US 20060094102A1

(19) **United States**

(12) **Patent Application Publication** (10) **Pub. No.: US 2006/0094102 A1**

Xue et al. (43) **Pub. Date: May 4, 2006**

(54) **AMMONIUM TRANSPORTER PROMOTER FOR GENE EXPRESSION IN OLEAGINOUS YEAST**

(76) Inventors: **Zhixiong Xue**, Chadds Ford, PA (US); **Quinn Qun Zhu**, West Chester, PA (US)

Correspondence Address:
E I DU PONT DE NEMOURS AND COMPANY
LEGAL PATENT RECORDS CENTER
BARLEY MILL PLAZA 25/1128
4417 LANCASTER PIKE
WILMINGTON, DE 19805 (US)

(21) Appl. No.: **11/185,301**

(22) Filed: **Jul. 20, 2005**

Related U.S. Application Data

(60) Provisional application No. 60/624,812, filed on Nov. 4, 2004.

Publication Classification

(51) **Int. Cl.**
C12N 9/26 (2006.01)
C12N 15/74 (2006.01)
C12N 1/18 (2006.01)

(52) **U.S. Cl.** **435/201**; 435/254.2; 435/483

(57) **ABSTRACT**

The promoter region associated with the *Yarrowia lipolytica* ammonium transporter (yat1) gene has been found to be particularly effective for the expression of heterologous genes in oleaginous yeast. The promoter regions of the instant invention have been shown to be advantageously inducible under oleaginous conditions (i.e., nitrogen limitation) and are useful to drive expression of genes involved in the production of ω-3 and ω-6 fatty acids.

Figure 1

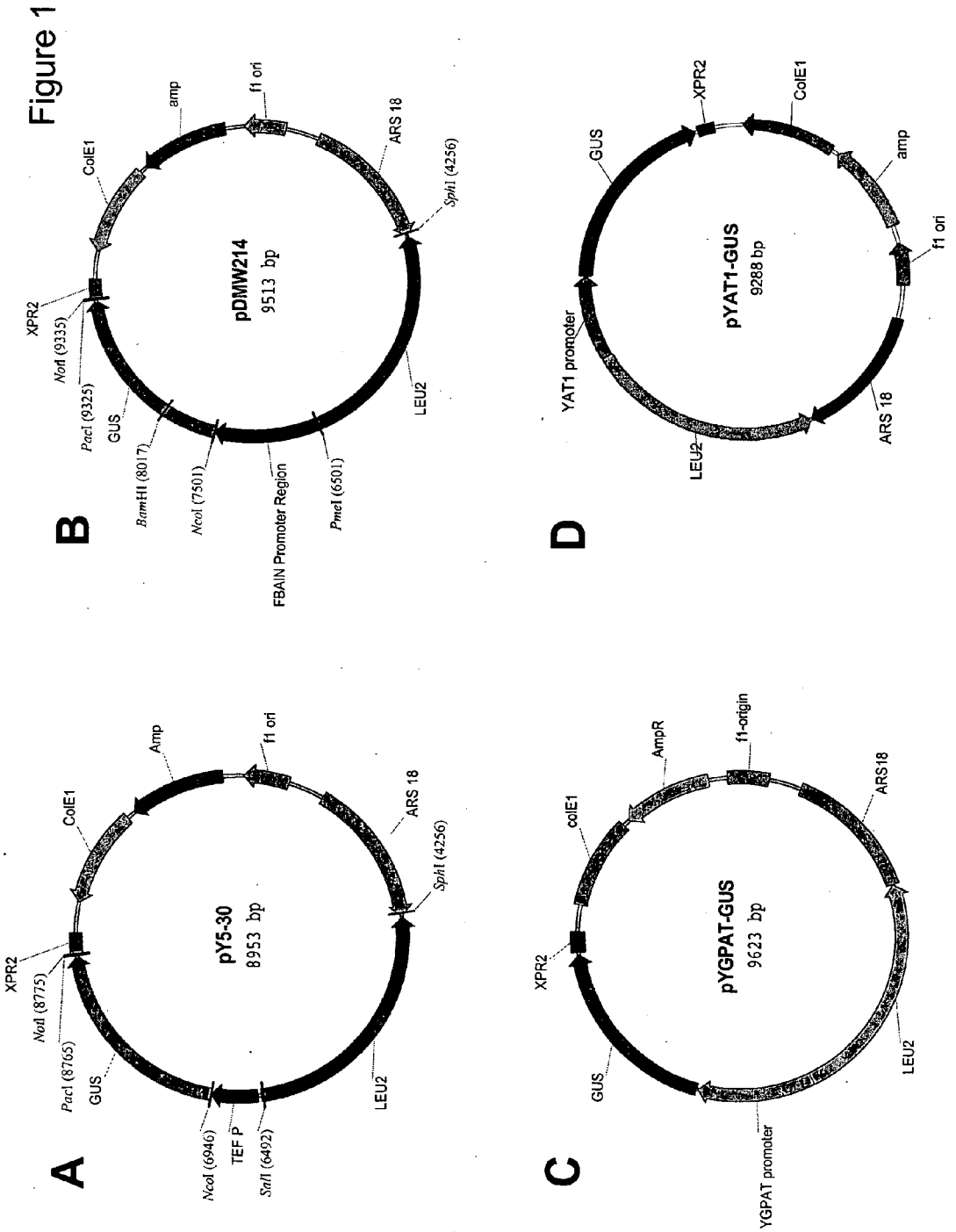

A
pY5-30
8953 bp
Amp
f1 ori
ColE1
ARS 18
XPR2
Not1 (8775)
Pac1 (8765)
GUS
Sph1 (4256)
LEU2
Nco1 (6946)
TEF P
Sal1 (6492)
B
pDMW214
9513 bp
amp
f1 ori
ColE1
ARS 18
XPR2
Not1 (9335)
Pac1 (9325)
GUS
Sph1 (4256)
LEU2
BamH1 (8017)
Nco1 (7501)
FBAIN Promoter Region
Pme1 (6501)
C
pYGPAT-GUS
9623 bp
AmpR
f1-origin
colE1
ARS18
XPR2
GUS
LEU2
YGPAT promoter
D
pYAT1-GUS
9288 bp
XPR2
ColE1
GUS
amp
f1 ori
YAT1 promoter
ARS 18
LEU2

Figure 2

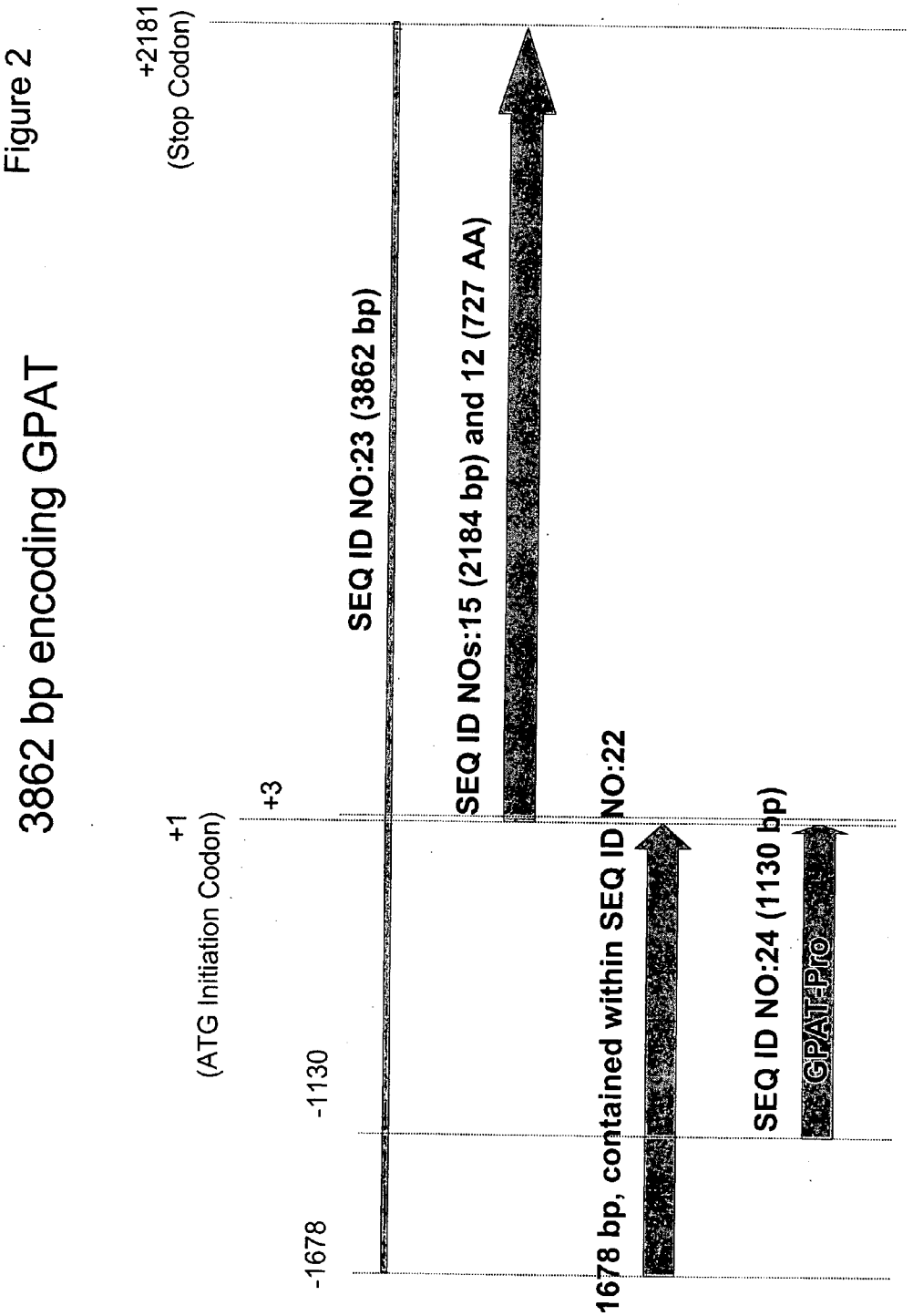

3862 bp encoding GPAT
+2181
(Stop Codon)
+1
(ATG Initiation Codon)
+3
-1130
-1678
SEQ ID NO:23 (3862 bp)
SEQ ID NOs:15 (2184 bp) and 12 (727 AA)
1678 bp, contained within SEQ ID NO:22
SEQ ID NO:24 (1130 bp)
GPAT-Pro

Figure 3

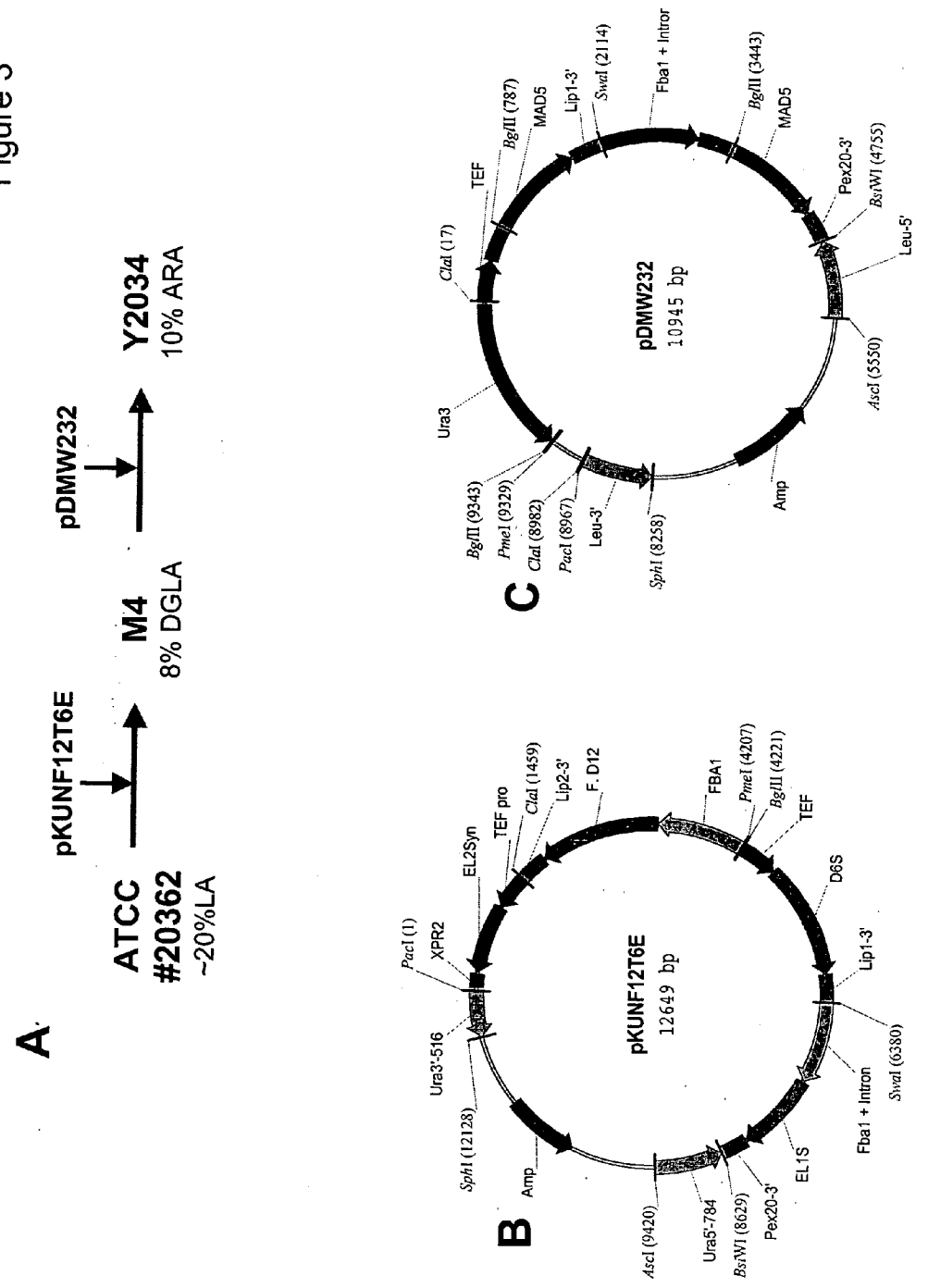
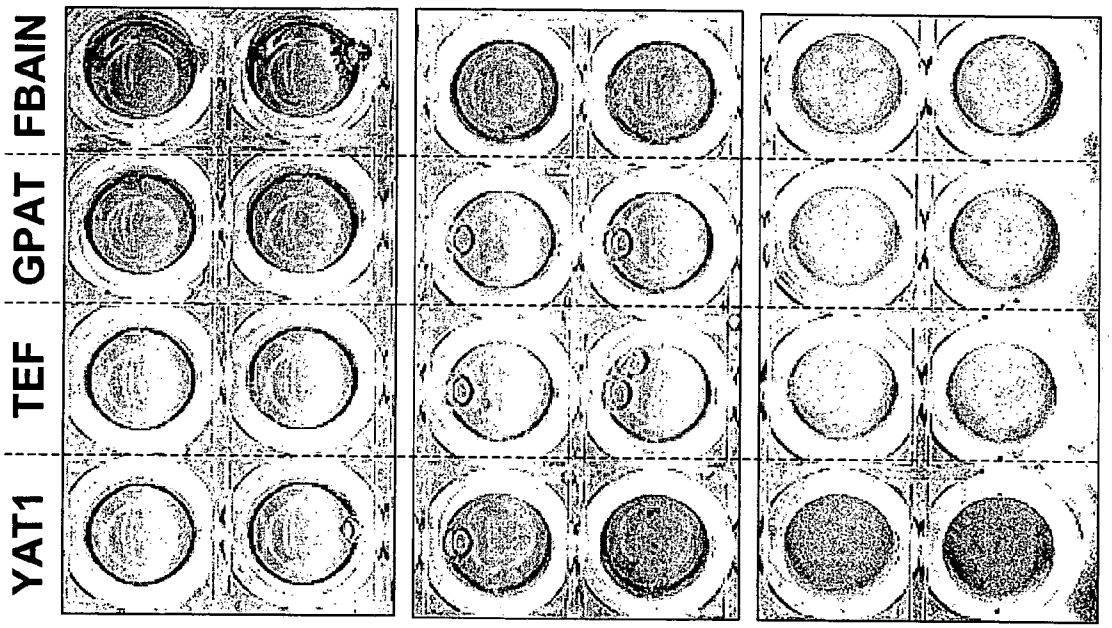

A
ATCC #20362
~20%LA
pKUNF12T6E
M4
8% DGLA
pDMW232
Y2034
10% ARA
B
pKUNF12T6E
12649 bp
PacI (1)
XPR2
Ura3'-516
SphI (12128)
Amp
AscI (9420)
Ura5'-784
BsiWI (8629)
Pex20-3'
EL1S
Fba1 + Intron
SwaI (6380)
Lip1-3'
D6S
TEF
BglII (3443)
FBA1
PmeI (4207)
BglII (4221)
F. D12
Lip2-3'
ClaI (1459)
TEF pro
EL2Syn
C
pDMW232
10945 bp
ClaI (17)
TEF
BglII (787)
MAD5
Lip1-3'
SwaI (2114)
Fba1 + Intron
BglII (3443)
MAD5
Pex20-3'
BsiWI (4755)
Leu-5'
AscI (5550)
Amp
SphI (8258)
Leu-3'
PacI (8967)
ClaI (8982)
PmeI (9329)
BglII (9343)
Ura3

Figure 4

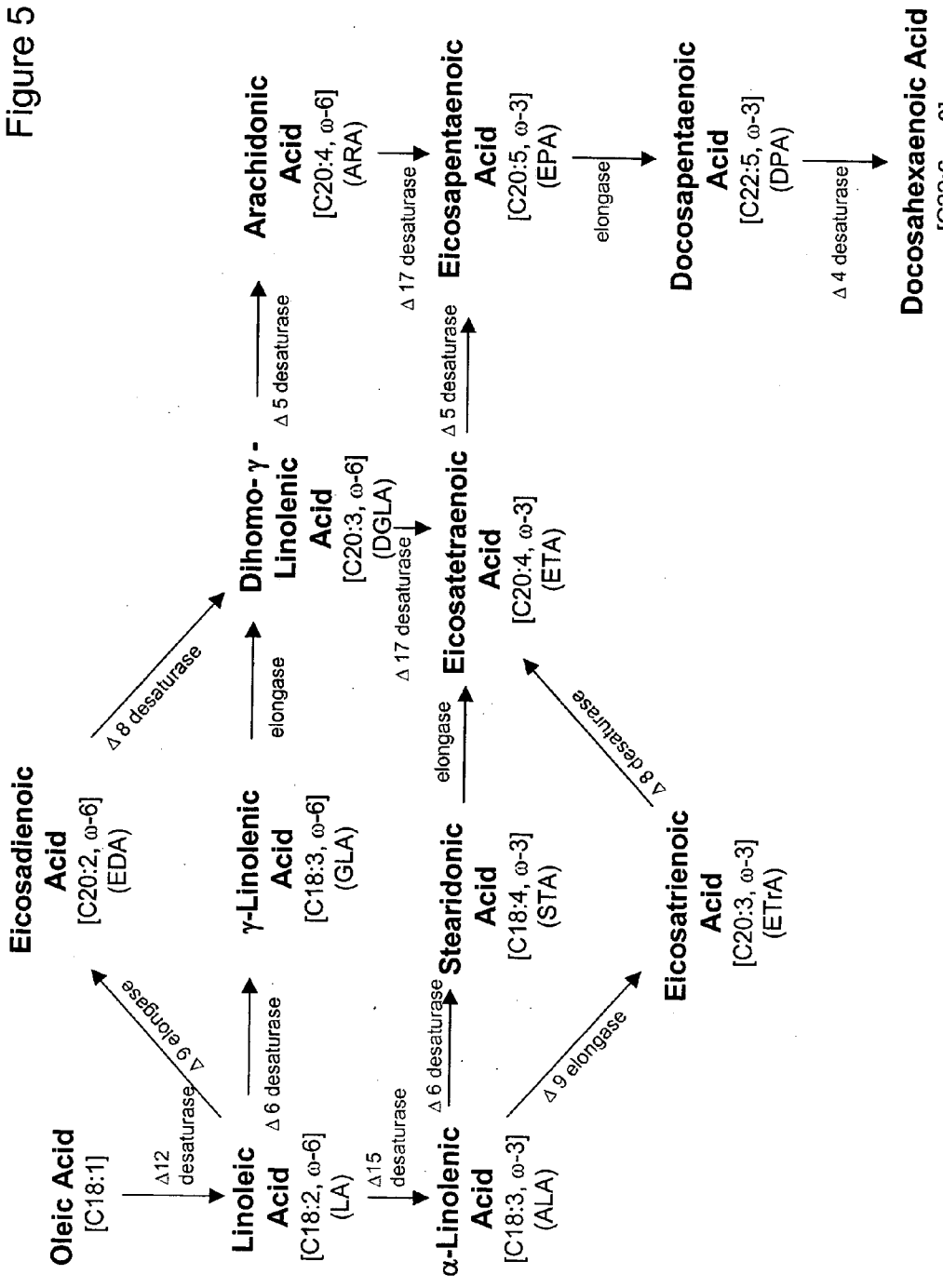


A: 24 hr in SD medium

B: 24 hr in HG medium

C: 120 hr in HG medium

Figure 5


Oleic Acid
[C18:1]
Δ12 desaturase
Linoleic Acid
[C18:2, ω-6]
(LA)
Δ 6 desaturase
γ-Linolenic Acid
[C18:3, ω-6]
(GLA)
elongase
Dihomo-γ-Linolenic Acid
[C20:3, ω-6]
(DGLA)
Δ 5 desaturase
Arachidonic Acid
[C20:4, ω-6]
(ARA)
Δ 17 desaturase
Eicosapentaenoic Acid
[C20:5, ω-3]
(EPA)
elongase
Docosapentaenoic Acid
[C22:5, ω-3]
(DPA)
Δ 4 desaturase
Docosahexaenoic Acid
[C22:6, ω-3]
(DHA)
Δ 9 elongase
Eicosadienoic Acid
[C20:2, ω-6]
(EDA)
Δ 8 desaturase
Δ15 desaturase
α-Linolenic Acid
[C18:3, ω-3]
(ALA)
Δ 6 desaturase
Stearidonic Acid
[C18:4, ω-3]
(STA)
elongase
Eicosatetraenoic Acid
[C20:4, ω-3]
(ETA)
Δ 5 desaturase
Δ 17 desaturase
Δ 9 elongase
Eicosatrienoic Acid
[C20:3, ω-3]
(ETrA)
Δ 8 desaturase

AMMONIUM TRANSPORTER PROMOTER FOR GENE EXPRESSION IN OLEAGINOUS YEAST

FIELD OF THE INVENTION

[0001] This invention is in the field of biotechnology. More specifically, this invention pertains to a promoter region isolated from *Yarrowia lipolytica* that is useful for gene expression in oleaginous yeast.

BACKGROUND OF THE INVENTION

[0002] Oleaginous yeast are defined as those organisms that are naturally capable of oil synthesis and accumulation, wherein oil accumulation ranges from at least about 25% up to about 80% of the cellular dry weight. Genera typically identified as oleaginous yeast include, but are not limited to: *Yarrowia, Candida, Rhodotorula, Rhodosporidium, Cryptococcus, Trichosporon* and *Lipomyces*. More specifically, illustrative oil-synthesizing yeast include: *Rhodosporidium toruloides, Lipomyces starkeyii, L. lipoferus, Candida revkaufi, C. pulcherrima, C. tropicalis, C. utilis, Trichosporon pullans, T. cutaneum, Rhodotorula glutinus, R. graminis* and *Yarrowia lipolytica* (formerly classified as *Candida lipolytica*).

[0003] The technology for growing oleaginous yeast with high oil content is well developed (for example, see EP 0 005 277B1; Ratledge, C., *Prog. Ind. Microbiol.* 16:119-206 (1982)). And, these organisms have been commercially used for a variety of purposes in the past. For example, various strains of *Yarrowa lipolytica* have historically been used for the manufacture and production of: isocitrate lyase; lipases; polyhydroxy-alkanoates; citric acid; erythritol; 2-oxoglutaric acid; γ-decalactone; γ-dodecalactone; and pyruvic acid. More recently, however, the natural abilities of oleaginous yeast have been enhanced by advances in genetic engineering, resulting in organisms capable of producing polyunsaturated fatty acids ("PUFAs"). Specifically, Picataggio et al. have demonstrated that *Y. lipolytica* can be engineered for production of ω-3 and ω-6 fatty acids, by introducing and expressing genes encoding the ω-3/ω-6 biosynthetic pathway (see WO 2004/101757 and co-pending U.S. patent application Ser. No. 60/624,812).

[0004] Recombinant production of any heterologous protein is generally accomplished by constructing an expression cassette in which the DNA coding for the protein of interest is placed under the control of appropriate regulatory sequences (i.e., promoters) suitable for the host cell. The expression cassette is then introduced into the host cell (usually by plasmid-mediated transformation or targeted integration into the host genome) and production of the heterologous protein is achieved by culturing the transformed host cell under conditions necessary for the proper function of the promoter contained within the expression cassette. Thus, the development of new host cells (e.g., oleaginous yeast) for recombinant production of proteins generally requires the availability of promoters that are suitable for controlling the expression of a protein of interest in the host cell.

[0005] A variety of strong promoters have been isolated from *Yarrowia lipolytica* that are useful for heterologous gene expression in yeast. For example, U.S. Pat. No. 4,937,189 and EP220864 (Davidow et al.) disclose the sequence of the XPR2 gene (which encodes an inducible alkaline extracellular protease) and upstream promoter region for use in expression of heterologous proteins. U.S. Pat. No. 6,265,185 (Muller et al.) describes promoters for the translation elongation factor EF1-α (TEF) protein and ribosomal protein S7 that are suitable for expression cloning in yeast and heterologous expression of proteins. These promoters were improved relative to the XPR2 promoter, when tested for yeast promoter activity on growth plates (Example 9, U.S. Pat. No. 6,265,185) and based on their activity in the pH range of 4-11. WO 2005/003310 describe regulatory sequences (e.g., promoters, introns) of the glyceraldehyde-3-phosphate dehydrogenase (gpd) and phosphoglycerate mutase (gpm) genes; WO 2005/049805 describes regulatory sequences (e.g., promoters, introns) of the fructose-bisphosphate aldolase (fba) gene; and, co-pending U.S. patent application Ser. No. 60/610,060 describes promoters of the glycerol-3-phosphate O-acyltransferase (gpat) gene. Similarly, Juretzek et al. (*Biotech. Bioprocess Eng.*, 5:320-326 (2000)) compares the glycerol-3-phosphate dehydrogenase (G3P), isocitrate lyase (ICL1), 3-oxo-acyl-CoA thiolase (POT1) and acyl-CoA oxidase (POX1, POX2 and POX5) promoters with respect to their regulation and activities during growth on different carbon sources.

[0006] Despite the utility of these known promoters, however, there is a need for new improved yeast promoters for metabolic engineering of yeast (oleaginous and non-oleaginous) and for controlling the expression of heterologous genes in yeast. Furthermore, possession of a suite of promoters that are regulatable under a variety of natural growth and induction conditions in yeast will play an important role in industrial settings, wherein it is desirable to express heterologous polypeptides in commercial quantities in said hosts for economical production of those polypeptides. Thus, it is an object of the present invention to provide such promoters that will be useful for gene expression in a variety of yeast cultures, and preferably in *Yarrowia* sp. cultures and other oleaginous yeast.

[0007] Applicants have solved the stated problem by identifying the YAT1 promoter from *Yarrowia lipolytica*, which is responsible for driving expression of the gene (yat1) encoding an ammonium transporter (YAT1). Advantageously, the promoter is useful for regulated expression of heterologous genes in *Yarrowia*, has improved activity with respect to the TEF promoter and is inducible under oleaginous conditions (i.e., nitrogen limitation).

SUMMARY OF THE INVENTION

[0008] The present invention provides methods for the regulated expression of a coding region of interest in a transformed yeast, using a promoter region of an ammonium transporter (yat1) gene. Accordingly the invention provides a method for the expression of a coding region of interest in a transformed yeast comprising:

[0009] a) providing a transformed yeast having a genetic construct comprising:

[0010] (i) a promoter region of a *Yarrowia* yat1 gene; and,

[0011] (ii) a coding region of interest expressible in the yeast cell;

[0012] wherein the promoter region is operably linked to the coding region of interest; and,

[0013] b) growing the transformed yeast of step (a) under conditions whereby the genetic construct of step (a) is expressed.

[0014] In another embodiment mutant yat1 promoters are provided exhibiting mutations that demonstrate between 20% and 400% increased promoter activity as compared with the wildtype promoter.

[0015] In an alternate embodiment the invention provides a method for the production of an ω-3 or an ω-6 fatty acid comprising:

[0016] a) providing a transformed oleaginous yeast comprising a genetic construct, comprising:

[0017] (i) a promoter region of a *Yarrowia* yat1 gene; and,

[0018] (ii) a coding region encoding at least one enzyme of the ω-3/ω-6 fatty acid biosynthetic pathway;

[0019] wherein the promoter region and coding region are operably linked;

[0020] b) culturing the transformed oleaginous yeast of step (a) under conditions whereby the at least one enzyme of the ω-3/ω-6 fatty acid biosynthetic pathway is expressed and a ω-3 or ω-6 fatty acid is produced; and,

[0021] c) optionally recovering the ω-3 or ω-6 fatty acid.

[0022] Preferred conditions suitable to induce expression of the genetic constructs of the invention are conditions of nitrogen-limitation; and, in a more preferred embodiment, the nitrogen-limiting conditions are those that induce oleaginy.

BRIEF DESCRIPTION OF THE DRAWINGS AND SEQUENCE DESCRIPTIONS

[0023] **FIG. 1** provides plasmid maps for the following: (A) pY5-30; (B) pDMW214; (C) pYGPAT-GUS; and (D) pYAT-GUS, respectively.

[0024] **FIG. 2** graphically represents the relationship between SEQ ID NOs:12, 15, 22, 23 and 24, each of which relates to glycerol-3-phosphate O-acyltransferase (GPAT) in *Y. lipolytica*.

[0025] **FIG. 3A** diagrams the development of *Yarrowia lipolytica* strain Y2034, producing up to 10% ARA in the total lipid fraction. **FIG. 3B** provides a plasmid map for pKUNF12T6E; and **FIG. 3C** provides a plasmid map for pDMW232.

[0026] **FIGS. 4A, 4B** and **4**C illustrate the relative promoter activities of YAT1, TEF, GPAT and FBAIN in *Y. lipolytica* grown in various media as determined by histochemical staining.

[0027] **FIG. 5** illustrates the ω-3/ω-6 fatty acid biosynthetic pathway.

[0028] The invention can be more fully understood from the following detailed description and the accompanying sequence descriptions, which form a part of this application.

[0029] The following sequences comply with 37 C.F.R. §1.821-1.825 ("Requirements for Patent Applications Containing Nucleotide Sequences and/or Amino Acid Sequence Disclosures—the Sequence Rules") and are consistent with World Intellectual Property Organization (WIPO) Standard ST.25 (1998) and the sequence listing requirements of the EPO and PCT (Rules 5.2 and 49.5(a-bis), and Section 208 and Annex C of the Administrative Instructions). The symbols and format used for nucleotide and amino acid sequence data comply with the rules set forth in 37 C.F.R. §1.822.

[0030] SEQ ID NOs:1-3, 10-12, 15, 23, 24, and 27-37 correspond to ORFs (i.e., encoding genes or proteins), promoters and plasmids, as identified in Table 1.

TABLE 1

Summary Of Nucleotide And Protein SEQ ID Numbers

Description	Nucleotide SEQ ID NO	Protein SEQ ID NO
Yarrowia lipolytica ammonium transporter (yat1) (GenBank Accession No. XM_504457)	1 (1461 bp)	2 (486 AA)
YAT1 promoter	3 (778 bp)	—
Plasmid pY5-30	10 (8953 bp)	—
FBAIN promoter	11 (973 bp)	—
Yarrowia lipolytica ORF YALI-CDS1055.1 (GenBank Accession No. CAG81570)	—	12 (727 AA)
Yarrowia lipolytica glycerol-3-phosphate O-acyltransferase (gpat) gene	15 (2184 bp)	—
Yarrowia lipolytica gpat gene: −1678 to +2181 region	23 (3862 bp)	—
GPAT promoter	24 (1130 bp)	—
Plasmid pKUNF12T6E	27 12,649 bp)	—
Synthetic elongase gene derived from *Mortierella alpina*, codon-optimized for expression in *Yarrowia lipolytica*	28 (957 bp)	29 (318 AA)
Synthetic Δ6 desaturase, derived from *Mortierella alpina*, codon-optimized for expression in *Yarrowia lipolytica*	30 (1374 bp)	31 (457 AA)
FBA promoter	32 (1001 bp)	—
Fusarium moniliforme Δ12 desaturase	33 (1434 bp)	34 (477 AA)
Synthetic elongase gene derived from *Thraustochytrium aureum*, codon-optimized for expression in *Yarrowia lipolytica*	35 (819 bp)	36 (272 AA)
Plasmid pDMW232	37 (10,945 bp)	—

[0031] SEQ ID NOs:4, 5 and 6 correspond to BD-Clontech's Creator Smart® cDNA library kit primers SMART IV oligonucleotide, CDSIII/3' PCR primer and 5'-PCR primer, respectively.

[0032] SEQ ID NO:7 corresponds to the M13 forward primer used for sequencing of *Y. lipolytica* cDNA.

[0033] SEQ ID NOs:8 and 9 correspond to primers 27203-F and 27203-R, respectively, used for amplifying the 5' upstream region of the yat1 gene.

[0034] SEQ ID NOs:13 and 14 correspond to the degenerate primers YGPAT-F and YGPAT-R, respectively, used for amplifying *Y. lipolytica* ORF YALI-CDS1055.1.

[0035] SEQ ID NOs:16 and 17 correspond to the Genome Walker adaptor used to isolate the GPAT promoter region by genome-walking.

[0036] SEQ ID NOs:18-21 correspond to the PCR primers used in genome-walking: Adaptor-1, YGPAT-5R-1, Nested Adaptor Primer 2 and YGPAT-5R-2, respectively.

