



(86) Date de dépôt PCT/PCT Filing Date: 2005/11/03

(87) Date publication PCT/PCT Publication Date: 2006/05/18

(85) Entrée phase nationale/National Entry: 2007/04/25

(86) N° demande PCT/PCT Application No.: US 2005/040345

(87) N° publication PCT/PCT Publication No.: 2006/052914

(30) Priorités/Priorities: 2004/11/04 (US60/624,812);
2004/12/29 (US11/024,544)

(51) Cl.Int./Int.Cl. *C12P 7/64* (2006.01),
C07H 21/04 (2006.01), *C12N 1/18* (2006.01),
C12N 15/74 (2006.01), *C12N 9/10* (2006.01),
C12P 21/06 (2006.01)

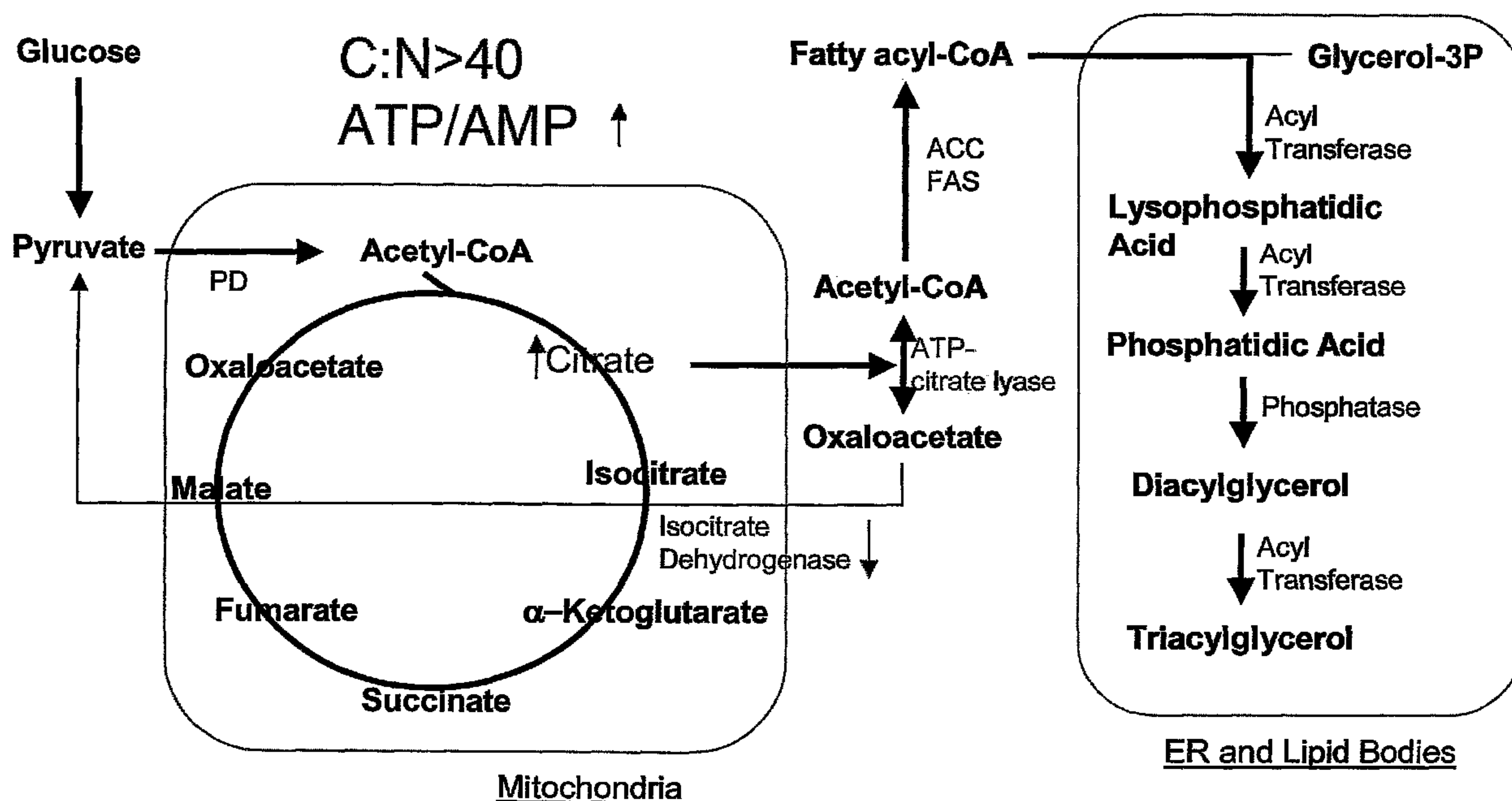
(71) Demandeur/Applicant:
E.I. DU PONT DE NEMOURS AND COMPANY, US

(72) Inventeurs/Inventors:
YADAV, NARENDRA S., US;
XUE, ZHIXIONG, US;
ZHANG, HONGXIANG, US

(74) Agent: BENNETT JONES LLP

(54) Titre : ACYLTRANSFERASES DE DIACYLGLYCEROL POUR LA MODIFICATION D'ACIDES GRAS POLY-
INSATURES ET DE LA TENEUR EN HUILE DANS DES ORGANISMES OLEAGINEUX

(54) Title: DIACYLGLYCEROL ACYLTRANSFERASES FOR ALTERATION OF POLYUNSATURATED FATTY ACIDS
AND OIL CONTENT IN OLEAGINOUS ORGANISMS



(57) Abrégé/Abstract:

Acyltransferases are provided, suitable for use in the manufacture of microbial oils enriched in omega fatty acids in oleaginous yeast (e.g., *Yarrowia lipolytica*). Specifically, genes encoding diacylglycerol acyltransferase (DGAT1) have been isolated from *Y. lipolytica* and *Mortierella alpina*. These genes encode enzymes that participate in the terminal step in oil biosynthesis in yeast. Each is expected to play a key role in altering the quantity of polyunsaturated fatty acids produced in oils of oleaginous yeasts.

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

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
18 May 2006 (18.05.2006)

PCT

(10) International Publication Number
WO 2006/052914 A2

(51) International Patent Classification:

C12P 7/64 (2006.01) *C12N 9/10* (2006.01)
C12P 21/06 (2006.01) *C12N 1/18* (2006.01)
C07H 21/04 (2006.01) *C12N 15/74* (2006.01)

(74) Agent: FELTHAM, S., Neil; E. I. DU PONT DE NEMOURS AND COMPANY, LEGAL PATENT RECORDS CENTER, 4417 Lancaster Pike, Wilmington, Delaware 19805 (US).

(21) International Application Number:

PCT/US2005/040345

(22) International Filing Date:

3 November 2005 (03.11.2005)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

60/624,812 4 November 2004 (04.11.2004) US
11/024,544 29 December 2004 (29.12.2004) US

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

(72) Inventors; and

(75) Inventors/Applicants (for US only): YADAV, Narendra, S. [US/US]; 3704 Knole Lane, Chadds Ford, Pennsylvania 19317 (US). XUE, Zhixiong [CN/US]; 111 Harvey Lane, Chadds Ford, Pennsylvania 19317-9728 (US). ZHANG, Hongxiang [CN/US]; 3808 Rotherfield Lane, Chadds Ford, Pennsylvania 19317 (US).

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

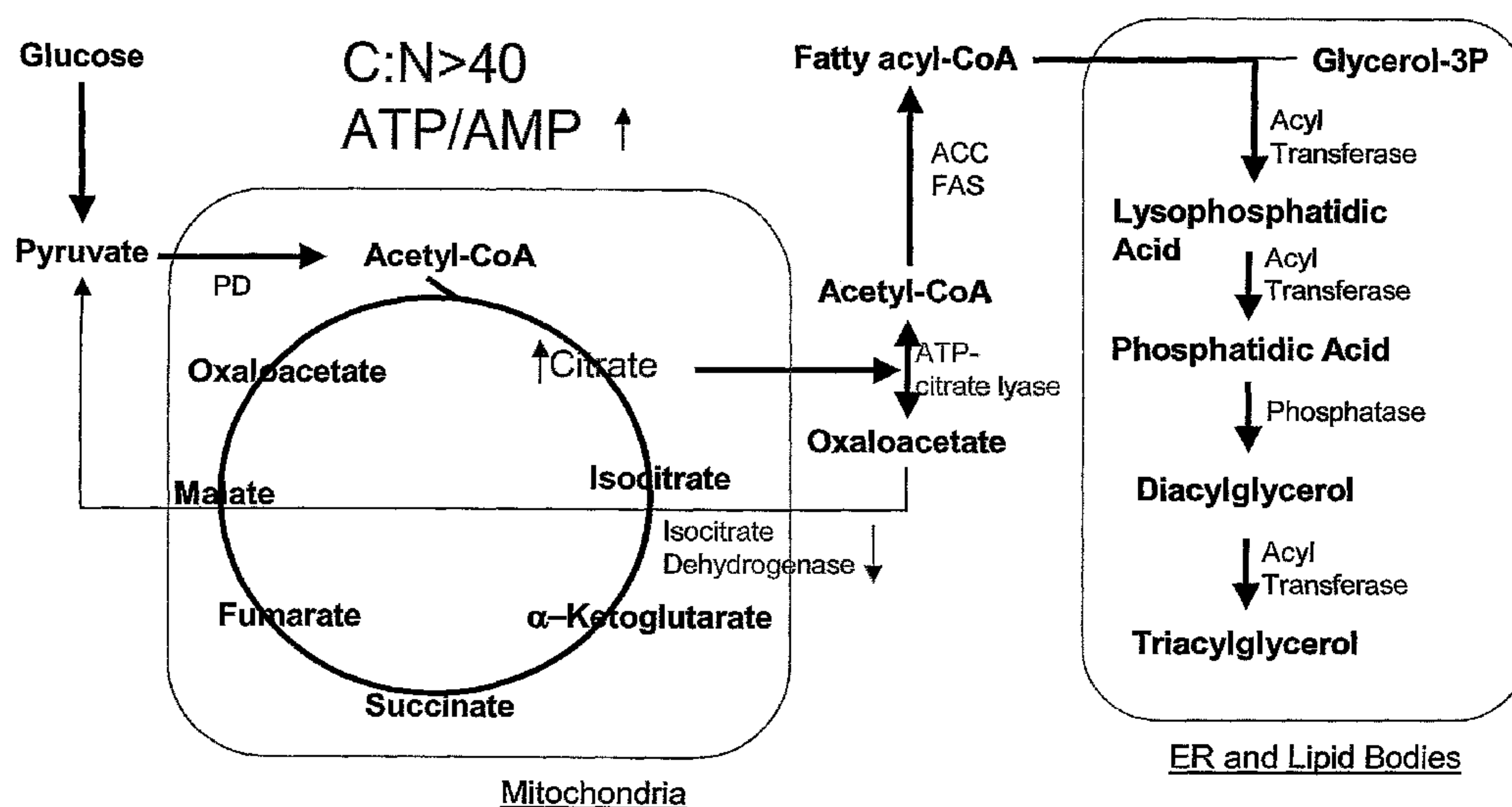
(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

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

[Continued on next page]

(54) Title: DIACYLGLYCEROL ACYLTRANSFERASES FOR ALTERATION OF POLYUNSATURATED FATTY ACIDS AND OIL CONTENT IN OLEAGINOUS ORGANISMS



(57) Abstract: Acyltransferases are provided, suitable for use in the manufacture of microbial oils enriched in omega fatty acids in oleaginous yeast (e.g., *Yarrowia lipolytica*). Specifically, genes encoding diacylglycerol acyltransferase (DGAT1) have been isolated from *Y. lipolytica* and *Mortierella alpina*. These genes encode enzymes that participate in the terminal step in oil biosynthesis in yeast. Each is expected to play a key role in altering the quantity of polyunsaturated fatty acids produced in oils of oleaginous yeasts.

WO 2006/052914 A2

WO 2006/052914 A2



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

DEMANDE OU BREVET VOLUMINEUX

LA PRÉSENTE PARTIE DE CETTE DEMANDE OU CE BREVET COMPREND PLUS D'UN TOME.

CECI EST LE TOME 1 DE 2
CONTENANT LES PAGES 1 À 110

NOTE : Pour les tomes additionels, veuillez contacter le Bureau canadien des brevets

JUMBO APPLICATIONS/PATENTS

THIS SECTION OF THE APPLICATION/PATENT CONTAINS MORE THAN ONE VOLUME

THIS IS VOLUME 1 OF 2
CONTAINING PAGES 1 TO 110

NOTE: For additional volumes, please contact the Canadian Patent Office

NOM DU FICHER / FILE NAME :

NOTE POUR LE TOME / VOLUME NOTE:

TITLEDIACYLGLYCEROL ACYLTRANSFERASES FOR ALTERATION OF
POLYUNSATURATED FATTY ACIDS AND OIL CONTENT IN
OLEAGINOUS ORGANISMS

5 This application claims the benefit of United States Provisional
Application 60/624812, filed November 4, 2004.

FIELD OF THE INVENTION

This invention is in the field of biotechnology. More specifically, this
invention pertains to the identification of nucleic acid fragments encoding
10 diacylglycerol acyltransferases (DGAT1) and acyl-CoA:sterol-
acyltransferases. These enzymes are useful for altering the quantity of oil
in oleaginous microorganisms, such as oleaginous yeasts.

BACKGROUND OF THE INVENTION

The present invention is directed toward the development of an
15 oleaginous yeast that accumulates oils enriched in long-chain ω -3 and/or
 ω -6 polyunsaturated fatty acids ("PUFAs"; e.g., 18:3, 18:4, 20:3, 20:4, 20:5
and 22:6 fatty acids). Toward this end, the natural abilities of oleaginous
yeast (mostly limited to 18:2 fatty acid production) have been enhanced by
advances in genetic engineering, leading to the production of 20:4
20 (arachidonic acid, or "ARA"), 20:5 (eicosapentaenoic acid, or "EPA") and
22:6 (docosahexaenoic acid, or "DHA") PUFAs in transformant *Yarrowia
lipolytica*. These ω -3 and ω -6 fatty acids were produced by introducing
and expressing heterologous genes encoding the ω -3/ ω -6 biosynthetic
pathway in the oleaginous host (see co-pending U.S. Patent Application
25 No. 10/840579, entirely incorporated herein by reference). However, in
addition to developing techniques to introduce the appropriate fatty acid
desaturases and elongases into these particular host organisms, it is also
necessary to increase the transfer of PUFAs into storage lipid pools
following their synthesis.

30 Most free fatty acids become esterified to coenzyme A (CoA), to
yield acyl-CoAs. These molecules are then substrates for glycerolipid
synthesis in the endoplasmic reticulum of the cell, where phosphatidic acid
and diacylglycerol (DAG) are produced. Either of these metabolic
intermediates may be directed to membrane phospholipids (e.g.,
35 phosphatidylglycerol, phosphatidylethanolamine, phosphatidylcholine) or
DAG may be directed to form triacylglycerols (TAGs), the primary storage
reserve of lipids in eukaryotic cells.

Two comprehensive mini-reviews on TAG biosynthesis in yeast, including details concerning the genes involved and the metabolic intermediates that lead to TAG synthesis are: D. Sorger and G. Daum, *Appl. Microbiol. Biotechnol.* 61:289-299 (2003); and H. Müllner and G. Daum, *Acta Biochimica Polonica*, 51(2):323-347 (2004). However, the authors acknowledge that most work performed thus far has focused on the yeast *Saccharomyces cerevisiae* and numerous questions regarding TAG formation and regulation remain.

Briefly, three pathways have been described for the synthesis of TAGs in *S. cerevisiae* (Sandager, L. et al., *J. Biol. Chem.* 277(8):6478-6482 (2002)). First, TAGs are mainly synthesized from DAG and acyl-CoAs by the activity of a diacylglycerol acyltransferase (i.e., DGAT2, encoded by the DGA1 gene). More recently, however, a phospholipid:diacylglycerol acyltransferase (i.e., PDAT, encoded by the LRO1 gene) has also been identified that is responsible for conversion of phospholipid and DAG to lysophospholipid and TAG, respectively, thus producing TAG via an acyl-CoA-independent mechanism (Dahlqvist et al., *PNAS.* 97(12):6487-6492 (2000)). Finally, two acyl-CoA:sterol-acyltransferases (encoded by the ARE1 and ARE2 genes) are known that utilize acyl-CoAs and sterols to produce sterol esters (and TAGs in low quantities; see Sandager, L. et al., *Biochem. Soc. Trans.* 28(6):700-702 (2000)). Together, PDAT and DGAT2 are responsible for approximately 95% of oil biosynthesis in *S. cerevisiae*.

Although homologs of each of the acyltransferase genes described above have been identified in various other organisms and disclosed in the public literature, few genes are available from organisms classified as oleaginous. With respect to yeast, those species included within the genera of *Yarrowia*, *Candida*, *Rhodotorula*, *Rhodospiridium*, *Cryptococcus*, *Trichosporon* and *Lipomyces* and that can accumulate at least 25% of their dry cell weight as oil are classified as oleaginous. Within this unique family of yeast, however, only two acyltransferases have been isolated and characterized. These include a DGAT2 and PDAT from *Yarrowia lipolytica* (see co-pending U.S. Patent Application No. 10/882760, entirely incorporated herein by reference). However, in contrast to the findings in *Saccharomyces cerevisiae*, the *Y. lipolytica* DGAT2 and PDAT were discovered to only be partially responsible for the organism's total oil biosynthesis.

There remains a need there for to identify genes encoding diacylglycerol acyltransferases (DGAT1) and acyl-CoA:sterol-acyltransferases useful for expression in oleaginous yeast for the production of PFUA's. The present work was conducted to identify and
5 characterize the additional gene(s) involved in oil biosynthesis in the oleaginous yeast, *Yarrowia lipolytica*. An understanding of the native mechanisms of oil biosynthesis in this organism is useful, prior to the development of techniques that modify the transfer of recombinantly produced fatty acids (e.g., long-chain PUFAs, such as ARA, EPA and
10 DHA) to the storage lipid pools (i.e., TAG fraction) within transformant oleaginous yeast.

Applicants have solved the stated problem by isolating the genes encoding a diacylglycerol acyltransferase (DGAT1) and an acyl-CoA:sterol-acyltransferase (ARE2) from the oleaginous yeast, *Yarrowia*
15 *lipolytica*. Together, the PDAT, DGAT2 and DGAT1 of *Yarrowia lipolytica* are responsible for up to ~95% of oil biosynthesis (while ARE2 may additionally be a minor contributor to oil biosynthesis). Additionally, an orthologous DGAT1 gene was cloned from *Mortierella alpina* (an oleaginous fungus) and four other fungal DGAT1 orthologs from public
20 sequence databases (i.e., *Neurospora crassa*, *Gibberella zeae* PH-1, *Magnaporthe grisea* and *Aspergillus nidulans*) were identified. With these fungal DGAT1 protein sequences, the Applicants have discovered diagnostic features that will be useful to identify subsequent genes within this family of proteins. These DGAT1 genes will be useful to enable one
25 to modify the transfer of long-chain free fatty acids (e.g., ω -3 and/or ω -6 fatty acids) to the TAG pool in oleaginous yeast.

SUMMARY OF THE INVENTION

The invention relates to the discovery of genes encoding acyltransferase enzymes. The genes and encoded enzymes are useful in
30 manipulating the production of commercially useful oils in microorganisms, and particularly in oleaginous yeasts. Accordingly the invention provides an isolated nucleic acid molecule encoding a diacylglycerol acyltransferase-1 enzyme, selected from the group consisting of:

- (a) an isolated nucleic acid molecule encoding the amino acid
35 sequence selected from the group consisting of SEQ ID NOs:14 and 18;

- (b) an isolated nucleic acid molecule that hybridizes with (a) under the following hybridization conditions: 0.1X SSC, 0.1% SDS, 65 °C and washed with 2X SSC, 0.1% SDS followed by 0.1X SSC, 0.1% SDS; or
- 5 (c) an isolated nucleic acid molecule that is completely complementary to (a) or (b).

Similarly, the invention provides an isolated nucleic acid molecule encoding an acyl-CoA:sterol-acyltransferase, selected from the group consisting of:

- 10 (a) an isolated nucleic acid molecule encoding the amino acid sequence as set forth in SEQ ID NO:16;
- (b) an isolated nucleic acid molecule that hybridizes with (a) under the following hybridization conditions: 0.1X SSC, 0.1% SDS, 65 °C and washed with 2X SSC, 0.1% SDS followed by 0.1X SSC, 0.1% SDS; or
- 15 (c) an isolated nucleic acid molecule that is completely complementary to (a) or (b).

Additionally the invention provides polypeptides encoded by the isolated nucleic acid molecules of the invention as well as genetic chimera and host cells expressing the same.

20

In another embodiment the invention provides an isolated nucleic acid molecule encoding an amino acid motif selected from the group consisting of:

- 25 a) SEQ ID NO:31;
b) SEQ ID NO:32;
c) SEQ ID NO:33;
d) SEQ ID NO:34;
e) SEQ ID NO:35;
f) SEQ ID NO:36;
30 g) SEQ ID NO:37;
h) SEQ ID NO:23;
i) SEQ ID NO:24;
j) SEQ ID NO:25;
k) SEQ ID NO:26;
35 l) SEQ ID NO:27;
m) SEQ ID NO:28;
n) SEQ ID NO:29; and

o) SEQ ID NO:30.

In another embodiment the invention provides an amino acid motif sequence selected from the group consisting of:

- a) SEQ ID NO:31;
- 5 b) SEQ ID NO:32;
- c) SEQ ID NO:33;
- d) SEQ ID NO:34;
- e) SEQ ID NO:35;
- f) SEQ ID NO:36;
- 10 g) SEQ ID NO:37;
- h) SEQ ID NO:23;
- i) SEQ ID NO:24;
- j) SEQ ID NO:25;
- k) SEQ ID NO:26;
- 15 l) SEQ ID NO:27;
- m) SEQ ID NO:28;
- n) SEQ ID NO:29; and
- o) SEQ ID NO:30.

In a preferred embodiment the invention provides a method of increasing triacylglycerol content in a transformed host cell comprising:

- (a) providing a transformed host cell comprising:
 - (i) at least one gene encoding a diacylglycerol acyltransferase 1 enzyme having the amino acid sequence selected from the group consisting of SEQ ID NOs:14, 18, 19, 20, 21 and 22 under the control of suitable regulatory sequences; and,
 - (ii) a source of fatty acids;
- (b) growing the cell of step (a) under conditions whereby the at least one gene encoding a diacylglycerol acyltransferase 1 enzyme is expressed, resulting in the transfer of the fatty acids to triacylglycerol; and
- 30 (c) optionally recovering the triacylglycerol of step (b).

In similar fashion the invention provides a method of increasing the ω -3 or ω -6 fatty acid content of triacylglycerols in a transformed host cell comprising:

- (a) providing a transformed host cell comprising:
- (i) genes encoding a functional ω -3/ ω -6 fatty acid biosynthetic pathway;
 - (ii) at least one gene encoding a diacylglycerol acyltransferase 1 enzyme having the amino acid sequence selected from the group consisting of SEQ ID NOs:14, 18, 19, 20, 21 and 22 under the control of suitable regulatory sequences;
- (b) growing the cell of step (a) under conditions whereby the genes of (i) and (ii) are expressed, resulting in the production of at least one ω -3 or ω -6 fatty acid and its transfer to triacylglycerol; and
- (c) optionally recovering the triacylglycerol of step (b).

Relatedly the invention provides a method of increasing triacylglycerol content in a transformed host cell comprising:

- (a) providing a transformed host cell comprising:
- (i) at least one gene encoding a diacylglycerol acyltransferase-1 enzyme comprising all of the amino acid motifs as set forth in:
 - 1) SEQ ID NO:31;
 - 2) SEQ ID NO:32;
 - 3) SEQ ID NO:33;
 - 4) SEQ ID NO:34;
 - 5) SEQ ID NO:35;
 - 6) SEQ ID NO:36; and
 - 7) SEQ ID NO:37;
 under the control of suitable regulatory sequences; and,
 - (ii) a source of fatty acids;
- (b) growing the cell of step (a) under conditions whereby the at least one gene encoding a diacylglycerol acyltransferase 1 enzyme is expressed, resulting in the transfer of the fatty acids to triacylglycerol; and
- (c) optionally recovering the triacylglycerol of step (b).

