZ324, double-crossover recombination occurred such that *Schizochytrium sp.* ATCC PTA-9695 *PFA3* was inserted into the native *orfC* locus. Homologous recombination with *Schizochytrium sp.* ATCC PTA-9695 *PFA3* (SEQ ID NO:5) restored PUFA production in the *Schizochytrium sp.* ATCC 20888 mutant lacking *orfC*. Cells were grown and analyzed for FAMES as described in Example 7. The native *Schizochytrium sp.* ATCC 20888

strain containing a functional *orfC* gene produced DHA and DPA n-6 in a ratio of 2.3:1. The recombinant strain containing *Schizochytrium* sp. ATCC PTA-9695 *PFA3* (SEQ ID NO:5) in place of the inactivated *orfC* gene produced DHA and DPA n-6 in a ratio of 14:9, further demonstrating that the PUFA profile of *Schizochytrium* can be altered by the nucleic acid molecules described herein. The EPA content of the recombinant strain was 1.2% of FAME, the DPA n-3 content was 0.2%, the DPA n-6 content was 2.9%, and the DHA content was 43.4%.

[0321] The *Schizochytrium* sp. ATCC 20888 mutant lacking functional *orfC* was also transformed with pREZ324 containing *PFA3*, such that *PFA3* was randomly integrated in the mutant and restored PUFA production. The EPA content of the recombinant strain was 1.2% of FAME, the DPA n-3 content was 0.2%, the DPA n-6 content was 2.5%, and the DHA content was 39.1%.

[0322] The native *orfC* gene in the daughter strain discussed in Example 10 was replaced by homologous recombination following transformation with a vector containing the paromomycin resistance marker surrounded by sequences from the *orfC* flanking region. A mutant strain was generated lacking a functional *orfC* gene. The mutant strain was auxotrophic and required PUFA supplementation for growth. The mutant lacking functional *orfC* was transformed with pREZ324. Double-crossover recombination occurred such that *Schizochytrium* sp. ATCC PTA-9695 *PFA3* was inserted into the native *orfC* locus of the mutant strain. Homologous recombination with *Schizochytrium* sp. ATCC PTA-9695 *PFA3* (SEQ ID NO:5) restored PUFA production in the the daughter strain mutant lacking *orfC*. Cells were grown and analyzed for FAMES as described in Example 7. The EPA content of the recombinant strain was 1.2% of FAME, the DPA n-3 content was 0.3%, the DPA n-6 content was 2.8%, and the DHA content was 43.1%.

[0323] The daughter strain mutant lacking functional *orfB* is also transformed with pREZ324 containing *PFA3*, such that *PFA3* is randomly integrated in the mutant and restores PUFA production.

EXAMPLE 13

[0324] *Thraustochytrium* sp. ATCC PTA-10212 *PFA3* (SEQ ID NO:72) was re-synthesized (DNA2.0) and codon-optimized for expression in *Schizochytrium* (SEQ ID

NO:122) and was cloned into expression vector pREZ22 to generate pREZ337. Codon-optimization occurred using the *Schizochytrium* codon usage table in FIG. 22. The expression vector contained approximately 2 kb of DNA from the flanking region of the native *orfC* gene locus from *Schizochytrium sp.* ATCC 20888.

[0325] The daughter strain mutant lacking functional *orfC* from Example 12 was transformed via electroporation with enzyme pretreatment (see Example 8) with pREZ337 containing codon-optimized *Thraustochytrium sp.* ATCC PTA-10212 *PFA3* (SEQ ID NO:122). Based on homologous regions flanking the ZeocinTM resistance marker in the mutant and flanking the *PFA3* gene in pREZ337, double-crossover recombination occurred such that codon-optimized *Thraustochytrium sp.* ATCC PTA-10212 *PFA3* (SEQ ID NO:122) was inserted into the native *orfC* locus. Recombination with codon-optimized *Thraustochytrium sp.* ATCC PTA-10212 *PFA3* (SEQ ID NO:122) restored PUFA production in the daughter strain mutant lacking *orfC*. Cells were grown and analyzed for FAMES as described in Example 7. The EPA content of the recombinant strain was 1.3% of FAME, the DPA n-3 content was 0.4%, the DPA n-6 content was 2.7%, and the DHA content was 50.2%.

[0326] In an experiment to be performed, the *Schizochytrium sp.* ATCC 20888 mutant lacking functional *orfC* from Example 12 is transformed via electroporation with enzyme pretreatment (see Example 8) with pREZ337 containing codon-optimized *Thraustochytrium sp.* ATCC PTA-10212 *PFA3* (SEQ ID NO:122). Based on homologous regions flanking the ZeocinTM resistance marker in the mutant and flanking the *PFA3* gene in pREZ337, double-crossover recombination occurs such that codon-optimized *Thraustochytrium sp.* ATCC PTA-10212 *PFA3* (SEQ ID NO:122) is inserted into the native *orfC* locus. Recombination with codon-optimized *Thraustochytrium sp.* ATCC PTA-10212 *PFA3* (SEQ ID NO:122) restores PUFA production in the *Schizochytrium sp.* ATCC 20888 mutant lacking *orfC*.

[0327] The *Schizochytrium sp.* ATCC 20888 and daughter strain mutants lacking functional *orfC* are also transformed with pREZ337 containing *PFA3*, such that *PFA3* is randomly integrated in the mutants and restores PUFA production in each of the mutants.

EXAMPLE 14

- [0328] Any two or all three of the *orfA*, *orfB*, and *orfC* genes in *Schizochytrium sp.* ATCC 20888 are replaced by homologous recombination following transformation with vectors containing either the ZeocinTM or paromomycin resistance marker surrounded by sequences from the appropriate *orf* flanking region. Mutant strains are generated lacking functional genes for any two or all three of *orfA*, *orfB*, and *orfC*. The mutant strains are auxotrophic and require PUFA supplementation for growth.
- [0329] The *Schizochytrium sp.* ATCC 20888 mutants lacking functional *orf* genes are transformed with one or more expression vectors containing corresponding *PFA* genes (one or more of SED ID NOs: 1, 3, 5, 120, 121, or 122). Based on homologous regions flanking the ZeocinTM or paromomycin resistance markers in the mutants and flanking the *PFA* genes in the respective expression vectors, double-crossover recombination can occur such that *PFA* genes are inserted into the native *orf* loci. Random integration of these expression vectors can also occur with the selection of transformants based solely on the restoration of PUFA production. Homologous recombination with *PFA* genes restores PUFA production in the mutants, such that native PUFA profiles are restored or altered based on the combination of *PFA* genes inserted into the mutants.
- [0330] In one performed experiment, the *Schizochytrium sp.* ATCC 20888 strain from Example 12 lacking a functional *orfC* gene and containing randomly integrated *Schizochytrium sp.* ATCC PTA-9695 *PFA3* (SEQ ID NO:5) was used for replacement of the *orfA* and *orfB* genes. The native *orfA* and *orfB* genes in the strain were replaced by homologous recombination following transformation with a vector containing the ZeocinTM resistance marker surrounded by sequences from the *orfA* and *orfB* flanking regions. A strain was generated lacking functional *orfA*, *orfB*, and *orfC*, and containing randomly integrated *Schizochytrium sp.* ATCC PTA-9695 *PFA3*. The strain was transformed with pREZ345 containing codon-optimized *Schizochytrium sp.* ATCC PTA-9695 *PFA1* (SEQ ID NO:1) and pREZ331 containing codon-optimized *Schizochytrium sp.* ATCC PTA-9695 *PFA2* (SEQ ID NO:3) such that random integration of *PFA1* and *PFA2* occurred. The resulting recombinant strain lacked functional *orfA*, *orfB*, and *orfC* and contained random integrations of *Schizochytrium sp.* ATCC PTA-9695 *PFA1*, *PFA2*, and *PFA3*. Cells were grown and analyzed for FAMES as described in Example 7. The

EPA content of the recombinant strain was 6.6% of FAME, the DPA n-3 content was 0.8%, the DPA n-6 content was 1.6%, and the DHA content was 20.9%.

[0331] In another performed experiment, the daughter strain from Example 12 lacking a functional *orfC* gene and containing *Schizochytrium* sp. ATCC PTA-9695 *PFA3* (SEQ ID NO:5) inserted into the native *orfC* locus was used for replacement of the *orfA* and *orfB* genes. The native *orfA* and *orfB* genes in the strain were replaced by homologous recombination following transformation with a vector containing the paromomycin resistance marker surrounded by sequences from the *orfA* and *orfB* flanking regions. A strain was generated lacking functional *orfA*, *orfB*, and *orfC*, and containing *Schizochytrium* sp. ATCC PTA-9695 *PFA3* inserted into the native *orfC* locus. The strain was transformed with pREZ345 containing codon-optimized *Schizochytrium* sp. ATCC PTA-9695 *PFA1* (SEQ ID NO:1) and pREZ331 containing codon-optimized *Schizochytrium* sp. ATCC PTA-9695 *PFA2* (SEQ ID NO:3). Double-crossover recombinations occurred such that *Schizochytrium* sp. ATCC PTA-9695 *PFA1* was inserted into the native *orfA* locus and *Schizochytrium* sp. ATCC PTA-9695 *PFA2* was inserted into the native *orfB* locus of the strain. The resulting recombinant strain lacked functional *orfA*, *orfB*, and *orfC* and contained *Schizochytrium* sp. ATCC PTA-9695 *PFA1*, *PFA2*, and *PFA3* inserted into the respective *orfA*, *orfB*, and *orfC* loci. Cells were grown and analyzed for FAMES as described in Example 7. The EPA content of the recombinant strain was 7.3% of FAME, the DPA n-3 content was 0.4%, the DPA n-6 content was 1.5%, and the DHA content was 23.9%.

[0332] In another performed experiment, the daughter strain from Example 12 lacking a functional *orfC* gene and containing randomly integrated *Schizochytrium* sp. ATCC PTA-9695 *PFA3* (SEQ ID NO:5) was used for replacement of the *orfA* and *orfB* genes. The native *orfA* and *orfB* genes in the strain were replaced by homologous recombination following transformation with a vector containing the ZeocinTM resistance marker surrounded by sequences from the *orfA* and *orfB* flanking regions. A strain was generated lacking functional *orfA*, *orfB*, and *orfC*, and containing randomly integrated *Schizochytrium* sp. ATCC PTA-9695 *PFA3*. The strain was transformed with pREZ345 containing codon-optimized *Schizochytrium* sp. ATCC PTA-9695 *PFA1* (SEQ ID NO:1) and pREZ331 containing codon-optimized *Schizochytrium* sp. ATCC PTA-9695 *PFA2* (SEQ ID NO:3) such that random integration of *PFA1* and *PFA2* occurred. The resulting

recombinant strain lacked functional *orfA*, *orfB*, and *orfC* and contained random integrations of *Schizochytrium sp.* ATCC PTA-9695 *PFA1*, *PFA2*, and *PFA3*. Cells were grown and analyzed for FAMES as described in Example 7. The EPA content of the recombinant strain was 6.2% of FAME, the DPA n-3 content was 1.3%, the DPA n-6 content was 0.9%, and the DHA content was 16.6%.

[0333] In another performed experiment, the daughter strain from Example 13 lacking a functional *orfC* gene and containing *Schizochytrium sp.* ATCC PTA-10212 *PFA3* (SEQ ID NO:122) inserted into the native *orfC* locus was used for replacement of the *orfA* and *orfB* genes. The native *orfA* and *orfB* genes in the strain were replaced by homologous recombination following transformation with a vector containing the paromomycin resistance marker surrounded by sequences from the *orfA* and *orfB* flanking regions. A strain was generated lacking functional *orfA*, *orfB*, and *orfC*, and containing *Schizochytrium sp.* ATCC PTA-10212 *PFA3* inserted into the native *orfC* locus. The strain was transformed with pLR95 containing codon-optimized *Schizochytrium sp.* ATCC PTA-10212 *PFA1* (SEQ ID NO:120) and pLR85 containing codon-optimized *Schizochytrium sp.* ATCC PTA-10212 *PFA2* (SEQ ID NO:121). Double-crossover recombinations occurred such that *Schizochytrium sp.* ATCC PTA-10212 *PFA1* was inserted into the native *orfA* locus and *Schizochytrium sp.* ATCC PTA-10212 *PFA2* was inserted into the native *orfB* locus of the strain. The resulting recombinant strain lacked functional *orfA*, *orfB*, and *orfC* and contained *Schizochytrium sp.* ATCC PTA-10212 *PFA1*, *PFA2*, and *PFA3* inserted into the respective *orfA*, *orfB*, and *orfC* loci. Cells were grown and analyzed for FAMES as described in Example 7. The EPA content of the recombinant strain was 5.2% of FAME, the DPA n-3 content was 0.6%, the DPA n-6 content was 2.1%, and the DHA content was 47.1%.

[0334] In another performed experiment, the daughter strain from Example 13 lacking a functional *orfC* gene and containing randomly integrated *Schizochytrium sp.* ATCC PTA-10212 *PFA3* (SEQ ID NO:122) was used for replacement of the *orfA* and *orfB* genes. The native *orfA* and *orfB* genes in the strain were replaced by homologous recombination following transformation with a vector containing the ZeocinTM resistance marker surrounded by sequences from the *orfA* and *orfB* flanking regions. A strain was generated lacking functional *orfA*, *orfB*, and *orfC*, and containing randomly integrated *Schizochytrium sp.* ATCC PTA-10212 *PFA3*. The strain was transformed with pLR95

containing codon-optimized *Schizochytrium* sp. ATCC PTA-10212 *PFA1* (SEQ ID NO:120) and pLR85 containing codon-optimized *Schizochytrium* sp. ATCC PTA-10212 *PFA2* (SEQ ID NO:121) such that random integration of *PFA1* and *PFA2* occurred. The resulting recombinant strain lacked functional *orfA*, *orfB*, and *orfC* and contained random integrations of *Schizochytrium* sp. ATCC PTA-10212 *PFA1*, *PFA2*, and *PFA3*. Cells were grown and analyzed for FAMES as described in Example 7. The EPA content of the recombinant strain was 1.8% of FAME, the DPA n-3 content was 1.8%, the DPA n-6 content was 2.3%, and the DHA content was 34.1%.

EXAMPLE 15

[0335] The *orfA*, *orfB*, and *orfC* genes from *Schizochytrium* sp. ATCC 20888 were cloned into a series of Duet vectors (Novagen). The Duet expression vectors are a set of compatible plasmids in which multiple target genes are cloned and co-expressed from the T7 inducible promoter in *E. coli*. Duet plasmid pREZ91 contained *Schizochytrium* sp. ATCC 20888 *orfA* in pETDuet-1; duet plasmid pREZ96 contained *Schizochytrium* sp. ATCC 20888 *orfB* in pCDFDuet-1; and duet plasmid pREZ101 contained *Schizochytrium* sp. ATCC 20888 *orfC* in pCOLADuet-1. Duet plasmids pREZ91, pREZ96, and pREZ101, along with plasmid pJK737, which contained the required accessory gene *HetI* (described in U.S. Patent No. 7,217,856, incorporated by reference herein in its entirety), were transformed into *E. coli* strain BLR(DE3), which contains an inducible T7 RNA polymerase gene. Upon cell growth and addition of IPTG, according to manufacturer's instructions (Novagen), DHA and DPA n-6 were produced. Briefly, 1 mM IPTG was added for induction when cells reached an optical density of about 0.5 at 600 nm. Cells were grown for 12 hours at 30°C in Luria broth and harvested. The fatty acids were converted to methyl-esters using standard techniques. Fatty acid profiles were determined using gas chromatography with flame ionization detection (GC-FID) as fatty acid methyl esters (FAME).

[0336] The *Schizochytrium* sp. ATCC PTA-9695 *PFA1* (SEQ ID NO:1) gene was cloned into the expression vector pETDuet-1, generating pREZ346. Duet plasmids pREZ346 (containing *Schizochytrium* sp. ATCC PTA-9695 *PFA1*), pREZ96 (containing *orfB*), and pREZ101 (containing *orfC*) were transformed into *E. coli* strain BLR(DE3) along with pJK737 (containing *HetI*). The *Schizochytrium* sp. ATCC PTA-9695 *PFA1* gene was

coexpressed with the *Schizochytrium* sp. ATCC 20888 *orfB* and *orfC* genes. The expression of *Schizochytrium* sp. ATCC PTA-9695 *PFA1*, in combination with *Schizochytrium* sp. ATCC 20888 *orfB* and *orfC*, supported DHA production in *E. coli* under induction conditions. The DHA content of the transformed *E. coli* was 2.8% of FAME, the DPA n-6 content was 1.1%, the DPA n-3 content was 0.6%, and the EPA content was 3.7%.

EXAMPLE 16

[0337] The codon-optimized *Thraustochytrium* sp. ATCC PTA-10212 *PFA1* (SEQ ID NO:120) gene was cloned into the expression vector pETDuet-1, generating pLR100. Duet plasmids pLR100 (containing codon-optimized *Thraustochytrium* sp. ATCC PTA-10212 *PFA1*), pREZ96 (containing *Schizochytrium* sp. ATCC 20888 *orfB*), and pREZ101 (containing *Schizochytrium* sp. ATCC 20888 *orfC*) are transformed into *E. coli* strain BLR(DE3) along with pJK737 (containing *HetI*). See Example 15. The *Thraustochytrium* sp. ATCC PTA-10212 *PFA1* gene is coexpressed with the *Schizochytrium* sp. ATCC 20888 *orfB* and *orfC* genes. The expression of *Thraustochytrium* sp. ATCC PTA-10212 *PFA1*, in combination with *Schizochytrium* sp. ATCC 20888 *orfB* and *orfC*, supports DHA and EPA production in *E. coli* under induction conditions.

EXAMPLE 17

[0338] The *Schizochytrium* sp. ATCC PTA-9695 *PFA3* (SEQ ID NO:5) gene was cloned into the expression vector pCOLADuet-1, generating pREZ326. Duet plasmids pREZ326 (containing *Schizochytrium* sp. ATCC PTA-9695 *PFA3*), pREZ91 (containing *Schizochytrium* sp. ATCC 20888 *orfA*), and pREZ96 (containing *Schizochytrium* sp. ATCC 20888 *orfB*) were transformed into *E. coli* strain BLR(DE3) along with pJK737 (containing *HetI*). See Example 15. The expression of *Schizochytrium* sp. ATCC PTA-9695 *PFA3*, in combination with *Schizochytrium* sp. ATCC 20888 *orfA* and *orfB*, supported DHA production in *E. coli* under induction conditions. Cells were grown and analyzed for FAMES as described in Example 15. The DHA content of the transformed *E. coli* was 0.3% of FAME.

EXAMPLE 18

[0339] The codon-optimized *Thraustochytrium* sp. ATCC PTA-10212 *PFA3* (SEQ ID NO:122) gene was cloned into the expression vector pCOLADuet-1, generating pREZ348. Duet plasmids pREZ348 (containing codon-optimized *Thraustochytrium* sp. ATCC PTA-10212 *PFA3*), pREZ91 (containing *Schizochytrium* sp. ATCC 20888 *orfA*), and pREZ96 (containing *Schizochytrium* sp. ATCC 20888 *orfB*) were transformed into *E. coli* strain BLR(DE3) along with pJK737 (containing *HetI*). See Example 15. The expression of *Thraustochytrium* sp. ATCC PTA-10212 *PFA3*, in combination with *Schizochytrium* sp. ATCC 20888 *orfA* and *orfB*, supported DHA production in *E. coli* under induction conditions. Cells were grown and analyzed for FAMES as described in Example 15. The DHA content of the transformed *E. coli* was 2.9% of FAME and the DPA n-6 content was 0.4%.

EXAMPLE 19

[0340] The *Schizochytrium* sp. ATCC PTA-9695 *PFA2* (SEQ ID NO:3) gene was cloned into the expression vector pCDFDuet-1, generating pREZ330. Duet plasmids pREZ330 (containing *Schizochytrium* sp. ATCC PTA-9695 *PFA2*), pREZ326 (containing *Schizochytrium* sp. ATCC PTA-9695 *PFA3*), and pREZ91 (containing *Schizochytrium* sp. ATCC 20888 *orfA*), were transformed into *E. coli* strain BLR(DE3) along with pJK737 (containing *HetI*). See Example 9. The expression of *Schizochytrium* sp. ATCC PTA-9695 *PFA2* and *PFA3*, in combination with *Schizochytrium* sp. ATCC 20888 *orfA*, supported DHA production in *E. coli* under induction conditions. Cells were grown and analyzed for FAMES as described in Example 15. The DHA content of the transformed *E. coli* was 0.8% of FAME and the DPA n-6 content was 0.2%.

EXAMPLE 20

[0341] The codon-optimized *Thraustochytrium* sp. ATCC PTA-10212 *PFA2* (SEQ ID NO:121) gene was cloned into the expression vector pCDFDuet-1, generating pLR87. Duet plasmids pLR87 (containing codon-optimized *Thraustochytrium* sp. ATCC PTA-10212 *PFA2*), pREZ348 (containing codon-optimized *Thraustochytrium* sp. ATCC PTA-10212 *PFA3*), and pREZ91 (containing *Schizochytrium* sp. ATCC 20888 *orfA*), were

transformed into *E. coli* strain BLR(DE3) along with pJK737 (containing *HetI*). See Example 15. The expression of codon-optimized *Thraustochytrium* sp. ATCC PTA-10212 *PFA2* and *PFA3*, in combination with *Schizochytrium* sp. ATCC 20888 *orfA*, supported DHA and low levels of EPA production in *E. coli* under induction conditions. Cells were grown and analyzed for FAMES as described in Example 15. The DHA content of the transformed *E. coli* was 4.4% of FAME, the DPA n-6 content was 1.1%, and the EPA content was 0.1%.

EXAMPLE 21

[0342] Duet plasmids pREZ346 (containing *Schizochytrium* sp. ATCC PTA-9695 *PFA1*), pREZ330 (containing *Schizochytrium* sp. ATCC PTA-9695 *PFA2*), and pREZ326 (containing *Schizochytrium* sp. ATCC PTA-9695 *PFA3*) were transformed into *E. coli* strain BLR(DE3) along with pJK737 (containing *HetI*). See Example 15. The expression of *Schizochytrium* sp. ATCC PTA-9695 *PFA1*, *PFA2*, and *PFA3* supported DHA production in *E. coli* under induction conditions. Cells were grown and analyzed for FAMES as described in Example 15. The DHA content of the transformed *E. coli* was 0.3% of FAME and the EPA content was 0.3%.

EXAMPLE 22

[0343] Duet plasmids pLR100 (containing codon-optimized *Thraustochytrium* sp. ATCC PTA-10212 *PFA1*), pLR87 (containing codon-optimized *Thraustochytrium* sp. ATCC PTA-10212 *PFA2*), and pREZ348 (containing codon-optimized *Thraustochytrium* sp. ATCC PTA-10212 *PFA3*) are transformed into *E. coli* strain BLR(DE3) along with pJK737 (containing *HetI*). See Example 15. The expression of codon-optimized *Thraustochytrium* sp. ATCC PTA-10212 *PFA1*, *PFA2*, and *PFA3* supports DHA and EPA production in *E. coli* under induction conditions.

EXAMPLE 23

[0344] Duet plasmids pREZ330 (containing *Schizochytrium* sp. ATCC PTA-9695 *PFA2*), pREZ91 (containing *Schizochytrium* sp. ATCC 20888 *orfA*), and pREZ101 (containing *Schizochytrium* sp. ATCC 20888 *orfC*) were transformed into *E. coli* strain BLR(DE3)

along with pJK737 (containing *HetI*). See Example 15. The expression of *Schizochytrium* sp. ATCC PTA-9695 *PFA2*, in combination with *Schizochytrium* sp. ATCC 20888 *orfA* and *orfC*, supported DHA production in *E. coli* under induction conditions. Cells were grown and analyzed for FAMES as described in Example 15. The DHA content of the transformed *E. coli* was 0.6% of FAME and the DPA n-6 content was 0.3%.

EXAMPLE 24

[0345] Duet plasmids pLR87 (containing codon-optimized *Thraustochytrium* sp. ATCC PTA-10212 *PFA2*), pREZ91 (containing *Schizochytrium* sp. ATCC 20888 *orfA*), and pREZ101 (containing *Schizochytrium* sp. ATCC 20888 *orfC*) were transformed into *E. coli* strain BLR(DE3) along with pJK737 (containing *HetI*). See Example 15. The expression of codon-optimized *Thraustochytrium* sp. ATCC PTA-10212 *PFA2*, in combination with *Schizochytrium* sp. ATCC 20888 *orfA* and *orfC*, supported DHA and low levels of EPA production in *E. coli* under induction conditions. Cells were grown and analyzed for FAMES as described in Example 15. The DHA content of the transformed *E. coli* was 1.7% of FAME, the DPA n-6 content was 0.9%, and the EPA content was 0.1%.

EXAMPLE 25

[0346] Duet plasmids pREZ346 (containing *Schizochytrium* sp. ATCC PTA-9695 *PFA1*), pREZ330 (containing *Schizochytrium* sp. ATCC PTA-9695 *PFA2*), and pREZ101 (containing *Schizochytrium* sp. ATCC 20888 *orfC*) were transformed into *E. coli* strain BLR(DE3) along with pJK737 (containing *HetI*). See Example 15. The expression of *PFA1* and *PFA2*, in combination with *Schizochytrium* sp. ATCC 20888 *orfC*, supported DHA production in *E. coli* under induction conditions. Cells were grown and analyzed for FAMES as described in Example 15. The DHA content of the transformed *E. coli* was 0.3% of FAME, the DPA n-6 content was 0.1%, and the EPA content was 0.5%.

EXAMPLE 26

[0347] Duet plasmids pLR100 (containing codon-optimized *Thraustochytrium* sp. ATCC PTA-10212 *PFA1*), pLR87 (containing codon-optimized *Thraustochytrium* sp. ATCC PTA-10212 *PFA2*), and pREZ101 (containing *Schizochytrium* sp. ATCC 20888 *orfC*) are transformed into *E. coli* strain BLR(DE3) along with pJK737 (containing *HetI*). See Example 15. The expression of codon-optimized *Thraustochytrium* sp. ATCC PTA-10212 *PFA1* and *PFA2*, in combination with *Schizochytrium* sp. ATCC 20888 *orfC*, supports DHA and EPA production in *E. coli* under induction conditions.

EXAMPLE 27

[0348] Duet plasmids pREZ346 (containing *Schizochytrium* sp. ATCC PTA-9695 *PFA1*), pREZ96 (containing *Schizochytrium* sp. ATCC 20888 *orfB*), and pREZ326 (containing *Schizochytrium* sp. ATCC PTA-9695 *PFA3*) were transformed into *E. coli* strain BLR(DE3) along with pJK737 (containing *HetI*). See Example 15. The expression of *Schizochytrium* sp. ATCC PTA-9695 *PFA1* and *PFA3*, in combination with *Schizochytrium* sp. ATCC 20888 *orfB*, supported DHA production in *E. coli* under induction conditions. Cells were grown and analyzed for FAMES as described in Example 15. The DHA content of the transformed *E. coli* was 0.1% of FAME and the EPA content was 0.1%.

EXAMPLE 28

[0349] Duet plasmids pLR100 (containing codon-optimized *Thraustochytrium* sp. ATCC PTA-10212 *PFA1*), pREZ96 (containing *Schizochytrium* sp. ATCC 20888 *orfB*), and pREZ348 (containing codon-optimized *Thraustochytrium* sp. ATCC PTA-10212 *PFA3*) are transformed into *E. coli* strain BLR(DE3) along with pJK737 (containing *HetI*). See Example 15. The expression of codon-optimized *Thraustochytrium* sp. ATCC PTA-10212 *PFA1* and *PFA3*, in combination with *Schizochytrium* sp. ATCC 20888 *orfB*, supports DHA and EPA production in *E. coli* under induction conditions.

EXAMPLE 29

- [0350] Pfa1p, Pfa2p, and Pfa3p PUFA synthase activities in *Schizochytrium sp.* ATCC PTA-9695 and *Thraustochytrium sp.* ATCC PTA-10212 are individually knocked-out by standard procedures. See, e.g., U.S. Patent No. 7,217,856, incorporated by reference herein in its entirety.
- [0351] The ZeocinTM, hygromycin, blasticidin, or other appropriate resistance marker is inserted into a restriction site of the *PFA1* gene (SEQ ID NO:1 or SEQ ID NO:68) that is contained in a plasmid. Following insertion of the resistance marker, the plasmid is introduced into *Schizochytrium sp.* ATCC PTA-9695 or *Thraustochytrium sp.* ATCC PTA-10212, respectively, by particle bombardment, electroporation, or other appropriate transformation method. Homologous recombination occurs, generating mutants in which the native *PFA1* gene is either replaced or disrupted by the ZeocinTM, hygromycin, blasticidin, or other appropriate resistance marker. Transformants are selected on plates containing ZeocinTM, hygromycin, blasticidin, or other appropriate selection agent, supplemented with PUFAs. Colonies are further examined for the capacity to grow in the absence of PUFA supplementation. Genomic DNA is isolated from the colonies that are resistant to the selection agent and unable to grow in the absence of PUFA supplementation. PCR and Southern Blot analysis of the DNA is performed to demonstrate that the *PFA1* gene is either deleted or disrupted.
- [0352] *PFA2* is knocked-out by similar procedures. Resultant knock-out mutants requiring PUFA supplementation are found to lack full-length *PFA2*.
- [0353] *PFA3* is knocked-out by similar procedures. Resultant knock-out mutants requiring PUFA supplementation are found to lack full-length *PFA3*.
- [0354] All of the various aspects, embodiments, and options described herein can be combined in any and all variations.
- [0355] All publications, patents, and patent applications mentioned in this specification are herein incorporated by reference to the same extent as if each individual publication, patent, or patent application was specifically and individually indicated to be incorporated by reference.

WHAT IS CLAIMED IS:

1. An isolated nucleic acid molecule selected from the group consisting of:
 - (a) a nucleic acid molecule comprising a polynucleotide sequence at least 80% identical to SEQ ID NO:1, wherein the polynucleotide sequence encodes a polypeptide comprising polyunsaturated fatty acid (PUFA) synthase activity selected from the group consisting of beta-keto acyl-ACP synthase (KS) activity, malonyl-CoA:ACP acyltransferase (MAT) activity, acyl carrier protein (ACP) activity, ketoreductase (KR) activity, beta-hydroxyacyl-ACP dehydrase (DH) activity, and combinations thereof;
 - (b) a nucleic acid molecule comprising a polynucleotide sequence at least 80% identical to SEQ ID NO:7, wherein the polynucleotide sequence encodes a polypeptide comprising KS activity;
 - (c) a nucleic acid molecule comprising a polynucleotide sequence at least 80% identical to SEQ ID NO:9, wherein the polynucleotide sequence encodes a polypeptide comprising MAT activity;
 - (d) a nucleic acid molecule comprising a polynucleotide sequence at least 80% identical to any one of SEQ ID NOs:13, 15, 17, 19, 21, or 23, wherein the polynucleotide sequence encodes a polypeptide comprising ACP activity;
 - (e) a nucleic acid molecule comprising a polynucleotide sequence at least 80% identical to SEQ ID NO:11, wherein the polynucleotide sequence encodes a polypeptide comprising ACP activity;
 - (f) a nucleic acid molecule comprising a polynucleotide sequence at least 80% identical to SEQ ID NO:25, wherein the polynucleotide sequence encodes a polypeptide comprising KR activity; and
 - (g) a nucleic acid molecule comprising a polynucleotide sequence at least 80% identical to SEQ ID NO:27, wherein the polynucleotide sequence encodes a polypeptide comprising DH activity.