[0037] SEQ ID NO:22 corresponds to a 1781 bp fragment contained within plasmid pEcoRV-G-5, the fragment containing a 1678 bp region upstream of the translation initiation codon 'ATG' of the gpat gene.

[0038] SEQ ID NOs:25 and 26 correspond to primers GPAT-5-1 and GPAT-5-2, respectively, used to amplify the GPAT promoter.

DETAILED DESCRIPTION OF THE INVENTION

[0039] In support of the invention described herein the following commonly owned patents, and patent applications are incorporated herein by reference in their entirety: 10/840,579 (filed May 6, 2004), 10/869,630 (filed Jun. 16, 2004), 10/987,548 (filed Nov. 12, 2004), U.S. Patent Provisional Application No. 60/610,060 (filed Sep. 15, 2004), and U.S. Patent Provisional Application No. 60/624,812 (filed Nov. 4, 2004).

[0040] In accordance with the subject invention, Applicants describe the characterization of a promoter from an oleaginous yeast, *Yarrowia lipolytica*. This promoter region, isolated upstream of the ammonium transporter (yat1) gene and inducible under oleaginous conditions (i.e., nitrogen limitation), is useful for genetic engineering in *Y. lipolytica* and other yeast for the production of heterologous polypeptides.

[0041] Preferred heterologous polypeptides of the present invention are those that are involved in the synthesis of microbial oils and particularly polyunsaturated fatty acids (PUFAs). PUFAs, or derivatives thereof, made by the methodology disclosed herein can be used in many applications. For example, the PUFAs can be used as dietary substitutes, or supplements, particularly infant formulas, for patients undergoing intravenous feeding or for preventing or treating malnutrition. Alternatively, the purified PUFAs (or derivatives thereof) may be incorporated into cooking oils, fats or margarines formulated so that in normal use the recipient would receive the desired amount for dietary supplementation. The PUFAs may also be incorporated into infant formulas, nutritional supplements or other food products, and may find use as anti-inflammatory or cholesterol lowering agents. Optionally, the compositions may be used for pharmaceutical use (human or veterinary). In this case, the PUFAs are generally administered orally but can be administered by any route by which they may be successfully absorbed, e.g., parenterally (e.g., subcutaneously, intramuscularly or intravenously), rectally, vaginally or topically (e.g., as a skin ointment or lotion).

[0042] Thus, the present invention advances the art by providing methods for the regulated expression of a coding region of interest in a transformed yeast comprising: a) providing a transformed yeast having a chimeric gene comprising (i) a promoter region of a *Yarrowia* yat1 gene; and, (ii) a coding region of interest expressible in the yeast, wherein the promoter region is operably linked to the coding region of interest; b) growing the transformed yeast of step (a) under conditions whereby the chimeric gene is expressed; and, c) optionally isolating the gene product from the cultivation medium. In preferred embodiments, the YAT1 promoter region comprises a sequence as set forth in SEQ ID NO:3.

Definitions

[0043] In this disclosure, a number of terms and abbreviations are used. The following definitions are provided.

[0044] "Ammonium transporter" is abbreviated YAT.

[0045] "Open reading frame" is abbreviated ORF.

[0046] "Polymerase chain reaction" is abbreviated PCR.

[0047] "Polyunsaturated fatty acid(s)" is abbreviated PUFA(s).

[0048] The term "oleaginous" refers to those organisms that tend to store their energy source in the form of lipid (Weete, In: Fungal Lipid Biochemistry, 2nd Ed., Plenum, 1980). Generally, the cellular PUFA content of these microorganisms follows a sigmoid curve, wherein the concentration of lipid increases until it reaches a maximum at the late logarithmic or early stationary growth phase and then gradually decreases during the late stationary and death phases (Yongmanitchai and Ward, *Appl. Environ. Microbiol.* 57:419-25 (1991)).

[0049] The term "oleaginous yeast" refers to those microorganisms classified as yeast that can accumulate at least 25% of their dry cell weight as oil. Examples of oleaginous yeast include (but are no means limited to) the following genera: *Yarrowia, Candida, Rhodotorula, Rhodosporidium, Cryptococcus, Trichosporon* and *Lipomyces*.

[0050] The term "ammonium transporter" refers to a family of transporters whose physiological role within a cell is to scavenge external ammonium for use as a nitrogen source (and, under some circumstances, to incorporate ammonium that leaks out of cells), based on studies of the MEP family of ammonium transporters in *Saccharomyces cerevisiae* (A. M. Marini et al., *EMBO J.*, 13(15):3456-3463 (1994); *Mol Cell Biol* 17(8):4282-93 (1997)). In general, the proteins are subject to nitrogen control and each has different kinetic properties, specificities and regulation; for example, the three *S. cerevisiae* isozymes are characterized as follows: MEP1, a low affinity/high capacity ammonia transporter (K_m, 5-10 μM); MEP2, a high affinity/low capacity ammonia transporter (K_m, 1-2 μM); and MEP3, a low affinity ammonia transporter (K_m, 1.4-2.1 mM). Other known members of this family include: the high affinity ammonium transporter (amt1), the ammonium and methylammonium transport system, the putative ammonium transporter amtB and nrgA.

[0051] As used herein, the term "YAT1" refers to an ammonium transporter enzyme (TC 2.A.49) encoded by the yat1 gene and isolated from *Yarrowia lipolytica* (GenBank Accession No. XM_504457). The sequence disclosed therein (/locus_tag="YALI0E27203g") was identified as a hypothetical protein having similarity "to sp|P41948 *Saccharomyces cerevisiae* YNL142w MEP2, hypothetical start". The *Y. lipolytica* yat1 gene is presented herein as SEQ ID NO:1, while the corresponding YAT1 protein is provided herein as SEQ ID NO:2.

[0052] The term "YAT1 promoter" or "YAT1 promoter region" refers to the 5' upstream untranslated region in front of the 'ATG' translation initiation codon of yat1 and that is necessary for expression. An example of a suitable YAT1 promoter region is provided as SEQ ID NO:3, but this is not intended to be limiting in nature. One skilled in the art will recognize that since the exact boundaries of the YAT1 promoter sequence have not been completely defined, DNA fragments of increased or diminished length may have identical promoter activity.

[0053] The term "GPAT" refers to a glycerol-3-phosphate O-acyltransferase enzyme (E.C. 2.3.1.15) encoded by the gpat gene and which converts acyl-CoA and sn-glycerol 3-phosphate to CoA and 1-acyl-sn-glycerol 3-phosphate (the first step of phospholipid biosynthesis). The term "GPAT promoter" or "GPAT promoter region" refers to the 5' upstream untranslated region in front of the 'ATG' translation initiation codon of gpat and that is necessary for expression. One example of a suitable GPAT promoter region is provided as SEQ ID NO:24, but this is not intended to be limiting in nature (see co-pending U.S. patent application Ser. No. 60/610,060).

[0054] The term "FBA1" refers to a fructose-bisphosphate aldolase enzyme (E.C. 4.1.2.13) encoded by the fba1 gene and which converts D-fructose 1,6-bisphosphate into glycerone phosphate and D-glyceraldehyde 3-phosphate. The term "FBAIN promoter" or "FBAIN promoter region" refers to the 5' upstream untranslated region in front of the 'ATG' translation initiation codon of fba1 and that is necessary for expression, plus a portion of 5' coding region comprising an intron of the fba1 gene. An example of a suitable FBAIN promoter region is provided as SEQ ID NO:11, but this is not intended to be limiting in nature (see WO 2005/049805).

[0055] The term "promoter activity" will refer to an assessment of the transcriptional efficiency of a promoter. This may, for instance, be determined directly by measurement of the amount of mRNA transcription from the promoter (e.g., by Northern blotting or primer extension methods) or indirectly by measuring the amount of gene product expressed from the promoter, as foe example by histochemical means.

[0056] The term "conditions of nitrogen-limitation" refers to a medium having a low concentration of nitrogen, wherein the nitrogen may be supplied from an inorganic (e.g., $(NH_4)_2SO_4$) or organic (e.g., urea or glutamate) source, or a medium having no nitrogen source. Although one skilled in the art will be able to determine an appropriate low concentration of nitrogen suitable to induce the YAT1 promoter of the present invention, in one embodiment a preferred medium would be one characterized as having a high carbon to nitrogen (i.e., C:N) ratio and about 0.1% or less ammonium sulfate or other suitable ammonium salts.

[0057] As used herein, an "isolated nucleic acid molecule" is a polymer of RNA or DNA that is single- or double-stranded, optionally containing synthetic, non-natural or altered nucleotide bases. An isolated nucleic acid molecule in the form of a polymer of DNA may be comprised of one or more segments of cDNA, genomic DNA or synthetic DNA.

[0058] The term "genetic construct" is a non-limiting term meaning any contiguous series of nucleic acids capable of being expressed in a host organism. A genetic construct may include but is not limited to an open reading frame, an open reading frame operably linked to regulatory sequences, or a wildtype or mutant gene. Genetic constructs may encode poylpeptides or be nucleic acid fragments or molecules that are oriented for antisense expression.

[0059] A nucleic acid molecule is "hybridizable" to another nucleic acid molecule, such as a cDNA, genomic DNA, or RNA molecule, 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 in Sambrook, J., Fritsch, E. F. and Maniatis, T. *Molecular Cloning: A Laboratory Manual,* 2nd ed., Cold Spring Harbor Laboratory: Cold Spring Harbor, N.Y. (1989), particularly Chapter 11 and Table 11.1 therein (entirely incorporated herein by reference). The conditions of temperature and ionic strength determine the "stringency" of the hybridization. Stringency conditions can be adjusted to screen for moderately similar fragments (such as homologous sequences from distantly related organisms), to highly similar fragments (such as genes that duplicate functional enzymes from closely related organisms). Post-hybridization washes determine stringency conditions. One set of preferred conditions uses a series of washes starting with 6×SSC, 0.5% SDS at room temperature for 15 min, then repeated with 2×SSC, 0.5% SDS at 45° C. for 30 min, and then repeated twice with 0.2×SSC, 0.5% SDS at 50° C. for 30 min. A more preferred set of stringent conditions uses higher temperatures in which the washes are identical to those above except for the temperature of the final two 30 min washes in 0.2×SSC, 0.5% SDS was increased to 60° C. Another preferred set of highly stringent conditions uses two final washes in 0.1×SSC, 0.1% SDS at 65 ° C. An additional set of stringent conditions include hybridization at 0.1×SSC, 0.1% SDS, 65° C. and washed with 2×SSC, 0.1% SDS followed by 0.1×SSC, 0.1% SDS, for example.

[0060] Hybridization requires that the two nucleic acids contain complementary sequences, although depending on the stringency of the hybridization, mismatches between bases are possible. The appropriate stringency for hybridizing nucleic acids depends on the length of the nucleic acids and the degree of complementation, variables well known in the art. The greater the degree of similarity or homology between two nucleotide sequences, the greater the value of Tm for hybrids of nucleic acids having those sequences. The relative stability (corresponding to higher Tm) of nucleic acid hybridizations decreases in the following order: RNA:RNA, DNA:RNA, DNA:DNA. For hybrids of greater than 100 nucleotides in length, equations for calculating Tm have been derived (see Sambrook et al., supra, 9.50-9.51). For hybridizations with shorter nucleic acids, i.e., oligonucleotides, the position of mismatches becomes more important, and the length of the oligonucleotide determines its specificity (see Sambrook et al., supra, 11.7-11.8). In one embodiment the length for a hybridizable nucleic acid is at least about 10 nucleotides. Preferably a minimum length for a hybridizable nucleic acid is at least about 15 nucleotides; more preferably at least about 20 nucleotides; and most preferably the length is at least about 30 nucleotides. Furthermore, the skilled artisan will recognize that the temperature and wash solution salt concentration may be adjusted as necessary according to factors such as length of the probe.

[0061] A "substantial portion" of an amino acid or nucleotide sequence is that portion comprising enough of the amino acid sequence of a polypeptide or the nucleotide sequence of a gene to putatively identify that polypeptide or gene, either by manual evaluation of the sequence by one skilled in the art, or by computer-automated sequence comparison and identification using algorithms such as BLAST (Basic Local Alignment Search Tool; Altschul, S. F., et al., *J. Mol. Biol.* 215:403-410 (1993)). In general, a sequence of ten or more contiguous amino acids or thirty or more nucleotides is necessary in order to identify putatively a polypeptide or nucleic acid sequence as homologous to a known protein or gene. Moreover, with respect to nucleotide sequences, gene-specific oligonucleotide probes comprising 20-30 contiguous nucleotides may be used in sequence-dependent methods of gene identification (e.g., Southern hybridization) and isolation (e.g., in situ hybridization of bacterial colonies or bacteriophage plaques). In addition, short oligonucleotides of 12-15 bases may be used as amplification primers in PCR in order to obtain a particular nucleic acid molecule comprising the primers. Accordingly, a "substantial portion" of a nucleotide sequence comprises enough of the sequence to specifically identify and/or isolate a nucleic acid molecule comprising the sequence.

[0062] The instant specification teaches nucleotide sequences encoding a particular microbial promoter region. The skilled artisan, having the benefit of the sequences as reported herein, may now use all or a substantial portion of the disclosed sequences for purposes known to those skilled in this art. Accordingly, the instant invention comprises the complete sequences as reported in the accompanying Sequence Listing, as well as substantial portions of those sequences as defined above.

[0063] The term "oligonucleotide" refers to a nucleic acid, generally of at least 18 nucleotides, that is hybridizable to a genomic DNA molecule, a cDNA molecule, or an mRNA molecule. In one embodiment, a labeled oligonucleotide can be used as a "probe" to detect the presence of a nucleic acid according to the invention. Thus, the term "probe" refers to a single-stranded nucleic acid molecule that can base pair with a complementary single-stranded target nucleic acid to form a double-stranded molecule. The term "label" will refer to any conventional molecule which can be readily attached to mRNA or DNA and which can produce a detectable signal, the intensity of which indicates the relative amount of hybridization of the labeled probe to the DNA fragment.

[0064] The term "complementary" is used to describe the relationship between nucleotide bases that are capable of hybridizing to one another. For example, with respect to DNA, adenosine is complementary to thymine and cytosine is complementary to guanine. Accordingly, the instant invention also includes isolated nucleic acid molecules that are complementary to the complete sequences as reported in the accompanying Sequence Listing, as well as those substantially similar nucleic acid sequences.

[0065] The term "percent identity", as known in the art, is a relationship between two or more polypeptide sequences or two or more polynucleotide sequences, as determined by comparing the sequences. In the art, "identity" also means the degree of sequence relatedness between polypeptide or polynucleotide sequences, as the case may be, as determined by the match between strings of such sequences. "Identity" and "similarity" can be readily calculated by known methods, including but not limited to those described in: 1.) *Computational Molecular Biology* (Lesk, A. M., Ed.) Oxford University: NY (1988); 2.) *Biocomputing: Informatics and Genome Projects* (Smith, D. W., Ed.) Academic: NY (1993); 3.) *Computer Analysis of Sequence Data, Part I* (Griffin, A. M., and Griffin, H. G., Eds.) Humania: NJ (1994); 4.) *Sequence Analysis in Molecular Biology* (von Heinje, G., Ed.) Academic (1987); and 5.) *Sequence Analysis Primer* (Gribskov, M. and Devereux, J., Eds.) Stockton: NY (1991). Preferred methods to determine identity are designed to give the best match between the sequences tested. Methods to determine identity and similarity are codified in publicly available computer programs. Sequence alignments and percent identity calculations may be performed using the Megalign program of the LASERGENE bioinformatics computing suite (DNASTAR Inc., Madison, Wis.). Multiple alignment of the sequences is performed using the Clustal method of alignment (Higgins and Sharp, *CABIOS.* 5:151-153 (1989)) with default parameters (GAP PENALTY=10, GAP LENGTH PENALTY=10). Default parameters for pairwise alignments using the Clustal method are: KTUPLE 1, GAP PENALTY=3, WINDOW=5 and DIAGONALS SAVED=5.

[0066] Suitable promoter regions (isolated polynucleotides of the present invention) encode promoter regions that are at least about 70% identical, preferably at least about 75% identical, and more preferably at least about 80% identical to the nucleotide sequences reported herein. Preferred nucleic acid molecules are about 85% identical to the nucleotide sequences reported herein, more preferred nucleic acid molecules are at least about 90% identical, and most preferred are nucleic acid molecules at least about 95% identical to the nucleotide sequences reported herein. Suitable promoter regions not only have the above homologies but typically are at least 50 nucleotides in length, more preferably at least 100 nucleotides in length, more preferably at least 250 nucleotides in length, and more preferably at least 500 nucleotides in length.

[0067] "Codon degeneracy" refers to the nature in the genetic code permitting variation of the nucleotide sequence without affecting the amino acid sequence of an encoded polypeptide. The skilled artisan is well aware of the "codon-bias" exhibited by a specific host cell in usage of nucleotide codons to specify a given amino acid. Therefore, when synthesizing a gene for improved expression in a host cell, it is desirable to design the gene such that its frequency of codon usage approaches the frequency of preferred codon usage of the host cell.

[0068] "Chemically synthesized", as related to a sequence of DNA, means that the component nucleotides were assembled in vitro. Manual chemical synthesis of DNA may be accomplished using well-established procedures; or, automated chemical synthesis can be performed using one of a number of commercially available machines. "Synthetic genes" can be assembled from oligonucleotide building blocks that are chemically synthesized using procedures known to those skilled in the art. These building blocks are ligated and annealed to form gene segments that are then enzymatically assembled to construct the entire gene. Accordingly, the genes can be tailored for optimal gene expression based on optimization of nucleotide sequence to reflect the codon bias of the host cell. The skilled artisan

appreciates the likelihood of successful gene expression if codon usage is biased towards those codons favored by the host. Determination of preferred codons can be based on a survey of genes derived from the host cell, where sequence information is available.

[0069] "Gene" refers to a nucleic acid molecule that expresses a specific protein, including regulatory sequences preceding (5' non-coding sequences) and following (3' non-coding sequences) the coding sequence. "Native gene" refers to a gene as found in nature with its own regulatory sequences. "Chimeric gene" refers to any gene that is not a native gene, comprising regulatory and coding sequences that are not found together in nature. Accordingly, a chimeric gene may comprise regulatory sequences and coding sequences that are derived from different sources, or regulatory sequences and coding sequences derived from the same source, but arranged in a manner different than that found in nature. Chimeric genes of the present invention will typically comprise a YAT1 promoter region operably linked to a coding region of interest. "Endogenous gene" refers to a native gene in its natural location in the genome of an organism. A "foreign" gene refers to a gene not normally found in the host organism, but that is introduced into the host organism by gene transfer. Foreign genes can comprise native genes inserted into a non-native organism, or chimeric genes. A "transgene" is a gene that has been introduced into the genome by a transformation procedure. A "codon-optimized gene" is a gene having its frequency of codon usage designed to mimic the frequency of preferred codon usage of the host cell.

[0070] "Coding sequence" refers to a DNA sequence that codes for a specific amino acid sequence.

[0071] "Suitable regulatory sequences" refer to transcriptional and translational nucleotide sequences located upstream (5' non-coding sequences), within, or downstream (3' non-coding sequences) of a coding sequence, and which influence the transcription, RNA processing or stability, or translation of the associated coding sequence. Regulatory sequences may include promoters, translation leader sequences, introns, polyadenylation recognition sequences, RNA processing sites, effector binding sites and stem-loop structures.

[0072] "Promoter" refers to a DNA sequence capable of controlling the expression of a coding sequence or functional RNA. In general, a coding sequence is located 3' to a promoter sequence. Promoters may be derived in their entirety from a native gene, or be composed of different elements derived from different promoters found in nature, or even comprise synthetic DNA segments. It is understood by those skilled in the art that different promoters may direct the expression of a gene in different tissues or cell types, or at different stages of development, or in response to different environmental or physiological conditions. Promoters that cause a gene to be expressed in most cell types at most times are commonly referred to as "constitutive promoters". It is further recognized that since in most cases the exact boundaries of regulatory sequences have not been completely defined, DNA fragments of different lengths may have identical promoter activity.

[0073] The term "mutant promoter" is defined herein as a promoter having a nucleotide sequence comprising a substitution, deletion, and/or insertion of one or more nucleotides relative to the parent promoter, wherein the mutant promoter has more or less promoter activity than the corresponding parent promoter. The term "mutant promoter" will encompass natural variants and in vitro generated variants obtained using methods well known in the art (e.g., classical mutagenesis, site-directed mutagenesis and "DNA shuffling").

[0074] The term "3'non-coding sequences" or "transcription terminator" refers to DNA sequences located downstream of a coding sequence. This includes polyadenylation recognition sequences and other sequences encoding regulatory signals capable of affecting mRNA processing or gene expression. The polyadenylation signal is usually characterized by affecting the addition of polyadenylic acid tracts to the 3' end of the mRNA precursor. The 3' region can influence the transcription, RNA processing or stability, or translation of the associated coding sequence.

[0075] "RNA transcript" refers to the product resulting from RNA polymerase-catalyzed transcription of a DNA sequence. When the RNA transcript is a perfect complementary copy of the DNA sequence, it is referred to as the primary transcript or it may be a RNA sequence derived from post-transcriptional processing of the primary transcript and is referred to as the mature RNA. "Messenger RNA" or "mRNA" refers to the RNA that is without introns and that can be translated into protein by the cell. "cDNA" refers to a double-stranded DNA that is complementary to, and derived from, mRNA. "Sense" RNA refers to RNA transcript that includes the mRNA and so can be translated into protein by the cell. "Antisense RNA" refers to a RNA transcript that is complementary to all or part of a target primary transcript or mRNA and that blocks the expression of a target gene (U.S. Pat. No. 5,107,065; WO 99/28508). The complementarity of an antisense RNA may be with any part of the specific gene transcript, i.e., at the 5' non-coding sequence, 3' non-coding sequence, or the coding sequence. "Functional RNA" refers to antisense RNA, ribozyme RNA, or other RNA that is not translated and yet has an effect on cellular processes.

[0076] The term "operably linked" refers to the association of nucleic acid sequences on a single nucleic acid molecule so that the function of one is affected by the other. For example, a promoter is operably linked with a coding sequence when it is capable of affecting the expression of that coding sequence (i.e., the coding sequence is under the transcriptional control of the promoter). Coding sequences can be operably linked to regulatory sequences in sense or antisense orientation.

[0077] The term "expression", as used herein, refers to the transcription and stable accumulation of sense (mRNA) or antisense RNA derived from a coding sequence. Expression may also refer to translation of mRNA into a polypeptide.

[0078] "Transformation" refers to the transfer of a nucleic acid molecule into a host organism, resulting in genetically stable inheritance. The nucleic acid molecule may be a plasmid that replicates autonomously, for example; or, it may integrate into the genome of the host organism. Host organisms containing the transformed nucleic acid molecules are referred to as "transgenic" or "recombinant" or "transformed" organisms.

[0079] The terms "plasmid", "vector" and "cassette" refer to an extra chromosomal element often carrying genes that

are not part of the central metabolism of the cell, and usually in the form of circular double-stranded DNA fragments. Such elements may be autonomously replicating sequences, genome integrating sequences, phage or nucleotide sequences, linear or circular, of a single- or double-stranded DNA or RNA, derived from any source, in which a number of nucleotide sequences have been joined or recombined into a unique construction which is capable of introducing a promoter fragment and DNA sequence for a selected gene product along with appropriate 3' untranslated sequence into a cell. "Expression cassette" refers to a specific vector containing a foreign gene and having elements in addition to the foreign gene that allow for enhanced expression of that gene in a foreign host.

[0080] The term "sequence analysis software" refers to any computer algorithm or software program that is useful for the analysis of nucleotide or amino acid sequences. "Sequence analysis software" may be commercially available or independently developed. Typical sequence analysis software will include, but is not limited to: 1.) the GCG suite of programs (Wisconsin Package Version 9.0, Genetics Computer Group (GCG), Madison, Wis.); 2.) BLASTP, BLASTN, BLASTX (Altschul et al., *J. Mol. Biol.* 215:403-410 (1990)); 3.) DNASTAR (DNASTAR, Inc. Madison, Wis.); and 4.) the FASTA program incorporating the Smith-Waterman algorithm (W. R. Pearson, *Comput. Methods Genome Res.*, [Proc. Int. Symp.] (1994), Meeting Date 1992, 111-20. Suhai, Sandor, Ed. Plenum: New York, N.Y.). Within the context of this application it will be understood that where sequence analysis software is used for analysis, that the results of the analysis will be based on the "default values" of the program referenced, unless otherwise specified. As used herein "default values" will mean any set of values or parameters that originally load with the software when first initialized.

[0081] Standard recombinant DNA and molecular cloning techniques used herein are well known in the art and are described by Sambrook, J., Fritsch, E. F. and Maniatis, T., *Molecular Cloning: A Laboratory Manual,* 2nd ed., Cold Spring Harbor Laboratory: Cold Spring Harbor, N.Y. (1989) (hereinafter "Maniatis"); by Silhavy, T. J., Bennan, M. L. and Enquist, L. W., *Experiments with Gene Fusions*, Cold Spring Harbor Laboratory: Cold Spring Harbor, N.Y. (1984); and by Ausubel, F. M. et al., *Current Protocols in Molecular Biology*, published by Greene Publishing Assoc. and Wiley-Interscience (1987).

Identification of YAT1 and the YAT1 Promoter in *Yarrowia lipolytica*

[0082] Ammonium transporters are a family of transporters whose physiological role within a cell is to scavenge external ammonium for use as a nitrogen source (and, under some circumstances, to incorporate ammonium that leaks out of cells). In general, the proteins are subject to nitrogen control and each have different kinetic properties, specificities and regulation; for example, the three *S. cerevisiae* MEP1, MEP2, and MEP3 isozymes have been extensively characterized by A. M. Marini et al. (*EMBO J.,* 13(15): 3456-3463 (1994); *Mol Cell Biol* 17(8):4282-93 (1997)) as a low affinity/high capacity ammonia transporter, a high affinity/low capacity ammonia transporter, and a low affinity ammonia transporter, respectively.

[0083] A cDNA library was constructed from *Yarrowia lipolytica* cells grown in an oleaginous medium containing no nitrogen source for 4 hrs (i.e., oleaginous conditions that promoter oil production). The relative abundance of each species of cDNA was then examined to identify those genes that were highly expressed under the oleaginous conditions. One of the genes that appeared multiple times (16/9984 or 0.16%) was GenBank Accession No. XM_504457, locus_tag="YALI0E27203g". This gene was annotated as a hypothetical protein having similarity "to sp|P41948 *Saccharomyces cerevisiae* YNL142w MEP2 [the high affinity, low capacity ammonia transporter; GenBank Accession No. X83608], hypothetical start".

[0084] Based on sequence comparison to the MEP1, MEP2 and MEP3 genes (supra), the Applicants hypothesized that the *Y. lipolytica* ORF likely encoded an ammonium transporter and designated the gene as the "yat1" gene (SEQ ID NOs:1 and 2). The "YAT1 promoter" was identified as the 5' upstream untranslated region in front of the 'ATG' translation initiation codon of yat1 and that is necessary for expression. According to the invention here, this putative YAT1 promoter region will be useful for driving expression of any suitable coding region of interest in a transformed yeast cell. In general, a promoter useful in an oleaginous yeast should meet the following criteria:

[0085] 1.) Strength. A strong yeast promoter is a necessary premise for a high expression level, and the low copy number of the ars18 (Fournier, P. et al., *Yeast* 7:25-36 (1991)) based expression vectors or chimeric genes integrated into the genome makes this demand even more important when *Y. lipolytica* is used as the host organism.

[0086] 2.) Activity in a medium suitable for expression of the coding region of interest, and high enzymatic activity of that coding region of interest.

[0087] 3.) pH Tolerance. If the coding region of interest is known to be produced only in e.g., an acidic environment, then the promoter operably linked to said coding region of interest must function at the appropriate pH. pH tolerance is of course limited by the tolerance of the host organism.

[0088] 4.) Inducibility. A tightly regulated yeast promoter makes it possible to separate the growth stage from the expression stage, thereby enabling expression of products that are known to inhibit cell growth.

[0089] 5.) Activity in the stationary phase of growth in oleaginous yeast hosts for accumulation of PUFAs.

[0090] Additionally, it is preferable for novel yeast promoters to possess differences in activity with respect to the known *Yarrowia lipolytica* TEF (U.S. Pat. No. 6,265,185), XPR2 (U.S. Pat. No. 4,937,189; EP220864; EP832258), GPD, GPM (U.S. Ser. No. 10/869,630), FBA, FBAIN (U.S. Ser. No. 10/987,548), and GPAT (U.S. Provisional Patent Application No. 60/610,060) promoters and/or the G3P, ICL1, POT1, POX1, POX2 and POX5 promoters (Juretzek et al., *Biotech. Bioprocess Eng.,* 5:320-326 (2000)).

[0091] A comparative study of the TEF, FBAIN and GPAT promoters and the YAT1 promoter of the instant invention is provided in Examples 8 and 9. As shown in Table 7 of Example 9, the yeast promoter of the present invention has improved activity compared to the TEF promoter (i.e., 1.3:1 nmoles of 4-methylumbelliferone per minute per mg of total

protein) under conditions where nitrogen is not limiting (i.e., minimal medium containing ammonium sulfate as the nitrogen source). Furthermore, under conditions of nitrogen limitation, the YAT1 promoter is significantly induced such that the activity is ~20-28 times greater than that of TEF; thus, when grown in such medium, the activity of the YAT1 promoter is comparable to that of the strong FBAIN promoter. Based on these results, the Applicants characterize the YAT1 promoter as the first promoter identified within *Yarrowia* that is inducible under oleaginous conditions (i.e., nitrogen limitation).

[0092] An example of a suitable YAT1 promoter region is provided as SEQ ID NO:3 (comprising the −778 to −1 region of the *Y. lipolytica* yat1 gene (wherein the 'A' position of the 'ATG' translation initiation codon is designated as +1)), but this is not intended to be limiting in nature. One skilled in the art will recognize that since the exact boundaries of the YAT1 promoter sequence have not been completely defined, DNA fragments of increased or diminished length may have identical promoter activity. For example, in an alternate embodiment, the YAT1 promoter will comprise nucleotides −500 to −1 of SEQ ID NO:3, thereby permitting relatively strong promoter activity; in another embodiment, the −100 to −1 region of SEQ ID NO:3 should be sufficient for basal activity of the promoter. Likewise, the promoter region of the invention may comprise additional nucleotides to those specified above. For example, the promoter sequences of the invention may be constructed on the basis of the −1000 to +1 region of the yat1 gene (e.g., nucleotide bases 3,222,879 to 3,223,879 of GenBank Accession No. CR382131, comprising the complete nucleotide sequence of chromosome E of strain CLIB99 of *Y. lipolytica*).

[0093] In alternate embodiments mutant promoters may be constructed, wherein the DNA sequence of the promoter has one or more nucleotide substitutions (i.e., deletions, insertions, substitutions, or addition of one or more nucleotides in the sequence) which do not effect (in particular impair) the yeast promoter activity. Regions that can be modified without significantly affecting the yeast promoter activity can be identified by deletion studies. A mutant promoter of the present invention has at least about 20%, preferably at least about 40%, more preferably at least about 60%, more preferably at least about 80%, more preferably at least about 90%, more preferably at least about 100%, more preferably at least about 200%, more preferably at least about 300% and most preferably at least about 400% of the promoter activity of the YAT1 promoter region described herein as SEQ ID NO:3.

[0094] Methods for mutagenesis are well known in the art and suitable for the generation of mutant promoters. For example, in vitro mutagenesis and selection, PCR based random mutagenesis, site-directed mutagenesis or other means can be employed to obtain mutations of the naturally occurring promoter of the instant invention (wherein such mutations may include deletions, insertions and point mutations, or combinations thereof). This would permit production of a putative promoter having a more desirable level of promoter activity in the host cell. Or, if desired, the regions of a nucleotide of interest important for promoter activity can be determined through routine mutagenesis, expression of the resulting mutant promoters and determination of their activities. An overview of these techniques is described in U.S. Ser. No. 10/869630. All such mutant promoters that are derived from the instant YAT1 promoter described herein are within the scope of the present invention.

[0095] Promoter activity is typically measured against the activity of the wild type promoter under similar conditions. Promoter activity is generally measured as a function of gene expression and may be determined in a variety of ways including gene expression profiling, measurement of the level of protein expression by SDS gel or other means, or the measurement of reporter activity where reporter gene fusions have been created.

Isolation of Homologs of YAT1 and the YAT1 Putative Promoter Region

[0096] It will be appreciated by a person of skill in the art that the promoter regions of the present invention have homologs in a variety of yeast species; and, the use of the promoters for regulated, heterologous gene expression are not limited to those promoters derived from *Yarrowia lipolytica*, but extend to homologs in other yeast species. For example, the invention encompasses homologs derived from oleaginous genera including, but not limited to: *Yarrowia, Candida, Rhodotorula, Rhodosporidium, Cryptococcus, Trichosporon* and *Lipomyces*; examples of preferred species within these genera include: *Rhodosporidium toruloides, Lipomyces starkeyii, L. lipoferus, Candida revkaufi, C. pulcherrima, C. tropicalis, C. utilis, Trichosporon pullans, T. cutaneum, Rhodotorula glutinus* and *R. graminis.*

[0097] Homology typically is measured using sequence analysis software, wherein the term "sequence analysis software" refers to any computer algorithm or software program (commercially available or independently developed) that is useful for the analysis of nucleotide or amino acid sequences. In general, such computer software matches similar sequences by assigning degrees of homology to various substitutions, deletions and other modifications.

[0098] As is well known in the art, isolation of homologous promoter regions using sequence-dependent protocols is readily possible using various techniques; and, these techniques can rely on either the direct identification of a promoter having homology to the YAT1 promoter of the invention or the indirect identification of a promoter by initial identification of gene having significant homology to the yat1 gene and then analysis of the 5' upstream sequence of the homologous gene. Examples of sequence-dependent protocols include, but are not limited to: 1.) methods of nucleic acid hybridization; 2.) methods of DNA and RNA amplification, as exemplified by various uses of nucleic acid amplification technologies [e.g., polymerase chain reaction (PCR), Mullis et al., U.S. 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)]; and 3.) methods of library construction and screening by complementation.