In similar fashion the invention provides a method of increasing triacylglycerol content in a transformed host cell comprising:

- (a) providing a transformed host cell comprising:
- (i) at least one gene encoding a diacylglycerol acyltransferase 1 enzyme comprising all of the amino acid motifs as set forth in:
- 5 1) SEQ ID NO:23;
2) SEQ ID NO:24;
3) SEQ ID NO:25;
4) SEQ ID NO:26;
5) SEQ ID NO:27;
10 6) SEQ ID NO:28;
7) SEQ ID NO:29; and
8) SEQ ID NO:30;
- under the control of suitable regulatory sequences; and,
- (ii) a source of fatty acids;
- 15 (b) growing the cell of step (a) under conditions whereby the at least one gene encoding a diacylglycerol acyltransferase-1 enzyme is expressed, resulting in the transfer of the fatty acids to triacylglycerol; and
- (c) optionally recovering the triacylglycerol of step (b).
- 20 Alternatively the invention provides a method of increasing the ω -3 or ω -6 fatty acid content of triacylglycerols in a transformed host cell comprising:
- (a) providing a transformed host cell comprising:
- (i) genes encoding a functional ω -3/ ω -6 fatty acid biosynthetic pathway; and
- 25 (ii) at least one gene encoding a diacylglycerol acyltransferase 1 enzyme comprising all of the amino acid motifs as set forth in:
- 30 1) SEQ ID NO:31;
2) SEQ ID NO:32;
3) SEQ ID NO:33;
4) SEQ ID NO:34;
5) SEQ ID NO:35;
6) SEQ ID NO:36; and
35 7) SEQ ID NO:37;
- under the control of suitable regulatory sequences;

- (b) growing the cell of step (a) under conditions whereby the genes of (i) and (ii) are expressed, resulting in the production of at least one ω -3 or ω -6 fatty acid and its transfer to triacylglycerol; and

- 5 (c) optionally recovering the triacylglycerol of step (b).

In another embodiment the invention provides a method of increasing the ω -3 or ω -6 fatty acid content of triacylglycerols in a transformed host cell comprising:

- (a) providing a transformed host cell comprising:
- 10 (i) genes encoding a functional ω -3/ ω -6 fatty acid biosynthetic pathway; and
- (ii) at least one gene encoding a diacylglycerol acyltransferase 1 enzyme comprising all of the amino acid motifs as set forth in:
- 15 1) SEQ ID NO:23;
- 2) SEQ ID NO:24;
- 3) SEQ ID NO:25;
- 4) SEQ ID NO:26;
- 5) SEQ ID NO:27;
- 20 6) SEQ ID NO:28;
- 7) SEQ ID NO:29; and
- 8) SEQ ID NO:30;

under the control of suitable regulatory sequences;

- (b) growing the cell of step (a) under conditions whereby the genes of (i) and (ii) are expressed, resulting in the production of at least one ω -3 or ω -6 fatty acid and its transfer to triacylglycerol; and
- 25 (c) optionally recovering the triacylglycerol of step (b).

In one embodiment the invention provides a method for the identification of a polypeptide having diacylglycerol acyltransferase-1 activity comprising:

30

- a) obtaining the amino acid sequence of a polypeptide suspected of having diacylglycerol acyltransferase-1 activity; and,
- b) identifying, in the amino acid sequence of the polypeptide of step (a), the presence of all of the amino acid motif sequences as set forth in:
- 35

- 1) SEQ ID NO:31;

- 2) SEQ ID NO:32;
- 3) SEQ ID NO:33;
- 4) SEQ ID NO:34;
- 5) SEQ ID NO:35;
- 6) SEQ ID NO:36; and
- 7) SEQ ID NO:37;

5

wherein the presence of all of the motif sequences of step (a) in the polypeptide is indicative of diacylglycerol acyltransferase-1 activity.

In another embodiment the invention provides a method for the identification of a fungal polypeptide having diacylglycerol acyltransferase-1 activity comprising:

10

- a) obtaining the amino acid sequence of a fungal polypeptide suspected of having diacylglycerol acyltransferase-1 activity; and,
- b) identifying, in the amino acid sequence of the polypeptide of step (a), the presence of all of the amino acid motif sequences as set forth in:

15

- 1) SEQ ID NO:23;
- 2) SEQ ID NO:24;
- 3) SEQ ID NO:25;
- 4) SEQ ID NO:26;
- 5) SEQ ID NO:27;
- 6) SEQ ID NO:28;
- 7) SEQ ID NO:29; and
- 8) SEQ ID NO:30;

20

25

wherein the presence of all of the motif sequences of step (a) in the polypeptide is indicative of diacylglycerol acyltransferase-1 activity.

BRIEF DESCRIPTION OF THE DRAWINGS AND SEQUENCE

30

DESCRIPTIONS

Figure 1 shows a schematic illustration of the biochemical mechanism for lipid accumulation in oleaginous yeast.

Figure 2 illustrates the ω -3 and ω -6 fatty acid biosynthetic pathways.

35

Figure 3 illustrates the construction of plasmid vector pY5 for gene expression in *Yarrowia lipolytica*.

Figure 4 provides plasmid maps for the following plasmids: (A) pY20; (B) pLV13; (C) pMDGAT1-17; and (D) pZUF-Mod-1.

Figure 5 diagrams the development of various *Yarrowia lipolytica* strains producing up to 15% EPA in the total lipid fraction.

5 Figure 6 provides plasmid maps for the following: (A) pKUNF12T6E; (B) pDMW232; (C) pZP3L37; (D) pY37/F15; and (E) pKO2UM26E.

Figure 7 provides plasmid maps for the following: (A) pZKUM; (B) pKO2UF2PE; (C) pZKUT16; (D) pZP2I7+Ura; and (E) pZUF17.

10 Figure 8a, 8b, 8c, 8d, 8e, 8f, 8g and 8h are an alignment of DGAT1 proteins using the Megalign program of DNASTAR using Clustal W.

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

15 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
20 (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.

25 SEQ ID NOs:1-38, 112-115, 117-120, 122, 123, 125, 126, 130, 131, 133-136, 140, 141 and 169-174 are ORFs encoding genes or proteins (or portions thereof) or protein motifs, as identified in Table 1.

Table 1
Summary of Gene and Protein SEQ ID Numbers

30

Description and Abbreviation	Nucleic acid SEQ ID NO.	Protein SEQ ID NO.
<i>Yarrowia lipolytica</i> DGAT2 ("YI DGAT2")	1 (2119 bp)	2 (514 AA)
	3 (1380 bp)	4 (459 AA)
	5 (1068 bp)	6 (355 AA)
<i>Yarrowia lipolytica</i> PDAT ("YI PDAT")	7 (2326 bp)	8 (648 AA)

<i>Yarrowia lipolytica</i> YALI-CDS2011.1 (see also GenBank Accession No. NC_006072, bases 974607-976238, locus_tag="YALI0F06578g")	9 (1632 bp)	10 (543 AA)
<i>Yarrowia lipolytica</i> YALI-CDS2141.1 (see also GenBank Accession No. CR382130, bases 1026155-1027735, locus_tag="YALI0D07986g")	11 (1581 bp)	12 (526 AA)
<i>Yarrowia lipolytica</i> DGAT1 (“YI DGAT1”)	13 (1578 bp)	14 (526 AA)
<i>Yarrowia lipolytica</i> ARE2 (“YI ARE2”)	15 (1632 bp)	16 (543 AA)
<i>Mortierella alpina</i> DGAT1 (“Ma DGAT1”)	17 (1578 bp)	18 (525 AA)
<i>Mortierella alpina</i> DGAT1 –internal cDNA fragment	175 (604 bp)	--
<i>Neurospora crassa</i> DGAT1 (“Nc DAGAT1”)	--	19 (533 AA)
<i>Gibberella zeae</i> DGAT1 (“Fm DAGAT1”)	--	20 (499 AA)
<i>Magnaporthe grisea</i> DGAT1 (“Mg DAGAT1”)	--	21 (503 AA)
<i>Aspergillus nidulans</i> DGAT1 (“An DAGAT1”)	--	22 (458 AA)
Fungal DGAT1 motif #1	--	23
Fungal DGAT1 motif #2	--	24
Fungal DGAT1 motif #3	--	25
Fungal DGAT1 motif #4	--	26
Fungal DGAT1 motif #5	--	27
Fungal DGAT1 motif #6	--	28
Fungal DGAT1 motif #7	--	29
Fungal DGAT1 motif #8	--	30
Universal DGAT1 motif #1	--	31
Universal DGAT1 motif #3	--	32
Universal DGAT1 motif #4	--	33
Universal DGAT1 motif #5	--	34
Universal DGAT1 motif #6	--	35
Universal DGAT1 motif #7	--	36

Universal DGAT1 motif #8	--	37
Fungal DGAT2 motif	--	38
Synthetic elongase gene derived from <i>Mortierella alpina</i> , codon-optimized for expression in <i>Yarrowia lipolytica</i>	112 (957 bp)	113 (318 AA)
Synthetic $\Delta 6$ desaturase, derived from <i>Mortierella alpina</i> , codon-optimized for expression in <i>Yarrowia lipolytica</i>	114 (1374 bp)	115 (457 AA)
<i>Fusarium moniliforme</i> $\Delta 12$ desaturase	117 (1434 bp)	118 (477 AA)
Synthetic elongase gene derived from <i>Thraustochytrium aureum</i> , codon-optimized for expression in <i>Yarrowia lipolytica</i>	119 (819 bp)	120 (272 AA)
<i>Mortierella alpina</i> $\Delta 5$ desaturase	122 (1341 bp)	123 (446 AA)
Synthetic $\Delta 17$ desaturase gene derived from <i>Saprolegnia diclina</i> , codon-optimized for expression in <i>Yarrowia lipolytica</i>	125 (1077 bp)	126 (358 AA)
<i>Yarrowia lipolytica</i> $\Delta 12$ desaturase	130 (1936 bp)	131 (419 AA)
<i>Mortierella isabellina</i> $\Delta 12$ desaturase	133 (1203 bp)	134 (400 AA)
<i>Mortierella alpina</i> $\Delta 6$ desaturase "B"	135 (1521 bp)	136 (458 AA)
Synthetic C ₁₆ elongase gene derived from <i>Rattus norvegicus</i> , codon-optimized for expression in <i>Yarrowia lipolytica</i>	140 (804 bp)	141 (267 AA)
<i>Mus musculus</i> DGAT2 ("Mm DGAT1")	--	169 (388 AA)
<i>Glycine max</i> DGAT1 ("Gm DGAT1")	--	170 (504 AA)
<i>Arabidopsis thaliana</i> DGAT1 ("At DGAT1")	--	171 (520 AA)
<i>Oryza sativa</i> DGAT1 ("Os DGAT1")	--	172 (500 AA)
<i>Perilla frutescens</i> DGAT1 ("Pf DGAT1")	--	173 (534 AA)
<i>Triticum aestivum</i> DGAT1 (Ta DGAT1")	--	174 (508 AA)

SEQ ID NOs:55, 82, 110, 121, 124, 128, 129, 137-139, 144, 164, 165 and 168 are plasmids as identified in Table 2.

Table 2
Summary of Plasmid SEQ ID Numbers

Plasmid	Corresponding Figure	SEQ ID NO
pY20	4A	55 (8,196 bp)
pLV13	4B	82 (5,105 bp)
pKUNF12T6E	6A	110 (12,649 bp)
pDMW232	6B	121 (10,945 bp)
pZP3L37	6C	124 (12,690 bp)
pY37/F15	6D	128 (8,194 bp)
pKO2UM26E	6E	129 (10,448 bp)
pZKUM	7A	137 (4,313 bp)
pKO2UF2PE	7B	138 (10,838 bp)
pZKUT16	7C	139 (5,833 bp)
pZP2I7 + Ura	7D	144 (7,822 bp)
pZUF17	7E	164 (8,165 bp)
pMDGAT1-17	4C	165 (8,666 bp)
pZUF-MOD-1	4D	168 (7,323 bp)

5 SEQ ID NOs:39 and 40 correspond to primers TEF5' and TEF3', respectively, used to isolate the TEF promoter.

 SEQ ID NOs:41 and 42 correspond to primers XPR5' and XPR3', respectively, used to isolate the XPR2 transcriptional terminator.

10 SEQ ID NOs:43-54 correspond to primers YL5, YL6, YL9, YL10, YL7, YL8, YL3, YL4, YL1, YL2, YL61 and YL62, respectively, used for plasmid construction.

 SEQ ID NO:56 corresponds to a 1 kB DNA fragment (amino acid sequence provided as SEQ ID NO:57) containing the *E. coli* hygromycin resistance gene.

15 SEQ ID NO:58 corresponds to a 1.7 kB DNA fragment containing the *Yarrowia* Ura3 gene (amino acid sequence provided as SEQ ID NO:59), which was amplified with primers KU5 and KU3 (SEQ ID NOs:60 and 61, respectively).

20 SEQ ID NOs:62 and 64 are the degenerate primers identified as P7 and P8, respectively, used for the isolation of a *Yarrowia lipolytica* DGAT2.

SEQ ID NOs:63 and 65 are the amino acid consensus sequences that correspond to the degenerate primers P7 and P8, respectively.

SEQ ID NOs:66-68 correspond to primers P80, P81 and LinkAmp Primer1, respectively, used for chromosome walking.

5 SEQ ID NOs:69-72 correspond to primers P95, P96, P97 and P98, respectively, used for targeted disruption of the *Y. lipolytica* DGAT2 gene.

SEQ ID NOs:73-75 correspond to primers P115, P116 and P112, respectively, used to screen for targeted integration of the disrupted *Y. lipolytica* DGAT2 gene.

10 SEQ ID NOs:76 and 78 are the degenerate primers identified as P26 and P27, respectively, used for the isolation of the *Y. lipolytica* PDAT.

SEQ ID NOs:77 and 79 are the amino acid consensus sequences that correspond to degenerate primers P26 and P27, respectively.

15 SEQ ID NOs:80, 81, 83 and 84 correspond to primers P39, P42, P41 and P40, respectively, used for targeted disruption of the *Y. lipolytica* PDAT gene.

SEQ ID NOs:85-88 correspond to primers P51, P52, P37 and P38, respectively, used to screen for targeted integration of the disrupted *Y. lipolytica* PDAT gene.

20 SEQ ID NO:89 corresponds to primer P79, used to amplify the full-length *Y. lipolytica* DGAT2 gene from rescued plasmids.

SEQ ID NOs:90 and 91 correspond to primers P84 and P85, respectively, used to amplify the full-length *Y. lipolytica* PDAT gene from rescued plasmids.

25 SEQ ID NOs:92 and 93 are the degenerate primers identified as P201 and P203, respectively, used for the isolation of the *Y. lipolytica* DGAT1.

30 SEQ ID NOs:94-99 correspond to primers P214, P215, P216, P217, P218 and P219, respectively, used for targeted disruption of the *Y. lipolytica* DGAT1 gene.

SEQ ID NOs:100 and 101 correspond to primers P226 and P227, respectively, used to screen for targeted integration of the disrupted *Y. lipolytica* DGAT1 gene.

35 SEQ ID NOs:102 and 103 are the degenerate primers identified as P205 and P208, respectively, used for isolation of the *Y. lipolytica* ARE2.

SEQ ID NOs:104-109 correspond to primers P220, P221, P222, P223, P224 and P225, respectively, used for targeted disruption of the *Y. lipolytica* ARE2 gene.

5 SEQ ID NOs:111, 116, 127 and 132 correspond to the following *Yarrowia lipolytica* promoters, respectively: fructose-bisphosphate aldolase + intron (FBAIN; 973 bp), fructose-bisphosphate aldolase (FBA; 1001 bp), fructose-bisphosphate aldolase + modified intron (FBAINm; 924 bp), glycerol-3-phosphate acyltransferase (GPAT; 1130 bp).

10 SEQ ID NOs:142 and 143 correspond to primers P239 and P240, respectively, used for sequencing of the *Y. lipolytica* DGAT1 ORF.

SEQ ID NOs:145-147 correspond to BD-Clontech Creator Smart[®] cDNA library kit primers SMART IV oligonucleotide, CDSIII/3' PCR primer and 5'-PCR primer.

15 SEQ ID NO:148 corresponds to the M13 forward primer used for sequencing of the *M. alpina* cDNA library.

SEQ ID NO:149 corresponds to the partial cDNA sequence (601 bp) encoding the putative *M. alpina* DGAT1 gene.

20 SEQ ID NOs:150 and 151 correspond to primers MARE2-N1 and MARE2-N2, respectively, used for cloning the 5'-end region of the putative *M. alpina* DGAT1 gene.

SEQ ID NOs:152 and 153 correspond to the Genome Walker adaptor from ClonTech's Universal GenomeWalker[™] Kit, used for genome-walking.

25 SEQ ID NOs:154 and 155 correspond to primers AP1 and AP2, respectively, used for genome-walking to isolate the 5'-end region of the *M. alpina* DGAT1.

SEQ ID NO:156 corresponds to the 5'-end sequence (1683 bp) of the *M. alpina* DGAT1 cDNA fragment.

30 SEQ ID NOs:157 and 158 correspond to primers ARE-N3-1 and ARE-N3-2, respectively, used for cloning the 3'-end region of the putative *M. alpina* DGAT1 gene.

SEQ ID NOs:159 and 160 correspond to primers AP and UAP, respectively, used for genome-walking to isolate the 3'-end region of the *M. alpina* DGAT1.

35 SEQ ID NO:161 corresponds to the 3'-end sequence (184 bp) of the *M. alpina* DGAT1 cDNA fragment.

SEQ ID NOs:162 and 163 correspond to primers MACAT-F1 and MACAT-R, respectively, used for cloning of the *M. alpina* DGAT1 ORF.

SEQ ID NOs:166 and 167 correspond to primers pzuf-mod1 and pzuf-mod2, respectively, used for creating "control" plasmid pZUF-MOD-1.

5 DETAILED DESCRIPTION OF THE INVENTION

In accordance with the subject invention, Applicants have isolated and confirmed the identity of a *Yarrowia lipolytica* gene encoding a diacylglycerol acyltransferase (DGAT1) enzyme useful for transferring fatty acids into storage triacylglycerols (TAGs). Orthologous genes have also
10 been isolated from *Mortierella alpina* and identified in *Neurospora crassa*, *Gibberella zeae* PH-1, *Magnaporthe grisea* and *Aspergillus nidulans*. Furthermore, the present invention provides motifs for readily identifying other DGAT1 genes from fungal organisms. In another embodiment, the *Yarrowia lipolytica* gene encoding an acyl-CoA:sterol-acyltransferase
15 (ARE2) enzyme has been isolated. Each of these genes may be useful to alter the quantity of long chain polyunsaturated fatty acids (PUFAs) produced in transformant oleaginous yeasts.

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

As such, the subject invention finds many applications. PUFAs, or derivatives thereof, accumulated by the methodology disclosed herein can be used as dietary substitutes, or supplements, particularly infant formulas, for patients undergoing intravenous feeding or for preventing or treating malnutrition. Alternatively, the purified PUFAs (or derivatives thereof) may be incorporated into cooking oils, fats or margarines formulated so that in normal use the recipient would receive the desired amount for dietary supplementation. The PUFAs may also be incorporated into infant formulas, nutritional supplements or other food products and may find use as anti-inflammatory or cholesterol lowering agents. Optionally, the compositions may be used for pharmaceutical use (human or veterinary). 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).

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

25 Definitions

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

"Open reading frame" is abbreviated ORF.

"Polymerase chain reaction" is abbreviated PCR.

30 "American Type Culture Collection" is abbreviated ATCC.

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

"Acyl-CoA:sterol-acyltransferase" is abbreviated ARE2.

"Phospholipid:diacylglycerol acyltransferase" is abbreviated PDAT.

"Diacylglycerol acyltransferase" is abbreviated DAG AT or DGAT.

35 "Diacylglycerol" is abbreviated DAG.

"Triacylglycerols" are abbreviated TAGs.

"Co-enzyme A" is abbreviated CoA.

The term "fatty acids" refers to long chain aliphatic acids (alkanoic acids) of varying chain length, from about C₁₂ to C₂₂ (although both longer and shorter chain-length acids are known). The predominant chain lengths are between C₁₆ and C₂₂. The structure of a fatty acid is represented by a simple notation system of "X:Y", where X is the total number of carbon (C) atoms in the particular fatty acid and Y is the number of double bonds.

Generally, fatty acids are classified as saturated or unsaturated. The term "saturated fatty acids" refers to those fatty acids that have no "double bonds" between their carbon backbone. In contrast, "unsaturated fatty acids" have "double bonds" along their carbon backbones (which are most commonly in the *cis*-configuration). "Monounsaturated fatty acids" have only one "double bond" along the carbon backbone (e.g., usually between the 9th and 10th carbon atom as for palmitoleic acid (16:1) and oleic acid (18:1)), while "polyunsaturated fatty acids" (or "PUFAs") have at least two double bonds along the carbon backbone (e.g., between the 9th and 10th, and 12th and 13th carbon atoms for linoleic acid (18:2); and between the 9th and 10th, 12th and 13th, and 15th and 16th for α -linolenic acid (18:3)).

"PUFAs" can be classified into two major families (depending on the position (*n*) of the first double bond nearest the methyl end of the fatty acid carbon chain). Thus, the "omega-6 fatty acids" (ω -6 or *n*-6) have the first unsaturated double bond six carbon atoms from the omega (methyl) end of the molecule and additionally have a total of two or more double bonds, with each subsequent unsaturation occurring 3 additional carbon atoms toward the carboxyl end of the molecule. In contrast, the "omega-3 fatty acids" (ω -3 or *n*-3) have the first unsaturated double bond three carbon atoms away from the omega end of the molecule and additionally have a total of three or more double bonds, with each subsequent unsaturation occurring 3 additional carbon atoms toward the carboxyl end of the molecule.

For the purposes of the present disclosure, the omega-reference system will be 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). This nomenclature is shown below in Table 3, in the column titled "Shorthand Notation". The remainder of the Table summarizes the

common names of ω -3 and ω -6 fatty acids, the abbreviations that will be used throughout the specification and each compounds' chemical name.

Table 3

5

Nomenclature Of Polyunsaturated Fatty Acids

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

"Microbial oils" or "single cell oils" are those oils naturally produced by microorganisms (e.g., algae, oleaginous yeasts and filamentous fungi) during their lifespan. The term "oil" refers to a lipid substance that is liquid at 25 °C and usually polyunsaturated. In contrast, the term "fat" refers to a lipid substance that is solid at 25 °C and usually saturated.

"Lipid bodies" refer to lipid droplets that usually are bounded by specific proteins and a monolayer of phospholipid. These organelles are sites where most organisms transport/store neutral lipids. Lipid bodies are thought to arise from microdomains of the endoplasmic reticulum that contain TAG-biosynthesis enzymes; and, their synthesis and size appear to be controlled by specific protein components.

“Neutral lipids” refer to those lipids commonly found in cells in lipid bodies as storage fats and oils and are so called because at cellular pH, the lipids bear no charged groups. Generally, they are completely non-polar with no affinity for water. Neutral lipids generally refer to mono-, di-, and/or triesters of glycerol with fatty acids, also called monoacylglycerol, diacylglycerol or TAG, respectively (or collectively, acylglycerols). A hydolysis reaction must occur to release free fatty acids from acylglycerols.

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

The term “DAG AT” refers to a diacylglycerol acyltransferase (also known as an acyl-CoA-diacylglycerol acyltransferase or a diacylglycerol O-acyltransferase) (EC 2.3.1.20). This enzyme is responsible for the conversion of acyl-CoA and 1,2-diacylglycerol to TAG and CoA (thereby involved in the terminal step of TAG biosynthesis). Two families of DAG AT enzymes exist: DGAT1 and DGAT2. The former family shares homology with the acyl-CoA:cholesterol acyltransferase (ACAT) gene family, while the latter family is unrelated (Lardizabal et al., *J. Biol. Chem.* 276(42):38862-28869 (2001)).

The term “PDAT” refers to a phospholipid:diacylglycerol acyltransferase enzyme (EC 2.3.1.158). This enzyme is responsible for the transfer of an acyl group from the *sn*-2 position of a phospholipid to the *sn*-3 position of 1,2-diacylglycerol, thus resulting in lysophospholipid and TAG (thereby involved in the terminal step of TAG biosynthesis). This enzyme differs from DGAT (EC 2.3.1.20) by synthesizing TAG via an acyl-CoA-independent mechanism.