2. The isolated nucleic acid molecule of claim 1, wherein the polynucleotide sequences are at least 90% identical to SEQ ID NOs:1, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, and 27, respectively.
3. The isolated nucleic acid molecule of claim 1, wherein the polynucleotide sequences are at least 95% identical to SEQ ID NOs:1, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, and 27, respectively.
4. The isolated nucleic acid molecule of claim 1, wherein the nucleic acid molecules comprise the polynucleotide sequences of SEQ ID NOs:1, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, and 27, respectively.
5. An isolated nucleic acid molecule selected from the group consisting of:
 - (a) a nucleic acid molecule comprising a polynucleotide sequence encoding a polypeptide, wherein the polypeptide comprises an amino acid sequence at least 80% identical to SEQ ID NO:2, and wherein the polypeptide comprises a PUFA synthase activity selected from the group consisting of KS activity, MAT activity, ACP activity, KR activity, DH activity, and combinations thereof;
 - (b) a nucleic acid molecule comprising a polynucleotide sequence encoding a polypeptide, wherein the polypeptide comprises an amino acid sequence at least 80% identical to SEQ ID NO:8, and wherein the polypeptide comprises KS activity;
 - (c) a nucleic acid molecule comprising a polynucleotide sequence encoding a polypeptide, wherein the polypeptide comprises an amino acid sequence at least 80% identical to SEQ ID NO:10, and wherein the polypeptide comprises MAT activity;
 - (d) a nucleic acid molecule comprising a polynucleotide sequence encoding a polypeptide, wherein the polypeptide comprises an amino acid sequence at least 80% identical to any one of SEQ ID NOs:14, 16, 18, 20, or 24, and wherein the polypeptide comprises ACP activity;

- (e) a nucleic acid molecule comprising a polynucleotide sequence encoding a polypeptide, wherein the polypeptide comprises an amino acid sequence at least 80% identical to SEQ ID NO:12, and wherein the polypeptide comprises ACP activity;
 - (f) a nucleic acid molecule comprising a polynucleotide sequence encoding a polypeptide, wherein the polypeptide comprises an amino acid sequence at least 80% identical to SEQ ID NO:26, and wherein the polypeptide comprises KR activity; and
 - (g) a nucleic acid molecule comprising a polynucleotide sequence encoding a polypeptide, wherein the polypeptide comprises an amino acid sequence at least 80% identical to SEQ ID NO:28, and wherein the polypeptide comprises DH activity.
- 6. The isolated nucleic acid molecule of claim 5, wherein the amino acid sequences are at least 90% identical to SEQ ID NOs:2, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, and 28, respectively.
 - 7. The isolated nucleic acid molecule of claim 5, wherein the amino acid sequences are at least 95% identical to SEQ ID NOs:2, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, and 28, respectively.
 - 8. The isolated nucleic acid molecule of claim 5, wherein the polypeptides comprise the amino acid sequences of SEQ ID NOs:2, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, and 28, respectively.
 - 9. An isolated nucleic acid molecule selected from the group consisting of:
 - (a) a nucleic acid molecule comprising a polynucleotide sequence at least 80% identical to SEQ ID NO:3, wherein the polynucleotide sequence encodes a polypeptide comprising PUFA synthase activity selected from the group consisting of KS activity, chain length factor (CLF) activity, acyltransferase (AT) activity, enoyl-ACP reductase (ER) activity, and combinations thereof;

- (b) a nucleic acid molecule comprising a polynucleotide sequence at least 80% identical to SEQ ID NO:29, wherein the polynucleotide sequence encodes a polypeptide comprising KS activity;
 - (c) a nucleic acid molecule comprising a polynucleotide sequence at least 80% identical to SEQ ID NO:31, wherein the polynucleotide sequence encodes a polypeptide comprising CLF activity;
 - (d) a nucleic acid molecule comprising a polynucleotide sequence at least 80% identical to SEQ ID NO:33, wherein the polynucleotide sequence encodes a polypeptide comprising AT activity; and
 - (e) a nucleic acid molecule comprising a polynucleotide sequence at least 80% identical to SEQ ID NO:35 wherein the polynucleotide sequence encodes a polypeptide comprising ER activity;
10. The isolated nucleic acid molecule of claim 9, wherein the polynucleotide sequences are at least 90% identical to SEQ ID NOs:3, 29, 31, 33, and 35, respectively.
11. The isolated nucleic acid molecule of claim 9, wherein the polynucleotide sequences are at least 95% identical to SEQ ID NOs:3, 29, 31, 33, and 35, respectively.
12. The isolated nucleic acid molecule of claim 9, wherein the nucleic acid molecules comprise the polynucleotide sequences of SEQ ID NOs:3, 29, 31, 33, and 35, respectively.
13. An isolated nucleic acid molecule selected from the group consisting of:
- (a) a nucleic acid molecule comprising a polynucleotide sequence encoding a polypeptide, wherein the polypeptide comprises an amino acid sequence at least 80% identical to SEQ ID NO:4, and wherein the polypeptide comprises a PUFA synthase activity selected from the group consisting of KS activity, CLF activity, AT activity, ER activity, and combinations thereof;
 - (b) a nucleic acid molecule comprising a polynucleotide sequence encoding a polypeptide, wherein the polypeptide comprises an amino acid sequence at least

- 80% identical to SEQ ID NO:30, and wherein the polypeptide comprises KS activity;
- (c) a nucleic acid molecule comprising a polynucleotide sequence encoding a polypeptide, wherein the polypeptide comprises an amino acid sequence at least 80% identical to SEQ ID NO:32, and wherein the polypeptide comprises CLF activity;
 - (d) a nucleic acid molecule comprising a polynucleotide sequence encoding a polypeptide, wherein the polypeptide comprises an amino acid sequence at least 80% identical to SEQ ID NO:34, and wherein the polypeptide comprises AT activity; and
 - (e) a nucleic acid molecule comprising a polynucleotide sequence encoding a polypeptide, wherein the polypeptide comprises an amino acid sequence at least 80% identical to SEQ ID NO:36, and wherein the polypeptide comprises ER activity.
14. The isolated nucleic acid molecule of claim 13, wherein the amino acid sequences are at least 90% identical to SEQ ID NOs:4, 30, 32, 34, and 36, respectively.
 15. The isolated nucleic acid molecule of claim 13, wherein the amino acid sequences are at least 95% identical to SEQ ID NOs:4, 30, 32, 34, and 36, respectively.
 16. The isolated nucleic acid molecule of claim 13, wherein the polypeptides comprise the amino acid sequence of SEQ ID NOs:4, 30, 32, 34, and 36, respectively.
 17. An isolated nucleic acid molecule selected from the group consisting of:
 - (a) an nucleic acid molecule comprising a polynucleotide sequence at least 80% identical to SEQ ID NO:5, wherein the polynucleotide sequence encodes a polypeptide comprising PUFA synthase activity selected from the group consisting of DH activity, ER activity, and combinations thereof;

- (b) a nucleic acid molecule comprising a polynucleotide sequence at least 80% identical to SEQ ID NO:37, wherein the polynucleotide sequence encodes a polypeptide comprising DH activity;
 - (c) a nucleic acid molecule comprising a polynucleotide sequence at least 80% identical to SEQ ID NO:39, wherein the polynucleotide sequence encodes a polypeptide comprising DH activity; and
 - (d) a nucleic acid molecule comprising a polynucleotide sequence at least 80% identical to SEQ ID NO:41, wherein the polynucleotide sequence encodes a polypeptide comprising ER activity.
18. The isolated nucleic acid molecule of claim 17, wherein the polynucleotide sequences are at least 90% identical to SEQ ID NOs:5, 37, 39, and 41, respectively.
19. The isolated nucleic acid molecule of claim 17, wherein the polynucleotide sequences are at least 95% identical to SEQ ID NOs:5, 37, 39, and 41, respectively.
20. The isolated nucleic acid molecule of claim 17, wherein the nucleic acid molecules comprise the polynucleotide sequences of SEQ ID NOs:5, 37, 39, and 41, respectively.
21. An isolated nucleic acid molecule selected from the group consisting of:
- (a) a nucleic acid molecule comprising a polynucleotide sequence encoding a polypeptide, wherein the polypeptide comprises an amino acid sequence at least 80% identical to SEQ ID NO:6, wherein the polypeptide comprises a PUFA synthase activity selected from the group consisting of DH activity, ER activity, and combinations thereof;
 - (b) a nucleic acid molecule comprising a polynucleotide sequence encoding a polypeptide, wherein the polypeptide comprises an amino acid sequence at least 80% identical to SEQ ID NO:38, and wherein the polypeptide comprises DH activity;
 - (c) a nucleic acid molecule comprising a polynucleotide sequence encoding a polypeptide, wherein the polypeptide comprises an amino acid sequence at least

80% identical to SEQ ID NO:40, and wherein the polypeptide comprises DH activity; and

- (d) a nucleic acid molecule comprising a polynucleotide sequence encoding a polypeptide, wherein the polypeptide comprises an amino acid sequence at least 80% identical to SEQ ID NO:42, and wherein the polypeptide comprises ER activity.

22. The isolated nucleic acid molecule of claim 21, wherein the amino acid sequences are at least 90% identical to SEQ ID NOs:6, 38, 40, and 42, respectively.
23. The isolated nucleic acid molecule of claim 21, wherein the amino acid sequences are at least 95% identical to SEQ ID NOs:6, 38, 40, and 42, respectively.
24. The isolated nucleic acid molecule of claim 21, wherein the polypeptides comprise the amino acid sequences of SEQ ID NOs:6, 38, 40, and 42, respectively.
25. An isolated nucleic acid molecule comprising a polynucleotide sequence encoding a polypeptide comprising PUFA synthase activity selected from the group consisting of KS activity, MAT activity, ACP activity, KR activity, CLF activity, AT activity, ER activity, DH activity, and combinations thereof,

wherein the polynucleotide hybridizes under stringent conditions to the complement of the polynucleotide sequence of any one of claims 1-24.

26. An isolated nucleic acid molecule comprising a polynucleotide sequence that is fully complementary to the polynucleotide sequence of any one of claims 1-24.
27. A recombinant nucleic acid molecule comprising the nucleic acid molecule of any one of claims 1-25 or combinations thereof and a transcription control sequence.
28. The recombinant nucleic acid molecule of claim 27, wherein the recombinant nucleic acid molecule is a recombinant vector.
29. A host cell that expresses the nucleic acid molecule of any one of claims 1-25, the recombinant nucleic acid molecule of claim 27 or claim 28, or combinations thereof.

30. The host cell of claim 29, wherein the host cell is selected from the group consisting of a plant cell, a microbial cell, and an animal cell.
31. The host cell of claim 30, wherein the host cell is a microbial cell.
32. The microbial cell of claim 31, wherein the microbial cell is a bacterium.
33. The microbial cell of claim 31, wherein the microbial cell is a thraustochytrid.
34. The microbial cell of claim 33, wherein the thraustochytrid is a *Schizochytrium* or a *Thraustochytrium*.
35. A method to produce at least one PUFA, comprising:
 - expressing a PUFA synthase gene in a host cell under conditions effective to produce PUFA,
 - wherein the PUFA synthase gene comprises the isolated nucleic acid molecule of any one of claims 1-25, the recombinant nucleic acid molecule of claim 27 or claim 28, or combinations thereof, and
 - wherein at least one PUFA is produced.
36. The method of claim 35, wherein the host cell is selected from the group consisting of a plant cell, an isolated animal cell, and a microbial cell.
37. The method of claim 35, wherein the at least one PUFA comprises docosahexaenoic acid (DHA).
38. A method to produce lipids enriched for DHA, comprising:
 - expressing a PUFA synthase gene in a host cell under conditions effective to produce lipids, wherein the PUFA synthase gene comprises the isolated nucleic acid molecule of any one of claims 1-25, the recombinant nucleic acid molecule of claim 27 or claim 28, or combinations thereof in the host cell, and
 - wherein lipids enriched with DHA are produced.
39. A method of making a recombinant vector comprising inserting the isolated nucleic acid molecule of any one of claims 1-25 into a vector.

40. A method of making a recombinant host cell comprising introducing the recombinant vector of claim 39 into a host cell.
41. The method of claim 40, wherein the host cell is selected from the group consisting of a plant cell, an isolated animal cell, and a microbial cell.
42. An isolated polypeptide encoded by the polynucleotide sequence of any one of claims 1-25.
43. An isolated polypeptide selected from the group consisting of:
 - (a) a polypeptide comprising an amino acid sequence at least 80% identical to SEQ ID NO:2, wherein the polypeptide comprises a PUFA synthase activity selected from the group consisting of KS activity, MAT activity, ACP activity, KR activity, DH activity, and combinations thereof;
 - (b) a polypeptide comprising an amino acid sequence at least 80% identical to SEQ ID NO:8, wherein the polypeptide comprises KS activity;
 - (c) a polypeptide comprising an amino acid sequence at least 80% identical to SEQ ID NO:10, wherein the polypeptide comprises MAT activity;
 - (d) a polypeptide comprising an amino acid sequence at least 80% identical to any one of SEQ ID NOs:14, 16, 18, 20, 22, or 24, wherein the polypeptide comprises ACP activity;
 - (e) a polypeptide comprising an amino acid sequence at least 80% identical to SEQ ID NO:12, wherein the polypeptide comprises ACP activity;
 - (f) a polypeptide comprising an amino acid sequence at least 80% identical to SEQ ID NO:26, wherein the polypeptide comprises KR activity; and
 - (g) a polypeptide comprising an amino acid sequence at least 80% identical to SEQ ID NO:28, wherein the polypeptide comprises DH activity.
44. The isolated polypeptide of claim 43, wherein the amino acid sequences are at least 90% identical to SEQ ID NOs:2, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, and 28, respectively.

45. The isolated polypeptide of claim 43, wherein the amino acid sequences are at least 95% identical to SEQ ID NOs:2, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, and 28, respectively.
46. The isolated polypeptide of claim 43, wherein the polypeptides comprise the amino acid sequences of SEQ ID NOs:2, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, and 28, respectively.
47. An isolated polypeptide selected from the group consisting of:
 - (a) a polypeptide comprising an amino acid sequence at least 80% identical to SEQ ID NO:4, wherein the polypeptide comprises a PUFA synthase activity selected from the group consisting of KS activity, CLF activity, AT activity, ER activity, and combinations thereof;
 - (b) a polypeptide comprising an amino acid sequence at least 80% identical to SEQ ID NO:30, wherein the polypeptide comprises KS activity;
 - (c) a polypeptide comprising an amino acid sequence at least 80% identical to SEQ ID NO:32, wherein the polypeptide comprises CLF activity;
 - (d) a polypeptide comprising an amino acid sequence at least 80% identical to SEQ ID NO:34, wherein the polypeptide comprises AT activity; and
 - (e) a polypeptide comprising an amino acid sequence at least 80% identical to SEQ ID NO:36, wherein the polypeptide comprises ER activity.
48. The isolated polypeptide of claim 47, wherein the amino acid sequences are at least 90% identical to SEQ ID NOs:4, 30, 32, 34, and 36, respectively.
49. The isolated polypeptide of claim 47, wherein the amino acid sequences are at least 95% identical to SEQ ID NOs:4, 30, 32, 34, and 36, respectively.
50. The isolated polypeptide of claim 47, wherein the polypeptides comprise the amino acid sequence of SEQ ID NOs:4, 30, 32, 34, and 36, respectively.

51. An isolated polypeptide selected from the group consisting of:
- (a) a polypeptide comprising an amino acid sequence at least 80% identical to SEQ ID NO:6, wherein the polypeptide comprises a PUFA synthase activity selected from the group consisting of DH activity, ER activity, and combinations thereof;
 - (b) a polypeptide comprising an amino acid sequence at least 80% identical to SEQ ID NO:38, wherein the polypeptide comprises DH activity;
 - (c) a polypeptide comprising an amino acid sequence at least 80% identical to SEQ ID NO:40, wherein the polypeptide comprises DH activity; and
 - (d) a polypeptide comprising an amino acid sequence at least 80% identical to SEQ ID NO:42, wherein the polypeptide comprises ER activity.
52. The isolated polypeptide of claim 51, wherein the amino acid sequences are at least 90% identical to SEQ ID NOs:6, 38, 40, and 42, respectively.
53. The isolated polypeptide of claim 51, wherein the amino acid sequences are at least 95% identical to SEQ ID NOs:6, 38, 40, and 42, respectively.
54. The isolated polypeptide of claim 51, wherein the polypeptides comprise the amino acid sequences of SEQ ID NOs:6, 38, 40, and 42, respectively.
55. The isolated polypeptide of any one of claims 42-54, wherein the polypeptide is a fusion polypeptide.
56. A composition comprising the polypeptide of any one of claims 42-54 and a biologically acceptable carrier.
57. A method of increasing DHA production in an organism having PUFA synthase activity, comprising:
- expressing the isolated nucleic acid molecule of any one of claims 1-25, the recombinant nucleic acid molecule of claim 27 or claim 28, or combinations thereof in the organism under conditions effective to produce DHA,

wherein the PUFA synthase activity replaces an inactive or deleted activity, introduces a new activity, or enhances an existing activity in the organism, and wherein DHA production in the organism is increased.

58. A method of isolating lipids from a host cell, comprising:
- (a) expressing a PUFA synthase gene in the host cell under conditions effective to produce lipids, wherein the PKS system comprises the isolated nucleic acid molecule of any one of claims 1-25, the recombinant nucleic acid molecule of claim 27 or claim 28, or combinations thereof in the host cell, and
 - (b) isolating lipids from the host cell.
59. The method of claim 58, wherein the host cell is selected from the group consisting of a plant cell, an isolated animal cell, and a microbial cell.
60. The method of claim 58, wherein the lipids comprise DHA.
61. An isolated nucleic acid molecule selected from the group consisting of:
- (a) a nucleic acid molecule comprising a polynucleotide sequence at least 80% identical to SEQ ID NO:68 or SEQ ID NO:120, wherein the polynucleotide sequence encodes a polypeptide comprising PUFA synthase activity selected from the group consisting of KS activity, MAT activity, ACP activity, KR activity, DH activity, and combinations thereof;
 - (b) a nucleic acid molecule comprising a polynucleotide sequence at least 80% identical to SEQ ID NO:74, wherein the polynucleotide sequence encodes a polypeptide comprising KS activity;
 - (c) a nucleic acid molecule comprising a polynucleotide sequence at least 80% identical to SEQ ID NO:76, wherein the polynucleotide sequence encodes a polypeptide comprising MAT activity;
 - (d) a nucleic acid molecule comprising a polynucleotide sequence at least 80% identical to any one of SEQ ID NOs:80, 82, 84, 86, 88, 90, 92, 94, 96, or 98,

wherein the polynucleotide sequence encodes a polypeptide comprising ACP activity;

- (e) a nucleic acid molecule comprising a polynucleotide sequence at least 80% identical to SEQ ID NO:78, wherein the polynucleotide sequence encodes a polypeptide comprising ACP activity;
 - (f) a nucleic acid molecule comprising a polynucleotide sequence at least 80% identical to SEQ ID NO:100, wherein the polynucleotide sequence encodes a polypeptide comprising KR activity; and
 - (g) a nucleic acid molecule comprising a polynucleotide sequence at least 80% identical to SEQ ID NO:118, wherein the polynucleotide sequence encodes a polypeptide comprising DH activity.
62. The isolated nucleic acid molecule of claim 61, wherein the polynucleotide sequences are at least 90% identical to SEQ ID NOs:68, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96, 98, 100, 118, and 120, respectively.
63. The isolated nucleic acid molecule of claim 61, wherein the polynucleotide sequences are at least 95% identical to SEQ ID NOs:68, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96, 98, 100, 118 and 120, respectively.
64. The isolated nucleic acid molecule of claim 61, wherein the nucleic acid molecules comprise the polynucleotide sequences of SEQ ID NOs:68, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96, 98, 100, 118, and 120, respectively.
65. An isolated nucleic acid molecule selected from the group consisting of:
- (a) a nucleic acid molecule comprising a polynucleotide sequence encoding a polypeptide, wherein the polypeptide comprises an amino acid sequence at least 80% identical to SEQ ID NO:69, and wherein the polypeptide comprises a PUFA synthase activity selected from the group consisting of KS activity, MAT activity, ACP activity, KR activity, DH activity, and combinations thereof;

- (b) a nucleic acid molecule comprising a polynucleotide sequence encoding a polypeptide, wherein the polypeptide comprises an amino acid sequence at least 80% identical to SEQ ID NO:75, and wherein the polypeptide comprises KS activity;
 - (c) a nucleic acid molecule comprising a polynucleotide sequence encoding a polypeptide, wherein the polypeptide comprises an amino acid sequence at least 80% identical to SEQ ID NO:77, and wherein the polypeptide comprises MAT activity;
 - (d) a nucleic acid molecule comprising a polynucleotide sequence encoding a polypeptide, wherein the polypeptide comprises an amino acid sequence at least 80% identical to any one of SEQ ID NOs:81, 83, 85, 87, 89, 91, 93, 95, 97, or 99, and wherein the polypeptide comprises ACP activity;
 - (e) a nucleic acid molecule comprising a polynucleotide sequence encoding a polypeptide, wherein the polypeptide comprises an amino acid sequence at least 80% identical to SEQ ID NO:79, and wherein the polypeptide comprises ACP activity;
 - (f) a nucleic acid molecule comprising a polynucleotide sequence encoding a polypeptide, wherein the polypeptide comprises an amino acid sequence at least 80% identical to SEQ ID NO:101, and wherein the polypeptide comprises KR activity; and
 - (g) a nucleic acid molecule comprising a polynucleotide sequence encoding a polypeptide, wherein the polypeptide comprises an amino acid sequence at least 80% identical to SEQ ID NO:119, and wherein the polypeptide comprises DH activity.
66. The isolated nucleic acid molecule of claim 65, wherein the amino acid sequences are at least 90% identical to SEQ ID NOs:69, 75, 77, 79, 81, 83, 85, 87, 89, 91, 93, 95, 97, 99, 101, and 119, respectively.

67. The isolated nucleic acid molecule of claim 65, wherein the amino acid sequences are at least 95% identical to SEQ ID NOs:69, 75, 77, 79, 81, 83, 85, 87, 89, 91, 93, 95, 97, 99, 101, and 119, respectively.
68. The isolated nucleic acid molecule of claim 65, wherein the polypeptides comprise the amino acid sequences of SEQ ID NOs:69, 75, 77, 79, 81, 83, 85, 87, 89, 91, 93, 95, 97, 99, 101, and 119, respectively.
69. An isolated nucleic acid molecule selected from the group consisting of:
- (a) a nucleic acid molecule comprising a polynucleotide sequence at least 80% identical to SEQ ID NO:70 or SEQ ID NO:121, wherein the polynucleotide sequence encodes a polypeptide comprising PUFA synthase activity selected from the group consisting of KS activity, CLF activity, AT activity, ER activity, and combinations thereof;
 - (b) a nucleic acid molecule comprising a polynucleotide sequence at least 80% identical to SEQ ID NO:102, wherein the polynucleotide sequence encodes a polypeptide comprising KS activity;
 - (c) a nucleic acid molecule comprising a polynucleotide sequence at least 80% identical to SEQ ID NO:104, wherein the polynucleotide sequence encodes a polypeptide comprising CLF activity;
 - (d) a nucleic acid molecule comprising a polynucleotide sequence at least 80% identical to SEQ ID NO:106, wherein the polynucleotide sequence encodes a polypeptide comprising AT activity; and
 - (e) a nucleic acid molecule comprising a polynucleotide sequence at least 80% identical to SEQ ID NO:108, wherein the polynucleotide sequence encodes a polypeptide comprising ER activity;
70. The isolated nucleic acid molecule of claim 69, wherein the polynucleotide sequences are at least 90% identical to SEQ ID NOs:70, 102, 104, 106, 108, and 121, respectively.

71. The isolated nucleic acid molecule of claim 69, wherein the polynucleotide sequences are at least 95% identical to SEQ ID NOs:70, 102, 104, 106, 108, and 121, respectively.
72. The isolated nucleic acid molecule of claim 69, wherein the nucleic acid molecules comprise the polynucleotide sequences of SEQ ID NOs:70, 102, 104, 106, 108, and 121, respectively.
73. An isolated nucleic acid molecule selected from the group consisting of:
 - (a) a nucleic acid molecule comprising a polynucleotide sequence encoding a polypeptide, wherein the polypeptide comprises an amino acid sequence at least 80% identical to SEQ ID NO:71, and wherein the polypeptide comprises a PUFA synthase activity selected from the group consisting of KS activity, CLF activity, AT activity, ER activity, and combinations thereof;
 - (b) a nucleic acid molecule comprising a polynucleotide sequence encoding a polypeptide, wherein the polypeptide comprises an amino acid sequence at least 80% identical to SEQ ID NO:103, and wherein the polypeptide comprises KS activity;
 - (c) a nucleic acid molecule comprising a polynucleotide sequence encoding a polypeptide, wherein the polypeptide comprises an amino acid sequence at least 80% identical to SEQ ID NO:105, and wherein the polypeptide comprises CLF activity;
 - (d) a nucleic acid molecule comprising a polynucleotide sequence encoding a polypeptide, wherein the polypeptide comprises an amino acid sequence at least 80% identical to SEQ ID NO:107, and wherein the polypeptide comprises AT activity; and
 - (e) a nucleic acid molecule comprising a polynucleotide sequence encoding a polypeptide, wherein the polypeptide comprises an amino acid sequence at least 80% identical to SEQ ID NO:109, and wherein the polypeptide comprises ER activity.

74. The isolated nucleic acid molecule of claim 73, wherein the amino acid sequences are at least 90% identical to SEQ ID NOs:71, 103, 105, 107, and 109, respectively.
75. The isolated nucleic acid molecule of claim 73, wherein the amino acid sequences are at least 95% identical to SEQ ID NOs:71, 103, 105, 107, and 109, respectively.
76. The isolated nucleic acid molecule of claim 73, wherein the polypeptides comprise the amino acid sequence of SEQ ID NOs:71, 103, 105, 107, and 109, respectively.
77. An isolated nucleic acid molecule selected from the group consisting of:
 - (a) an nucleic acid molecule comprising a polynucleotide sequence at least 80% identical to SEQ ID NO:72 or SEQ ID NO:122, wherein the polynucleotide sequence encodes a polypeptide comprising PUFA synthase activity selected from the group consisting of DH activity, ER activity, and combinations thereof;
 - (b) a nucleic acid molecule comprising a polynucleotide sequence at least 80% identical to SEQ ID NO:110, wherein the polynucleotide sequence encodes a polypeptide comprising DH activity;
 - (c) a nucleic acid molecule comprising a polynucleotide sequence at least 80% identical to SEQ ID NO:112, wherein the polynucleotide sequence encodes a polypeptide comprising DH activity; and
 - (d) a nucleic acid molecule comprising a polynucleotide sequence at least 80% identical to SEQ ID NO:114, wherein the polynucleotide sequence encodes a polypeptide comprising ER activity.
78. The isolated nucleic acid molecule of claim 77, wherein the polynucleotide sequences are at least 90% identical to SEQ ID NOs:72, 110, 112, 114, and 122, respectively.
79. The isolated nucleic acid molecule of claim 77, wherein the polynucleotide sequences are at least 95% identical to SEQ ID NOs:72, 110, 112, 114, and 122, respectively.

80. The isolated nucleic acid molecule of claim 77, wherein the nucleic acid molecules comprise the polynucleotide sequences of SEQ ID NOs:72, 110, 112, 114, and 122, respectively.
81. An isolated nucleic acid molecule selected from the group consisting of:
- (a) a nucleic acid molecule comprising a polynucleotide sequence encoding a polypeptide, wherein the polypeptide comprises an amino acid sequence at least 80% identical to SEQ ID NO:73, wherein the polypeptide comprises a PUFA synthase activity selected from the group consisting of DH activity, ER activity, and combinations thereof;
 - (b) a nucleic acid molecule comprising a polynucleotide sequence encoding a polypeptide, wherein the polypeptide comprises an amino acid sequence at least 80% identical to SEQ ID NO:111, and wherein the polypeptide comprises DH activity;
 - (c) a nucleic acid molecule comprising a polynucleotide sequence encoding a polypeptide, wherein the polypeptide comprises an amino acid sequence at least 80% identical to SEQ ID NO:113, and wherein the polypeptide comprises DH activity; and
 - (d) a nucleic acid molecule comprising a polynucleotide sequence encoding a polypeptide, wherein the polypeptide comprises an amino acid sequence at least 80% identical to SEQ ID NO:115, and wherein the polypeptide comprises ER activity.
82. The isolated nucleic acid molecule of claim 81, wherein the amino acid sequences are at least 90% identical to SEQ ID NOs:73, 111, 113, and 115, respectively.
83. The isolated nucleic acid molecule of claim 81, wherein the amino acid sequences are at least 95% identical to SEQ ID NOs:73, 111, 113, and 115, respectively.
84. The isolated nucleic acid molecule of claim 81, wherein the polypeptides comprise the amino acid sequences of SEQ ID NOs:73, 111, 113, and 115, respectively.

85. An isolated nucleic acid molecule comprising a polynucleotide sequence encoding a polypeptide comprising PUFA synthase activity selected from the group consisting of KS activity, MAT activity, ACP activity, KR activity, CLF activity, AT activity, ER activity, DH activity, and combinations thereof,

wherein the polynucleotide hybridizes under stringent conditions to the complement of the polynucleotide sequence of any one of claims 61-84.

86. An isolated nucleic acid molecule comprising a polynucleotide sequence that is fully complementary to the polynucleotide sequence of any one of claims 61-84.
87. A recombinant nucleic acid molecule comprising the nucleic acid molecule of any one of claims 61-85 or combinations thereof and a transcription control sequence.
88. The recombinant nucleic acid molecule of claim 87, wherein the recombinant nucleic acid molecule is a recombinant vector.
89. A host cell that expresses the nucleic acid molecule of any one of claims 61-85, the recombinant nucleic acid molecule of claim 87 or claim 88, or combinations thereof.
90. The host cell of claim 89, wherein the host cell is selected from the group consisting of a plant cell, a microbial cell, and an animal cell.
91. The host cell of claim 90, wherein the host cell is a microbial cell.
92. The microbial cell of claim 91, wherein the microbial cell is a bacterium.
93. The microbial cell of claim 91, wherein the microbial cell is a thraustochytrid.
94. The microbial cell of claim 93, wherein the thraustochytrid is a *Schizochytrium* or a *Thraustochytrium*.
95. A method to produce at least one PUFA, comprising:

expressing a PUFA synthase gene in a host cell under conditions effective to produce PUFA,

wherein the PUFA synthase gene comprises the isolated nucleic acid molecule of any one of claims 61-85, the recombinant nucleic acid molecule of claim 87 or claim 88, or combinations thereof, and

wherein at least one PUFA is produced.

96. The method of claim 95, wherein the host cell is selected from the group consisting of a plant cell, an isolated animal cell, and a microbial cell.

97. The method of claim 95, wherein the at least one PUFA comprises DHA or eicosapentaenoic acid (EPA).

98. A method to produce lipids enriched for DHA, EPA, or a combination thereof comprising:

expressing a PUFA synthase gene in a host cell under conditions effective to produce lipids, wherein the PUFA synthase gene comprises the isolated nucleic acid molecule of any one of claims 61-85, the recombinant nucleic acid molecule of claim 87 or claim 88, or combinations thereof in the host cell, and

wherein lipids enriched with DHA, EPA, or a combination thereof are produced.