[0099] For example, putative promoter regions to those of the instant invention could be isolated by using all or a portion of the YAT1 nucleic acid molecules (e.g., corresponding to the promoter, gene or a combination thereof) as DNA hybridization probes to screen libraries from any desired microbe using methodology well known to those skilled in the art. Specific oligonucleotide probes based upon the instant nucleic acid sequences can be designed and

synthesized by methods known in the art (Maniatis, supra). Moreover, the entire sequences can be used directly to synthesize DNA probes by methods known to the skilled artisan (e.g., random primers DNA labeling, nick translation, or end-labeling techniques), or RNA probes using available in vitro transcription systems. In addition, specific primers can be designed and used to amplify a part of (or full-length of) the instant sequences. The resulting amplification products can be labeled directly during amplification reactions or labeled after amplification reactions, and used as probes to isolate full-length DNA fragments under conditions of appropriate stringency.

[0100] 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 (Thein and Wallace, "The use of oligonucleotides as specific hybridization probes in the Diagnosis of Genetic Disorders", in *Human Genetic Diseases: A Practical Approach*, K. E. Davis (Ed.), (1986) pp 33-50 IRL: Herndon, Va.; and Rychlik, W., In *Methods in Molecular Biology*, White, B. A. (Ed.), (1993) Vol. 15, pp 31-39, PCR Protocols: Current Methods and Applications. Humania: Totowa, N.J.).

[0101] Generally two short segments of the instant sequences may be used in PCR protocols to amplify longer nucleic acid molecules encoding homologous polynucleotides from DNA or RNA. The PCR may also be performed on a library of cloned nucleic acid molecules wherein the sequence of one primer is derived from the instant nucleic acid molecules, 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.

[0102] Alternatively, the instant sequences may be employed as hybridization reagents for the identification of homologs. The basic components of a nucleic acid hybridization test include a probe, a sample suspected of containing the nucleotide sequence of interest, and a specific hybridization method. Probes of the present invention are typically single-stranded nucleic acid sequences that are complementary to the nucleic acid sequences to be detected. Probes are "hybridizable" to the nucleic acid sequence to be detected. The probe length can vary from 5 bases to tens of thousands of bases, and will depend upon the specific test to be done. Typically a probe length of about 15 bases to about 30 bases is suitable. Only part of the probe molecule need be complementary to the nucleic acid sequence to be detected. In addition, the complementarity between the probe and the target sequence need not be perfect. Hybridization does occur between imperfectly complementary molecules with the result that a certain fraction of the bases in the hybridized region are not paired with the proper complementary base.

[0103] Hybridization methods are well defined. Typically the probe and sample must be mixed under conditions that will permit nucleic acid hybridization. This involves contacting the probe and sample in the presence of an inorganic or organic salt under the proper concentration and temperature conditions. The probe and sample nucleic acids must be in contact for a long enough time that any possible hybridization between the probe and sample nucleic acid may occur. The concentration of probe or target in the mixture will determine the time necessary for hybridization to occur. The higher the probe or target concentration, the shorter the hybridization incubation time needed. Optionally, a chaotropic agent may be added. The chaotropic agent stabilizes nucleic acids by inhibiting nuclease activity. Furthermore, the chaotropic agent allows sensitive and stringent hybridization of short oligonucleotide probes at room temperature (Van Ness and Chen, *Nucl. Acids Res.* 19:5143-5151 (1991)). Suitable chaotropic agents include guanidinium chloride, guanidinium thiocyanate, sodium thiocyanate, lithium tetrachloroacetate, sodium perchlorate, rubidium tetrachloroacetate, potassium iodide and cesium trifluoroacetate, among others. Typically, the chaotropic agent will be present at a final concentration of about 3 M. If desired, one can add formamide to the hybridization mixture, typically 30-50% (v/v).

[0104] Various hybridization solutions can be employed. Typically, these comprise from about 20 to 60% volume, preferably 30%, of a polar organic solvent. A common hybridization solution employs about 30-50% v/v formamide, about 0.15 to 1 M sodium chloride, about 0.05 to 0.1 M buffers (e.g., sodium citrate, Tris-HCl, PIPES or HEPES (pH range about 6-9)), about 0.05 to 0.2% detergent (e.g., sodium dodecylsulfate), or between 0.5-20 mM EDTA, FICOLL (Pharmacia Inc.) (about 300-500 kdal), polyvinylpyrrolidone (about 250-500 kdal) and serum albumin. Also included in the typical hybridization solution will be unlabeled carrier nucleic acids from about 0.1 to 5 mg/mL, fragmented nucleic DNA (e.g., calf thymus or salmon sperm DNA, or yeast RNA), and optionally from about 0.5 to 2% wt/vol glycine. Other additives may also be included, such as volume exclusion agents that include a variety of polar water-soluble or swellable agents (e.g., polyethylene glycol), anionic polymers (e.g., polyacrylate or polymethylacrylate) and anionic saccharidic polymers (e.g., dextran sulfate).

Recombinant Expression in Yeast

[0105] Initiation control regions or promoter regions that are useful to drive expression of a coding gene of interest in the desired host cell are selected from those derived from the upstream portion of the yat1 gene (SEQ ID NO:1). The promoter regions may be identified from the upstream sequences of yat1 and its homologs and isolated according to common methods (Maniatis, supra). Once a promoter region is identified and isolated (e.g., SEQ ID NO:3), it may be operably linked to a coding region of interest to be expressed in a suitable expression vector. These chimeric genes may then be expressed in natural host cells and heterologous host cells, particularly in the cells of oleaginous yeast hosts. Thus, one aspect of the present invention provides a recombinant expression vector comprising a yeast promoter of the invention.

[0106] In a further aspect, the invention provides a method of expressing a coding region of interest in a transformed yeast, wherein a transformed yeast is provided having a chimeric gene comprising: (i) a promoter region of a *Yarrowia* yat1 gene; and, (ii) a coding region of interest expressible in the yeast, wherein the promoter region is operably linked to the coding region of interest; and the transformed yeast is grown under conditions wherein the chimeric gene is expressed. The polypeptide so produced can optionally be recovered from the culture.

[0107] Microbial expression systems and expression vectors are well known to those skilled in the art. Any of these could be used to construct chimeric genes comprising a promoter region derived from the yat1 gene for production of any specific coding region of interest suitable for regulated expression in a desirable yeast host cell. These chimeric genes could then be introduced into appropriate microorganisms by integration via transformation to provide expression of the enzymes upon induction. Alternatively, the promoters can be cloned into a plasmid that is capable of transforming and replicating itself in the preferred yeast. The coding region of interest to be expressed can then be cloned downstream from the promoter. Once the recombinant host is established, regulated gene expression can be accomplished by growing the cells under suitable conditions (infra).

[0108] Suitable Coding Regions of Interest

[0109] Useful chimeric genes will include the promoter region of the yat1 gene as defined herein or a mutant promoter thereof, operably linked to a suitable coding region of interest to be expressed in a preferred host cell.

[0110] Coding regions of interest to be expressed in the recombinant yeast host may be either endogenous to the host or heterologous and must be compatible with the host organism. Genes encoding proteins of commercial value are particularly suitable for expression. For example, suitable coding regions of interest may include (but are not limited to) those encoding viral, bacterial, fungal, plant, insect, or vertebrate coding regions of interest, including mammalian polypeptides. Further, these coding regions of interest may be, for example, structural proteins, enzymes (e.g., oxidoreductases, transferases, hydrolyases, lyases, isomerases, ligases), or peptides. A non-limiting list includes genes encoding enzymes such as acyltransferases, aminopeptidases, amylases, carbohydrases, carboxypeptidases, catalyases, cellulases, chitinases, cutinases, cyclodextrin glycosyltransferases, deoxyribonucleases, esterases, α-galactosidases, β-glucanases, β-galactosidases, glucoamylases, α-glucosidases, β-glucosidases, invertases, laccases, lipases, mannosidases, mutanases, oxidases, pectinolytic enzymes, peroxidases, phospholipases, phytases, polyphenoloxidases, proteolytic enzymes, ribonucleases, transglutaminases or xylanases.

[0111] Preferred in the present invention in some embodiments are coding regions of the enzymes involved in the production of microbial oils, including ω-6 and ω-3 fatty acids. These coding regions include desaturases and elongases (e.g., see WO 2004/101757 for a partial review of available genes in GenBank and/or the patent literature and considerations for choosing a specific polypeptide having desaturase or elongase activity).

Components of Vectors/DNA Cassettes

[0112] Vectors or DNA cassettes useful for the transformation of suitable host cells are well known in the art. The specific choice of sequences present in the construct is dependent upon the desired expression products (supra), the nature of the host cell, and the proposed means of separating transformed cells versus non-transformed cells. Typically, however, the vector or cassette contains sequences directing transcription and translation of the relevant gene(s), a selectable marker, and sequences allowing autonomous replication or chromosomal integration. Suitable vectors comprise a region 5' of the gene that controls transcriptional initiation and a region 3' of the DNA fragment that controls transcriptional termination. It is most preferred when both control regions are derived from genes from the transformed host cell, although it is to be understood that such control regions need not be derived from the genes native to the specific species chosen as a production host.

[0113] Nucleotide sequences surrounding the translational initiation codon 'ATG' have been found to affect expression in yeast cells. If the desired polypeptide is poorly expressed in yeast, the nucleotide sequences of exogenous genes can be modified to include an efficient yeast translation initiation sequence motif to obtain optimal gene expression. For expression in yeast, this can be done by site-directed mutagenesis of an inefficiently expressed gene to include the favored translation initiation motif.

[0114] The termination region can be derived from the 3' region of the gene from which the initiation region was obtained or from a different gene. A large number of termination regions are known and function satisfactorily in a variety of hosts (when utilized both in the same and different genera and species from where they were derived). The termination region usually is selected more as a matter of convenience rather than because of any particular property. Preferably, the termination region is derived from a yeast gene, particularly *Saccharomyces, Schizosaccharomyces, Candida, Yarrowia* or *Kluyveromyces*. The 3'-regions of mammalian genes encoding γ-interferon and α-2 interferon are also known to function in yeast. Termination control regions may also be derived from various genes native to the preferred hosts. Optionally, a termination site may be unnecessary; however, it is most preferred if included.

[0115] As one of skill in the art is aware, merely inserting a chimeric gene into a cloning vector does not ensure that it will be successfully expressed at the level needed. In response to needs for high expression rates, many specialized expression vectors have been created by manipulating a number of different genetic elements that control aspects of transcription, translation, protein stability, oxygen limitation and secretion from the host cell. More specifically, some of the molecular features that have been manipulated to control gene expression include: 1.) the nature of the relevant transcriptional promoter and terminator sequences; 2.) whether the gene is plasmid-borne or integrated into the genome of the host cell and the number of copies of the cloned gene [e.g., additional copies a particular coding region of interest (operably linked to the promoter of the instant invention) may be introduced into the host to increase expression]; 3.) the final cellular location of the synthesized foreign protein; 4.) the efficiency of translation in the host organism; 5.) the intrinsic stability of the cloned gene protein within the host cell [e.g., expression of the coding region of interest can be increased by removing/deleting destabilizing sequences from either the mRNA or the encoded protein or by adding stabilizing sequences to the mRNA (U.S. Pat. No. 4,910,141)]; and 6.) the codon usage within the cloned gene, such that its frequency approaches the frequency of preferred codon usage of the host cell [e.g. translational efficiency of the encoded mRNAs can be increased by replacement of codons in the native gene with those for optimal gene expression in the selected host microorganism, to thereby substantially enhance the expres-

sion of the foreign gene encoding the polypeptide]. Each of these types of modifications are encompassed in the present invention, as means to further optimize expression of a chimeric gene comprising a promoter region of the yat1 gene as defined herein or a mutant promoter thereof, operably linked to a suitable coding region of interest.

[0116] Transformation of Yeast Cells

[0117] Once an appropriate chimeric gene has been constructed that is suitable for regulated expression in a yeast cell, it is placed in a plasmid vector capable of autonomous replication in a host cell or it is directly integrated into the genome of the host cell. Integration of expression cassettes can occur randomly within the host genome or can be targeted through the use of constructs containing regions of homology with the host genome sufficient to target recombination with the host locus. Where constructs are targeted to an endogenous locus, all or some of the transcriptional and translational regulatory regions can be provided by the endogenous locus.

[0118] Where two or more genes are expressed from separate replicating vectors, it is desirable that each vector has a different means of selection and should lack homology to the other constructs to maintain stable expression and prevent reassortment of elements among constructs. Judicious choice of regulatory regions, selection means and method of propagation of the introduced construct can be experimentally determined so that all introduced genes are expressed at the necessary levels to provide for synthesis of the desired products.

[0119] Constructs comprising a coding region of interest may be introduced into a host cell by any standard technique. These techniques include transformation (e.g., lithium acetate transformation [*Methods in Enzymology,* 194:186-187 (1991)]), protoplast fusion, biolistic impact, electroporation, microinjection, or any other method that introduces the gene of interest into the host cell. More specific teachings applicable for oleaginous yeast (i.e., *Yarrowia lipolytica*) include U.S. Pat. No. 4,880,741 and No. 5,071,764 and Chen, D. C. et al. (*Appl Microbiol Biotechnol.* 48(2):232-235 (1997)).

[0120] For convenience, a host cell that has been manipulated by any method to take up a DNA sequence (e.g., an expression cassette) will be referred to as "transformed" or "recombinant" herein. The transformed host will have at least one copy of the expression construct and may have two or more, depending upon whether the gene is integrated into the genome, amplified, or is present on an extrachromosomal element having multiple copy numbers. The transformed host cell can be identified by various selection techniques, as described in commonly owned U.S. Ser. No. 10/869,630.

Preferred Hosts

[0121] Preferred host cells for expression of coding regions of interest operably linked to the instant promoter fragments herein are yeast cells (where oleaginous yeast are most preferred where the desired use is for the production of microbial oils, infra). Oleaginous yeast are naturally capable of oil synthesis and accumulation, wherein the oil can comprise greater than about 25% of the cellular dry weight, more preferably greater than about 30% of the cellular dry weight, and most preferably greater than about 40% of the cellular dry weight. Genera typically identified as oleaginous yeast include, but are not limited to: *Yarrowia, Candida, Rhodotorula, Rhodosporidium, Cryptococcus, Trichosporon* and *Lipomyces*. More specifically, illustrative oil-synthesizing yeast include: *Rhodosporidium toruloides, Lipomyces starkeyii, L. lipoferus, Candida revkaufi, C. pulcherrima, C. tropicalis, C. utilis, Trichosporon pullans, T. cutaneum, Rhodotorula glutinus, R. graminis* and *Yarrowia lipolytica* (formerly classified as *Candida lipolytica*).

[0122] Most preferred is the oleaginous yeast *Yarrowia lipolytica*; and, in a further embodiment, most preferred are the *Y. lipolytica* strains designated as ATCC #20362, ATCC #8862, ATCC #18944, ATCC #76982 and/or LGAM S(7)1 (Papanikolaou S., and Aggelis G., *Bioresour. Technol.* 82(1):43-9 (2002)). The *Y. lipolytica* strain designated as ATCC #20362 was the particular strain from which the YAT1 promoter was isolated therefrom.

Industrial Production using Transformed Yeast Expressing a Suitable Coding Region of Interest

[0123] In general, media conditions that may be optimized for regulated expression of a particular coding region of interest include the type and amount of carbon source, the type and amount of nitrogen source, the carbon-to-nitrogen ratio, the oxygen level, growth temperature, pH, length of the biomass production phase and the time of cell harvest. Microorganisms of interest, such as oleaginous yeast, are grown in complex media (e.g., yeast extract-peptone-dextrose broth (YPD)) or a defined minimal media that lacks a component necessary for growth and thereby forces selection of the desired expression cassettes (e.g., Yeast Nitrogen Base (DIFCO Laboratories, Detroit, Mich.)).

[0124] Fermentation media in the present invention must contain a suitable carbon source. Suitable carbon sources may include, but are not limited to: monosaccharides (e.g., glucose, fructose), disaccharides (e.g., lactose, sucrose), oligosaccharides, polysaccharides (e.g., starch, cellulose or mixtures thereof), sugar alcohols (e.g., glycerol) or mixtures from renewable feedstocks (e.g., cheese whey permeate, cornsteep liquor, sugar beet molasses, barley malt). Additionally, carbon sources may include alkanes, fatty acids, esters of fatty acids, monoglycerides, diglycerides, triglycerides, phospholipids and various commercial sources of fatty acids including vegetable oils (e.g., soybean oil) and animal fats. Additionally, the carbon source may include one-carbon sources (e.g., carbon dioxide, methanol, formaldehyde, formate, carbon-containing amines) for which metabolic conversion into key biochemical intermediates has been demonstrated. Hence it is contemplated that the source of carbon utilized in the present invention may encompass a wide variety of carbon-containing sources and will only be limited by the choice of the host organism. Although all of the above mentioned carbon sources and mixtures thereof are expected to be suitable in the present invention, preferred carbon sources are sugars and/or fatty acids. Most preferred is glucose and/or fatty acids containing between 10-22 carbons.

[0125] Nitrogen may be supplied from an inorganic (e.g., $(NH_4)_2SO_4$) or organic source (e.g., urea or glutamate). Although the YAT1 promoter is active in media containing nitrogen (e.g., up to about 0.5% ammonium sulfate), the activity of the promoter increases when the host cell is grown in nitrogen-limiting conditions (e.g., in medium

containing very low levels of ammonium, or lacking ammonium). Thus, a preferred medium would be one that contains less than about 0.1% ammonium sulfate, or other suitable ammonium salts.

[0126] In a more preferred embodiment, the YAT1 promoter is induced when the host cell is grown in media with a high carbon to nitrogen (i.e., C:N) ratio, such as a high glucose medium containing about 8-12% glucose, and about 0.1% or less ammonium sulfate. These conditions are also sufficient to induce oleaginy in those yeast that are oleaginous (e.g., *Yarrowia lipolytica*).

[0127] In addition to appropriate carbon and nitrogen sources, the fermentation media must also contain suitable minerals, salts, cofactors, buffers, vitamins, and other components known to those skilled in the art suitable for the growth of the microorganism.

[0128] Preferred growth media in the present invention are common commercially prepared media, such as minimal media made with Yeast Nitrogen Base (DIFCO Laboratories, Detroit, Mich.). Other defined or synthetic growth media may also be used and the appropriate medium for growth of the particular microorganism will be known by one skilled in the art of microbiology or fermentation science. A suitable pH range for the fermentation is typically between about pH 4.0 to pH 8.0, wherein pH 5.5 to pH 7.0 is preferred as the range for the initial growth conditions. The fermentation may be conducted under aerobic or anaerobic conditions, wherein microaerobic conditions are preferred.

[0129] Host cells comprising a suitable coding region of interest operably linked to the promoters of the present invention may be cultured using methods known in the art. For example, the cell may be cultivated by shake flask cultivation, small-scale or large-scale fermentation in laboratory or industrial fermentors performed in a suitable medium and under conditions allowing regulated expression of the coding region of interest. Furthermore, where commercial production of a product that relies on the instant genetic chimera is desired, a variety of culture methodologies may be applied. For example, large-scale production of a specific gene product over-expressed from a recombinant host may be produced by a batch, fed-batch or continuous fermentation process, as is well known in the art (see, e.g., Thomas D. Brock in *Biotechnologq: A Textbook of Industrial Microbiology,* 2nd ed., (1989) Sinauer Associates: Sunderland, Mass.; or Deshpande, Mukund V., *Appl. Biochem. Biotechnol.,* 36:227 (1992), each herein incorporated by reference).

DESCRIPTION OF PREFERRED EMBODIMENTS

[0130] Although the promoters of the present invention will be suitable for regulated expression of any suitable coding region of interest in an oleaginous yeast, in a preferred embodiment the promoters will be utilized in the development of an oleaginous yeast that accumulates high levels of oils enriched in PUFAs. Toward this end, it is necessary to introduce and express e.g., desaturases and elongases that allow for the synthesis and accumulation of ω-3 and/or ω-6 fatty acids.

[0131] The term "fatty acids" refers to long chain aliphatic acids (alkanoic acids) of varying chain lengths, from about C_{12} to C_{22} (although both longer and shorter chain-length acids are known). The predominant chain lengths are between C_{16} and C_{22}. The structure of a fatty acid is represented by a simple notation system of "X:Y", where X is the total number of carbon (C) atoms and Y is the number of double bonds. Additional details concerning the differentiation between "saturated fatty acids" versus "unsaturated fatty acids", "monounsaturated fatty acids" versus "polyunsaturated fatty acids" (or "PUFAs"), and "ω-6 fatty acids" (ω-6 or n-6) versus "ω-3 fatty acids" (ω-3 or n-3) are provided in commonly owned U.S. Ser. No. 10/840,579.

[0132] Nomenclature used to describe PUFAs in the present disclosure is shown below in Table 2. In the column titled "Shorthand Notation", the omega-reference system is used to indicate the number of carbons, the number of double bonds and the position of the double bond closest to the omega carbon, counting from the omega carbon (which is numbered 1 for this purpose). The remainder of the Table summarizes the common names of ω-3 and c)-6 fatty acids, the abbreviations that will be used throughout the remainder of the specification, and each compounds' chemical name.

TABLE 2

Nomenclature Of Polyunsaturated Fatty Acids

Common Name	Abbreviation	Chemical Name	Shorthand Notation
Linoleic	LA	cis-9,12-octadecadienoic	18:2 ω-6
γ-Linoleic	GLA	cis-6,9,12-octadecatrienoic	18:3 ω-6
Eicosadienoic	EDA	cis-11,14-eicosadienoic	20:2 ω-6
Dihomo-γ-Linoleic	DGLA	cis-8,11,14-eicosatrienoic	20:3 ω-6
Arachidonic	ARA	cis-5,8,11,14-eicosatetraenoic	20:4 ω-6
α-Linolenic	ALA	cis-9,12,15-octadecatrienoic	18:3 ω-3
Stearidonic	STA	cis-6,9,12,15-octadecatetraenoic	18:4 ω-3
Eicosatrienoic	ETrA	cis-11,14,17-eicosatrienoic	20:3 ω-3
Eicosatetraenoic	ETA	cis-8,11,14,17-eicosatetraenoic	20:4 ω-3
Eicosapentaenoic	EPA	cis-5,8,11,14,17-eicosapentaenoic	20:5 ω-3
Docosapentaenoic	DPA	cis-7,10,13,16,19-docosapentaenoic	22:5 ω-3
Docosahexaenoic	DHA	cis-4,7,10,13,16,19-docosahexaenoic	22:6 ω-3

Microbial Biosynthesis of Omega-3 and Omega-6 Fatty Acids

[0133] The process of de novo synthesis of palmitate (16:0) in oleaginous microorganisms is described in commonly owned U.S. Ser. No. 10/840,579. This fatty acid is the precursor of longer-chain saturated and unsaturated fatty acid derivates, which are formed through the action of elongases and desaturases. For example, palmitate is converted to its unsaturated derivative [palmitoleic acid (16:1)] by the action of a Δ9 desaturase; similarly, palmitate is elongated to form stearic acid (18:0), which can be converted to its unsaturated derivative by a Δ9 desaturase to thereby yield oleic (18:1) acid.

[0134] The metabolic process that converts LA to GLA, DGLA and ARA (the ω-6 pathway) and ALA to STA, ETA,

EPA and DHA (the ω-3 pathway) is well described in the literature and is schematically depicted in **FIG. 5** (see also commonly owned U.S. Ser. No. 10/840,579 and U.S. Ser. No. 10/869,630). Simplistically, this process involves elongation of the carbon chain through the addition of carbon atoms and desaturation of the molecule through the addition of double bonds, via a series of special desaturation and elongation enzymes present in the endoplasmic reticulim membrane (and hereinafter referred to as "PUFA biosynthetic pathway enzymes").

[0135] More specifically, "PUFA biosynthetic pathway enzymes" or "ω-3/ω-6 biosynthetic pathway enzymes" will refer to any of the following enzymes (and genes which encode said enzymes) associated with the biosynthesis of a PUFA, including: a Δ4 desaturase, a Δ5 desaturase, a Δ6 desaturase, a Δ12 desaturase, a Δ15 desaturase, a Δ17 desaturase, a Δ9 desaturase, a Δ8 desaturase and/or an elongase(s). For further clarity within the present disclosure, the term "desaturase" refers to a polypeptide that can desaturate one or more fatty acids to produce a mono- or polyunsaturated fatty acid or precursor of interest. Thus, despite use of the omega-reference system to refer to specific fatty acids, it is more convenient to indicate the activity of a desaturase by counting from the carboxyl end of the source using the delta-system. For example, a Δ17 desaturase will desaturate a fatty acid between the 17th and 18th carbon atom numbered from the carboxyl-terminal end of the molecule and can, for example, catalyze the conversion of ARA to EPA and/or DGLA to ETA. In contrast, the term "elongase" refers to a polypeptide that can elongate a fatty acid carbon chain to produce a mono- or polyunsaturated fatty acid that is 2 carbons longer than the fatty acid source that the elongase acts upon. This process of elongation occurs in a multi-step mechanism in association with fatty acid synthase, whereby CoA is the acyl carrier (Lassner et al., *The Plant Cell* 8:281-292 (1996)).

[0136] As will be understood by one skilled in the art, the particular functionalities required to be introduced into a host organism for production of a particular PUFA final product will depend on the host cell (and its native PUFA profile and/or desaturase/elongase profile), the availability of substrate and the desired end product(s). As shown in **FIG. 5**, LA, GLA, EDA, DGLA, ARA, ALA, STA, ETrA, ETA, EPA, DPA and DHA may all be produced in oleaginous yeast, by introducing various combinations of the following PUFA enzyme functionalities: a Δ4 desaturase, a Δ5 desaturase, a Δ6 desaturase, a Δ12 desaturase, a Δ15 desaturase, a Δ17 desaturase, a Δ9 desaturase, a Δ8 desaturase and/or an elongase(s). One skilled in the art will be able to identify various candidate genes encoding each of the above enzymes, according to publicly available literature (e.g., GenBank), the patent literature, and experimental analysis of microorganisms having the ability to produce PUFAs. Thus, a variety of desaturases and elongases are suitable as coding regions of interest in the present invention. These coding regions of interest could be operably linked to the YAT1 promoters of the present invention or mutant promoters thereof, and used as chimeric genes for expression of various ω-6 and ω-3 fatty acids, using techniques well known to those skilled in the art (e.g., see WO 2004/101757 and co-pending U.S. patent application Ser. No. 60/624,812). As such, the invention provides a method for the production of ω-3 and/or ω-6 fatty acids comprising:

[0137] a) providing a transformed oleaginous yeast comprising a genetic construct, said construct comprising:

[0138] 1) a promoter region of a *Yarrowia* yat1 gene; and,

[0139] 2) a coding region of interest encoding at least one enzyme of the ω-3/ω-6 fatty acid biosynthetic pathway;

[0140] wherein the promoter region and coding region are operably linked;

[0141] b) culturing the transformed oleaginous yeast of step (a) under conditions whereby the at least one enzyme of the ω-3/ω-6 fatty acid biosynthetic pathway is expressed and a ω-3 or ω-6 fatty acid is produced; and,

[0142] c) optionally recovering the ω-3 or ω-6 fatty acid.

[0143] In preferred embodiments, the nucleic acid sequence of the promoter region is selected from the group consisting of: SEQ ID NO:3, and subsequences and mutant promoters thereof; the coding region of interest is any desaturase or elongase suitable for regulated expression in the oleaginous yeast for the production of ω-3 or ω-6 fatty acids; and, the transformed oleaginous yeast is cultured under conditions that induce oleaginy.

[0144] More specifically, for production of the greatest and the most economical yield of PUFAs, the transformed oleaginous yeast host cell is grown under conditions that optimize desaturase and elongase activities by optimizing expression of the chimeric genes of the present invention, wherein these chimeric genes comprise a promoter region of a yat1 gene and a coding region of interest encoding a PUFA biosynthetic pathway enzyme. Typically, accumulation of high levels of PUFAs in oleaginous yeast cells requires a two-stage process, since the metabolic state must be "balanced" between growth and synthesis/storage of fats. Thus, most preferably, a two-stage fermentation process is necessary for the production of PUFAs in oleaginous yeast. In this approach, the first stage of the fermentation is dedicated to the generation and accumulation of cell mass and is characterized by rapid cell growth and cell division. In the second stage of the fermentation, it is preferable to establish conditions of nitrogen deprivation in the culture to promote high levels of lipid accumulation.

[0145] The effect of this nitrogen deprivation is two-fold. First, the nitrogen deprivation reduces the effective concentration of AMP in the cells, thereby reducing the activity of the NAD-dependent isocitrate dehydrogenase of mitochondria. When this occurs, citric acid will accumulate, thus forming abundant pools of acetyl-CoA in the cytoplasm and priming fatty acid synthesis. Secondly, the nitrogen deprivation induces the YAT1 promoter of the invention, thereby promoting expression of any chimeric genes of the invention comprising the YAT1 promoter and a coding region of interest encoding an enzyme of the ω-3/ω-6 fatty acid biosynthetic pathway. Thus, this second phase of the 2-stage fermentation is characterized by the cessation of cell division followed by the synthesis of fatty acids and accumulation of oil.

[0146] Although cells are typically grown at about 30° C., some studies have shown increased synthesis of unsaturated fatty acids at lower temperatures (Yongmanitchai and Ward, *Appl. Environ. Microbiol.* 57:419-25 (1991)). Based on process economics, this temperature shift should likely occur after the first phase of the two-stage fermentation, when the bulk of the organisms' growth has occurred.

[0147] Additionally, particular attention is given to several metal ions (e.g., Mn^{+2}, Co^{+2}, Zn^{+2}, Mg^{+2}) that promote synthesis of lipids and PUFAs in the fermentation media (Nakahara, T. et al. *Ind. Appl. Single Cell Oils*, D. J. Kyle and R. Colin, eds. pp 61-97 (1992)).

Purification of PUFAs

[0148] The PUFAs produced in a host microorganism as described herein may be found as free fatty acids or in esterified forms such as acylglycerols, phospholipids, sulfolipids or glycolipids, and may be extracted from the host cell through a variety of means well-known in the art. One review of extraction techniques, quality analysis and acceptability standards for yeast lipids is that of Z. Jacobs (*Critical Reviews in Biotechnology* 12(5/6):463-491 (1992)). A brief review of downstream processing is also available by A. Singh and O. Ward (*Adv. Appl. Microbiol.* 45:271-312 (1997)).

[0149] In general, means for the purification of fatty acids (including PUFAs) may include extraction with organic solvents, sonication, supercritical fluid extraction (e.g., using carbon dioxide), saponification and physical means such as presses, or combinations thereof example of which are contained in commonly owned U.S. Ser. No. 10/840,579.

EXAMPLES

[0150] The present invention is further defined in the following Examples. It should be understood that these Examples, while indicating preferred embodiments of the invention, are given by way of illustration only. From the above discussion and these Examples, one skilled in the art can ascertain the essential characteristics of this invention, and without departing from the spirit and scope thereof, can make various changes and modifications of the invention to adapt it to various usages and conditions.

General Methods

[0151] Standard recombinant DNA and molecular cloning techniques used in the Examples are well known in the art and are described by: 1.) Sambrook, J., Fritsch, E. F. and Maniatis, T. *Molecular Cloning: A Laboratory Manual*; Cold Spring Harbor Laboratory: Cold Spring Harbor, N.Y. (1989) (hereinafter "Maniatis"); 2.) T. J. Silhavy, M. L. Bennan, and L. W. Enquist, Experiments with Gene Fusions; Cold Spring Harbor Laboratory: Cold Spring Harbor, N.Y. (1984); and 3.) Ausubel, F. M. et al., Current Protocols in Molecular Biology, published by Greene Publishing Assoc. and Wiley-Interscience (1987).

[0152] Materials and methods suitable for the maintenance and growth of microbial cultures are well known in the art. Techniques suitable for use in the following examples may be found as set out in *Manual of Methods for General Bacteriology* (Phillipp Gerhardt, R. G. E. Murray, Ralph N. Costilow, Eugene W. Nester, Willis A. Wood, Noel R. Krieg and G. Briggs Phillips, Eds), American Society for Microbiology: Washington, D.C. (1994)); or by Thomas D. Brock in *Biotechnology: A Textbook of Industrial Microbiology,* 2nd ed., Sinauer Associates: Sunderland, Mass. (1989). All reagents, restriction enzymes and materials used for the growth and maintenance of microbial cells were obtained from Aldrich Chemicals (Milwaukee, Wis.), DIFCO Laboratories (Detroit, Mich.), GIBCO/BRL (Gaithersburg, Md.) or Sigma Chemical Company (St. Louis, Mo.), unless otherwise specified.

[0153] General molecular cloning was performed according to standard methods (Sambrook et al., supra). Oligonucleotides were synthesized by Sigma-Genosys (Spring, Tex.). When polymerase chain reaction (PCR) was involved in subcloning, the constructs were sequenced to confirm that no errors had been introduced to the sequence. PCR products were cloned into Promega's pGEM-T-easy vector (Madison, Wis.).

[0154] Manipulations of genetic sequences were accomplished using the suite of programs available from the Genetics Computer Group Inc. (Wisconsin Package Version 9.0, Genetics Computer Group (GCG), Madison, Wis.). The GCG program "Pileup" was used with the gap creation default value of 12, and the gap extension default value of 4. The GCG "Gap" or "Besffit" programs were used with the default gap creation penalty of 50 and the default gap extension penalty of 3. Unless otherwise stated, in all other cases GCG program default parameters were used.

[0155] The meaning of abbreviations is as follows: "sec" means second(s), "min" means minute(s), "h" means hour(s), "d" means day(s), "μL" means microliter(s), "mL" means milliliter(s), "L" means liter(s), "μM" means micromolar, "mM" means millimolar, "M" means molar, "mmol" means millimole(s), "μmole" mean micromole(s), "g" means gram(s), "μg" means microgram(s), "ng" means nanogram(s), "U" means unit(s), "bp" means base pair(s) and "kB" means kilobase(s).

Transformation and Cultivation of *Yarrowia Lipolytica*

[0156] *Y. lipolytica* strain ATCC #20362 was purchased from the American Type Culture Collection (Rockville, Md.). Strains were usually grown at 28° C. on YPD agar (1% yeast extract, 2% bactopeptone, 2% glucose, 2% agar) or in YPD liquid medium (2% bacto-yeast extract, 3% bactopeptone, 2% glucose).