The term “ARE2” refers to an acyl-CoA:sterol-acyltransferase enzyme (EC 2.3.1.26; also known as a sterol-ester synthase 2 enzyme), catalyzing the following reaction: $\text{acyl-CoA} + \text{cholesterol} = \text{CoA} + \text{cholesterol ester}$.

The term “PUFA biosynthetic pathway enzyme” refers to any of the following enzymes (and genes which encode said enzymes) associated

with the biosynthesis of a PUFA, including: a $\Delta 4$ desaturase, a $\Delta 5$ desaturase, a $\Delta 6$ desaturase, a $\Delta 12$ desaturase, a $\Delta 15$ desaturase, a $\Delta 17$ desaturase, a $\Delta 9$ desaturase, a $\Delta 8$ desaturase and/or an elongase(s).

The term " ω -3/ ω -6 fatty acid biosynthetic pathway" refers to a set of genes which, when expressed under the appropriate conditions encode enzymes that catalyze the production of either or both ω -3 and ω -6 fatty acids. Typically the genes involved in the ω -3/ ω -6 fatty acid biosynthetic pathway encode some or all of the following enzymes: $\Delta 12$ desaturase, $\Delta 6$ desaturase, elongase(s), $\Delta 5$ desaturase, $\Delta 17$ desaturase, $\Delta 15$ desaturase, $\Delta 9$ desaturase, $\Delta 8$ desaturase and $\Delta 4$ desaturase. A representative pathway is illustrated in Figure 2, providing for the conversion of oleic acid through various intermediates to DHA, which demonstrates how both ω -3 and ω -6 fatty acids may be produced from a common source. The pathway is naturally divided into two portions where one portion will generate ω -3 fatty acids and the other portion, only ω -6 fatty acids. That portion that only generates ω -3 fatty acids will be referred to herein as the ω -3 fatty acid biosynthetic pathway, whereas that portion that generates only ω -6 fatty acids will be referred to herein as the ω -6 fatty acid biosynthetic pathway.

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

The term "desaturase" refers to a polypeptide that can desaturate, i.e., introduce a double bond, in one or more fatty acids to produce a mono- or polyunsaturated fatty acid. Despite use of the omega-reference system throughout the specification in reference to specific fatty acids, it is more convenient to indicate the activity of a desaturase by counting from the carboxyl end of the substrate using the delta-system. Of particular interest herein are: $\Delta 12$ desaturases that desaturate a fatty acid between the 12th and 13th carbon atoms numbered from the carboxyl-terminal end of the molecule and that catalyze the conversion of oleic acid to LA; $\Delta 15$ desaturases that catalyze the conversion of LA to ALA; $\Delta 17$ desaturases that catalyze the conversion of ARA to EPA and/or DGLA to ETA; $\Delta 6$

desaturases that catalyze the conversion of LA to GLA and/or ALA to STA;
Δ5 desaturases that catalyze the conversion of DGLA to ARA and/or ETA
to EPA; Δ4 desaturases that catalyze the conversion of DPA to DHA; Δ8
desaturases that catalyze the conversion of EDA to DGLA and/or ETrA to
5 ETA; and Δ9 desaturases that catalyze the conversion of palmitate to
palmitoleic acid (16:1) and/or stearate to oleic acid (18:1).

The term "elongase" refers to a polypeptide that can elongate a
fatty acid carbon chain to produce an acid that is 2 carbons longer than
the fatty acid substrate that the elongase acts upon. This process of
10 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)). Briefly, malonyl-CoA is condensed with a long-chain
acyl-CoA to yield CO₂ and a β-ketoacyl-CoA (where the acyl moiety has
been elongated by two carbon atoms). Subsequent reactions include
15 reduction to β-hydroxyacyl-CoA, dehydration to an enoyl-CoA and a
second reduction to yield the elongated acyl-CoA. Examples of reactions
catalyzed by elongases are the conversion of GLA to DGLA, STA to ETA,
and EPA to DPA. Accordingly, elongases can have different specificities.
For example, a C_{16/18} elongase will prefer a C₁₆ substrate, a C_{18/20}
20 elongase will prefer a C₁₈ substrate and a C_{20/22} elongase will prefer a C₂₀
substrate. In like manner, a Δ9 elongase is able to catalyze the
conversion of LA and ALA to EDA and ETrA, respectively.

The term "oleaginous" refers to those organisms that tend to store
their energy source in the form of lipid (Weete, In: *Fungal Lipid*
25 *Biochemistry*, 2nd Ed., Plenum, 1980). Generally, the cellular oil 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.*
30 *Microbiol.* 57:419-25 (1991)).

The term "oleaginous yeast" refers to those microorganisms
classified as yeasts that can make oil. Generally, the cellular oil or
triacylglycerol content of oleaginous microorganisms follows a sigmoid
curve, wherein the concentration of lipid increases until it reaches a
35 maximum at the late logarithmic or early stationary growth phase and then
gradually decreases during the late stationary and death phases
(Yongmanitchai and Ward, *Appl. Environ. Microbiol.* 57:419-25 (1991)). It

is not uncommon for oleaginous microorganisms to accumulate in excess of about 25% of their dry cell weight as oil. Examples of oleaginous yeast include, but are no means limited to, the following genera: *Yarrowia*, *Candida*, *Rhodotorula*, *Rhodospiridium*, *Cryptococcus*, *Trichosporon* and
5 *Lipomyces*.

The term "fermentable carbon source" means a carbon source that a microorganism will metabolize to derive energy. Typical carbon sources of the invention include, but are not limited to: monosaccharides, oligosaccharides, polysaccharides, alkanes, fatty acids, esters of fatty
10 acids, monoglycerides, carbon dioxide, methanol, formaldehyde, formate and carbon-containing amines.

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

A nucleic acid molecule is "hybridizable" to another nucleic acid molecule, such as a cDNA, genomic DNA, or RNA molecule, when a
20 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
25 Laboratory: Cold Spring Harbor, NY (1989), particularly Chapter 11 and Table 11.1 therein (entirely incorporated herein by reference). The conditions of temperature and ionic strength determine the "stringency" of the hybridization. Stringency conditions can be adjusted to screen for moderately similar fragments (such as homologous sequences from
30 distantly related organisms), to highly similar fragments (such as genes that duplicate functional enzymes from closely related organisms). Post-hybridization washes determine stringency conditions. One set of preferred conditions uses a series of washes starting with 6X SSC, 0.5% SDS at room temperature for 15 min, then repeated with 2X SSC, 0.5%
35 SDS at 45 °C for 30 min, and then repeated twice with 0.2X SSC, 0.5% SDS at 50 °C for 30 min. A more preferred set of stringent conditions uses higher temperatures in which the washes are identical to those

above except for the temperature of the final two 30 min washes in 0.2X SSC, 0.5% SDS was increased to 60 °C. Another preferred set of highly stringent conditions uses two final washes in 0.1X SSC, 0.1% SDS at 65 °C. An additional set of stringent conditions include hybridization at 0.1X
5 SSC, 0.1% SDS, 65 °C and washed with 2X SSC, 0.1% SDS followed by 0.1X SSC, 0.1% SDS, for example.

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

A "substantial portion" of an amino acid or nucleotide sequence is
30 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
35 Search Tool; Altschul, S. F., et al., *J. Mol. Biol.* 215:403-410 (1993)). In general, a sequence of ten or more contiguous amino acids or thirty or more nucleotides is necessary in order to putatively identify a polypeptide

or nucleic acid sequence as homologous to a known protein or gene.

Moreover, with respect to nucleotide sequences, gene specific oligonucleotide probes comprising 20-30 contiguous nucleotides may be used in sequence-dependent methods of gene identification (e.g.,

5 Southern hybridization) and isolation (e.g., *in situ* hybridization of bacterial colonies or bacteriophage plaques). In addition, short oligonucleotides of 12-15 bases may be used as amplification primers in PCR in order to obtain a particular nucleic acid fragment comprising the primers.

Accordingly, a "substantial portion" of a nucleotide sequence comprises
10 enough of the sequence to specifically identify and/or isolate a nucleic acid fragment comprising the sequence. The instant specification teaches partial or complete amino acid and nucleotide sequences encoding one or more particular yeast and fungal proteins. The skilled artisan, having the benefit of the sequences as reported herein, may now use all or a
15 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.

The term "complementary" is used to describe the relationship
20 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 fragments that are complementary to the complete sequences as reported in the
25 accompanying Sequence Listing, as well as those substantially similar nucleic acid sequences.

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.
30 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
35 (Lesk, A. M., Ed.) Oxford University: NY (1988); 2.) Biocomputing: Informatics and Genome Projects (Smith, D. W., Ed.) Academic: NY (1993); 3.) Computer Analysis of Sequence Data, Part I (Griffin, A. M., and

Griffin, H. G., Eds.) Humana: NJ (1994); 4.) Sequence Analysis in Molecular Biology (von Heinje, G., Ed.) Academic (1987); and 5.) Sequence Analysis Primer (Gribskov, M. and Devereux, J., Eds.) Stockton: NY (1991). Preferred methods to determine identity are designed to give the best match between the sequences tested. Methods to determine identity and similarity are codified in publicly available computer programs. Sequence alignments and percent identity calculations may be performed using the Megalign program of the LASERGENE bioinformatics computing suite (DNASTAR Inc., Madison, WI). Multiple alignment of the sequences is performed using the Clustal method of alignment (Higgins and Sharp, *CABIOS*. 5:151-153 (1989)) with default parameters (GAP PENALTY=10, GAP LENGTH PENALTY=10), unless otherwise specified. Default parameters for pairwise alignments using the Clustal method are: KTUPLE 1, GAP PENALTY=3, WINDOW=5 and DIAGONALS SAVED=5.

Suitable nucleic acid fragments (isolated polynucleotides of the present invention) encode polypeptides that are at least about 70% identical, preferably at least about 75% identical, and more preferably at least about 80% identical to the amino acid sequences reported herein. Preferred nucleic acid fragments encode amino acid sequences that are about 85% identical to the amino acid sequences reported herein. More preferred nucleic acid fragments encode amino acid sequences that are at least about 90% identical to the amino acid sequences reported herein. Most preferred are nucleic acid fragments that encode amino acid sequences that are at least about 95% identical to the amino acid sequences reported herein. Suitable nucleic acid fragments not only have the above homologies but typically encode a polypeptide having at least 50 amino acids, preferably at least 100 amino acids, more preferably at least 150 amino acids, still more preferably at least 200 amino acids, and most preferably at least 250 amino acids.

"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.

The term “codon-optimized”, as it refers to genes or coding regions of nucleic acid molecules, refers to modification of codons such that the altered codons reflect the typical codon usage of the host organism without altering the polypeptide for which the DNA codes.

“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.

“Gene” refers to a nucleic acid fragment 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. “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.

5 "Coding sequence" refers to a DNA sequence that codes for a specific amino acid sequence. "Suitable regulatory sequences" refer to nucleotide sequences located upstream (5' non-coding sequences), within, or downstream (3' non-coding sequences) of a coding sequence, and which influence the transcription, RNA processing or stability, or translation of the associated coding sequence. Regulatory sequences
10 may include promoters, translation leader sequences, introns, polyadenylation recognition sequences, RNA processing sites, effector binding sites and stem-loop structures.

"Promoter" refers to a DNA sequence capable of controlling the expression of a coding sequence or functional RNA. In general, a coding
15 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
20 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
25 have not been completely defined, DNA fragments of different lengths may have identical promoter activity.

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
30 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.

35 "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. 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.

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

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

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

The terms "plasmid", "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.

5 "Transformation cassette" refers to a specific vector containing a foreign gene(s) and having elements in addition to the foreign gene(s) that facilitate transformation of a particular host cell. "Expression cassette" refers to a specific vector containing a foreign gene(s) and having

10 elements in addition to the foreign gene(s) that allow for enhanced expression of that gene in a foreign host.

The term "homologous recombination" refers to the exchange of DNA fragments between two DNA molecules (during cross over). The fragments which are exchanged are flanked by sites of identical nucleotide sequences between the two DNA molecules (i.e., "regions of homology").

15 The term "regions of homology" refer to stretches of nucleotide sequence on nucleic acid fragments that participate in homologous recombination that have homology to each other. Effective homologous recombination will generally take place where these regions of homology are at least

20 about 10 bp in length where at least about 50 bp in length is preferred. Typically fragments that are intended for recombination contain at least two regions of homology where targeted gene disruption or replacement is desired.

The term "sequence analysis software" refers to any computer

25 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

30 (GCG), Madison, WI); 2.) BLASTP, BLASTN, BLASTX (Altschul et al., *J. Mol. Biol.* 215:403-410 (1990)); 3.) DNASTAR (DNASTAR, Inc. Madison, WI); 4.) Sequencher (Gene Codes Corporation, Ann Arbor, MI); and 5.) the FASTA program incorporating the Smith-Waterman algorithm (W. R. Pearson, *Comput. Methods Genome Res.*, [Proc. Int. Symp.]

35 (1994), Meeting Date 1992, 111-20. Suhai, Sandor, Ed. Plenum: New York, NY). Within the context of this application it will be understood that where sequence analysis software is used for analysis, that the results of

the analysis will be based on the "default values" of the program referenced, unless otherwise specified. As used herein "default values" will mean any set of values or parameters that originally load with the software when first initialized.

5 The term "conserved domain" or "motif" means a set of amino acids conserved at specific positions along an aligned sequence of evolutionarily related proteins. While amino acids at other positions can vary between homologous proteins, amino acids that are highly conserved at specific positions indicate amino acids that are essential in the structure, the
10 stability, or the activity of a protein. Because they are identified by their high degree of conservation in aligned sequences of a family of protein homologues, they can be used as identifiers, or "signatures", to determine if a protein with a newly determined sequence belongs to a previously identified protein family. Motifs that are universally found in DGAT1
15 enzymes (i.e., animal, plants and fungi) are provided as SEQ ID NOs:31-37; motifs found in DGAT1s that are specific to fungal organisms are provided as SEQ ID NOs:23-30.

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

Microbial Biosynthesis Of Fatty Acids and Triacylglycerols

In general, lipid accumulation in oleaginous microorganisms is triggered in response to the overall carbon to nitrogen ratio present in the
30 growth medium (Figure 1). When cells have exhausted available nitrogen supplies (e.g., when the carbon to nitrogen ratio is greater than about 40), the depletion of cellular adenosine monophosphate (AMP) leads to the cessation of AMP-dependent isocitrate dehydrogenase activity in the mitochondria and the accumulation of citrate, transport of citrate into the
35 cytosol, and subsequent cleavage of the citrate by ATP-citrate lyase to yield acetyl-CoA and oxaloacetate. Acetyl-CoA is the principle building block for *de novo* biosynthesis of fatty acids. Although any compound that

can effectively be metabolized to acetyl-CoA can serve as a precursor of fatty acids, glucose is the primary source of carbon in this type of reaction (Figure 1). Glucose is converted to pyruvate via glycolysis, and pyruvate is then transported into the mitochondria where it can be converted to acetyl-CoA by pyruvate dehydrogenase ("PD"). Since acetyl-CoA can not be transported directly across the mitochondrial membrane into the cytoplasm, the two carbons from acetyl-CoA condense with oxaloacetate to yield citrate (catalyzed by citrate synthase). Citrate is transported directly into the cytoplasm, where it is cleaved by ATP-citrate lyase to regenerate acetyl-CoA and oxaloacetate. The oxaloacetate reenters the tricarboxylic acid cycle, via conversion to malate.

The synthesis of malonyl-CoA is the first committed step of fatty acid biosynthesis, which takes place in the cytoplasm. Malonyl-CoA is produced via carboxylation of acetyl-CoA by acetyl-CoA carboxylase ("ACC"). Fatty acid synthesis is catalyzed by a multi-enzyme fatty acid synthase complex ("FAS") and occurs by the condensation of eight two-carbon fragments (acetyl groups from acetyl-CoA) to form a 16-carbon saturated fatty acid, palmitate. More specifically, FAS catalyzes a series of 7 reactions, which involve the following (Smith, S. *FASEB J*, 8(15):1248-59 (1994)):

1. Acetyl-CoA and malonyl-CoA are transferred to the acyl carrier peptide (ACP) of FAS. The acetyl group is then transferred to the malonyl group, forming β -ketobutyryl-ACP and releasing CO₂.
2. The β -ketobutyryl-ACP undergoes reduction (via β -ketoacyl reductase) and dehydration (via β -hydroxyacyl dehydratase) to form a *trans*-monounsaturated fatty acyl group.
3. The double bond is reduced by NADPH, yielding a saturated fatty-acyl group two carbons longer than the initial one. The butyryl-group's ability to condense with a new malonyl group and repeat the elongation process is then regenerated.
4. When the fatty acyl group becomes 16 carbons long, a thioesterase activity hydrolyses it, releasing free palmitate (16:0).

Whereas palmitate synthesis occurs in the cytosol, formation of longer chain saturated and unsaturated fatty acid derivatives occur in both the mitochondria and endoplasmic reticulum (ER), wherein the ER is the dominant system. Specifically, palmitate (16:0) is the precursor of stearic (18:0), palmitoleic (16:1) and oleic (18:1) acids through the action of

elongases and desaturases. For example, palmitate and stearate are converted to their unsaturated derivatives, palmitoleic (16:1) and oleic (18:1) acids, respectively, by the action of a $\Delta 9$ desaturase.

5 TAGs (the primary storage unit for fatty acids) are formed by a series of reactions that involve: 1.) the esterification of one molecule of acyl-CoA to glycerol-3-phosphate via an acyltransferase to produce lysophosphatidic acid; 2.) the esterification of a second molecule of acyl-CoA via an acyltransferase to yield 1,2-diacylglycerol phosphate (commonly identified as phosphatidic acid); 3.) removal of a phosphate by
10 phosphatidic acid phosphatase to yield 1,2-diacylglycerol (DAG); and 4.) the addition of a third fatty acid by the action of another acyltransferase (e.g., PDAT, DGAT1 or DGAT2) to form TAG (Figure 1).

A wide spectrum of fatty acids can be incorporated into TAGs, including saturated and unsaturated fatty acids and short-chain and long-
15 chain fatty acids. Some non-limiting examples of fatty acids that can be incorporated into TAGs by acyltransferases (e.g., DAG ATs or PDAT) include: capric (10:0), lauric (12:0), myristic (14:0), palmitic (16:0), palmitoleic (16:1), stearic (18:0), oleic (18:1), vaccenic (18:1), linoleic (18:2), eleostearic (18:3), γ -linolenic (18:3), α -linolenic (18:3), stearidonic
20 (18:4), arachidic (20:0), eicosadienoic (20:2), dihomo- γ -linoleic (20:3), eicosatrienoic (20:3), arachidonic (20:4), eicosa-tetraenoic (20:4), eicosa-pentaenoic (20:5), behenic (22:0), docosa-pentaenoic (22:5), docosa-hexaenoic (22:6), lignoceric (24:0), nervonic (24:1), cerotic (26:0) and montanic (28:0) fatty acids. In preferred embodiments of the present
25 invention, incorporation of PUFAs into TAG is most desirable.

Acyltransferases And Their Role In The Terminal Step of TAG Biosynthesis

Genes Encoding DGAT1

Historically, DGAT1 (responsible for the third acyl transferase
30 reaction, wherein an acyl-CoA group is transferred from acyl-CoA to the *sn*-3 position of DAG to form TAG) was thought to be the only enzyme specifically involved in TAG synthesis. This enzyme was known to be homologous to acyl-CoA:cholesterol acyltransferases (ACATs); however, recent studies have identified a new family of DAG acyltransferase (DAG
35 AT) enzymes that are unrelated to the ACAT gene family. Thus, nomenclature now distinguishes between the DAG AT enzymes that are related to the ACAT gene family (DGAT1 family) versus those that are

unrelated (DGAT2 family) (Lardizabal et al., *J. Biol. Chem.* 276(42):38862-28869 (2001)).

Many genes encoding DGAT1 enzymes have been identified through genetic means and the DNA sequences of some of these genes are publicly available. For example, some non-limiting examples include the following GenBank Accession Numbers: AY445635 (olive); AF384160 (mouse); NM_053437 (Norway rat); NM_174693 (cow); AY116586 (pig); AY327327 and AY327326 (*Toxoplasma gondii*); AF298815 (*Perilla frutescens*); and AF164434 (*Brassica napus*). Additionally, the patent literature provides many additional DNA sequences of DGAT1 genes (and/or details concerning several of the genes above and their methods of isolation). See, for example: U.S. 6,100,077 (human); U.S. 6,552,250 (*Brassica*); U.S. 6,344,548 (human, mouse, *Arabidopsis*); US 2004/0088759A1 (plant); and US 2004/0078836A1 (Farese et al.).

15 Genes Encoding DGAT2

Members of the DGAT2 family appear to be present in all major phyla of eukaryotes (fungi, plants, animals and basal eukaryotes). As such, many genes encoding DGAT2 enzymes have been identified through genetic means and the DNA sequences of some of these genes are publicly available. For example, some non-limiting examples include the following GenBank Accession Numbers: NC_001147 (locus NP_014888; *Saccharomyces cerevisiae*); NM_012079 (human); NM_127503, AF051849 and AJ238008 (*Arabidopsis thaliana*); NM_026384, NM_010046 and AB057816 (mouse); AY093657 (pig); AB062762 (rat); AF221132 (*Caenorhabditis elegans*); AF391089 and AF391090 (*Mortierella ramanniana*); AF129003 (*Nicotiana tabacum*); and, AF251794 and AF164434 (*Brassica napus*). Additionally, the patent literature provides many additional DNA sequences of DGAT2 genes (and/or details concerning several of the genes above and their methods of isolation). See, for example: US 2003/124126 (Cases et al.); WO 2001/034814 (Banas et al.); and US 2003/115632, US 2003/0028923 and US 2004/0107459 (Lardizabal et al.). The work of Lardizabal et al. includes DNA sequences of DGAT2s from, e.g., *Mortierella ramanniana*, *Neurospora crassa*, *Saccharomyces cerevisiae*, *Hordeum vulgare*, *Zea mays*, *Glycine max*, *Triticum aestivum*, *Drosophila*, *Homo sapiens*, *Schizosaccharomyces pombe*, *Candida albicans* and *Arabidopsis thaliana*.

Most recently, a DGAT2 enzyme from the oleaginous yeast *Yarrowia lipolytica* has been isolated and characterized in co-pending U.S. Patent Application No. 10/882760 (incorporated entirely herein by reference). Briefly, following cloning of a partial putative DGAT2 DNA
5 fragment from *Y. lipolytica*, targeted disruption of the endogenous *Y. lipolytica* gene was carried out to test the identity of the fragment. Lower oil content in the disrupted strain confirmed that the native DGAT2 activity was eliminated. Subsequently, a full-length *Y. lipolytica* DGAT2 gene (2119 bp; SEQ ID NO:1) was assembled, which included three nested
10 open reading frames: 1.) ORF 1: nucleotides +291 to +1835 of SEQ ID NO:1, corresponding to the protein encoded by SEQ ID NO:2 (514 amino acid residues); 2.) ORF 2: nucleotides +456 to +1835 of SEQ ID NO:1, corresponding to SEQ ID NO:3 (1380 bases) and the protein encoded by SEQ ID NO:4 (459 amino acid residues); and 3.) ORF 3: nucleotides +768
15 to +1835 of SEQ ID NO:1, corresponding to SEQ ID NO:5 (1068 bases) and the protein encoded by SEQ ID NO:6 (355 amino acid residues).

Genes Encoding PDAT

TAG synthesis can also occur in the absence of acyl-CoA, via the acyl-CoA-independent PDAT enzyme, as recently discovered by Dahlqvist
20 et al. (*Proc. Nat. Acad. Sci. (USA)* 97:6487-6492 (2000)) and Oelkers et al. (*J. Biol. Chem.* 275:15609-15612 (2000)). Specifically, PDAT removes an acyl group from the *sn*-2 position of a phosphatidylcholine substrate for transfer to the *sn*-3 position of DAG to produce TAG; and, although the function of PDAT is not as well characterized as DGAT2, PDAT has been
25 postulated to play a major role in removing "unusual" fatty acids from phospholipids in some oilseed plants (Banas, A. et al., *Biochem. Soc. Trans.* 28(6):703-705 (2000)).