99. A method of making a recombinant vector comprising inserting the isolated nucleic acid molecule of any one of claims 61-85 into a vector.

100. A method of making a recombinant host cell comprising introducing the recombinant vector of claim 99 into a host cell.

101. The method of claim 100, wherein the host cell is selected from the group consisting of a plant cell, an isolated animal cell, and a microbial cell.

102. An isolated polypeptide encoded by the polynucleotide sequence of any one of claims 61-85.

103. An isolated polypeptide selected from the group consisting of:

(a) a polypeptide comprising an amino acid sequence at least 80% identical to SEQ ID NO:69, wherein the polypeptide comprises a PUFA synthase activity selected

from the group consisting of KS activity, MAT activity, ACP activity, KR activity, DH activity, and combinations thereof;

- (b) a polypeptide comprising an amino acid sequence at least 80% identical to SEQ ID NO:75, wherein the polypeptide comprises KS activity;
 - (c) a polypeptide comprising an amino acid sequence at least 80% identical to SEQ ID NO:77, wherein the polypeptide comprises MAT activity;
 - (d) a polypeptide comprising an amino acid sequence at least 80% identical to any one of SEQ ID NOs:81, 83, 85, 87, 89, 91, 93, 95, 97, or 99, wherein the polypeptide comprises ACP activity;
 - (e) a polypeptide comprising an amino acid sequence at least 80% identical to SEQ ID NO:79, wherein the polypeptide comprises ACP activity;
 - (f) a polypeptide comprising an amino acid sequence at least 80% identical to SEQ ID NO:101, wherein the polypeptide comprises KR activity; and
 - (g) a polypeptide comprising an amino acid sequence at least 80% identical to SEQ ID NO:119, wherein the polypeptide comprises DH activity.
104. The isolated polypeptide of claim 103, wherein the amino acid sequences are at least 90% identical to SEQ ID NOs:69, 75, 77, 79, 81, 83, 85, 87, 89, 91, 93, 95, 97, 99, 101, and 119, respectively.
105. The isolated polypeptide of claim 103, wherein the amino acid sequences are at least 95% identical to SEQ ID NOs:69, 75, 77, 79, 81, 83, 85, 87, 89, 91, 93, 95, 97, 99, 101, and 119, respectively.
106. The isolated polypeptide of claim 103, wherein the polypeptides comprise the amino acid sequences of SEQ ID NOs:69, 75, 77, 79, 81, 83, 85, 87, 89, 91, 93, 95, 97, 99, 101, and 119, respectively.

107. An isolated polypeptide selected from the group consisting of:
- (a) a polypeptide comprising an amino acid sequence at least 80% identical to SEQ ID NO:71, wherein the polypeptide comprises a PUFA synthase activity selected from the group consisting of KS activity, CLF activity, AT activity, ER activity, and combinations thereof;
 - (b) a polypeptide comprising an amino acid sequence at least 80% identical to SEQ ID NO:103, wherein the polypeptide comprises KS activity;
 - (c) a polypeptide comprising an amino acid sequence at least 80% identical to SEQ ID NO:105, wherein the polypeptide comprises CLF activity;
 - (d) a polypeptide comprising an amino acid sequence at least 80% identical to SEQ ID NO:107, wherein the polypeptide comprises AT activity; and
 - (e) a polypeptide comprising an amino acid sequence at least 80% identical to SEQ ID NO:109, wherein the polypeptide comprises ER activity.
108. The isolated polypeptide of claim 107, wherein the amino acid sequences are at least 90% identical to SEQ ID NOs:71, 103, 105, 107, and 109, respectively.
109. The isolated polypeptide of claim 107, wherein the amino acid sequences are at least 95% identical to SEQ ID NOs:71, 103, 105, 107, and 109, respectively.
110. The isolated polypeptide of claim 107, wherein the polypeptides comprise the amino acid sequence of SEQ ID NOs:71, 103, 105, 107, and 109, respectively.
111. An isolated polypeptide selected from the group consisting of:
- (a) a polypeptide comprising an amino acid sequence at least 80% identical to SEQ ID NO:73, wherein the polypeptide comprises a PUFA synthase activity selected from the group consisting of DH activity, ER activity, and combinations thereof;
 - (b) a polypeptide comprising an amino acid sequence at least 80% identical to SEQ ID NO:111, wherein the polypeptide comprises DH activity;

- (c) a polypeptide comprising an amino acid sequence at least 80% identical to SEQ ID NO:113, wherein the polypeptide comprises DH activity; and
 - (d) a polypeptide comprising an amino acid sequence at least 80% identical to SEQ ID NO:115, wherein the polypeptide comprises ER activity.
112. The isolated polypeptide of claim 111, wherein the amino acid sequences are at least 90% identical to SEQ ID NOs:73, 111, 113, and 115, respectively.
113. The isolated polypeptide of claim 111, wherein the amino acid sequences are at least 95% identical to SEQ ID NOs:73, 111, 113, and 115, respectively.
114. The isolated polypeptide of claim 111, wherein the polypeptides comprise the amino acid sequences of SEQ ID NOs:73, 111, 113, and 115, respectively.
115. The isolated polypeptide of any one of claims 102-114, wherein the polypeptide is a fusion polypeptide.
116. A composition comprising the polypeptide of any one of claims 102-114 and a biologically acceptable carrier.
117. A method of increasing production of DHA, EPA, or a combination thereof in an organism having PUFA synthase activity, comprising:
- expressing the isolated nucleic acid molecule of any one of claims 61-85, the recombinant nucleic acid molecule of claim 87 or claim 88, or combinations thereof in the organism under conditions effective to produce DHA, EPA, or a combination thereof,
 - wherein the PUFA synthase activity replaces an inactive or deleted activity, introduces a new activity, or enhances an existing activity in the organism, and
 - wherein production of DHA, EPA, or a combination thereof in the organism is increased.
118. A method of isolating lipids from a host cell, comprising:
- (a) expressing a PUFA synthase gene in the host cell under conditions effective to produce lipids, wherein the PKS system comprises the isolated nucleic acid

molecule of any one of claims 61-85, the recombinant nucleic acid molecule of claim 87 or claim 88, or combinations thereof in the host cell, and

(b) isolating lipids from the host cell.

119. The method of claim 118, wherein the host cell is selected from the group consisting of a plant cell, an isolated animal cell, and a microbial cell.
120. The method of claim 118, wherein the lipids comprise DHA, EPA, or a combination thereof.

Figure 1. *Schizochytrium* sp. ATCC PTA-9695 Gene architecture

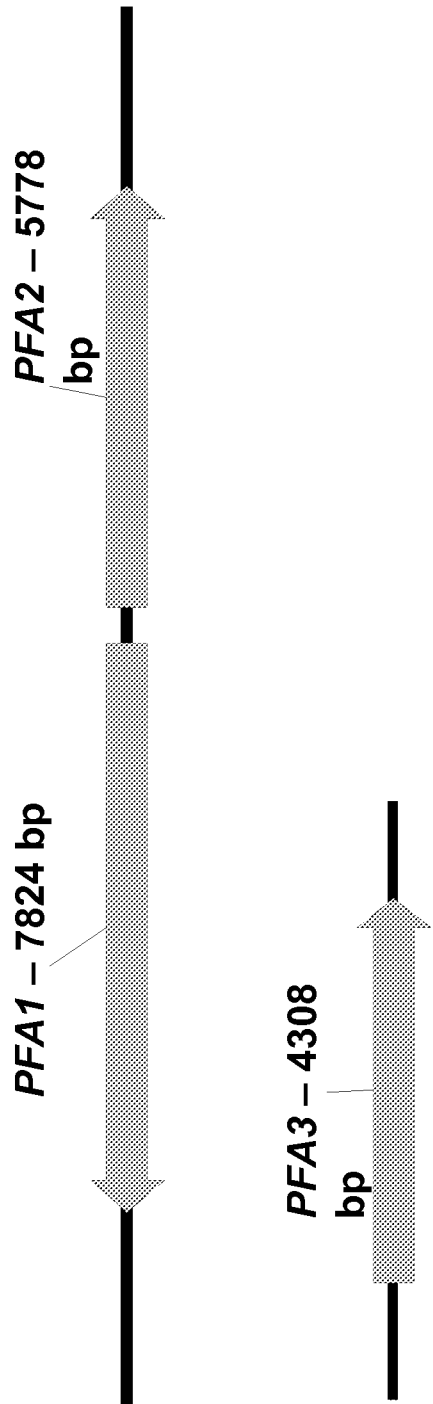


Figure 2. *Thraustochytrium* sp. ATCC PTA-10212 Gene architecture

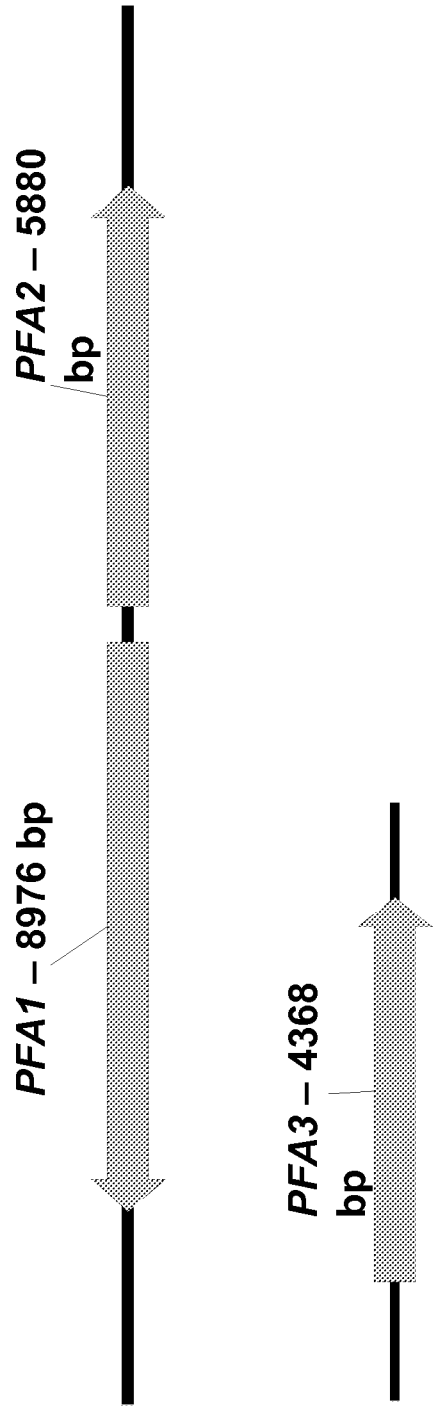


Figure 3. Domain architecture:

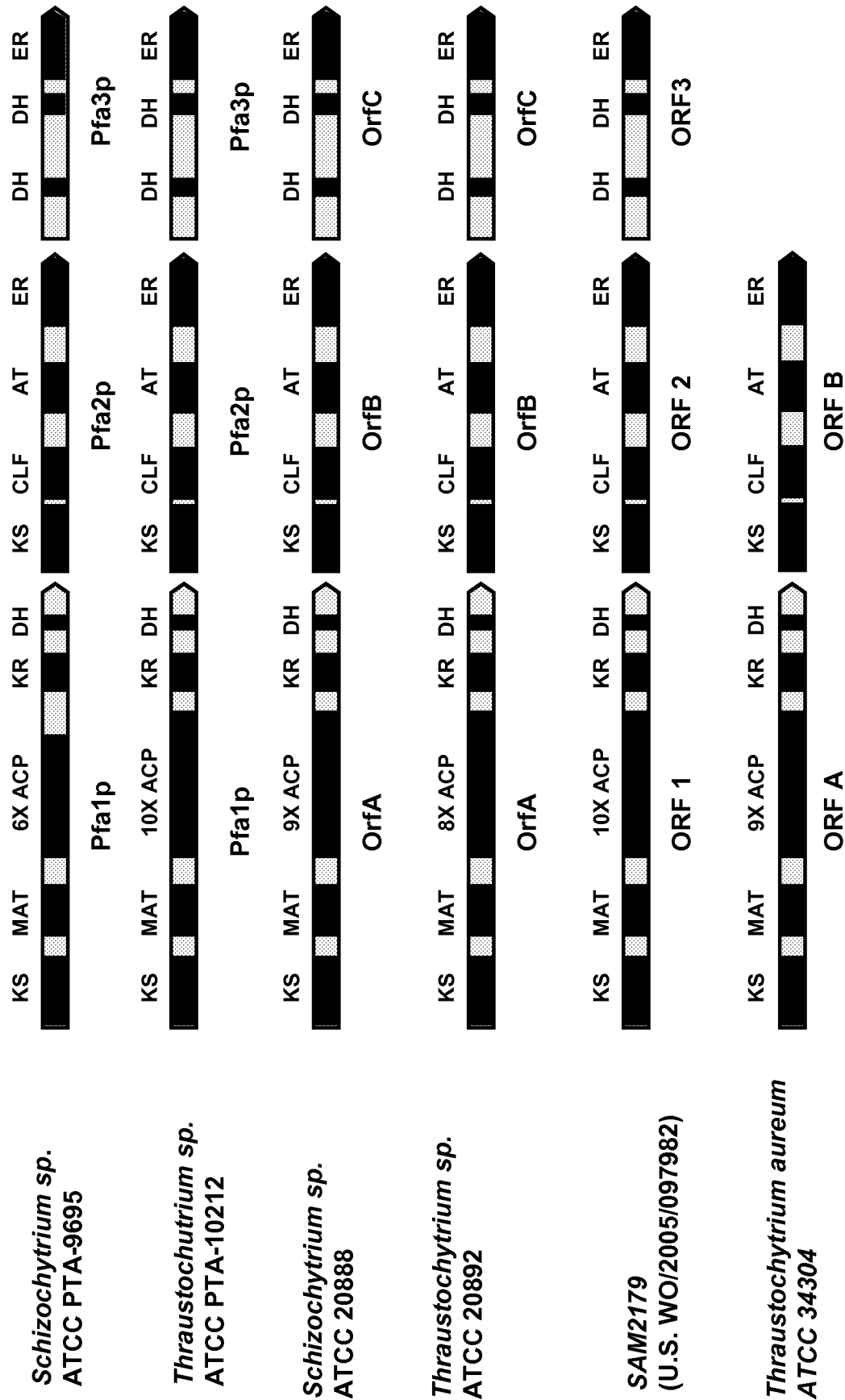


FIG. 4

	1	50
Sch. sp. 9695 Pfalp	(1)	-----
Thr. aureum ORf1	(1)	RKCIRPSLGHHWAIIGVLGRALRIVRPIRYEATNLRRLPRSGWLVALGLF
Sch. sp. 20888 OrfA	(1)	-----
Thr. sp. 10212 Pfalp	(1)	-----
Thr. sp. 20892 OrfA	(1)	-----
	51	100
Sch. sp. 9695 Pfalp	(1)	-----M
Thr. aureum ORf1	(51)	CDLSSCAGKLDLQTRDTAKDPCCRKWSASRAPPRPRAEADKASNEMETK
Sch. sp. 20888 OrfA	(1)	-----MAARLQEQKGEM
Thr. sp. 10212 Pfalp	(1)	-----ME
Thr. sp. 20892 OrfA	(1)	-----MKDME
	101	150
Sch. sp. 9695 Pfalp	(2)	DTRIATVGMSSAILPSCGENVRESWEAIRDGLDCLSDLPADRVDTAYYINPE
Thr. aureum ORf1	(101)	DDRTAIVGMSAILPSCGSEVRESWEAIREGLDCLQDLPADRVDTAYYDPN
Sch. sp. 20888 OrfA	(14)	DTRIATVGMSSAILPSCGTTVRESWETIRAGIDCLSDLPEDRVDTAYYDPV
Thr. sp. 10212 Pfalp	(3)	DQRTAIVGMSAILPSCGENVRESWEAIRDGLNCLSDLPADRVDTAYYINPT
Thr. sp. 20892 OrfA	(6)	DRRTAIVGMSAHLPCGTDVRESWQAIRDGLDCLSDLPADRVDTAYYINPN
	151	200
Sch. sp. 9695 Pfalp	(52)	KTTKDKIYCKRGGFIPEYDFDAREFGLNMFQMEDSDANQTTISLLKVKKEAL
Thr. aureum ORf1	(151)	KTTKDKIYCKRGGFIPEYDFDAREFGLNMFQMEDSDANQTTISLLKVKKEAL
Sch. sp. 20888 OrfA	(64)	KTTKDKIYCKRGGFIPEYDFDAREFGLNMFQMEDSDANQTTISLLKVKKEAL
Thr. sp. 10212 Pfalp	(53)	KGVKDKIYCKRGGFIPEYDFDAREFGLNMLQMEDSDANQTTISLLKVKKEAL
Thr. sp. 20892 OrfA	(56)	KATKDKIYCKRGGFIPEYDFDAREFGLNMFQMEDSDANQTTISLLKVKKEAL
	201	250
Sch. sp. 9695 Pfalp	(102)	TDANIPTATSGKKNIGCVLGIGGGQKASHEFYSSLNLYVVVDKVLKRMGLP
Thr. aureum ORf1	(201)	SDAGVPTATSGKKNIGCVLGIGGGQKASHEFYSSLNLYVVVEKVLKRMGLP
Sch. sp. 20888 OrfA	(114)	QDAGINALGKEKKNIGCVLGIGGGQKASHEFYSSLNLYVVVEKVLKRMGLP
Thr. sp. 10212 Pfalp	(103)	SDANIPTATSGKKNIGCVLGIGGGQKASHEFYSSLNLYVVVDKVLKRMGLP
Thr. sp. 20892 OrfA	(106)	SDASVPTATSGKKNIGCVLGIGGGQKASHEFYSSLNLYVVVEKVLKRMGLP
	251	300
Sch. sp. 9695 Pfalp	(152)	SEDVAAVVKYKASFPEWRLDSFPGFLGNVTAGRCNTFNMEGMNCVVDA
Thr. aureum ORf1	(251)	DEVVEAAVEKYKANFPEWRLDSFPGFLGNVTAGRCNTFNMEGMNCVVDA
Sch. sp. 20888 OrfA	(164)	SEDVKVAVEKYKANFPEWRLDSFPGFLGNVTAGRCNTFNMEGMNCVVDA
Thr. sp. 10212 Pfalp	(153)	DEVVEAVEKYKANFPEWRLDSFPGFLGNVTAGRCNTFNMEGMNCVVDA
Thr. sp. 20892 OrfA	(156)	DADVEAVEKYKANFPEWRLDSFPGFLGNVTAGRCNTFNMEGMNCVVDA
	301	350
Sch. sp. 9695 Pfalp	(202)	ACASSLIAIKVAIDELLGDCDAMIAGATCTDNTSGMYMAFSKTPVFSTD
Thr. aureum ORf1	(301)	ACASSLIAIKVAIDELLGDCDTMIAGATCTDNTSGMYMAFSKTPVFSTD
Sch. sp. 20888 OrfA	(214)	ACASSLIAIKVAIDELLGDCDMMTGATCTDNTSGMYMAFSKTPVFSTD
Thr. sp. 10212 Pfalp	(203)	ACASSLIAIKVAIDELLGDCDAMIAGATCTDNTSGMYMAFSKTPVFSTD
Thr. sp. 20892 OrfA	(206)	ACASSLIAIKVAIDELLGDCDTMIAGATCTDNTSGMYMAFSKTPVFSTD
	351	400
Sch. sp. 9695 Pfalp	(252)	PSVYAYDAATKGMLIGESAMFVLKRYADAVRDGDTTHAVIRSCASSSDG
Thr. aureum ORf1	(351)	QSVYAYDAATKGMLIGESAMFVLKRYADAVRDGDETHAVIRSCASSSDG
Sch. sp. 20888 OrfA	(264)	PSVYAYDEKTKGMLIGESAMFVLKRYADAVRDGDETHAVIRSCASSSDG
Thr. sp. 10212 Pfalp	(253)	QSCLAYDEKTKGMLIGESAMFVLKRYADAVRDGDTTHAVIRSCASSSDG
Thr. sp. 20892 OrfA	(256)	PSVYAYDEKTKGMLIGESAMFVLKRYADAVRDGDTTHAVIRSCASSSDG
	401	450
Sch. sp. 9695 Pfalp	(302)	KAGGIYPTTISGOEEALRRAYARANVDPATVTLVEGHGTGTPVGDRIELT
Thr. aureum ORf1	(401)	KAGGIYAPTISGOEEALRRAYARAGVDPSTVTLVEGHGTGTPVGDRIELT
Sch. sp. 20888 OrfA	(314)	KAGGIYPTTISGOEEALRRAYNRACVDPATVTLVEGHGTGTPVGDRIELT
Thr. sp. 10212 Pfalp	(303)	KAGGIYPTTISGOEEALRRAYRAGVSPNTTTLVEGHGTGTPVGDRIELT
Thr. sp. 20892 OrfA	(306)	KAGGIYPTTISGOEEALRRAYARAGVCPSTTGLVEGHGTGTPVGDRIELT
	451	500
Sch. sp. 9695 Pfalp	(352)	ALSNLFSKASANGGGAEAEQAVGSIKSOIGHLKAVAGLAGLVKVYMA
Thr. aureum ORf1	(451)	ALSNLFDANKGR-----KETVAVGSIKSOIGHLKAVAGFAGLVKVYMA
Sch. sp. 20888 OrfA	(364)	ALSNLFDKANGEG-----NTEKAVGSIKSSIGHLKAVAGLAGLVKVYMA
Thr. sp. 10212 Pfalp	(353)	ALSNLFDKANGPG-----HKEEAVGSIKSOIGHLKAVAGCAGLVKVYMA
Thr. sp. 20892 OrfA	(356)	ALSNLFDKANGSK-----KEQAVGSIKSOIGHLKAVAGFAGLVKVYMA

FIG. 4 (cont'd)

	501	550
Sch. sp. 9695 Pfalp	(402)	LKHKTLPOTINVDKPPSLVDGTRIQSSPLYNTNMNRPWETPVGVPRRAGV
Thr. aureum ORf1	(495)	LKHKTLPOTINVDHPPALHDGSPIDSSLYINTNMNRPWETAPGVPRRAGI
Sch. sp. 20888 OrfA	(409)	LKHKTLPGTINVDNPPNLYDNTPEINSSLYINTNMNRPWEPPEGVPRRAGI
Thr. sp. 10212 Pfalp	(398)	LKHKTLPQINVENPPNLDGTVISDTILYINTNMNRPWITKPGVPRRAGI
Thr. sp. 20892 OrfA	(400)	LKHKTLPQINVDQPLLYDGTQIDSSLYINKTNRPWETONKLPRRAGV
	551	600
Sch. sp. 9695 Pfalp	(452)	SSFGFGGANYHAVLEEFEPHESESAYRYNNLPQVALLHAGDVATLAATVRA
Thr. aureum ORf1	(545)	SSFGFGGANYHAVLEEAPEHAKPYRMNOVPOPVLLHASSASALASICDA
Sch. sp. 20888 OrfA	(459)	SSFGFGGANYHAVLEEAPEHTTAYRLNKRPOPVLLMAATPAALOSLCEA
Thr. sp. 10212 Pfalp	(448)	SSFGFGGANYHAVLEEFEPQTKPYRINVSAQPMILHAVNANSLOKLCED
Thr. sp. 20892 OrfA	(450)	SSFGFGGANYHAVLEEFEPHEKEYRLNTVGHEVLLYAPSVEALKVLICND
	601	650
Sch. sp. 9695 Pfalp	(502)	KIALATAEQEEARVVKHADYIAYHRFLDECKLRGAVPOAHARVGLVRLDL
Thr. aureum ORf1	(595)	QADAEQAAVSP-EASKHADYRAIVAFHEAFKLPAGVPAGHARIGFVSGSA
Sch. sp. 20888 OrfA	(509)	OLKEFEAAIKENETVKNITAYIKCVKFGEOFKFPGSIPATNARLGLFLVDA
Thr. sp. 10212 Pfalp	(498)	OLKLEKESREKCVNTKNTDYAFSKFQDSFKLGSVPSOHARVGFASKSI
Thr. sp. 20892 OrfA	(500)	QLAELTIALEEAKTHKNVDKVCYKFFIDEFOLOGSCPPENPRVGFLATLP
	651	700
Sch. sp. 9695 Pfalp	(552)	SLLIIVLEAAAPKLAGESATEWTVSVATGEAAFRVRGVATEANVAALFS
Thr. aureum ORf1	(644)	ANTLAVLRAASAKLKQSSATLEWTLLE--GVTYRSAAHHTPGSVAALFA
Sch. sp. 20888 OrfA	(559)	EDACSTLRAICAQFAKIVTKEAWRLPRE-GVSFRAKGIATN-GAVAALFS
Thr. sp. 10212 Pfalp	(548)	EDTISTLSAIVNRFQKIIITTSWALPKE-GAIFRSTALINDNKSVAALFS
Thr. sp. 20892 OrfA	(550)	TENLIIVALKAILAQDLAKPDAKKWDLPHKKAFGATFASSSVKGSVAALFA
	701	750
Sch. sp. 9695 Pfalp	(602)	GOGAQYTHMFSADVAMWPPFRESVAAMDRAQRERF--SPAKRVSSVLYP
Thr. aureum ORf1	(692)	GOGAQYTHMFADVAMWPPFRSAVOEMDAAQVTAAP----KRLSEVLYP
Sch. sp. 20888 OrfA	(607)	GOGAQYTHMFSEVAMWPPQFRSLAAMDAAQSKVAGSDIDFERVSQVLYP
Thr. sp. 10212 Pfalp	(597)	GOGAQYTHMFNDVAMWPPQFRLCVNDMEKAQEEVIN-DSVKRISQVMEP
Thr. sp. 20892 OrfA	(600)	GQGTOYLNMFSDVAMWPPFRDSIVAMEEAOTEVFEG--QVEPIISKVLEP
	751	800
Sch. sp. 9695 Pfalp	(650)	RKPYGDEPRQDHKEISOTRISOPATLACSVGAFDIFKAAGLAPSEFAAGHS
Thr. aureum ORf1	(738)	RKPYAAEPEQDNKAISMTINSOPALMACAAGAFEVFRQAGLAPDHVAGHS
Sch. sp. 20888 OrfA	(657)	RKPYEREPEONPKKISITAYSOPSTLACALGAFFIFKEAGETPDFAAGHS
Thr. sp. 10212 Pfalp	(646)	RKPYARESPLDNKEISKTEYSQTTVASSVGLFELFRDAGFAPAFVAGHS
Thr. sp. 20892 OrfA	(648)	RERYASESEQGNELLCITEYSQPTTIAAAGAFDIFKAAGEKPDPMVGGHS
	801	850
Sch. sp. 9695 Pfalp	(700)	LGEFAALYAAGSLDRDAVEDLVCARAKAMSDFTAQASSSGGAMAAVIGAK
Thr. aureum ORf1	(788)	LGEFGALLAAGCASREELFRLVCSRAKAMQDVPPQ-----
Sch. sp. 20888 OrfA	(707)	LGEFAALYAAGCVDRDELFLVCCRARIIMGGKDAP-ATPKGCMAAVIGEN
Thr. sp. 10212 Pfalp	(696)	LGEFSALYAAGLIDREDLFLVLCNRAMAMRDAPK--KSADGAMAAVIGEN
Thr. sp. 20892 OrfA	(698)	LGEFAALYAAGSLSRDDLFLVCKRAKAMANASDG-----AMAAVIGPD
	851	900
Sch. sp. 9695 Pfalp	(750)	ADQISLGGAPPVVLANSNSPSQTVITGTAEAVAAASDKLRCSGNFRVVEL
Thr. aureum ORf1	(823)	-----DG-----AWLANCNSPSQVVISGDKTAVERRSSRLAGLG-FRIIFL
Sch. sp. 20888 OrfA	(756)	AENIKVQ-AANVVLGNSNSPSQTVITGSVEGTOAESARLQKEG-FRVVEL
Thr. sp. 10212 Pfalp	(744)	ASSIKES-APEVWVANNNSPSQTVITGANSQVQAEISKLKLTQG-FRVVHL
Thr. sp. 20892 OrfA	(742)	ARLITPQ-NSDVIYVNFNSATQVVISGTVQGVKEESKLLISKG-FRVIFL
	901	950
Sch. sp. 9695 Pfalp	(800)	ACEAFHSPHMRGAEOIFASALAQAPVSAPAAARFYSNVTGGAAVTSPAD
Thr. aureum ORf1	(863)	ACEGAFHSPHMTAAQATFOALDSEKISTPTNGARLYNNVSGKTCRSLGE
Sch. sp. 20888 OrfA	(804)	ACESAFHSPOMENASSAFKDVISKVSFRTPKAE--TKLFSNVSGETYPTD
Thr. sp. 10212 Pfalp	(792)	ACDGAHSPHMTAENAEKQFQKALSAYKFNKPTGS--SPKIFSNVTGGVFTD
Thr. sp. 20892 OrfA	(790)	KCOGAFHSPMLGPSDESKFLVETCTISPPKNVK-FFCNVSGKESP---N
	951	1000
Sch. sp. 9695 Pfalp	(850)	VKTNLGSHMTSPVOFVQOVRAMHAAGARVFVEFGPKQVLSKLVKETLGEA
Thr. aureum ORf1	(913)	LDCLGSHMTSPVLFOAQVENMYAAGARIFVEFGPKQVLSKLVGETLADK
Sch. sp. 20888 OrfA	(852)	AREMLTQHMTSSVKFETQVRNMHAGARIFVEFGPKQVLSKLVSETLKDD
Thr. sp. 10212 Pfalp	(840)	PKTALSSHMTSSVOFETQIRNMYAAGARVFIEFGPKQVLSKLVNETFPGD
Thr. sp. 20892 OrfA	(836)	PKQTLKSHMTSSVOFEEQIRNMYDAGARVFLEFGPQVLAKLIAENFP--

FIG. 4 (cont'd)