[0157] Transformation of *Y. lipolytica* was performed according to the method of Chen, D. C. et al. (*Appl. Microbiol Biotechnol.* 48(2):232-235 (1997)), unless otherwise noted. Briefly, *Yarrowia* was streaked onto a YPD plate and grown at 30° C. for approximately 18 hr. Several large loopfuls of cells were scraped from the plate and resuspended in 1 mL of transformation buffer containing: 2.25 mL of 50% PEG, average MW 3350; 0.125 mL of 2 M Li acetate, pH 6.0; 0.125 mL of 2 M DTT; and 50 μg sheared salmon sperm DNA. Then, approximately 500 ng of linearized plasmid DNA was incubated in 100 μl of resuspended cells, and maintained at 39° C. for 1 hr with vortex mixing at 15 min intervals. The cells were plated onto selection media plates and maintained at 30° C. for 2 to 3 days.

[0158] For selection of transformants, minimal medium ("MM") was generally used; the composition of MM is as follows: 0.17% yeast nitrogen base (DIFCO Laboratories, Detroit, Mich.) without ammonium sulfate or amino acids,

2% glucose, 0.1% proline, pH 6.1 and 20 g/L agar). Supplements of uracil or leucine were added as appropriate to a final concentration of 0.01% (thereby producing "MMU" selection media or "MMLe" selection media).

[0159] Alternatively, transformants were selected on 5-fluoroorotic acid ("FOA"; also 5-fluorouracil-6-carboxylic acid monohydrate) selection media, comprising: 0.17% yeast nitrogen base (DIFCO Laboratories) without ammonium sulfate or amino acids, 2% glucose, 0.1% proline, 75 mg/L uracil, 75 mg/L uridine, 900 mg/L FOA (Zymo Research Corp., Orange, Calif.) and 20 g/L agar.

[0160] "SD" media comprises: 0.67% yeast nitrogen base with ammonium sulfate, without amino acids and 2% glucose. And finally, to promote conditions of oleaginy, High Glucose Media ("HGM") was prepared as follows: 14 g/L KH_2PO_4, 4 g/L K_2HPO_4, 2 g/L $MgSO_4.7H_2O$, 80 g/L glucose (pH 6.5).

Fatty Acid Analysis of *Yarrowia Lipolytica*

[0161] For fatty acid analysis, cells were collected by centrifugation and lipids were extracted as described in Bligh, E. G. & Dyer, W. J. (*Can. J. Biochem. Physiol.* 37:911-917 (1959)). Fatty acid methyl esters were prepared by transesterification of the lipid extract with sodium methoxide (Roughan, G., and Nishida I. *Arch Biochem Biophys.* 276(1):3846 (1990)) and subsequently analyzed with a Hewlett-Packard 6890 GC fitted with a 30-m×0.25 mm (i.d.) HP-INNOWAX (Hewlett-Packard) column. The oven temperature was from 170° C. (25 min hold) to 185° C. at 3.5° C./min.

[0162] For direct base transesterification, *Yarrowia* culture (3 mL) was harvested, washed once in distilled water, and dried under vacuum in a Speed-Vac for 5-10 min. Sodium methoxide (100 μl of 1%) was added to the sample, and then the sample was vortexed and rocked for 20 min. After adding 3 drops of 1 M NaCl and 400 μl hexane, the sample was vortexed and spun. The upper layer was removed and analyzed by GC as described above.

Example 1

Construction and Sequencing of a *Yarrowia Lipolytica* cDNA Library

[0163] The present Example describes the construction of a cDNA library of *Y. lipolytica*, grown under oleaginous conditions. More specifically, *Y. lipolytica* cells accumulate considerable amounts of oil when switched to a medium with a high carbon to nitrogen ratio (>80). A cDNA library was constructed from cells grown in an oleaginous medium containing no nitrogen source for 4 hrs, using the BD-Clontech Creator Smart® cDNA library kit (Mississauga, ON, Canada) according to the manufacturer's protocol.

[0164] Synthesis of the cDNA library first required growth of *Y. lipolytica* strain ATCC #20362 in 30 mL YPD medium overnight at 30° C. with shaking. Cells were diluted in two cultures of 100 mL fresh YPD to OD_{600}=0.4, then grown until OD_{600}=1.5 (Perkin-Elmer Lambda 20 UV/VIS Spectrophotometer). Cells from each culture were pelleted twice by centrifugation at 3750 rpm in a Beckman GH3.8 rotor for 5 min and washed with sterile water, then resuspended in 100 mL HGM. Cells were then shaken at 30° C. for 4 hrs. Each 100 mL culture was divided into three equal portions and pelleted by centrifugation at 3750 rpm in a Beckman GH3.8 rotor for 5 min.

[0165] Total RNA was extracted from the pellets using the Qiagen Rneasy Midi Kit. Cells were resusupended in 600 μl buffer RLT from the kit, with β-mercaptoethanol added at a concentration of 10 μl/mL. Resuspended cells were transferred to six 2 mL screw cap tubes each containing 0.6 mL of 0.5 mm glass beads. The cells were homogenized at the HOMOGENIZE-setting on a Biospec (Bartlesville, Okla.) mini bead beater for 2 min. The tubes were briefly spun to settle the beads. Liquid was transferred to 4 fresh 2 mL microfuge tubes and 600 μl of the RLT/BME mix was added to the beads. The bead tubes were vortexed and all liquid above the beads was transferred to the fresh 2 mL tubes. The fresh tubes were spun for 2 min in a microfuge to pellet cells debris, and the supernatent from each set of three tubes that came from a single culture were pooled in a 10 mL screwcap centrifuge tube. The procedure was then completed following the manufacturer's protocol. The two total RNA samples were then combined and mRNA was isolated from the combined sample using the Qiagen Oligotex Midi kit, following the manufacturer's protocol. Purified poly(A)+RNA was obtained with a concentration of 35.7 ng/μl.

[0166] cDNA was generated, using the LD-PCR method specified by BD-Clontech and 0.1 μg of polyA(+) RNA sample. Specifically, for 1st strand cDNA synthesis, 3 μl of the poly(A)+RNA sample was mixed with 1 μl of SMART IV oligo nucleotide (SEQ ID NO:4) and 1 μl of CDSIII/3' PCR primer (SEQ ID NO:5). The mixture was heated at 72° C. for 2 min and cooled on ice for 2 min. To the tube was added the following: 2 μl 1st strand buffer, 1 μl 20 mM DTT, 1 μl 10 mM dNTP mix and 1 μl Powerscript reverse transcriptase. The mixture was incubated at 42° C. for 1 hr and cooled on ice.

[0167] The 1st strand cDNA synthesis mixture was used as template for the PCR reaction. Specifically, the reaction mixture contained the following: 2 μl of the 1st strand cDNA mixture, 2 μl 5'-PCR primer (SEQ ID NO:6), 2 μl CDSIII/3'-PCR primer (SEQ ID NO:5), 80 μl water, 10 μl 10× Advantage 2 PCR buffer, 2 μl 50×dNTP mix and 2 μl 50× Advantage 2 polymerase mix. The thermocycler conditions were set for 95° C. for 20 sec, followed by 14 cycles of 95° C. for 5 sec and 68° C. for 6 min on a GenAmp 9600 instrument. PCR product was quantitated by agarose gel electrophoresis and ethidium bromide staining.

[0168] Seventy-five μl of the above PCR products (cDNA) were mixed with 3 μl of 20 μg/μl proteinase K supplied with the kit. The mixture was incubated at 45° C. for 20 min, then 75 μl of water was added and the mixture was extracted with 150 μl phenol:chloroform:isoamyl alcohol mixture (25:24:1). The aqueous phase was further extracted with 150 μl chloroform:isoamyl alcohol (25:1). The aqueous phase was then mixed with 15 μl of 3 M sodium acetate, 2 μl of 20 μg/μl glycogen and 400 μl of 100% ethanol. The mixture was immediately centrifuged at room temperature for 20 min at 14000 rpm in a microfuge. The pellet was washed once with 150 μl of 80% ethanol, air dried and dissolved in 79 μl of water.

[0169] Dissolved cDNA was subsequently digested with SfiI (79 μl of the cDNA was mixed with 10 μl of 10× SfiI buffer, 10 μl of SfiI enzyme and 1 μl of 100×BSA and the

mixture was incubated at 50° C. for 2 hrs). Xylene cyanol dye (2 μl of 1%) was added. The mixture was then fractionated on the Chroma Spin-400 column provided with the kit, following the manufacturer's procedure exactly. Fractions collected from the column were analyzed by agarose gel electrophoresis. The first three fractions containing cDNA were pooled and cDNA precipitated with ethanol. The precipitated cDNA was redissolved in 7 μl of water, and ligated into kit-supplied pDNR-LIB.

Library Sequencing

[0170] The ligation products were used to transform *E. coli* XL-1 Blue electroporation competent cells (Stratagene). An estimated total of 5.4×10^6 colonies was obtained. Sequencing of the cDNA library was carried out by Agencourt Bioscience Corporation (Beverly, Mass.), using a M13 forward primer (SEQ ID NO:7)

Example 2

Identification of the *Yarrowia lipolytica* Gene Encoding YAT1 as a Highly Expressed Gene Under Oleaginous Conditions

[0171] This example describes the identification of YAT1 as a highly expressed gene under oleaginous conditions. Specifically, the relative abundance of each species of cDNA was examined (based on the sequencing results from Example 1) to identify those genes that were highly expressed under oleaginous conditions. One of the genes that appeared multiple times (16/9984 or 0.16%) was Yali0E27203 (GenBank Accession No. XM_504457; SEQ ID NO:1 herein), annotated as a homolog of the high affinity, low capacity ammonia transporter MEP2 of *Saccharomyces cerevisiae* (GenBank Accession No. X83608; see also A. M. Marini et al., *EMBO J.*, 13(15):3456-3463 (1994); A. M. Marini et al., *Mol Cell Biol* 17(8):4282-93 (1997)).

[0172] BLAST (Basic Local Alignment Search Tool; Altschul, S. F., et al., *J. Mol. Biol.* 215:403-410 (1993)) searches were conducted using the corresponding amino acid sequence of Yali0E27203 (SEQ ID NO:2) as a query against the *S. cerevisiae* MEP1, MEP2 and MEP3-isozymes. The results of these BLAST comparisons are shown below in Table 3 and are reported according to the % identity, % similarity and Expectation value.

TABLE 3

Comparison Of *Yarrowia lipolytica* Homolog To *Saccharomyces cerevisiae* MEP1, MEP2 And MEP3

Similarity Identified	% Identity[a]	% Similarity[b]	E-value[c]
MEP1 [GenBank Accession No. X77608]	35	56	1.8e−79
MEP2 [GenBank Accession No. X83608]	52	66	1.9e−132
MEP3 [GenBank Accession No. P53390]	37	57	3.6e−83

[a]% Identity is defined as percentage of amino acids that are identical between the two proteins.
[b]% Similarity is defined as percentage of amino acids that are identical or conserved between the two proteins.
[c]Expect value. The Expect value estimates the statistical significance of the match, specifying the number of matches, with a given score, that are expected in a search of a database of this size absolutely by chance.

It was hypothesized that YALI0E27203 encoded a *Y. lipolytica* ammonia transporter of high affinity; thus, YALI0E27203 was tentatively named yat1.

Example 3

Isolation of the 5' Upstream Region of YAT1 from *Yarrowia Lipolytica*

[0173] To isolate the YAT1 promoter region upstream of the yat1 gene identified in Example 2, primers 27203-F and 27203-R (SEQ ID NOs:8 and 9) were designed, based on the sequence of *Y. lipolytica* chromosome E between positions 3,222,879 and 3,223,875. These primers were expected to amplify 778 bp of the 5' upstream untranslated region in front of the 'ATG' translation initiation codon of the yat1 gene, from −778 to −1 (wherein the 'A' position of the 'ATG' translation initiation codon was designated as +1).

[0174] PCR amplification was performed in a 50 μl total volume using a 1:1 dilution of a premixed 2×PCR solution (TaKaRa ExTaq, TaKaRa Bio Inc., Otsu Shiga, 520-2193, Japan). The final composition contained 25 mM TAPS, pH 9.3, 50 mM KCl, 2 mM $MgCl_2$, 1 mM 2-mercaptoethanol, 200 μM each deoxyribonucleotide triphosphate, 10 pmole each primer (supra), 50 ng *Y. lipolytica* ATCC #20362 genomic DNA, and 1.25 units of TaKaRa ExTaq™ DNA polymerase. The reaction mixture was first heated to 94° C. for 150 sec. Amplification was carried out for 30 cycles at 94° C. for 30 sec, 55° C. for 30 sec and 72° C. for 1 min. This was followed by a final extension for 7 min at 72° C.

[0175] The PCR product was analyzed by agarose gel electrophoresis and was shown to contain a single DNA fragment of ~800 bp. This fragment was purified using a Qiagen PCR purification kit following the manufacturer's protocol.

Example 4

Synthesis of Plasmids pY5-30, pYGPAT-GUS and pDMW214

[0176] A series of plasmids were created comprising a chimeric gene composed of various native *Y. lipolytica* promoters and the "GUS" reporter gene, wherein "GUS" corresponds to the *E. coli* gene encoding β-glucuronidase (Jefferson, R. A. *Nature.* 342(6251):837-838 (1989)). This was required for comparative studies investigating the promoter activity of TEF, YAT1, FBAIN and GPAT, as described in Example 8.

Synthesis of Plasmid pY5-30 (TEF::GUS::XPR)

[0177] The synthesis of plasmid pY5-30, comprising a TEF::GUS::XPR chimeric gene, is described in WO2005/003310. More specifically, plasmid pY5-30 (**FIG. 1A**; SEQ ID NO:10) contained: a *Yarrowia* autonomous replication sequence (ARS18); a ColE1 plasmid origin of replication; an ampicillin-resistance gene (Amp^R) for selection in *E. coli*; a *Yarrowia* LEU2 gene for selection in *Yarrowia*; and the chimeric TEF::GUS::XPR gene.

Synthesis of Plasmid pDMW214 (FBAIN::GUS::XPR)

[0178] The synthesis of plasmid pDMW214, comprising a FBAIN::GUS::XPR chimeric gene, is described in WO 2005/049805. Briefly, however, the FBAIN promoter region (SEQ ID NO:11; which includes both an upstream DNA

sequence and a downstream sequence from the putative 'ATG' translation initiation codon of the fructose-bisphosphate aldolase (fba1) gene [wherein the downstream region comprises an intron]) was amplified by PCR, digested with NcoI and SalI, and then purified following gel electrophoresis. The NcoI/SalIII-digested PCR products were ligated to NcoI/SalI digested pY5-30 vector to produce plasmid "pDMW214" (**FIG. 1B**).

Synthesis of pYGPAT-GUS (GPAT::GUS::XPR)

[**0179**] The synthesis of plasmid pYGPAT-GUS, comprising a GPAT::GUS::XPR chimeric gene, is described in U.S. patent application Ser. No. 60/610,060 (filed Sep. 15, 2004) and will be reiterated below. In general, synthesis of the plasmid required identification and isolation of the *Y. lipolytica* glycerol-3-phosphate O-acyltransferase (gpat) gene, isolation of the promoter region upstream of the putative 'ATG' translation initiation codon via genome-walking, and then cloning of the promoter into a derivative of pY5-30 (supra).

[**0180**] Isolation and Sequencing of the Putative *Yarrowia Lipolytica* GPAT

[**0181**] Based on the gene sequences encoding two isozymes of GPAT in *Saccharomyces cerevisiae* (GAT1 [GenBank Accession No. AJ311354] and GAT2(SCT1) [GenBank Accession No. AJ314608]; see also Zheng and Zou. *J. Biol. Chem.* 276(45):4171041716 (2001) and WO 02/08391 A2), BLAST searches (supra, Example 2) were conducted against the Genolevures genome database of *Y. lipolytica* ORFs (sponsored by the Center for Bioinformatics, LaBRI, bâtiment A30, Université Bordeaux 1, 351, cours de la Libération, 33405 Talence Cedex, France) to identify any similar sequences contained therein. The results of the BLAST comparisons identified one homolog (*Yarrowia lipolytica* ORF YAL1-CDS1055.1, a protein of 727 amino acids; SEQ ID NO:12; also GenBank Accession No. CAG81570) annotated as a protein having similarity to *S. cerevisiae* GPATs.

[**0182**] Following the tentative identification of ORF YAL1-CDS1055.1, this amino acid sequence was BLASTed against the *S. cerevisiae* isozymes encoding GPAT. The results of these BLAST comparisons are shown below and are reported according to the % identity, % similarity, and Expectation value (supra, Example 2).

TABLE 4

Comparison Of *Yarrowia lipolytica* ORF YAL1-CDS1055.1 To *Saccharomyces cerevisiae* GAT1 And GAT2(SCT1)

Similarity Identified	% Identity[a]	% Similarity[b]	E-value[c]
GAT1 [GenBank Accession No. AJ311354]	35	54	1.2e-113
GAT2(SCT1) [GenBank Accession No. AJ314608]	43	59	3.4e-141

Thus, it was hypothesized that ORF YALI-CDS1055.1 encoded a *Y. lipolytica* GPAT.

[**0183**] Degenerated oligonucleotides, as shown below, were designed to amplify the entire coding region of ORF YALI-CDS1055.1.

```
Degenerated Oligonucleotide YGPAT-F
                                        (SEQ ID NO:13)
ATGTCNGAGACYGACCAYCTNCTN

Degenerated oligonucleotide YGPAT-R
                                        (SEQ ID NO:14)
YTCYTCRTCYTGYTCTCGYCGYTT
```

[Note: The nucleic acid degeneracy code used for SEQ ID NOs:13 and 14 was as follows: R = A/G; Y = C/T; and N = A/C/T/G.]

[**0184**] The PCR amplification was carried out in a 50 μl total volume using a 1:1 dilution of a premixed 2×PCR solution and 1.25 units of TaKaRa Ex Taq™ DNA polymerase (Takara Mirus Bio, Madison, Wis.), according to the PCR composition described in Example 3. The thermocycler conditions were set for 30 cycles at 94° C. for 2.5 min, 55° C. for 30 sec and 72° C. for 2.5 min, followed by a final extension at 72° C. for 6 min.

[**0185**] The PCR products were separated by gel electrophoresis in 1% (w/v) agarose. A 2.2 kB DNA fragment was excised and purified using a QiaexII gel purification kit (Qiagen, Valencia, Calif.). Subequently, the purified 2.2 kB DNA fragment was cloned into the pGEM-T-easy vector (Promega, Madison, Wis.). The ligated DNA was used to transform cells of *E. coli* Top10 and transformants were selected on LB (1% bacto-tryptone, 0.5% bacto-yeast extract and 1% NaCl) agar containing ampicillin (100 μg/mL). Analysis of the plasmid DNA from one transformant confirmed the presence of a plasmid of the expected size, designated as "pGPAT-1".

[**0186**] Sequence analyses of pGPAT-1 showed that it contained a 2184 bp fragment of *Y. lipolytica* DNA encoding GPAT (SEQ ID NO:15). Specifically, identity of this gene sequence was determined by conducting BLAST searches for similarity to sequences contained in the BLAST "nr" database (comprising all non-redundant GenBank CDS translations, sequences derived from the 3-dimensional structure Brookhaven Protein Data Bank, the SWISS-PROT protein sequence database, EMBL and DDBJ databases). The sequence was analyzed for similarity to all publicly available DNA sequences contained in the "nr" database using the BLASTN algorithm provided by the National Center for Biotechnology Information (NCBI). The DNA sequence was translated in all reading frames and compared for similarity to all publicly available protein sequences contained in the "nr" database, using the BLASTX algorithm (Gish, W. and States, D. J. *Nature Genetics* 3:266-272 (1993)) provided by the NCBI. The BLAST searches revealed that the translated product of SEQ ID NO:15 (comprising the putative gpat gene) had the highest BLAST hits to annotated GPATs from: (1) *S. cerevisiae* (SwissPro P36148, Entrez CAA82146): 43% identity, 59% similarity, E-146; and (2) *Schizosaccharomyces pombe* SCT1(GAT2) homolog (GPAT): 48% identity, 66% similarity, E-148.

[**0187**] Furthermore, the translated product of SEQ ID NO:15 was 100% identical to the amino acid sequence of ORF YAL1-CDS1055.1 (SEQ ID NO:12; also GenBank Accession No. CAG81570).

[**0188**] Isolation of the 5' Upstream Region of GPAT from *Y. Lipolytica*

[**0189**] To isolate the GPAT promoter region upstream of the gene identified in Example 1, a genome-walking tech-

nique (Universal GenomeWalker, ClonTech, CA) was utilized, following the manufacturer's protocol.

[0190] Briefly, genomic DNA of *Y. lipolytica* was digested with DraI, EcoRV, PvuII or StuI individually, and the digested DNA samples were ligated with Genome Walker adaptor (SEQ ID NOs:16 [top strand] and 17 [bottom strand]), as shown below:

```
5'-GTAATACGACTCACTATAGGGCACGCGTGGTCGACGGCCCGGGCTGGT-3'

                                    3'-H2N-CCCGACCA-5'
```

[0191] PCR reactions were then carried out using the ligation products as templates, and the following oligonucleotides as primers: Adaptor-1 and YGPAT-5R-1 (SEQ ID NOs:18 and 19). The PCR amplification was carried out in a 50 μl total volume using the components and conditions previously described, with the exception that the template used was 1 μl of ligation product (versus 50 ng genomic DNA). Second PCR reactions were then carried out using 1 μl of 1:50 diluted first PCR product as template, and Nested Adaptor Primer 2 and YGPAT-5R-2 (SEQ ID NOs:20 and 21) as primers.

[0192] A 1.7 kB DNA fragment, amplified from the EcoRV digested sample, was purified using a Qiagen PCR purification kit and cloned into the pGEM-T-easy vector (Promega, Madison, Wis.). The ligated DNA was used to transform *E. coli* Top10 and transformants were selected on LB agar containing ampicillin (100 μg/mL).

[0193] Analysis of the plasmid DNA from one transformant confirmed the presence of the expected plasmid, designated "pEcoRV-G-5". Sequence analyses showed that pEcoRV-G-5 contained a fragment of 1781 bp (SEQ ID NO:22), which included 1678 bp of 5' upstream sequence from the nucleotide 'A' of the translation initiation codon 'ATG' of the GPAT gene. Assembly of DNA corresponding to overlapping SEQ ID NOs:15 and 22 yielded a single contig of DNA represented as SEQ ID NO:23 (**FIG. 2**; 3862 bp total length). This contig therefore contained the −1678 to +2181 region of the GPAT gene, wherein the 'A' position of the 'ATG' translation initiation codon was designated as +1.

[0194] Construction of pYGPAT-GUS

[0195] The nucleotide region between the −1130 position and the 'ATG' translation initiation site of the gpat gene was considered to represent the putative GPAT promoter region ("GPAT-Pro", provided as SEQ ID NO:24).

[0196] GPAT-Pro was amplified by PCR using primers GPAT-5-1 and GPAT-5-2 (SEQ ID NOs:25 and 26), and pEcoRV-G5 (supra) as template. The PCR product was then purified using a Qiagen PCR purification kit and was completely digested with SalI and NcoI. The digestion product was again purified with a Qiagen PCR purification kit and ligated to NcoI/SalI digested pY5-30 vector (supra, wherein the NcoI/SalI digestion had excised the TEF promoter from the pY5-30 vector backbone). Ligated DNA was then used to individually transform *E. coli* Top10. Transformants were selected on LB agar containing ampicillin (100 μg/mL).

[0197] Analysis of the plasmid DNA from one transformant containing GPAT-Pro confirmed the presence of the expected plasmid, designated "pYGPAT-GUS". Thus, this plasmid contained a chimeric gene comprising a GPAT promoter, GUS reporter gene and XPR terminator (**FIG. 1C**).

Example 5

Synthesis of pYAT-GUS

[0198] The present Example describes the synthesis of pYAT-GUS (comprising a YAT1::GUS::XPR chimeric gene). Synthesis of this plasmid first required amplification of the putative YAT1 promoter region. Then, the putative promoter region was cloned into pYGPAT-GUS (supra, Example 4).

[0199] The purified YAT1 PCR product from Example 3 was digested with HindIII and SalI and a ~600 bp fragment was isolated by agarose gel electrophoresis followed by purification with Qiagen MinElute Gel purification kit according to the manufacturer's protocol. Furthermore, the YAT1 PCR product was also digested with NcoI and HindIII, and a ~200 bp fragment was isolated and purified as above. Finally, plasmid pYGPAT-GUS was digested with SalI and NcoI, and a ~9.5 kB fragment was isolated and purified. The three DNA fragments were ligated together to create pYAT-GUS (**FIG. 1D**). In the resulting plasmid, the YAT1 promoter region (SEQ ID NO:3; corresponding to 778 bp of the 5' upstream region of YALI0E27203 (−778 to −1)) was fused to the GUS gene, followed by the XPR terminator region.

Example 6

Generation of ARA-Producing *Yarrowia lipolytica* ATCC #20362 Strain Y2034

[0200] The present Example describes the construction of strain Y2034, derived from *Y. lipolytica* ATCC #20362, capable of producing significant concentrations of ARA relative to the total lipids. Comparison of the TEF, YAT1, GPAT and FBAIN promoters was examined in this ARA-producing strain based on analysis of GUS expression, as described in Example 8 (infra).

[0201] The development of strain Y2034 herein, producing 10% ARA, required the construction of strain M4 (producing 8% DGLA) (**FIG. 3A**).

Construction of Strain M4 Producing 8% DGLA

[0202] Construct pKUNF12T6E (**FIG. 3B**; SEQ ID NO:27) was generated to integrate four chimeric genes (comprising a Δ12 desaturase, a Δ6 desaturase and 2 elongases) into the Ura3 loci of wild type *Yarrowia* strain ATCC #20362, to thereby enable production of DGLA. The pKUNF12T6E plasmid contained the following components:

TABLE 5

Description of Plasmid pKUNF12T6E (SEQ ID NO: 27)

RE Sites And Nucleotides Within SEQ ID NO: 27	Description Of Fragment And Chimeric Gene Components
AscI/BsiWI (9420-8629)	784 bp 5' part of *Yarrowia* Ura3 gene (GenBank Accession No. AJ306421)
SphI/PacI (12128-1)	516 bp 3' part of *Yarrowia* Ura3 gene (GenBank Accession No. AJ306421)
SwaI/BsiWI (6380-8629)	FBAIN::EL1S:Pex20, comprising: FBAIN: FBAIN promoter (SEQ ID NO: 11) EL1S: codon-optimized elongase 1 gene (SEQ ID NO: 28), derived from *Mortierella alpina* (GenBank Accession No. AX464731) Pex20: Pex20 terminator sequence from *Yarrowia* Pex20 gene (GenBank Accession No. AF054613)
BglII/SwaI (4221-6380)	TEF::Δ6S::Lip1, comprising: TEF: TEF promoter (GenBank Accession No. AF054508) Δ6S: codon-optimized Δ6 desaturase gene (SEQ ID NO: 30), derived from *Mortierella alpina* (GenBank Accession No. AF465281) Lip1: Lip1 terminator sequence from *Yarrowia* Lip1 gene (GenBank Accession No. Z50020)
PmeI/ClaI (4207-1459)	FBA::F.Δ12::Lip2, comprising: FBA: FBA promoter (SEQ ID NO: 32) F.Δ12: *Fusarium moniliforme* Δ12 desaturase gene (SEQ ID NO: 33) Lip2: Lip2 terminator sequence from *Yarrowia* Lip2 gene (GenBank Accession No. AJ012632)
ClaI/PacI (1459-1)	TEF::EL2S::XPR, comprising: TEF: TEF promoter (GenBank Accession No. AF054508) EL2S: codon-optimized elongase gene (SEQ ID NO: 35), derived from *Thraustochytrium aureum* (U.S. Pat. No. 6,677,145) XPR: XPR terminator sequence of *Yarrowia* Xpr gene (GenBank Accession No. M17741)

[0203] The pKUNF12T6E plasmid was digested with AscI/SphI, and then used for transformation of wild type *Y. lipolytica* ATCC #20362 according to the General Methods. The transformant cells were plated onto FOA selection media plates and maintained at 30° C. for 2 to 3 days. The FOA resistant colonies were picked and streaked onto MM and MMU selection plates. The colonies that could grow on MMU plates but not on MM plates were selected as Ura- strains. Single colonies of Ura- strains were then inoculated into liquid MMU at 30° C. and shaken at 250 rpm/min for 2 days. The cells were collected by centrifugation, lipids were extracted, and fatty acid methyl esters were prepared by trans-esterification, and subsequently analyzed with a Hewlett-Packard 6890 GC.

[0204] GC analyses showed the presence of DGLA in the transformants containing the 4 chimeric genes of pKUNF12T6E (**FIG. 3B**), but not in the wild type *Yarrowia* control strain. Most of the selected 32 Ura^- strains produced about 6% DGLA of total lipids. There were 2 strains (i.e., strains M4 and 13-8) that produced about 8% DGLA of total lipids.

Construction of Strain Y2034 Producing 10% ARA

[0205] Construct pDMW232 (**FIG. 3C**; SEQ ID NO:37) was generated to integrate two Δ5 chimeric genes into the Leu2 gene of *Yarrowia* strain M4.

[0206] Plasmid pDMW232 contained the following components, as described in Table 6:

TABLE 6

Description of Plasmid pDMW232 (SEQ ID NO: 37)

RE Sites And Nucleotides Within SEQ ID NO: 37	Description Of Fragment And Chimeric Gene Components
AscI/BsiWI (5550-4755)	788 bp 5' part of *Yarrowia* Leu2 gene (GenBank Accession No. AF260230)
SphI/PacI (8258-8967)	703 bp 3' part of *Yarrowia* Leu2 gene (GenBank Accession No. AF260230)
SwaI/BsiWI (2114-4755)	FBAIN::MAΔ5::Pex20, comprising: FBAIN: FBAIN Promoter (SEQ ID NO: 11) MAΔ5: *Mortierella alpina* Δ5 desaturase gene (GenBank Accession No. AF067654) Pex20: Pex20 terminator sequence of *Yarrowia* Pex20 gene (GenBank Accession No. AF054613)
SwaI/ClaI (2114-17)	TEF::MAΔ5::Lip1, comprising: TEF: TEF promoter (GenBank Accession No. AF054508) MAΔ5: as described for FBAIN::MAΔ5::Pex20 (supra) Lip1: Lip1 terminator sequence of *Yarrowia* Lip1 gene (GenBank Accession No. Z50020)
PmeI/ClaI (5550-4755)	*Yarrowia* Ura3 gene (GenBank Accession No. AJ306421)

[0207] Plasmid pDMW232 was digested with AscI/SphI, and then used to transform strain M4 (Example 4) separately according to the General Methods. Following transformation, the cells were plated onto MMLe plates and maintained at 30° C. for 2 to 3 days. The individual colonies grown on MMLe plates were picked and streaked onto MM and MMLe plates. Those colonies that could grow on MMLe plates but not on MM plates were selected as $Leu2^-$ strains. Single colonies of $Leu2^-$ strains were then inoculated into liquid MMLe media at 30° C. and shaken at 250 rpm/min for 2 days. The cells were collected by centrifugation, lipids were extracted, and fatty acid methyl esters were prepared by trans-esterification, and subsequently analyzed with a Hewlett-Packard 6890 GC.

[0208] GC analyses showed the presence of ARA in pDMW232 transformants, but not in the parental M4 strain. Specifically, among the 48 selected $Leu2^-$ transformants with pDMW232, there were 34 strains that produced less than 5% ARA, 11 strains that produced 6-8% ARA, and 3 strains that produced about 10% ARA of total lipids in the engineered *Yarrowia*. One of the strains that produced 10% ARA was named "Y2034" and was the parental strain in Example 7.

Example 7

Transformation of *Y. lipolytica* with pY5-30, pYAT-GUS, pYGPAT-GUS and pDMW214

[0209] The plasmids pY5-30 (Example 4; comprising a TEF::GUS::XPR chimeric gene), pYAT-GUS (Example 5; comprising a YAT1::GUS::XPR chimeric gene), pYGPAT-GUS (Example 4; comprising a GPAT::GUS::XPR chimeric gene) and pDMW214 (Example 4; comprising a FBAIN-::GUS::XPR chimeric gene) were transformed separately into *Y. lipolytica* ATCC #20362 strain Y2034, according to the General Methods. Selection was performed on SD medium comprising 2.5% agar.

[0210] Using this technique, transformants were obtained that contained pY5-30, pYAT-GUS, pYGPAT-GUS and pDMW214, respectively.

Example 8

Comparative Analysis of the TEF, YAT1, GPAT and FBAIN Promoter Activities in *Yarrowia Lipolytica*, as Determined by Histochemical Assay

[0211] The activity of the TEF, YAT1, GPAT and FBAIN promoters was determined in *Y. lipolytica* containing the pY5-30, pYAT-GUS, pYGPAT-GUS and pDMW214 constructs, each of which possessed a GUS reporter gene and an XPR terminator (from Example 7). GUS activity in each expressed construct was measured by histochemical assays (Jefferson, R. A. *Plant Mol. Biol. Reporter* 5:387-405 (1987)).

[0212] Specifically, *Y. lipolytica* strains containing plasmids pY5-30, pYAT-GUS, pYGPAT-GUS and pDMW214, respectively, were grown from single colonies in 5 mL SD media at 30° C. for 24 hrs to an OD_{600}~8.0. Then, 1 mL of cells were collected by centrifugation. The remaining cultures were centrifuged and washed 2× with HGM, resuspended in 5 mL each of HGM and allowed to grow at 30° C. further. After 24 and 120 hrs, ~0.25 mL of each culture were centrifuged to collect the cells. Cell samples were resuspended individually in 100 μl of histochemical staining buffer [Staining buffer prepared by dissolving 5 mg of 5-bromo-4-chloro-3-indolyl glucuronide (X-Gluc) in 50 μl dimethyl formamide, followed by addition of 5 mL 50 mM $NaPO_4$, pH 7.0.]. Zymolase 20T (5 μl of 1 mg/mL; ICN Biomedicals, Costa Mesa, Calif.) was added to each, and the mixture incubated at 30° C.