PDAT is structurally related to the lecithin:cholesterol acyltransferase (LCAT) family of proteins. Several genes encoding PDAT
30 enzymes have been identified through genetic means and the DNA sequences of some of these genes are publicly available. For example, some non-limiting examples include the following GenBank Accession Numbers: P40345 (*Saccharomyces cerevisiae*); O94680 and NP_596330 (*Schizosaccharomyces pombe*); and, NP_190069 and AB006704
35 [gi:2351069] (*Arabidopsis thaliana*). Additionally, the patent literature provides many additional DNA sequences of PDAT genes (and/or details

concerning several of the genes above and their methods of isolation); see, for example, WO 2000/060095 (Dahlqvist et al.).

And, particularly relevant to the disclosure herein, a PDAT enzyme from the oleaginous yeast *Yarrowia lipolytica* has been isolated and characterized in co-pending U.S. Patent Application No. 10/882760 in a manner similar to that described above for DGAT2. Again, lower oil content in a strain having a disruption in the putative PDAT gene confirmed that the native PDAT activity was eliminated. Subsequently, a full-length *Y. lipolytica* PDAT gene (2326 bp; SEQ ID NO:7) was assembled.

Genes Encoding ARE2

The process of sterol esterification in yeast was first studied by H. Yang et al. (*Science*. 272(5266):1353-1356 (1996)), wherein it was discovered that two genes (ARE1 and ARE2) encode ACAT-related enzymes that permit the esterification of cholesterol. The DNA sequences of only a few ARE2 genes are publicly available. For example, see GenBank Accession Numbers: Q876L2 (*Saccharomyces bayanus*), P53629 (*S. cerevisiae*) and Q10269 (*Schizosaccharomyces pombe*).

Interaction Between PDAT, DGAT1, DGAT2 And ARE2

In *S. cerevisiae*, four genes (i.e., ARE1, ARE2, DGA1 [encoding DGAT2] and LRO1 [encoding PDAT]), contribute to oil biosynthesis. PDAT and DGAT2 are responsible for up to approximately 95 % of oil biosynthesis (Sandager, L. et al., *J. Biol. Chem.* 277(8):6478-6482 (2002); Oelkers et. al. *J. Biol. Chem.* 277:8877 (2002)).

Surprisingly, according to the work described in co-pending U.S. Patent Application No. 10/882760 in *Yarrowia lipolytica*, PDAT and DGAT2 appeared to only be partially responsible for oil biosynthesis. Thus, it was apparent that at least one other DAG AT must play a role in TAG formation. As described in the Application herein, oil biosynthesis in the yeast *Yarrowia lipolytica* requires the activity of PDAT, DGAT1 and DGAT2, while ARE2 may additionally be a minor contributor to oil biosynthesis. This is in marked contrast to the enzymes responsible for oil biosynthesis in *S. cerevisiae*, where only DGAT2 and PDAT are the major DAG ATs.

Biosynthesis Of Omega-3 And Omega-6 Polyunsaturated Fatty Acids

The metabolic process that converts LA to GLA, DGLA and ARA (the ω -6 pathway) and ALA to STA, ETA, EPA, DPA and DHA (the ω -3

pathway) involves elongation of the carbon chain through the addition of two-carbon units and desaturation of the molecule through the addition of double bonds (Figure 2). This requires a series of desaturation and elongation enzymes. Specifically, oleic acid is converted to LA (18:2), the first of the ω -6 fatty acids, by the action of a Δ 12 desaturase. Subsequent ω -6 fatty acids are produced as follows: 1.) LA is converted to GLA by the action of a Δ 6 desaturase; 2.) GLA is converted to DGLA by the action of an elongase; and 3.) DGLA is converted to ARA by the action of a Δ 5 desaturase. In like manner, linoleic acid (LA) is converted to ALA, the first of the ω -3 fatty acids, by the action of a Δ 15 desaturase. Subsequent ω -3 fatty acids are produced in a series of steps similar to that for the ω -6 fatty acids. Specifically, 1.) ALA is converted to STA by the activity of a Δ 6 desaturase; 2.) STA is converted to ETA by the activity of an elongase; and 3.) ETA is converted to EPA by the activity of a Δ 5 desaturase. Alternatively, ETA and EPA can be produced from DGLA and ARA, respectively, by the activity of a Δ 17 desaturase. EPA can be further converted to DHA by the activity of an elongase and a Δ 4 desaturase.

In alternate embodiments, a Δ 9 elongase is able to catalyze the conversion of LA and ALA to EDA and ETrA, respectively. A Δ 8 desaturase then converts these products to DGLA and ETA, respectively.

Many microorganisms, including algae, bacteria, molds, fungi and yeasts, can synthesize PUFAs and omega fatty acids in the ordinary course of cellular metabolism. Particularly well-studied are fungi including *Schizochytrium aggregatum*, species of the genus *Thraustochytrium* and *Mortierella alpina*. Additionally, many dinoflagellates (Dinophyceae) naturally produce high concentrations of PUFAs. As such, a variety of desaturase and elongase genes involved in PUFA production have been identified through genetic means and the DNA sequences of some of these genes are publicly available (non-limiting examples are as follows):

AY131238, Y055118, AY055117, AF296076, AF007561, L11421, NM_031344, AF465283, AF465281, AF110510, AF465282, AF419296, AB052086, AJ250735, AF126799, AF126798 (Δ 6 desaturases); AF199596, AF226273, AF320509, AB072976, AF489588, AJ510244, AF419297, AF07879, AF067654, AB022097 (Δ 5 desaturases); AF489589.1, AY332747 (Δ 4 fatty acid desaturases); AAG36933, AF110509, AB020033, AAL13300, AF417244, AF161219, AY332747,

AAG36933, AF110509, AB020033, AAL13300, AF417244, AF161219, X86736, AF240777, AB007640, AB075526, AP002063 ($\Delta 12$ desaturases); NP_441622, BAA18302, BAA02924, AAL36934 ($\Delta 15$ desaturases); AF338466, AF438199, E11368, E11367, D83185, U90417, AF085500, 5 AY504633, NM_069854, AF230693 ($\Delta 9$ desaturases); and AX464731, NM_119617, NM_134255, NM_134383, NM_134382, NM_068396, NM_068392, NM_070713, NM_068746, NM_064685 (elongases).

Additionally, the patent literature provides many additional DNA sequences of genes (and/or details concerning several of the genes above and their methods of isolation) involved in PUFA production. See, for 10 example: U.S. 5,968,809 ($\Delta 6$ desaturases); U.S. 5,972,664 and U.S. 6,075,183 ($\Delta 5$ desaturases); WO 91/13972 and U.S. 5,057,419 ($\Delta 9$ desaturases); WO 93/11245 ($\Delta 15$ desaturases); WO 94/11516, U.S. 5,443,974 and WO 03/099216 ($\Delta 12$ desaturases); WO 00/12720 and U.S. 15 2002/0139974A1 (elongases); U.S. 2003/0196217 A1 ($\Delta 17$ desaturase); WO 00/34439 ($\Delta 8$ desaturases); and, WO 02/090493 ($\Delta 4$ desaturases). Each of these patents and applications are herein incorporated by reference in their entirety.

Depending upon the host cell, the availability of substrate, and the 20 desired end product(s), several desaturases and elongases are of interest for use in production of PUFAs. Considerations for choosing a specific polypeptide having desaturase or elongase activity include: 1.) the substrate specificity of the polypeptide; 2.) whether the polypeptide or a component thereof is a rate-limiting enzyme; 3.) whether the desaturase or 25 elongase is essential for synthesis of a desired PUFA; and/or 4.) co-factors required by the polypeptide. The expressed polypeptide preferably has parameters compatible with the biochemical environment of its location in the host cell. For example, the polypeptide may have to compete for substrate with other enzymes in the host cell. Analyses of the 30 K_M and specific activity of the polypeptide are therefore considered in determining the suitability of a given polypeptide for modifying PUFA production in a given host cell. The polypeptide used in a particular host cell is one that can function under the biochemical conditions present in the intended host cell but otherwise can be any polypeptide having 35 desaturase or elongase activity capable of modifying the desired fatty acid substrate.

Sequence Identification Of DGAT1 Acyltransferases

Despite the availability of several genes encoding DGAT1 (*supra*) that could be used for heterologous expression in oleaginous yeast (e.g., *Yarrowia lipolytica*), expression of a native enzyme is sometimes preferred
5 over a heterologous (or "foreign") enzyme. This preference may occur because: 1.) the native enzyme is optimized for interaction with other enzymes and proteins in the cell; and 2.) heterologous genes are unlikely to share the same codon preference in the host organism. Knowledge of the sequences of a host organism's native DGAT1 also facilitates
10 disruption of the homologous chromosomal genes by targeted disruption. And, as the present invention has shown, understanding of the complete complement of acyltransferase genes that enable TAG synthesis in an organism enables one to readily manipulate the oil content that the host organism produces in a variety of ways.

15 Comparison of the *Yarrowia lipolytica* DGAT1 nucleotide base (SEQ ID NO:13) and deduced amino acid (SEQ ID NO:14) sequences to public databases reveals that the most similar known sequences are about 55% identical to the amino acid sequence of DGAT1 reported herein over a length of 526 amino acids using the BLASTP method of alignment with a
20 low complexity filter and the following parameters: Expect value = 10, matrix = Blosum 62 (Altschul, et al., *Nucleic Acids Res.* 25:3389-3402(1997))).

Comparison of the *Mortierella alpina* DGAT1 nucleotide base (SEQ ID NO:17) and deduced amino acid (SEQ ID NO:18) sequences to public
25 databases reveals that the most similar known sequences are about 49% identical to the amino acid sequence of DGAT1 reported herein over a length of 525 amino acids using the BLASTP method of alignment (Altschul, et al., *supra*). Furthermore, comparison of the *Y. lipolytica* DGAT1 to the *M. alpina* DGAT1 revealed the two proteins to be 32.4%
30 identical.

Similarly, comparison of the *Neurospora crassa* DGAT1 amino acid (SEQ ID NO:19) sequence to the *Yarrowia lipolytica* DGAT1 amino acid (SEQ ID NO:14) reveals about 37% identity over a length of 533 amino acids; comparison of the *Gibberella zeae* PH-1 DGAT1 amino acid (SEQ
35 ID NO:20) sequence to the *Y. lipolytica* DGAT1 amino acid (SEQ ID NO:14) reveals about 38.1% identity over a length of 499 amino acids; comparison of the *Magnaporthe grisea* DGAT1 amino acid (SEQ ID

NO:21) sequence to the *Y. lipolytica* DGAT1 amino acid (SEQ ID NO:14) reveals about 36.2% identity over a length of 503 amino acids; and, comparison of the *Aspergillus nidulans* DGAT1 amino acid (SEQ ID NO:22) sequence to the *Y. lipolytica* DGAT1 amino acid (SEQ ID NO:14) reveals about 41.7% identity over a length of 458 amino acids.

More preferred DGAT1 amino acid fragments are at least about 70%-80% identical to the sequences herein (i.e., SEQ ID NOs:14, 18, 19, 20, 21, 22), where those sequences that are 85%-90% identical are particularly suitable and those sequences that are about 95% identical are most preferred. Similarly, preferred DGAT1 encoding nucleic acid sequences corresponding to the instant ORFs (i.e., SEQ ID NOs:13 and 17) are those encoding active proteins and which are at least about 70%-80% identical to the nucleic acid sequences encoding DGAT1 reported herein, where those sequences that are 85%-90% identical are particularly suitable and those sequences that are about 95% identical are most preferred.

Sequence Identification Of An ARE2 Acyltransferase

Comparison of the *Yarrowia lipolytica* ARE2 nucleotide base (SEQ ID NO:15) and deduced amino acid (SEQ ID NO:16) sequences to public databases reveals that the most similar known sequences are about 44% identical to the amino acid sequence of ARE2 reported herein over a length of 543 amino acids using the BLASTP method of alignment with a low complexity filter and the following parameters: Expect value = 10, matrix = Blosum 62 (Altschul, et al., *Nucleic Acids Res.* 25:3389-3402(1997)). More preferred ARE2 amino acid fragments are at least about 70%-80% identical to the sequences herein, where those sequences that are 85%-90% identical are particularly suitable and those sequences that are about 95% identical are most preferred. Similarly, preferred ARE2 encoding nucleic acid sequences corresponding to the instant ORFs are those encoding active proteins and which are at least about 70%-80% identical to the nucleic acid sequences encoding ARE2 reported herein, where those sequences that are 85%-90% identical are particularly suitable and those sequences that are about 95% identical are most preferred.

Isolation Of Homologs

Each of the acyltransferase nucleic acid fragments of the instant invention may be used to isolate genes encoding homologous proteins

from the same or other microbial species. Isolation of homologous genes using sequence-dependent protocols is well known in the art. Examples of sequence-dependent protocols include, but are not limited to:

- 1.) methods of nucleic acid hybridization; 2.) methods of DNA and RNA amplification, as exemplified by various uses of nucleic acid amplification technologies [e.g., polymerase chain reaction (PCR), Mullis et al., U.S. Patent 4,683,202; ligase chain reaction (LCR), Tabor, S. et al., *Proc. 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.

For example, genes encoding similar proteins or polypeptides to the DGAT1s and ARE2s described herein could be isolated directly by using all or a portion of the instant nucleic acid fragments as DNA hybridization probes to screen libraries from any desired yeast or fungus 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.

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. Humana: Totowa, NJ).

Generally two short segments of the instant sequences may be used in polymerase chain reaction protocols to amplify longer nucleic acid fragments encoding homologous genes from DNA or RNA. The polymerase chain reaction may also be performed on a library of cloned
5 nucleic acid fragments wherein the sequence of one primer is derived from the instant nucleic acid fragments, and the sequence of the other primer takes advantage of the presence of the polyadenylic acid tracts to the 3' end of the mRNA precursor encoding microbial genes.

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

Alternatively, the instant DGAT1 and ARE2 sequences may be
20 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 gene or gene fragment of interest, and a specific hybridization method. Probes of the present invention are typically single-stranded nucleic acid sequences that are complementary
25 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
30 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.

35 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).

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).

Nucleic acid hybridization is adaptable to a variety of assay formats. One of the most suitable is the sandwich assay format. The sandwich assay is particularly adaptable to hybridization under non-denaturing conditions. A primary component of a sandwich-type assay is a solid

support. The solid support has adsorbed to it or covalently coupled to it immobilized nucleic acid probe that is unlabeled and complementary to one portion of the sequence.

Availability of the instant nucleotide and deduced amino acid sequences facilitates immunological screening of DNA expression libraries. Synthetic peptides representing portions of the instant amino acid sequences may be synthesized. These peptides can be used to immunize animals to produce polyclonal or monoclonal antibodies with specificity for peptides or proteins comprising the amino acid sequences. These antibodies can then be used to screen DNA expression libraries to isolate full-length DNA clones of interest (Lerner, R. A. *Adv. Immunol.* 36:1 (1984); Maniatis, *supra*).

Gene Optimization For Improved Heterologous Expression

It may be desirable to modify the expression of the instant acyltransferases and/or PUFA biosynthetic pathway enzymes to achieve optimal conversion efficiency of each, according to the specific TAG composition that is desired within a specific host organism. As such, a variety of techniques can be utilized to improve and/or optimize the expression of a polypeptide of interest in an alternative host. Two such techniques include codon-optimization and mutagenesis of the gene.

Codon Optimization

As will be appreciated by one skilled in the art, it is frequently useful to modify a portion of the codons encoding a particular polypeptide that is to be expressed in a foreign host, such that the modified polypeptide uses codons that are preferred by the alternate host. Use of host-preferred codons can substantially enhance the expression of the foreign gene encoding the polypeptide.

In general, host-preferred codons can be determined within a particular host species of interest by examining codon usage in proteins (preferably those expressed in the largest amount) and determining which codons are used with highest frequency. Thus, the coding sequence for a polypeptide having acyltransferase activity can be synthesized in whole or in part using the codons preferred in the host species. All (or portions) of the DNA also can be synthesized to remove any destabilizing sequences or regions of secondary structure that would be present in the transcribed mRNA. All (or portions) of the DNA also can be synthesized to alter the base composition to one more preferable in the desired host cell.

Thus, for example, it may be desirable to modify a portion of the codons encoding the *Mortierella* polypeptide having DGAT1 activity, to enhance the expression of the gene in *Yarrowia lipolytica*, should the enzyme's substrate specificity be different than that of the *Y. lipolytica* DGAT1 disclosed herein. The codon usage profile and the consensus sequence around the 'ATG' translation initiation codon for this particular organism (i.e., *Y. lipolytica*) are taught in co-pending U.S. Patent Application No. 10/840478 (herein incorporated entirely by reference); likewise, a method for rapid synthesis of genes optimized for expression in *Y. lipolytica* is also provided.

Mutagenesis

Methods for synthesizing sequences and bringing sequences together are well established in the literature. For example, *in vitro* mutagenesis and selection, site-directed mutagenesis, error prone PCR (Melnikov et al., *Nucleic Acids Research*, 27(4):1056-1062 (February 15, 1999)), "gene shuffling" or other means can be employed to obtain mutations of naturally occurring acyltransferase genes. This would permit production of a polypeptide having acyltransferase activity *in vivo* with more desirable physical and kinetic parameters for function in the host cell (e.g., a longer half-life or a higher rate of synthesis of TAGs from fatty acids).

If desired, the regions of a DGAT1 or ARE2 polypeptide important for enzymatic activity can be determined through routine mutagenesis, expression of the resulting mutant polypeptides and determination of their activities. Mutants may include deletions, insertions and point mutations, or combinations thereof. A typical functional analysis begins with deletion mutagenesis to determine the N- and C-terminal limits of the protein necessary for function, and then internal deletions, insertions or point mutants are made to further determine regions necessary for function. Other techniques such as cassette mutagenesis or total synthesis also can be used. Deletion mutagenesis is accomplished, for example, by using exonucleases to sequentially remove the 5' or 3' coding regions. Kits are available for such techniques. After deletion, the coding region is completed by ligating oligonucleotides containing start or stop codons to the deleted coding region after the 5' or 3' deletion, respectively. Alternatively, oligonucleotides encoding start or stop codons are inserted into the coding region by a variety of methods including site-directed

mutagenesis, mutagenic PCR or by ligation onto DNA digested at existing restriction sites. Internal deletions can similarly be made through a variety of methods including the use of existing restriction sites in the DNA, by use of mutagenic primers via site-directed mutagenesis or mutagenic
5 PCR. Insertions are made through methods such as linker-scanning mutagenesis, site-directed mutagenesis or mutagenic PCR. Point mutations are made through techniques such as site-directed mutagenesis or mutagenic PCR.

Chemical mutagenesis also can be used for identifying regions of
10 an acyltransferase polypeptide important for activity. A mutated construct is expressed, and the ability of the resulting altered protein to function as an acyltransferase is assayed. Such structure-function analysis can determine which regions may be deleted, which regions tolerate insertions, and which point mutations allow the mutant protein to function
15 in substantially the same way as the native acyltransferase.

All such mutant proteins and nucleotide sequences encoding them that are derived from the DGAT1 and ARE2 genes described herein are within the scope of the present invention.

Microbial Production Of Fatty Acids And Triacylglycerols

20 Microbial production of fatty acids and TAGs has several advantages over purification from natural sources such as fish or plants. For example:

- 1.) Many microbes are known with greatly simplified oil compositions compared with those of higher organisms, making purification of
25 desired components easier;
- 2.) Microbial production is not subject to fluctuations caused by external variables, such as weather and food supply;
- 3.) Microbially produced oil is substantially free of contamination by environmental pollutants; and,
- 30 4.) Microbial oil production can be manipulated by controlling culture conditions, notably by providing particular substrates for microbially expressed enzymes, or by addition of compounds or genetic engineering approaches to suppress undesired biochemical pathways.

35 With respect to the production of ω -3 and/or ω -6 fatty acids in particular, and TAGs containing those PUFAs, additional advantages are incurred since microbes can provide fatty acids in particular forms that may have

specific uses; and, recombinant microbes provide the ability to alter the naturally occurring microbial fatty acid profile by providing new biosynthetic pathways in the host or by suppressing undesired pathways, thereby increasing levels of desired PUFAs, or conjugated forms thereof, and decreasing levels of undesired PUFAs.

Thus, knowledge of the sequences of the present acyltransferase genes will be useful for manipulating fatty acid biosynthesis and accumulation in oleaginous yeasts, and particularly, in *Yarrowia lipolytica*. This may require metabolic engineering directly within the fatty acid or TAG biosynthetic pathways or additional manipulation of pathways that contribute carbon to the fatty acid biosynthetic pathway. Methods useful for manipulating biochemical pathways are well known to those skilled in the art.

Metabolic Engineering To Up-Regulate Genes And Biosynthetic Pathways Affecting Fatty Acid Synthesis And Oil Accumulation In Oleaginous Yeast

It is expected that introduction of chimeric genes encoding the acyltransferases described herein, under the control of the appropriate promoters, will result in increased transfer of fatty acids to storage TAGs. As such, the present invention encompasses a method for increasing the TAG content in an oleaginous yeast comprising expressing at least one acyltransferase enzyme of the present invention in a transformed oleaginous yeast host cell producing a fatty acid, such that the fatty acid is transferred to the TAG pool.

Additional copies of acyltransferase genes (e.g., DGAT1, ARE2) may be introduced into the host to increase the transfer of fatty acids to the TAG fraction. Expression of the genes also can be increased at the transcriptional level through the use of a stronger promoter (either regulated or constitutive) to cause increased expression, by removing/deleting destabilizing sequences from either the mRNA or the encoded protein, or by adding stabilizing sequences to the mRNA (U.S. 4,910,141). Yet another approach to increase expression of heterologous genes is to increase the translational efficiency of the encoded mRNAs by replacement of codons in the native gene with those for optimal gene expression in the selected host microorganism.

In one specific embodiment, the present invention encompasses a method of increasing the ω -3 and/or ω -6 fatty acid content of TAGs in an

oleaginous yeast, since it is possible to introduce an expression cassette encoding each of the enzymes necessary for ω -3 and/or ω -6 fatty acid biosynthesis into the organism (since naturally produced PUFAs in these organisms are limited to 18:2 (i.e., LA), and less commonly 18:3 (i.e., ALA) fatty acids). Thus, the method comprises:

- a) providing a transformed oleaginous yeast host cell possessing genes encoding a functional ω -3/ ω -6 fatty acid biosynthetic pathway and at least one acyltransferase enzyme of the present invention;
- b) growing the yeast cells of step (a) in the presence of a fermentable carbon substrate, whereby the gene(s) of the ω -3/ ω -6 fatty acid biosynthetic pathway and the acyltransferase(s) are expressed, whereby a ω -3 and/or ω -6 fatty acid is produced, and whereby the ω -3 and/or ω -6 fatty acid is transferred to TAGs.

A variety of PUFA products can be produced (prior to their transfer to TAGs), depending on the fatty acid substrate and the particular genes of the ω -3/ ω -6 fatty acid biosynthetic pathway that are transformed into the host cell. As such, production of the desired fatty acid product can occur directly (wherein the fatty acid substrate is converted directly into the desired fatty acid product without any intermediate steps or pathway intermediates) or indirectly (wherein multiple genes encoding the PUFA biosynthetic pathway may be used in combination, such that a series of reactions occur to produce a desired PUFA). Specifically, for example, it may be desirable to transform an oleaginous yeast with an expression cassette comprising a Δ 12 desaturase, Δ 6 desaturase, an elongase, a Δ 5 desaturase and a Δ 17 desaturase for the overproduction of EPA. As is well known to one skilled in the art, various other combinations of the following enzymatic activities may be useful to express in a host in conjunction with the acyltransferases described herein: a Δ 15 desaturase, a Δ 4 desaturase, a Δ 5 desaturase, a Δ 6 desaturase, a Δ 17 desaturase, a Δ 9 desaturase, a Δ 8 desaturase and/or an elongase(s) (see Figure 2). The particular genes included within a particular expression cassette will depend on the host cell (and its PUFA profile and/or desaturase profile), the availability of substrate and the desired end product(s).

Thus, within the context of the present invention, it may be useful to modulate the expression of the TAG biosynthetic pathway by any one of

the methods described above. For example, the present invention provides genes encoding key enzymes in the fatty acid biosynthetic pathway leading to the storage of TAGs. These genes encode the DGAT1 and ARE2 enzymes. It will be particularly useful to modify the expression levels of these genes in oleaginous yeasts to maximize production and accumulation of TAGs using various means for metabolic engineering of the host organism. In preferred embodiments, modification of the expression levels of these genes in combination with expression of ω -3/ ω -6 biosynthetic genes can be utilized to maximize production and accumulation of preferred PUFAs in the TAG pool.