		1001		1050
Sch. sp. 9695 Pfalp	(900)	GDVVTVAVNEDSAKSDSTQLRQAALTLAVAGVPLKDFDRWOLPDAIRLEP		
Thr. aureum ORF1	(963)	SDFVTVAVNSSSSKDSQVLRQAAKLAVLGVPLANFDPWELCDARRIRE		
Sch. sp. 20888 OrfA	(902)	PSVVTIVSNPASGTDSDIQLRDAAVQLVVAGVNLQGFDFWDAPDATRMQA		
Thr. sp. 10212 Pfalp	(890)	TSVLTVSNPASAKSDIQLRQAAVQNAVAGVALTDFDKWELKDPTRMKE		
Thr. sp. 20892 OrfA	(884)	-SCTAISVNPASSGDSDVQLRLAAVKFAVSGAALSTFDPWEYRKPDILIL		
		1051		1100
Sch. sp. 9695 Pfalp	(950)	VKKKTTLLLSAATYVSAKTLRQREAVLNDGYTVSGAT-----AV		
Thr. aureum ORF1	(1013)	CPISKTTLLLSAATYVSNTKLAAREKVMEDNCDFSSIFASG-----P		
Sch. sp. 20888 OrfA	(952)	IKKKTTLLLSAATYVSDKTKKVRDAAMNDGRCVTYIKGAAPLIKAPPEPV		
Thr. sp. 10212 Pfalp	(940)	FPPKTTLLLSAATYVSKKTLQERERIMNDGRTVSCVQ-----R		
Thr. sp. 20892 OrfA	(933)	RPPKTTALVLSAATYVSPKTLAERKKAMEDIKLVSIPTPR-----D		
		1101		1150
Sch. sp. 9695 Pfalp	(990)	VKEVDTANEELVRQAODLQRLAEASTAAQAQSKVAELERTIQDLERK		
Thr. aureum ORF1	(1055)	ASQEMEREIANLRAELEAAQRLDTAKTQLARKOVQDPTADRQRMIAKH		
Sch. sp. 20888 OrfA	(1002)	VDEAAKREAPERLQELQLODAQRLDDAKRAAEANSKLAAKEEAKTAAS		
Thr. sp. 10212 Pfalp	(979)	IENTNTGELKLLKQLODKENEVVRVQALATOASADLONTKAELOKAQAT		
Thr. sp. 20892 OrfA	(973)	SMVSIGLIAQEVTAQPLETEIRRLNKELEHLKRELAAKASVKSASKS		
		1151		1200
Sch. sp. 9695 Pfalp	(1040)	-----VQQQQQOE-----		
Thr. aureum ORF1	(1105)	RSTLAAMVKEFEALAS-----GSPCAVPPAPVVDTAEDVPFADKV		
Sch. sp. 20888 OrfA	(1052)	-AKPAVDTAVVEKHRATLKSMLELDGYGSVDASSIQQQQQQQTAPAPVK		
Thr. sp. 10212 Pfalp	(1029)	KSSNAASDAVAKHKATLLAMLEELTGTGKAVDYSSFSKGOVASPATVRVV		
Thr. sp. 20892 OrfA	(1023)	----SKERSVLKHRALLQNILODYDDLVRVPPAVRSVAVDNTAPYADQV		
		1201		1250
Sch. sp. 9695 Pfalp	(1047)	-----K-GEN--S-----		
Thr. aureum ORF1	(1146)	STPPQ-----VTSAPIELARAEAVVMEVLAAKTGYEVDMIIEADM		
Sch. sp. 20888 OrfA	(1101)	AAAPAAP---VASAPAPAVSNELLEKAETVMEVLAAKTGYETDMIEADM		
Thr. sp. 10212 Pfalp	(1079)	SAPVQAA---APVQVSASVDSGLLAKAEQVVLEVLASKTGYETELIELDM		
Thr. sp. 20892 OrfA	(1069)	STFASERSASPLFEKRSSVSSARLAEAEAAVLSVLADKTGYDSSMIEMDM		
		1251		1300
Sch. sp. 9695 Pfalp	(1052)	-----DSNAAAEVLRRHKELLQRLQDCDEQAVPVAT		
Thr. aureum ORF1	(1187)	LDLAEELGIDSVKRIETLLAAVQAOLGVEAKDVALSRTRTVGEVVDAMKAE		
Sch. sp. 20888 OrfA	(1148)	ELETELGIDSIKRVEILSEVQAMINVEAKDVALSRTRTVGEVVDAMKAE		
Thr. sp. 10212 Pfalp	(1126)	ELETELGIDSIKRVEILSEVQALNVEAKDVALSRTRTVGEVIDAMKAE		
Thr. sp. 20892 OrfA	(1119)	DLESELGVDSIKRVEIMSEVQTLISVEVSDVDALSRITKTGVGVIEAMKLE		
		1301		1350
Sch. sp. 9695 Pfalp	(1084)	VVPTPTSSPTPTSSPVSGN-----SKSTRGSADLQALLAKAETV		
Thr. aureum ORF1	(1237)	IGGQATSAPSPMAQPOASAPSPSPTASVLPKEVALPASVDPAKLAPAEAV		
Sch. sp. 20888 OrfA	(1198)	IAGSSAPAPAAAAPAPAKA-----APAAAAPAVSNELLEKAETV		
Thr. sp. 10212 Pfalp	(1176)	IAGGQPAAPVQVAAPTQVV-----APVQASAPVDSGLLAKAEQV		
Thr. sp. 20892 OrfA	(1169)	LGPGQGTLTAESIRQPPVSEPAVPTS----SSSIANVSSARLAEAEAA		
		1351		1400
Sch. sp. 9695 Pfalp	(1123)	VMAVLAAKTGYEADMIEADMLEAELGIDSIKRVEILSEVQOLGVEAKD		
Thr. aureum ORF1	(1287)	VMEVLAAKTGYEVDMIIEADMLLDLAEELGIDSVKRIETLLAAVQAOLGVEAKD		
Sch. sp. 20888 OrfA	(1237)	VMEVLAAKTGYETDMIESDMELETELGIDSIKRVEILSEVQAMINVEAKD		
Thr. sp. 10212 Pfalp	(1215)	VLEVLASKTGYETELIELDMLELETELGIDSIKRVEILSEVQAOLSVAKD		
Thr. sp. 20892 OrfA	(1215)	VLSVLADKTGYDSSMIEMDMLESELGVDSIKRVEIMSEVQTLISVEVSD		
		1401		1450
Sch. sp. 9695 Pfalp	(1173)	VDALSRTRTVGEVVDAMKAEIVAAAS-----		
Thr. aureum ORF1	(1337)	VDALSRTRTVGEVVDAMKAEIIGGQATSAPASVAQPOASAPSPSATTASVL		
Sch. sp. 20888 OrfA	(1287)	VDALSRTRTVGEVVDAMKAEIAGGSAPAPAAAAPGPAAAA-----		
Thr. sp. 10212 Pfalp	(1265)	VDALSRTRTVGEVIDAMKAEIAGGQPAAPVQVAAPTQVVA-----		
Thr. sp. 20892 OrfA	(1265)	VDALSRTRTVGEVIDAMKLEIIGGPGQGTTLTAESIRQPPVSEPAVPTS---		
		1451		1500
Sch. sp. 9695 Pfalp	(1198)	-----		
Thr. aureum ORF1	(1387)	EKPVAAPTSADPAKLAPAEAVVMEVLAAKTGYEVDMIIEADMLLDLAEELGID		
Sch. sp. 20888 OrfA	(1327)	PAPAAAAPAVSNELLEKAETVMEVLAAKTGYETDMIESDMELETELGID		
Thr. sp. 10212 Pfalp	(1305)	PVQASAP--VDSGLLAKAEQVVLEVLASKTGYETELIELDMLELETELGID		
Thr. sp. 20892 OrfA	(1312)	-S-SSSIANVSSARLAEAEAAVLSVLADKTGYDSSMIEMDMLESELGV		

FIG. 4 (cont'd)

		1501		1550
Sch. sp. 9695 Pfalp	(1198)	-----	-----	GGSAPAV
Thr. aureum ORf1	(1437)	SVKRIEILAAVQAQLGVEAKDVDALSRTRTVGEVVEAMKAEIIGGOATSAP		
Sch. sp. 20888 OrfA	(1377)	SIKRVEILSEVQAMLNVEAKDVDALSRTRTVGEVVDAMKAEIAGGSAPAP		
Thr. sp. 10212 Pfalp	(1353)	SIKRVEILSEVQAQLSVEAKDVDALSRTRTVGEVIDAMKAEISGGQPAAP		
Thr. sp. 20892 OrfA	(1360)	SIKRVEIMSEVQTLLSVEVSDVDALSRTRTVGDVIEAMKLELGGPQGQTL		
		1551		1600
Sch. sp. 9695 Pfalp	(1205)	PSAPAASAPTPAASTAPS-----DLOALLSKAETVVMVLAAKTGYES		
Thr. aureum ORf1	(1487)	ASVAQPOISVSPTPLAASPS-----ADPAKLARAEAVVMEVLAAKTGYES		
Sch. sp. 20888 OrfA	(1427)	AAAPAPAAAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAP		
Thr. sp. 10212 Pfalp	(1403)	VQVAAPTQIVAPVQVSAP-----VDSGLLAKAEQVVLEVLAASKTGYES		
Thr. sp. 20892 OrfA	(1410)	TLTAEIRQPPVSEPAVPTSSSSSIANVLSARLAEAEAAVLSVLADKTGYDS		
		1601		1650
Sch. sp. 9695 Pfalp	(1250)	DMVEADMILEAELGIDSIKRVEILSEVQQLGVEAKDVDALSRTRTVGEV		
Thr. aureum ORf1	(1532)	DMIEADMILLDAELGIDSVKRIEILAAVQAQLGVEAKDVDALSRTRTVGEV		
Sch. sp. 20888 OrfA	(1472)	DMIESDMLELETELGIDSIKRVEILSEVQAMLNVEAKDVDALSRTRTVGEV		
Thr. sp. 10212 Pfalp	(1446)	ELIELDMLELETELGIDSIKRVEILSEVQAQLSVEAKDVDALSRTRTVGEV		
Thr. sp. 20892 OrfA	(1460)	SMIEMDMLESELGVDSIKRVEIMSEVQTLLSVEVSDVDALSRTRTVGDV		
		1651		1700
Sch. sp. 9695 Pfalp	(1300)	VDAMKAEIVAAASAGSAP----APAVPSAPAAASAPTPAASTAPSDLOAL		
Thr. aureum ORf1	(1582)	VDAMKAEIIGGOATSAPASVAQPOASAPSPSATASVLEKFEVAAPTSADPAK		
Sch. sp. 20888 OrfA	(1522)	VDAMKAEIAGGSAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAP		
Thr. sp. 10212 Pfalp	(1496)	VDAMKAEISGGQP-----TAPVQVAAPTQIVAPVQVSAPVDSGL		
Thr. sp. 20892 OrfA	(1510)	IEAMKLELGGPQG-Q---TLTAEIRQPPVSEPAVPTSSSSSIANVSSAR		
		1701		1750
Sch. sp. 9695 Pfalp	(1346)	LSKAETVVMVLAAKTGYESADMVEADMILEAELGIDSIKRVEILSEVQQL		
Thr. aureum ORf1	(1632)	LARAEAVVMEVLAAKTGYESVDMIEADMILLDAELGIDSVKRIEILAAVQAQ		
Sch. sp. 20888 OrfA	(1568)	LEKAETVVMVLAAKTGYESDMIESDMLELETELGIDSIKRVEILSEVQAM		
Thr. sp. 10212 Pfalp	(1535)	LAKAEQVVLEVLAASKTGYESTELIELDMLELETELGIDSIKRVEILSEVQAQ		
Thr. sp. 20892 OrfA	(1556)	LAEAEAAVLSVLADKTGYDSSMIEMDMLESELGVDSIKRVEIMSEVQTL		
		1751		1800
Sch. sp. 9695 Pfalp	(1396)	LGVEAKDVDALSRTRTVGEVVDAMKAEIVAAAGGSAPAPAVPSAPAAASAA		
Thr. aureum ORf1	(1682)	LGVEAKDVDALSRTRTVGEVVEAMKAEIIGGOAT-----SAPASMAQPOIS		
Sch. sp. 20888 OrfA	(1618)	LNVEAKDVDALSRTRTVGEVVDAMKAEIAGSS-----ASAPAPAAAPAPAA		
Thr. sp. 10212 Pfalp	(1585)	LSVEAKDVDALSRTRTVGEVIDAMKAEISGGQ-----PAAPVQVAAPTQI		
Thr. sp. 20892 OrfA	(1606)	LSVEVSDVDALSRTRTVGDVIEAMKLELGGPQGQTLTSEPIHOPPVSAPA		
		1801		1850
Sch. sp. 9695 Pfalp	(1446)	PTPAATAPSADLOALLAKAETVVMVLAAKTGYESADMVEADMILEAELG		
Thr. aureum ORf1	(1727)	VSPTPLAASPSADPAKLARAEAVVMEVLAAKTGYESVDMIEADMILLDAELG		
Sch. sp. 20888 OrfA	(1663)	APAPAPAAAPAVSNELLEKAETVVMVLAAKTGYESDMIESDMLELETELG		
Thr. sp. 10212 Pfalp	(1630)	VAPVQASAP--VDSGLLAKAEQVVLEVLAASKTGYESTELIELDMLELETELG		
Thr. sp. 20892 OrfA	(1656)	VPTSSSSSIANVSSARLAEAEAAVLSVLADKTGYDSSMIEMDMLESELG		
		1851		1900
Sch. sp. 9695 Pfalp	(1496)	IDSIKRVEILSEVQQLGVEAKDVDALSRTRTVGEVVDAMKAEIVAAASAG		
Thr. aureum ORf1	(1777)	IDSVKRIEILAAVQAQLGVEAKDVDALSRTRTVGEVVDAMKAEIIGGOATS		
Sch. sp. 20888 OrfA	(1713)	IDSIKRVEILSEVQAMLNVEAKDVDALSRTRTVGEVVDAMKAEIAGGSAP		
Thr. sp. 10212 Pfalp	(1678)	IDSIKRVEILSEVQAQLSVEAKDVDALSRTRTVGEVIDAMKAEISGGQPA		
Thr. sp. 20892 OrfA	(1706)	VDSIKRVEIMSEVQTLLSVEVSDVDALSRTRTVGDVIEAMKMEIIGGPQGQ		
		1901		1950
Sch. sp. 9695 Pfalp	(1546)	SAPAPAVE-----		
Thr. aureum ORf1	(1827)	APASVAQP-----		
Sch. sp. 20888 OrfA	(1763)	APAAAP-----APAAAAPAVSNELLEKAETVVMVLAAKTGYESDMIES		
Thr. sp. 10212 Pfalp	(1728)	APVQVAAPTQIVAPVQVSAPVDSGLLAKAEQVVLEVLAASKTGYESTELIEL		
Thr. sp. 20892 OrfA	(1756)	TLTAEIR-----		
		1951		2000
Sch. sp. 9695 Pfalp	(1554)	-----		
Thr. aureum ORf1	(1835)	-----QASAP-----		
Sch. sp. 20888 OrfA	(1808)	DMELETELGIDSIKRVEILSEVQAMLNVEAKDVDALSRTRTVGEVVDAMK		
Thr. sp. 10212 Pfalp	(1778)	DMELETELGIDSIKRVEILSEVQAQLNVEAKDVDALSRTRTVGEVIDAMK		
Thr. sp. 20892 OrfA	(1764)	-----		

FIG. 4 (cont'd)

		2001		2050
Sch. sp. 9695 Pfalp	(1554)	-----SAPAASAAPTPAASTAPSA	DLQALLSKAETVVM	AVLAA
Thr. aureum ORf1	(1840)	-----SPSATASAPVTPLAAPAS	VDPAKLANAEAVVME	VLAA
Sch. sp. 20888 OrfA	(1858)	AEIAGSSAPAPAAAPAPAPAAA	APAVSSELLEKAETV	VMEVLAA
Thr. sp. 10212 Pfalp	(1828)	AEIAGGQP--AAPVQVAAPAPV	VPVQVSTPVD	SGLLAKAEQVVL
Thr. sp. 20892 OrfA	(1764)	-----QPEVSEPAVPTSSSSSI	ANVSSARLAEAEAAV	ESVLAD
		2051		2100
Sch. sp. 9695 Pfalp	(1592)	KTGYEADMVEADMILEAELGID	SIKRVEILSEVOGOL	GEAKDVDALSRT
Thr. aureum ORf1	(1877)	KTGYEADMIEADMILLDAELGID	SVKRIEILAAVOAOL	GEAKDVDALSRT
Sch. sp. 20888 OrfA	(1908)	KTGYETDMIESDMILETELGLD	SIKRVEILSEVOAML	NVEAKDVDALSRT
Thr. sp. 10212 Pfalp	(1876)	KTGYETELIELDMELETELGLD	SIKRVEILSEVOAOL	SVEAKDVDALSRT
Thr. sp. 20892 OrfA	(1802)	KTGYDSSMIEMDMILESELGV	DSIKRVEINSEVOAL	LSVEVSDVDALSRT
		2101		2150
Sch. sp. 9695 Pfalp	(1642)	RTVGEVVDAMKAEIVAAAGGS	SAPAAVPSAPAA	SA-----APTATAPS
Thr. aureum ORf1	(1927)	RTVGEVVDAMKAEIGGOATS	SAPASVAQPOAS	APSPSATASVLPKEV
Sch. sp. 20888 OrfA	(1958)	RTVGEVVDAMKAEIAGGS	-----APAPAAAP	-----APAAAP
Thr. sp. 10212 Pfalp	(1926)	RTVGEVIDAMKAEISGGQ	-----PTAPVQVAPTQ	-----VVAPVKVST
Thr. sp. 20892 OrfA	(1852)	KTVGDIIEAMKMEISGPGQ	----QTLTAESIREPP	VSEPAVPTSSSSIA
		2151		2200
Sch. sp. 9695 Pfalp	(1686)	DLQALLSKAETVVM	AVLAAKTGYEADMVE	ADMILEAELGID
Thr. aureum ORf1	(1977)	SVDPAKLANAEAVVME	VLAAKTGYEADMID	ADMILLDAELGID
Sch. sp. 20888 OrfA	(1992)	AVSNELLEKAETVME	VLAAKTGYETDMIES	DMILETELGLD
Thr. sp. 10212 Pfalp	(1965)	PVDSGLLAKAEQVVL	EVLASKTGYETELI	ELDMELETELGLD
Thr. sp. 20892 OrfA	(1898)	NYSSARLAEAEAAV	ESVLADKTGYDSS	MIEMDMILESELGV
		2201		2250
Sch. sp. 9695 Pfalp	(1736)	SEVOGOLGEAKDVD	ALSRTRTVGEVVDAM	KAEIVAAAGGS
Thr. aureum ORf1	(2027)	AAVOAOLGEAKDVD	ALSRTRTVGEVVEA	MKAEIGAAGPND
Sch. sp. 20888 OrfA	(2042)	SEVOAMLNVEAKDVD	ALSRTRTVGEVVDAM	KAEIAGGSAPAP
Thr. sp. 10212 Pfalp	(2015)	SEVOAOLNVEAKDVD	ALSRTRTVGEVIDAM	KAEIAGDQAPAV
Thr. sp. 20892 OrfA	(1948)	SEVQTLSSVEVSDV	DALSRTKTVGDIIEA	MKLEISGSSSIET
		2251		2300
Sch. sp. 9695 Pfalp	(1786)	LLPT-----		
Thr. aureum ORf1	(2073)	-----		
Sch. sp. 20888 OrfA	(2092)	GAAP-----		
Thr. sp. 10212 Pfalp	(2065)	GVANPALLAKAEQV	VLEVLASKTGYETEL	IELDMELETELGLD
Thr. sp. 20892 OrfA	(1998)	TSYKS-----		
		2301		2350
Sch. sp. 9695 Pfalp	(1790)	-----		
Thr. aureum ORf1	(2073)	-----G-----		
Sch. sp. 20888 OrfA	(2096)	-----		
Thr. sp. 10212 Pfalp	(2115)	LSEVQAELSVEAKD	VDALSRTRTVGEVID	AMKAEIAGSAVT
Thr. sp. 20892 OrfA	(2003)	-----		
		2351		2400
Sch. sp. 9695 Pfalp	(1790)	-----LFGSECEDLSL	TFPVITTLPLPAEL	VLAEGGAR
Thr. aureum ORf1	(2074)	-----HLFGTGCE	DLSLCSASVVEIAR	CSELALERPMDRP
Sch. sp. 20888 OrfA	(2096)	-----AVKIDSVH	GADCDLSLMHAKV	VDIRREDELILERP
Thr. sp. 10212 Pfalp	(2165)	MEETDDEDEDFILY	DHYGSECEDLSLS	FSVKSIPRADKL
Thr. sp. 20892 OrfA	(2003)	-----VKASGCEN	VDTRFAKVVOISL	PSKLKSTVSHDRP
		2401		2450
Sch. sp. 9695 Pfalp	(1824)	VVVDGGSALTS	SLVSSLGDRAVLLO	VQSSSACS
Thr. aureum ORf1	(2109)	ILIVSDGSALPA	ALASRLGSCAVILT	TAGETDQSVR---
Sch. sp. 20888 OrfA	(2136)	VIVVDGSELTLA	LVVLGACAVLTTE	EGLQLAQRAGAA
Thr. sp. 10212 Pfalp	(2215)	IVIVDCGKLTTE	LAKAIGERAVVAT	ESAOSLSVSG--
Thr. sp. 20892 OrfA	(2037)	VIVVDGGLPLTE	LCKIILGGNIVVLS	YQKGPAGPRG-----
		2451		2500
Sch. sp. 9695 Pfalp	(1873)	FAALQAALTSVEA	QFGKVGGFVFOFG	-DDIVQAOLGWALL
Thr. aureum ORf1	(2156)	EADLVRALEAVE	SREGVPGGVVLER	ASETARDQLGFALL
Sch. sp. 20888 OrfA	(2186)	AESAERATKEAE	QRFGLGGFISQA	ERFPAEILGFTLM
Thr. sp. 10212 Pfalp	(2263)	ESETERMSSIES	SYGKIGGFVYOH	FHDSYGMOLGWAL
Thr. sp. 20892 OrfA	(2080)	EEALIOALALIR	STYGVPIGFICQ	OVSNVSTKAQLC

FIG. 4 (cont'd)

		2501		2550
Sch. sp. 9695 Pfalp	(1922)	EQIEGGRTFFVAVARLDGQLGSG--K---STTATVDLSRAQOGSVFGLC		
Thr. aureum ORf1	(2206)	QQIPGGRACFVGVSRIDGKLGSG--ACAKGKGWAEAAEIAQOGAVAGLC		
Sch. sp. 20888 OrfA	(2236)	TAVAGGRPAFIGVARLDGFLGFT-----SQGTSDALKRAORGAIFGLC		
Thr. sp. 10212 Pfalp	(2313)	DPIKNGRTFFFLAVARMNGKLGMDN--ASVHDQGIVESCGIAERGAIFGLC		
Thr. sp. 20892 OrfA	(2130)	AVLPDSRSFFVGVVRINGKLGTFENISDFSFKFDLTKALDYGORGSLGLC		
		2551		2600
Sch. sp. 9695 Pfalp	(1967)	KTLDLEWFA--VFCRGIDLAADLDAAQAARCLLGELSDDVAVRESGYSA		
Thr. aureum ORf1	(2254)	KTLDLEWPH--VFARSDIELGANEETAQAIFEEISCDFETVREAGYTK		
Sch. sp. 20888 OrfA	(2279)	KTIGLEWSESDVFSRGNDIAQGMHPEDAATAVAVREMACADIRIREVGIGA		
Thr. sp. 10212 Pfalp	(2361)	KTLDLEWPN--VFARGNDIAEGNSYSLAELINDEISCANLSTRESGYTI		
Thr. sp. 20892 OrfA	(2180)	KSLDLEWEQ--VFCRGIDLACDLMPLQAARILRNELQCPNMRIREVGYDI		
		2601		2650
Sch. sp. 9695 Pfalp	(2015)	SGQRCTTTTKSLTIGKPHQPISSSDLEFLVSGGARGITPLCVRELAQRVGG		
Thr. aureum ORf1	(2302)	DGKRNTTEARPVGLGPKQALRSSDVELVSGGARGITPVCVRELAQSTISG		
Sch. sp. 20888 OrfA	(2329)	NQORCTIRAALKEITGNPQROIADDDVLLVSGGARGITPLCIREITROIAG		
Thr. sp. 10212 Pfalp	(2409)	SGERPTTEAHKLVTGKPHAPIKKKDAFLVSGGARGITPLCIREIAKAVKG		
Thr. sp. 20892 OrfA	(2228)	SGARYTISTDDLICGSPKAKVERADLEFLVTGGARGITPHCVREIASRSPG		
		2651		2700
Sch. sp. 9695 Pfalp	(2065)	GTYVLIGRSELPTTEPAWAVGVESGKPLEKAALAFLEKAEFAAGRGAKPTP		
Thr. aureum ORf1	(2352)	GTFVLIGRS-PIADDPAWACGV-EEANIGTAAMAHLEKAEFAAGRGPKPTP		
Sch. sp. 20888 OrfA	(2379)	GKYLIGRSKVSASEPAWCAGITDEKAVOKAATQELKRAFSAGEGPKPTP		
Thr. sp. 10212 Pfalp	(2459)	GTYILNGRS-ALADEPLWANGK-SGKDLQKAGLAFLEKAEFAAGRGSKPTP		
Thr. sp. 20892 OrfA	(2278)	TTFVLIGRS-EMSDPEDWAVGH-YNKDLQSTMKHLKATHAAG-GVKPTP		
		2701		2750
Sch. sp. 9695 Pfalp	(2115)	MLHKKLVGAVVGAREVPASLAEITAQATAVYESCDVSSAAKVRERVERV		
Thr. aureum ORf1	(2400)	KAHKLVGSVLGAREVLGSLESIRAQGAAREYVSCDVSCAERVKAIVDDI		
Sch. sp. 20888 OrfA	(2429)	RAVTKLVGSVLGAREVFSIAAIEALGGKAIYSSCDVNSAADVAKAVRDA		
Thr. sp. 10212 Pfalp	(2507)	KVHKSLIDKVLGIREVRASTIANIEAHGAKAIYLSCDVSSAEKVKAAVQKV		
Thr. sp. 20892 OrfA	(2325)	KAHPALNVRVTGSREVBESLRAIQEAGANVEYIACDVSDENKVRQIVQRV		
		2751		2800
Sch. sp. 9695 Pfalp	(2165)	QQQGRPVSGVFHASGVLRDKLVENKSLADFSAVYDTKVGGGLINELACVD		
Thr. aureum ORf1	(2450)	ERRVG-AVTGVVHASGVLRDKSVRELEADFEVVYGTQKVDGLINELQAVD		
Sch. sp. 20888 OrfA	(2479)	ESQLCARVSGIVHASGVLRDRLEIEKKLPDEFDAVFGTKVTGLENLLAAVD		
Thr. sp. 10212 Pfalp	(2557)	EKEHLVRITGIVHASGVLRDKLVENKTLDDFNAVYGTQKVTGLYNLLSAVN		
Thr. sp. 20892 OrfA	(2375)	EOKYCEITGIWHASGVLRDKLVECKTTDDFEAVFGTKVTGLYNIVSQVN		
		2801		2850
Sch. sp. 9695 Pfalp	(2215)	LAQLRHLLVLFSSLAGFHHGNGVQSDYAMANEALNKLAAHLSAVHPQLCAR		
Thr. aureum ORf1	(2499)	RPKLRHLLVLFSSLAGFHHGNTGQAVYAMANEALNKNMFHLETAMPGLSVKT		
Sch. sp. 20888 OrfA	(2529)	RANLKHMLVLFSSLAGFHHGNGVQSDYAMANEALNKNMLELAKDVS---VKS		
Thr. sp. 10212 Pfalp	(2607)	MNFVRHLLVLFSSLAGYHGNVQSDYAMANESLNKIGFRLGAAYSQLCVK		
Thr. sp. 20892 OrfA	(2425)	MSKLRHFLVLFSSLAGFHHGNGQTDYATANEALNKNIAHTLSAFLPKLNAKV		
		2851		2900
Sch. sp. 9695 Pfalp	(2265)	ICFGPWDG-GMVTPLKANFIRMGIQIIPREGGAQTVANMLVSSSPGOLL		
Thr. aureum ORf1	(2549)	IGFGPWDG-GMVNDALKAHFASMGVQIIPLDGGAGTVSRITGACSPTOVL		
Sch. sp. 20888 OrfA	(2576)	ICFGPWDG-GMVTPLKQFQEMGVQIIPREGGATVARIVLGSSPAEIL		
Thr. sp. 10212 Pfalp	(2657)	ICFGPWDG-GMVTPLKQFQSMGVQIIPREGGATVARIVLSSNPSONVL		
Thr. sp. 20892 OrfA	(2475)	EDFGPWWGSGMVTETLEKHFKAAGVQIIPLEGARTVAQIILLASSPQSL		
		2901		2950
Sch. sp. 9695 Pfalp	(2314)	VGNWGVPPVPSATEHTVLOTLRQS-DNPFLLDSHVIOGRVLPMTLAVGX		
Thr. aureum ORf1	(2598)	VGNWGVPPVPSASVHKITVRLGGESANPFLLSSHTIOGRVLPMTVALGL		
Sch. sp. 20888 OrfA	(2625)	VGNWRTPSKKVGSDEITLHRKISAK-SNPFLEDHVIOGRVLPMTLAIGS		
Thr. sp. 10212 Pfalp	(2706)	VGNWGVPPVPSPLKSAITVQFTPE-LNPFLKSHQIHGKNVLPMTVAIGX		
Thr. sp. 20892 OrfA	(2525)	VGNWGFPAFKPLQRSNVVTGTLSP-EIEFTADHKIOGRVLPMMATIGF		
		2951		3000
Sch. sp. 9695 Pfalp	(2363)	MAHQAAQSTYAGHQLWAVEDAQLFKGIAIDNGADVPRVVELSRRKEEQEDA		
Thr. aureum ORf1	(2648)	LAEAARGLYVGHQVVGIEDAQVFGGVLDKGATCEVQLRRESSTASP---		
Sch. sp. 20888 OrfA	(2674)	LAETCLGLFPGYSLWADDAQLFKGVTVDGDVNCEVTLTPSTAPSG----		
Thr. sp. 10212 Pfalp	(2755)	LAHLVKNFYAGHHLNGVEDAQLFSGVVIDHVAQAVKLTEQSLDDDG---		
Thr. sp. 20892 OrfA	(2574)	MASIAEGLYPGYNLOGVENAQLFQGLTINQETKFOITLIEHNSEN---		

FIG. 4 (cont'd)

		3001		3050
Sch. sp. 9695 Pfalp	(2413)	GKVVVKVOVLLKSQVN-GKSVPAYKATVVLS-PAPPPSVITRDFDIT---		
Thr. aureum ORF1	(2695)	---SEVVLASALNVFAAGKVVPAVPAHVVLGASGPPETGGVQLLKLKDLGVD		
Sch. sp. 20888 OrfA	(2720)	---FVNVQATLKTFFSS-GKLVPAVPAVIVLSNQGAPPANATMQPPSL---		
Thr. sp. 10212 Pfalp	(2802)	---FVKVQAVLTASNDNGKMPAYKAVIVLG-KTSRPAFILEKDFSLQ---		
Thr. sp. 20892 OrfA	(2621)	----LDVLTSLGVMLESGKVLPAVRCVVCLNTTQQQPKLSPKILNLEVDP		
		3051		3100
Sch. sp. 9695 Pfalp	(2458)	--PDPACTEHDLYDGKTLFHGKAFQIEQVLSATPKOLTAKCNLPLTPE		
Thr. aureum ORF1	(2742)	ADPACSVGKGALYDGRTLFHGPAFOYMDVLRCSPAELAVRCHVVPASAAQ		
Sch. sp. 20888 OrfA	(2763)	--DADPALQGSVYDGKTLFHGPAFRGIDVLSCTKSQLVAKCSAVPGSDA		
Thr. sp. 10212 Pfalp	(2845)	--EENRSRADELYDGKTLFHGPELFRGITKLLNVSDTSLTTQCTNIDITAT		
Thr. sp. 20892 OrfA	(2667)	-----ACEVNPYDGKSLFHGELLQFVQOVLHSSTKGLVAKCPALPIKEA		
		3101		3150
Sch. sp. 9695 Pfalp	(2506)	QRGQFVVNLSQODPFOADIAFOAMLVWARMIRQSAALPNICERFDIYKPM		
Thr. aureum ORF1	(2792)	DRGQFVSRGVLYDPFINDTVFOALLVWARLVDRDSASLPSNVERTISFHGQP		
Sch. sp. 20888 OrfA	(2811)	ARGEFATDTDANDPFIYNDIAFOAMLVWVRRTLGOAALPNSTQRTIVQHRPV		
Thr. sp. 10212 Pfalp	(2893)	ER-GQFADIEPVNPFMADAAFOAMLVWVRNLPNSASLPNICERVDIYKPI		
Thr. sp. 20892 OrfA	(2711)	IR-GPFIKQTLHDPIEDDVIIFQLMLVWCRNALGSASLPNRIEKMSYFGNV		
		3151		3200
Sch. sp. 9695 Pfalp	(2556)	AP-GATYYTTSVKLASAP-LVDSVCKCTVAMHDEQGEVYFSARASVVLNK		
Thr. aureum ORF1	(2842)	PSEGEVYTTTLKLDASAAGPLDPIAKAQFFLHRACGAVFASGRASVVLNK		
Sch. sp. 20888 OrfA	(2861)	PQ-DKPFYITLPSNQSG---GHSQHKHALQFHNEQGDLFIDVOASVIATD		
Thr. sp. 10212 Pfalp	(2942)	AP-GEKYYTTLQALGNT---SGSVLKSVEFYMHDEQGEVFLSGPASVVVND		
Thr. sp. 20892 OrfA	(2760)	SE-GSTFEASVTPVGPR-VPKDPVIKMOFLLODESGNTFSSGEGSVVLSD		
		3201		
Sch. sp. 9695 Pfalp	(2604)	TLTY-		
Thr. aureum ORF1	(2892)	ALSF-		
Sch. sp. 20888 OrfA	(2907)	SLAF-		
Thr. sp. 10212 Pfalp	(2988)	KNEF-		
Thr. sp. 20892 OrfA	(2808)	ELVF-		