[0213] The results of histochemical staining showed that the YAT1 promoter in construct pYAT-GUS was active. Comparatively, the YAT1 promoter appeared to be stronger than the TEF promoter (**FIG. 4A**) but significantly weaker than the FBAIN promoter and GPAT promoter, when cells were grown in SD medium for 24 hrs. More interestingly, however, it appeared that the YAT1 promoter was stronger than the GPAT promoter and comparable with FBAIN promoter in cells grown in HGM for 24 hrs (**FIG. 4B**). This remained true after 120 hrs in HGM (**FIG. 4C**). Thus, the YAT1 promoter appeared to be induced in HGM, a medium that promotes oleaginous growth conditions due to nitrogen limitation.

Example 9

Comparative Analysis of the TEF, YAT1, FBAIN and GPAT Promoter Activities in *Yarrowia Lipolytica*, as Determined by Fluorometric Assay

[0214] A variety of methods are available to compare the activity of various promoters, to thereby facilitate determination of each promoter's strength for use in future applications wherein a suite of promoters would be necessary to construct chimeric genes. Thus, although it may be useful to indirectly quantitate promoter activity based on reporter gene expression using histochemical staining (Example 8), quantification of GUS expression using more quantitative means may be desirable. One suitable method to assay GUS activity is by fluorometric determination of the production of 4-methylumbelliferone (4-MU) from the corresponding β-glucuronide (4-MUG; see Jefferson, R. A., *Plant Mol. Biol. Reporter* 5:387-405 (1987)).

[0215] *Yarrowia lipolytica* strain Y2034 containing plasmids pY5-30, pYAT-GUS, pYGPAT-GUS and pDMW214 constructs, respectively (from Example 7), were grown from single colonies in 10 mL SD medium at 30° C. for 48 hrs to an OD_{600} ~5.0. Two mL of each culture was collected for GUS activity assays, as described below, while 5 mL of each culture was switched into HGM.

[0216] Specifically, cells from the 5 mL aliquot were collected by centrifugation, washed once with 5 mL of HGM and resuspended in HG medium. The cultures in HGM were then grown in a shaking incubator at 30° C. for 24 hrs. Two mL of each HGM culture were collected for GUS activity assay, while the remaining culture was allowed to grow for an additional 96 hrs before collecting an additional 2 mL of each culture for the assay.

[0217] Each 2 mL culture sample in SD medium was resuspended in 1 mL of 0.5×cell culture lysis reagent (Promega). Resuspended cells were mixed with 0.6 mL of glass beads (0.5 mm diameter) in a 2.0 mL screw cap tube with a rubber O-ring. The cells were then homogenized in a Biospec mini beadbeater (Bartlesville, Okla.) at the highest setting for 90 sec. The homogenization mixtures were centrifuged for 2 min at 14,000 rpm in an Eppendof centrifuge to remove cell debris and beads. The supernatant was used for GUS assay and protein determination.

[0218] For each fluorometric assay, 200 μl of extract was added to 800 μl of GUS assay buffer (2 mM 4-methylumbelliferyl-β-D-glucuronide ("MUG") in extraction buffer) and placed at 37° C. Aliquots of 100 μl were taken at 0, 30 and 60 min time points and added to 900 μl of stop buffer (1 M Na_2CO_3). Each time point was read using a Fluorimeter (CytoFluorR Series 4000, Framingham, Mass.) set to an excitation wavelength of 360 nm and an emission wavelength of 455 nm. Total protein concentration of each sample was determined using 20 μl of extract and 980 μl of BioRad Bradford reagent (Bradford, M. M. *Anal. Biochem.* 72:248-254 (1976)). GUS activity is expressed as nmoles of 4-MU per minute per mg of total protein.

[0219] As shown in the Table below, the activity of the YAT1 promoter was highly induced after 24 hrs in HGM.

TABLE 7

Comparison of TEF, FBAIN, YAT1 And GPAT Promoter Activity Under Various Growth Conditions

Culture	Promoter			
Conditions	TEF	FBAIN	YAT1	GPAT
48 hr, SD	0.401	43.333	0.536	5.252
24 hr, HGM	0.942	30.694	19.154	2.969
120 hr HGM	0.466	17.200	13.400	3.050

[0220] Based on the data above wherein the activity of the YAT1 promoter was quantitated based on GUS activity of cell extracts, the activity of the YAT1 promoter increased by ~37 fold when cells were switched from SD medium into HGM and grown for 24 hrs. After 120 hrs in HGM, the activity was reduced somewhat but was still 25× higher than the activity in SD medium. In contrast, the activity of the

FBAIN promoter and the GPAT promoter was reduced by 30% and 40%, respectively, when switched from SD medium into HGM for 24 hrs. The activity of the TEF promoter increased by 2.3 fold after 24 hrs in HGM. Thus, the YAT1 promoter is inducible under oleaginous conditions and will thereby function as an effective promoter for the regulated expression of heterologous genes in *Yarrowia lipolytica*.

SEQUENCE LISTING

<160> NUMBER OF SEQ ID NOS: 37

<210> SEQ ID NO 1
<211> LENGTH: 1461
<212> TYPE: DNA
<213> ORGANISM: Yarrowia lipolytica (GenBank Accession No. XM_504457)
<220> FEATURE:
<221> NAME/KEY: CDS
<222> LOCATION: (1)..(1461)

<400> SEQUENCE: 1

```
atg gct gac gct acc tca acg caa gct ccc ctt ccg acc act ggc aat       48
Met Ala Asp Ala Thr Ser Thr Gln Ala Pro Leu Pro Thr Thr Gly Asn
1               5                   10                  15

gga gga gac tcc ctc aca caa aac ctg aac gtc ccc ttc ctg ggt gcc       96
Gly Gly Asp Ser Leu Thr Gln Asn Leu Asn Val Pro Phe Leu Gly Ala
            20                  25                  30

gat atg gtc tgg atc atg acc tct tcg gca ctg gtc tgg atc atg att      144
Asp Met Val Trp Ile Met Thr Ser Ser Ala Leu Val Trp Ile Met Ile
        35                  40                  45

cca gga gtc ggt ctg ctc tac tct ggt atg tcg cgt aag cac cac gcc      192
Pro Gly Val Gly Leu Leu Tyr Ser Gly Met Ser Arg Lys His His Ala
    50                  55                  60

ctg tct ctc ctg tgg gca tcc atc atg tgc tgt gcc ctc gtc tct ttc      240
Leu Ser Leu Leu Trp Ala Ser Ile Met Cys Cys Ala Leu Val Ser Phe
65                  70                  75                  80

gag tgg ttt ttc tgg ggt tac act ctg gca ttc tct cac aag gcc gga      288
Glu Trp Phe Phe Trp Gly Tyr Thr Leu Ala Phe Ser His Lys Ala Gly
                85                  90                  95

aag ttc att gga acc atg gat aac ttc ggt ctt atg aat gtt ctc gct      336
Lys Phe Ile Gly Thr Met Asp Asn Phe Gly Leu Met Asn Val Leu Ala
            100                 105                 110

gct ccc tct gtt ggc tcc tct gct gtc ccc gat atc ctg tat atg ttc      384
Ala Pro Ser Val Gly Ser Ser Ala Val Pro Asp Ile Leu Tyr Met Phe
        115                 120                 125

tac cag ggc atg ttc gca tgc atc acc ggt atg ctc atg gtt ggt gga      432
Tyr Gln Gly Met Phe Ala Cys Ile Thr Gly Met Leu Met Val Gly Gly
    130                 135                 140

gct cac gag cga gct cga ctc ggc ccc atg atg gta tat ctc ttc att      480
Ala His Glu Arg Ala Arg Leu Gly Pro Met Met Val Tyr Leu Phe Ile
145                 150                 155                 160

tgg atg act gtc gtc tac tct cct att gca tgc tgg aca tgg aac cct      528
Trp Met Thr Val Val Tyr Ser Pro Ile Ala Cys Trp Thr Trp Asn Pro
                165                 170                 175

agt gga tgg ctc gcc gtc ctt gga gga ctt gat ttc gca ggt gga gga      576
Ser Gly Trp Leu Ala Val Leu Gly Gly Leu Asp Phe Ala Gly Gly Gly
            180                 185                 190

cct gtt cac atg tct tcc ggt gcg ggt gcc ctt gcc tat gct ctc tgg      624
Pro Val His Met Ser Ser Gly Ala Gly Ala Leu Ala Tyr Ala Leu Trp
        195                 200                 205

tgt ggt aag aga cgt gac cct gct gtt gag aag ctg cct cac tac cgg      672
Cys Gly Lys Arg Arg Asp Pro Ala Val Glu Lys Leu Pro His Tyr Arg
    210                 215                 220
```

-continued

```
ccc tct tcc gtt act tcc gtt gtt ctc ggc act gtt ttg ctt tgg ttc      720
Pro Ser Ser Val Thr Ser Val Val Leu Gly Thr Val Leu Leu Trp Phe
225                 230                 235                 240

gga tgg ttc gga ttc aac ggt ggt tcc tct ggt aac gcc tcc atc cga      768
Gly Trp Phe Gly Phe Asn Gly Gly Ser Ser Gly Asn Ala Ser Ile Arg
                245                 250                 255

ggc ttc tac gct gcc gct aat act aac ctt gct gct gct tgc ggt gct      816
Gly Phe Tyr Ala Ala Ala Asn Thr Asn Leu Ala Ala Ala Cys Gly Ala
            260                 265                 270

ctc gct tgg atg tgt gtc gac ttc ttc cga aag ggc cga aag tgg tcc      864
Leu Ala Trp Met Cys Val Asp Phe Phe Arg Lys Gly Arg Lys Trp Ser
        275                 280                 285

act gtt ggt ctc tgt tct ggt gcc atc gca ggt ctc gtt ggc atc acc      912
Thr Val Gly Leu Cys Ser Gly Ala Ile Ala Gly Leu Val Gly Ile Thr
    290                 295                 300

ccc gcc gcc ggc ttc gtc cct atc tgg tct gct gtc cct att ggt atc      960
Pro Ala Ala Gly Phe Val Pro Ile Trp Ser Ala Val Pro Ile Gly Ile
305                 310                 315                 320

atc acc gcc gtc ttc gct aac att tct ggt gac ctt aag aat ctg ctc     1008
Ile Thr Ala Val Phe Ala Asn Ile Ser Gly Asp Leu Lys Asn Leu Leu
                325                 330                 335

cga att gat gat ggt ctc gat gtc ttc tct ctc cat ggt gtg gga gga     1056
Arg Ile Asp Asp Gly Leu Asp Val Phe Ser Leu His Gly Val Gly Gly
            340                 345                 350

ttc tgc ggc tct gtt ctt act gcc ttc ttt gct gct gac tac att gcc     1104
Phe Cys Gly Ser Val Leu Thr Ala Phe Phe Ala Ala Asp Tyr Ile Ala
        355                 360                 365

cat ctg gat ggt gcc aca gag atc aag ggt gga tgg ctc aac cac cac     1152
His Leu Asp Gly Ala Thr Glu Ile Lys Gly Gly Trp Leu Asn His His
    370                 375                 380

tgg gct cag ctg ggt tac cag ctg gct ggt gcg ttt gct acc ctc ggc     1200
Trp Ala Gln Leu Gly Tyr Gln Leu Ala Gly Ala Phe Ala Thr Leu Gly
385                 390                 395                 400

tac tcc ttt gtg gtc tct tca gtt att ctt gtc atc atg aac aga atc     1248
Tyr Ser Phe Val Val Ser Ser Val Ile Leu Val Ile Met Asn Arg Ile
                405                 410                 415

ccc tac ctc aac gtc cga atg acc gag gag gag gag atg ctt gga acc     1296
Pro Tyr Leu Asn Val Arg Met Thr Glu Glu Glu Glu Met Leu Gly Thr
            420                 425                 430

gac atg gcc cag atc ggc gag ttt gcc ttc gac tgg gag gac tct gga     1344
Asp Met Ala Gln Ile Gly Glu Phe Ala Phe Asp Trp Glu Asp Ser Gly
        435                 440                 445

gtc ctg gac ctg cat ggc cag aac ccc aat ggt atg ggc gtc acc ccc     1392
Val Leu Asp Leu His Gly Gln Asn Pro Asn Gly Met Gly Val Thr Pro
    450                 455                 460

aac gtc cag act ccc aag ccc agc agc atc aac gaa aac aag gag gct     1440
Asn Val Gln Thr Pro Lys Pro Ser Ser Ile Asn Glu Asn Lys Glu Ala
465                 470                 475                 480

gct gag agt gat tcg gtt taa                                        1461
Ala Glu Ser Asp Ser Val
                485
```

<210> SEQ ID NO 2
<211> LENGTH: 486
<212> TYPE: PRT
<213> ORGANISM: Yarrowia lipolytica (GenBank Accession No. XM_504457)

<400> SEQUENCE: 2

-continued

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

Gly Gly Asp Ser Leu Thr Gln Asn Leu Asn Val Pro Phe Leu Gly Ala
            20                  25                  30

Asp Met Val Trp Ile Met Thr Ser Ser Ala Leu Val Trp Ile Met Ile
        35                  40                  45

Pro Gly Val Gly Leu Leu Tyr Ser Gly Met Ser Arg Lys His His Ala
    50                  55                  60

Leu Ser Leu Leu Trp Ala Ser Ile Met Cys Cys Ala Leu Val Ser Phe
65                  70                  75                  80

Glu Trp Phe Phe Trp Gly Tyr Thr Leu Ala Phe Ser His Lys Ala Gly
                85                  90                  95

Lys Phe Ile Gly Thr Met Asp Asn Phe Gly Leu Met Asn Val Leu Ala
            100                 105                 110

Ala Pro Ser Val Gly Ser Ser Ala Val Pro Asp Ile Leu Tyr Met Phe
        115                 120                 125

Tyr Gln Gly Met Phe Ala Cys Ile Thr Gly Met Leu Met Val Gly Gly
    130                 135                 140

Ala His Glu Arg Ala Arg Leu Gly Pro Met Met Val Tyr Leu Phe Ile
145                 150                 155                 160

Trp Met Thr Val Val Tyr Ser Pro Ile Ala Cys Trp Thr Trp Asn Pro
                165                 170                 175

Ser Gly Trp Leu Ala Val Leu Gly Gly Leu Asp Phe Ala Gly Gly Gly
            180                 185                 190

Pro Val His Met Ser Ser Gly Ala Gly Ala Leu Ala Tyr Ala Leu Trp
        195                 200                 205

Cys Gly Lys Arg Arg Asp Pro Ala Val Glu Lys Leu Pro His Tyr Arg
    210                 215                 220

Pro Ser Ser Val Thr Ser Val Val Leu Gly Thr Val Leu Leu Trp Phe
225                 230                 235                 240

Gly Trp Phe Gly Phe Asn Gly Gly Ser Ser Gly Asn Ala Ser Ile Arg
                245                 250                 255

Gly Phe Tyr Ala Ala Ala Asn Thr Asn Leu Ala Ala Ala Cys Gly Ala
            260                 265                 270

Leu Ala Trp Met Cys Val Asp Phe Phe Arg Lys Gly Arg Lys Trp Ser
        275                 280                 285

Thr Val Gly Leu Cys Ser Gly Ala Ile Ala Gly Leu Val Gly Ile Thr
    290                 295                 300

Pro Ala Ala Gly Phe Val Pro Ile Trp Ser Ala Val Pro Ile Gly Ile
305                 310                 315                 320

Ile Thr Ala Val Phe Ala Asn Ile Ser Gly Asp Leu Lys Asn Leu Leu
                325                 330                 335

Arg Ile Asp Asp Gly Leu Asp Val Phe Ser Leu His Gly Val Gly Gly
            340                 345                 350

Phe Cys Gly Ser Val Leu Thr Ala Phe Phe Ala Ala Asp Tyr Ile Ala
        355                 360                 365

His Leu Asp Gly Ala Thr Glu Ile Lys Gly Gly Trp Leu Asn His His
    370                 375                 380

Trp Ala Gln Leu Gly Tyr Gln Leu Ala Gly Ala Phe Ala Thr Leu Gly
385                 390                 395                 400
```

-continued

```
Tyr Ser Phe Val Val Ser Ser Val Ile Leu Val Ile Met Asn Arg Ile
                405                 410                 415

Pro Tyr Leu Asn Val Arg Met Thr Glu Glu Glu Glu Met Leu Gly Thr
            420                 425                 430

Asp Met Ala Gln Ile Gly Glu Phe Ala Phe Asp Trp Glu Asp Ser Gly
        435                 440                 445

Val Leu Asp Leu His Gly Gln Asn Pro Asn Gly Met Gly Val Thr Pro
    450                 455                 460

Asn Val Gln Thr Pro Lys Pro Ser Ser Ile Asn Glu Asn Lys Glu Ala
465                 470                 475                 480

Ala Glu Ser Asp Ser Val
                485
```

<210> SEQ ID NO 3
<211> LENGTH: 778
<212> TYPE: DNA
<213> ORGANISM: Yarrowia lipolytica

<400> SEQUENCE: 3

```
ataagtttgc aaaaagatcg tattatagtt ggagcaaggg agaaatgtag agtgtgaaag     60

actcactatg gtccgggctt atctcgacca atagccaaag tctggagttt ctgagagaaa    120

aaggcaagat acgtatgtaa caaagcgacg catggtacaa taataccgga ggcatgtatc    180

atagagagtt agtggttcga tgatggcact ggtgcctggt atgactttat acggctgact    240

acatatttgt cctcagacat acaattacag tcaagcactt acccttggac atctgtaggt    300

acccccggc caagacgatc tcagcgtgtc gtatgtcgga ttggcgtagc tccctcgctc    360

gtcaattggc tcccatctac tttcttctgc ttggctacac ccagcatgtc tgccatggct    420

cgttttcgtg ccttatctat cctcccagta ttaccaactc taaatgacat gatgtgattg    480

ggtctacact ttcatatcag agataaggag tagcacagtt gcataaaaag cccaactcta    540

atcagcttct tcctttcttg taattagtac aaaggtgatt agcgaaatct ggaagcttag    600

ttggccctaa aaaaatcaaa aaaagcaaaa aacgaaaaac gaaaaaccac agttttgaga    660

acagggaggt aacgaaggat cgtatatata tatatatata tatataccca cggatcccga    720

gaccggcctt tgattcttcc ctacaaccaa ccattctcac caccctaatt cacaaatg      778
```

<210> SEQ ID NO 4
<211> LENGTH: 39
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: SMART IV oligonucleotide

<400> SEQUENCE: 4

```
aagcagtggt atcaacgcag agtggccatt acggccggg                            39
```

<210> SEQ ID NO 5
<211> LENGTH: 59
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: CDSIII/3'PCR primer
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (28)..(57)
<223> OTHER INFORMATION: thymidine (dT); see BD Biosciences Clontech's
SMART cDNA technology

-continued

<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (59)..(59)
<223> OTHER INFORMATION: n is a, c, g, or t

<400> SEQUENCE: 5

attctagagg ccgaggcggc cgacatgttt ttttttttt ttttttttt tttttttvn 59

<210> SEQ ID NO 6
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: 5'-PCR primer

<400> SEQUENCE: 6

aagcagtggt atcaacgcag agt 23

<210> SEQ ID NO 7
<211> LENGTH: 18
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Primer M13

<400> SEQUENCE: 7

tgtaaaacga cggccagt 18

<210> SEQ ID NO 8
<211> LENGTH: 32
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Primer 27203-F

<400> SEQUENCE: 8

gatcgtcgac ataagtttgc aaaaagatcg ta 32

<210> SEQ ID NO 9
<211> LENGTH: 34
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Primer 27203-R

<400> SEQUENCE: 9

gatcccatgg ttgtgaatta gggtggtgag aatg 34

<210> SEQ ID NO 10
<211> LENGTH: 8953
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Plasmid pY5-30

<400> SEQUENCE: 10

ggtggagctc cagcttttgt tccctttagt gagggttaat ttcgagcttg gcgtaatcat 60

ggtcatagct gtttcctgtg tgaaattgtt atccgctcac aattccacac aacatacgag 120

ccggaagcat aaagtgtaaa gcctggggtg cctaatgagt gagctaactc acattaattg 180

cgttgcgctc actgcccgct ttccagtcgg gaaacctgtc gtgccagctg cattaatgaa 240

tcggccaacg cgcggggaga ggcggtttgc gtattgggcg ctcttccgct tcctcgctca 300

-continued

ctgactcgct gcgctcggtc gttcggctgc ggcgagcggt atcagctcac tcaaaggcgg 360

taatacggtt atccacagaa tcaggggata acgcaggaaa gaacatgtga gcaaaaggcc 420

agcaaaaggc caggaaccgt aaaaaggccg cgttgctggc gtttttccat aggctccgcc 480

cccctgacga gcatcacaaa aatcgacgct caagtcagag gtggcgaaac ccgacaggac 540

tataaagata ccaggcgttt cccccctggaa gctccctcgt gcgctctcct gttccgaccc 600

tgccgcttac cggatacctg tccgcctttc tcccttcggg aagcgtggcg ctttctcata 660

gctcacgctg taggtatctc agttcggtgt aggtcgttcg ctccaagctg ggctgtgtgc 720

acgaaccccc cgttcagccc gaccgctgcg ccttatccgg taactatcgt cttgagtcca 780

acccggtaag acacgactta tcgccactgg cagcagccac tggtaacagg attagcagag 840

cgaggtatgt aggcggtgct acagagttct tgaagtggtg gcctaactac ggctacacta 900

gaaggacagt atttggtatc tgcgctctgc tgaagccagt taccttcgga aaaagagttg 960

gtagctcttg atccggcaaa caaaccaccg ctggtagcgg tggtttttt gtttgcaagc 1020

agcagattac gcgcagaaaa aaaggatctc aagaagatcc tttgatcttt tctacggggt 1080

ctgacgctca gtggaacgaa aactcacgtt aagggatttt ggtcatgaga ttatcaaaaa 1140

ggatcttcac ctagatcctt ttaaattaaa aatgaagttt taaatcaatc taaagtatat 1200

atgagtaaac ttggtctgac agttaccaat gcttaatcag tgaggcacct atctcagcga 1260

tctgtctatt tcgttcatcc atagttgcct gactccccgt cgtgtagata actacgatac 1320

gggagggctt accatctggc cccagtgctg caatgatacc gcgagaccca cgctcaccgg 1380

ctccagattt atcagcaata aaccagccag ccggaagggc cgagcgcaga agtggtcctg 1440

caactttatc cgcctccatc cagtctatta attgttgccg ggaagctaga gtaagtagtt 1500

cgccagttaa tagtttgcgc aacgttgttg ccattgctac aggcatcgtg gtgtcacgct 1560

cgtcgtttgg tatggcttca ttcagctccg gttcccaacg atcaaggcga gttacatgat 1620

cccccatgtt gtgcaaaaaa gcggttagct ccttcggtcc tccgatcgtt gtcagaagta 1680

agttggccgc agtgttatca ctcatggtta tggcagcact gcataattct cttactgtca 1740

tgccatccgt aagatgcttt tctgtgactg gtgagtactc aaccaagtca ttctgagaat 1800

agtgtatgcg gcgaccgagt tgctcttgcc cggcgtcaat acgggataat accgcgccac 1860

atagcagaac tttaaaagtg ctcatcattg gaaaacgttc ttcggggcga aaactctcaa 1920

ggatcttacc gctgttgaga tccagttcga tgtaacccac tcgtgcaccc aactgatctt 1980

cagcatcttt tactttcacc agcgtttctg ggtgagcaaa aacaggaagg caaaatgccg 2040

caaaaaaggg aataagggcg acacggaaat gttgaatact catactcttc ctttttcaat 2100

attattgaag catttatcag ggttattgtc tcatgagcgg atacatattt gaatgtattt 2160

agaaaaataa acaaataggg gttccgcgca catttccccg aaaagtgcca cctgacgcgc 2220

cctgtagcgg cgcattaagc gcggcgggtg tggtggttac gcgcagcgtg accgctacac 2280

ttgccagcgc cctagcgccc gctcctttcg ctttcttccc ttcctttctc gccacgttcg 2340

ccggctttcc ccgtcaagct ctaaatcggg ggctcccttt agggttccga tttagtgctt 2400

tacggcacct cgaccccaaa aaacttgatt agggtgatgg ttcacgtagt gggccatcgc 2460

cctgatagac ggtttttcgc cctttgacgt tggagtccac gttctttaat agtggactct 2520

tgttccaaac tggaacaaca ctcaacccta tctcggtcta ttcttttgat ttataaggga 2580

-continued

ttttgccgat ttcggcctat tggttaaaaa atgagctgat ttaacaaaaa tttaacgcga 2640

attttaacaa aatattaacg cttacaattt ccattcgcca ttcaggctgc gcaactgttg 2700

ggaagggcga tcggtgcggg cctcttcgct attacgccag ctggcgaaag ggggatgtgc 2760

tgcaaggcga ttaagttggg taacgccagg gttttcccag tcacgacgtt gtaaaacgac 2820

ggccagtgaa ttgtaatacg actcactata gggcgaattg ggtaccgggc ccccctcga 2880

ggtcgatggt gtcgataagc ttgatatcga attcatgtca cacaaaccga tcttcgcctc 2940

aaggaaacct aattctacat ccgagagact gccgagatcc agtctacact gattaatttt 3000

cgggccaata atttaaaaaa atcgtgttat ataatattat atgtattata tatatacatc 3060

atgatgatac tgacagtcat gtcccattgc taaatagaca gactccatct gccgcctcca 3120

actgatgttc tcaatattta aggggtcatc tcgcattgtt taataataaa cagactccat 3180

ctaccgcctc caaatgatgt tctcaaaata tattgtatga acttatttt attacttagt 3240

attattagac aacttacttg ctttatgaaa aacacttcct atttaggaaa caatttataa 3300

tggcagttcg ttcatttaac aatttatgta gaataaatgt tataaatgcg tatgggaaat 3360

cttaaatatg gatagcataa atgatatctg cattgcctaa ttcgaaatca acagcaacga 3420

aaaaaatccc ttgtacaaca taaatagtca tcgagaaata tcaactatca aagaacagct 3480

attcacacgt tactattgag attattattg gacgagaatc acacactcaa ctgtctttct 3540

ctcttctaga aatacaggta caagtatgta ctattctcat tgttcatact tctagtcatt 3600

tcatcccaca tattccttgg atttctctcc aatgaatgac attctatctt gcaaattcaa 3660

caattataat aagatatacc aaagtagcgg tatagtggca atcaaaaagc ttctctggtg 3720

tgcttctcgt atttatttt attctaatga tccattaaag gtatatattt atttcttgtt 3780

atataatcct tttgtttatt acatgggctg gatacataaa ggtattttga tttaattttt 3840

tgcttaaatt caatcccccc tcgttcagtg tcaactgtaa tggtaggaaa ttaccatact 3900

tttgaagaag caaaaaaaat gaaagaaaaa aaaaatcgta tttccaggtt agacgttccg 3960

cagaatctag aatgcggtat gcggtacatt gttcttcgaa cgtaaaagtt gcgctccctg 4020

agatattgta catttttgct tttacaagta caagtacatc gtacaactat gtactactgt 4080

tgatgcatcc acaacagttt gttttgtttt tttttgtttt ttttttttct aatgattcat 4140

taccgctatg tatacctact tgtacttgta gtaagccggg ttattggcgt tcaattaatc 4200

atagacttat gaatctgcac ggtgtgcgct gcgagttact tttagcttat gcatgctact 4260

tgggtgtaat attgggatct gttcggaaat caacggatgc tcaaccgatt tcgacagtaa 4320

taatttgaat cgaatcggag cctaaaatga acccgagtat atctcataaa attctcggtg 4380

agaggtctgt gactgtcagt acaaggtgcc ttcattatgc cctcaacctt accatacctc 4440

actgaatgta gtgtacctct aaaaatgaaa tacagtgcca aaagccaagg cactgagctc 4500

gtctaacgga cttgatatac aaccaattaa aacaaatgaa aagaaataca gttctttgta 4560

tcatttgtaa caattaccct gtacaaacta aggtattgaa atcccacaat attcccaaag 4620

tccacccctt tccaaattgt catgcctaca actcatatac caagcactaa cctaccaaac 4680

accactaaaa ccccacaaaa tatatcttac cgaatataca gtaacaagct accaccacac 4740

tcgttgggtg cagtcgccag cttaaagata tctatccaca tcagccacaa ctcccttcct 4800

ttaataaacc gactacaccc ttggctattg aggttatgag tgaatatact gtagacaaga 4860

-continued

```
cactttcaag aagactgttt ccaaaacgta ccactgtcct ccactacaaa cacacccaat   4920

ctgcttcttc tagtcaaggt tgctacaccg gtaaattata aatcatcatt tcattagcag   4980

ggcagggccc tttttataga gtcttataca ctagcggacc ctgccggtag accaacccgc   5040

aggcgcgtca gtttgctcct tccatcaatg cgtcgtagaa acgacttact ccttcttgag   5100

cagctccttg accttgttgg caacaagtct ccgacctcgg aggtggagga agagcctccg   5160

atatcggcgg tagtgatacc agcctcgacg gactccttga cggcagcctc aacagcgtca   5220

ccggcgggct tcatgttaag agagaacttg agcatcatgg cggcagacag aatggtggca   5280

atggggttga ccttctgctt gccgagatcg ggggcagatc cgtgacaggg ctcgtacaga   5340

ccgaacgcct cgttggtgtc gggcagagaa gccagagagg cggagggcag cagacccaga   5400

gaaccgggga tgacggaggc ctcgtcggag atgatatcgc caaacatgtt ggtggtgatg   5460

atgataccat tcatcttgga gggctgcttg atgaggatca tggcggccga gtcgatcagc   5520

tggtggttga gctcgagctg ggggaattcg tccttgagga ctcgagtgac agtctttcgc   5580

caaagtcgag aggaggccag cacgttggcc ttgtcaagag accacacggg aagaggggag   5640

ttgtgctgaa gggccaggaa ggcggccatt cgggcaattc gctcaacctc aggaacggag   5700

taggtctcgg tgtcggaagc gacgccagat ccgtcatcct cctttcgctc tccaaagtag   5760

atacctccga cgagctctcg gacaatgatg aagtcggtgc cctcaacgtt tcggatgggg   5820

gagagatcgg cgagcttggg cgacagcagc tggcagggtc gcaggttggc gtacaggttc   5880

aggtcctttc gcagcttgag gagaccctgc tcgggtcgca cgtcggttcg tccgtcggga   5940

gtggtccata cggtgttggc agcgcctccg acagcaccga gcataataga gtcagccttt   6000

cggcagatgt cgagagtagc gtcggtgatg ggctcgccct ccttctcaat ggcagctcct   6060

ccaatgagtc ggtcctcaaa cacaaactcg gtgccggagg cctcagcaac agacttgagc   6120

accttgacgg cctcggcaat cacctcgggg ccacagaagt cgccgccgag aagaacaatc   6180

ttcttggagt cagtcttggt cttcttagtt tcgggttcca ttgtggatgt gtgtggttgt   6240

atgtgtgatg tggtgtgtgg agtgaaaatc tgtggctggc aaacgctctt gtatatatac   6300

gcacttttgc ccgtgctatg tggaagacta aacctccgaa gattgtgact caggtagtgc   6360

ggtatcggct agggacccaa accttgtcga tgccgatagc gctatcgaac gtaccccagc   6420

cggccgggag tatgtcggag gggacatacg agatcgtcaa gggtttgtgg ccaactggta   6480

aataaatgat gtcgactcag gcgacgacgg aattcctgca gcccatctgc agaattcagg   6540

agagaccggg ttggcggcgt atttgtgtcc caaaaaacag ccccaattgc cccaattgac   6600

cccaaattga cccagtagcg ggcccaaccc cggcgagagc ccccttcacc ccacatatca   6660

aacctccccc ggttcccaca cttgccgtta agggcgtagg gtactgcagt ctggaatcta   6720

cgcttgttca gactttgtac tagtttcttt gtctggccat ccgggtaacc catgccggac   6780

gcaaaataga ctactgaaaa ttttttgct ttgtggttgg gactttagcc aagggtataa   6840

aagaccaccg tccccgaatt acctttcctc ttcttttctc tctctccttg tcaactcaca   6900

cccgaaatcg ttaagcattt ccttctgagt ataagaatca ttcaccatgg atggtacgtc   6960

ctgtagaaac cccaacccgt gaaatcaaaa aactcgacgg cctgtgggca ttcagtctgg   7020

atcgcgaaaa ctgtggaatt gatcagcgtt ggtgggaaag cgcgttacaa gaaagccggg   7080

caattgctgt gccaggcagt tttaacgatc agttcgccga tgcagatatt cgtaattatg   7140
```

-continued

cgggcaacgt ctggtatcag cgcgaagtct ttataccgaa aggttgggca ggccagcgta 7200

tcgtgctgcg tttcgatgcg gtcactcatt acggcaaagt gtgggtcaat aatcaggaag 7260

tgatggagca tcagggcggc tatacgccat ttgaagccga tgtcacgccg tatgttattg 7320

ccgggaaaag tgtacgtatc accgtttgtg tgaacaacga actgaactgg cagactatcc 7380

cgccgggaat ggtgattacc gacgaaaacg gcaagaaaaa gcagtcttac ttccatgatt 7440

tctttaacta tgccgggatc catcgcagcg taatgctcta caccacgccg aacacctggg 7500

tggacgatat caccgtggtg acgcatgtcg cgcaagactg taaccacgcg tctgttgact 7560

ggcaggtggt ggccaatggt gatgtcagcg ttgaactgcg tgatgcggat caacaggtgg 7620

ttgcaactgg acaaggcact agcgggactt tgcaagtggt gaatccgcac ctctggcaac 7680

cgggtgaagg ttatctctat gaactgtgcg tcacagccaa aagccagaca gagtgtgata 7740

tctacccgct tcgcgtcggc atccggtcag tggcagtgaa gggcgaacag ttcctgatta 7800

accacaaacc gttctacttt actggctttg gtcgtcatga agatgcggac ttacgtggca 7860

aaggattcga taacgtgctg atggtgcacg accacgcatt aatggactgg attggggcca 7920

actcctaccg tacctcgcat tacccttacg ctgaagagat gctcgactgg gcagatgaac 7980

atggcatcgt ggtgattgat gaaactgctg ctgtcggctt taacctctct ttaggcattg 8040

gtttcgaagc gggcaacaag ccgaaagaac tgtacagcga agaggcagtc aacggggaaa 8100

ctcagcaagc gcacttacag gcgattaaag agctgatagc gcgtgacaaa aaccacccaa 8160

gcgtggtgat gtggagtatt gccaacgaac cggatacccg tccgcaagtg cacgggaata 8220

tttcgccact ggcggaagca acgcgtaaac tcgacccgac gcgtccgatc acctgcgtca 8280

atgtaatgtt ctgcgacgct cacaccgata ccatcagcga tctctttgat gtgctgtgcc 8340

tgaaccgtta ttacggatgg tatgtccaaa gcggcgattt ggaaacggca gagaaggtac 8400

tggaaaaaga acttctggcc tggcaggaga aactgcatca gccgattatc atcaccgaat 8460

acggcgtgga tacgttagcc gggctgcact caatgtacac cgacatgtgg agtgaagagt 8520

atcagtgtgc atggctggat atgtatcacc gcgtctttga tcgcgtcagc gccgtcgtcg 8580

gtgaacaggt atggaatttc gccgattttg cgacctcgca aggcatattg cgcgttggcg 8640

gtaacaagaa agggatcttc actcgcgacc gcaaaccgaa gtcggcggct tttctgctgc 8700

aaaaacgctg gactggcatg aacttcggtg aaaaaccgca gcagggaggc aaacaatgat 8760

taattaacta gagcggccgc caccgcggcc cgagattccg gcctcttcgg ccgccaagcg 8820

acccgggtgg acgtctagag gtacctagca attaacagat agtttgccgg tgataattct 8880

cttaacctcc cacactcctt tgacataacg atttatgtaa cgaaactgaa atttgaccag 8940

atattgtgtc cgc 8953

<210> SEQ ID NO 11
<211> LENGTH: 973
<212> TYPE: DNA
<213> ORGANISM: Yarrowia lipolytica
<220> FEATURE:
<221> NAME/KEY: misc_feature
<223> OTHER INFORMATION: Promoter FBAIN

<400> SEQUENCE: 11

aaattgcccc ggagaagacg gccaggccgc ctagatgaca aattcaacaa ctcacagctg 60

-continued