Metabolic Engineering To Down-Regulate Undesirable Genes And Biosynthetic Pathways Affecting Fatty Acid Synthesis And Oil Accumulation In Oleaginous Yeast

In some embodiments, it may be useful to disrupt or inactivate a host organism's native DGAT1 and/or ARE2, based on the complete sequences described herein, the complement of those complete sequences, substantial portions of those sequences, codon-optimized acyltransferases derived therefrom, and those sequences that are substantially homologous thereto.

For gene disruption, a foreign DNA fragment (typically a selectable marker gene) is inserted into the structural gene to be disrupted in order to interrupt its coding sequence and thereby functionally inactivate the gene. Transformation of the disruption cassette into the host cell results in replacement of the functional native gene by homologous recombination with the non-functional disrupted gene (see, for example: Hamilton et al., *J. Bacteriol.* 171:4617-4622 (1989); Balbas et al., *Gene* 136:211-213 (1993); Gueldener et al., *Nucleic Acids Res.* 24:2519-2524 (1996); and Smith et al., *Methods Mol. Cell. Biol.* 5:270-277(1996)).

Antisense technology is another method of down-regulating genes when the sequence of the target gene is known. To accomplish this, a nucleic acid segment from the desired gene is cloned and operably linked to a promoter such that the anti-sense strand of RNA will be transcribed. This construct is then introduced into the host cell and the antisense strand of RNA is produced. Antisense RNA inhibits gene expression by preventing the accumulation of mRNA that encodes the protein of interest. The person skilled in the art will know that special considerations are associated with the use of antisense technologies in order to reduce

expression of particular genes. For example, the proper level of expression of antisense genes may require the use of different chimeric genes utilizing different regulatory elements known to the skilled artisan.

Although targeted gene disruption and antisense technology offer effective means of down-regulating genes where the sequence is known, other less specific methodologies have been developed that are not sequence-based. For example, cells may be exposed to UV radiation and then screened for the desired phenotype. Mutagenesis with chemical agents is also effective for generating mutants and commonly used substances include chemicals that affect nonreplicating DNA (e.g., HNO₂ and NH₂OH), as well as agents that affect replicating DNA (e.g., acridine dyes, notable for causing frameshift mutations). Specific methods for creating mutants using radiation or chemical agents are well documented in the art. See, for example: Thomas D. Brock in Biotechnology: A Textbook of Industrial Microbiology, 2nd ed. (1989) Sinauer Associates: Sunderland, MA; or Deshpande, Mukund V., *Appl. Biochem. Biotechnol.*, 36:227 (1992).

Another non-specific method of gene disruption is the use of transposable elements or transposons. Transposons are genetic elements that insert randomly into DNA but can be later retrieved on the basis of sequence to determine where the insertion has occurred. Both *in vivo* and *in vitro* transposition methods are known. Both methods involve the use of a transposable element in combination with a transposase enzyme. When the transposable element or transposon is contacted with a nucleic acid fragment in the presence of the transposase, the transposable element will randomly insert into the nucleic acid fragment. The technique is useful for random mutagenesis and for gene isolation, since the disrupted gene may be identified on the basis of the sequence of the transposable element. Kits for *in vitro* transposition are commercially available [see, for example: 1.) The Primer Island Transposition Kit, available from Perkin Elmer Applied Biosystems, Branchburg, NJ, based upon the yeast Ty1 element; 2.) The Genome Priming System, available from New England Biolabs, Beverly, MA, based upon the bacterial transposon Tn7; and 3.) the EZ::TN Transposon Insertion Systems, available from Epicentre Technologies, Madison, WI, based upon the Tn5 bacterial transposable element].

Thus, within the context of the present invention, it may be useful to disrupt one of the acyltransferase genes of the invention. For example, it

may be necessary to disrupt genes and pathways that diminish the existing fatty acid pool and/or that hydrolyze TAGs to regulate (and/or maximize) TAG accumulation.

Expression Systems, Cassettes And Vectors

5 The genes and gene products of the instant sequences described herein may be produced in microbial host cells, particularly in the cells of oleaginous yeasts (e.g., *Yarrowia lipolytica*). Expression in recombinant microbial hosts may be useful for the transfer of various fatty acids to TAGs.

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

 Vectors or DNA cassettes useful for the transformation of suitable host cells are well known in the art. The specific choice of sequences present in the construct is dependent upon the desired expression
20 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.
25 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
30 native to the specific species chosen as a production host.

 Initiation control regions or promoters which are useful to drive expression of the instant ORFs in the desired host cell are numerous and familiar to those skilled in the art. Virtually any promoter capable of directing expression of these genes in the selected host cell is suitable for
35 the present invention. Expression in a host cell can be accomplished in a transient or stable fashion. Transient expression can be accomplished by inducing the activity of a regulatable promoter operably linked to the gene

of interest. Stable expression can be achieved by the use of a constitutive promoter operably linked to the gene of interest. As an example, when the host cell is yeast, transcriptional and translational regions functional in yeast cells are provided, particularly from the host species. The
5 transcriptional initiation regulatory regions can be obtained, for example, from: 1.) genes in the glycolytic pathway, such as alcohol dehydrogenase, glyceraldehyde-3-phosphate-dehydrogenase (see U.S. Patent Application Number 10/869630, incorporated herein by reference), phosphoglycerate mutase (see U.S. Patent Application Number 10/869630), fructose-
10 bisphosphate aldolase (see U.S. Patent Application Number 60/519971, incorporated herein by reference), phosphoglucose-isomerase, phosphoglycerate kinase, glycerol-3-phosphate O-acyltransferase (see U.S. Patent Application No. 60/610060), etc.; or, 2.) regulatable genes such as acid phosphatase, lactase, metallothionein, glucoamylase, the
15 translation elongation factor EF1- α (TEF) protein (U.S. 6,265,185), ribosomal protein S7 (U.S. 6,265,185), etc. Any one of a number of regulatory sequences can be used, depending upon whether constitutive or induced transcription is desired, the efficiency of the promoter in expressing the ORF of interest, the ease of construction and the like.

20 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 to obtain optimal gene expression. For expression in
25 yeast, this can be done by site-directed mutagenesis of an inefficiently expressed gene by fusing it in-frame to an endogenous yeast gene, preferably a highly expressed gene. Alternatively, one can determine the consensus translation initiation sequence in the host and engineer this sequence into heterologous genes for their optimal expression in the host
30 of interest (see, e.g., U.S. Patent Application No. 10/840478 for specific teachings applicable for *Yarrowia lipolytica*).

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
35 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.

As one of skill in the art is aware, merely inserting a gene into a cloning vector does not ensure that it will be successfully expressed at the level needed. In response to the need for a high expression rate, many specialized expression vectors have been created by manipulating a number of different genetic elements that control aspects of transcription, translation, protein stability, oxygen limitation and secretion from the host cell. More specifically, some of the molecular features that have been manipulated to control gene expression include: 1.) the nature of the relevant transcriptional promoter and terminator sequences; 2.) the number of copies of the cloned gene and whether the gene is plasmid-borne or integrated into the genome of the host cell; 3.) the final cellular location of the synthesized foreign protein; 4.) the efficiency of translation in the host organism; 5.) the intrinsic stability of the cloned gene protein within the host cell; and 6.) the codon usage within the cloned gene, such that its frequency approaches the frequency of preferred codon usage of the host cell. Each of these types of modifications are encompassed in the present invention, as means to further optimize expression of the acyltransferase enzymes.

Preferred Microbial Hosts For Recombinant Expression Of Acyltransferases

Host cells for expression of the instant DGAT1 and ARE2 genes and nucleic acid fragments may include microbial hosts that grow on a variety of feedstocks, including simple or complex carbohydrates, organic acids and alcohols and/or hydrocarbons over a wide range of temperature and pH values. Although the genes described in the instant invention have been isolated for expression in an oleaginous yeast, and in particular *Yarrowia lipolytica*, it is contemplated that because transcription, translation and the protein biosynthetic apparatus is highly conserved, any

bacteria, yeast, algae and/or filamentous fungus will be a suitable host for expression of the present nucleic acid fragments.

Preferred microbial hosts are oleaginous organisms, such as oleaginous yeasts. These oleaginous organisms are naturally capable of oil synthesis and accumulation, wherein the total oil content can comprise greater than about 25% of the cellular dry weight, more preferably greater than about 30% of the cellular dry weight and most preferably greater than about 40% of the cellular dry weight. Additionally, there is basis for the use of these organisms for the production of PUFAs, as seen in co-pending U.S. Patent Application No. 10/840579, herein incorporated entirely by reference.

Genera typically identified as oleaginous yeast include, but are not limited to: *Yarrowia*, *Candida*, *Rhodotorula*, *Rhodospiridium*, *Cryptococcus*, *Trichosporon* and *Lipomyces*. More specifically, illustrative oil-synthesizing yeasts include: *Rhodospiridium 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*).

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, ATCC #90812 and/or LGAM S(7)1 (Papanikolaou S., and Aggelis G., *Bioresour. Technol.* 82(1):43-9 (2002)).

Transformation Of Microbial Hosts

Once the DNA encoding a polypeptide suitable for expression in an oleaginous yeast has been obtained, 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 within the host locus. Where constructs are targeted to an endogenous locus, all or some of the transcriptional and translational regulatory regions can be provided by the endogenous locus.

Where two or more genes are expressed from separate replicating vectors, it is desirable that each vector has a different means of selection and should lack homology to the other construct(s) to maintain stable expression and prevent reassortment of elements among constructs.

Judicious choice of regulatory regions, selection means and method of propagation of the introduced construct(s) can be experimentally determined so that all introduced genes are expressed at the necessary levels to provide for synthesis of the desired products.

5 Constructs comprising the gene of interest may be introduced into a host cell by any standard technique. These techniques include transformation (e.g., lithium acetate transformation [*Methods in Enzymology*, 194:186-187 (1991)]), protoplast fusion, biolistic impact, electroporation, microinjection, or any other method that introduces the
10 gene of interest into the host cell. More specific teachings applicable for oleaginous yeasts (i.e., *Yarrowia lipolytica*) include U.S. Patent Nos. 4,880,741 and 5,071,764 and Chen, D. C. et al. (*Appl Microbiol Biotechnol.* 48(2):232-235-(1997)).

15 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 extra-chromosomal element having
20 multiple copy numbers. The transformed host cell can be identified by selection for a marker contained on the introduced construct. Alternatively, a separate marker construct may be co-transformed with the desired construct, as many transformation techniques introduce many DNA molecules into host cells. Typically, transformed hosts are selected
25 for their ability to grow on selective media. Selective media may incorporate an antibiotic or lack a factor necessary for growth of the untransformed host, such as a nutrient or growth factor. An introduced marker gene may confer antibiotic resistance, or encode an essential growth factor or enzyme, thereby permitting growth on selective media
30 when expressed in the transformed host. Selection of a transformed host can also occur when the expressed marker protein can be detected, either directly or indirectly. The marker protein may be expressed alone or as a fusion to another protein. The marker protein can be detected by: 1.) its enzymatic activity (e.g., β -galactosidase can convert the substrate X-gal [5-bromo-4-chloro-3-indolyl- β -D-galactopyranoside] to a colored product;
35 luciferase can convert luciferin to a light-emitting product); or 2.) its light-producing or modifying characteristics (e.g., the green fluorescent protein

of *Aequorea victoria* fluoresces when illuminated with blue light).

Alternatively, antibodies can be used to detect the marker protein or a molecular tag on, for example, a protein of interest. Cells expressing the marker protein or tag can be selected, for example, visually, or by techniques such as FACS or panning using antibodies. For selection of yeast transformants, any marker that functions in yeast may be used. Desirably, resistance to kanamycin, hygromycin and the amino glycoside G418 are of interest, as well as ability to grow on media lacking uracil or leucine.

Following transformation, substrates suitable for the gene products of the instant sequences (and optionally other PUFA enzymes that are expressed within the host cell), may be produced by the host either naturally or transgenically, or they may be provided exogenously.

Fermentation Processes For Triacylglycerol Biosynthesis And Accumulation

The transformed microbial host cell is grown under conditions that optimize activity of fatty acid biosynthetic genes and acyltransferase genes. This leads to production of the greatest and the most economical yield of fatty acids, which can in turn be transferred to TAGs for storage. In general, media conditions that may be optimized include the type and amount of carbon source, the type and amount of nitrogen source, the carbon-to-nitrogen ratio, the oxygen level, growth temperature, pH, length of the biomass production phase, length of the oil accumulation 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, MI)).

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 substrate may include one-carbon substrates (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 substrates and will only be limited by the choice of the host organism. Although all of the above mentioned carbon substrates and mixtures thereof are expected to be suitable in the present invention, preferred carbon substrates are sugars and/or fatty acids. Most preferred is glucose and/or fatty acids containing between 10-22 carbons.

Nitrogen may be supplied from an inorganic (e.g., $(\text{NH}_4)_2\text{SO}_4$) or organic source (e.g., urea, glutamate). 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 and promotion of the enzymatic pathways necessary for fatty acid production. 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 (Nakahara, T. et al., *Ind. Appl. Single Cell Oils*, D. J. Kyle and R. Colin, eds. pp 61-97 (1992)).

Preferred growth media in the present invention are common commercially prepared media, such as Yeast Nitrogen Base (DIFCO Laboratories, Detroit, MI). 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.

Typically, accumulation of high levels of fatty acids and TAGs in oleaginous yeast cells requires a two-stage process, since the metabolic state must be "balanced" between growth and synthesis/storage of fats. Thus, most preferably, a two-stage fermentation process is necessary for the production of oils 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. The effect of this nitrogen deprivation is to reduce 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. Thus, this phase is characterized by the cessation of cell division followed by the synthesis of fatty acids and accumulation of TAGs.

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.

It is contemplated that a variety of fermentation process designs may be applied, where commercial production of fatty acids and TAGs using the instant acyltransferase genes is desired. For example, commercial production of TAGs containing PUFAs from a recombinant microbial host may be produced by a batch, fed-batch or continuous fermentation process.

A batch fermentation process is a closed system wherein the media composition is set at the beginning of the process and not subject to further additions beyond those required for maintenance of pH and oxygen level during the process. Thus, at the beginning of the culturing process the media is inoculated with the desired organism and growth or metabolic activity is permitted to occur without adding additional substrates (i.e., carbon and nitrogen sources) to the medium. In batch processes the metabolite and biomass compositions of the system change constantly up to the time the culture is terminated. In a typical batch process, cells moderate through a static lag phase to a high-growth log phase and finally to a stationary phase, wherein the growth rate is diminished or halted. Left untreated, cells in the stationary phase will eventually die. A variation of the standard batch process is the fed-batch process, wherein the substrate is continually added to the fermentor over the course of the fermentation process. A fed-batch process is also suitable in the present

invention. Fed-batch processes are useful when catabolite repression is apt to inhibit the metabolism of the cells or where it is desirable to have limited amounts of substrate in the media at any one time. Measurement of the substrate concentration in fed-batch systems is difficult and therefore may be estimated on the basis of the changes of measurable factors such as pH, dissolved oxygen and the partial pressure of waste gases (e.g., CO₂). Batch and fed-batch culturing methods are common and well known in the art and examples may be found in Thomas D. Brock in Biotechnology: A Textbook of Industrial Microbiology, 2nd ed., (1989) Sinauer Associates: Sunderland, MA; or Deshpande, Mukund V., *Appl. Biochem. Biotechnol.*, 36:227 (1992), herein incorporated by reference.

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

Purification Of Fatty Acids

Fatty acids, including PUFAs, may be found in the host microorganism 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)).

In general, means for the purification of fatty acids (including PUFAs) may include extraction with organic solvents, sonication, 5 supercritical fluid extraction (e.g., using carbon dioxide), saponification and physical means such as presses, or combinations thereof. Of particular interest is extraction with methanol and chloroform in the presence of water (E. G. Bligh & W. J. Dyer, *Can. J. Biochem. Physiol.* 37:911-917 (1959)). Where desirable, the aqueous layer can be acidified to protonate 10 negatively-charged moieties and thereby increase partitioning of desired products into the organic layer. After extraction, the organic solvents can be removed by evaporation under a stream of nitrogen. When isolated in conjugated forms, the products may be enzymatically or chemically cleaved to release the free fatty acid or a less complex conjugate of 15 interest, and can then be subject to further manipulations to produce a desired end product. Desirably, conjugated forms of fatty acids are cleaved with potassium hydroxide.

If further purification is necessary, standard methods can be employed. Such methods may include extraction, treatment with urea, 20 fractional crystallization, HPLC, fractional distillation, silica gel chromatography, high-speed centrifugation or distillation, or combinations of these techniques. Protection of reactive groups, such as the acid or alkenyl groups, may be done at any step through known techniques (e.g., alkylation, iodination). Methods used include methylation of the fatty acids 25 to produce methyl esters. Similarly, protecting groups may be removed at any step. Desirably, purification of fractions containing GLA, STA, ARA, DHA and EPA may be accomplished by treatment with urea and/or fractional distillation.

DESCRIPTION OF PREFERRED EMBODIMENTS

30 The ultimate goal of the work described herein is the development of an oleaginous yeast that accumulates TAGs enriched in ω -3 and/or ω -6 PUFAs. Toward this end, acyltransferases must be identified that function efficiently in oleaginous yeasts, to enable synthesis and high accumulation of preferred TAGs in storage lipid pools. Specifically, modification of the 35 expression levels of these acyltransferases will enable increased transfer of fatty acids (and particularly, PUFAs) to TAGs. Thus, identification of efficient acyltransferases is necessary for the manipulation of the amount

of ω -3/ ω -6 PUFAs incorporated into the TAG fraction produced in host cells.

In the present invention, Applicants have isolated and cloned genes from *Yarrowia lipolytica* that encode DGAT1 and ARE2. This work was undertaken, following the surprising discovery that the *Y. lipolytica* DGAT2 and PDAT are only partially responsible for oil biosynthesis (based on analysis of a double knockout mutant; see co-pending U.S. Patent Application No. 10/882760). Confirmation of the DGAT1 gene's activity was provided herein based upon lower oil content (total fatty acids as a % of dry cell weight) in *Yarrowia* strains wherein disruption of the native DGAT1 had occurred by targeted gene replacement through homologous recombination (Example 9). Additionally, over-expression of the DGAT1 and ARE2 genes of the invention are expected to increase oil content (total fatty acids as a % of dry cell weight), based on results obtained wherein the *Yarrowia* DGAT1 was overexpressed (Example 12).

Following identification of the *Y. lipolytica* DGAT1 (SEQ ID NOs:13 and 14), a proprietary database of *Mortierella alpina* cDNA sequences was searched to identify an orthologous gene(s). This lead to the identification of a DGAT1 homolog, whose full nucleotide sequence was thus determined (SEQ ID NO:17). Based on these two novel sequences, the Applicants were then able to identify a suite of unique fungal DGAT1 orthologs that are involved in the synthesis of TAGs. More specifically, these DGAT1s have been identified from *Neurospora crassa* (SEQ ID NO:19), *Gibberella zeae PH-1* (SEQ ID NO:20), *Magnaporthe grisea* (SEQ ID NO:21) and *Aspergillus nidulans* (SEQ ID NO:22), based on comparison of the *Yarrowia* and *Mortierella* DNA sequences to the GenBank database using the BLAST algorithm, well known to those skilled in the art.

Analysis of these six DGAT1 sequences, when aligned with DGAT1 sequences from mouse (SEQ ID NO:169), soy (SEQ ID NO:170), *Arabidopsis* (SEQ ID NO:171), rice (SEQ ID NO:172), *Perilla* (SEQ ID NO:173) and wheat (SEQ ID NO:174), however, reveals unique characteristics that are specific for fungi and absent in DGAT1s from the non-fungal organisms described above (i.e., fungal motifs #1-8, herein described as SEQ ID NOs:23-30). Furthermore, in a broader context, common motifs universally found in DGAT1 enzymes were also discovered (SEQ ID NOs:31-37). These unique fungal and universal

conserved domains are between amino acids 97-105, 278-284, 334-341, 364-374, 418-424, 415-424, 456-466 and 513-519, with respect to SEQ ID NO:14. As is well known to one of skill in the art, these motifs will therefore be diagnostic for DGAT1s and will permit rapid identification of novel DGAT1s. The motifs described herein are to be distinguished from the 'FxxPxYR' motif (SEQ ID NO:38) that was recently described by Lardizabal et al. (US 04/0107459 A1), which is preferably useful to identify DGAT2s of fungal origin.

The Applicants conclude that these DGAT1 and ARE2 acyltransferase genes are useful for expression in various microbial hosts, and particularly for over-expression in oleaginous yeasts (e.g., *Yarrowia lipolytica*). Additional benefits may result, since expression of the acyltransferases can also be put under the control of strong constitutive or regulated promoters that do not have the regulatory constraints of the native gene.

EXAMPLES

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.

25 GENERAL METHODS

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

35 Materials and methods suitable for the maintenance and growth of microbial cultures are well known in the art. Techniques suitable for use in the following Examples may be found as set out in Manual of Methods for

General Bacteriology (Phillipp Gerhardt, R. G. E. Murray, Ralph N. Costilow, Eugene W. Nester, Willis A. Wood, Noel R. Krieg and G. Briggs Phillips, Eds), American Society for Microbiology: Washington, D.C. (1994); or by Thomas D. Brock in Biotechnology: A Textbook of Industrial Microbiology, 2nd ed., Sinauer Associates: Sunderland, MA (1989). All reagents, restriction enzymes and materials used for the growth and maintenance of microbial cells were obtained from Aldrich Chemicals (Milwaukee, WI), DIFCO Laboratories (Detroit, MI), GIBCO/BRL (Gaithersburg, MD) or Sigma Chemical Company (St. Louis, MO), unless otherwise specified.

E. coli TOP10 cells and *E. coli* Electromax DH10B cells were obtained from Invitrogen (Carlsbad, CA). Max Efficiency competent cells of *E. coli* DH5 α were obtained from GIBCO/BRL (Gaithersburg, MD). *E. coli* (XL1-Blue) competent cells were purchased from the Stratagene Company (San Diego, CA). *E. coli* strains were typically grown at 37 °C on Luria Bertani (LB) plates. General molecular cloning was performed according to standard methods (Sambrook et al., *supra*). Oligonucleotides were synthesized by Sigma-Genosys (Spring, TX).

All polymerase chain reactions (PCRs) were performed in a thermocycler using DNA polymerase in a buffer recommended by the manufacturer of the polymerase. Unless specified otherwise, amplification was carried out as follows: initial denaturation at 95 °C for 1 min, followed by 30 cycles of denaturation at 95 °C for 30 sec, annealing at 55 °C for 1 min, and elongation at 72 °C for 1 min. A final elongation cycle at 72 °C for 10 min was carried out, followed by reaction termination at 4 °C. PCR products were cloned into Promega's pGEM-T-easy vector (Madison, WI), unless otherwise noted.

DNA sequence was generated on an ABI Automatic sequencer using dye terminator technology (U.S. 5,366,860; EP 272,007) using a combination of vector and insert-specific primers. Sequence editing was performed in Sequencher (Gene Codes Corporation, Ann Arbor, MI). All sequences represent coverage at least two times in both directions. Comparisons of genetic sequences were accomplished using DNASTAR software (DNASTAR, Inc., Madison, WI). The percent identities between these proteins were determined by the Megalign program of DNASTAR using Clustal W with the following parameters: gap penalty =10, gap

length penalty = 0.2, delay divergent seqs (%) = 30, DNA transition weight = 0.5 and protein weight matrix by Gonnet series.

The meaning of abbreviations is as follows: "sec" means second(s), "min" means minute(s), "h" means hour(s), "d" means day(s),
5 "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).

10 Transformation And Cultivation Of *Yarrowia lipolytica*

Yarrowia lipolytica strains ATCC #20362, #76982 and #90812 were purchased from the American Type Culture Collection (Rockville, MD). *Y. lipolytica* strains were usually grown at 28 °C on YPD agar (1% yeast extract, 2% bactopectone, 2% glucose, 2% agar).

15 Transformation of *Yarrowia 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
20 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
25 cells were plated onto selection media plates and maintained at 30 °C for 2 to 3 days.