FIG. 5

	1	50
Sch. sp. 9695 Pfa2p	(1)	-----MPCDNI
Sch. sp. 20888 OrfB	(1)	-----MAARNVSAAHMHDEKRI
Thr. sp. 10212 Pfa2p	(1)	-----MVKLSVGDNICHQOPV
Thr. aureum ORF2	(1)	QAIGHRAARWSCRSKSKARGHKAQKEMNQGGNDEGVSVARADPCPTRI
Thr. sp. 20892 OrfB	(1)	-----MQLPPAHSADENRI
	51	100
Sch. sp. 9695 Pfa2p	(7)	AVVGMAVOYAGCNIQDEFWDTLMRKEINSSPISAERLGRYRDLHSHFOR
Sch. sp. 20888 OrfB	(19)	AVVGMAVOYAGCNIQDEFWEVLNMGKVESKVISDKRLGSENYRAEHYKAER
Thr. sp. 10212 Pfa2p	(17)	AVVGMAVMYAGCNIQHEFWQSLQGNMISKSISQNLGSEYRSEHKKFER
Thr. aureum ORF2	(51)	AVVGMAVEYAGCNGKSAFWNTLMGKINSACISDDRSGSARRSEHYAPER
Thr. sp. 20892 OrfB	(15)	AVVGMAVKYAGCDIKKEFWKTLNGSINTKSISAARLGSNKRDSEHYVFER
	101	150
Sch. sp. 9695 Pfa2p	(57)	SKYADTFCNDRYGCNEASVDN-EHELLADLARRAILLDAGIN--LDDASTT
Sch. sp. 20888 OrfB	(69)	SKYADTFCNETYGTLDENEIDNEHELLLNLAQALASTS-----V
Thr. sp. 10212 Pfa2p	(67)	SKYSDTFCNERYGCIDENYQS-EHELLLKLADALADTKG-----S
Thr. aureum ORF2	(101)	SKYADTFCNERYGCIDPKVDN-EHDLGLLGLAAALODADRRSDGGKFDP
Thr. sp. 20892 OrfB	(65)	SKYADTFCNERYGCIQOGTDN-EHDLGLLGLAQEALADAAGRMEKQ-PSEA
	151	200
Sch. sp. 9695 Pfa2p	(104)	ANLRDFGIVSGCLSFPMNDLQCELLNLYQVHVENVYGAQRFDRSR-PWSE
Sch. sp. 20888 OrfB	(109)	KDSTCGIVSGCLSFPMNDLQCELLNLYQNHVEKKLGARVFKDAS-HWSE
Thr. sp. 10212 Pfa2p	(107)	IDENLTGIVSGCLSFPMNDLQCELLNLYQCHIEKNGPNALKVDN-LWSK
Thr. aureum ORF2	(150)	AOLKTCGIVSGCLSFPMNDLQCELLNLYQAHAEPRIKHKCFADOT-PWST
Thr. sp. 20892 OrfB	(113)	FDLENTGIVSGCLSFPMNDLQCELLNLYQSHVEKQPPSALVEAVKLWSE
	201	250
Sch. sp. 9695 Pfa2p	(153)	RPVSVSEASDPFVSDPASFVANQLGLGPNRYSLDAACASALYCLKLAS
Sch. sp. 20888 OrfB	(158)	REQNKEPAGDRRIIMDPASFVAELNLGALHYSNDAACATALLYVLELAQ
Thr. sp. 10212 Pfa2p	(156)	RTTNG---KDDKCAFDPASFVAEOLDNGFLHYSLDAACASALYVLELAQ
Thr. aureum ORF2	(199)	RTALHLELPDGPPTHDRDPASFVAGQLGLGELHYSLDAACASALYVLELAQ
Thr. sp. 20892 OrfB	(163)	ROKSTKAHAGDRRFIDPASFVAOKLNLGELHYAIDAACASALYVLELAQ
	251	300
Sch. sp. 9695 Pfa2p	(203)	DHLLSRADYMLCGATCFPPDPFFILSGFSTFOAMPVGGPDDNPNSSVPLRQ
Sch. sp. 20888 OrfB	(208)	DHLLSGAADYMLCGATCLPEPFFILSGFSTFOAMPVGTGQ--NVSNPLHK
Thr. sp. 10212 Pfa2p	(203)	DHLLSGAADTMLCGASCLPEPFFILSGFSTFHAMPVSG---DVSAPLHK
Thr. aureum ORF2	(249)	DHLLSGEADYMLCGATCFPEPFFILSGFSTFHAMPVGEN---GVSNPFHF
Thr. sp. 20892 OrfB	(213)	DHLLSGAVDYMMLCGATCFPEPFFILSGFSTFOAMPXGAD---GVSNPLHK
	301	350
Sch. sp. 9695 Pfa2p	(253)	GSQGLTPGEGGAIMVLKRLEDAIRDGDRIYGTLLGTSLSNAGCGLPLSPH
Sch. sp. 20888 OrfB	(256)	DSQGLTPGEGGSIMVLKRLDDAIRDGDHIYGTLLGANVSNSTGLPLKPL
Thr. sp. 10212 Pfa2p	(249)	TSQGLTPGEGGAIMVLKRLNDAIRDGDRIYGTLLGAELSNAGCGLPLSPH
Thr. aureum ORF2	(296)	DSQGLTPGEGGSIMVLKRLADAERDGDHIYGTLLGASLSNAGCGLPLKPH
Thr. sp. 20892 OrfB	(260)	TSAGLTPGEGGSIMVLKRLKDAIRDGNHIYGVILLEANLSNAGCGLPLSPH
	351	400
Sch. sp. 9695 Pfa2p	(303)	LPSEKSCHEDLATSVGIDF-SENQYVECHATGTPOGDRVVEALNHCFRG
Sch. sp. 20888 OrfB	(306)	LPSEKKCIMDTYTFINVHE-HKIQYVECHATGTPOGDRVEIDAVNACFEG
Thr. sp. 10212 Pfa2p	(299)	NPSEFDCMEKALQVRHRL-SSIQYVECHATGTPOGDKVEIDANTKCFGE
Thr. aureum ORF2	(346)	QPSEESCCKATYELGVPE-RDQYVECHATGTPOGDTVELQAVNACFEG
Thr. sp. 20892 OrfB	(310)	LPSEESCRRDTYRAGVAADQSIQYVECHATGTPRGDRVVEISAVERVFKK
	401	450
Sch. sp. 9695 Pfa2p	(352)	NTDHPPRNGSTKGNFGHTLVAAGFAGMAKVLLSMOHTIPTTPGVDRSN-
Sch. sp. 20888 OrfB	(355)	---KSPRFGSTKGNFGHTLVAAGFAGMCKVLLSMKHGIIPTTPGIDDETK
Thr. sp. 10212 Pfa2p	(348)	---HSPRFGSTKGNFGHTLVAAGFAGMCKVLLSMOGEIPTTPGLENPDN
Thr. aureum ORF2	(395)	---ASPRFGSTKGNFGHTLVAAGFAGMCKVLLMERGIIPTTPGVDSGT-
Thr. sp. 20892 OrfB	(360)	---NTPRFGSTKGNFGHTLVAAGFAGMAKILLMEHGVIIPTTPGLDASN-
	451	500
Sch. sp. 9695 Pfa2p	(401)	CSDPLVYDEATWPYSSQARAGKPGDELKCASLSAFGFGGTNAHCVFRE
Sch. sp. 20888 OrfB	(402)	MDPLVSGEATWPPTNGEPRK-----AGLSAFGFGGTNAHAVFEE
Thr. sp. 10212 Pfa2p	(395)	INDHLVYTETIWPNTNGDLK-----ACLSAFGFGGTNAHAVFEE
Thr. aureum ORF2	(441)	QSDPLVYTAAPWPDTRGGPK-----AGLSAFGFGGTNAHAVFEE
Thr. sp. 20892 OrfB	(406)	QASEHVYKAITWPTTHGAPK-----AGLSAFGFGGTNAHAVFEE

FIG. 5 (cont'd)

	501		550
Sch. sp. 9695 Pfa2p	(451)	HRQIAATATASPVLP-----EVTGPPIAIIIGMDATFGSLKGLDAFEQA	
Sch. sp. 20888 OrfB	(443)	HDPSNACTGHDSISALSARCGGESNMPIAITGMDATFGALKGLDAFERA	
Thr. sp. 10212 Pfa2p	(436)	YRSDLOANKTLENE--S-KSHEIFSSFKIAIVGMESEFGTLKGLQEFFERA	
Thr. aureum ORF2	(482)	HIPSRAPPVAVLCQPRLG-----SGPNRFLAIVGMDATFGSLKGLSALEAA	
Thr. sp. 20892 OrfB	(447)	FNAEGISYRPGKPP-----VESNTRPSVITGMDCTFGSLEGIDAFETA	
	551		600
Sch. sp. 9695 Pfa2p	(494)	LYKGTDGASDLPSKRWRFLGADTDFLTANGLDAVPRGCIYRDVDVDYKRL	
Sch. sp. 20888 OrfB	(493)	IYTCAGCAIPLPEKRWRFLGKDKDFLDLCGVKATPHGCIYEDVDVDFORL	
Thr. sp. 10212 Pfa2p	(483)	IYNGGHCACDLPENRWRFLGEDKEFLQACGLQKLPKGCYIKEVETDFKRL	
Thr. aureum ORF2	(527)	IYEARRHARPLPAKRWRFLGGDESFLHEIGLECSPHGCIYEDVDVDFKRL	
Thr. sp. 20892 OrfB	(491)	IYEGRDAARDLPAKRWRFLGEDLEFLRAIRLKEKPGECIVEVDVDFKRL	
	601		650
Sch. sp. 9695 Pfa2p	(544)	RSPMIPEDVLRPOQLLAVATMDRAIODAGMATGGKVAVLVGLGTDTELYR	
Sch. sp. 20888 OrfB	(543)	PTPMTPEDMLLPOQLLAVTTIDRAILDSGMKKGGNVAVFVGLGTDDELYR	
Thr. sp. 10212 Pfa2p	(533)	RLPMIOEDILRPLLOLLAVSIIDRAINASGVKPNQKVAVLVGLGTDDELYR	
Thr. aureum ORF2	(577)	PTPMVPEDILRPOQLLAVSTIDKAILDSGLAKGGNVAVLVGLGTDDELYR	
Thr. sp. 20892 OrfB	(541)	KTPITPEMLRPOQLLAVSTMDRAIIDAGLIKGOHVAVLVGLGTDDELYR	
	651		700
Sch. sp. 9695 Pfa2p	(594)	HRARVTLKERIDP---AAFSPEQVQEMMDYINDCGTSTSYTSYIGNLVAT	
Sch. sp. 20888 OrfB	(593)	HRARVALKERVR-----PEASKKINDIMQYINDCGTSTSYTSYIGNLVAT	
Thr. sp. 10212 Pfa2p	(583)	HRARVALKERLQTA--KEDIPLLEKIMNYVNDRGTSYTSYIGNLVAT	
Thr. aureum ORF2	(627)	HRARVALKERLQGLVRSAGGALTSRLMNYINDSGTSTSYTSYIGNLVAT	
Thr. sp. 20892 OrfB	(591)	HRARVALKEVLHPSL--KSDTAILOKIMQYVNDAGTSTSYTSYIGNLVAT	
	701		750
Sch. sp. 9695 Pfa2p	(641)	RVSSQWGFTGPSFTVTEGANSVYRCLELGKFLDTHQVDVAVVVAGVDLCA	
Sch. sp. 20888 OrfB	(638)	RVSSQWGFTGPSFTITEGNSVYRCALGKYLLETGEVDGVVVAGVDLCC	
Thr. sp. 10212 Pfa2p	(631)	RVSSLWGFTGPSFTITEGENSVYRCLELGRVFLANGEVDVAVVVAGVDLCC	
Thr. aureum ORF2	(677)	RVSSQWGFTGPSFTVTEGANSVYRCAQLAKYMLDGEVDVAVVVAGVDLCC	
Thr. sp. 20892 OrfB	(639)	RVSSQWGFTGPSFTVTEGNSVYRCAQLAKDMLQVNRVDVAVVIAGVDLNG	
	751		800
Sch. sp. 9695 Pfa2p	(691)	SAENLYLKARRSAISRQDHPRANFEASADGYFAGEGSGALVLEKROADVGS	
Sch. sp. 20888 OrfB	(688)	SAENLYVKSRRFKVSTSDTPRASFDAAADGYFVGECCGAFVLEKRETSCTK	
Thr. sp. 10212 Pfa2p	(681)	SAENLYVKSRRSKVSTQNEPFANFESNADGYFAGDGGGALVLEKRLSDCTD	
Thr. aureum ORF2	(727)	SAEAFVRSRRMOISKSORPAAPFDRAADGFFAGEGCGALVFKRLIDCVS	
Thr. sp. 20892 OrfB	(689)	SAESFFVRANRQKISKLSHPCASFDRDADGFFAGEGCGALVFKRLIEDCAP	
	801		850
Sch. sp. 9695 Pfa2p	(741)	-DDRYVYAGVAGITCAQPAEAVSPLLLQVHNDNEKRVVEMVELAADSGR	
Sch. sp. 20888 OrfB	(738)	-DDRIYACMDATVPGNVPSACIPEALDOARVKPGD---IEMLELSADSAR	
Thr. sp. 10212 Pfa2p	(731)	STEKIYATVDSIAVGEVGPITKQALKNASTAAKD---IELAELSASSGK	
Thr. aureum ORF2	(777)	-GERIYASLDSVVATTTPRAALPAAAGSARVDPAS---IDMVELSADSHR	
Thr. sp. 20892 OrfB	(739)	-QEKIYASIDSIAIDKEPTSSAVKAVYQSDSSLSD---IELLEISGDSKR	
	851		900
Sch. sp. 9695 Pfa2p	(790)	HAPHLANSPLSAESQLEQVSKLLAHQVPG-----SVAIGSVRANVGDVGYA	
Sch. sp. 20888 OrfB	(784)	HLKDPSVLPKELTAEEETGGIQTILRDDDKLPRNVATGSKVATVGDVGYA	
Thr. sp. 10212 Pfa2p	(778)	HHSGRITCEDEINELGFIIFNEG-----IQRVAIGSVRANVGDVGYA	
Thr. aureum ORF2	(823)	FVRAPGTVAQPLTAEEVGVAREVIGTAGRGSRSVAVGSVRANVGDAGEA	
Thr. sp. 20892 OrfB	(785)	FAAFEGAVEIQSSVEAQLKGLSKVLEPAK--GGQVAVGSTRAATVGDIGYA	
	901		950
Sch. sp. 9695 Pfa2p	(836)	SGAASLIKALCLNRYLE---ANPOWERPVAAPVSEALFTCPRSRAWLKN	
Sch. sp. 20888 OrfB	(834)	SGAASLIKAALCLYNRYLEPSNGDDWDEPAPEAPWDSTLFACQTSRAWLKN	
Thr. sp. 10212 Pfa2p	(819)	SGAASLIKALCLNRYLE-KLPNNKPKTKDVENSKSFFVCEHSRAWLKN	
Thr. aureum ORF2	(873)	SGAASLYKTALCLNRYLA-ATPGWDAPAAGVDEGRELIVCRESRRAWLKN	
Thr. sp. 20892 OrfB	(833)	SGAASLIKALCLNRYLE-ALANWSGFCEQSAWGSNNFVCHETRPWNKN	
	951		1000
Sch. sp. 9695 Pfa2p	(883)	PGESRLAAVASASESGSCFGVLLTDEYATHESNRLSLDDAAPKLIATRG	
Sch. sp. 20888 OrfB	(884)	PGERRYAAVSGVSETRSCYSVLLSEAEGHYERENRISLDEEAPKLIVIRA	
Thr. sp. 10212 Pfa2p	(868)	VDENRHAVVSGVCENGSCYIGIVMSDVQGHHEESNLVSLDKNEPKVLGIYG	
Thr. aureum ORF2	(922)	AGVARHAATISGVDEGGSCYGLVLSVPGQYETGNRISLQAESPKILLLSA	
Thr. sp. 20892 OrfB	(882)	QNEKRCALISGTDPSHTCFSLVLSDTG-CYEEHNRTCFDVOAPQVLVLIHG	

FIG. 5 (cont'd)

		1001		1050
Sch. sp. 9695 Pfa2p	(933)	DTVDDINAKVNAELALIRAHAEIGSATDDDDPAAAVAFTAHRLRFLRLNGE		
Sch. sp. 20888 OrfB	(934)	DSHEEILGHLDKIRERFLOPTGAAPRESELKAQARRIFLE--LIGETIAQ		
Thr. sp. 10212 Pfa2p	(918)	DSVDDIEVQINKYLEKFLOETGTAAAAQKVKSPPTIDIDSN--VFAEMINL		
Thr. aureum ORF2	(972)	PDHAALDDVAELAALAEQADGLSAAAAAVDR-----LLGESILVG		
Thr. sp. 20892 OrfB	(931)	FDGKTIYRLEGYLLELVEGHASPS-----EYFHK--LIGQSILE		
		1051		1100
Sch. sp. 9695 Pfa2p	(983)	TVASHGATATCLALLTIPEKLEKELELAAGVPRSAKAGFNWMSPSGSA		
Sch. sp. 20888 OrfB	(982)	DAASSGSOKPLALSIVSTPSKLOREVELAAGKIPRCIMFRDWSSPAGSR		
Thr. sp. 10212 Pfa2p	(966)	PODKN---KKFAVALVTIPNKLOREIELAVKGIIPRCVKAARDWCSPSGSI		
Thr. aureum ORF2	(1012)	CAAGS---GGITLCIVASPASLHKEIALAHFGIIPRCIKARRDWASPAAGSY		
Thr. sp. 20892 OrfB	(969)	NSKES---K-LTSLVNCNPNQLOKELMLAIKGVORSMLTGKDWVSPSGSC		
		1101		1150
Sch. sp. 9695 Pfa2p	(1033)	FAPTPVTSDFVAFMYGEGRSPYYGVGLDLHRIWPALHERINDKTAALWEN		
Sch. sp. 20888 OrfB	(1032)	FAPEPLASDFVAFMYGEGRSPYYGITQDIHRIWPELHEVINEKTNRLWAE		
Thr. sp. 10212 Pfa2p	(1013)	FACNPLKSDNEAFMYGEGRSPYAGLGYDLHRIWPMHLVNNPTTELWDQ		
Thr. aureum ORF2	(1059)	FAPEPIASDFVAFMYGEGRSPYCGVGRDLHRIWPALHERVNAKTVNWLGD		
Thr. sp. 20892 OrfB	(1015)	FAPNPLSSAKVAFMYGEGRSPYCGVGLGLHRIWPLHENVNNKTVDLWTE		
		1151		1200
Sch. sp. 9695 Pfa2p	(1083)	GDSWIMPRAVDASISQRAVQTAFDADQIEMFRTGIFVSICTDYARDVLGV		
Sch. sp. 20888 OrfB	(1082)	GDRWIMPRASFKESELESQQQEFDRNMIEMFRLGILTSTIAFTNLARDVLNI		
Thr. sp. 10212 Pfa2p	(1063)	GDSWYLPRESSVAEKEKVFQDFDKNOIEMFRLGIFVSMCFTDMATELLGL		
Thr. aureum ORF2	(1109)	GDWLLPRATSAREEEQLCRNFDSNOVEMFRTGVYISMCLTDIAPSLIGL		
Thr. sp. 20892 OrfB	(1065)	GDGWLYPRTLREHTKAIESFNANOIEMFRAGIFISMCQTDYVMNVLCV		
		1201		1250
Sch. sp. 9695 Pfa2p	(1133)	QPKACFGLSLGEISMIFALSRRNCGLSDOLTORLRTSPVWSTOLAVEFOA		
Sch. sp. 20888 OrfB	(1132)	TPKAAFGLSLGEISMIFAFSKKNGLSDLTKDLRESDVWNKALAVEFNA		
Thr. sp. 10212 Pfa2p	(1113)	KPKAAFGLSLGEISMIFAFSKKNTKLSKELTERLKEAKVWASOLAVEFAA		
Thr. aureum ORF2	(1159)	GPKAAFGLSLGEISMIFALSSENCRLSEEMTERLSESPVWNSELAVEFNA		
Thr. sp. 20892 OrfB	(1115)	QPKAGFGLSLGEISMIFAMSKENCROSOEMTNRLGSEPVWSELAINFNA		
		1251		1300
Sch. sp. 9695 Pfa2p	(1183)	LRKLWNYPADAPVESFWQGYIVRASRAETIEKAIGPDNPFVRLLIIVNDS		
Sch. sp. 20888 OrfB	(1182)	LREAWGIPQSVPKDEFWQGYIVRGTKODIEAAIAPDSKYVRLTIINDANT		
Thr. sp. 10212 Pfa2p	(1163)	TRDLWNPADKSIDDEFWQGYFVYANPTLVENTIG--ENKPFVRLLIIVNDS		
Thr. aureum ORF2	(1209)	LRKLWGPAPGAPVDSFWQGYIVRATRAQVEQAIGEDNOFVRLLIIVNDS		
Thr. sp. 20892 OrfB	(1165)	TRKLWKIPRGAPLESFWQGYIVHGTRREEVEHAIGLSEPVRLLIIVNDS		
		1301		1350
Sch. sp. 9695 Pfa2p	(1233)	ALIAGKPAECLRVLEPLGGRLPPMPVKQGMIGHCPEVAPYTPGIARIHET		
Sch. sp. 20888 OrfB	(1232)	ALISGKPDACKAAIAPLGGNIPALPVTQGMCGHCPEVGPIYTKDIKIHAN		
Thr. sp. 10212 Pfa2p	(1212)	CLIAAGKPDCEQKVIKHLHLKLPVPVVTQGMIGHCPEAIPYLDQISHIHE		
Thr. aureum ORF2	(1259)	VLIAGKPAACEAVIARIGSILPPLQVSQGMVGHCAEVLPTTSEIGRIHNN		
Thr. sp. 20892 OrfB	(1215)	ALIAGKPDACQAVISRLNSIFPSPVPVKQGMIGHCPEVRAFIKDIGYIHET		
		1351		1400
Sch. sp. 9695 Pfa2p	(1283)	LEIFDSPVKMYTSTVTNAELRG-----GNSNSTEFVQKL		
Sch. sp. 20888 OrfB	(1282)	LEFFVVVGLDLWTTINQIRLV-----PRATGAKDEWAPSSFGEXAQOL		
Thr. sp. 10212 Pfa2p	(1262)	LEIFKPKENVKLFITISENF-----ELVSMKDSVSKLVAEI		
Thr. aureum ORF2	(1309)	LRFSQDETGGCKMYSSVNSRIGPVEESQMGPGTELVFSPSMEDEFAQOL		
Thr. sp. 20892 OrfB	(1265)	LRFSNDYSDCQLFSAVTKG-----ALDSSSTMEIKHEVGEV		
		1401		1450
Sch. sp. 9695 Pfa2p	(1317)	YTRIADFPGIVDKVSRDG--HDVFVEVGPNMMSAAVSDILGKAATPHVSV		
Sch. sp. 20888 OrfB	(1325)	YEQANFPQIVETIYKQN--YDVFEVGPNMHRSTAVRTTLGEPQR-NHLAG		
Thr. sp. 10212 Pfa2p	(1296)	YQHVAADFPMIYNKYKETCKTDIFTELGSNNYRSGAVKTLGEP---EIVSV		
Thr. aureum ORF2	(1359)	YSPVADFPAPTEAVYQOG--HDVFVEVGPDHRSRAAVSTLGPETR-RHIAV		
Thr. sp. 20892 OrfB	(1300)	YSPVADFPQIVNTVHSAG--YDVFEILGCDASRSAAVQNTILGGQG-KFLST		
		1451		1500
Sch. sp. 9695 Pfa2p	(1366)	AIDRPSESAWTOTLKSTALLTAHRVPLHNPTLFDADLYHPTFLTAIDSAMQ		
Sch. sp. 20888 OrfB	(1373)	AIDKQNEADAWETIVKLVASLKAHLVPGVTISPLYHSKLVAAEQACYAALC		
Thr. sp. 10212 Pfa2p	(1343)	AIDKQNETAWGOLMKMVASLISHRVPGVELKKLYHPPELLKFDPAQAKPNRF		
Thr. aureum ORF2	(1407)	ANDKKGESAWSOLLKMLATLASHRVPGLDLSSMYHPAVVERCRLALAAQR		
Thr. sp. 20892 OrfB	(1348)	AIDKKGHSAWSOVLRATASLAHRVPGISILDLPHENFREMCTMATT--		

FIG. 5 (cont'd)

		1501		1550
Sch. sp. 9695 Pfa2p	(1416)	E-P ~~PPKPNRFLRS~~ ESNGYFCPDGISKQVAAASAKPSTHCMVRLHPAKAV		
Sch. sp. 20888 OrfB	(1423)	KGEKPKK ~~NKFVRKIQINGP~~ ENSKADP ~~ISSADLIASFP~~ ADPAIEAAISSRI		
Thr. sp. 10212 Pfa2p	(1393)	IRNIELNGFFDR ~~TNITVDQLSPADPKLAEIV~~ NNRNMPKDNVYVPIERVK		
Thr. aureum ORF2	(1457)	SGQ ~~PEQRNKF~~ LR ~~TIEVNGFY~~ DPADAT ~~IPEAVATILPATAAIS~~ PPKLGAPH		
Thr. sp. 20892 OrfB	(1396)	---PKVED ~~KFLRTIQINGP~~ EKEMIHLED ~~TTLSCLPAPSEANIAA~~ IOSRS		
		1551		1600
Sch. sp. 9695 Pfa2p	(1465)	~~VVAAAGAVVADSTPVVKAKOT~~ SSS-----		
Sch. sp. 20888 OrfB	(1473)	~~NKPVA~~ PKFYARLNIDEQDETRDPILNK ~~DNAPSSSSSSSSSSSSSSSS~~ SPSPA		
Thr. sp. 10212 Pfa2p	(1443)	TMIK ~~AEPLNQVSVGSKPVVTERISSDN~~ LFEKLSEITKSFDG-----		
Thr. aureum ORF2	(1507)	DSQPE ~~EARPVGEASVPRRATSSSKLARTLAID~~ ACDSDVRAALLDLDAPI		
Thr. sp. 20892 OrfB	(1443)	~~IRSAARS~~ QSHDCASHSHEENK ~~DSCPEKLKLD~~ SVSVAINF ~~DN~~ -----		
		1601		1650
Sch. sp. 9695 Pfa2p	(1489)	-----		
Sch. sp. 20888 OrfB	(1523)	PSAPVQKKAAPAAETKAVASADALRSALLDLDSMLALSSASASGNLVETA		
Thr. sp. 10212 Pfa2p	(1486)	-----VN-----		
Thr. aureum ORF2	(1557)	AVG-----G-----		
Thr. sp. 20892 OrfB	(1486)	-----		
		1651		1700
Sch. sp. 9695 Pfa2p	(1489)	-----LLYGDDAF ~~LCYDVWPLYMGAMAEGISSV~~ DLVVAAG		
Sch. sp. 20888 OrfB	(1573)	PSDASVIVPPCN ~~IADLGSRAFN~~ ITYGVSAPLYTGAMAKGIASADLVIAAG		
Thr. sp. 10212 Pfa2p	(1488)	-----ACTEAM ~~IGDSG~~ ---FLKTYEVDYPLYTGAMAKGIASADLVIAAG		
Thr. aureum ORF2	(1561)	--SSRAQVPPCP ~~ISALGSAAFRAAHGV~~ DYALYMGAMAKGVASAENVIAAG		
Thr. sp. 20892 OrfB	(1486)	-----DDRIQLGHAGFREMYNTRYSLYTGAMAKGIASADLVIAAG		
		1701		1750
Sch. sp. 9695 Pfa2p	(1526)	EAPMLASFGAARLP ~~MDVELOIREIQORTSN~~ -AFAVNLM ~~PGPDEAAT~~ ---		
Sch. sp. 20888 OrfB	(1623)	~~EOGILASFGAGGLPMQVVR~~ ESIEKIQAA ~~LNGPYAVNLIHSPFDSN~~ LEKG		
Thr. sp. 10212 Pfa2p	(1529)	~~KSKILASFGAGGLALQVVEDAIKQIKAE~~ LNGPF ~~AVNLIHSPFDS~~ PLEKG		
Thr. aureum ORF2	(1609)	~~KAPMLASFGAGGLPLGEVEEADIKIQAA~~ IEGPF ~~AVNLIHSPFDPN~~ LEEG		
Thr. sp. 20892 OrfB	(1526)	~~KEGILASFGAGGLEPLATVRKGIDKIQQA~~ LESGPY ~~AVNLIHSPFDSN~~ LEQG		
		1751		1800
Sch. sp. 9695 Pfa2p	(1572)	-VDALLTE ~~GVSI~~ VEASGYTGALSAD ~~IVRYRVTGLRRTSCGASVS~~ ATHRVV		
Sch. sp. 20888 OrfB	(1673)	~~NVDLFLEKGVTFVEASAFM~~ -TLTPQV ~~VRYRAAGLTENAD~~ -GSVNIRNRIT		
Thr. sp. 10212 Pfa2p	(1579)	~~NVDLFLEKYNVRFVEVS~~ AFM-SLTPQV ~~VRYRAAGLA~~ KARD-GSVKIQNRIT		
Thr. aureum ORF2	(1659)	~~NVLEFLRRGIRLVEASAFM~~ -SVTPSL ~~VRYRVAGLERGPG~~ -GTARVLRVI		
Thr. sp. 20892 OrfB	(1576)	~~NVDLFLEKNVRYAEC~~ SAFT-TLTPV ~~VVHYRAAGLVRRD~~ -GSLIKNRIT		
		1801		1850
Sch. sp. 9695 Pfa2p	(1621)	AKVSRTEVAEHFL ~~RPAPAAVLEALVAAKQITPEQAALAS~~ EVAMADDVAVE		
Sch. sp. 20888 OrfB	(1721)	GKVSRT ~~ELAEMFMPPAPEHLLQKL~~ IASGEINQ ~~EQAEIARVFPV~~ ADDIAVE		
Thr. sp. 10212 Pfa2p	(1627)	AKESRT ~~ELAEFLPAPKNILDALVADGSISQEQ~~ AQLALLVPMADDITVE		
Thr. aureum ORF2	(1707)	GKVSRA ~~ELAEMFMPPPAATVSKLLAQGLVTEE~~ QASLAEIVPLVDDVAIE		
Thr. sp. 20892 OrfB	(1624)	AKVSRTE ~~LAEMFLRPAPQIILEKLVA~~ EIISSDOARMAAKVPMADDIAVE		
		1851		1900
Sch. sp. 9695 Pfa2p	(1671)	ADSGGHTDNRPIH ~~VILPLIV~~ AQRNR---WRHLVDTPVRV ~~GAGGGIACPR~~		
Sch. sp. 20888 OrfB	(1771)	ADSGGHTDNRPIH ~~VILPLIINLRDPIHRECGYFANLR~~ VRVGAGGGIGCPQ		
Thr. sp. 10212 Pfa2p	(1677)	ADSGGHTDNRPIH ~~VILPLIIQORN~~ --RICKQY ~~PKHLKVRIGAAGGIGCPK~~		
Thr. aureum ORF2	(1757)	ADSGGHTDNRPIH ~~VILPVVLAIRDPMRECKYF~~ AANRVRVGAGGGIGCPA		
Thr. sp. 20892 OrfB	(1674)	ADSGGHTDNRPIH ~~VILPLIIQLRNTILAEYGCATAF~~ TRIGAGGGIGCPS		
		1901		1950
Sch. sp. 9695 Pfa2p	(1717)	AALAFS ~~SGAAFVVTG~~ SVNQLAPEAGTSDAVRL ~~ELATATYS~~ SDVAMAPGG-		
Sch. sp. 20888 OrfB	(1821)	AALATF ~~NMGASFIVTGT~~ VNOVAKOSGTC ~~DNVRKOLAKATYS~~ SDVCMAPAAD		
Thr. sp. 10212 Pfa2p	(1725)	AAFAF ~~FEMGAAYIATGT~~ VNOLSKEAGTCDYVR ~~KVLNKATYS~~ SDVTMAPAAD		
Thr. aureum ORF2	(1807)	AARAF ~~DMGAAFVLTGS~~ INOLTRQAGTSDSVRAAL ~~ARATYS~~ SDVTMAPAAD		
Thr. sp. 20892 OrfB	(1724)	AALAF ~~DMGASFVVTGS~~ INQICREAGTCDTVREILAN ~~SSYS~~ SDVTMAPAAD		
		1951		2000
Sch. sp. 9695 Pfa2p	(1766)	-----YQVLKKQTMFA ~~ARATMLAQLQAK~~ GSF ~~DAVPEPOLK~~ FLERSVF		
Sch. sp. 20888 OrfB	(1871)	~~MTTEGVKLQVLKKG~~ TMFPSRANKLYEL ~~FCKYDSFESMPPAE~~ LARVEKRIE		
Thr. sp. 10212 Pfa2p	(1775)	~~MFHDHGV~~ ELQVLKKG ~~TMFPSRAKKLYDLF~~ KKYKSIEELPADEV ~~VKKLEQ~~ KVF		
Thr. aureum ORF2	(1857)	~~MFDOGVKLQVLKRG~~ TMFPARANKLYEL ~~FTTYQSLDAIPRAELAR~~ LEKPVF		
Thr. sp. 20892 OrfB	(1774)	~~MFDOGVKLQVLKRG~~ TMFPSRANKLRKLFV ~~NYESLETIPS~~ KELKYLENIIF		