```
actttctgcc attgccacta gggggggcc tttttatatg gccaagccaa gctctccacg    120

tcggttgggc tgcacccaac aataaatggg tagggttgca ccaacaaagg gatgggatgg    180

ggggtagaag atacgaggat aacggggctc aatggcacaa ataagaacga atactgccat    240

taagactcgt gatccagcga ctgacaccat tgcatcatct aagggcctca aaactacctc    300

ggaactgctg cgctgatctg gacaccacag aggttccgag cactttaggt tgcaccaaat    360

gtcccaccag gtgcaggcag aaaacgctgg aacagcgtgt acagtttgtc ttaacaaaaa    420

gtgagggcgc tgaggtcgag cagggtggtg tgacttgtta tagcctttag agctgcgaaa    480

gcgcgtatgg atttggctca tcaggccaga ttgagggtct gtggacacat gtcatgttag    540

tgtacttcaa tcgccccctg gatatagccc cgacaatagg ccgtggcctc atttttttgc    600

cttccgcaca tttccattgc tcggtaccca caccttgctt ctcctgcact tgccaacctt    660

aatactggtt tacattgacc aacatcttac aagcgggggg cttgtctagg gtatatataa    720

acagtggctc tcccaatcgg ttgccagtct ctttttcct ttctttcccc acagattcga    780

aatctaaact acacatcaca caatgcctgt tactgacgtc cttaagcgaa agtccggtgt    840

catcgtcggc gacgatgtcc gagccgtgag tatccacgac aagatcagtg tcgagacgac    900

gcgttttgtg taatgacaca atccgaaagt cgctagcaac acacactctc tacacaaact    960

aacccagctc tcc                                                      973
```

<210> SEQ ID NO 12
<211> LENGTH: 727
<212> TYPE: PRT
<213> ORGANISM: Yarrowia lipolytica (ORF YALI-CDS1055.1 in the Genolevures database of Y. lipolytica ORFs (sponsored by the Center for Bioinformatics; http://cbi.labri.fr/Genolevures/Y_lipolytica.php)

<400> SEQUENCE: 12

```
Met Ser Glu Thr Asp His Leu Leu Ala Ala Glu Pro Val Ala Glu Tyr
1               5                   10                  15

Pro Gln Tyr Thr Pro Trp Pro Asn Ser Arg Lys Ser Val Asp Thr Glu
            20                  25                  30

Phe Ser Ala Thr Ser Trp Ile Tyr Asp Leu Val Leu Trp Ile Phe Thr
        35                  40                  45

Ala Cys Phe Asp Ile Phe Phe Arg Glu Ile Arg Pro Arg Gly Ala Phe
    50                  55                  60

Arg Ile Pro Arg Lys Gly Pro Val Leu Phe Val Ala Ala Pro His Ala
65                  70                  75                  80

Asn Gln Phe Val Asp Pro Val Ile Leu Met Asn Gln Val Lys Gln Glu
                85                  90                  95

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

Ala Ala Val Gly Arg Met Ala Arg Ser Met Asn Ser Ile Pro Val Val
        115                 120                 125

Arg Ala Gln Asp Asn Ala Lys Lys Gly Glu Gly Lys Ile Tyr Val Asp
    130                 135                 140

Ala Glu Asp Pro Thr Lys Ile His Gly Ile Gly Thr Gln Phe Thr Lys
145                 150                 155                 160

Gln Cys Glu Val Arg Gly Leu Val Val Cys Ser Ser Ser Val Gly Ser
                165                 170                 175
```

-continued

```
Ile Asp Val Ala Glu Ile Val Ser Asp Thr Leu Leu Ile Ala Arg Lys
            180                 185                 190

Glu Phe Lys Gly Pro Lys Ala Lys Glu Ala Leu Lys Glu Ser Asn Gly
        195                 200                 205

Gly Ile Thr Tyr Lys Tyr Ala Asp Tyr Val Asn Gln Ala Thr Val Tyr
    210                 215                 220

Arg Ser Val Phe Asp Lys Leu His His Gly Gly Cys Val Gly Ile Phe
225                 230                 235                 240

Pro Glu Gly Gly Ser His Asp Arg Thr Glu Leu Leu Pro Leu Lys Ala
                245                 250                 255

Gly Val Ala Ile Met Ala Leu Gly Ala Leu Ala Glu Asp Pro Ser Cys
            260                 265                 270

Gly Val Arg Ile Val Pro Cys Gly Leu Asn Tyr Phe His Ala Tyr Lys
        275                 280                 285

Phe Arg Ser Arg Ala Val Val Glu Phe Gly Ser Pro Ile Ala Ile Pro
    290                 295                 300

Pro Asp Leu Val Glu Lys Tyr Lys Ala Gly Gly Glu Ala Lys Arg Glu
305                 310                 315                 320

Ala Val Lys Thr Val Leu Asp Ile Thr Ala Ala Gly Leu Lys Ser Val
                325                 330                 335

Thr Val Gln Val Gln Asp Phe Asp Thr Leu Met Leu Ile Gln Ala Ile
            340                 345                 350

Arg Arg Leu Tyr Arg Pro Pro Gly Lys Lys Ile Pro Leu Pro Met Val
        355                 360                 365

Val Glu Leu Asn Arg Arg Leu Val Tyr Ala Tyr Asn His Tyr Lys Asp
    370                 375                 380

Asp Pro Arg Ile Glu Glu Met Lys Gln Glu Ile Arg Lys Tyr Asn Lys
385                 390                 395                 400

Phe Leu Gln Ala Met Gly Leu Lys Asp His Gln Val Glu Lys Ala Arg
                405                 410                 415

Ile Ser Lys Ile Glu Ile Leu Gly Arg Leu Leu Tyr Arg Ser Ile Lys
            420                 425                 430

Leu Val Phe Leu Ser Ile Gly Cys Leu Pro Gly Leu Leu Leu Phe Ser
        435                 440                 445

Pro Ile Phe Ile Ile Ser Lys Ser Ile Ser Lys Thr Lys Ala Lys Glu
    450                 455                 460

Ala Leu Lys Ala Ser Ser Val Lys Ile Lys Ala Asn Asp Val Val Ala
465                 470                 475                 480

Thr Trp Lys Val Leu Val Ala Met Gly Leu Thr Pro Val Leu Tyr Ile
                485                 490                 495

Leu Tyr Ser Leu Val Gly Ser Val Val Ile Arg Lys Leu Asp Leu Ile
            500                 505                 510

Ser Trp Phe Pro Thr Ile Leu Leu Pro Gly Leu Val Leu Ser Ile Ile
        515                 520                 525

Ile Thr Thr Ser Tyr Ala Ala Leu Ala Met Gly Glu Ala Gly Met Asp
    530                 535                 540

Ile Phe Lys Ser Leu Arg Pro Leu Ala Leu Ala Leu Asn Pro Ser Thr
545                 550                 555                 560

Lys Asn Ser Leu Leu Lys Leu Gln Asn Glu Arg Lys Arg Leu Val Leu
                565                 570                 575

Lys Ser Ser Glu Leu Val Thr Ser Leu Gly Pro Glu Leu Phe Pro Asp
```

-continued

```
            580                 585                 590

Phe Pro Glu Asn Ser Ile Leu Gln Gly Ser Asp Lys Phe Glu Asp Glu
        595                 600                 605

Glu Asn Tyr Glu Asn Glu Lys Arg Ser His Ser Arg Ser Thr Ser Ala
    610                 615                 620

Thr Ser Leu Ser Ala Met Ser Glu Gly Asp Gly Asp Glu Leu Val Arg
625                 630                 635                 640

Glu Val Arg Lys Gly Ala Ser Tyr Phe Pro Val Ser Thr Thr Ser Glu
                645                 650                 655

Asp Glu Asp Gln Ala Ile Ser Arg Val Gly Ser Glu Ala Ser Leu Ala
            660                 665                 670

Asp Ile Pro Leu Phe Gly Met Ser Arg Ser Gln Ser Gly Ala Ser Leu
        675                 680                 685

Ser Glu Ala Ser Thr His Gly Ser Ser Thr Gly Ala Asp Ala Glu Glu
    690                 695                 700

Ala Lys Thr Glu Val Thr Arg Arg Ile Ala Leu Ala Met Glu Glu Lys
705                 710                 715                 720

Arg Arg Glu Gln Asp Glu Glu
                725
```

<210> SEQ ID NO 13
<211> LENGTH: 24
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Primer YGPAT-F
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (6)..(6)
<223> OTHER INFORMATION: n is a, c, g, or t
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (21)..(21)
<223> OTHER INFORMATION: n is a, c, g, or t
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (24)..(24)
<223> OTHER INFORMATION: n is a, c, g, or t

<400> SEQUENCE: 13

atgtcngaga cygaccayct nctn 24

<210> SEQ ID NO 14
<211> LENGTH: 24
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Primer YGPAT-R

<400> SEQUENCE: 14

ytcytcrtcy tgytctcgyc gytt 24

<210> SEQ ID NO 15
<211> LENGTH: 2184
<212> TYPE: DNA
<213> ORGANISM: Yarrowia lipolytica

<400> SEQUENCE: 15

atgtccgaaa ccgaccatct gctggccgcc gagcccgtgg ctgagtaccc ccagtacacg 60

ccttggccca actcccgaaa atcagtggac acggagtttt ccgcaacctc gtggatttac 120

-continued

gacttggttc tgtggatttt cacggcttgc tttgacattt ttttcagaga aatccggcca 180

cgtggtgcct tccgaatccc cagaaagggc cccgtgctgt tcgtggctgc cccccacgca 240

aaccagtttg tggaccccgt catcctcatg aaccaggtca aacaggaggc cggacgacga 300

atctccttcc ttgtggccga gaagtccatg cgacgagctg cagtcggacg aatggcccga 360

agcatgaact caattcctgt cgtgcgagct caggacaatg caaaaaaggg agagggaaag 420

atttacgtcg acgcagagga ccccacaaag atccacggaa tcggcaccca gttcacgaag 480

cagtgcgagg tgcgaggcct cgtggtctgc tcgtcctctg tcggctcaat tgacgtggct 540

gagattgtgt ccgacactct gctcattgca agaaaggaat tcaagggccc caaagccaag 600

gaggctctca aggaatccaa cggaggaatc acatacaagt acgccgacta cgtcaaccag 660

gccacagtct accgatccgt gttcgacaaa ttgcaccatg gaggctgtgt gggtatcttc 720

ccagagggag gatctcatga ccgaaccgaa ctgctgcccc ttaaggccgg tgttgctatc 780

atggctctgg gggctctcgc agaggacccc tcttgtggtg tgcgaatcgt cccctgtggt 840

ctcaactact tccacgccta caagttccga tctcgggccg tggtggagtt tggctctcct 900

attgccattc ctccggatct cgtggagaag tacaaggcag gaggagaggc caagcgggag 960

gctgtcaaga ccgttctaga cattactgcc gctggtctca agtctgtgac tgttcaggtg 1020

caggatttcg acaccctgat gctgatccag gccattcgac gactctaccg acctcccgga 1080

aagaagattc ctctgcccat ggttgtagag ctcaaccgtc gacttgtata cgcctacaac 1140

cactacaagg acgatccccg tatcgaggag atgaagcagg agattcgaaa gtacaacaag 1200

ttcctgcagg ccatgggtct caaggaccat caggtagaga aggcccgaat ctccaagatt 1260

gagattctgg gccggcttct gtaccggtcc atcaagcttg tgttcttgtc cattggctgt 1320

ctccccggtc tgcttttgtt ttctcccatc ttcatcattt ctaagtccat ttccaaaacc 1380

aaggccaagg aggctctcaa ggcctccagt gtcaaaatca aggctaacga tgtggttgcc 1440

acttggaagg tgctggttgc aatgggtctg accccagttc tttacattct ctattcactg 1500

gttggatctg tggtgattcg aaagctcgat ctcatctcct ggttccccac aattcttctt 1560

cccggcctcg ttttaagcat catcatcaca acctcatacg ccgccctggc tatgggagag 1620

gccggtatgg acattttcaa gtctcttcga ccacttgcat tggctctcaa cccttccacc 1680

aaaaactctc tgctcaagct gcaaaatgaa cgaaagcgac ttgtgctcaa gtcttccgag 1740

ctcgttacct ctttgggccc tgagctgttc cccgacttcc ccgagaactc cattctgcag 1800

ggaagcgata agtttgagga cgaggagaac tacgaaaacg agaagcgatc gcattccaga 1860

tccacttctg ccacttctct atctgccatg agcgagggag acggtgatga gcttgttcgg 1920

gaggtccgaa agggtgctag ctacttccct gtgagtacca cttctgagga cgaagaccaa 1980

gccatctcgc gagtgggctc tgaggcatct cttgctgaca ttcctctgtt tggtatgtcc 2040

cgatcacaat ctggagcttc tctttcggaa gcctccacac acggctcttc tactggagct 2100

gatgccgagg aggctaagac ggaggtgact cgcagaattg cattggcgat ggaggaaaaa 2160

cgacgagagc aggatgagga ataa 2184

<210> SEQ ID NO 16
<211> LENGTH: 44
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence

-continued

<220> FEATURE:
<223> OTHER INFORMATION: Genome Walker adaptor-1

<400> SEQUENCE: 16

gtaatacgac tatagggcac gcgtggtcga cggcccgggc tggt 44

<210> SEQ ID NO 17
<211> LENGTH: 8
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Genome Walker adaptor-2
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(1)
<223> OTHER INFORMATION: 5' end is associated with a -PO4 group
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (8)..(8)
<223> OTHER INFORMATION: 3' end is associated with a -H2N group

<400> SEQUENCE: 17

accagccc 8

<210> SEQ ID NO 18
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Primer Adaptor-1

<400> SEQUENCE: 18

gtaatacgac tcactatagg gc 22

<210> SEQ ID NO 19
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Primer YGPAT-5R-1

<400> SEQUENCE: 19

gtcaaagcaa gcccgtgaaa atc 23

<210> SEQ ID NO 20
<211> LENGTH: 19
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Nested Adaptor Primer 2

<400> SEQUENCE: 20

actatagggc acgcgtggt 19

<210> SEQ ID NO 21
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Primer YGPAT-5R-2

<400> SEQUENCE: 21

gtcgtaaatc cacgaggttg 20

-continued

<210> SEQ ID NO 22
<211> LENGTH: 1781
<212> TYPE: DNA
<213> ORGANISM: Yarrowia lipolytica

<400> SEQUENCE: 22

atccagacga cggcatggga gaacggattg gtttttttt gagttcatca tgtcctgcca 60

acccactgaa aacagggtcc gatgcggaat gatggctttt tactgtagtc agtgtctttt 120

acagaaatac acccccagcc atcatatatc ctctccaggc atggaccact gggtacactc 180

gtgacgtgtc ataatcactg agccagtata cgaacaaaat acaatcctgc catcgtactc 240

actcaaacaa ccgtgttcaa taaggaagct gggccagaca gctgaaatgt accaaccagt 300

cgaggtgaag cacagaacaa gaagacacaa accacgatct ttacaagaat gtactgtagt 360

acatacagca cgagtagtgt acgaagtttt tgtcagacac gtgacttgct cgtaaattag 420

ccaatatgac acaccgccaa ggacctcgga tgcgtgcggt gcgtgcggtg cgtgcggtgt 480

gtgcggtgcg tgcgtatgta ccccactctg tgtcgatcag ccgacagtga tgttccggaa 540

gtgcggtaca acttttcttg tcgacctgag ataccgaggt tgcgcagggg atcaactttt 600

gtgtctcaga gggacccaag tgcgtacgga gagtacagta catactgtag ctaacggtag 660

caggcgaact actggtacat acctccccg gaatatgtac aggcataatg cgtatctgtg 720

ggacatgtgg tcgttgcgcc attatgtaag cagcgtgtac tcctctgact gtccatatgg 780

tttgctccat ctcaccctca tcgttttcat tgttcacagg cggccacaaa aaaactgtct 840

tctctccttc tctcttcgcc ttagtctact cggaccagtt ttagtttagc ttggcgccac 900

tggataaatg agacctcagg ccttgtgatg aggaggtcac ttatgaagca tgttaggagg 960

tgcttgtatg gatagagaag cacccaaaat aataagaata ataataaaac agggggcgtt 1020

gtcatttcat atcgtgtttt caccatcaat acacctccaa acaatgccct tcatgtggcc 1080

agccccaata ttgtcctgta gttcaactct atgcagctcg tatcttattg agcaagtaaa 1140

actctgtcag ccgatattgc ccgacccgcg acaagggtca acaaggtggt gtaaggcctt 1200

cgcagaagtc aaaactgtgc caaacaaaca tctagagtct ctttggtgtt tctcgcatat 1260

atttaatcgg ctgtcttacg tatttggcct cggtaccgga ctaatttcgg atcatcccca 1320

atacgctttt tcttcgcagc tgtcaacagt gtccatgatc tatccaccta aatgggtcat 1380

atgaggcgta taatttcgtg gtgctgataa taattcccat atatttgaca caaaacttcc 1440

ccccctagac atacatctca caatctcact tcttgtgctt ctgtcacaca tctcctccag 1500

ctgacttcaa ctcacacctc tgccccagtt ggtctacagc ggtataaggt ttctccgcat 1560

agaggtgcac cactcctccc gatacttgtt tgtgtgactt gtgggtcacg acatatatat 1620

ctacacacat tgcgccaccc tttggttctt ccagcacaac aaaaacacga cacgctaaat 1680

gtccgaaacc gaccatctgc tggccgccga gcccgtggct gagtaccccc agtacacgcc 1740

ttggcccaac tcccgaaaat cagtggacac ggagttttcc g 1781

<210> SEQ ID NO 23
<211> LENGTH: 3862
<212> TYPE: DNA
<213> ORGANISM: Yarrowia lipolytica

<400> SEQUENCE: 23

atccagacga cggcatggga gaacggattg gtttttttt gagttcatca tgtcctgcca 60

-continued