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

Alternatively, transformants were selected on 5-fluoroorotic acid
35 ("FOA"; also 5-fluorouracil-6-carboxylic acid monohydrate) selection media, comprising: 0.17% yeast nitrogen base (DIFCO Laboratories, Detroit, MI) without ammonium sulfate or amino acids, 2% glucose, 0.1%

proline, 75 mg/L uracil, 75 mg/L uridine, 900 mg/L FOA (Zymo Research Corp., Orange, CA) and 20 g/L agar.

To promote oleaginous conditions, High Glucose Media ("HGM") was prepared as follows: 14 g/L KH_2PO_4 , 4 g/L LK_2HPO_4 , 2 g/L $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$,
 5 80 g/L glucose (pH 6.5).

Fatty Acid Analysis Of *Yarrowia lipolytica*

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

15 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
 20 vortexed and spun. The upper layer was removed and analyzed by GC as described above.

EXAMPLE 1

Construction Of Plasmids Suitable For Gene Expression In *Yarrowia lipolytica*

25 The present Example describes the construction of plasmids pY5, pY5-13, pY20 and pLV5.

Construction Of Plasmid pY5

The plasmid pY5, a derivative of pINA532 (a gift from Dr. Claude Gaillardin, Institut National Agronomique, Centre de biotechnologie Agro-
 30 Industrielle, laboratoire de Genetique Moleculaire et Cellulaire INRA-CNRS, F-78850 Thiverval-Grignon, France), was constructed for expression of heterologous genes in *Yarrowia lipolytica*, as diagrammed in Figure 3.

First, the partially-digested 3598 bp *EcoRI* fragment containing the ARS18
 35 sequence and LEU2 gene of pINA532 was subcloned into the *EcoRI* site of pBluescript (Stratagene, San Diego, CA) to generate pY2. The TEF promoter (Muller S. et al., *Yeast*, 14:1267-1283 (1998)) was amplified from

Y. lipolytica genomic DNA by PCR using TEF5' (SEQ ID NO:39) and TEF3' (SEQ ID NO:40) as primers. PCR amplification was carried out in a 50 µl total volume containing: 100 ng *Yarrowia* genomic DNA, PCR buffer containing 10 mM KCl, 10 mM (NH₄)₂SO₄, 20 mM Tris-HCl (pH 8.75), 2 mM MgSO₄, 0.1% Triton X-100, 100 µg/mL BSA (final concentration), 200 µM each deoxyribonucleotide triphosphate, 10 pmole of each primer and 1 µl of PfuTurbo DNA polymerase (Stratagene). Amplification was carried out as follows: initial denaturation at 95 °C for 3 min, followed by 35 cycles of the following: 95 °C for 1 min, 56 °C for 30 sec, 72 °C for 1 min. A final extension cycle of 72 °C for 10 min was carried out, followed by reaction termination at 4 °C. The 418 bp PCR product was ligated into pCR-Blunt to generate pIP-tef. The *Bam*HI/*Eco*RV fragment of pIP-tef was subcloned into the *Bam*HI/*Sma*I sites of pY2 to generate pY4.

The XPR2 transcriptional terminator was amplified by PCR using pINA532 as template and XPR5' (SEQ ID NO:41) and XPR3' (SEQ ID NO:42) as primers. The PCR amplification was carried out in a 50 µl total volume, using the components and conditions described above. The 179 bp PCR product was digested with *Sac*II and then ligated into the *Sac*II site of pY4 to generate pY5. Thus, pY5 (shown in Figure 3) is useful as a *Yarrowia*-*E. coli* shuttle plasmid containing: 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 (E.C. 1.1.1.85, encoding isopropylmalate isomerase) for selection in *Yarrowia*; the translation elongation promoter (TEF) for expression of heterologous genes in *Yarrowia*; and the extracellular protease gene terminator (XPR2) for transcriptional termination of heterologous gene expression in *Yarrowia*.

Construction Of Plasmid pY5-13

pY5-13 was constructed as a derivative of pY5 to facilitate subcloning and heterologous gene expression in *Yarrowia lipolytica*. Specifically, pY5-13 was constructed by 6 rounds of site-directed mutagenesis using pY5 as template. Both *Sa*II and *Cla*I sites were eliminated from pY5 by site-directed mutagenesis using oligonucleotides YL5 and YL6 (SEQ ID NOs:43 and 44) to generate pY5-5. A *Sa*II site was introduced into pY5-5 between the LEU2 gene and the TEF promoter by site-directed mutagenesis using oligonucleotides YL9 and YL10 (SEQ ID NOs:45 and 46) to generate pY5-6. A *Pac*I site was introduced into pY5-6

between the LEU2 gene and ARS18 using oligonucleotides YL7 and YL8 (SEQ ID NOs:47 and 48) to generate pY5-8. A *NcoI* site was introduced into pY5-8 around the translation start codon of the TEF promoter using oligonucleotides YL3 and YL4 (SEQ ID NOs:49 and 50) to generate pY5-9. The *NcoI* site inside the LEU2 gene of pY5-9 was eliminated using YL1 and YL2 oligonucleotides (SEQ ID NOs:51 and 52) to generate pY5-12. Finally, a *BsiWI* site was introduced into pY5-12 between the ColEI and XPR2 region using oligonucleotides YL61 and YL62 (SEQ ID NOs:53 and 54) to generate pY5-13.

10 Construction Of Plasmids pY20 and pLV5

Plasmid pY20 (SEQ ID NO:55) is a derivative of pY5. It was constructed by inserting a *Not I* fragment containing a chimeric hygromycin resistance gene into the *Not I* site of pY5. Specifically, the *E. coli* hygromycin resistance gene (SEQ ID NO:56; "HPT"; Kaster, K.R. et al., *Nucleic Acids Res.* 11:6895-6911 (1983)) was PCR amplified for expression. The chimeric gene had the hygromycin resistance ORF under the control of the *Y. lipolytica* TEF promoter.

Plasmid pLV5 is a derivative of pY20. It was constructed by replacing the hygromycin resistant gene with the *Yarrowia* Ura3 gene. A 1.7 kB DNA fragment (SEQ ID NO:58) containing the *Yarrowia* Ura3 gene was PCR amplified using oligonucleotides KU5 and KU3 (SEQ ID NOs:60 and 61) as primers and *Yarrowia* genomic DNA as template.

EXAMPLE 2

25 Cloning Of A Partial *Yarrowia lipolytica* Acyl-CoA:Diacylglycerol Acyltransferase (DGAT2) Gene And Disruption Of The Endogenous DGAT2 Gene

The present Example describes the use of degenerate PCR primers to isolate a partial coding sequence of the *Yarrowia lipolytica* DGAT2 and the use of the partial sequence to disrupt the native gene in *Y. lipolytica*.

Cloning Of A Partial Putative DGAT2 Sequence From *Yarrowia lipolytica* By PCR Using Degenerate PCR Primers And Chromosome Walking

Genomic DNA was isolated from *Y. lipolytica* (ATCC #76982) using a DNeasy Tissue Kit (Qiagen, Catalog # 69504) and resuspended in kit buffer AE at a DNA concentration of 0.5 µg/µl. PCR amplifications were performed using the genomic DNA as template and several sets of degenerate primers designed to encode conserved amino acid sequences

among different known DGAT2s (i.e., GenBank Accession Nos. NC_001147 [*Saccharomyces cerevisiae*] and AF391089 and AF391090 [*Mortierella ramanniana*]). The best results were obtained with degenerate primers P7 and P8, as shown in the Table below.

5

Table 4
Degenerate Primers Used For Amplification Of A Partial Putative DGAT2

Primer Set	Description	Degenerate Nucleotide Sequence	Corresponding Amino Acid Sequence
P7	(32) 29-mers	5'-AACTACATCTTCGGCTA YCA YCCNCA YGG-3' (SEQ ID NO:62)	NYIFGYHPHG (SEQ ID NO:63)
P8	(48) 29-mers	5'-AGGGACTCGGAGGCGC CGCCNCANACDAT-3' (SEQ ID NO:64)	complementary to IVVGGASESL (SEQ ID NO:65)

10 [Note: Abbreviations are standard for nucleotides and proteins. The nucleic acid degeneracy code used is as follows: Y=C/T; D=A/G/T; and N=A/C/G/T.]

The PCR was carried out in a RoboCycler Gradient 40 PCR machine (Stratagene) using the manufacturer's recommendations and Accuprime Taq polymerase (Invitrogen). Amplification was carried out as described in the General Methods.

15 The expected PCR product (ca. 264 bp) was detected by 4% NuSieve (FMC) agarose gel electrophoresis, isolated, purified, cloned into the TOPO[®] cloning vector (Invitrogen) and sequenced. The resultant sequence (contained within SEQ ID NO:1) had homology to known
20 DGAT2s, based on BLAST program analysis (Basic Local Alignment Search Tool; Altschul, S. F. et al., *J. Mol. Biol.* 215:403-410 (1993)).

Using the 264 bp fragment as an initiation point, a 673 bp fragment was obtained by chromosome walking using the TOPO[®] Walker Kit (Invitrogen, Catalog #K8000-01). The chromosome walking was carried
25 out in 6 steps, as described briefly below:

- 1.) Genomic DNA (5 µg) was digested with restriction enzymes *Pst* I or *Sac* I, leaving a 3' overhang;

- 2.) Digested DNA was treated with 0.1 U calf intestinal alkaline phosphatase to dephosphorylate DNA;
- 3.) Primer extension was performed, using the DGAT2 specific primer P80 (SEQ ID NO:66) and *Taq* polymerase;
- 5 4.) TOPO[®] Linker (1 μ l) was added and the reaction was incubated at 37 °C for 5 min to ligate TOPO[®] Linker to the DNA;
- 5.) PCR was performed using the DGAT2 gene specific primer, P81 (SEQ ID NO:67) and LinkAmp primer 1 (SEQ ID NO:68); and
- 10 6.) The newly amplified fragment was sequenced with primer P81 and LinkAmp primer 1.

The sequence of the 673 bp fragment obtained by chromosome walking also showed homology to known DGAT2 sequences.

Targeted Disruption Of The *Yarrowia lipolytica* DGAT2 Gene

Targeted disruption of the DGAT2 gene in *Y. lipolytica* ATCC #90812 and
 15 ATCC #76982 was carried out by homologous recombination-mediated replacement of the endogenous DGAT2 gene with a targeting cassette designated as plasmid pY21DGAT2. pY21DGAT2 was derived from plasmid pY20 (Example 1; SEQ ID NO:55). Specifically, pY21DGAT2 was created by inserting a 570 bp *Hind III/Eco RI* fragment into similarly linearized pY20. The
 20 570 bp DNA fragment contained (in 5' to 3' orientation): 3' homologous sequence from position +1090 to +1464 (of the coding sequence (ORF) in SEQ ID NO:1), a *Bgl II* restriction site and 5' homologous sequence from position +906 to +1089 (of the coding sequence (ORF) shown in SEQ ID NO:1). The fragment was prepared by PCR amplification of 3' and 5' sequences from the 673 bp DGAT2
 25 PCR product obtained by chromosome walking using two pairs of PCR primers, P95 and P96 (SEQ ID NOs:69 and 70), and P97 and P98 (SEQ ID NOs:71 and 72), respectively.

pY21DGAT2 was linearized by *Bgl II* restriction digestion and transformed into mid-log phase *Y. lipolytica* ATCC #90812 and ATCC #76982 cells, as
 30 described in the General Methods. The cells were plated onto YPD hygromycin selection plates and maintained at 30 °C for 2 to 3 days.

Four *Y. lipolytica* ATCC #76982 hygromycin-resistant colonies and fourteen *Y. lipolytica* ATCC #90812 hygromycin-resistant colonies were isolated and screened for targeted disruption by PCR. One set of PCR
 35 primers (P115 and P116 [SEQ ID NOs:73 and 74, respectively]) was designed to amplify a specific junction fragment following homologous

recombination. Another pair of PCR primers (P115 and P112 [SEQ ID NO:75]) was designed to detect the native gene.

All (4 of 4) of the hygromycin-resistant colonies of ATCC #76982 strains were positive for the junction fragment and negative for the native fragment; and, 2 of the 14 hygromycin-resistant colonies of ATCC #90812 strains were positive for the junction fragment and negative for the native fragment. Thus, targeted integration was confirmed in these 6 strains. Disruption of the gene was further confirmed by GC analysis of total lipids of one of the disrupted strains, designated as "S-D2" (see Example 9).

EXAMPLE 3

Cloning Of A Partial *Yarrowia lipolytica* Phospholipid:Diacylglycerol Acyltransferase (PDAT) Gene And Disruption Of The Endogenous PDAT Gene

The present Example describes the use of degenerate PCR primers to isolate a partial coding sequence of *Y. lipolytica* PDAT and the use of the partial sequence to disrupt the native gene in *Y. lipolytica*.

Cloning Of A Partial Putative PDAT Sequence From *Yarrowia lipolytica* By PCR Using Degenerate PCR Primers And Chromosome Walking

Genomic DNA was isolated from *Y. lipolytica* (ATCC #76982) using a DNeasy Tissue Kit (Qiagen, Catalog #69504) and resuspended in kit buffer AE at a DNA concentration of 0.5 µg/µl. PCR amplifications were performed using genomic DNA as the template and several pairs of degenerate primers encoding conserved amino acid sequences in different known PDATs (GenBank Accession Nos. NP 190069 and AB006704 [(gi:2351069)*Arabidopsis thaliana*], and NP_596330 [*Schizosaccharomyces pombe*]; and the *Saccharomyces cerevisiae* Lro 1 gene [Dahlqvist et al., *Proc. Natl. Acad. Sci. USA* 97:6487 (2000)]). The best results were obtained with degenerate primers P26 and P27, as shown in the Table below.

Table 5

Degenerate Primers Used For Amplification Of A Partial Putative PDAT

Primer Set	Description	Degenerate Nucleotide Sequence	Corresponding Amino Acid Sequence
------------	-------------	--------------------------------	-----------------------------------

P26	(32) 29-mers	5'-ATGCTGGACAAGGAG ACCGGNCTNGAYCC-3' (SEQ ID NO:76)	MLDKETGLDP (SEQ ID NO:77)
P27	(16) 33-mers	5'- CCAGATGACGTCGCCG CCCTTGGGNARCATNGA-3' (SEQ ID NO:78)	complementary to SMLPKGGEVIW (SEQ ID NO:79)

[Note: Abbreviations are standard for nucleotides and proteins. The nucleic acid degeneracy code used is as follows: R= A/G; Y=C/T; and N=A/C/G/T.]

The PCR was carried out in a RoboCycler Gradient 40 PCR machine (Stratagene), using the amplification conditions described in the General Methods. The expected PCR product (ca. 600 bp) was detected by 4% NuSieve (FMC) agarose gel electrophoresis, isolated, purified, cloned into the TOPO[®] cloning vector (Invitrogen) and sequenced. The resultant sequence (contained within SEQ ID NO:7) had homology to known PDATs, based on BLAST program analysis (Altschul, S. F. et al., *J. Mol. Biol.* 215:403-410 (1993)).

Targeted Disruption Of *Yarrowia lipolytica* PDAT Gene

Following the sequencing of this ca. 600 bp partial coding region for PDAT, a larger DNA fragment encoding this sequence was discovered in the public *Y. lipolytica* protein database of the "Yeast project *Genolevures*" (Center for Bioinformatics, LaBRI, Talence Cedex, France) (see also Dujon, B. et al., *Nature* 430 (6995):35-44 (2004)). This allowed isolation of a 1008 bp genomic DNA fragment comprising a portion of the PDAT gene from *Y. lipolytica* ATCC #90812 using PCR primers P39 and P42 (SEQ ID NOs:80 and 81).

Targeted disruption of the PDAT gene in *Y. lipolytica* ATCC #90812 was carried out by homologous recombination-mediated replacement of the endogenous PDAT gene with a targeting cassette designated as pLV13 (SEQ ID NO:82). pLV13 was derived from plasmid pLV5 (Example 1). Specifically, pLV13 was created by inserting a 992 bp *Bam* *HI*/*Eco* *RI* fragment into similarly linearized pLV5. The 992 bp DNA fragment contained (in 5' to 3' orientation): 3' homologous sequence from position +877 to +1371 (of the coding sequence (ORF) in SEQ ID NO:7), a *Bgl* *II* restriction site and 5' homologous sequence from position +390 to +876 (of the coding sequence (ORF) in SEQ ID NO:7). The fragment was prepared by PCR amplification of 3' and 5' sequences from the 1008 bp PCR product described above, using PCR primers P39 and P41

(SEQ ID NOs:80 and 83) and P40 and P42 (SEQ ID NOs:84 and 81), respectively.

pLV13 was linearized by *Bgl* II restriction digestion and was transformed into mid-log phase *Y. lipolytica* ATCC #90812 cells by the lithium acetate method (General Methods). The cells were plated onto Bio101 DOB/CSM-Ura selection plates and maintained at 30 °C for 2 to 3 days.

Ten *Y. lipolytica* ATCC #90812 colonies were isolated and screened for targeted disruption by PCR. One set of PCR primers (P51 and P52 [SEQ ID NOs:85 and 86, respectively]) was designed to amplify the targeting cassette. Another set of PCR primers (P37 and P38 [SEQ ID NOs:87 and 88, respectively]) was designed to detect the native gene. Ten of the ten strains were positive for the junction fragment and 3 of the 10 strains were negative for the native fragment, thus confirming successful targeted integration in these 3 strains. Disruption of the gene was further confirmed by GC analysis of total lipids in one of the disrupted strains, designated as "S-P" (see Example 9).

EXAMPLE 4

Construction Of A *Yarrowia lipolytica* Double Knockout Strain Containing Disruptions In Both PDAT And DGAT2 Genes

The present Example describes the creation of a double knockout strain that was disrupted in both PDAT and DGAT2 genes.

Specifically, the *Y. lipolytica* ATCC #90812 hygromycin-resistant "S-D2" mutant (containing the DGAT2 disruption from Example 2) was transformed with plasmid pLV13 (from Example 3) and transformants were screened by PCR, as described in Example 3. Two of twelve transformants were confirmed to be disrupted in both the DGAT2 and PDAT genes. Disruption of the gene was further confirmed by GC analysis of total lipids in one of the disrupted strains, designated as "S-D2-P" (see Example 9).

EXAMPLE 5

Cloning Of Full-Length *Yarrowia lipolytica* DGAT2 And PDAT Genes

The present Example describes the recovery of the genomic sequences flanking the disrupted DGAT2 and PDAT genes by plasmid rescue, using the sequence in the rescued plasmid to PCR the intact ORF of the native gene. The full-length genes and their deduced amino acid sequences are compared to other fungal DGAT2 and PDAT sequences, respectively.

Plasmid Rescue Of *Yarrowia lipolytica* DGAT2 And PDAT Genes

Since the acyltransferase genes were disrupted by the insertion of the entire pY21DGAT2 and pLV13 vectors that each contained an *E. coli* ampicillin-resistant gene and *E. coli* ori, it was possible to rescue the flanking PDAT and DGAT2 sequences in *E. coli*. For this, genomic DNA of *Y. lipolytica* strain "S-D2" (carrying the disrupted DGAT2 gene; Example 2) and *Y. lipolytica* strain "S-P" (carrying the disrupted PDAT gene; Example 3) was isolated using the DNeasy Tissue Kit. Specifically, 10 µg of the genomic DNA was digested with 50 U of the following restriction enzymes in a reaction volume of 200 µl: for DGAT2--*Age I* and *Nhe I*; for PDAT--*Kpn I*, *Pac I* and *Sac I*. Digested DNA was extracted with phenol:chloroform and resuspended in 40 µl deionized water. The digested DNA (10 µl) was self-ligated in a 200 µl ligation mixture containing 3 U T4 DNA ligase. Each ligation reaction was carried out at 16 °C for 12 hrs. The ligated DNA was extracted with phenol:chloroform and resuspended in 40 µl deionized water. Finally, 1 µl of the resuspended ligated DNA was used to transform *E. coli* by electroporation and plated on LB containing ampicillin (Ap). Ap-resistant transformants were isolated and analyzed for the presence of plasmids. The following insert sizes were found in the recovered or rescued plasmids (Tables 5 and 6):

Table 6

Insert Sizes Of Recovered DGAT2 Plasmids, According To Restriction Enzyme

Enzyme	plasmid insert size (kB)
<i>AgeI</i>	2.3
<i>NheI</i>	9.5

Table 7

Insert Sizes Of Recovered PDAT Plasmids, According To Restriction Enzyme

Enzyme	plasmid insert size (kB)
<i>Kpn I</i>	6.9
<i>Sac I</i>	5.4

<i>Sph I</i>	7.0
--------------	-----

Sequencing of the DGAT2 rescued plasmids was initiated with sequencing primers P79 (SEQ ID NO:89) and P95 (SEQ ID NO:69). In contrast, sequencing of the PDAT plasmids was initiated with sequencing primers P84 (SEQ ID NO:90) and P85 (SEQ ID NO:91).

Based on the sequencing results, a full-length gene encoding the *Y. lipolytica* DGAT2 gene was assembled (2119 bp; SEQ ID NO:1). Specifically, the sequence encoded an open reading frame (ORF) of 1545 bases (nucleotides +291 to +1835 of SEQ ID NO:1), while the deduced amino acid sequence was 514 residues in length (SEQ ID NO:2). Since this ORF has an initiation codon ('ATG') at position 1, as well as at positions 56 and 160, it contains at least two additional nested (smaller) ORFs. Specifically, one ORF is 1380 bases long (nucleotides +456 to +1835 of SEQ ID NO:1, corresponding to SEQ ID NO:3), with a deduced amino acid sequence of 459 residues (SEQ ID NO:4); another ORF is 1068 bases long (nucleotides +768 to +1835 of SEQ ID NO:1, corresponding to SEQ ID NO:5) with a deduced amino acid sequence of 355 residues (SEQ ID NO:6).

The ORF encoded by SEQ ID NO:5 has a high degree of similarity to other known DGAT2 enzymes and because disruption in SEQ ID NO:5 eliminated DAG AT function of the native gene (see Example 9), the polypeptide of SEQ ID NO:6 has been identified as clearly having DGAT2 functionality.

Following sequencing and analysis of the DGAT2 protein described above, a *Yarrowia lipolytica* DGAT2 protein sequence was published within the public *Y. lipolytica* protein database of the "Yeast project Genolevures" (sponsored by the Center for Bioinformatics, LaBRI, bâtiment A30, Université Bordeaux 1, 351, cours de la Libération, 33405 Talence Cedex, France) (see also Dujon, B. et al., *Nature* 430 (6995):35-44 (2004)). Specifically, the sequence disclosed therein was identified as ORF YALI-CDS2240.1, encoding 514 amino acids, and the protein was reported to share some similarities with tr|Q08650 *Saccharomyces cerevisiae* YOR245C DGA1 acyl-CoA:diacylglycerol acyltransferase.

In a manner similar to that used to deduce the full-length sequence of DGAT2, a full-length gene encoding the *Y. lipolytica* PDAT gene was assembled (2326 bp; SEQ ID NO:7) based on sequencing results.

Specifically, the sequence encoded an open reading frame of 1944 bases (nucleotides +274 to +2217 of SEQ ID NO:7), while the deduced amino acid sequence was 648 residues in length (SEQ ID NO:8).

Following sequencing and analysis of the PDAT protein described above, the *Yarrowia lipolytica* PDAT protein sequence was published as part of the public *Y. lipolytica* protein database of the "Yeast project *Genolevures*" (*supra*). The PDAT sequence disclosed therein was identified as ORF YALI-CDS1359.1, encoding 648 amino acids, and the protein was reported to share some similarities to sp|P40345 *Saccharomyces cerevisiae* YNR008w LRO1, a lecithin cholesterol acyltransferase-like gene which mediates diacylglycerol esterification.

EXAMPLE 6

Identification of Additional Putative *Yarrowia lipolytica* DAG ATs

In order to identify additional DAG ATs in *Yarrowia*, the public *Y. lipolytica* protein database of the "Yeast project *Genolevures*" (*supra*) was searched using the *Saccharomyces cerevisiae* ARE1 (Sc ARE1; GenBank Accession No. CAA42296) and ARE2 (Sc ARE2; GenBank Accession No. P53629) protein sequences (Yang, H. et al., *Science*. 272(5266):1353-1356 (1996)). Both searches identified the following *Y. lipolytica* ORFs as the first and second hits, respectively:

- (1) YALI-CDS2011.1: annotated as "similar to sp|P53629 *Saccharomyces cerevisiae* YNR019w ARE2 acyl-CoA sterol acyltransferase, hypothetical start"; 543 amino acids in length (SEQ ID NOs:9 and 10); and
- (2) YALI-CDS2141.1: annotated as "unnamed protein product; weakly similar to tr|Q9FUL6 *Perilla frutescens* Diacylglycerol acyltransferase (Pf DGAT1), hypothetical start"; 526 amino acids in length (SEQ ID NOs:11 and 12).