FIG. 5 (cont'd)

		2001		2050
Sch. sp. 9695 Pfa2p	(1809)	KQSVADVWAAAREKEFGVDATAAS-----PQERMALCVRWYMSQSSRW		
Sch. sp. 20888 OrfB	(1921)	SRALFEVWDETKNFYINRLHNPEKIQPAERDPEKLKMSLCERWYLSLASRW		
Thr. sp. 10212 Pfa2p	(1825)	KKSFDEVWDETKNFYINRLHSPEKIERAERDAKLKMSLCERWYLSKSSRW		
Thr. aureum ORF2	(1907)	RMSIDEVWNETKQFYETPLNNPAKVARAERDPEKLKMSLCERWYLSKSSRW		
Thr. sp. 20892 OrfB	(1824)	KQAVDQVWEETKRFYCEKLNPPDKIARAMKDEPKLKMSLCERWYLSKSSGW		
		2051		2100
Sch. sp. 9695 Pfa2p	(1851)	ATEATSARKADYQIWC GPAIGSFNDFVRGTKLDATEAGTGEFPRVVDINQH		
Sch. sp. 20888 OrfB	(1971)	ANTGASDRVNDYQVWCGPAIGSFNDFING-TYLDPAVANEXPCVVQINKQ		
Thr. sp. 10212 Pfa2p	(1875)	ANTGESGRVQDYQIWC GPAIGSFNDFAKGSPCLDPEILGSEPSVVQINKH		
Thr. aureum ORF2	(1957)	ASTGQVGRELDYQVWCGPTIGAENEFTYKSSSLDAEACGGREFPCVVRVQNE		
Thr. sp. 20892 OrfB	(1874)	ANAGIKSRAIDYQIWC GPAIGSFNNFASG-TSLDWKVTGVEPGVAENVMA		
		2101		2140
Sch. sp. 9695 Pfa2p	(1901)	ILLGASHYRPFVQVQQDDDDVEYIIV-----		
Sch. sp. 20888 OrfB	(2020)	ILRGACFLRPLEILRNARLSGAAALVASIDDTYVPAEKL		
Thr. sp. 10212 Pfa2p	(1925)	ILRGACFYQPLSOLKYLNFNYEELDTLTYSASNFI----		
Thr. aureum ORF2	(2007)	ILCGAAEQLARFMLLAGRESADALAYTVAEAR-----		
Thr. sp. 20892 OrfB	(1923)	ILDGARELAAKR-----		

FIG. 6

		1	50
Sch. sp. 9695 Pfa3p	(1)	---MTSSKKTPVWEMSKKEELLDGKTVVFDYNELLEFAEGDNGOVFGPEFD	
Sch. sp. 20888 OrfC	(1)	MAEPVKINKKPCWEMSKKEELTSGKTEVFNYEELLEFAEGDIKAVFGPEFA	
Thr. sp. 20892 OrfC	(1)	MAEPVASGKVPWEMSKSELCDRTTVVFDYEELLEFAEGDTSKVFGPEFK	
Thr. sp. 10212 Pfa3p	(1)	-MAGLQMKKKPVWEMSKKEEQSSGKNVVFDDYELLEFAEGDIGKVFGPKFD	
		51	100
Sch. sp. 9695 Pfa3p	(48)	LDKYRRRVRLPAREYLLVSRVTLMDAEVNNFRVGSRMVTEYDVPVNGEL	
Sch. sp. 20888 OrfC	(51)	LDKYPRRVRLPAREYLLVTRVTLMDAEVNNFRVGARMVTEYDVPVNGEL	
Thr. sp. 20892 OrfC	(51)	LDGKRRRVRLPAREYLLVTRVTLMDAEVGNFRVGARMVTEYDVPVNGEL	
Thr. sp. 10212 Pfa3p	(50)	LDKYSRRRVRLPAREYLLVTRVTLMDAEVGNFRVGSRMVTEYDVPVNGEL	
		101	150
Sch. sp. 9695 Pfa3p	(98)	SEGGDVPWAVLVESGQCDDLILISYMGIDFOCKGDRVYRLNNTLTTFEGVA	
Sch. sp. 20888 OrfC	(101)	SEGGDVPWAVLVESGQCDDLILISYMGIDFONOGDRVYRLNNTLTTFEGVA	
Thr. sp. 20892 OrfC	(101)	SEGGDVPWAVLVESGQCDDLILISYMGIDFOCKGDRVYRLNNTLTTFEGVA	
Thr. sp. 10212 Pfa3p	(100)	SQGGDVPWAVLVESGQCDDLILISYMGIDFOCKGDRVYRLNNTLTTFEGVA	
		151	200
Sch. sp. 9695 Pfa3p	(148)	HEGETLYDIRVTGFAKGAGGEISMFFFEYDCYVDGRLLIEMRDGCAGFF	
Sch. sp. 20888 OrfC	(151)	HEGETLEYDIRVTGXAKFLDGGISMFFFEYDCYVNGHLLIEMRDGCAGFF	
Thr. sp. 20892 OrfC	(151)	KEGETLYDIRVTGFAKLPDGNISMFFFEYDCYVNGHLLIEMRDGSAGFF	
Thr. sp. 10212 Pfa3p	(150)	HEGETLYDIRVTGFAKGNHGEISMFFFEYDCYVNGHLLIEMRDGCAGFF	
		201	250
Sch. sp. 9695 Pfa3p	(198)	TDAELAAGKGVLTVAELAARAQIQKQDIAPFAPAPCSHKTSLEDAEMRL	
Sch. sp. 20888 OrfC	(201)	TNEELDAGKGVVFTFGDLAARAKIPKQDVSPYAVAPCLHKTCLNEFEMOT	
Thr. sp. 20892 OrfC	(201)	TDEELAAGKGVVTRAQONMRDKIVROSIEPFALAACHTKTLNESIMOS	
Thr. sp. 10212 Pfa3p	(200)	TDEELAAGKGVLTVAELHKRSIVPKSIKPFALNPAVHKTMFSENDMEK	
		251	300
Sch. sp. 9695 Pfa3p	(248)	LYDPQWAVRVEG--SGMAGIDYKLCARKMLMIDRVTHLDPPGGANGLGLLV	
Sch. sp. 20888 OrfC	(251)	LYDPQDWASVEGSKNGMPEINNYKLCARKMLMIDRVTSIDHGGVYGLGQLV	
Thr. sp. 20892 OrfC	(251)	LYDPQWAVRVEG--SNKMAEINNYKLCARKMLMIDRVTHIDHGGAYGLGLLV	
Thr. sp. 10212 Pfa3p	(250)	LCSPQWENVLGG--SGMAGIDYKLCARKMLMIDRVTKIQHNGGAYGLGLLV	
		301	350
Sch. sp. 9695 Pfa3p	(296)	GEKILERDHWYFPCHFVNDQVMAGSLVSDGCSOLLKLYMWLGLHHTTGA	
Sch. sp. 20888 OrfC	(301)	GEKILERDHWYFPCHFVKDQVMAGSLVSDGCSOLLKLYMWLGLHHTTGP	
Thr. sp. 20892 OrfC	(301)	GEKILDRNHWFYFPCHFVNDQVMAGSLVSDGCSOLLKLYMWLGLHHTKNE	
Thr. sp. 10212 Pfa3p	(298)	GEKILERDHWYFPCHFVKDQVMAGSLVSDGCSOLLKLYMWLGLHDTVPP	
		351	400
Sch. sp. 9695 Pfa3p	(346)	FDLFPVSGHANKVRCRGQISPHKGKLVYVMEIKEMGFDAKTGDPFAIADV	
Sch. sp. 20888 OrfC	(351)	FDLFPVNGHANKVRCRGQISPHKGKLVYVMEIKEMGFEDDN-DPYAIADV	
Thr. sp. 20892 OrfC	(351)	FDLFPVSGHKNKVRRCRGQISPHKGKLVYVMEIKKMGYDOAGSPYAIADV	
Thr. sp. 10212 Pfa3p	(348)	FQFLFPVPGQPNKVRRCRGQISPHKGKLVYVMEIREMGFNESTGOPYAIADV	
		401	450
Sch. sp. 9695 Pfa3p	(396)	DIIDVNFEEGQAFAGVEDHSYGGDLRKKIVVDFKGIALSLQPKKEQQK	
Sch. sp. 20888 OrfC	(400)	NIIDVDFEKGQDFS-LDRISDYGGDLNKKIVVDFKGIALKNQKSTNNH	
Thr. sp. 20892 OrfC	(401)	DIIDVNEELGOSFD-INDLSYGGDLKKIVVDFKGIALOLKGFASFAM	
Thr. sp. 10212 Pfa3p	(398)	DIIDVNEELGOSFD-MADISYGGGNLSKKIVVDFKGIALONEGTVKSSN	
		451	500
Sch. sp. 9695 Pfa3p	(446)	ES-----MTTITTTTITTSRYIA	
Sch. sp. 20888 OrfC	(449)	PSKVQPVFANGAATVGPEASKASSGASASASAAAPKPAFSADVLAPNPYA	
Thr. sp. 20892 OrfC	(450)	SS-----SSSLNEGWCQVPKPSQRMHEHQ	
Thr. sp. 10212 Pfa3p	(447)	-----ITDSSP--KSTTIQ	
		501	550
Sch. sp. 9695 Pfa3p	(463)	PPSGCLYGDPTAPESVTWHPMAEGNGGPGPTPSFSPSAYEPRAVCFSPFP	
Sch. sp. 20888 OrfC	(499)	LPEHILKGDALAPKENSWHPMARLGG--NPTPSFAPSAYKPRNIAFXFPF	
Thr. sp. 20892 OrfC	(474)	PPHCLASDEAPSVTWHPMSKILGG--NPTPFFSPSSYPRAICFIPFP	
Thr. sp. 10212 Pfa3p	(459)	PPPNCLSGDPLAPSOVTWHPMAGYNG--APAPSFSPSDYEPRAVCFKFPF	
		551	600
Sch. sp. 9695 Pfa3p	(513)	NNPLDNDHTPGKMPLTWENMSEFMCGKVSNCGLPEFARFDASKTSRSPAF	
Sch. sp. 20888 OrfC	(547)	GNNPDNDHTPGKMPLTWENMAEFMAGKVSNCGLPEFARFDDSNITSRSPAF	
Thr. sp. 20892 OrfC	(522)	GNNPLDNNCKAGEMPLNWNMSEFMCGKVSNCGLPEFARFDKSNITSRSPAF	
Thr. sp. 10212 Pfa3p	(507)	GNNPLDNDHTPGKMPLTWENMSEFMCGKVSNCGLPEFKRFDNSKTSRSPAF	

FIG. 6 (cont'd)

	601		650
Sch. sp. 9695 Pfa3p	(563)	DLALVTRVTSVADMEHGPFYNDVNPGQGTMTVGEFDCPADAWFFGASSRD	
Sch. sp. 20888 OrfC	(597)	DLALVTRAVSVSDLKHVNYRNIIDLPSSKGTMTVGEFDCPADAWFFKGACND	
Thr. sp. 20892 OrfC	(572)	DLALVTRVVEVTNMEHGKFLNVDNPNPSKGTMTVGEFDCPADAWFFDGSCND	
Thr. sp. 10212 Pfa3p	(557)	DLALVTRVSVSDMEFKPHLNIDVNPNPSKGTMTGEFDCPADAWFFGQSCND	
	651		700
Sch. sp. 9695 Pfa3p	(613)	DHMPYSILMEIALOTSGVLTSVLKAPLTMKDDILFRNLDAEIVGDAM	
Sch. sp. 20888 OrfC	(647)	DHMPYSILMEIALOTSGVLTSVLKAPLTMKDDILFRNLDAEIVFV-AD	
Thr. sp. 20892 OrfC	(622)	DHMPYSILMEISLOTSGVLTSVLKAPLTMKDDILFRNLDAEIVFV-PD	
Thr. sp. 10212 Pfa3p	(607)	DHMPYSILMEIALOTSGVLTSVLKAPLTMKDDILFRNLDAEIVFV-SD	
	701		750
Sch. sp. 9695 Pfa3p	(663)	PDYRGKTIINFTKCTGYSMLGKMGHHRFTFELSVDGAVFYKGSTSFGWFT	
Sch. sp. 20888 OrfC	(696)	PDYRGKTIINFTKCTGYSMLGEMGVHRFTFELYVDDVLFYKGSTSFGWFT	
Thr. sp. 20892 OrfC	(671)	PDYRGKTIINFTKCTGYSMLGKMGHHRFTFELSVDGVVIFYKGSTSFGWFT	
Thr. sp. 10212 Pfa3p	(656)	PDYRGKTIINFTQCTGYSMLGKMGHHRFTFELSVDGVVIFYKGSTSFGWFT	
	751		800
Sch. sp. 9695 Pfa3p	(713)	PEVFESQTLGDNKGPRLPWYREI-----NVAVDTLSAPASASSAQ	
Sch. sp. 20888 OrfC	(746)	PEVFASQAGLDNGKSEPWEIFENKVPASOVSSFDVRPNNGSGRTAIFINAP	
Thr. sp. 20892 OrfC	(721)	PEVFAQQAGLDNGKITEPWCKTNN--TSVRRVEIASAKGKEQLTEKLPA	
Thr. sp. 10212 Pfa3p	(706)	PEVFESQVGLDNGKIKVQPYLQKS--SHVVTYDVASTAGKDKLFSKIGS	
	801		850
Sch. sp. 9695 Pfa3p	(753)	GQLOLQRRGSQAQFLDTIHLGSGAGVHGQGYAHGKAVNKQDWFFSCHF	
Sch. sp. 20888 OrfC	(796)	SGAQLNRRIDQGYLDADIVSG-SGKSLGYAHGSKTVNPNWFFSCHF	
Thr. sp. 20892 OrfC	(769)	TNAQVLRPSEQCEYLDYINLAPD-SGLHGKGYAHGKDVNPQDWFFSCHF	
Thr. sp. 10212 Pfa3p	(754)	KDAQVQRRNTOCEFLDTMHLIPN-TGKYNKGYAHGKKNPNWFFSCHF	
	851		900
Sch. sp. 9695 Pfa3p	(803)	WFDEVPMPGSLGIESMFQLEAFCVKQGLAARHGIAHPVFAHAP-GATSWK	
Sch. sp. 20888 OrfC	(845)	WFDSVMPGSLGVESMFQLEAIAHEDLAGKARHCOPHLCARPRASSWK	
Thr. sp. 20892 OrfC	(818)	WFDEVPMPGSLGIESMFQLEAFVNDQNPGEYNVSNPTFAHAP-GKTAWK	
Thr. sp. 10212 Pfa3p	(803)	WFDEVPMPGSLGIESMFQLEAFSIDQGIASKHGIVNPTFAHNS-GKTSWK	
	901		950
Sch. sp. 9695 Pfa3p	(852)	YRGQLTPKNDKMDSEVHIKSVAAFS--VVDVADGFLVDGLRVYSAD	
Sch. sp. 20888 OrfC	(895)	YRGQLTPKSKMDSEVHIKSVDAHDG---VVDLADGFLVADSLRVYSVS	
Thr. sp. 20892 OrfC	(867)	YRGQLTPKIRAMDCEVHIKSVITASPENGGVVDIVADGALVVDGLRVYEAK	
Thr. sp. 10212 Pfa3p	(852)	YRGQLNNKGRMDSEIHIKDIVKNADG--TVDLADGFLVDSLRVYSAD	
	951		1000
Sch. sp. 9695 Pfa3p	(899)	NLRVRIQAGSHVEEQEVAAKATTN-----SSIADVDVADLQAL	
Sch. sp. 20888 OrfC	(942)	NLRVRIASGEAPAAASSAASVGSSASSVERTRSSPAVASGPAQTIDIKQL	
Thr. sp. 20892 OrfC	(917)	EERVPVYSAPQAIQPDVQQQPPSAKADPG-----KTGVALSPTQL	
Thr. sp. 10212 Pfa3p	(900)	DLRVNIVPGTKAAPKSVAAAPRHVATPIPGV-----PSNTSSVEISLES	
	1001		1050
Sch. sp. 9695 Pfa3p	(939)	KQALLTLERPLQLDAG-----SEVPACAVSDLGDRGFM	
Sch. sp. 20888 OrfC	(992)	KTELLELDAPLYLSQDPTSGQLKKHTDVASGOATVQPCTLGDLGDRSFM	
Thr. sp. 20892 OrfC	(957)	KDVLLEVDPNPLYLGVDNINLVQFESKPATSSRIVSIKPCSIDLGDKSFM	
Thr. sp. 10212 Pfa3p	(945)	KKELLNLEKPLYLETSHI--VKQFGDVNNGQASVIPPCTINDLGERSFM	
	1051		1100
Sch. sp. 9695 Pfa3p	(972)	ETYGVVAPLYSGAMAKGSIADLVIAAGRKILGSFGAGGLPMHIVRAAI	
Sch. sp. 20888 OrfC	(1042)	ETYGVVAPLYTGAMAKGSIADLVIAAGRKILGSFGAGGLPMHIVRAAL	
Thr. sp. 20892 OrfC	(1007)	ETYNVSAPLYTGAMAKGSIADLVIAAGRKILGSFGAGGLPISIVREAL	
Thr. sp. 10212 Pfa3p	(993)	ETYNVAPLYTGAMAKGSIADLVIAAGRKILGSFGAGGLPMHIVRASV	
	1101		1150
Sch. sp. 9695 Pfa3p	(1022)	EKIQAALPAGPYAVNLIHSPFDANLEKGNVDLFLEKGVVVVEASAFMELT	
Sch. sp. 20888 OrfC	(1092)	EKIQAALPOGPYAVNLIHSPFDSNLEKGNVDLFLEKGVTVVVEASAFMTLT	
Thr. sp. 20892 OrfC	(1057)	EKIQOHLPHGPYAVNLIHSPFDSNLEKGNVDLFLEMGVTVVVECSAFMELT	
Thr. sp. 10212 Pfa3p	(1043)	EKIQAALPEGPYAVNLIHSPFDSNLEKGNVDLFLEKGVHVVEASAFALT	
	1151		1200
Sch. sp. 9695 Pfa3p	(1072)	QOVVRYRAGLSRDARGGSVRTAHKIIIGKVSRTELAEMFIRPAPQAILDK	
Sch. sp. 20888 OrfC	(1142)	QOVVRYRAAGLSRNADG-SVNIRNRIIGKVSRTELAEMFIRPAPHEILDK	
Thr. sp. 20892 OrfC	(1107)	AQVVRYRAGLSKSDAG-SIRIAHRIIGKVSRTELAEMFIRPAPQHLLQK	
Thr. sp. 10212 Pfa3p	(1093)	QOVVRYRACGLSPAKDG-SVLIKNRIIGKVSRTELAEMFFRPAPONLLDK	

FIG. 6 (cont'd)

		1201		1250
Sch. sp. 9695 Pfa3p	(1122)	LVASGETTPEQAALALEVPMADDIAVEADSGGHTDNRPIHVILPLIISLR		
Sch. sp. 20888 OrfC	(1191)	LVASGETTQEQAEELARRVPVADDIAVEADSGGHTDNRPIHVILPLIINLR		
Thr. sp. 20892 OrfC	(1156)	LVASGETTAEQAELATQVPVADDIAVEADSGGHTDNRPIHVILPLIINLR		
Thr. sp. 10212 Pfa3p	(1142)	LVASGETTKEQASLAEVPMADDIAVEADSGGHTDNRPIHVILPLIINLR		
		1251		1300
Sch. sp. 9695 Pfa3p	(1172)	NRLQREKYPARHRVRVGAGGGIGCPQAAALGAFHMGAAFFVTGTVNQLSP		
Sch. sp. 20888 OrfC	(1241)	NRLHRECGYPALHRVRVGAGGGIGCPQAAAAALTMGAAFFVTGTVNQVAK		
Thr. sp. 20892 OrfC	(1206)	NRLHKEEDYPSHLRVRVGAGGGIGCPQAAALAFOMGAAFLITGTVNQLAP		
Thr. sp. 10212 Pfa3p	(1192)	NRLHKECGEPALHRVRVGAGGGIGCPSAAYAFNMGAAFLITGSVNQVSK		
		1301		1350
Sch. sp. 9695 Pfa3p	(1222)	QAGTCDIVRRQLSRATYSDIITMAPAADMFEGVLEQVLKKGTMFPPSRAKK		
Sch. sp. 20888 OrfC	(1291)	QSGTCDIVRKQLSQATYSDICMAPAADMFEEGVKLQVLKKGTMFPPSRANK		
Thr. sp. 20892 OrfC	(1256)	ESGTCDIVRLQLSKATYSDFCMAPAADMFQGVLEQVLKKGTMFPPSRAKK		
Thr. sp. 10212 Pfa3p	(1242)	QSGTCDIVRKQLSEASYSDIITMAPAADMFQGVLEQVLKKGTMFPPSRAKK		
		1351		1400
Sch. sp. 9695 Pfa3p	(1272)	LYELFHKYDSFEAMPADLARVEKRIFSKSLAEVWAETKDFYITRLNPE		
Sch. sp. 20888 OrfC	(1341)	LYELFCKYDSFDSMPPAELERIEKRIFKRALQEVWEETKDFYINGIKNPE		
Thr. sp. 20892 OrfC	(1306)	LYELFCKYDSFEAMPAEELORVEKRIFOKSLAEVWQETSDFYIHRKKNPE		
Thr. sp. 10212 Pfa3p	(1292)	LYELFCMYNSFDMPKSELORIEKRIFOKSLAEVWEETKDFYINRLNPE		
		1401		1450
Sch. sp. 9695 Pfa3p	(1322)	KIRKAENEDPKLKMSLCFRWYLGLSSEFWANNGIADRTMDYQIWC GPAIGA		
Sch. sp. 20888 OrfC	(1391)	-KIQRAEHDPKLKMSLCFRWYLGASRWANMGAPDRVMDYQVWC GPAIGA		
Thr. sp. 20892 OrfC	(1356)	-KINRAASDGKLKMSLCFRWYLGLSSEFWANSQAQDRVMDYQIWC GPAIGA		
Thr. sp. 10212 Pfa3p	(1342)	KIEHASKKDPKLKMSLCFRWYLGLSSEFWANNGIKERSMDYQIWC GPAIGS		
		1451		1500
Sch. sp. 9695 Pfa3p	(1372)	FNDFIADSYLDNAVSGETPDVVQINLQILSGAAYLQRLLSVKLA--PRID		
Sch. sp. 20888 OrfC	(1440)	FNDFIKGTYLDPAVSNEYPCVVQINLQILRGACYLRRINALND--PRID		
Thr. sp. 20892 OrfC	(1405)	FNDFITGTYLDTVAKSYPCVAQINLQILOGAAYLKRLGVIRFDRMLLQA		
Thr. sp. 10212 Pfa3p	(1392)	XNDFVKGTYLDPAVAGSYPCVVQINNQILRGACFLQVRRAIKHH--PRID		
		1501		1517
Sch. sp. 9695 Pfa3p	(1420)	VDTEDDFTYRDPDHAL-		
Sch. sp. 20888 OrfC	(1488)	LETEDAAFVYEPTNAL-		
Thr. sp. 20892 OrfC	(1455)	VDIDDPVFTYVPTQPL-		
Thr. sp. 10212 Pfa3p	(1440)	IDVDEDVFTYRPESTL-		

19/38

[illegible]

FIG. 8

MDTRIAVMSAILPSGENVRESWEAIRDGLDCLSDLPADRVDTVAYYNPEKTTKDIYCKRGFIPEYDFDAREFGLNMFQMEDSDANQTISLLKV
 KEALTDANIPAFSSGKKNIGCVLGIGGQKASHEFYSRNLNVVVDKVLKMGLEEDVAAAVDKYKASFPEWRLDSFPGFLGNVTAGRCNFTNM
 EGMNCVVDAAACASSLIAVKVAIEELLYGDCDAMIAGATCTDNSIGMYMAFSKTPVFSTDPVSKAYDAATKGMFIGESAMLVLKRYADAVRDGD
 TVHAVIKGCASSDGGKAAAGIYPTISGQEEALRRAYARANVDPATVTLVEGHGTGTPVGDKIELTALSNLFSKAFSANGGAEAEQVAVGSIKSQI
 GHLKAVAGLAGLVKVVLAALKHKTLPTQNTINVDKPPSLVDGTPIQQSPLYVNTMNRPWFTPVGVPRRAGVSSFGFGGANHYHAVLEEFEPHESESAYRY
 NNLPQVALLHAGDVATLAAATVRAKLALATAEQEEARVVKNADYIAYHRFLDECKLRGAVPQAHARVGLLVRDLSSLIAVLEAAAKLAGESAT
 EWTVSVATGEAAFRVRGVATEANVAALFSGGQAQYTHMFSDVAMNWPFRFRESVAMMDRAQRRERFGRPAKRVSSVLYPRKPYGDEPRQDHKEIS
 QTRYSPATLACSVGAFDIFKAAGLAPSFAGHSLGEFAALYAAAGSLDRDAVFDLVCARAKAMSDFTAQASSSGGAMAAVIGAKADQLSLGGAPD
 VWLANSNPSQTVITGTAEAVAAASDKLRCSGNFRVPLACEAAFHSPHMRGAEQTFASALAQAPVSAPAAARFYSNVTGGAAVTSADVKTNLG
 KHMTSPVQVQVQVRAMHAAAGARVFVEFGPKQVLSRLVKETLGEAGDVTVAVNPDSAKDSDTQLRQAALTLAVAGVPLKDFDRWQLPDATRLE
 PVKKKKTTLRLSAATYVSAKTLRQREAVLNDGYTVSGATAVVKEDTANEERLVRQAQDLRQLAEASTAAQAAQSKVAELERTIQDLERKVQQ
 QQKEKGENSDSNAAAEVLRHKELLQMLQDCDEQAVPVATVPTPTSSPTTSSPVSGNSKSTRGSADLQALLAKAETVVMAVLAAKTGYEAD
 MVEADMIDLEAELGIDSIRVEILSEVQQQLGVEAKDLDALSRTRTVGEVVDAMKAEIVAASGGSAPAVPSAPAAASAAPTAASTAPSADLQALLSK
 AETVVMAVLAAKTGYEADMVEADMIDLEAELGIDSIRVEILSEVQQQLGVEAKDLDALSRTRTVGEVVDAMKAEIVAASGGSAPAVPSAPAA
 SAAPTAASTAPSADLQALLSKAETVVMAVLAAKTGYEADMVEADMIDLEAELGIDSIRVEILSEVQQQLGVEAKDLDALSRTRTVGEVVDAMK
 AEIVAASGGSAPAVPSAPAAASAAPTAAATAPSADLQALLAKAETVVMAVLAAKTGYEADMVEADMIDLEAELGIDSIRVEILSEVQQQLGVE
 AKDLDALSRTRTVGEVVDAMKAEIVAASGGSAPAVPSAPAAASAAPTAASTAPSADLQALLSKAETVVMAVLAAKTGYEADMVEADMIDLEA
 ELGIDSIRVEILSEVQQQLGVEAKDLDALSRTRTVGEVVDAMKAEIVAASGGSAPAAASAAPTAASTAPSADLQALLSKAETVVMAVL
 AKTGYEADMVEADMIDLEAELGIDSIRVEILSEVQQQLGVEAKDLDALSRTRTVGEVVDAMKAEIVAASGGSAPAAAPSAPALLPTLFGSECEDSL
 TFPVITTLPLPAELVLAEGGARPVVVDDGSALTSSLVSSLDRAVLLVQVSSSACSPRSTTHKLVTVADRSEALQAALTSVEAQFGKVGGFVQF
 GDDDVQAQLGWALLAKHLKTSLSQIEGGRTFFVAVARLDGQLGSLGKSTTATVDLSRAQQGSVFGCLKTLDLEWPAVFCRGIDLAAADLDAQ
 AARCLLGELSDPDVAVRESGYASGQRCTTTTKSLTTGKPHQPISSDDLFLVSGGARGITPLCVRELAQRVGGGTYYVIGRSELPTTEPAWAVGVESG
 KPLEKAALAFLEKAEFAAGRGAKPTPMLHKKLVGAVVGAREVRAASLAEITAQGATAVYESCDVSSAAKVRREVERVQQQGGRRRVSGVFHASGVL
 RDKLVENKSLADFSVYDTKVGGLINLLACVDLAQLRHLVLFSSLAGFHGNVQSDYAMANEALNKLAAHLSAVHPQLCARSCICFGPWDDGGMVT
 PALKANFIRMGIQIPRQGGAQTVANMLVSSSPGQLLVGNWGVPPVPSATEHTVLQTLRQSDNPFLDSHVIGRRVLPMTLAVGYMAHQASIIYA
 GHQLWAVEDAQLFKGIAIDNGADVPVRVELSRRKEEQEDAGKVKVQVLLKQVNGKSVPAKYKATVVLSPAPRPSVITRDFDLTPDPACTEHD
 YDGKTLFHGKAFQIEQVLSATPKQLTAKCRNLPLTPEQRGQFVVNLSSQDPFQADIAFQAMLVWARMLRQSAALPNNCERFDYKPMAPGATY
 YTSVKLASASPLVDSVCKCTVAMHDEQGEVYFSARASVVLNKTLLTY