```
acccactgaa aacagggtcc gatgcggaat gatggctttt tactgtagtc agtgtctttt    120
acagaaatac acccccagcc atcatatatc ctctccaggc atggaccact gggtacactc    180
gtgacgtgtc ataatcactg agccagtata cgaacaaaat acaatcctgc catcgtactc    240
actcaaacaa ccgtgttcaa taaggaagct gggccagaca gctgaaatgt accaaccagt    300
cgaggtgaag cacagaacaa gaagacacaa accacgatct ttacaagaat gtactgtagt    360
acatacagca cgagtagtgt acgaagtttt tgtcagacac gtgacttgct cgtaaattag    420
ccaatatgac acaccgccaa ggacctcgga tgcgtgcggt gcgtgcggtg cgtgcggtgt    480
gtgcggtgcg tgcgtatgta ccccactctg tgtcgatcag ccgacagtga tgttccggaa    540
gtgcggtaca acttttcttg tcgacctgag ataccgaggt tgcgcagggg atcaactttt    600
gtgtctcaga gggacccaag tgcgtacgga gagtacagta catactgtag ctaacggtag    660
caggcgaact actggtacat acctccccccg gaatatgtac aggcataatg cgtatctgtg    720
ggacatgtgg tcgttgcgcc attatgtaag cagcgtgtac tcctctgact gtccatatgg    780
tttgctccat ctcaccctca tcgttttcat tgttcacagg cggccacaaa aaaactgtct    840
tctctccttc tctcttcgcc ttagtctact cggaccagtt ttagtttagc ttggcgccac    900
tggataaatg agacctcagg ccttgtgatg aggaggtcac ttatgaagca tgttaggagg    960
tgcttgtatg gatagagaag cacccaaaat aataagaata ataataaaac agggggcgtt   1020
gtcatttcat atcgtgtttt caccatcaat acacctccaa acaatgccct tcatgtggcc   1080
agccccaata ttgtcctgta gttcaactct atgcagctcg tatcttattg agcaagtaaa   1140
actctgtcag ccgatattgc ccgacccgcg acaagggtca acaaggtggt gtaaggcctt   1200
cgcagaagtc aaaactgtgc caaacaaaca tctagagtct ctttggtgtt tctcgcatat   1260
atttaatcgg ctgtcttacg tatttggcct cggtaccgga ctaatttcgg atcatcccca   1320
atacgctttt tcttcgcagc tgtcaacagt gtccatgatc tatccaccta aatgggtcat   1380
atgaggcgta taatttcgtg gtgctgataa taattcccat atatttgaca caaaacttcc   1440
ccccctagac atacatctca caatctcact tcttgtgctt ctgtcacaca tctcctccag   1500
ctgacttcaa ctcacacctc tgccccagtt ggtctacagc ggtataaggt ttctccgcat   1560
agaggtgcac cactcctccc gatacttgtt tgtgtgactt gtgggtcacg acatatatat   1620
ctacacacat tgcgccaccc tttggttctt ccagcacaac aaaaacacga cacgctaaat   1680
gtccgaaacc gaccatctgc tggccgccga gcccgtggct gagtaccccc agtacacgcc   1740
ttggcccaac tcccgaaaat cagtggacac ggagttttcc gcaacctcgt ggatttacga   1800
cttggttctg tggattttca cggcttgctt tgacattttt ttcagagaaa tccggccacg   1860
tggtgccttc cgaatcccca gaaagggccc cgtgctgttc gtggctgccc cccacgcaaa   1920
ccagtttgtg gaccccgtca tcctcatgaa ccaggtcaaa caggaggccg gacgacgaat   1980
ctccttcctt gtggccgaga agtccatgcg acgagctgca gtcggacgaa tggcccgaag   2040
catgaactca attcctgtcg tgcgagctca ggacaatgca aaaaagggag agggaaagat   2100
ttacgtcgac gcagaggacc ccacaaagat ccacggaatc ggcacccagt tcacgaagca   2160
gtgcgaggtg cgaggcctcg tggtctgctc gtcctctgtc ggctcaattg acgtggctga   2220
gattgtgtcc gacactctgc tcattgcaag aaaggaattc aagggcccca aagccaagga   2280
ggctctcaag gaatccaacg gaggaatcac atacaagtac gccgactacg tcaaccaggc   2340
```

-continued

cacagtctac cgatccgtgt tcgacaaatt gcaccatgga ggctgtgtgg gtatcttccc 2400

agagggagga tctcatgacc gaaccgaact gctgcccctt aaggccggtg ttgctatcat 2460

ggctctgggg gctctcgcag aggacccctc ttgtggtgtg cgaatcgtcc cctgtggtct 2520

caactacttc cacgcctaca agttccgatc tcgggccgtg gtggagtttg gctctcctat 2580

tgccattcct ccggatctcg tggagaagta caaggcagga ggagaggcca agcgggaggc 2640

tgtcaagacc gttctagaca ttactgccgc tggtctcaag tctgtgactg ttcaggtgca 2700

ggatttcgac accctgatgc tgatccaggc cattcgacga ctctaccgac ctcccggaaa 2760

gaagattcct ctgcccatgg ttgtagagct caaccgtcga cttgtatacg cctacaacca 2820

ctacaaggac gatccccgta tcgaggagat gaagcaggag attcgaaagt acaacaagtt 2880

cctgcaggcc atgggtctca aggaccatca ggtagagaag gcccgaatct ccaagattga 2940

gattctgggc cggcttctgt accggtccat caagcttgtg ttcttgtcca ttggctgtct 3000

ccccggtctg cttttgtttt ctcccatctt catcatttct aagtccattt ccaaaaccaa 3060

ggccaaggag gctctcaagg cctccagtgt caaaatcaag gctaacgatg tggttgccac 3120

ttggaaggtg ctggttgcaa tgggtctgac cccagttctt tacattctct attcactggt 3180

tggatctgtg gtgattcgaa agctcgatct catctcctgg ttccccacaa ttcttcttcc 3240

cggcctcgtt ttaagcatca tcatcacaac ctcatacgcc gccctggcta tgggagaggc 3300

cggtatggac attttcaagt ctcttcgacc acttgcattg gctctcaacc cttccaccaa 3360

aaactctctg ctcaagctgc aaaatgaacg aaagcgactt gtgctcaagt cttccgagct 3420

cgttacctct ttgggccctg agctgttccc cgacttcccc gagaactcca ttctgcaggg 3480

aagcgataag tttgaggacg aggagaacta cgaaaacgag aagcgatcgc attccagatc 3540

cacttctgcc acttctctat ctgccatgag cgagggagac ggtgatgagc ttgttcggga 3600

ggtccgaaag ggtgctagct acttccctgt gagtaccact tctgaggacg aagaccaagc 3660

catctcgcga gtgggctctg aggcatctct tgctgacatt cctctgtttg gtatgtcccg 3720

atcacaatct ggagcttctc tttcggaagc ctccacacac ggctcttcta ctggagctga 3780

tgccgaggag gctaagacgg aggtgactcg cagaattgca ttggcgatgg aggaaaaacg 3840

acgagagcag gatgaggaat aa 3862

<210> SEQ ID NO 24
<211> LENGTH: 1130
<212> TYPE: DNA
<213> ORGANISM: Yarrowia lipolytica
<220> FEATURE:
<221> NAME/KEY: misc_feature
<223> OTHER INFORMATION: GPAT promoter

<400> SEQUENCE: 24

caacttttct tgtcgacctg agataccgag gttgcgcagg ggatcaactt ttgtgtctca 60

gagggaccca agtgcgtacg gagagtacag tacatactgt agctaacggt agcaggcgaa 120

ctactggtac atacctcccc cggaatatgt acaggcataa tgcgtatctg tgggacatgt 180

ggtcgttgcg ccattatgta agcagcgtgt actcctctga ctgtccatat ggtttgctcc 240

atctcaccct catcgttttc attgttcaca ggcggccaca aaaaaactgt cttctctcct 300

tctctcttcg ccttagtcta ctcggaccag ttttagttta gcttggcgcc actggataaa 360

-continued

tgagacctca ggccttgtga tgaggaggtc acttatgaag catgttagga ggtgcttgta 420

tggatagaga agcacccaaa ataataagaa taataataaa acaggggcg ttgtcatttc 480

atatcgtgtt ttcaccatca atacacctcc aaacaatgcc cttcatgtgg ccagccccaa 540

tattgtcctg tagttcaact ctatgcagct cgtatcttat tgagcaagta aaactctgtc 600

agccgatatt gcccgacccg cgacaagggt caacaaggtg gtgtaaggcc ttcgcagaag 660

tcaaaactgt gccaaacaaa catctagagt ctctttggtg tttctcgcat atatttaatc 720

ggctgtctta cgtatttggc ctcggtaccg gactaatttc ggatcatccc caatacgctt 780

tttcttcgca gctgtcaaca gtgtccatga tctatccacc taaatgggtc atatgaggcg 840

tataatttcg tggtgctgat aataattccc atatatttga cacaaaactt ccccccctag 900

acatacatct cacaatctca cttcttgtgc ttctgtcaca catctcctcc agctgacttc 960

aactcacacc tctgccccag ttggtctaca gcggtataag gtttctccgc atagaggtgc 1020

accactcctc ccgatacttg tttgtgtgac ttgtgggtca cgacatatat atctacacac 1080

attgcgccac cctttggttc ttccagcaca acaaaaacac gacacgctaa 1130

<210> SEQ ID NO 25
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Primer GPAT-5-1

<400> SEQUENCE: 25

caacttttct tgtcgacctg ag 22

<210> SEQ ID NO 26
<211> LENGTH: 34
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Primer GPAT-5-2

<400> SEQUENCE: 26

gatcccatgg ttagcgtgtc gtgtttttgt tgtg 34

<210> SEQ ID NO 27
<211> LENGTH: 12649
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Plasmid pKUNF12T6E
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (2507)..(2507)
<223> OTHER INFORMATION: n is a, c, g, or t
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (2512)..(2515)
<223> OTHER INFORMATION: n is a, c, g, or t

<400> SEQUENCE: 27

taaccctcac taaagggaac aaaagctgga gctccaccgc ggacacaata tctggtcaaa 60

tttcagtttc gttacataaa tcgttatgtc aaaggagtgt gggaggttaa gagaattatc 120

accggcaaac tatctgttaa ttgctaggta cctctagacg tccacccggg tcgcttggcg 180

gccgaagagg ccggaatctc gggccgcggt ggcggccgct tagttggtct tggacttctt 240

-continued

gggcttcttc aggtaggact ggacaaagaa gttgccgaac agagcgagca gggtgatcat 300

gtacacgccg agcagctgga ccagagcctg agggtagtcg caggggaaga ggtagtcgta 360

cagggactgc accagcatag ccatgaactg ggtcatctgc agagtggtga tgtagggctt 420

gatgggcttg acgaagccga agccctgaga ggaaaagaag tagtaggcgt acatgacggt 480

gtggacgaag gagttgagga tgacggagaa gtaggcgtcg ccaccaggag cgtacttggc 540

aatagcccac cagatggcga agatggtggc atggtggtac acgtgcagga aggagacctg 600

gttgaacttc ttgcacagga tcatgatagc ggtgtccagg aactcgtagg ccttggagac 660

gtagaacacg tagacgattc gggacatgcc ctgagcgtgg gactcgttgc ccttctccat 720

gtcgttgccg aagaccttgt agccacccag gatagcctgt cggatggtct cgacgcacat 780

gtagagggac agtccgaaga ggaacaggtt gtggagcagc ttgatggtct tcagctcgaa 840

gggcttctcc atctgcttca tgatgggaat gccgaagagc agcatggcca tgtagccgac 900

ctcgaaggcg agcatggtgg agacgtccat catgggcaga ccgtcggtca gagcgtaggg 960

cttagctccg tccatccact ggtcgacacc ggtctcgact cgtccgacca cgtcgtccca 1020

gacagaggag ttggccatgg tgaatgattc ttatactcag aaggaaatgc ttaacgattt 1080

cgggtgtgag ttgacaagga gagagagaaa agaagaggaa aggtaattcg gggacggtgg 1140

tcttttatac ccttggctaa agtcccaacc acaaagcaaa aaaattttca gtagtctatt 1200

ttgcgtccgg catgggttac ccggatggcc agacaaagaa actagtacaa agtctgaaca 1260

agcgtagatt ccagactgca gtaccctacg cccttaacgg caagtgtggg aaccgggga 1320

ggtttgatat gtggggtgaa gggggctctc gccgggggttg ggcccgctac tgggtcaatt 1380

tggggtcaat tggggcaatt ggggctgttt tttgggacac aaatacgccg ccaacccggt 1440

ctctcctgaa ttctgcatcg atcgaggaag aggacaagcg gctgcttctt aagtttgtga 1500

catcagtatc caaggcacca ttgcaaggat tcaaggcttt gaacccgtca tttgccattc 1560

gtaacgctgg tagacaggtt gatcggttcc ctacggcctc cacctgtgtc aatcttctca 1620

agctgcctga ctatcaggac attgatcaac ttcggaagaa acttttgtat gccattcgat 1680

cacatgctgg tttcgatttg tcttagagga acgcatatac agtaatcata gagaataaac 1740

gatattcatt tattaaagta gatagttgag gtagaagttg taaagagtga taaatagcgg 1800

ccgcgcctac ttaagcaacg ggcttgataa cagcgggggg ggtgcccacg ttgttgcggt 1860

tgcggaagaa cagaacaccc ttaccagcac cctcggcacc agcgctgggc tcaacccact 1920

ggcacatacg cgcactgcgg tacatggcgc ggatgaagcc acgaggacca tcctggacat 1980

cagcccggta gtgcttgccc atgatgggct taatggcctc ggtggcctcg tccgcgttgt 2040

agaaggggat gctgctgacg tagtggtgga ggacatgagt ctcgatgatg ccgtggagaa 2100

ggtggcggcc gatgaagccc atctcacggt caatggtagc agcggcacca cggacgaagt 2160

tccactcgtc gttggtgtag tggggaaggg tagggtcggt gtgctggagg aaggtgatgg 2220

caacgagcca gtggttaacc cagaggtagg gaacaaagta ccagatggcc atgttgtaga 2280

aaccgaactt ctgaacgagg aagtacagag cagtggccat cagaccgata ccaatatcgc 2340

tgaggacgat gagcttagcg tcactgttct cgtacagagg gctgcgggga tcgaagtggt 2400

taacaccacc gccgaggccg ttatgcttgc ccttgccgcg accctcacgc tggcgctcgt 2460

ggtagttgtg gccggtaaca ttggtgatga ggtagttggg ccagccnacg annnnctcag 2520

-continued

taagatgagc gagctcgtgg gtcatctttc cgagacgagt agcctgctgc tcgcgggttc 2580

ggggaacgaa gaccatgtca cgctccatgt tgccagtggc cttgtggtgc tttcggtggg 2640

agatttgcca gctgaagtag gggacaagga gggaagagtg aagaacccag ccagtaatgt 2700

cgttgatgat gcgagaatcg gagaaagcac cgtgaccgca ctcatgggca ataacccaga 2760

gaccagtacc gaaaagaccc tgaagaacgg tgtacacggc ccacagacca gcgcgggcgg 2820

gggtggaggg gatatattcg ggggtcacaa agttgtacca gatgctgaaa gtggtagtca 2880

ggaggacaat gtcgcggagg atataaccgt atcccttgag agcggagcgc ttgaagcagt 2940

gcttagggat ggcattgtag atgtccttga tggtaaagtc gggaacctcg aactggttgc 3000

cgtaggtgtc gagcatgaca ccatactcgg acttgggctt ggcgatatca acctcggaca 3060

tggacgagag cgatgtggaa gaggccgagt ggcggggaga gtctgaagga gagacggcgg 3120

cagactcaga atccgtcaca gtagttgagg tgacggtgcg tctaagcgca gggttctgct 3180

tgggcagagc cgaagtggac gccatggaga gctgggttag tttgtgtaga gagtgtgtgt 3240

tgctagcgac tttcggattg tgtcattaca caaaacgcgt cgtctcgaca ctgatcttgt 3300

cgtggatact cacggctcgg acatcgtcgc cgacgatgac accggacttt cgcttaagga 3360

cgtcagtaac aggcattgtg tgatgtgtag tttagatttc gaatctgtgg ggaaagaaag 3420

gaaaaaagag actggcaacc gattgggaga gccactgttt atatataccc tagacaagcc 3480

ccccgcttgt aagatgttgg tcaatgtaaa ccagtattaa ggttggcaag tgcaggagaa 3540

gcaaggtgtg ggtaccgagc aatggaaatg tgcggaaggc aaaaaaatga ggccacggcc 3600

tattgtcggg gctatatcca gggggcgatt gaagtacact aacatgacat gtgtccacag 3660

accctcaatc tggcctgatg agccaaatcc atacgcgctt tcgcagctct aaaggctata 3720

acaagtcaca ccaccctgct cgacctcagc gccctcactt tttgttaaga caaactgtac 3780

acgctgttcc agcgttttct gcctgcacct ggtgggacat ttggtgcaac ctaaagtgct 3840

cggaacctct gtggtgtcca gatcagcgca gcagttccga ggtagttttg aggcccttag 3900

atgatgcaat ggtgtcagtc gctggatcac gagtcttaat ggcagtattc gttcttattt 3960

gtgccattga gccccgttat cctcgtatct tctacccccc atcccatccc tttgttggtg 4020

caaccctacc catttattgt tgggtgcagc ccaaccgacg tggagagctt ggcttggcca 4080

tataaaaagg ccccccccta gtggcaatgg cagaaagtca gctgtgagtt gttgaatttg 4140

tcatctaggc ggcctggccg tcttctccgg ggcaattgtt cctctatagt actgcgtaca 4200

ctgtttaaac agtgtacgca gatctgcgac gacggaattc ctgcagccca tctgcagaat 4260

tcaggagaga ccgggttggc ggcgtatttg tgtcccaaaa aacagcccca attgccccaa 4320

ttgaccccaa attgacccag tagcgggccc aaccccggcg agagccccct tcaccccaca 4380

tatcaaacct cccccggttc ccacacttgc cgttaagggc gtagggtact gcagtctgga 4440

atctacgctt gttcagactt tgtactagtt tctttgtctg gccatccggg taacccatgc 4500

cggacgcaaa atagactact gaaaattttt ttgctttgtg gttgggactt tagccaaggg 4560

tataaaagac caccgtcccc gaattacctt tcctcttctt ttctctctct ccttgtcaac 4620

tcacacccga aatcgttaag catttccttc tgagtataag aatcattcac catggctgcc 4680

gctccctctg tgcgaacctt tacccgagcc gaggttctga acgctgaggc tctgaacgag 4740

ggcaagaagg acgctgaggc tcccttcctg atgatcatcg acaacaaggt gtacgacgtc 4800

-continued

cgagagttcg tccctgacca tcctggaggc tccgtgattc tcacccacgt tggcaaggac 4860

ggcaccgacg tctttgacac ctttcatccc gaggctgctt gggagactct cgccaacttc 4920

tacgttggag acattgacga gtccgaccga gacatcaaga acgatgactt tgccgctgag 4980

gtccgaaagc tgcgaaccct gttccagtct ctcggctact acgactcctc taaggcctac 5040

tacgccttca aggtctcctt caacctctgc atctggggac tgtccaccgt cattgtggcc 5100

aagtggggtc agacctccac cctcgccaac gtgctctctg ctgccctgct cggcctgttc 5160

tggcagcagt gcggatggct ggctcacgac tttctgcacc accaggtctt ccaggaccga 5220

ttctggggtg atctcttcgg agccttcctg ggaggtgtct gccagggctt ctcctcttcc 5280

tggtggaagg acaagcacaa cactcaccat gccgctccca acgtgcatgg cgaggatcct 5340

gacattgaca cccaccctct cctgacctgg tccgagcacg ctctggagat gttctccgac 5400

gtccccgatg aggagctgac ccgaatgtgg tctcgattca tggtcctgaa ccagacctgg 5460

ttctacttcc ccattctctc cttcgctcga ctgtcttggt gcctccagtc cattctcttt 5520

gtgctgccca acggtcaggc tcacaagccc tccggagctc gagtgcccat ctccctggtc 5580

gagcagctgt ccctcgccat gcactggacc tggtacctcg ctaccatgtt cctgttcatc 5640

aaggatcctg tcaacatgct cgtgtacttc ctggtgtctc aggctgtgtg cggaaacctg 5700

ctcgccatcg tgttctccct caaccacaac ggtatgcctg tgatctccaa ggaggaggct 5760

gtcgacatgg atttctttac caagcagatc atcactggtc gagatgtcca tcctggactg 5820

ttcgccaact ggttcaccgg tggcctgaac taccagatcg agcatcacct gttcccttcc 5880

atgcctcgac acaacttctc caagatccag cctgccgtcg agaccctgtg caagaagtac 5940

aacgtccgat accacaccac tggtatgatc gagggaactg ccgaggtctt ctcccgactg 6000

aacgaggtct ccaaggccac ctccaagatg ggcaaggctc agtaagcggc cgcatgagaa 6060

gataaatata taaatacatt gagatattaa atgcgctaga ttagagagcc tcatactgct 6120

cggagagaag ccaagacgag tactcaaagg ggattacacc atccatatcc acagacacaa 6180

gctgggaaa ggttctatat acactttccg gaataccgta gtttccgatg ttatcaatgg 6240

gggcagccag gatttcaggc acttcggtgt ctcggggtga aatggcgttc ttggcctcca 6300

tcaagtcgta ccatgtcttc atttgcctgt caaagtaaaa cagaagcaga tgaagaatga 6360

acttgaagtg aaggaattta aattgccccg gagaagacgg ccaggccgcc tagatgacaa 6420

attcaacaac tcacagctga ctttctgcca ttgccactag ggggggcct ttttatatgg 6480

ccaagccaag ctctccacgt cggttgggct gcacccaaca ataaatgggt agggttgcac 6540

caacaaaggg atgggatggg gggtagaaga tacgaggata acggggctca atggcacaaa 6600

taagaacgaa tactgccatt aagactcgtg atccagcgac tgacaccatt gcatcatcta 6660

agggcctcaa aactacctcg gaactgctgc gctgatctgg acaccacaga ggttccgagc 6720

actttaggtt gcaccaaatg tcccaccagg tgcaggcaga aaacgctgga acagcgtgta 6780

cagtttgtct taacaaaaag tgagggcgct gaggtcgagc agggtggtgt gacttgttat 6840

agcctttaga gctgcgaaag cgcgtatgga tttggctcat caggccagat tgagggtctg 6900

tggacacatg tcatgttagt gtacttcaat cgccccctgg atatagcccc gacaataggc 6960

cgtggcctca ttttttgcc ttccgcacat ttccattgct cggtacccac accttgcttc 7020

tcctgcactt gccaacctta atactggttt acattgacca acatcttaca agcgggggc 7080

-continued

ttgtctaggg tatatataaa cagtggctct cccaatcggt tgccagtctc ttttttcctt 7140

tctttcccca cagattcgaa atctaaacta cacatcacac aatgcctgtt actgacgtcc 7200

ttaagcgaaa gtccggtgtc atcgtcggcg acgatgtccg agccgtgagt atccacgaca 7260

agatcagtgt cgagacgacg cgttttgtgt aatgacacaa tccgaaagtc gctagcaaca 7320

cacactctct acacaaacta acccagctct ccatggagtc cattgctccc ttcctgccct 7380

ccaagatgcc tcaggacctg ttcatggacc tcgccagcgc tatcggtgtc cgagctgctc 7440

cctacgtcga tcccctggag gctgccctgg ttgcccaggc cgagaagtac attcccacca 7500

ttgtccatca cactcgaggc ttcctggttg ccgtggagtc tcccctggct cgagagctgc 7560

ctctgatgaa ccccttccac gtgctcctga tcgtgctcgc ctacctggtc accgtgtttg 7620

tgggtatgca gatcatgaag aactttgaac gattcgaggt caagaccttc tccctcctgc 7680

acaacttctg tctggtctcc atctccgcct acatgtgcgg tggcatcctg tacgaggctt 7740

atcaggccaa ctatggactg tttgagaacg ctgccgatca caccttcaag ggtctcccta 7800

tggctaagat gatctggctc ttctacttct ccaagatcat ggagtttgtc gacaccatga 7860

tcatggtcct caagaagaac aaccgacaga tttcctttct gcacgtgtac caccactctt 7920

ccatcttcac catctggtgg ctggtcacct tcgttgctcc caacggtgaa gcctacttct 7980

ctgctgccct gaactccttc atccacgtca tcatgtacgg ctactacttt ctgtctgccc 8040

tgggcttcaa gcaggtgtcg ttcatcaagt tctacatcac tcgatcccag atgacccagt 8100

tctgcatgat gtctgtccag tcttcctggg acatgtacgc catgaaggtc cttggccgac 8160

ctggataccc cttcttcatc accgctctgc tctggttcta catgtggacc atgctcggtc 8220

tcttctacaa cttttaccga aagaacgcca agctcgccaa gcaggccaag gctgacgctg 8280

ccaaggagaa ggccagaaag ctccagtaag cggccgcaag tgtggatggg gaagtgagtg 8340

cccggttctg tgtgcacaat tggcaatcca agatggatgg attcaacaca gggatatagc 8400

gagctacgtg gtggtgcgag gatatagcaa cggatattta tgtttgacac ttgagaatgt 8460

acgatacaag cactgtccaa gtacaatact aaacatactg tacatactca tactcgtacc 8520

cgggcaacgg tttcacttga gtgcagtggc tagtgctctt actcgtacag tgtgcaatac 8580

tgcgtatcat agtctttgat gtatatcgta ttcattcatg ttagttgcgt acgaagtcgt 8640

caatgatgtc gatatgggtt ttgatcatgc acacataagg tccgacctta tcggcaagct 8700

caatgagctc cttggtggtg gtaacatcca gagaagcaca caggttggtt ttcttggctg 8760

ccacgagctt gagcactcga gcggcaaagg cggacttgtg gacgttagct cgagcttcgt 8820

aggagggcat tttggtggtg aagaggagac tgaaataaat ttagtctgca gaacttttta 8880

tcggaacctt atctggggca gtgaagtata tgttatggta atagttacga gttagttgaa 8940

cttatagata gactggacta tacggctatc ggtccaaatt agaaagaacg tcaatggctc 9000

tctgggcgtc gcctttgccg acaaaaatgt gatcatgatg aaagccagca atgacgttgc 9060

agctgatatt gttgtcggcc aaccgcgccg aaaacgcagc tgtcagaccc acagcctcca 9120

acgaagaatg tatcgtcaaa gtgatccaag cacactcata gttggagtcg tactccaaag 9180

gcggcaatga cgagtcagac agatactcgt cgacctttc cttgggaacc accaccgtca 9240

gcccttctga ctcacgtatt gtagccaccg acacaggcaa cagtccgtgg atagcagaat 9300

atgtcttgtc ggtccatttc tcaccaactt taggcgtcaa gtgaatgttg cagaagaagt 9360

-continued

atgtgccttc attgagaatc ggtgttgctg atttcaataa agtcttgaga tcagtttggc 9420

gcgccagctg cattaatgaa tcggccaacg cgcggggaga ggcggtttgc gtattgggcg 9480

ctcttccgct tcctcgctca ctgactcgct gcgctcggtc gttcggctgc ggcgagcggt 9540

atcagctcac tcaaaggcgg taatacggtt atccacagaa tcaggggata acgcaggaaa 9600

gaacatgtga gcaaaaggcc agcaaaaggc caggaaccgt aaaaaggccg cgttgctggc 9660

gtttttccat aggctccgcc cccctgacga gcatcacaaa aatcgacgct caagtcagag 9720

gtggcgaaac ccgacaggac tataaagata ccaggcgttt ccccctggaa gctccctcgt 9780

gcgctctcct gttccgaccc tgccgcttac cggatacctg tccgcctttc tcccttcggg 9840

aagcgtggcg ctttctcata gctcacgctg taggtatctc agttcggtgt aggtcgttcg 9900

ctccaagctg ggctgtgtgc acgaaccccc cgttcagccc gaccgctgcg ccttatccgg 9960

taactatcgt cttgagtcca acccggtaag acacgactta tcgccactgg cagcagccac 10020

tggtaacagg attagcagag cgaggtatgt aggcggtgct acagagttct tgaagtggtg 10080

gcctaactac ggctacacta gaagaacagt atttggtatc tgcgctctgc tgaagccagt 10140

taccttcgga aaaagagttg gtagctcttg atccggcaaa caaaccaccg ctggtagcgg 10200

tggtttttt gtttgcaagc agcagattac gcgcagaaaa aaaggatctc aagaagatcc 10260

tttgatcttt tctacggggt ctgacgctca gtggaacgaa aactcacgtt aagggatttt 10320

ggtcatgaga ttatcaaaaa ggatcttcac ctagatcctt ttaaattaaa aatgaagttt 10380

taaatcaatc taaagtatat atgagtaaac ttggtctgac agttaccaat gcttaatcag 10440

tgaggcacct atctcagcga tctgtctatt tcgttcatcc atagttgcct gactccccgt 10500

cgtgtagata actacgatac gggagggctt accatctggc cccagtgctg caatgatacc 10560

gcgagaccca cgctcaccgg ctccagattt atcagcaata aaccagccag ccggaagggc 10620

cgagcgcaga agtggtcctg caactttatc cgcctccatc cagtctatta attgttgccg 10680

ggaagctaga gtaagtagtt cgccagttaa tagtttgcgc aacgttgttg ccattgctac 10740

aggcatcgtg gtgtcacgct cgtcgtttgg tatggcttca ttcagctccg gttcccaacg 10800

atcaaggcga gttacatgat cccccatgtt gtgcaaaaaa gcggttagct ccttcggtcc 10860

tccgatcgtt gtcagaagta agttggccgc agtgttatca ctcatggtta tggcagcact 10920

gcataattct cttactgtca tgccatccgt aagatgcttt tctgtgactg gtgagtactc 10980

aaccaagtca ttctgagaat agtgtatgcg gcgaccgagt tgctcttgcc cggcgtcaat 11040

acgggataat accgcgccac atagcagaac tttaaaagtg ctcatcattg gaaaacgttc 11100

ttcggggcga aaactctcaa ggatcttacc gctgttgaga tccagttcga tgtaacccac 11160

tcgtgcaccc aactgatctt cagcatcttt tactttcacc agcgtttctg ggtgagcaaa 11220

aacaggaagg caaaatgccg caaaaaaggg aataagggcg acacggaaat gttgaatact 11280

catactcttc ctttttcaat attattgaag catttatcag ggttattgtc tcatgagcgg 11340

atacatattt gaatgtattt agaaaaataa acaaataggg gttccgcgca catttccccg 11400

aaaagtgcca cctgatgcgg tgtgaaatac cgcacagatg cgtaaggaga aaataccgca 11460

tcaggaaatt gtaagcgtta atattttgtt aaaattcgcg ttaaattttt gttaaatcag 11520

ctcatttttt aaccaatagg ccgaaatcgg caaaatccct tataaatcaa aagaatagac 11580

cgagataggg ttgagtgttg ttccagtttg gaacaagagt ccactattaa agaacgtgga 11640

-continued

ctccaacgtc aaagggcgaa aaaccgtcta tcagggcgat ggcccactac gtgaaccatc 11700

accctaatca agttttttgg ggtcgaggtg ccgtaaagca ctaaatcgga accctaaagg 11760

gagcccccga tttagagctt gacggggaaa gccggcgaac gtggcgagaa aggaagggaa 11820

gaaagcgaaa ggagcgggcg ctagggcgct ggcaagtgta gcggtcacgc tgcgcgtaac 11880

caccacaccc gccgcgctta atgcgccgct acagggcgcg tccattcgcc attcaggctg 11940

cgcaactgtt gggaagggcg atcggtgcgg gcctcttcgc tattacgcca gctggcgaaa 12000

gggggatgtg ctgcaaggcg attaagttgg gtaacgccag ggttttccca gtcacgacgt 12060

tgtaaaacga cggccagtga attgtaatac gactcactat agggcgaatt gggcccgacg 12120

tcgcatgcag tggtggtatt gtgactgggg atgtagttga gaataagtca tacacaagtc 12180

agctttcttc gagcctcata taagtataag tagttcaacg tattagcact gtacccagca 12240

tctccgtatc gagaaacaca acaacatgcc ccattggaca gatcatgcgg atacacaggt 12300

tgtgcagtat catacatact cgatcagaca ggtcgtctga ccatcataca agctgaacaa 12360

gcgctccata cttgcacgct ctctatatac acagttaaat tacatatcca tagtctaacc 12420

tctaacagtt aatcttctgg taagcctccc agccagcctt ctggtatcgc ttggcctcct 12480

caataggatc tcggttctgg ccgtacagac ctcggccgac aattatgata tccgttccgg 12540

tagacatgac atcctcaaca gttcggtact gctgtccgag agcgtctccc ttgtcgtcaa 12600

gacccacccc gggggtcaga ataagccagt cctcagagtc gcccttaat 12649

<210> SEQ ID NO 28
<211> LENGTH: 957
<212> TYPE: DNA
<213> ORGANISM: Mortierella alpina
<220> FEATURE:
<221> NAME/KEY: misc_feature
<223> OTHER INFORMATION: synthetic elongase 1 (codon-optimized)

<400> SEQUENCE: 28

atggagtcca ttgctccctt cctgccctcc aagatgcctc aggacctgtt catggacctc 60

gccagcgcta tcggtgtccg agctgctccc tacgtcgatc ccctggaggc tgccctggtt 120

gcccaggccg agaagtacat tcccaccatt gtccatcaca ctcgaggctt cctggttgcc 180

gtggagtctc ccctggctcg agagctgcct ctgatgaacc ccttccacgt gctcctgatc 240

gtgctcgcct acctggtcac cgtgtttgtg ggtatgcaga tcatgaagaa ctttgaacga 300

ttcgaggtca agaccttctc cctcctgcac aacttctgtc tggtctccat ctccgcctac 360

atgtgcggtg gcatcctgta cgaggcttat caggccaact atggactgtt tgagaacgct 420

gccgatcaca ccttcaaggg tctccctatg gctaagatga tctggctctt ctacttctcc 480

aagatcatgg agtttgtcga caccatgatc atggtcctca agaagaacaa ccgacagatt 540

tcctttctgc acgtgtacca ccactcttcc atcttcacca tctggtggct ggtcaccttc 600

gttgctccca acggtgaagc ctacttctct gctgccctga actccttcat ccacgtcatc 660

atgtacggct actactttct gtctgccctg ggcttcaagc aggtgtcgtt catcaagttc 720

tacatcactc gatcccagat gacccagttc tgcatgatgt ctgtccagtc ttcctgggac 780

atgtacgcca tgaaggtcct tggccgacct ggataccct tcttcatcac cgctctgctc 840

tggttctaca tgtggaccat gctcggtctc ttctacaact tttaccgaaa gaacgccaag 900

ctcgccaagc aggccaaggc tgacgctgcc aaggagaagg ccagaaagct ccagtaa 957

-continued

<210> SEQ ID NO 29
<211> LENGTH: 318
<212> TYPE: PRT
<213> ORGANISM: Mortierella alpina AX464731

<400> SEQUENCE: 29

Met Glu Ser Ile Ala Pro Phe Leu Pro Ser Lys Met Pro Gln Asp Leu
1 5 10 15

Phe Met Asp Leu Ala Thr Ala Ile Gly Val Arg Ala Ala Pro Tyr Val
20 25 30

Asp Pro Leu Glu Ala Ala Leu Val Ala Gln Ala Glu Lys Tyr Ile Pro
35 40 45

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

Leu Ala Arg Glu Leu Pro Leu Met Asn Pro Phe His Val Leu Leu Ile
65 70 75 80

Val Leu Ala Tyr Leu Val Thr Val Phe Val Gly Met Gln Ile Met Lys
85 90 95

Asn Phe Glu Arg Phe Glu Val Lys Thr Phe Ser Leu Leu His Asn Phe
100 105 110

Cys Leu Val Ser Ile Ser Ala Tyr Met Cys Gly Gly Ile Leu Tyr Glu
115 120 125

Ala Tyr Gln Ala Asn Tyr Gly Leu Phe Glu Asn Ala Ala Asp His Thr
130 135 140

Phe Lys Gly Leu Pro Met Ala Lys Met Ile Trp Leu Phe Tyr Phe Ser
145 150 155 160

Lys Ile Met Glu Phe Val Asp Thr Met Ile Met Val Leu Lys Lys Asn
165 170 175

Asn Arg Gln Ile Ser Phe Leu His Val Tyr His His Ser Ser Ile Phe
180 185 190

Thr Ile Trp Trp Leu Val Thr Phe Val Ala Pro Asn Gly Glu Ala Tyr
195 200 205

Phe Ser Ala Ala Leu Asn Ser Phe Ile His Val Ile Met Tyr Gly Tyr
210 215 220

Tyr Phe Leu Ser Ala Leu Gly Phe Lys Gln Val Ser Phe Ile Lys Phe
225 230 235 240

Tyr Ile Thr Arg Ser Gln Met Thr Gln Phe Cys Met Met Ser Val Gln
245 250 255

Ser Ser Trp Asp Met Tyr Ala Met Lys Val Leu Gly Arg Pro Gly Tyr
260 265 270

Pro Phe Phe Ile Thr Ala Leu Leu Trp Phe Tyr Met Trp Thr Met Leu
275 280 285

Gly Leu Phe Tyr Asn Phe Tyr Arg Lys Asn Ala Lys Leu Ala Lys Gln
290 295 300

Ala Lys Ala Asp Ala Ala Lys Glu Lys Ala Arg Lys Leu Gln
305 310 315

<210> SEQ ID NO 30
<211> LENGTH: 1374
<212> TYPE: DNA
<213> ORGANISM: Mortierella alpina
<220> FEATURE:
<221> NAME/KEY: misc_feature

-continued

<223> OTHER INFORMATION: synthetic delta-6 desaturase (codon-optimized)

<400> SEQUENCE: 30

atggctgccg ctccctctgt gcgaaccttt acccgagccg aggttctgaa cgctgaggct 60

ctgaacgagg gcaagaagga cgctgaggct cccttcctga tgatcatcga caacaaggtg 120

tacgacgtcc gagagttcgt ccctgaccat cctggaggct ccgtgattct cacccacgtt 180

ggcaaggacg gcaccgacgt ctttgacacc tttcatcccg aggctgcttg ggagactctc 240

gccaacttct acgttggaga cattgacgag tccgaccgag acatcaagaa cgatgacttt 300

gccgctgagg tccgaaagct gcgaaccctg ttccagtctc tcggctacta cgactcctct 360

aaggcctact acgccttcaa ggtctccttc aacctctgca tctggggact gtccaccgtc 420

attgtggcca agtggggtca gacctccacc ctcgccaacg tgctctctgc tgccctgctc 480

ggcctgttct ggcagcagtg cggatggctg gctcacgact ttctgcacca ccaggtcttc 540

caggaccgat tctggggtga tctcttcgga gccttcctgg gaggtgtctg ccagggcttc 600

tcctcttcct ggtggaagga caagcacaac actcaccatg ccgctcccaa cgtgcatggc 660

gaggatcctg acattgacac ccaccctctc ctgacctggt ccgagcacgc tctggagatg 720

ttctccgacg tccccgatga ggagctgacc cgaatgtggt ctcgattcat ggtcctgaac 780

cagacctggt tctacttccc cattctctcc ttcgctcgac tgtcttggtg cctccagtcc 840

attctctttg tgctgcccaa cggtcaggct cacaagccct ccggagctcg agtgcccatc 900

tccctggtcg agcagctgtc cctcgccatg cactggacct ggtacctcgc taccatgttc 960

ctgttcatca aggatcctgt caacatgctc gtgtacttcc tggtgtctca ggctgtgtgc 1020

ggaaacctgc tcgccatcgt gttctccctc aaccacaacg gtatgcctgt gatctccaag 1080

gaggaggctg tcgacatgga tttctttacc aagcagatca tcactggtcg agatgtccat 1140

cctggactgt tcgccaactg gttcaccggt ggcctgaact accagatcga gcatcacctg 1200

ttcccttcca tgcctcgaca caacttctcc aagatccagc ctgccgtcga gaccctgtgc 1260

aagaagtaca acgtccgata ccacaccact ggtatgatcg agggaactgc cgaggtcttc 1320

tcccgactga acgaggtctc caaggccacc tccaagatgg gcaaggctca gtaa 1374

<210> SEQ ID NO 31
<211> LENGTH: 457
<212> TYPE: PRT
<213> ORGANISM: Mortierella alpina AF465281

<400> SEQUENCE: 31

Met Ala Ala Ala Pro Ser Val Arg Thr Phe Thr Arg Ala Glu Val Leu
1 5 10 15

Asn Ala Glu Ala Leu Asn Glu Gly Lys Lys Asp Ala Glu Ala Pro Phe
20 25 30

Leu Met Ile Ile Asp Asn Lys Val Tyr Asp Val Arg Glu Phe Val Pro
35 40 45

Asp His Pro Gly Gly Ser Val Ile Leu Thr His Val Gly Lys Asp Gly
50 55 60

Thr Asp Val Phe Asp Thr Phe His Pro Glu Ala Ala Trp Glu Thr Leu
65 70 75 80

Ala Asn Phe Tyr Val Gly Asp Ile Asp Glu Ser Asp Arg Asp Ile Lys
85 90 95

-continued

```
Asn Asp Asp Phe Ala Ala Glu Val Arg Lys Leu Arg Thr Leu Phe Gln
            100                 105                 110

Ser Leu Gly Tyr Tyr Asp Ser Ser Lys Ala Tyr Tyr Ala Phe Lys Val
        115                 120                 125

Ser Phe Asn Leu Cys Ile Trp Gly Leu Ser Thr Val Ile Val Ala Lys
    130                 135                 140

Trp Gly Gln Thr Ser Thr Leu Ala Asn Val Leu Ser Ala Ala Leu Leu
145                 150                 155                 160

Gly Leu Phe Trp Gln Gln Cys Gly Trp Leu Ala His Asp Phe Leu His
                165                 170                 175

His Gln Val Phe Gln Asp Arg Phe Trp Gly Asp Leu Phe Gly Ala Phe
            180                 185                 190

Leu Gly Gly Val Cys Gln Gly Phe Ser Ser Ser Trp Trp Lys Asp Lys
        195                 200                 205

His Asn Thr His His Ala Ala Pro Asn Val His Gly Glu Asp Pro Asp
    210                 215                 220

Ile Asp Thr His Pro Leu Leu Thr Trp Ser Glu His Ala Leu Glu Met
225                 230                 235                 240

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

Met Val Leu Asn Gln Thr Trp Phe Tyr Phe Pro Ile Leu Ser Phe Ala
            260                 265                 270

Arg Leu Ser Trp Cys Leu Gln Ser Ile Leu Phe Val Leu Pro Asn Gly
        275                 280                 285

Gln Ala His Lys Pro Ser Gly Ala Arg Val Pro Ile Ser Leu Val Glu
    290                 295                 300

Gln Leu Ser Leu Ala Met His Trp Thr Trp Tyr Leu Ala Thr Met Phe
305                 310                 315                 320

Leu Phe Ile Lys Asp Pro Val Asn Met Leu Val Tyr Phe Leu Val Ser
                325                 330                 335

Gln Ala Val Cys Gly Asn Leu Leu Ala Ile Val Phe Ser Leu Asn His
            340                 345                 350

Asn Gly Met Pro Val Ile Ser Lys Glu Glu Ala Val Asp Met Asp Phe
        355                 360                 365

Phe Thr Lys Gln Ile Ile Thr Gly Arg Asp Val His Pro Gly Leu Phe
    370                 375                 380

Ala Asn Trp Phe Thr Gly Gly Leu Asn Tyr Gln Ile Glu His His Leu
385                 390                 395                 400

Phe Pro Ser Met Pro Arg His Asn Phe Ser Lys Ile Gln Pro Ala Val
                405                 410                 415

Glu Thr Leu Cys Lys Lys Tyr Asn Val Arg Tyr His Thr Thr Gly Met
            420                 425                 430

Ile Glu Gly Thr Ala Glu Val Phe Ser Arg Leu Asn Glu Val Ser Lys
        435                 440                 445

Ala Thr Ser Lys Met Gly Lys Ala Gln
    450                 455


<210> SEQ ID NO 32
<211> LENGTH: 1001
<212> TYPE: DNA
<213> ORGANISM: Yarrowia lipolytica
<220> FEATURE:
<221> NAME/KEY: misc_feature
```

-continued

<223> OTHER INFORMATION: Promoter FBA

<400> SEQUENCE: 32

taaacagtgt acgcagtact atagaggaac aattgccccg gagaagacgg ccaggccgcc 60

tagatgacaa attcaacaac tcacagctga ctttctgcca ttgccactag gggggcctt 120

tttatatggc caagccaagc tctccacgtc ggttgggctg cacccaacaa taaatgggta 180

gggttgcacc aacaaaggga tgggatgggg ggtagaagat acgaggataa cggggctcaa 240

tggcacaaat aagaacgaat actgccatta agactcgtga tccagcgact gacaccattg 300

catcatctaa gggcctcaaa actacctcgg aactgctgcg ctgatctgga caccacagag 360

gttccgagca ctttaggttg caccaaatgt cccaccaggt gcaggcagaa aacgctggaa 420

cagcgtgtac agtttgtctt aacaaaaagt gagggcgctg aggtcgagca gggtggtgtg 480

acttgttata gcctttagag ctgcgaaagc gcgtatggat ttggctcatc aggccagatt 540

gagggtctgt ggacacatgt catgttagtg tacttcaatc gccccctgga tatagccccg 600

acaataggcc gtggcctcat ttttttgcct tccgcacatt tccattgctc ggtacccaca 660

ccttgcttct cctgcacttg ccaaccttaa tactggttta cattgaccaa catcttacaa 720

gcgggggct tgtctagggt atatataaac agtggctctc ccaatcggtt gccagtctct 780

tttttccttt ctttccccac agattcgaaa tctaaactac acatcacaca atgcctgtta 840

ctgacgtcct taagcgaaag tccggtgtca tcgtcggcga cgatgtccga gccgtgagta 900

tccacgacaa gatcagtgtc gagacgacgc gttttgtgta atgacacaat ccgaaagtcg 960

ctagcaacac acactctcta cacaaactaa cccagctctc c 1001

<210> SEQ ID NO 33
<211> LENGTH: 1434
<212> TYPE: DNA
<213> ORGANISM: Fusarium monoliforme
<220> FEATURE:
<221> NAME/KEY: misc_feature
<223> OTHER INFORMATION: delta-12 desaturase

<400> SEQUENCE: 33

atggcgtcca cttcggctct gcccaagcag aaccctgcgc ttagacgcac cgtcacctca 60

actactgtga cggattctga gtctgccgcc gtctctcctt cagactctcc ccgccactcg 120

gcctcttcca catcgctctc gtccatgtcc gaggttgata tcgccaagcc caagtccgag 180

tatggtgtca tgctcgacac ctacggcaac cagttcgagg ttcccgactt taccatcaag 240

gacatctaca atgccatccc taagcactgc ttcaagcgct ccgctctcaa gggatacggt 300

tatatcctcc gcgacattgt cctcctgact accactttca gcatctggta caactttgtg 360

acccccgaat atatcccctc caccccgcc cgcgctggtc tgtgggccgt gtacaccgtt 420

cttcagggtc ttttcggtac tggtctctgg gttattgccc atgagtgcgg tcacggtgct 480

ttctccgatt ctcgcatcat caacgacatt actggctggg ttcttcactc ttccctcctt 540

gtcccctact tcagctggca aatctcccac cgaaagcacc acaaggccac tggcaacatg 600

gagcgtgaca tggtcttcgt tccccgaacc cgcgagcagc aggctactcg tctcggaaag 660

atgacccacg agctcgctca tcttactgag gagacccccg ctttcactct tctcatgctc 720

gtccttcagc agctcgttgg ctggcccaac tacctcatca ccaatgttac cggccacaac 780

taccacgagc gccagcgtga gggtcgcggc aagggcaagc ataacggcct cggcggtggt 840

-continued

gttaaccact tcgatccccg cagccctctg tacgagaaca gtgacgctaa gctcatcgtc 900

ctcagcgata ttggtatcgg tctgatggcc actgctctgt acttcctcgt tcagaagttc 960

ggtttctaca acatggccat ctggtacttt gttccctacc tctgggttaa ccactggctc 1020

gttgccatca ccttcctcca gcacaccgac cctacccttc cccactacac caacgacgag 1080

tggaacttcg tccgtggtgc cgctgctacc attgaccgtg agatgggctt catcggccgc 1140

caccttctcc acggcatcat cgagactcat gtcctccacc actacgtcag cagcatcccc 1200

ttctacaacg cggacgaggc caccgaggcc attaagccca tcatgggcaa gcactaccgg 1260

gctgatgtcc aggatggtcc tcgtggcttc atccgcgcca tgtaccgcag tgcgcgtatg 1320

tgccagtggg ttgagcccag cgctggtgcc gagggtgctg gtaagggtgt tctgttcttc 1380

cgcaaccgca acaacgtggg caccccccgc gctgttatca agcccgttgc ttaa 1434

<210> SEQ ID NO 34
<211> LENGTH: 477
<212> TYPE: PRT
<213> ORGANISM: Fusarium monoliforme

<400> SEQUENCE: 34

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

Thr Val Thr Ser Thr Thr Val Thr Asp Ser Glu Ser Ala Ala Val Ser
20 25 30

Pro Ser Asp Ser Pro Arg His Ser Ala Ser Ser Thr Ser Leu Ser Ser
35 40 45

Met Ser Glu Val Asp Ile Ala Lys Pro Lys Ser Glu Tyr Gly Val Met
50 55 60

Leu Asp Thr Tyr Gly Asn Gln Phe Glu Val Pro Asp Phe Thr Ile Lys
65 70 75 80

Asp Ile Tyr Asn Ala Ile Pro Lys His Cys Phe Lys Arg Ser Ala Leu
85 90 95

Lys Gly Tyr Gly Tyr Ile Leu Arg Asp Ile Val Leu Leu Thr Thr Thr
100 105 110

Phe Ser Ile Trp Tyr Asn Phe Val Thr Pro Glu Tyr Ile Pro Ser Thr
115 120 125

Pro Ala Arg Ala Gly Leu Trp Ala Val Tyr Thr Val Leu Gln Gly Leu
130 135 140

Phe Gly Thr Gly Leu Trp Val Ile Ala His Glu Cys Gly His Gly Ala
145 150 155 160

Phe Ser Asp Ser Arg Ile Ile Asn Asp Ile Thr Gly Trp Val Leu His
165 170 175

Ser Ser Leu Leu Val Pro Tyr Phe Ser Trp Gln Ile Ser His Arg Lys
180 185 190

His His Lys Ala Thr Gly Asn Met Glu Arg Asp Met Val Phe Val Pro
195 200 205

Arg Thr Arg Glu Gln Gln Ala Thr Arg Leu Gly Lys Met Thr His Glu
210 215 220

Leu Ala His Leu Thr Glu Glu Thr Pro Ala Phe Thr Leu Leu Met Leu
225 230 235 240

Val Leu Gln Gln Leu Val Gly Trp Pro Asn Tyr Leu Ile Thr Asn Val
245 250 255

-continued