The percent identities between these proteins were determined by the Megalign program of DNASTAR using Clustal W according to the parameters described in the General Methods. The percent (%) identities are shown below, wherein the % identity is defined as percentage of amino acids that are identical between the two proteins:

Table 8
Percent Identities Between Known Acyltransferases And *Yarrowia*
lipolytica ORFs

	Sc ARE1	Sc ARE2	Pf DGAT1
YALI-CDS2141.1	16.6	14.5	29.5
YALI-CDS2011.1	32.6	33.8	18.4

5

Based on this comparison, YALI-CDS2141.1 and YALI-CDS2011.1 (designated herein as "YI DGAT1" and "YI ARE2", respectively) were candidates ORFs that were likely to encode proteins having DAG AT functionality in *Yarrowia*.

10

Following the analysis of the proteins described above, the *Yarrowia lipolytica* strain CLIB99 complete genome was published in GenBank as part of the *Genolevures* project. Thus, the ORF identified as YALI-CDS2011.1 corresponds to GenBank Accession No. NC_006072, locus_tag="YALI0F06578g" and the ORF identified as YALI-CDS2141.1

15

corresponds to GenBank Accession No. CR382130, locus_tag="YALI0D07986g".

EXAMPLE 7

Cloning Of A *Yarrowia lipolytica* Acyl-CoA:Diacylglycerol Acyltransferase (DGAT1) Gene And Disruption Of The Endogenous DGAT1 Gene

20

The present Example describes the use of degenerate PCR primers to isolate the full-length coding sequence of the *Yarrowia lipolytica* DGAT1 (encoded by ORF YALI-CDS2011.1 (Example 6)) and the use of the sequence to disrupt the native gene in *Y. lipolytica*.

Cloning Of A Putative DGAT1 Sequence From *Yarrowia lipolytica* By PCR Using Degenerate PCR Primers

25

The full-length YI DGAT1 ORF was cloned by PCR using degenerate PCR primers P201 and P203 (SEQ ID NOs:92 and 93, respectively) and *Y. lipolytica* ATCC #76982 (from Example 2) genomic DNA as template. The degenerate primers were required, since the

30

nucleotide sequence encoding YI DGAT1 was not known. The PCR was carried out in a RoboCycler Gradient 40 PCR machine, using the components and thermocycler conditions described in the General Methods. The expected PCR product (ca. 1.6 kB) was detected by agarose gel electrophoresis, isolated, purified, cloned into the

TOPO[®] cloning vector (Invitrogen), and partially sequenced to confirm its identity.

Targeted Disruption Of The *Yarrowia lipolytica* DGAT1 Gene

5 Targeted disruption of the putative DGAT1 gene in *Y. lipolytica* ATCC #90812 was carried out by homologous recombination-mediated replacement of the endogenous DGAT1 gene with a targeting cassette (using the methodology described in Example 2). Specifically, the 1.6 kB isolated YI DGAT1 ORF (SEQ ID NO:13) was used as a PCR template molecule to construct a YI DGAT1 targeting cassette consisting of: 5' homologous YI DGAT1 sequence, the *Yarrowia* Leucine 2 (*Leu2*) gene, and 3' homologous YI DGAT1 sequence. For this, each portion of the targeting cassette was first individually amplified, using the primers set forth below:

- 15 • Upper primer P214 and lower primer P215 (SEQ ID NOs:94 and 95, respectively), for amplification of the 5' homologous DGAT1 sequence;
- Upper primer P216 and lower primer P217 (SEQ ID NOs:96 and 97, respectively), for amplification of the 3' homologous DGAT1 sequence; and,
- 20 • Upper primer P218 and lower primer P219 (SEQ ID NOs:98 and 99, respectively), for amplification of the *Leu2* gene (GenBank Accession No. AAA35244).

The PCRs were performed using *Pfu Ultra* polymerase (Stratagene, Catalog #600630), as described in the General Methods, and purified.

25 The three correct-sized, purified fragments were mixed together as template molecules for a second PCR reaction using PCR primers P214 and P219 (SEQ ID NOs:94 and 99) to obtain the YI DGAT1 disruption cassette.

The targeting cassette was gel purified and used to transform mid-log phase wildtype *Y. lipolytica* (ATCC #90812). Transformation was performed as described in the General Methods.

Transformants were plated onto Bio101 DOB/CSM-Leu selection plates and maintained at 30 °C for 2 to 3 days. Several leucine prototrophs were screened by PCR to confirm the targeted DGAT1

35 disruption. Specifically, one set of PCR primers (P226 and P227 [SEQ ID NOs:100 and 101, respectively]) was designed to amplify a junction between the disruption cassette and native target gene. Another set of

PCR primers (P214 and P217 [SEQ ID NOs:94 and 97, respectively]) was designed to detect the native gene.

5 All of the leucine prototroph colonies were positive for the junction fragment and negative for the native fragment. Thus, targeted integration was confirmed in these strains. Disruption of the gene was further confirmed by GC analysis of total lipids of one of the disrupted strains, designated as "S-D1" (see Example 9).

10 In a similar manner, the DGAT1 targeting cassette was used to disrupt the DGAT1 gene in strains containing single disruptions in either PDAT ("S-P" from Example 3), DGAT2 ("S-D2" from Example 2), or double disruptions in PDAT and DGAT2 ("S-D2-P" from Example 4). This resulted in the creation of strains with double knockouts in DGAT1 and PDAT ("S-D1-P"), in DGAT2 and DGAT1 ("S-D2-D1") and triple knockouts in DGAT2, DGAT1 and PDAT ("S-D2-D1-P").

15 EXAMPLE 8

Cloning Of *Yarrowia lipolytica* Acyl-CoA: Sterol-Acyltransferase (ARE2) Gene And Disruption Of The Endogenous ARE2 Gene

The present Example describes the use of degenerate PCR primers to isolate the full-length coding sequence of the *Yarrowia lipolytica* ARE2 (encoded by ORF YALI-CDS2141.1 (Example 6)) and the use of the sequence to disrupt the native gene in *Y. lipolytica*.

Cloning Of A Putative ARE2 Sequence From *Yarrowia lipolytica* By PCR Using Degenerate PCR Primers

25 The full length YI ARE2 ORF was cloned by PCR using degenerate PCR primers P205 and P208 (SEQ ID NOs:102 and 103, respectively) and *Y. lipolytica* ATCC #76982 (from Example 2) genomic DNA as template. The degenerate primers were required, since the nucleotide sequence encoding YI ARE2 was not known. The PCR was performed using the protocol described in Example 7. A PCR product of the expected size was detected.

Targeted Disruption Of The *Yarrowia lipolytica* ARE2 Gene (Prophetic)

35 Targeted disruption of the ARE2 gene in *Y. lipolytica* ATCC #90812 will be carried out by homologous recombination-mediated replacement of the endogenous ARE2 gene with a targeting cassette (as described in Example 7). Specifically, the ~1.6 kB isolated ORF encoding the putative YI ARE2 protein (SEQ ID NO:15) will be used as a PCR template molecule to construct a YI ARE2 targeting cassette consisting of: 5'

homologous YI ARE2 sequence, the *Yarrowia* Leucine 2 (*Leu2*) gene, and 3' homologous YI ARE2 sequence. For this, each portion of the targeting cassette will be first individually amplified, as described in Example 7, using the primers set forth below:

- 5 • Upper primer P220 and lower primer P221 (SEQ ID NOs:104 and 105, respectively), for amplification of the 5' homologous ARE2 sequence;
- Upper primer P222 and lower primer P223 (SEQ ID NOs:106 and 107, respectively), for amplification of the 3' homologous ARE2 sequence; and,
- 10 • Upper primer P224 and lower primer P225 (SEQ ID NOs:108 and 109, respectively), for amplification of the *Leu2* gene.

Following purification, each of the correct-sized fragments will be mixed and utilized as template molecules in a PCR reaction using primers P220 and P223 to obtain the targeting cassette. Following gel purification of the product, the targeting cassette will be used to transform mid-log phase wildtype and mutant *Y. lipolytica* (ATCC #90812) strains containing single disruptions in PDAT ("S-P" from Example 3), DGAT2 ("S-D2" from Example 2), DGAT1 ("S-D1" from Example 7), double disruptions in PDAT and DGAT2 ("S-D2-P" from Example 4), double disruptions in PDAT and DGAT1 ("S-D1-P" from Example 7), double disruptions in DGAT1 and DGAT2 ("S-D1-D2" from Example 7) or triple disruptions in PDAT, DGAT2 and DGAT1 ("S-D1-D2-P" from Example 7). Transformation will be performed as described in the General Methods.

25 Transformants will be plated onto Bio101 DOB/CSM-Leu selection plates and maintained at 30 °C for 2 to 3 days. Several leucine prototrophs will be screened by PCR to confirm the targeted ARE2 disruption, using the methodology described in Example 7.

EXAMPLE 9

30 Determination Of TAG Content In Mutant And Wildtype *Yarrowia lipolytica* Strains (ATCC #90812)

The present Example describes a comparison of TAG content in wildtype and mutant *Y. lipolytica* ATCC #90812 containing: (1) single disruptions in PDAT, DGAT2 and DGAT1; (2) double disruptions in PDAT and DGAT2, DGAT1 and PDAT, and DGAT1 and DGAT2; and (3) triple disruptions in PDAT, DGAT2 and DGAT1.

Specifically, single colonies of wildtype and mutant *Y. lipolytica* containing single disruptions in PDAT ("S-P", from Example 3), DGAT2 ("S-D2", from Example 2), DGAT1 ("S-D1", from Example 7), double disruptions in PDAT and DGAT2 ("S-D2-P", from Example 4), DGAT1 and PDAT ("S-D1-P", from Example 7), DGAT1 and DGAT2 ("S-D1-D2", from Example 7), and triple disruptions ("S-D1-D2-P", from Example 7) were separately grown using conditions that induce oleaginy. One loopful of cells from each culture was each individually inoculated into 3 mL YPD medium and grown overnight on a shaker (300 rpm) at 30 °C. The cells were harvested and washed once in 0.9% NaCl and resuspended in 50 mL of HGM. Cells were then grown on a shaker for 48 hrs. Cells were washed in water and the cell pellet was lyophilized. Twenty (20) mg of dry cell weight was used for total fatty acid by GC analysis and the oil fraction following TLC (*infra*) and GC analysis.

15 Thin Layer Chromatography (TLC)

The methodology used for TLC is described below in the following five steps: (1) The internal standard of 15:0 fatty acid (10 µl of 10 mg/mL) was added to 2 to 3 mg dry cell mass, followed by extraction of the total lipid using a methanol/chloroform method. (2) Extracted lipid (50 µl) was blotted across a light pencil line drawn approximately 1 inch from the bottom of a 5x20 cm silica gel 60 plate, using 25-50 µl micropipettes. (3) The TLC plate was then dried under N₂ and was inserted into a tank containing about ~100 mL 80:20:1 hexane:ethyl ether:acetic acid solvent. (4) After separation of bands, a vapor of iodine was blown over one side of the plate to identify the bands. This permitted samples on the other side of the plate to be scraped using a razor blade for further analysis. (5) Basic transesterification of the scraped samples and GC analysis was performed, as described in the General Methods.

Results From GC Analysis

GC results are shown below in Table 9. Cultures are described as the "S" strain (wildtype), "S-P" (PDAT knockout), "S-D1" (DGAT1 knockout), "S-D2" (DGAT2 knockout), "S-D1-D2" (DGAT1 and DGAT2 knockout), "S-P-D1" (PDAT and DGAT1 knockout), "S-P-D2" (PDAT and DGAT2 knockout) and "S-P-D1-D2" (PDAT, DGAT1 and DGAT2 knockout). Abbreviations utilized are: "WT" = wildtype; "FAs" = fatty acids; "dcw" = dry cell weight; and, "FAs % dcw, % WT" = FAs % dcw relative to the % in wildtype, wherein the "S" strain is wildtype.

Table 9

Lipid Content In *Yarrowia* ATCC #90812 Strains With Single, Double Or Triple Disruptions In PDAT, DGAT2 And DGAT1

Strain	Residual DAG AT	dcw, mg	Total Fatty Acids			TAG Fraction		
			FAs, μg	FAs % dcw	FAs % dcw, % WT	FAs, μg	FAs % dcw	FAs % dcw, % WT
S	D1, D2, P	32.0	797	15.9	100	697	13.9	100
S-D1	D2, P	78.8	723	13.6	86	617	11.6	83
S-D2	D1, P	37.5	329	6.4	40	227	4.4	32
S-P	D1, D2	28.8	318	6.0	38	212	4.0	29
S-D1-D2	P	64.6	219	4.1	26	113	2.1	15
S-D1-P	D2	76.2	778	13.4	84	662	11.4	82
S-D2-P	D1	31.2	228	4.3	27	122	2.3	17
S-D1-D2-P	None	52.2	139	2.4	15	25	0.4	3

5

The results in Table 9 provide evidence that DGAT1 is a DAG AT, since its disruption resulted in lower oil content compared to the wild type strain. The results shown above also indicate the relative contribution of the three DAG ATs to oil biosynthesis. DGAT2 contributes the most, while PDAT and DGAT1 contribute equally but less than DGAT2. The residual oil content ca. 3% in the triple knockout strain may be the contribution of the YI ARE2 (see Example 8).

10

EXAMPLE 10

Generation Of EPA-Producing *Yarrowia lipolytica* ATCC #20362 Strain

15

MU and Strain Y2067U

The present Example describes the construction of strain MU and strain Y2067U, each derived from *Yarrowia lipolytica* ATCC #20362, wherein each strain was capable of producing significant concentrations of EPA relative to the total lipids (Figure 5). The affect of various acyltransferase knockouts and acyltransferase gene overexpression was examined in these EPA producing strains based on analysis of TAG content and/or fatty acid composition, as described in Examples 11, 12 and 15 (*infra*).

20

The development of strain MU herein required the construction of strain M4 (producing 8% DGLA), strain Y2034 (producing 10% ARA),

25

strain E (producing 10% EPA), strain EU (producing 10% EPA) and strain M26 (producing 14%). The development of strain Y2067U first required the creation of a derivative of strain EU, designated as strain Y2067 (producing 15% EPA).

5 Construction Of Strain M4 Producing 8% DGLA

Construct pKUNF12T6E (Figure 6A; SEQ ID NO:110) was generated to integrate four chimeric genes (comprising a $\Delta 12$ desaturase, a $\Delta 6$ desaturase and 2 elongases) into the *Ura3* loci of wild type *Yarrowia* strain ATCC #20362, to thereby enable production of DGLA. The

10 pKUNF12T6E plasmid contained the following components:

Table 10

Description of Plasmid pKUNF12T6E (SEQ ID NO:110)

RE Sites And Nucleotides Within SEQ ID NO:110	Description Of Fragment And Chimeric Gene Components
<i>AscI/BsiWI</i> (9420-8629)	784 bp 5' part of <i>Yarrowia Ura3</i> gene (GenBank Accession No. AJ306421)
<i>SphI/PacI</i> (12128-1)	516 bp 3' part of <i>Yarrowia Ura3</i> gene (GenBank Accession No. AJ306421)
<i>SwaI/BsiWI</i> (6380-8629)	FBAIN::EL1S:Pex20, comprising: • FBAIN: FBAIN promoter (SEQ ID NO:111; see also U.S. Patent Application No. 60/519971) • EL1S: codon-optimized elongase 1 gene (SEQ ID NO:112), derived from <i>Mortierella alpina</i> (GenBank Accession No. AX464731) • Pex20: Pex20 terminator sequence from <i>Yarrowia Pex20</i> gene (GenBank Accession No. AF054613)
<i>BglII/SwaI</i> (4221-6380)	TEF:: $\Delta 6$ S::Lip1, comprising: • TEF: TEF promoter (GenBank Accession No. AF054508) • $\Delta 6$S: codon-optimized $\Delta 6$ desaturase gene (SEQ ID NO:114), derived from <i>Mortierella alpina</i> (GenBank Accession No. AF465281) • Lip1: Lip1 terminator sequence from <i>Yarrowia Lip1</i> gene (GenBank Accession No. Z50020)

<i>PmeI/Clal</i> (4207-1459)	FBA::F. Δ 12::Lip2, comprising: • FBA: FBA promoter (SEQ ID NO:116; see also U.S. Patent Application No. 60/519971) • F.Δ12: <i>Fusarium moniliforme</i> Δ12 desaturase gene (SEQ ID NO:117) • Lip2: Lip2 terminator sequence from <i>Yarrowia Lip2</i> gene (GenBank Accession No. AJ012632)
<i>Clal/PacI</i> (1459-1)	TEF::EL2S::XPR, comprising: • TEF: TEF promoter (GenBank Accession No. AF054508) • EL2S: codon-optimized elongase gene (SEQ ID NO:119), derived from <i>Thraustochytrium aureum</i> (U.S. 6,677,145) • XPR: XPR terminator sequence of <i>Yarrowia Xpr</i> gene (GenBank Accession No. M17741)

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.

GC analyses showed the presence of DGLA in the transformants containing the 4 chimeric genes of pKUNF12T6E (Figure 6A), 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 About 10% ARA

Constructs pDMW232 (Figure 6B; SEQ ID NO:121) was generated to integrate two Δ 5 chimeric genes into the *Leu2* gene of *Yarrowia* strain M4. The plasmid pDMW232 contained the following components:

Table 11
Description of Plasmid pDMW232 (SEQ ID NO:121)

RE Sites And Nucleotides Within SEQ ID NO:121	Description Of Fragment And Chimeric Gene Components
<i>AscI/BsiWI</i> (5550-4755)	788 bp 5' part of <i>Yarrowia Leu2</i> gene (GenBank Accession No. AF260230)
<i>SphI/PacI</i> (8258-8967)	703 bp 3' part of <i>Yarrowia Leu2</i> gene (GenBank Accession No. AF260230)
<i>SwaI/BsiWI</i> (2114-4755)	FBAIN::MA Δ 5::Pex20, comprising: • FBAIN: FBAIN Promoter (SEQ ID NO:111; see also U.S. Patent Application No. 60/519971) • MAΔ5: <i>Mortierella alpina</i> Δ5 desaturase gene (SEQ ID NO:122) (GenBank Accession No. AF067654) • Pex20: Pex20 terminator sequence of <i>Yarrowia Pex20</i> gene (GenBank Accession No. AF054613)
<i>SwaI/Clal</i> (2114-17)	TEF::MA Δ 5::Lip1, comprising: • TEF: TEF promoter (GenBank Accession No. AF054508) • MAΔ5: as described for FBAIN::MAΔ5::Pex20 (<i>supra</i>) • Lip1: Lip1 terminator sequence of <i>Yarrowia Lip1</i> gene (GenBank Accession No. Z50020)
<i>PmeI/Clal</i> (5550-4755)	<i>Yarrowia Ura3</i> gene (GenBank Accession No. AJ306421)

5

Plasmid pDMW232 was digested with *AscI/SphI*, and then used to transform strain M4 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 from each transformation 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,

10

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

GC analyses showed the presence of ARA in pDMW232 transformants, but not in the parental M4 strain. Specifically, among the
5 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".

10 Construction Of Strain E, Producing About 10% EPA

Construct pZP3L37 (Figure 6C; SEQ ID NO:124) was created to integrate three synthetic $\Delta 17$ desaturase chimeric genes into the acyl-CoA oxidase 3 (i.e., POX3) gene of the Y2034 strain. The plasmid pZP3L37 contained the following components:

15

Table 12

Description of Plasmid pZP3L37 (SEQ ID NO:124)

RE Sites And Nucleotides Within SEQ ID NO:124	Description Of Fragment And Chimeric Gene Components
--	--

<i>AscI/BsiWI</i> (6813-6043)	763 bp 5' part of <i>Yarrowia Pox3</i> gene (GenBank Accession No. AJ001301)
<i>SphI/PacI</i> (9521-10345)	818 bp 3' part of <i>Yarrowia Pox3</i> gene (GenBank Accession No. AJ001301)
<i>Clal/BsiWI</i> (4233-6043)	TEF:: Δ 17S::Pex20, comprising: • TEF: TEF promoter (GenBank Accession No. AF054508) • Δ17S: codon-optimized Δ17 desaturase gene (SEQ ID NO:125), derived from <i>S. diclina</i> (US 2003/0196217 A1) • Pex20: Pex20 terminator sequence of <i>Yarrowia Pex20</i> gene (GenBank Accession No. AF054613)
<i>Clal/PmeI</i> (4233-1811)	FBAIN:: Δ 17S::Lip2, comprising: • FBAIN: FBAIN promoter (SEQ ID NO:111; see also U.S. Patent Application No. 60/519971) • Δ17S: SEQ ID NO:125 (<i>supra</i>) • Lip2: Lip2 terminator sequence of <i>Yarrowia Lip2</i> gene (GenBank Accession No. AJ012632)
<i>PmeI/SwaI</i> (1811-1)	<i>Yarrowia Leu2</i> gene (GenBank Accession No. AF260230)
<i>PacI/SwaI</i> (10345-1)	FBAINm:: Δ 17S::Pex16, comprising: • FBAINm: FBAINm promoter (SEQ ID NO:127; see also U.S. Patent Application No. 60/519971) • Δ17S: SEQ ID NO:125 (<i>supra</i>) • Pex16: Pex16 terminator sequence of <i>Yarrowia Pex16</i> gene (GenBank Accession No. U75433)

Plasmid pZP3L37 was digested with *AscI/SphI*, and then used to transform strain Y2034 according to the General Methods. Following transformation, the cells were plated onto MM plates and maintained at 30 °C for 2 to 3 days. A total of 48 transformants grown on the MM plates were picked and re-streaked onto fresh MM plates. Once grown, these strains were individually inoculated into liquid MM at 30 °C and shaken at 250 rpm/min for 2 days. The cells were collected by centrifugation, lipids were extracted, and fatty acid methyl esters were prepared by transesterification, and subsequently analyzed with a Hewlett-Packard 6890 GC.

GC analyses showed the presence of EPA in most of the transformants with pZP3L37, but not in the parental strain (i.e., Y2034). Among the 48 selected transformants with pZP3L37, there were 18 strains that produced less than 2% EPA, 14 strains that produced 2-3% EPA, and
5 1 strain that produced about 7% EPA of total lipids in the engineered *Yarrowia*.

The strain that produced 7% EPA was further analyzed after culturing the strain as follows ("two-stage growth conditions"). First, cells were grown in triplicate in liquid MM at 30 °C with shaking at 250 rpm/min
10 for 48 hrs. The cells were collected by centrifugation and the liquid supernatant was extracted. The pelleted cells were resuspended in HGM and grown for 72 hrs at 30 °C with shaking at 250 rpm/min. The cells were again collected by centrifugation and the liquid supernatant was extracted.

GC analyses showed that the engineered strain produced about
15 10% EPA of total lipids after the two-stage growth. The strain was designated as the "E" strain.

Construction Of Strain EU Producing About 10% EPA

Strain EU (*Ura*⁻) was created by identifying mutant cells of strain E that were 5-FOA resistant. Specifically, one loop of *Yarrowia* E strain cells
20 were inoculated into 3 mL YPD medium and grown at 30 °C with shaking at 250 rpm for 24 hrs. The culture was diluted with YPD to an OD₆₀₀ of 0.4 and then incubated for an additional 4 hrs. The culture was plated (100 µl/plate) onto FOA selection plates and maintained at 30 °C for 2 to 3 days. A total of 16 FOA resistant colonies were picked and streaked onto MM
25 and FOA selection plates. From these, 10 colonies grew on FOA selection plates but not on MM plates and were selected as potential *Ura*⁻ strains.

One of these strains was used as host for transformation with pY37/F15, comprising a chimeric GPD::*Fusarium moniliforme* Δ15::XPR2
30 gene and a *Ura3* gene as a selection marker (Figure 6D; SEQ ID NO:128). After three days of selection on MM plates, hundreds of colonies had grown on the plates and there was no colony growth of the transformation control that carried no plasmid. This 5-FOA resistant strain was designated as strain "EU".