(SEQ ID NO:2)

FIG. 10

MPCDNIAVVGMAVQYAGCKNQDEFWDTLMRKEINSSPISAERLGTRYRDLHFHPQRSKYADTFCNDRYGCVDASVDNEHDLADLARRALLDAG
 INLDDASTTANLRDFGIVSGCLSPFMDNLQGELLNLYQVHVNRVGAQFRDRSPWSEPRPRAVSPEASDPRVYSDPASFVANQLGLGPVRYSLDAA
 CASALYCLKLASDHLLSRSAVMLCGATCFDPFFILSGFSTFQAMPGLGGPDDNPLSVPLRQGSQGLTPGEGGAIMVLKRLEDAVRDGDRIYGTLLG
 TSLSNAGCGLPLSPHLPSEKSCMEDLYTSVGIDPSEVQYVECHATGTPQGDVVEVEALRHCFRGNTHDPPRMGSKGNFGHTLVAAAGFAGMAKVL
 LSMQHGTTPPTPGVDRSNCIDPLVVEAIPWPYSSAQAQAGKPGDELKCSLSAFGFGGTNAHCVFREHRQIAATATASPVLPVPTPGPIAIGMDAT
 FGTCLKGLDAFEQAIIYKGTGDGASDLPSKRWRFGLGADTDFTLTMGLDAVPRGCYVRDVEDYKRLRSPMIPEDVLRPQQLA VATMDRALQDAGM
 ATGGKVAVLVGLGTDTELYRHRARVTLKERLDPAAFSPEQVQEMMDYINDCGTSTSYTSYIGNLVATRVSQWGTGTPSFTVTEGANSVYRCLEL
 GKFLDTHQVDVAVVAGVDLCATAENLYLKARRSAISRQDHPRANFEASADGYFAGEGSGALVLKRQADVGSDDKVYASVAGLTCAAQPAEAV
 SPLLLQVHNDNEKRVVEMVELAADSGRHAPHLANSPLSAESQLEQVSKLLAHQVPGSVAIGSVRANVGDVGYASGAASLIKALCLHNRYL PAN
 PPOWERPVAPVSEALFTCPRSRAWLKNPGESRLAAVASASESGSCFGVLLTDEYATHESNRSLDADAAPKLIAIRGDTVDDIMAKVNAELALLRAH
 AETGSATDDDPAAA V AFTAHRLRFLRLVGETVASHGATATLCLALLTTPEKLEKELELA AKGVPRSAKAGRNWMSPSGSAFAPTPVTSDRVAFMY
 GEGRSPYYGVGLDLHRLWPALHERINDKTAALWENGDSWLMPRADVADSQRAVQTAFDADQIEMFRTGIFVSICLTDYARDVLGVQPKACFGLSL
 GEISMLFALSRRNCGLSQDLTQRLRTSPVWSTQLAVEFQALRKLWNPADAPVESFWQGYLVASRAEIEKAIGPDNRFVRLLVNDSSSALIAGKP
 AECLRVLERLGGRLPPMPVKQGMIGHCEVAPYTPGIAHIEILEIPDSPVKMYTSVTNAELRGGSSSITEFVQKLYTRIADFPGIVDKVSRDGHDF
 VEVGPNNMRSAAVSDILGKAATPHVSVALDRPSESAWTQTLKSLALLTAHRVPLHNPTLFAADLYHPTFLTADSAMQEPKPNRFLRSVEVNGYFC
 PDGISKQVAAASAKPSTHCMVRLHPAKAVVAAAGAVVADSTPVVKAKQTSLSLLVGDDAFLRCYDWDWPLYMGAMAEGISSVDLVVAAA EAR
 MLASFGAARLPMDQVELQIREIQRTSNFAFVNLMPGPDEAATVDALLRTGV SIVEASGYTGALSADLVRYRVTGLRRTSCGASVSATHRVVAKV
 SRTEVAEHFLRPAPAAVLEALVAAKQITPEQAALASRVAMADDVAVEADSGGHTDNRPHVLLPLVVAQRNRWRHLVDTPTVRVGAGGGIACPR
 ALLAFSLGA AFVVTGSVNQLAREAGTSDA VRLLLATATYSDVAMAPGGVQVLKKQTMFAARATMLAQLQAKFGSFDVPEPQLRKLERSVFKQS
 VADVWAAAREKFGVDATAAASPQERMALCVRWYMSQSSRWATEATSARKADYQIWC GPAIGSFNDFVRGTKLDATAGTGEFPRVVDINQHILLGA
 SHYRRVQQQQQDDDDVEYIV

(SEQ ID NO:4)

FIG. 12

MTSSKKTVPWEMSKEELLDGKTVVFDYNELLEFAEGDVGVGFPEFDIIDKYRRRVRLPAREYLLVSRVTLMDAEVNNFRVGSRMVTEYDVPVNG
 ELSEGGDVPWAVLVESGQCDDLMLISYMGIDFQCKGDRVYRLNLTSLTFFGVAHEGETLVYDIRVTGFAKGAGGEISMFFFEYDCFVDGRLLIEMRD
 GCAGFTDAELAAGKGVLTKAELAARAIQKQDIAPFAPPCSHKTSLDAREMRLLVDRQWARVFGSGMAGIDYKLCARKMLMIDRVTHLDP
 GGAHGLLLIGEKVLERDHWYFPCHFVRDEVMAAGSLVSDGCSQLLKVYMLWLGHTTVGAFDFRPVSGHANKVRCRGQISPHKGLVYVMEIKE
 MGFDAKTGDPFAIADVDIIDVNFEEGQAFAGVEDLHSYGQGDRLKKIVVDFKGIALSLQKRKEQQKESMTVTTTTTTTSRVIAPPSGCLKGDP
 SVTWHIPMAEENGPGPTPSFSPSAYPPRAVCFSPFPNNPLDNDHTPGQMPLTWFNMFSEFMCGKVSNCCLGPEFARFDASKTSRSPAFDLALVTRVTS
 VADMEHGPFFYNVDVNPQGQTMVGEFDCPADAWFFGASSRDDHMPYSILMEIALQTSGLTSVLKAPLTMDKDDILFRNLDAEELVGDAMPDVR
 GKTIRNFTKCTGYSMGLGKMGHRTFELSVDGAVFYKGSTSFGWFFVPEVFESQTLGDNKGKPRLPWYRENNVAVDTLSAPASASSAQGLQQRG
 SQAQFLDTIHLAGSGAGVHGQYAHGEKAVNKQDWWFFSCHFWFDPVMPGSLGIESMFQLVEAWCVKQGLAARHGIAHPVFAHAPGATSWKYRG
 QLTPKNDRMDSEVHIKSVAAFSSWVDVADGFLFVDGLRVYSADNLRVRIQTGAGHVEEQEVAAKATTKNSSIADVADLQALKQALLTLERPL
 QLDAGSEVPACAVSDLGDRGFMETYGVVAPLYSGAMAKGASADLVIAMGQRKMLGSFGAGGLPMHVVRAGIEKIQAAALPAGPYAVNLIHSPFD
 ANLEKGNVDLFLEKGVVVVEASAFMELTPQVVRYRATGLSRDARGSVRTAHKIIKVSRTELAEMFIRPAPQAILDKLVASGEITPEQAAALAEVP
 MADDIAVEADSGGHTDNRPIHVILPLILSLRNRLQRELKYPARHRVRVGAGGGIGCPQAAALGAFHMGAAFVVTGTNVNQLSRQAGTCDNVRRQLSR
 ATYSBITMAPAADMFEOQVELQVLKKGTMFPSRAKKLFLFHKYDSFEAMPADELARVEKRIFSLSAEVWAETKDFYITRLNNPEKIRKAENEDP
 KLKMSLCFRWYLGSSFWANNNGIADRTMDYQIWC GPAIGAFNDFIADSYLDVAVSGEFPDVVQINLQILSGAAYLQRLLSVKLAPRIDVDTEDDLF
 TYRPDHAL

(SEQ ID NO:6)

FIG. 13 (cont'd)

AATTGGATATGGAATTGGAACCCGAACTTGGTATTGATTCTATCAAGAGAGTAGAAAATTTCTTCTGAAGTTCAAGCTCAATTGAGTGTGAAGCTAAAGATGTAGATGCTCTTA
 GTAGAACTCGTACTGTTGGTGAAGTATTGATGCAATGAAAGCCGAAATTTCTGGTGTGAGCCAGCTGCTCTGTTCAAAGTTGCAAGCTCTACTCAAAATAGTTGCTCCTGTTC
 AAGCATCTGCTCTGTTGATTCTGGTTTGGCAAAAGCGGAACAAAGTTGATTGGAAGTTAGCATCCAAGACTGGTTATGAAACTGAGTTGATTGAAATAGATATGGAAT
 TGGAAACCGAACTTGGTATTGATTCTATCAAGAGAGTAGAAAATTTCTTCTGAAGTTCAAGCTCAATTGAGTGTGAAGCTAAAGATGTAGATGCTCTTAAGTAACTCGTACTG
 TTGGTGAAGTATTGATGCAATGAAAGCTGAAATTTCTGGTGGTCAACCAGCTGCTCTGTTCAAGTTGCAGCTCTACTCAAAATAGTTGCTCTCTCAAGTATCTACTCTGTTGATTCTGTTTGT
 TGATTCTGTTTGTAGCAAAAGCGGAACAAAGTTGATTGGAAGTATTGGCATCTAAGCATCTAAGACTGTTATGAGACTGATTGATTGAATTTGGATATGGAATTTGGAACCTGAACTTG
 GTATTGATTCTATCAAGAGAGTAGAAAATTTCTTCTGAAGTTCAAGCTCAATTGATTGTAAGCTAAAGATGTAGATGCTCTTAGTAGAACTGCTACTGTTGTTGTTGAAAGTATTG
 ATGCAATGAAAGCCGAAATTTGCTGGTGTCAACCAGCTGCTCTGTTCAAGTTGCAGCTCTGCTCCAGTAGTTGCTCTCAAGTATCTACTCTGTTGATTCTGTTTGT
 GGCAAAAGCGGAACAAAGTTGATTGGAAGTATTGAGTATGAAACTGAGTTGATTGAATTTGGATATGGAATTTGGAACCTGGAATTTGCAATTTGGAATTTGATTCTATCA
 AGAGAGTAGAAAATTTCTTCTGAAAGTTCAAGCTCAATTGAGTTGGAAGCTAAAGATGTAGATGCTCTTAGTAGAACTCGTACTGTTGGTGAAGTATTGATGCAATGAAAGCC
 GAAATTTCTGGTGTCAACCAACTGCTCTGTTCAAGTTGCAGCTCTACTCAAGTAGTTGCTCTGTTAAAGTATCTACTCTGTTGATTCTGTTGTTAGCAAAAGCGGAAC
 AAGTAGTATTGGAAGTATTGGCATCTAAGACTGGTTATGAACTGAGTTGATTGAATTAGATATGGAATTTGGAACCTGGAATTTGATTTGATTCTATCAAGAGAGTAGAAAAT
 CTTTCTGAAAGTTCAAGCTCAATTGAAATGTGGAAGCTAAAGATGTGGATGCTCTTAGTAGAACTCGTACTGTTGGTGAAGTATTGCAATTTGGAACCTGGAATTTGCTGGTGA
 TCAACTGCTCCAGCTGATTCTCAGTTCAAGCTAAGAGTGGTGTAGCCAACTGCACTTTTGGCAAGCGGAACAAAGTAGTATTGGAAGTATTGGAAGTATTGGAAGTATTGGAAGT
 ATGAAACTGAGCTGATTGAATTTGGATATGGAATTTGGAACCTGGAATTTGATTAATCAAGAGAGTAGAAAATTTCTGCGAAGTTCAGCAGAAATTTGAGTGTGTTGAAAGCA
 AAAGATGTAGACGCTCTAAGTAGAACCCGCTACTGTTGGGGAAGTAGCATGCAATGAAAGCTGAAATTTGCTGGCAGTGTCTACGGTTGCAACTTTTGGATGATTCAACAAT
 TATGGAGGAGACAGATGATGAAGATGAAGACTTTATTTATACGATCATGTATACGGAAGCGAATGTGAAGATCTTAGTCTGAGTTTTCATCCCTGAAGAGCATCCCGCGCG
 CTGATAAACTTTTGTGGATAACATTGCTGAAGGCCAAATTGTTATTGTGATTGTGGAACAAAGCTTACAACTGAACCTTGCAAAAGCTATTGGAGAACCTGCGGTGTTGCTA
 CATTACGTCACAGAGCTTGTATCCCGTGGATTCTGTTGTTAAATCATTTACTTAGAAAATACAGAAAGTAGAGATGAAAGTAGGTTTCAAGCATTTGAATCTTCTGTTAT
 GGAAAATTTGGTGGCTTTGTTATCAACAATTTTCATGATGACGACTTGGTAGCAATTTGGATGGCGTTAATGGCAGCGAAACAATTGGAAGAGTCCCTCAACGACCCGAT
 TAAGAAATGGAAGAACCTTTCTTTGGCTTTTGGCTGTTGCGCTGATGAATGTAACCTTGGTATGGACAATGCTTCAGTTGATGATCAAGGAATAGTGAATCATCGCGGTATCGCCGAAC
 GTGGTGTATCTTTGGTGTGTGCAAAACTTTGGATTTGGAATGGCTTAATGTTTGTCTGCTGGTGTGATATTGCTGAAGTATGAGTTATAGTTGGCTGCGGAATTTGATTGT
 TGATGAGATTCTTGTGCAAACTTTTCCATTCCGGAAATCTGGTTACACGATTAGCGGAGAAAGATTCAACACTGAAGCTCACAAATTTGGTACTGGAAGCCCTCATGCTCCGAT
 TAAGAAAGAGGATGCTTTCTAGTATCTGGTGGTGTCTGGTGTATTACTCCACTTTGTATTCTGTGAAATTTGCTAAAGCAGTGAAGGTGGCACTTACATTTTGTATGGTCTGATC
 AGCTTTGGCTGATGAACCCCTTGTGGCTAATGGTAAATCCGGAAAGATTTAGATAAAGCTGGTTTGGCATTTTGAAGGAAGTGTTCAGCTGGCGTGGTAGTAACCAA
 CTCCAAAAGTTCACAAATCTTTGATTGATAAAGTGTCTGGTATTAGGGAGGTTAGAGCATCTATTGCAAAATATAGAAGCCCATGGAGCAAAAGCTATATATTGTTCTTGGCAT
 GTATCTCCGCTGAGAAAGTAAAGGCTGCGATGCAAAAGTGTGAAGAGGAGCATCTAGTTCTGTAATTACTGGTATTGTCATCAGGCGTTTGGAGGATATAAATTTGGTTGA
 GAACAAAACTTTGATGATTCAAGCGCATATATGGAAACCAAGTAACTGGACTAGTAAACTTGTCTGTCAGCAGTGAACATGAAATTTTGTCTGTCATTTGGTTATGTTAGTTC
 TTTGGCTGGATATCATGGAAATGTTGGTCAATCTGATTATGCAATGGCTAACGAATCACTTAAACAGATTGGTTTAGATTGGTGCAGCTTATCTCAATTTGTGTTAAATCT
 ATTTGTTTGGACCTTGGATGGTGAATGGTAACTCCAGCTTGAAGAAACAAATTTCAATCAATGGTGTCCAGATTATCTCTGTAAGGTGCGCGGAGACTGTTGCAAG
 AATAGCTTATCTCAAACTCTTCAAGTTTATGTTGGCAACTGGGTGTCTCCAGTTTCACTTTGTCAAAATCGGCAACTATTGTTCAAACTTTTACCCCTGAGTTAAATC
 CATTTCTAAAGTCTCAAAATTCATGGTAAAAATGTTTGGCTATGCTCATCTGTTGAAGATTTTATGCTGGTCACTATTTGCTGGTCACTATTTGTTGGGAGT
 TGAAGATGCTCAATTTGTCAGTGGTGTGTAATTGACCATTGCGGTGCAAGCTCAAGTGAATTAACGGAACAGAGTTTGGATGATGAGCAAGGTAAAGTTCAAGCTGTTCT
 TGACTGCTTCAACAGTAAATGGTAAATGGTACCTGCATACAAAGCAGTATTGTTTGGGAAACAAAGTAGACCTGCGTTATTTTGAAGAGATTTTTCATTGCAAGATCT
 AATCTCGCAGTGTGATGAGTGAATGTTTGTTCATGTTCCATTATTTCTGTGGAATTAACCAAGTGTGTAATGATCTGATCTACTCACTAACAACCTCAATGTA
 CCAATATTGATTGACTGCTACTGAACGTGGTCAATTTGCGGATATCGAACCTGTGAATCTTTATGGCGGATGCTGCAATTTCAAGCTATGCTTGTATGGGTAGAAATTTAA
 GGAATAGTGCATCTTTACCAACAATTTGTGAAGAGTAGATATCTATAACCAATAGCACCTGGTGAAGAAAGTATTACACTCTTGTGCAAGCTTTGGGTAAATACCTCCGGTCTG
 TTCTCAAGTCTGATTATATGACGATGAACAAGGAGAGTATTCTATCTGGAAGAGCTAGTGTGTTGTTGTAATGACAAAGATGGAGTTTAG

(SEQ ID NO:68)

FIG. 16

ATGGTGAAATTAAGTGTGGTGATAAATATTGTGTCATGATCAACGGTGTGGTGTGGTATGTATGGTGTGTCAAAATCAACATGAATTTTGGCAATCTTAC
 AAGGTAAAAATATGAATTCAAAATCGAATTCACAAAATCGTTTAGTCTGAGTATAGAGAAAGAACATTTTAAACCTGAAAGAAAGTAAATATTCGGATACCTTTTGTAAATGAA
 AGATATGGTGTATTGATGAGAAATGTTCAAGTGAACATGAACCTTTTATTAAACCTTGCAAAAGATGCTATTGCGGATACAAAAGGTTCTATTGATTGAAATAAAACCCGGAAT
 CGTTAGTGGTGTCTTATCTTTTCCAAATGGATAATTAACAGGTGATTTAATAATTTGTATCAATGTCAATGAAAGAAAATTTGGGCCAAATGCATTAAGAGATGTGAATTT
 ATGGTCTAAAAGAACCCACCAACGGAAAAGATGATAAAAAGCTTATTTTGATCCTGCCCTCTTCGTAGCTGAACAATTAGATATGGGACCATTACATTATAGTTTAGATGCTG
 CTTGTGGCTGCACTTTATGTATTAAAGACTTGGTCAAGATCATTTAATTAAGTGGTCTGCTGATACAATGTTATGTGGTGCATCTTGTACCTGAACCTTTTATTATTTATCT
 GGTTTTCTACTTTTCATGCAATGCCATTATCTGGTGATGTTCTGCTCTTGCAATAAACCCTTACAAAGGCTTACACCTGGTGAAGGTGGTGTCTATTATGGTACTTAAACGAT
 TAAATGATGCAATCCGTGATGGTGATAGAAATTTAGGTACTTTACTTGTGCTGAATTAAGTAATAGTGTGTTACCATTTAGTGCACATATGCCAAGTGAATTTGATT
 GTATGGAATAAGCTTTACAAGAGTACACAGATTACCATCATCTTATTCATATGTTGAGTGTCTACCACTGTGTACACAAAGTGATAAAAGTTGAAATTTGATGCTATGACA
 AAATGTTTGGTGAACATTTACCAAGGTTTGGTTCACAGAAAGGAAATTTTGGTTCATACACTTGTGCTGCTGTTTGGTGTAAAGTTTATATATCAATGCCAATATG
 GTGAAATACCAACCACTCCAGGCTTGAAATCCAGACAATATTATGCATGATTTAGTGTACTGAACAATTCATGGCTTAATACAAATGGTGAATTTGAAACGTGCATGTT
 TATCTGCTTTTGGATTGGGTACTAATGCATGCTGATCTGATGAGAGTATCGTTCAGATTACAAAGCAAAATAAAACCTCTTGAAATGAAAGTAAAAGTCAAGTCAAGAAATCTTTT
 CTTCATTTAAAATTGCTATTGTTGGTATGGAATCTGAAATTTGGTACTTTGAAAGGATTACAAGAAATTTGAACGTGCTATTACAAATGGTGGTCAATGGTGCATGGTATTACCTGA
 AAATAGATGGAGATTTCTTGGAGAAAGATAAAGAAATTTTACAAGCTTGTGGTTACAAGAAATTTACAAAGAGTTGTTATATTAAAGAAAGTGGAAACTGATTTTAAAAGGTTAC
 GTTTACCAATGATACAGGAGGATATTCTAAGACCCTTACAGTTGTAGCTGTTTCGATTATCGACAGAGCACCTTAACGCATCTGGTGTAAACCAATGGCAAGTTGCAAGTTT
 TAGTTGGATTAGGTACTGATCTTGAATTTATATCGTCACTCGTGTGTCATTAAGGAACCGCTCCAAACCTCGGTCAAGAAAGATATCTCTTACTTGAAAGAAATTAATGA
 ACTATGTCATGATAGAGGTACAAAGTACATCATCTTATATTTGGAATTTGGTTCGAACCTGAGTTTCATCATATTATGGGGTTTACTGGTCCATTCACGATTACCTGA
 AGGTGAAATAATCCGTATATCGTTGCTGATTGGAAGATGGTCTTACGTAATGTTGGAATAGATGCTGTTGTTGCCGGGGTGTGATTTATGTTGATGCTGCTGAAAATCT
 TTTTGTAAAATCTCGTAGAAGTAAAGTTTCCACAAAAATGAACCAATTTGCAAAATTTGAATCAAAATGCTGATGGATATTTGCTGGAGATGGTGTGGAGCTTTGTTTGA
 ACGATTGAGTGTACGGATTCAACTGAAAAATTTATGCAACGGTGGATTCAATGCTGTGTGATGAAGTTGGCCCAACTATTAAAACAAGCTTTGAAGAAATGCATCCA
 TAGCAGCGAAAGATAATTGAACCTGGCAGAGCTATCAGCAAGTTCAGCAAACTATTCCTGTAAGATCACTTGTGAAGATGAACATAATGAACCTGGGTGAAATTTTCAATGA
 AGGTATACAAAGAGTTGCAATTTGGTAGTGAAAGTAAATGTTGGAGATGTTGGATATGATCTGTGTCAGCAAGTTTAAATCAAAACGGCTTTGTGCTGTACAAACCGGATAT
 TACCAAAAGTTACCAAAATGGAATAAGCCAAACGAAAGATGTTGAATGGTCCAAATCAATTTTGTATGTGAACATCTAGAGCATGGTTGAAAAATGTTGATGAAAAATAGACAT
 GCTGCTGTTTCTGGAGTTTGCGAATAATGGTTCGTGTTATGGAATCGTAATGCTGATGTACAAAGGACATCATGAAGAAATCGAAATCTGTAGTTTAGACAAAAAATGAACCCAA
 AGTACTGGGTATTTACGGAGATTCAGTTGATGATATCTGATGTTCAAGTCAACAAATATCTTGAAAAAATCTCTCAAGAACTGGAAAGGCTGCACAAAAAGTTAAAT
 CACCTACAATAGATATTGACTCCAATGTTGCTGAGATGCTTAATCTACCGCAGGATAAAACAAAAAATTTGCGTCCGATTGGTTACCAACACCAAAATAAACTCCAGCGT
 GAAATAGAACTTGCTGAAAGGTAATCCACGTTGCGTAAAGCAAAAGAGATTGGTGTCTCCATCTGGAAAGTATTTTGTGTTAAATCCACTCAAAAGTGATAAATATTGCA
 TTTATGATGGTGAAGGCCGAAGCCCATATGCTGGACTGGGATATGATTGCAATCGAATTTGGCCTATGCTACACGAGTTGGTTAAACAATAGAACTACAGAACTTTTGGGATCA
 AGGTGATAGTTGGTATTTACCTCGATCTAGCTGTGCTGAAAAAGAAAAAGTCTTCGGAGATTTTGATAAGAAATCAAAATGAAATGTTTAGATTGGGTATTTTGTATCAAT
 GTGTTTCACTGATATGGCCACTGAATTTTGGGTTTAAACCCAAAGCCGCTTTGTTTAAAGTTTGGGTGAAATATCTATGCTTTTGTGCAATTTTCTAAAAAGAAATACCAAGTTG
 TCCAAAGAAATTGACCCGTCGTAAAGAAAGCAAAAGTTTGGGCATCACAATAGCTGTGAAATTTGCAGCTATTCGAGATTGTGGAAATATCCAGCTGATAAAATCTATTGAT
 GAAATTTGGCAAGGGTATTTGTTTACGCAAAATCGAACCTTGTCGAGAACACAATTTGGGGAGATAAATTTGTTGTTGATTGTAATGATTCGCAAGAAAGTTGTCTAAT
 GCCGGGAAACCATGATGAATGTCAAAAAGTTATTGAGAAAGTATGAGAAAGTACCGGGGTTCCAGTAACCTCAGGGTATGATCGGTCATATGCCAGAAAGCAATTCCTTATCT
 AGATCAAAATCAGTCATATTCAATGTGAAATTCCAAAACCCGAAATGTGAATTTTACAACATGTGAACAAACAGAGAAATAGTGTGATGAAAGATTCCGCTGCA
 AATTGGTTGCTGAGATTATCAGCATGTTGCTGATTTTCCAAACATCGTGAACAAAGGTTAAAGAAACCTTGCAAACTGATATATTAGTAATGGGATCGAACAATATTCGAT
 CTGGAGCTGTCAAAAACAATTTTAGGTCCAGAAATCTGTTTGTGCAATTGATAGGCAAAATGAAACTGCATGGGGTCAACTAATGAAGATGGTTGTCATCTGTGATAAGTCAAT
 CGAGTCCGGGTGTGAAATGAAAAAECTATATCTGAAATTTGCTGAATTTGATCCACAGGCAAAACCGAAATCGTTTTCATCAGAAATATAGAACTGAATGGATTTTGTGAT
 CGTACGAATATTATTGTTGATAAGCAACTATCCCTCGGATCCGAACTCGCTGAAATTTGTGAACAATCGAAATATGCCATAAGATAATGTTTATGTACCAATTTGAACGGGT
 GAAAAACGATGATAAAGCGGAACCAAGTAAATTACAAAGTCAGCGTGGGAAGTAAACCAAGTTGTACTGAAAGAAATAGTTCGGACGATAATCTATTTGAAAAAGTTGTCAGAA
 ATTACAAAAATCTTTTGTGATGTTGATGCTGAAAGCAATGTTGGGAGACTCTGGATTCTCAAAAACATATGAGGTGACTATCTTTGTACACAGGTGCAATGGCTAAA
 GGAATTCGCTGCTGATTGTTGTTATTGCTGCTGTAATCAAGATCTTGGCATCATTTGGAGCTGGTGGGTTGGCCCTTACAAAGTGTGAAGATGCCATTAAACAAATTA
 GCTGAATTGGGGAACGGTCCGTTTGTCTGTAATTTGATTCAATCCACCATCGATCCTAGCTTGGAGAAAGGTAACGTTGATCTTTTCTAAAAATATAACGTTCTGATTGTTGAAG

FIG. 16 (cont'd)

TATCCGCATTTATGTCATTAAACCCCTCAGGTTGTACGATACAGAGCCCGTGGTTTGGCCAAAGCAAGAGATGGATCTGTGAAAAATTCAAAATCGTATTATTGGCCAAAATTTCAA
GAACAGAGTTAGCGGAACGTGTTCTTGAAACCAAGCACCCAAAATATTTAGATGCATTTGGTTGCGGATGGATCTATTAGTCAAGAACAAAGCCCAACTTGCATTTACTTGTGCCA
ATGGCTGATGATATTACTGTGGAAGCTGATTCTGTGGGCATACAGACAATCGACCAATTCATGTTTGTACCTTTTGATAAATTCAGCAAGAAATAGAAATTTGTAAACAAATAC
CCAAAACATTTAAAAGTTTCGAATCGGAGCAGCTGGTGGTATTGGATGCCCCGAAGGCAGCATTTGCTGCGTTTGGAGATGGGTGCTGCATACATTTGCAACTGGAACGGTAAATCA
ACTTTCAAAAGGAAGCAGGTACTTTGTGACTATGTACGTAAAGTATTGAATAAAGCTACATATTCGGATGTTACCATGGCTCCAGCCGCAGATATGTTCCGATCATGGTGTGTTGAATT
ACAAAGTTTGAAGAAAGGTACTATGTTTCTTCACGTGCTAAAAAACTATACGATTTGTTCAAAAAAATACAAATCGATTGAGGAATTAACCAAGCAGATGAGGTGAAAAAACTTG
AGCAAAAAAGTTTCAAAAAAGTCGTTTGATGAAGTATGGGATGAGACCAAGAAATTAATAATCGTTTACATTTCTCCGAAAAAAATTTGAACGTGCTGAAAGAGATGCAAAA
CTTAAAATGTCGTTATGTTTTCGTTGGTATTTGTCGAAGTCTTCCAGATGGGCTAATACCGGTGAATCTGGAAGAGTGCAGGATTATCAAAATTTGGTGTGGTCCAGCAATTGGG
TCATATAATGATTTTGCAGAAAGGATCACCATGTTTGGATCCTGAGATTTTGGTAGTTTCCAAAGTGTGTTTCAGATTAAATAAACATATTTTACGTGGTGTGCTTGTTCATCAAA
GACTCTCAGTTGAAATATCTGAATTTTAACTATGAGGAATTAGATACGTTAAACATACTCTGCATCCGAATTTTATTAA

(SEQ ID NO:70)