```
Thr Gly His Asn Tyr His Glu Arg Gln Arg Glu Gly Arg Gly Lys Gly
            260                 265                 270

Lys His Asn Gly Leu Gly Gly Gly Val Asn His Phe Asp Pro Arg Ser
        275                 280                 285

Pro Leu Tyr Glu Asn Ser Asp Ala Lys Leu Ile Val Leu Ser Asp Ile
    290                 295                 300

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

Gly Phe Tyr Asn Met Ala Ile Trp Tyr Phe Val Pro Tyr Leu Trp Val
                325                 330                 335

Asn His Trp Leu Val Ala Ile Thr Phe Leu Gln His Thr Asp Pro Thr
            340                 345                 350

Leu Pro His Tyr Thr Asn Asp Glu Trp Asn Phe Val Arg Gly Ala Ala
        355                 360                 365

Ala Thr Ile Asp Arg Glu Met Gly Phe Ile Gly Arg His Leu Leu His
    370                 375                 380

Gly Ile Ile Glu Thr His Val Leu His His Tyr Val Ser Ser Ile Pro
385                 390                 395                 400

Phe Tyr Asn Ala Asp Glu Ala Thr Glu Ala Ile Lys Pro Ile Met Gly
                405                 410                 415

Lys His Tyr Arg Ala Asp Val Gln Asp Gly Pro Arg Gly Phe Ile Arg
            420                 425                 430

Ala Met Tyr Arg Ser Ala Arg Met Cys Gln Trp Val Glu Pro Ser Ala
        435                 440                 445

Gly Ala Glu Gly Ala Gly Lys Gly Val Leu Phe Phe Arg Asn Arg Asn
    450                 455                 460

Asn Val Gly Thr Pro Pro Ala Val Ile Lys Pro Val Ala
465                 470                 475
```

<210> SEQ ID NO 35
<211> LENGTH: 819
<212> TYPE: DNA
<213> ORGANISM: Thraustochytrium aureum
<220> FEATURE:
<221> NAME/KEY: misc_feature
<223> OTHER INFORMATION: synthetic elongase (codon-optimized)

<400> SEQUENCE: 35

```
atggccaact cctctgtctg ggacgacgtg gtcggacgag tcgagaccgg tgtcgaccag     60

tggatggacg gagctaagcc ctacgctctg accgacggtc tgcccatgat ggacgtctcc    120

accatgctcg ccttcgaggt cggctacatg gccatgctgc tcttcggcat tcccatcatg    180

aagcagatgg agaagccctt cgagctgaag accatcaagc tgctccacaa cctgttcctc    240

ttcggactgt ccctctacat gtgcgtcgag accatccgac aggctatcct gggtggctac    300

aaggtcttcg gcaacgacat ggagaagggc aacgagtccc acgctcaggg catgtcccga    360

atcgtctacg tgttctacgt ctccaaggcc tacgagttcc tggacaccgc tatcatgatc    420

ctgtgcaaga agttcaacca ggtctccttc ctgcacgtgt accaccatgc caccatcttc    480

gccatctggt gggctattgc caagtacgct cctggtggcg acgcctactt ctccgtcatc    540

ctcaactcct tcgtccacac cgtcatgtac gcctactact tcttttcctc tcagggcttc    600

ggcttcgtca agcccatcaa gccctacatc accactctgc agatgaccca gttcatggct    660
```

-continued

atgctggtgc agtccctgta cgactacctc ttcccctgcg actaccctca ggctctggtc 720

cagctgctcg gcgtgtacat gatcaccctg ctcgctctgt tcggcaactt ctttgtccag 780

tcctacctga agaagcccaa gaagtccaag accaactaa 819

<210> SEQ ID NO 36
<211> LENGTH: 272
<212> TYPE: PRT
<213> ORGANISM: Thraustochytrium aureum

<400> SEQUENCE: 36

Met Ala Asn Ser Ser Val Trp Asp Asp Val Val Gly Arg Val Glu Thr
1 5 10 15

Gly Val Asp Gln Trp Met Asp Gly Ala Lys Pro Tyr Ala Leu Thr Asp
20 25 30

Gly Leu Pro Met Met Asp Val Ser Thr Met Leu Ala Phe Glu Val Gly
35 40 45

Tyr Met Ala Met Leu Leu Phe Gly Ile Pro Ile Met Lys Gln Met Glu
50 55 60

Lys Pro Phe Glu Leu Lys Thr Ile Lys Leu Leu His Asn Leu Phe Leu
65 70 75 80

Phe Gly Leu Ser Leu Tyr Met Cys Val Glu Thr Ile Arg Gln Ala Ile
85 90 95

Leu Gly Gly Tyr Lys Val Phe Gly Asn Asp Met Glu Lys Gly Asn Glu
100 105 110

Ser His Ala Gln Gly Met Ser Arg Ile Val Tyr Val Phe Tyr Val Ser
115 120 125

Lys Ala Tyr Glu Phe Leu Asp Thr Ala Ile Met Ile Leu Cys Lys Lys
130 135 140

Phe Asn Gln Val Ser Phe Leu His Val Tyr His His Ala Thr Ile Phe
145 150 155 160

Ala Ile Trp Trp Ala Ile Ala Lys Tyr Ala Pro Gly Gly Asp Ala Tyr
165 170 175

Phe Ser Val Ile Leu Asn Ser Phe Val His Thr Val Met Tyr Ala Tyr
180 185 190

Tyr Phe Phe Ser Ser Gln Gly Phe Gly Phe Val Lys Pro Ile Lys Pro
195 200 205

Tyr Ile Thr Thr Leu Gln Met Thr Gln Phe Met Ala Met Leu Val Gln
210 215 220

Ser Leu Tyr Asp Tyr Leu Phe Pro Cys Asp Tyr Pro Gln Ala Leu Val
225 230 235 240

Gln Leu Leu Gly Val Tyr Met Ile Thr Leu Leu Ala Leu Phe Gly Asn
245 250 255

Phe Phe Val Gln Ser Tyr Leu Lys Lys Pro Lys Lys Ser Lys Thr Asn
260 265 270

<210> SEQ ID NO 37
<211> LENGTH: 10945
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Plasmid pDMW232

<400> SEQUENCE: 37

aattcctgca gcccatcgat caggagagac cgggttggcg gcgtatttgt gtcccaaaaa 60

-continued

acagccccaa ttgccccaat tgaccccaaa ttgacccagt agcgggccca accccggcga 120

gagccccctt caccccacat atcaaacctc ccccggttcc cacacttgcc gttaagggcg 180

tagggtactg cagtctggaa tctacgcttg ttcagacttt gtactagttt ctttgtctgg 240

ccatccgggt aacccatgcc ggacgcaaaa tagactactg aaaatttttt tgctttgtgg 300

ttgggacttt agccaagggt ataaaagacc accgtccccg aattaccttt cctcttcttt 360

tctctctctc cttgtcaact cacacccgaa atcgttaagc atttccttct gagtataaga 420

atcattcacc atgggaacgg accaaggaaa aaccttcacc tgggaagagc tggcggccca 480

taacaccaag gacgacctac tcttggccat ccgcggcagg gtgtacgatg tcacaaagtt 540

cttgagccgc catcctggtg gagtggacac tctcctgctc ggagctggcc gagatgttac 600

tccggtcttt gagatgtatc acgcgtttgg ggctgcagat gccattatga agaagtacta 660

tgtcggtaca ctggtctcga atgagctgcc catcttcccg gagccaacgg tgttccacaa 720

aaccatcaag acgagagtcg agggctactt tacggatcgg aacattgatc ccaagaatag 780

accagagatc tggggacgat acgctcttat ctttggatcc ttgatcgctt cctactacgc 840

gcagctcttt gtgcctttcg ttgtcgaacg cacatggctt caggtggtgt ttgcaatcat 900

catgggattt gcgtgcgcac aagtcggact caaccctctt catgatgcgt ctcacttttc 960

agtgacccac aaccccactg tctggaagat tctgggagcc acgcacgact ttttcaacgg 1020

agcatcgtac ctggtgtgga tgtaccaaca tatgctcggc catcacccct acaccaacat 1080

tgctggagca gatcccgacg tgtcgacgtc tgagcccgat gttcgtcgta tcaagcccaa 1140

ccaaaagtgg tttgtcaacc acatcaacca gcacatgttt gttcctttcc tgtacggact 1200

gctggcgttc aaggtgcgca ttcaggacat caacattttg tactttgtca agaccaatga 1260

cgctattcgt gtcaatccca tctcgacatg gcacactgtg atgttctggg gcggcaaggc 1320

tttctttgtc tggtatcgcc tgattgttcc cctgcagtat ctgcccctgg gcaaggtgct 1380

gctcttgttc acggtcgcgg acatggtgtc gtcttactgg ctggcgctga ccttccaggc 1440

gaaccacgtt gttgaggaag ttcagtggcc gttgcctgac gagaacggga tcatccaaaa 1500

ggactgggca gctatgcagg tcgagactac gcaggattac gcacacgatt cgcacctctg 1560

gaccagcatc actggcagct tgaactacca ggctgtgcac catctgttcc ccaacgtgtc 1620

gcagcaccat tatcccgata ttctggccat catcaagaac acctgcagcg agtacaaggt 1680

tccatacctt gtcaaggata cgttttggca agcatttgct tcacatttgg agcacttgcg 1740

tgttcttgga ctccgtccca aggaagagta ggcagctaag cggccgcatg agaagataaa 1800

tatataaata cattgagata ttaaatgcgc tagattagag agcctcatac tgctcggaga 1860

gaagccaaga cgagtactca aaggggatta caccatccat atccacagac acaagctggg 1920

gaaaggttct atatacactt tccggaatac cgtagtttcc gatgttatca atgggggcag 1980

ccaggatttc aggcacttcg gtgtctcggg gtgaaatggc gttcttggcc tccatcaagt 2040

cgtaccatgt cttcatttgc ctgtcaaagt aaaacagaag cagatgaaga atgaacttga 2100

agtgaaggaa tttaaattgc cccggagaag acggccaggc cgcctagatg acaaattcaa 2160

caactcacag ctgactttct gccattgcca ctaggggggg gccttttttat atggccaagc 2220

caagctctcc acgtcggttg ggctgcaccc aacaataaat gggtagggtt gcaccaacaa 2280

agggatggga tgggggggtag aagatacgag gataacgggg ctcaatggca caaataagaa 2340

-continued

```
cgaatactgc cattaagact cgtgatccag cgactgacac cattgcatca tctaagggcc   2400

tcaaaactac ctcggaactg ctgcgctgat ctggacacca cagaggttcc gagcacttta   2460

ggttgcacca aatgtcccac caggtgcagg cagaaaacgc tggaacagcg tgtacagttt   2520

gtcttaacaa aaagtgaggg cgctgaggtc gagcagggtg gtgtgacttg ttatagcctt   2580

tagagctgcg aaagcgcgta tggatttggc tcatcaggcc agattgaggg tctgtggaca   2640

catgtcatgt tagtgtactt caatcgcccc ctggatatag ccccgacaat aggccgtggc   2700

ctcatttttt tgccttccgc acatttccat tgctcggtac ccacaccttg cttctcctgc   2760

acttgccaac cttaatactg gtttacattg accaacatct tacaagcggg gggcttgtct   2820

agggtatata taaacagtgg ctctcccaat cggttgccag tctcttttt cctttctttc   2880

cccacagatt cgaaatctaa actacacatc acacaatgcc tgttactgac gtccttaagc   2940

gaaagtccgg tgtcatcgtc ggcgacgatg tccgagccgt gagtatccac gacaagatca   3000

gtgtcgagac gacgcgtttt gtgtaatgac acaatccgaa agtcgctagc aacacacact   3060

ctctacacaa actaacccag ctctccatgg gaacggacca aggaaaaacc ttcacctggg   3120

aagagctggc ggcccataac accaaggacg acctactctt ggccatccgc ggcagggtgt   3180

acgatgtcac aaagttcttg agccgccatc ctggtggagt ggacactctc ctgctcggag   3240

ctggccgaga tgttactccg gtctttgaga tgtatcacgc gtttggggct gcagatgcca   3300

ttatgaagaa gtactatgtc ggtacactgg tctcgaatga gctgcccatc ttcccggagc   3360

caacggtgtt ccacaaaacc atcaagacga gagtcgaggg ctactttacg gatcggaaca   3420

ttgatcccaa gaatagacca gagatctggg gacgatacgc tcttatcttt ggatccttga   3480

tcgcttccta ctacgcgcag ctctttgtgc ctttcgttgt cgaacgcaca tggcttcagg   3540

tggtgtttgc aatcatcatg ggatttgcgt gcgcacaagt cggactcaac cctcttcatg   3600

atgcgtctca cttttcagtg acccacaacc ccactgtctg gaagattctg ggagccacgc   3660

acgacttttt caacggagca tcgtacctgg tgtggatgta ccaacatatg ctcggccatc   3720

acccctacac caacattgct ggagcagatc ccgacgtgtc gacgtctgag cccgatgttc   3780

gtcgtatcaa gcccaaccaa aagtggtttg tcaaccacat caaccagcac atgtttgttc   3840

ctttcctgta cggactgctg gcgttcaagg tgcgcattca ggacatcaac attttgtact   3900

ttgtcaagac caatgacgct attcgtgtca atcccatctc gacatggcac actgtgatgt   3960

tctggggcgg caaggctttc tttgtctggt atcgcctgat tgttcccctg cagtatctgc   4020

ccctgggcaa ggtgctgctc ttgttcacgg tcgcggacat ggtgtcgtct tactggctgg   4080

cgctgacctt ccaggcgaac cacgttgttg aggaagttca gtggccgttg cctgacgaga   4140

acgggatcat ccaaaaggac tgggcagcta tgcaggtcga gactacgcag gattacgcac   4200

acgattcgca cctctggacc agcatcactg gcagcttgaa ctaccaggct gtgcaccatc   4260

tgttccccaa cgtgtcgcag caccattatc ccgatattct ggccatcatc aagaacacct   4320

gcagcgagta caaggttcca taccttgtca aggatacgtt ttggcaagca tttgcttcac   4380

atttggagca cttgcgtgtt cttggactcc gtcccaagga agagtaggca gctaagcggc   4440

cgcaagtgtg gatggggaag tgagtgcccg gttctgtgtg cacaattggc aatccaagat   4500

ggatggattc aacacaggga tatagcgagc tacgtggtgg tgcgaggata tagcaacgga   4560

tatttatgtt tgacacttga gaatgtacga tacaagcact gtccaagtac aatactaaac   4620
```

-continued

atactgtaca tactcatact cgtacccggg caacggtttc acttgagtgc agtggctagt 4680

gctcttactc gtacagtgtg caatactgcg tatcatagtc tttgatgtat atcgtattca 4740

ttcatgttag ttgcgtacgc caccattctg tctgccgcca tgatgctcaa gttctctctt 4800

aacatgaagc ccgccggtga cgctgttgag gctgccgtca aggagtccgt cgaggctggt 4860

atcactaccg ccgatatcgg aggctcttcc tccacctccg aggtcggaga cttgttgcca 4920

acaaggtcaa ggagctgctc aagaaggagt aagtcgtttc tacgacgcat tgatggaagg 4980

agcaaactga cgcgcctgcg ggttggtcta ccggcagggt ccgctagtgt ataagactct 5040

ataaaaaggg ccctgccctg ctaatgaaat gatgatttat aatttaccgg tgtagcaacc 5100

ttgactagaa gaagcagatt gggtgtgttt gtagtggagg acagtggtac gttttggaaa 5160

cagtcttctt gaaagtgtct tgtctacagt atattcactc ataacctcaa tagccaaggg 5220

tgtagtcggt ttattaaagg aagggagttg tggctgatgt ggatagatat ctttaagctg 5280

gcgactgcac ccaacgagtg tggtggtagc ttgttactgt atattcggta agatatattt 5340

tgtggggttt tagtggtgtt tggtaggtta gtgcttggta tatgagttgt aggcatgaca 5400

atttggaaag gggtggactt tgggaatatt gtgggatttc aataccttag tttgtacagg 5460

gtaattgtta caaatgatac aaagaactgt atttcttttc atttgtttta attggttgta 5520

tatcaagtcc gttagacgag ctcagtgggc gcgccagctg cattaatgaa tcggccaacg 5580

cgcgggagga ggcggtttgc gtattgggcg ctcttccgct tcctcgctca ctgactcgct 5640

gcgctcggtc gttcggctgc ggcgagcggt atcagctcac tcaaaggcgg taatacggtt 5700

atccacagaa tcaggggata acgcaggaaa gaacatgtga gcaaaaggcc agcaaaaggc 5760

caggaaccgt aaaaaggccg cgttgctggc gtttttccat aggctccgcc cccctgacga 5820

gcatcacaaa aatcgacgct caagtcagag gtggcgaaac ccgacaggac tataaagata 5880

ccaggcgttt cccccctggaa gctccctcgt gcgctctcct gttccgaccc tgccgcttac 5940

cggatacctg tccgcctttc tcccttcggg aagcgtggcg ctttctcata gctcacgctg 6000

taggtatctc agttcggtgt aggtcgttcg ctccaagctg ggctgtgtgc acgaaccccc 6060

cgttcagccc gaccgctgcg ccttatccgg taactatcgt cttgagtcca acccggtaag 6120

acacgactta tcgccactgg cagcagccac tggtaacagg attagcagag cgaggtatgt 6180

aggcggtgct acagagttct tgaagtggtg gcctaactac ggctacacta gaagaacagt 6240

atttggtatc tgcgctctgc tgaagccagt taccttcgga aaaagagttg gtagctcttg 6300

atccggcaaa caaaccaccg ctggtagcgg tggttttttt gtttgcaagc agcagattac 6360

gcgcagaaaa aaaggatctc aagaagatcc tttgatcttt tctacggggt ctgacgctca 6420

gtggaacgaa aactcacgtt aagggatttt ggtcatgaga ttatcaaaaa ggatcttcac 6480

ctagatcctt ttaaattaaa aatgaagttt taaatcaatc taaagtatat atgagtaaac 6540

ttggtctgac agttaccaat gcttaatcag tgaggcacct atctcagcga tctgtctatt 6600

tcgttcatcc atagttgcct gactccccgt cgtgtagata actacgatac gggagggctt 6660

accatctggc cccagtgctg caatgatacc gcgagaccca cgctcaccgg ctccagattt 6720

atcagcaata aaccagccag ccggaagggc cgagcgcaga agtggtcctg caactttatc 6780

cgcctccatc cagtctatta attgttgccg ggaagctaga gtaagtagtt cgccagttaa 6840

tagtttgcgc aacgttgttg ccattgctac aggcatcgtg gtgtcacgct cgtcgtttgg 6900

-continued

tatggcttca ttcagctccg gttcccaacg atcaaggcga gttacatgat cccccatgtt 6960

gtgcaaaaaa gcggttagct ccttcggtcc tccgatcgtt gtcagaagta agttggccgc 7020

agtgttatca ctcatggtta tggcagcact gcataattct cttactgtca tgccatccgt 7080

aagatgcttt tctgtgactg gtgagtactc aaccaagtca ttctgagaat agtgtatgcg 7140

gcgaccgagt tgctcttgcc cggcgtcaat acgggataat accgcgccac atagcagaac 7200

tttaaaagtg ctcatcattg gaaaacgttc ttcggggcga aaactctcaa ggatcttacc 7260

gctgttgaga tccagttcga tgtaacccac tcgtgcaccc aactgatctt cagcatcttt 7320

tactttcacc agcgtttctg ggtgagcaaa aacaggaagg caaaatgccg caaaaaaggg 7380

aataagggcg acacggaaat gttgaatact catactcttc ctttttcaat attattgaag 7440

catttatcag ggttattgtc tcatgagcgg atacatattt gaatgtattt agaaaaataa 7500

acaaataggg gttccgcgca catttccccg aaaagtgcca cctgatgcgg tgtgaaatac 7560

cgcacagatg cgtaaggaga aaataccgca tcaggaaatt gtaagcgtta atattttgtt 7620

aaaattcgcg ttaaattttt gttaaatcag ctcatttttt aaccaatagg ccgaaatcgg 7680

caaaatccct tataaatcaa aagaatagac cgagataggg ttgagtgttg ttccagtttg 7740

gaacaagagt ccactattaa agaacgtgga ctccaacgtc aaagggcgaa aaaccgtcta 7800

tcagggcgat ggcccactac gtgaaccatc accctaatca agtttttttgg ggtcgaggtg 7860

ccgtaaagca ctaaatcgga accctaaagg gagcccccga tttagagctt gacggggaaa 7920

gccggcgaac gtggcgagaa aggaagggaa gaaagcgaaa ggagcgggcg ctagggcgct 7980

ggcaagtgta gcggtcacgc tgcgcgtaac caccacaccc gccgcgctta atgcgccgct 8040

acagggcgcg tccattcgcc attcaggctg cgcaactgtt gggaagggcg atcggtgcgg 8100

gcctcttcgc tattacgcca gctggcgaaa gggggatgtg ctgcaaggcg attaagttgg 8160

gtaacgccag ggttttccca gtcacgacgt tgtaaaacga cggccagtga attgtaatac 8220

gactcactat agggcgaatt gggcccgacg tcgcatgcta tcggcatcga caaggtttgg 8280

gtccctagcc gataccgcac tacctgagtc acaatcttcg gaggtttagt cttccacata 8340

gcacgggcaa aagtgcgtat atatacaaga gcgtttgcca gccacagatt ttcactccac 8400

acaccacatc acacatacaa ccacacacat ccacaatgga acccgaaact aagaagacca 8460

agactgactc caagaagatt gttcttctcg gcggcgactt ctgtggcccc gaggtgattg 8520

ccgaggccgt caaggtgctc aagtctgttg ctgaggcctc cggcaccgag tttgtgtttg 8580

aggaccgact cattggagga gctgccattg agaaggaggg cgagcccatc accgacgcta 8640

ctctcgacat ctgccgaaag gctgactcta ttatgctcgg tgctgtcgga ggcgctgcca 8700

acaccgtatg gaccactccc gacggacgaa ccgacgtgcg acccgagcag ggtctcctca 8760

agctgcgaaa ggacctgaac ctgtacgcca acctgcgacc ctgccagctg ctgtcgccca 8820

agctcgccga tctctccccc atccgaaacg ttgagggcac cgacttcatc attgtccgag 8880

agctcgtcgg aggtatctac tttggagagc gaaaggagga tgacggatct ggcgtcgctt 8940

ccgacaccga gacctactcc gttaattaat ttgaatcgaa tcgatgagcc taaaatgaac 9000

ccgagtatat ctcataaaat tctcggtgag aggtctgtga ctgtcagtac aaggtgcctt 9060

cattatgccc tcaaccttac catacctcac tgaatgtagt gtacctctaa aaatgaaata 9120

cagtgccaaa agccaaggca ctgagctcgt ctaacggact tgatatacaa ccaattaaaa 9180

-continued

```
caaatgaaaa gaaatacagt tctttgtatc atttgtaaca attaccctgt acaaactaag   9240

gtattgaaat cccacaatat tcccaaagtc cacccctttc caaattgtca tgcctacaac   9300

tcatatacca agcactaacc taccgtttaa acagtgtacg cagatctggt gtagtggtag   9360

tgcagtggtg gtattgtgac tggggatgta gttgagaata agtcatacac aagtcagctt   9420

tcttcgagcc tcatataagt ataagtagtt caacgtatta gcactgtacc cagcatctcc   9480

gtatcgagaa acacaacaac atgccccatt ggacagatca tgcggataca caggttgtgc   9540

agtatcatac atactcgatc agacaggtcg tctgaccatc atacaagctg aacaagcgct   9600

ccatacttgc acgctctcta tatacacagt taaattacat atccatagtc taacctctaa   9660

cagttaatct tctggtaagc ctcccagcca gccttctggt atcgcttggc ctcctcaata   9720

ggatctcggt tctggccgta cagacctcgg ccgacaatta tgatatccgt tccggtagac   9780

atgacatcct caacagttcg gtactgctgt ccgagagcgt ctcccttgtc gtcaagaccc   9840

accccggggg tcagaataag ccagtcctca gagtcgccct taggtcggtt ctgggcaatg   9900

aagccaacca caaactcggg gtcggatcgg gcaagctcaa tggtctgctt ggagtactcg   9960

ccagtggcca gagagccctt gcaagacagc tcggccagca tgagcagacc tctggccagc  10020

ttctcgttgg gagaggggac taggaactcc ttgtactggg agttctcgta gtcagagacg  10080

tcctccttct tctgttcaga gacagtttcc tcggcaccag ctcgcaggcc agcaatgatt  10140

ccggttccgg gtacaccgtg ggcgttggtg atatcggacc actcggcgat tcggtgacac  10200

cggtactggt gcttgacagt gttgccaata tctgcgaact ttctgtcctc gaacaggaag  10260

aaaccgtgct taagagcaag ttccttgagg gggagcacag tgccggcgta ggtgaagtcg  10320

tcaatgatgt cgatatgggt tttgatcatg cacacataag gtccgacctt atcggcaagc  10380

tcaatgagct ccttggtggt ggtaacatcc agagaagcac acaggttggt tttcttggct  10440

gccacgagct tgagcactcg agcggcaaag gcggacttgt ggacgttagc tcgagcttcg  10500

taggagggca ttttggtggt gaagaggaga ctgaaataaa tttagtctgc agaacttttt  10560

atcggaacct tatctggggc agtgaagtat atgttatggt aatagttacg agttagttga  10620

acttatagat agactggact atacggctat cggtccaaat tagaaagaac gtcaatggct  10680

ctctgggcgt cgcctttgcc gacaaaaatg tgatcatgat gaaagccagc aatgacgttg  10740

cagctgatat tgttgtcggc caaccgcgcc gaaaacgcag ctgtcagacc cacagcctcc  10800

aacgaagaat gtatcgtcaa agtgatccaa gcacactcat agttggagtc gtactccaaa  10860

ggcggcaatg acgagtcaga cagatactcg tcgacctttt ccttgggaac caccaccgtc  10920

agcccttctg actcacgtat tgtag                                       10945
```

What is claimed is:

1. A method for the expression of a coding region of interest in a transformed yeast comprising:

c) providing a transformed yeast having a genetic construct comprising:

(i) a promoter region of a *Yarrowia* yat1 gene; and,

(ii) a coding region of interest expressible in the yeast cell;

wherein the promoter region is operably linked to the coding region of interest; and,

d) growing the transformed yeast of step (a) under conditions whereby the genetic construct of step (a) is expressed.

2. A method according to claim 1 wherein the *Yarrowia* yat1 gene is isolated from *Yarrowia lipolytica*.

3. A method according to claim 1 wherein the promoter region of the yat1 gene is contained within a nucleic acid molecule as set forth in SEQ ID NO:3.

4. A method according to claim 3 wherein the promoter region contains at least one mutation that does not diminish its promoter activity.

5. A method according to claim 4 wherein the promoter activity is at least about 20% to at least about 400% of the promoter activity of the wildtype promoter activity.

6. A method according to claim 5 wherein the promoter activity is determined by histochemical assay.

7. A method according to claim 1 wherein the transformed yeast is an oleaginous yeast.

8. A method of claim 7, wherein the oleaginous yeast is a member of a genus selected from the group consisting of *Yarrowia, Candida, Rhodotorula, Rhodosporidium, Cryptococcus, Trichosporon* and *Lipomyces*.

9. A method according to claim 8, wherein the oleaginous yeast is selected from the group consisting of: *Yarrowia lipolytica* ATCC #20362, *Yarrowia lipolytica* ATCC #8862, *Yarrowia lipolytica* ATCC #18944, *Yarrowia lipolytica* ATCC #76982 and *Yarrowia lipolytica* LGAM S(7)1.

10. A method according to claim 1 wherein the coding region of interest encodes a polypeptide selected from the group consisting of: desaturases, elongases, acyltransferases, aminopeptidases, amylases, carbohydrases, carboxypeptidases, catalyases, cellulases, chitinases, cutinases, cyclodextrin glycosyltransferases, deoxyribonucleases, esterases, α-galactosidases, β-galactosidases, glucoamylases, α-glucosidases, β-glucanases, β-glucosidases, invertases, laccases, lipases, mannosidases, mutanases, oxidases, pectinolytic enzymes, peroxidases, phospholipases, phytases, polyphenoloxidases, proteolytic enzymes, ribonucleases, transglutaminases and xylanases.

11. A method according to claim 1 wherein the conditions whereby the genetic construct of step (a) is expressed are those of nitrogen-limitation.

12. A method according to claim 11 wherein the conditions of nitrogen-limitation are conditions that induce oleaginy in the transformed yeast.

13. A method for the production of an ω-3 or an ω-6 fatty acid comprising:

d) providing a transformed oleaginous yeast comprising a genetic construct, comprising:

(i) a promoter region of a *Yarrowia* yat1 gene; and,

(ii) a coding region encoding at least one enzyme of the ω-3/ω-6 fatty acid biosynthetic pathway;

wherein the promoter region and coding region are operably linked;

e) culturing the transformed oleaginous yeast of step (a) under conditions whereby the at least one enzyme of the ω-3/ω-6 fatty acid biosynthetic pathway is expressed and a ω-3 or ω-6 fatty acid is produced; and,

f) optionally recovering the ω-3 or ω-6 fatty acid.

14. A method according to claim 13 wherein the *Yarrowia* yat1 gene is isolated from *Yarrowia lipolytica*.

15. A method according to claim 13 wherein the promoter region of the yat1 gene is contained within a nucleic acid molecule as set forth in SEQ ID NO:3.

16. A method according to claim 13 wherein the coding region of interest encodes a polypeptide selected from the group consisting of: desaturases and elongases.

17. A method according to claim 16 wherein the desaturase is selected from the group consisting of: Δ9 desaturase, Δ12 desaturase, Δ6 desaturase, Δ5 desaturase, Δ17 desaturase, Δ15 desaturase, Δ8 desaturase and Δ4 desaturase.

18. A method according to claim 13 wherein the oleaginous yeast is a member of a genus selected from the group of consisting of: *Yarrowia, Candida, Rhodotorula, Rhodosporidium, Cryptococcus, Trichosporon* and *Lipomyces*.

19. A method according to claim 18 wherein the oleaginous yeast is *Yarrowia lipolytica*.

20. A method according to claim 19, wherein the *Yarrowia lipolytica* is a strain selected from the group consisting of: *Yarrowia lipolytica* ATCC #20362, *Yarrowia lipolytica* ATCC #8862, *Yarrowia lipolytica* ATCC #18944, *Yarrowia lipolytica* ATCC #76982 and *Yarrowia lipolytica* LGAM S(7)1.

21. A method according to claim 13 wherein the ω-3 or ω-6 fatty acid is selected from the group consisting of: linoleic acid, α-linolenic acid, γ-linolenic acid, stearidonic acid, dihomo-γ-linoleic acid, eicosatetraenoic acid, arachidonic acid, eicosapentaenoic acid, docosapentaenoic acid, docosahexaenoic acid, eicosadienoic acid and eicosatrienoic acid.

22. A method according to claim 13 wherein the conditions whereby the at least one enzyme of the ω-3/ω-6 fatty acid biosynthetic pathway is expressed are those of nitrogen limitation.

23. A method according to claim 22 wherein the conditions of nitrogen-limitation are conditions that induce oleaginy in the transformed oleaginous yeast.

* * * * *