FIG. 17

ATGGTGAAGCTTTCCGTTGGTGACAAACATTTGGCCACGATCAGCGCTCGCCGTGGTCGGCATGTACGCCGGGTGCCCAGAACACGACGACGAGTTTGGCAGAGCCT
 CCAGGTTAAGAACATGAACAGCAGAGCATCAGCCAGAACCGCTGGGCTCCGAGTACCGGAGGAGCCTTTAAGCCGGAGCGCTCGAAGTACAGCGACACCTTCTGCAAC
 GAGCGTTACGGCTGCATCGACAGAGAACGTCAGAGCGAGCATGAGCTCTCCTGAAGCTCGCTAAGGACGCGATCGCCGATACCAAGGCGAGCATCGACCTTAACAAGAACCG
 GCATTGCTCCGGCTGCCTCTCGTTCCCTATGGATAACCTCCAGGGGACCTTCTCAACCTCTACAGTGCCTATATTGAGAAGAGATCGGCCGAAAGCCCTCAAGGATGTCA
 ACCTCTGGTGAAGCGCACGACCAACGGTAAGGACGATAAGAAAGCCCTACTTCGATCCCGCAGCTTCGTGCTGAGCAGCTTGACATGGTCCCTCCACTACTCGCTCGTGAC
 GCTGCTGCGCTCCGCTCTCTACGTCTCCGCTCGCCAGGACCACTCTCAGCGGTGCGCGACACCATGTCTGCGCGCTCGTGCTCCCGGAGCCCTTTTTCATCC
 TTTCGGGCTTTTGGACCTTCCACGCCATGCCCTTTTGGGTGAGGTGCGGCCCTCTTCAACAGACGAGCCAGGCGCTACTCCGGCGAGGGCGTGTATCATGTGCTCTGA
 AGCGCTCAAGCATGCCATTCCGAGCGGACCGCATCTACGGACGCTCTGGGCGCGAGCTTTCCAAACGGGGTTTGGGCTCCCGCTCCCGCGACATGCCGCTCCGAG
 TTCGACTGCATGGAGAAAGCCCTCCAGCGGTTACCCGCTCCCATCAGTACGTGGAGTGCACGCCACTGCAACCCCGAGGCGGACAAAGTCCGATCGAGATCGACGC
 CATGACGAAGTGTCTGCGGAGCATCTGCGCTCGCTTCCGCTCCACAGGTAACCTCGGCCACACCTCTGTGCTGTCTGGGATGTGCAAGGTCTCCTCTCGAT
 GCAGTACGTTGAGATTCTCTACGCTTGCGTGGAGAACCCGACAACTATTGACAGTCTTGTGTTACCGAGACTATTCCCTGGCCGAACACCAACGGCGATCTTAAAGC
 GTGCTGCCCTCAGCGCTTTGGCTTTGGCGGTACTAACGCCACGCGGTGTTGAGGAGTACCGCAGGACCTTCAGGCCAACAAAGACCTTGAGAACGAGAGCAAGTCCAC
 GAGATCTTTCTCTTAAGATTGCCATTGTGGCATGGATCGGAGTTTGGCACTCTCAAGGCCCTCCAGAGTTGAGCGTGCCATCTACAACGGCGGCCACGGCGGTGC
 GACCTTCGGAGAACCGCTGGCGCTTCTCGGTGAGGACAAAGGAGTTCTCCAGGCTGCGGCTCCAGAAAGTCCCGGTGCTGTACATCAAGGAGGTGAGACTGACTT
 TAAGCGCTTCGCTCCCATGATCCAGGAGACATCTCCGCCCCCTCCAGCTCTCGCGCTGCGTGATCATCGACCGCCCTCAACGCCAGCGGCTTAAGCCCAACGGCA
 AGGTGCGCTCGTGGGCTCGGCACCGATCTTGAAGTCTACCGCCACCGGCTCGGCTCGCCCTGAAGGAGCGCTTCAAGCCGCTCAAGGAGGACATCCCCCTGCTG
 GAGAACTCATGAACAGACCGCGGCACTCCACGCTACACCTGTACATCGCAACCTCTGCGACCCGCTCAGCTCGCTCTGGGCTTCAACGGGCTTCAACGGGCTAG
 CTTACGATCACGGAGGGGAGAACTCGGTTTACCGCTCGACCTCGGCTGTTCTCGCAACGGTGAGTCTGATCGCGTGGTGTCTGCTGGCGTGATCTCTGCGG
 CTCGGCCGAGAACCTGTTCGTCAAGTCGCGCTCCAAAGTGTCCACCCAGAACGAGCCCTTTGTAACTTTGAGTCGAACGCCGACGGCTACTTCGCCGGCGACGGTGGG
 GTGCCCTGTTCTCAAGCGCTTTCGACTGCATGACTCCACCGAGAGATCTACCGGACCGTGGACAGCATTTGCTGCGGACGAGGTGGGCGGCACTATAAGCAGGCC
 CTGAAGAACCGCTCGATCGCCGAAAGGACATGAGCTCGCGAGCTCTCCGCTCCAGCGCAAGCACCACTCCGGCCGATCACTCGGAGCAGGCTTAACGAGCTCG
 GCGAGATCTTCAACGAGGCGATTACGCGGTGGCCATCGGACGCTCAAGGCCAACGTCGGCGACGTCGGCTACGCCCTCGGCTGCTCAAGACGCGCCCTCTGC
 CTCTAACCGCTACCTCCCAAGCTCCCAAGTGGAAACAGCCGACAAAGGAGCTGAGTGGTCGAAGAGCTTCTTTGCTGCGAGCACTCGCGCGCTGGCTCAAGAACGT
 GGACGAGAACCGCCACCGGCTCGTGAAGCGGCTCTGCGAGAACGGCTCTGCTACGGCATCTGTCATGAGCGACTCCAGGCCACCATGAGGAGTGAACCTGTGCTCCCTC
 GATAAGAACAGGCCAAAGTGCTCGGTATCTACGGCGATTCCGTGGACGATATTCTGTCCAGCTGAACAGTAACTGAGGAGTCTTCAAGGAGTGGCACTGGCACTGGCGG
 TGCGCAGAAAGGTGAAGAGCCCTACCATTGACATCGACTCGAACGCTTTTGCCGAGATGCTGAACCTTCCCAAGGACAAAGAAAGTTTGGCGTCTGGTCAAGGACCC
 CCAACAGCTCCAGCGGAGATTGAGCTCGCGTTAAGGGCATCCCTCGCTGCGTGAAGGCCAAGCGGACTGGTGTCTCCCTCCGGCAGCATCTTTTGGTGAACCCGCTC
 AAGTCGGACAAACATTGCTTTATGACGGGAGGGCGCTCGCTTACGGCGGCTCGGCTAGCATCTCCACCGCATCTGGCCCATGCTTACGAGCTGTGAACACCCGAC
 GACTGAGCTGTGGACCAAGGTGACTCGTGTACTCGCGCGCAGCTCTCCGTGGCCGAGAAAGGAGAGGTCTTTTGGCGACTTCGACAAAGAACAGATCGAGATGTTCGCG
 CTCGGTAATTTCTCAGCATGTCTTACCAGATGCGGACGGAGCTCTCGGCTTAAAGGAAAGCGGCTTTCCGGCTCTCCCTCGCGGAGATCAGCATGCTCTTGTCTTCT
 CGAAGAGAACACCAAGCTCTCCAAGGAGCTTACTCGCGCTCAAGGAGGCCAAGGTGTGGGCTGCGAGCTGGCGGACTTCGCCGCCATCCCGGCACTTCGGAACAT
 CCGCGCCGACAACTGATGTTGCGGAGTTACTTCGTTACGCCAACCTACGCTGTGGAGAACCACTATGCGGAGAACAGTTCGTCGCGCTCTTCTGCTCCCTTACGTCAA
 CGACTCCAGTCTGCCTATTGCGGTAAGCCCGATGAGTGGCAGAGTCCACCTTAAGCTCCCGCGCTCCCGCTACCTTAAGCTCCCGCTCCCGCTACCTTATGGCCACT
 GCCGGAGGCCATTCCCTACCTCGACCAAGATCAGCCACATCCACGAGATGCTTGAGATCCCGAAAGCTGAGAACGTCAGGCTCTTCAACGACGTCGAGAACCGCGAGCTGTG
 TCGATGAAGGACTCCGTTAGCAAGCTGTCGGGAGATCTACAGCAGCTGCTGACTTCCCAACATTGTCAACAGGTCAAGGAGCTTGCAAGACGGACATTTTCAATCGA
 GCTGGGACGAAACAACTACCGTTCGGTCCGTCAGACTATCTCGTCCGAGATCGTGAGCTGTCATCGACCGTCAAGACGAGACTGCTTGGGCCAGCTCATGAAGA
 TGGTCGCCAGCTGATCTCCACCGGTCCCGGCTCGAGCTCAAGAGCTGTACCATCCGGAGCTCTGAAGTTCGATCCCGAGGCCAAGGCCAAGCCGCTTATCCGCAAC
 ATCGAGCTCAACGGCTTTTTCGACCGCACGAACATCATGCTGATAAAGAGCTTTCCCTCGCGGACCCGAAGCTCGCCGAGATGCTCAACACCGCAACATGCCGAAGGATAA
 CGGTACGTCCCCATTGAGCGGTCAAGGCCGAGCCGCTAACTCCAGGTGCTCGGTGCGAGCCGCTGACCGGAGTATCTCGTGGAGGACA
 ACCTCTTGAGAAAGCTCTCGGAGATCACTAAGTCTTCGACCGTGCACCGGCTGACCGGAGCCATGCTCGGCGATTCGGGCTTTCTCAAGACGTACGAGGTGACTACCCCG
 TCTACACCGGCGCTATGGCCAAAGGTTATCGCCTCCGCGCACCTCTGCTATTGCTGCTGCGGCGGTAAAGTCGAAGATCCTTGGTGTCTGGCGCTCGCTCTCCAGGTGCTCG

FIG. 17 (cont'd)

AGGATGCCATTAAAGCAGATCAAGGCTGAGCTTGGCAACGGTCCCTTTGCCGTCAACCTCATCCACTCGCCTTTCGACCCCTCGCTTGAGAAAGGCAACGTTGACCTTTTCCTCA
 AGTACAACGTCCGCTTTGTGAGGTGAGCGCTTCATGAGCCTCACCCCTCAGGTCGTTCCGCTACCGCGCTGCCGCTTGCCAAGGCCGTGACGGCTCGGTCAAGATTTCAG
 AACCGCATCATGCCAAGATTTCGCGACGGAGCTGGCCGAGCTCTCCCTCAAGCCCGCTCCGAAGAAC/ATCCTCGATGCCCTCGTTGCCGACGGCTCGATTTCCTCCAGGAGCA
 GGCTCAGCTCGGCTCCTCGTCCCTATGGCCGATGACATCAACCGTTGAGGCCGACTCCGGTGGCCACACCGGACAAACCGCCCAATTCAATGTGCTCTCCCTCATCATCCAGCA
 GCGCAACCGCATTTGCAAGCAGTACCCGAAGCACCTCAAGTCCGATCGGCGTGGCATCGGTTGCCCTAAGCGGCTTTTGCCGCTTTGAGATGGGTGCGGCT
 ACATCGCCACGGGCAACCGTTAACCAAGCTCTCGAAGGAGGCCGGCACCTGCGACTACGTGCGCAAGGTGCTCAACAAGGCCACCTACTCCGACGTCACGATGGCTCCCCGTGCC
 GACATGTTGACACCGGTGTCGAGCTCCAGGTTCTCAAGAAGGGCACCATGTTCCGTGCGGCCAAGAAAGCTCTACGACCTCTTAAAGAAAGTACAAAGTCGATCGAGGAGCT
 CCCTGCCGACGAGGTCAAGAAGCTGGAGCAGAAAGGTTTAAAGATCGTTCGACGAGGTCTGGGACGAGACTAAGAACTACTACATTAAACCGCTCCACTCCCTGAGAAG
 ATCGAGCGCGGAGCGTGACGCCAAGCTGAAGATGTCGCTCTGCTTTCGTTGGTACCTGAGCAAGTCTCCCGCTGGGCCAACACCGGCGAGTCGGGCGGTGTCCAGGACTA
 CCAGATCTGTTGCGGCCCCCATCGGCTCGTACAAACGACTTCGCGAAGGGCTCGCCCTGACCTTGAGATCCTTGGCTCGTTCCCGTCCAGATCAACAAGCA
 TATTCTGCGCGGCGCTTGCTTCTACCAAGCTCTTTTCGCAAGTCAAGTACCTTAACCTCAACTACGAGGAGCTCGATACGCTACCTACAGCGCTAGCAACCTTATCTAA

(SEQ ID NO: 121)

FIG. 18

MVKLSVGDNICHQORVAVVGMVAVMYAGCQNHQHEFWQSLQGNMNSKSIQNRGLGSEYREEHFKEPERSKYSDTFCNERYGCIDENVQSEHELLLKLAKDALADTKGSIDLNKTGIVS
 GCLSPMDNLQGDLLNLYQCHIEKKIGPNALKDVNLWSKRTTNGKDDKKAYFDPASFVAEQLDMGPLHYSLDAACASALYVLRQAQDHLHLSGAADTMLCGASCLPEPFFILSGFSTF
 HAMPKSGDVSAPLHKTSQGLTPGEGGAIMVLKRLNDAIRDGDRIYGTLLGAELSNAGCGLPLSPHMPSEFDCMEKALQRVHRLPSSIQYVECHATGTTPQGDKVEIDAMTKCFGEHLPR
 FGSTKGNFGHTLVAAAGFAGMCKVLLSMQYGEIPPTPGLNPDNIMHDLVVTEIPWPNTNGDLKRALSAFGFGGTNAHAVFEEYRSDQANKTLENESKSHIEFSSFKIAVGMSEF
 GTLKGLQEFERAIYNGHGAACDLPENRWRFLEDKEFLQACGLQKLPKRGYIKEVETDFKRLRPMIQEDILRPLQLLAVSIIDRALNASGVKPNGKVAVLVGLGTDLELYRHRARVA
 LKERLQTAVKEDIPLEKLMNVVNDRGTSYTSYIGNLVATRVSLLWGTGTPSTTTEGENSVYRCLDLGRWFLANGEVDAVVAGVDLCGSAENLFVKSRRSKVSTQNEPFFANFES
 NADGYFAGDGCALVLKRLSDCTDSTEKIYATVDSIAVGDDEVGPTIKQALKNASIAAKDIELAELSSASSGKHHSGRITCEDELNELGEIFNEGIQORVAIGSVKANVGDVGYASGAASLI
 KTALCLYNRYLPKLPNWNKPTKDVESKSFVCEHSRAWLKNVDENRHAUVSGVCENGSCYGVMSDVQGHHEESNLVSLDKNEPKVLGIYGDSVDDILVQLNKYLEKFLQETGT
 AAAAQKVKSPPTIDDSNVFAEMLNLPQDKNKKFAVALVTTTPNKLQREIELAVKGIPRCVKAKRDWCSPSGSIFACNPLKSDNIAFMYGEGRSPYAGLGYDLHRIWPMHLELVNNRTTE
 LWDQGDWSYLPRSSVAEKEKVFGDFDKNQIEMFRLGIFVSMCFDMMATELLGLPKAAFGLSLGEISMLFAFSKKNTKLSKELTRRLKEAKVVASQLAVEFAAIRDLWNIPADKSID
 EFWQGYFVYANRTLVENTIGENKFVRLLVNDQSQCCLIAQKPDCECKVIEKLHLKLPAPVPTQGMIGHCPEAIPYLDQISHIHEMLEIPKENVKLTTSSENRELVSMKDSVSKLVAEY
 QHVADFPNIVNKVKECTKTDIFIELGSNNYRSGAVKTIILGPEIVSAIDRQNETAWGQMKMVASLISHRVPGVVELKLYHPELLKFPDQAKPNRFRNIELNGFFDRNTIIVDKQLSPA
 DPKLAEIVNNRNMPKDNVYVPIERVKTMIKAEANLQVSVGSKPVTTERISSDDNLFELKSEITKSFDFGVNACTEAMLGDSGFLKTYEVVYPLTYTGAMAKGASADLVIAAGKSKILA
 SFGAGGLALQVVEDAIKQIAELGNGPFAVNLJHSPDPSEKGNVDLFLKYNVRFVVSFMSLTPQVVRYRAAGLAKARDGSVKIQNRHIAKISRTELAELFLKPAKPNILDALVAD
 GSISQEQALALLVPMADDITVEADSGGHTDNRPIHVLLPLIQQRNRICKQYPKHLKVRIGAAAGGIGCPKAAFAFEMGAAYIATGTVNLQSKEAGTCDYVRKVLNKAATYSYSDVTMAP
 AADMFDHGVLEQLVKKGTMFPSPRAKKLYDLFKYKSIEELPADEVKKLEQKVFKKSFDEVDFTKNYYINRLHSPKIERAERDAKLKMSLCFRWYLSKSSRWANTGESGRVQDY
 QIWC GPAIGSYNDFAKGSPCLDPEILGSFSPSVVQINKHILRGACFYQRLSQLKYLNFNYEELDTLTYASASNFI

(SEQ ID NO:71)

(SEQ ID NO:72)

GATGAAGATGTATTACCTATCGTCCAGAATCAACCCCTATAG

FIG. 20

ATGTTGGCTGCAGATGAAGAAGCCCTGTGTGGAGATGTCGAAGGAGGAGCAGTCGTCCGGCAAGAACGTCGTTTGAACACGAGAGCTCCTCGAATTCGCGGAGG
 GTGACATCGGCAAGGTGTTCGGCCCAAGTTTGACATCATGACAAGTACAGCGCCGTGTGCGCTCCCGCCGCGAGTACCTCTCTGACACCGGTGACGTCATGGAT
 GCCGAGTGGCAACTTCGGCTGCGCTCGCATGGTCAACGAGTACGAGTCCCGGTGAACGGCGAGCTTTCCAGGGCGGAGCTTCCAGGGCGGAGCTTCCCTGGGCTGCTCTGTCGAGTTC
 GGGCAGTGCACCTCATGTTATCTCTGATCATGGGCAATTGACTTCAAGGTGACCGCGTTTACCGCCTTCTCAACACGACCTCAAGTCTACGGTGTGCGCCACGA
 GGGCGAGACTCTCGTTTACGACATCCGCTCACTGGTTTCCAAAGGGCATGACGGCGAGATTAGCATGTTCTTCTCGAGTACGACTGCTACGTCAACGGCCGCTGCTCAT
 CGAGATGCGGACGTTGCGTCTTCTACGGAGGAGTTCGCCGGGCAAGGGCGTTCATCAAGAGCCTGCTGAGTCCACAAAGCGCAAGTCTGATTGTGCCAAG
 TCGATCAAGCCTTTTGGCCTCAACCCCGCTCCACAAGAGATGTTACGGGAGAACGACATGAGAACTTTGCGAGCCAGTGGGAGAACGTCCTGCGGCTCCGGCTCCA
 GGGCATCGATCAAGCTGTGGCCCGCAAGATGCTCATGATCGACCGCATCAAGAGATCCAGACAAACGGCGGTGCTAGCGCCTGGCTCTGTTGGCGAGAAGATTC
 TTGAGCGGACCATTTGCTTCCCTTGCCACTTCGTCAAGGACCAAGTATGGCGGTCCCTCGTTAGCGACGGTGTCTCGAGTGTCTAAAGCTTTACATGCTTTGGCTCG
 GCCTCCAGCAGTGTCCCGGATTTCCAGTTCCGTCTCTGTCCGTGGCAGCCCAACAAAGTGGCTGCCGTGGCCAGATAGCCCCCATCTGTGGCAAGCTGTGATGATGG
 AGATTCGCGAGATGGTTTCAACGAGTCCACGGCCAGCCCTACGGATGCTGAGCTTACATGATGTGAACACTACGAGCTCGGCCAGTCTTTGACATGGCCGACATC
 GACTCGTACGGCCGTGGCAACCTCTCCAAAGAAATTGTCTCGTCAATTCAAGGGCATTTGCTCCAGATGGAGGGCACCGTCAAGAGCTCCAACATCATGATTCGTCCCCAA
 GTCCACGATTATCCAGCCGCGCCCAACTGCTCCGGCGATCCTCTCGCCCTCGCAGGTCACTGGCACCCGATGGCCGTGTCAACGGCGCCCTCCCTCTTCAG
 CCCGTGCGATTACCTCTCTGTGCCCTTTCCTTAAAGCCCTTCCTGGCAACCCCTCGACAACGATCATACGCCGGGCAAGATGCCGTGACCTGGTTTAACTATGTCGGAGTT
 TATGTGGGCAAGGTCAAGCACTGGCTTGGCCCTGAGTTTAAAGCGTTTCGACAACTCCAAGAGAGCGCTCCCGCCCTTCGACCTGGCCCTGGTTACGGCGGTGGTGTCCGT
 CAGCGATATGGAGTTCAAGCCCACTCAACATCAGCTCAACCCGTCAAGGGCACGATGATTGGCGAGTTCGACTGCCCCGTGACGCTGCTTCTTTTTCAGGGCTCTCTGCA
 ACGACGGCCACATGCGGTACAGCATCTGATGAGATCGCCCTTCAGACCAGGCTGCTCACTCCGTCTCAAGGCCCGCTCACTATGGACAAGGACGACATTCCTTTTC
 GCAACCTCGACGCCAACCGCGAGATGGTCTCCGACGTGATTCGCCGGTAAGACCATCAAGAACTTCAACCAAGTCAACGGCTACAGCATGCTTGGCAAGATGGGCATC
 CACCGTTCACCTTTGAGCTCTCGTGCATGACGTGCTGTTTACAAGGGCTCGACAGCTTTTGGTTGTTTACGCCGAGGTGTTGAGTCCGAGGTGCGGCTCGATAACGGC
 AAGAAAGTCCAGCCGTGATCTGGAGCAGAAGTCTCGAACGTTGTGACGTACGATGTCCGCTCGACCCCGCAAGGACAAGCTCTTCTCGAAGATCGGCTCGAAGGACG
 CTCAGGTCCAGCTCGAACACCAAGTGGAGTTTCTCGACACGATGACATTTCCGAACACCGCAAGTACAAACAGGGCTACGCGACGGTGAGAAGAAAGGTCAACCC
 CAACGACTGTTCTCTGCCACTTTTGGTTTCGACCCGGTATGCCGGTCTCTCGGTATGAGTCCATGTTCCAGCTCATCGAGGCTTTTTCGATTGACCAAGGTATCGCG
 TCCAAGCATGGCATCGTAACCTTCTCGCATCGAACCGCAAGACCTCGTGGAAGTACCGGGCCAGCTCAACAAGGCAAGGCCATGGACAGCGAGATTCAACA
 TCAAGGATATTGTCAAGAACGCGGACGGTACTGTGATCTCATCGCCGATGTTTCTCTGTGAGTCTGCTCGGTTTACAGCGCCGATGACCTCCGCTCAAGATCGTCC
 CCGCCTAAAGCTGTCCCAAGAGCTCGCGCCGCTCCGCGCATGTGGCCACTCCGATCCCGGCTCCCAACCTCTCGGTGGAGATCTCGTTGAGTCCCTTA
 AGAAGAGCTCTCAACCTGGAGAAGCCCTCTACCTTGAGACTTCCAACACATCGTCAAGAGTTTCGGGACGTTTAAACAGGCCAGGCTTCCGTATCCCTCCGTGCAAC
 ATTAACGATCTCGGTGAGCGCTGTTATGGAGACTTACAACGTCGTGCTCCCTTACACCGGCGCATGGCGAAGGGCATCGTTTCGGCGACCTCGTCATCGCTGCCGGT
 AAGCGCAAGATCTCGGACGCTTCGGCGCGGTGGCTCCGATGCACCTCGTGGCGCTCGCTGGTGAGAAAGATCCAGGCCCTCCCGAGGGCCCGTACGCGGTCAACCT
 CATCCACTCGCTTCGACTCGAACCTTGAGAAAGGTTAACGTGGACCTTCTTCGGAGAAGGGGTCCACGTGGTCTGAGGCTCCGCTTACCGCTTACCGCTCACGACCCAGGTGCT
 TCGTACCGCGCTGCGGCTCTCGGTGCTAAGGAGGTTTCGTGTTGATTAAAGAACCGCATCATCGTAAAGTCAAGCGCACGGAGCTTGGCGAGATGTTCTTTCGCCCTGC
 CCCCAGAACCTCTCGATAAGCTCATCGCCAGGGCGAGATCAACAAGGACGAGCGCTCCCTGCTTGAAGTTCTTATGGCCGACGATGTGCTGTTGAGGCCGACTCCG
 GCGGCCACACCGATAACCGTCCATCCACGTACTCCCGTATTAACTCCGCAACCGTATCCACAAGGATGCGGCTTCTTCTGCGCTCTCCGCGTCCGCTGCGGCG
 CTGGTGGTGGCATCGGTTGCCCTTCGGCGCTGTGCGGCTTCAACATGGGCGGCTTCTCTGATCACCGGCTCCGTTAACAGGTGAGCAAGCATCCGGCACGTGCGAC
 ATTGTGCGCAAGCACTTACGAGGCCAGTACTCCGACATCAGATGGTCCCGCTGATGTTGAGCTGAGCTCGAGGTGGAGTCCAGGTCTCAAGAAGGTACGATGTT
 TCCCTCGCGGCCAAGAAGCTCTACGAGCTTTTGTGATGATCAACAGCTTTGACGACATGCCGAAGTCCGAGCTCCAGCCTGGAGAAGCGCATTTTCCAGAAGAGCTCG
 CCGAGGTCTGGAGGAGACTAAGGACTTTTACATCAACCGCTCAACAACCGGAGAGATCGAGACGCGGAGAAGAGGACCCCAAGCTCAAGATGTCCCTTTGCTTTCG
 CTGGTATCTCGGCTTTCGAGCTTTTGGGCCAACAACGGCATCAAGGAGCGCAGCATGGATTACAGATTGTTGTGCGGCTCCGGCCATTGGCAGCTACAACGACTTCGTGAAG
 GCACCTACCTCGACCCCGCTGTCGCTGATCCCTGTGCTGAGATCAACATGACAGATCTCCGCGGTGCTGCTCTCTCCAGCGGTCCGCGCATTAAGCAGCAGC
 CGCGCTCGATATCGACGTTGATGAGGACGCTTTTACCTACCGCCCGGAGACCCCTCTAA

(SEQ ID NO:122)

FIG. 21

MVGLQMKKKPVWEMSKEEQSSGKNVVFDDYDELLEFAEGDIGKVFGPKFDIIDKYRRVRRLPAREYLLVTRVTLMDAEVGNFRVGSRMVTEYDVPVNGELSQQGDVPWAVLVESGQ
 CDLMLISYMGIDFQCKGDRVYRLNLTTLTFYGAHEGETLVYDIRVTGFAKGMHGEISMEFFEYDCYVNGRLLIEMRDGCAGFFTDEELAAGKGVIKTV AELHKRKSIIVPKSIKPFAL
 NPAVHKTMFSENDMEKLCERQWENVLGSGLQGIDYKLCARCKMLMIDRITIKIQHNGGAYGLGLLVGEKILERDHWYFPCHFVKDQVMAGSLVSDGCSQLLKLYMLWGLHHDVVVPD
 FQFRPVPGQPNKVRRCRGQISPHRGKLVYVMEIREMGFNESTGQPYAIADVDIIDVNYELGQSFDMAIDISYGRGNLSKKIVVDFKGIALQMEGTVKSSNIIDSSPKSTIIQPPPNCLRGDP
 LAPSQV'TWHPMAGVNGAPAPSFSPSDYPPRAVCFKPPGNPLDNDHTPGKMPLTWFNMFSEFCGKVSNCCLGPEFKRFDNSKTSRSPAFDLALVTRVVSVDMEFKPHLNDVNPSSKG
 TMIGEEFCPADAWFFQGSNCNDGHMPYSIVMEIALQTSGLVTSVLKAPLTMDKDDILFRNLDAEMVRSVDVDCRGKTIKNFTQCTGYSMLGKMGHIRTFFELSVDDDVVFYKGSTSG
 WFTPEVFESQVGLDNGKKVQPWYLFQKSSNVVTYDVASTAGKDKLFSKIGSKDAQVQRNNTQCEFLDTMHIIPNTGKYNKGYAHGEKKVNPNDWFFSCHFWFDFVMPGSLGIESMF
 QLIEAFSIDDQGLASKKHGIVNPTFAHNSNGKTSWKYRGQLNNKGRMDSEIHIKDIVKNADGTVDLIADGFLLVDSL RVYSADDLRVKI VPGTKAAPKSVAAAPRHVATPIPGVPSNTSSV
 EISLESKKELLEPLYLETSNHHVKQFGDVNNGQASVIPPCTINDLGERSEMETYNVVAPLYTGAMAKGIASADLVIAAGKRKILGSFGAGGLPMHLVRASVEKIQAAALPEGPYAV
 NLIHSPFDSNLEKGNVDLFLEKGVHVVEASFTALTQVVRYRACGLSRAKDGSVLIKNRHIGKVSRTELAEMFFRPAPQNLLDKLIASGEITKEQASLAEVPMADDDVAVEADSGGHT
 DNRPIHVILPLIINLRNRHKECGFPAALRVV'GAGGGIGCPSAAVAAFNMGAFLITGSVNQVSKQSGTCDIVRKQLSEASYSIDITMAPAADMFDQGVELQVLKKGTMFPSPRAKKLYE
 LFCMYNSFDDMPKSELQRLEKRIFQKSLAEVWEEETKDFYINRLNPNPEKIEHAEKKDKPKLKMSLCFRWYLGLSFFWANNGIKERSMDYQIWC'GPAIGSYNDFVKGTYLDP'AVAGSYPC
 VVQINMQILRGACFLQ'RV'RAIKHDPRLDIDVD'EDVFTYRPESTL

(SEQ ID NO:73)

FIG. 22

***Schizochytrium* Codon Usage**

AA*	Codon	Fraction	AA*	Codon	Fraction	AA*	Codon	Fraction	AA*	Codon	Fraction
Ala	GCC	0.64	End	TAA	0.34	Leu	CTT	0.16	Ser	TCG	0.33
Ala	GCA	0.03	End	TGA	0.33	Leu	TTG	0.02	Ser	TCC	0.31
Ala	GCT	0.18	End	TAG	0.33	Leu	CTG	0.12	Ser	AGT	0.03
Ala	GCG	0.16	Gln	CAA	0.08	Leu	CTC	0.69	Ser	TCA	0
Arg	CGG	0.01	Gln	CAG	0.92	Leu	TTA	0	Ser	TCT	0.09
Arg	AGA	0	Glu	GAA	0.09	Leu	CTA	0	Thr	ACG	0.3
Arg	CGC	0.8	Glu	GAG	0.91	Lys	AAA	0.04	Thr	ACC	0.54
Arg	CGA	0.01	Gly	GGA	0.1	Lys	AAG	0.96	Thr	ACA	0.02
Arg	AGG	0	Gly	GGT	0.2	Met	ATG	1	Thr	ACT	0.14
Arg	CGT	0.17	Gly	GGG	0	Phe	TTT	0.45	Trp	TGG	1
Asn	AAC	0.94	Gly	GGC	0.7	Phe	TTC	0.55	Tyr	TAC	0.94
Asn	AAT	0.06	His	CAC	0.83	Pro	CCT	0.21	Tyr	TAT	0.06
Asp	GAT	0.24	His	CAT	0.17	Pro	CCG	0.34	Val	GTC	0.62
Asp	GAC	0.76	Ile	ATC	0.7	Pro	CCC	0.43	Val	GTA	0
Cys	TGC	0.95	Ile	ATA	0	Pro	CCA	0.02	Val	GTT	0.14
Cys	TGT	0.05	Ile	ATT	0.3	Ser	AGC	0.24	Val	GTG	0.24

* AA=Amino Acid