35 Single colonies of the EU strain were then inoculated into liquid MMU additionally containing 0.1 g/L uridine and cultured for 2 days at 30 °C with shaking at 250 rpm/min. The cells were collected by

centrifugation, lipids were extracted, and fatty acid methyl esters were prepared by trans-esterification and subsequently analyzed with a Hewlett-Packard 6890 GC. GC analyses showed that the EU strain produced about 10% EPA of total lipids.

5 Construction Of Strain M26, Producing 14% EPA

Construct pZKO2UM26E (Figure 6E, SEQ ID NO:129) was used to integrate a cluster of three chimeric genes (comprising an elongase, a $\Delta 6$ desaturase and a $\Delta 12$ desaturase) and a *Ura3* gene into the *Yarrowia* $\Delta 12$ desaturase gene site of strain EU. Plasmid pKO2UM26E contained the
10 following components:

Table 13
Description of Plasmid pKO2UM26E (SEQ ID NO:129)

RE Sites And Nucleotides Within SEQ ID NO:129	Description Of Fragment And Chimeric Gene Components
<i>HindIII</i> / <i>AscI</i> (1-728)	728 bp 5' part of <i>Yarrowia</i> $\Delta 12$ desaturase gene (SEQ ID NO:130)
<i>SphI</i> / <i>EcoRI</i> 3436-3992)	556 bp 3' part of <i>Yarrowia</i> $\Delta 12$ desaturase gene (SEQ ID NO:130)
<i>BsiWI</i> / <i>HindIII</i> (11095-1)	GPAT::EL1S::XPR, comprising: • GPAT: GPAT promoter (SEQ ID NO:132; see also U.S. Patent Application No. 60/610060) • EL1S: codon-optimized elongase 1 gene (SEQ ID NO:112), derived from <i>Mortierella alpina</i> (GenBank Accession No. AX464731) • XPR: terminator sequence of <i>Yarrowia Xpr2</i> gene (GenBank Accession No. M17741)
<i>BglII</i> / <i>BsiWI</i> (8578-11095)	FBAIN::M. $\Delta 12$.Pex20, comprising: • FBAIN: FBAIN promoter (SEQ ID NO:111; see also U.S. Patent Application No. 60/519971) • M.$\Delta 12$: <i>Mortierella isabellina</i> $\Delta 12$ desaturase gene (GenBank Accession No. AF417245; SEQ ID NO:133) ▪ Pex20: Pex20 terminator sequence of <i>Yarrowia Pex20</i> gene (GenBank Accession No. AF054613)

<i>Sall/PacI</i> (6704-8202)	<i>Yarrowia</i> Ura3 gene (GenBank Accession No. AJ306421)
<i>EcoRI/Sall</i> (3992-6704)	FBAIN::M. Δ 6B::Pex20, comprising: • FBAIN: FBAIN promoter (SEQ ID NO:111; see also U.S. Patent Application No. 60/519971) • M.Δ6B: <i>Mortierella alpina</i> Δ6 desaturase gene "B" (GenBank Accession No. AB070555; SEQ ID NO:135) • Pex20: Pex20 terminator sequence of <i>Yarrowia</i> Pex20 gene (GenBank Accession No. AF054613)

The plasmid pKO2UM26E was digested with *SphI/AscI*, and then used to transform EU strain according to the General Methods. Following transformation, cells were plated onto MM plates and maintained at 30 °C for 2 to 3 days.

A total of 48 transformants grown on MM plates were picked and re-streaked onto fresh MM plates. Once grown, these strains were individually inoculated into liquid MM at 30 °C and grown with shaking at 250 rpm/min for 1 day. The cells were collected by centrifugation, lipids were extracted, and fatty acid methyl esters were prepared by transesterification, and subsequently analyzed with a Hewlett-Packard 6890 GC.

GC analyses showed that EPA was produced in almost all transformants with pKO2UM26E after one-day growth in MM media. Among the 48 selected transformants, 5 strains produced less than 4% EPA, 23 strains produced 4-5.9% EPA, 9 strains produced 6-6.9% EPA and 11 strains produced 7-8.2% EPA of total lipids in the engineered *Yarrowia*. The strain that produced 8.2% EPA was selected for further analysis using a two-stage growth procedure (i.e., 48 hrs MM + 96 hrs in HGM). GC analyses showed that the engineered strain produced about 14% EPA of total lipids. The strain was designated as strain "M26". The final genotype of the M26 strain with respect to wildtype *Yarrowia lipolytica* ATCC #20362 was as follows: *Pox3-*, *Y. Δ 12-*, *FBA::F. Δ 12::Lip2*, *FBAIN::M Δ 12::Pex20*, *TEF:: Δ 6S::Lip1*, *FBAIN:: Δ 6B::Pex20*, *FBAIN::E1S::Pex20*; *GPAT::E1S::Xpr*, *TEF::E2S::Xpr*; *FBAIN::MA Δ 5::Pex20*, *TEF::MA Δ 5::Lip1*, *FBAIN:: Δ 17S::Lip2*, *FBAINm:: Δ 17S::Pex16* and *TEF:: Δ 17S::Pex20*.

Construction Of Strain MU, Producing EPA

Strain MU was a *Ura* auxotroph of strain M26. This strain was made by transforming strain M26 with 5 µg of plasmid pZKUM (Figure 7A; SEQ ID NO:137) that had been digested with *PacI* and *HincII*. Transformation was performed using the Frozen-EZ Yeast Transformation kit (Zymo Research Corporation, Orange, CA) and transformants were selected by plating 100 µl of the transformed cell mix on an agar plate with the following medium: 6.7 g/L yeast Nitrogen Base (DIFCO Laboratories, Detroit, MI), 20 g/L dextrose, 50 mg/L uracil and 800 mg/L FOA. After 7 days, small colonies appeared that were plated on MM and MMU agar plates. All were *Ura* auxotrophs. One of the strains was designated "MU".

Construction Of Strain Y2067 Producing About 15% EPA

Plasmid pKO2UF2PE (Figure 7B; SEQ ID NO:138) was created to integrate a cluster containing two chimeric genes (comprising a heterologous $\Delta 12$ desaturase and an elongase) and a *Ura3* gene into the native *Yarrowia* $\Delta 12$ desaturase gene of strain EU (*supra*). Plasmid pKO2UF2PE contained the following components:

Table 14

Description of Plasmid pKO2UF2PE (SEQ ID NO:138)

20

RE Sites And Nucleotides Within SEQ ID NO:138	Description Of Fragment And Chimeric Gene Components
<i>AscI/BsiWI</i> (3382-2645)	730 bp 5' part of <i>Yarrowia</i> $\Delta 12$ desaturase gene (SEQ ID NO:130)
<i>SphI/EcoRI</i> (6090-6646)	556 bp 3' part of <i>Yarrowia</i> $\Delta 12$ desaturase gene (SEQ ID NO:130)
<i>Swal/BsiWI</i> (1-2645)	FBAINm::F. $\Delta 12$ DS::Pex20, comprising: • FBAINm: FBAINm promoter (SEQ ID NO:127; see also U.S. Patent Application No. 60/519971) • F.$\Delta 12$: <i>Fusarium moniliforme</i> $\Delta 12$ desaturase gene (SEQ ID NO:117) • Pex20: Pex20 terminator sequence of <i>Yarrowia</i> <i>Pex20</i> gene (GenBank Accession No. AF054613)

<i>SwaI/PmeI</i> (1-8525)	GPAT::EL1S::OCT, comprising: • GPAT: GPAT promoter (SEQ ID NO:132; see also U.S. Patent Application No. 60/610060) • EL1S: codon-optimized elongase 1 gene (SEQ ID NO:112), derived from <i>Mortierella alpina</i> (GenBank Accession No. AX464731) • OCT: OCT terminator sequence of <i>Yarrowia</i> OCT gene (GenBank Accession No. X69988)
<i>EcoRI/PacI</i> (6646-8163)	<i>Yarrowia Ura3</i> gene (GenBank Accession No. AJ306421)

Plasmid pKO2UF2PE was digested with *AscI/SphI* and then used to transform strain EU according to the General Methods (although strain EU was streaked onto a YPD plate and grown for approximately 36 hr prior to suspension in transformation buffer [versus 18 hrs]). Following transformation, cells were plated onto MM plates and maintained at 30 °C for 2 to 3 days. A total of 72 transformants grown on MM plates were picked and re-streaked separately onto fresh MM plates. Once grown, these strains were individually inoculated into liquid MM at 30 °C and shaken at 250 rpm/min for 2 days. The cells were collected by centrifugation, lipids were extracted, and fatty acid methyl esters were prepared by trans-esterification, and subsequently analyzed with a Hewlett-Packard 6890 GC.

GC analyses showed the presence of EPA in almost all of the transformants with pKO2UF2PE. More specifically, among the 72 selected transformants, there were 17 strains that produced 8-9.9% EPA, 27 strains that produced 10-10.9% EPA, 16 strains that produced 11-11.9% EPA, and 7 strains that produced 12-12.7% EPA of total lipids in the engineered *Yarrowia*. The strain that produced 12.7% EPA was further analyzed by using two-stage growth conditions. GC analyses showed that the engineered strain produced about 15% EPA of total lipids after the two-stage growth. The strain was designated as strain "Y2067".
Construction Of Strain Y2067U Producing About 14% EPA With *Ura*-Phenotype

In order to disrupt the *Ura3* gene in Y2067 strain, construct pZKUT16 (Figure 7C; SEQ ID NO:139) was created to integrate a TEF::rELO2S::Pex20 chimeric gene into the *Ura3* gene of strain Y2067.

rELO2S is a codon-optimized rELO gene encoding a rat hepatic enzyme that elongates 16:0 to 18:0. The plasmid pZKUT16 contained the following components:

5

Table 15

Description of Plasmid pZKUT16 (SEQ ID NO:139)

RE Sites And Nucleotides Within SEQ ID NO:139	Description Of Fragment And Chimeric Gene Components
<i>BsiWI/PacI</i> (1-721)	721 bp 5' part of <i>Yarrowia</i> Ura3 gene (GenBank Accession No. AJ306421)
<i>Sall/ClaI</i> (3565-4289)	724 bp 3' part of <i>Yarrowia</i> Ura3 gene (GenBank Accession No. AJ306421)
<i>ClaI/BsiWI</i> (4289-1)	TEF::rELO2S::Pex20, comprising: • TEF: TEF Promoter (GenBank Accession No. AF054508) • rELO2S: codon-optimized rELO2 elongase gene (SEQ ID NO:140), derived from rat (GenBank Accession No. AB071986) • Pex 20: Pex20 terminator sequence of <i>Yarrowia Pex20</i> gene (GenBank Accession No. AF054613)

10 The plasmid pZKUT16 was digested with *Sall/PacI*, and then used to transform Y2067 strain according to the General Methods. Following transformation, cells were plated onto FOA selection plates and maintained at 30 °C for 2 to 3 days.

15 A total of 24 transformants grown on FOA plates were picked and re-streaked onto MM plates and FOA plates, separately. The strains that could grow on FOA plates, but not on MM plates, were selected as *Ura*-strains. A total of 10 *Ura*- strains were individually inoculated into liquid MMU media at 30 °C and grown with shaking at 250 rpm/min for 1 day. The cells were collected by centrifugation, lipids were extracted, and fatty acid methyl esters were prepared by trans-esterification, and subsequently
20 analyzed with a Hewlett-Packard 6890 GC.

GC analyses showed the presence of 5 to 7% EPA in all of the transformants with pZKUT16 after one day growth in MMU media. The strain that produced 6.2% EPA was further analyzed using two-stage growth conditions (48 hrs MM + 96 hrs in HGM). GC analyses showed that the engineered strain produced about 14% EPA of total lipids. The strain was designated as strain "Y2067U". The final genotype of this strain with respect to wildtype *Yarrowia lipolytica* ATCC #20362 was as follows: *Ura3-*, *Pox3-*, *Y.Δ12-*, *FBA::F.Δ12::Lip2*, *FBAINm::F. Δ12::Pex20*, *TEF::Δ6S::Lip1*, *FBAIN::E1S::Pex20*; *GPAT::E1S:: Oct*, *TEF::E2S::Xpr*; *FBAIN::MAΔ5::Pex20*, *TEF::MAΔ5::Lip1*, *FBAIN::Δ17S::Lip2*, *FBAINm::Δ17S::Pex16*, *TEF::Δ17S::Pex20* and *TEF::rELO2S::Pex20*.

EXAMPLE 11

Determination Of TAG Content In Mutant And Wildtype *Yarrowia lipolytica* Strain MU Engineered For EPA Biosynthesis

The present Example describes TAG content and fatty acid composition in various acyltransferase knockout strains of strain MU (an engineered strain of *Yarrowia lipolytica* ATCC #20362, capable of producing greater than 10% EPA). More specifically, single disruptions in PDAT, DGAT2 and DGAT1 and double disruptions in PDAT and DGAT2 were created in strain MU. Lipid content and composition is compared in each of these strains, following growth in 4 different growth conditions.

More specifically, single disruptions in PDAT, DGAT2, DGAT1 were created in strain MU (*supra*, Example 10), using the methodology described in Examples 2, 3, and 7 (with the exception that selection for the DGAT1 disruption relied on the URA3 gene). This resulted in single knockout strains identified as "MU-D1" (disrupted in DGAT1), "MU-D2" (disrupted in DGAT2) and "MU-P" (disrupted in PDAT). Individual knockout strains were confirmed by PCR. Additionally, the MU-D2 strain was disrupted for the PDAT gene and the disruption confirmed by PCR. The resulting double knockout strain was designated "MU-D2-P".

The MU-D1, MU-D2, MU-P and M-D2-P knockout strains were analyzed to determine each knockout's effect on lipid content and composition, as described below. Furthermore, the growth conditions promoting oleaginity were also explored to determine their effect on total lipid content. Thus, in total, four different experiments were conducted, identified as "Experiment A", "Experiment B", "Experiment C" and "Experiment E". Specifically, three loops of cells from plates containing each strain above were inoculated into MMU [3 mL for Experiments B and

C; and 50 mL for Experiments A and E] and grown in a shaker at 30 °C for 24 hrs (for Experiments A, B and C) or 48 hrs (for Experiment E). Cells were harvested, washed once in HGM, resuspended in either HGM (50 mL for Experiments A and E; and 3 mL for Experiment B) or HGM with uracil ("HGMU") (3 mL for Experiment C) and cultured as above for 4 days. One aliquot (1 mL) was used for lipid analysis by GC as described according to the General Methods, while a second aliquot was used for determining the culture OD at 600 nm. The remaining culture in Experiments A and E was harvested, washed once in water and lyophilized for dry cell weight (dcw) determination. In contrast, the dcw in Experiments B and C were determined from their OD₆₀₀ using the equation showing their relationship. The fatty acid compositions of each of the different strains in Experiments A, B, C and E were also determined.

The results are shown in Table 16 below. Cultures are described as the "MU" strain (the parent EPA producing strain), "MU-P" (PDAT knockout), "MU-D1" (DGAT1 knockout), "MU-D2" (DGAT2 knockout) and "MU-D2-P" (DGAT2 and PDAT knockouts). Abbreviations utilized are: "WT" = wildtype (i.e., MU); "OD" = optical density; "dcw" = dry cell weight; "TFAs" = total fatty acids; and, "TFAs % dcw, % WT" = TFAs % dcw relative to the wild type ("MU") strain. Fatty acids are identified as 16:0, 16:1, 18:0, 18:1 (oleic acid), 18:2 (LA), GLA, DGLA, ARA, ETA and EPA; and the composition of each is presented as a % of the total fatty acids.

Table 16
Lipid And EPA Content In *Yarrowia* Strain MU With Various Acyltransferase Disruptions

Expt	Strain	Residual DAG AT	1 st Phase Growth Condition	2 nd Phase Growth Condition	OD	dcw (mg)	TFAs (μg)	TFAs % dcw	TFAs % dcw, % WT	% 16:0	% 16:1	% 18:0	% 18:1	% 18:2	% GLA	% DGLA	% ARA	% ETA	% EPA
A	MU	D1, D2, P	1 day, 50 mL MMU	4 days, 50 mL HGM	4.0	91	374	20.1	100	17	10	2	18	10	22	7	1	3	9.7
A	MU-D2	D1, P			3.1	75	160	10.4	52	16	12	0	8	9	23	7	0	8	17.4
A	MU-D1	D2, P			4.3	104	217	10.2	51	15	10	2	11	10	22	7	0	7	17.4
A	MU-P	D1, D2			4.4	100	238	11.7	58	16	9	2	11	7	24	7	1	6	17.5
B	MU	D1, D2, P	1 day, 3 mL MMU	4 days, 3 mL HGM	5.9	118	581	24.1	100	17	9	3	18	10	22	8	1	3	9.1
B	MU-D2	D1, P			4.6	102	248	11.9	50	16	10	0	7	10	24	7	1	7	17.8
B	MU-D1	D2, P			6.1	120	369	15.0	62	18	9	3	14	11	20	7	1	5	12.0
B	MU-P	D1, D2			6.4	124	443	17.5	72	15	8	3	16	10	25	6	1	4	11.9
C	MU	D1, D2, P	1 day, 3 mL MMU	4 days, 3 mL HGMU	6.8	129	522	19.9	100	16	10	2	13	11	21	10	1	4	12.6
C	MU-D2	D1, P			5.6	115	239	10.2	51	17	9	1	6	11	21	8	1	7	18.9
C	MU-D1	D2, P			6.9	129	395	15.0	75	15	9	2	12	12	20	10	1	5	13.5
C	MU-P	D1, D2			7.1	131	448	16.8	84	17	8	3	14	11	20	10	1	4	11.3
E	MU	D1, D2, P	2 days, 50 mL MM	4 days, 50 mL HGM	4.6	89	314	17.3	100	16	12	2	18	9	22	7	1	4	11.2
E	MU-D2	D1, P			2.8	62	109	8.5	49	14	12	1	6	8	25	6	0	7	20.0
E	MU-P	D2, P			5.0	99	232	11.5	66	16	10	2	14	7	24	7	1	5	15.8
E	MU-D2-P	D1			4.2	98	98	4.9	28	18	10	0	7	12	20	5	0	6	22.5

The data showed that the lipid content within the transformed cells varied according to the growth conditions. Furthermore, the contribution of each acyltransferase on lipid content also varied. Specifically, in Experiments B, C and E, DGAT2 contributed more to oil biosynthesis than either PDAT or DGAT1. In contrast, as demonstrated in Experiment A, a single knockout in DGAT2, DGAT1 and PDAT resulted in approximately equivalent losses in lipid content (i.e., 48%, 49% and 42% loss, respectively [see "TFAs % dcw, % WT"]).

EXAMPLE 12

Sequencing Of *Yarrowia lipolytica* DGAT1 And ORF Expression Under The Control Of A *Yarrowia* Promoter

The present Example describes the sequencing of YI DGAT1 and the over-expression of a chimeric gene comprising the *Yarrowia lipolytica* TEF promoter, YI DGAT1, and *Yarrowia lipolytica* peroxin (Pex20) terminator (i.e., a TEF::YI DGAT1::Pex20 gene) in a wild type *Yarrowia* strain.

Sequencing Of The *Y. lipolytica* DGAT1

First, the ORF of *Y. lipolytica* DGAT1 was PCR-amplified using degenerate primers P201 and P203 (SEQ ID NOs:92 and 93) and genomic DNA of *Y. lipolytica* ATCC #90812 as template (from Example 2). The PCR was performed using the Expand High Fidelity PCR System of Roche Applied Sciences (Indianapolis, IN), as described in the General Methods.

The expected 1.6 kB fragment was isolated by standard agarose gel electrophoresis, purified, and cloned into pCR4-TOPO vector from Invitrogen (Carlsbad, CA) to yield plasmid pYAP42-23. Plasmid pYAP42-23 was transformed into *E. coli* XL2; and, transformants comprising pYAP42-23 were confirmed by plasmid miniprep analysis and restriction enzyme digestions with either *NotI* or *NcoI*. The DNA insert in plasmid pYAP42-23 was sequenced according to the methodology described in the General Methods using sequencing primers T7, T3, P239 (SEQ ID NO:142) and P240 (SEQ ID NO:143), to obtain the complete nucleotide sequence of the YI DGAT1 ORF.

The nucleotide sequence of the YI DGAT1 ORF is provided as SEQ ID NO:13; the translated product has the amino acid sequence provided in SEQ ID NO:14. The resultant sequence was compared to other known proteins, based on BLAST program analysis (Basic Local Alignment Search Tool; Altschul, S. F. et al., *J. Mol. Biol.* 215:403-410 (1993)). In

particular, SEQ ID NO:13 was identical to the YI DGAT1 partial sequence that was obtained in Example 7, except for the presence of 6 silent mutations in the region of the degenerate PCR primers. These mutations included: an A-to-G mutation at position 6; an A-to-G mutation at position 21; an A-to-G mutation at position 24; a T-to-C mutation at position 1548; a C-to-T mutation at position 1552; and a T-to-C mutation at position 1557. Since these mutations resulted from the use of degenerate PCR primers, the deduced amino acid sequence of SEQ ID NO:13, i.e., SEQ ID NO:14 is identical to ORF YALI-CDS2141.1 (SEQ ID NO:12, corresponding to GenBank Accession No. NC_006072, locus_tag="YALI0F06578g").

Construction Of A *Y. lipolytica* DGAT1 Chimeric Gene

Plasmid pYAP23-42 was digested with *Nco*I and *Not*I for 1 hr and the 1.6 kB fragment containing YI DGAT1 was isolated and inserted into *Nco*I- and *Not*I-digested pZP2I7+Ura (SEQ ID NO:144; Figure 7DA), such that the ORF was cloned under the control of the TEF promoter and the PEX20-3' terminator region in an integrating vector targeted into the *Yarrowia* POX2 gene. Correct transformants were confirmed by miniprep analysis and the resultant plasmid was designated as "pYDA1".

Plasmids pZP2I7+Ura and pYDA1 were transformed into strain "MU-D2" of *Y. lipolytica* (*supra*, Example 11), according to the General Methods. Transformants were plated onto MM and single colonies were picked, purified, and analyzed to determine the effect of the overexpressed DGAT1 on lipid content. Specifically, lipid content was analyzed in the following cultures: "MU" ("wildtype"), MU-D2 transformed with pZP2I7+Ura, and MU-D2 transformed with pYDA1 (clones #5, 6, 7 and 16). Several loops of cells from each of the strains above were inoculated into 50 mL MM and grown in a shaker at 30 °C for 48 hrs. Cells were harvested, washed once in HGM, resuspended in 30 mL HGM medium, and grown as above for another 4 days. After growth, 100 µL aliquots from each culture were used to determine absorbance at 600 nm (OD₆₀₀) and a 1 mL aliquot was used for GC analysis. For this, the 1 mL sample was harvested, washed once in water, spun down, and the pellet used for lipid determination following direct transesterification by base method and GC analysis (as described in the General Methods). The remaining culture was harvested, washed once in water and lyophilized to obtain the dry cell weight.

The results are shown in the Table below. Cultures are described as the "MU" strain ("wildtype") and "MU-D2" (DGAT2 knockout).

Abbreviations utilized are: "WT" = wildtype (i.e., strain MU-D2 having a DGAT2 knockout); "OD" = optical density; "dcw" = dry cell weight; "TFAs" = total fatty acids; and, "TFAs % dcw, % WT" = TFAs % dcw relative to the wild type ("MU-D2") strain.

Table 17

Lipid Content In *Yarrowia* Strains Engineered To Produce EPA And Overexpressing DGAT1

Strain	OD	dcw, mg	TFAs, μ g	TFAs % dcw	TFAs % dcw, % WT
MU	2.1	33	195	17.0	
MU-D2 + pZP2I7+Ura	1.8	31	82	7.7	100
MU-D2 + pYDA1, clone #5	2.1	31	192	17.8	231
MU-D2 + pYDA1, clone #6	1.3	22	338	21.3	278
MU-D2 + pYDA1, clone #7	4.4	72	146	5.9	76
MU-D2 + pYDA1, clone #16	2.2	34	195	16.9	220

In general, the results showed that overexpression of DGAT1 was able to compensate for the lack of DGAT2 activity in strain MU-D2, to result in lipid content approximately equal or greater than in strain MU (i.e., MU-D2 + pYDA1, clones #5, 6 and 16 have a lipid content (measured as TFAs % dcw) approximately equal to or greater than that of strain MU). This lipid content in MU-D2 + pYDA1, clones #5, 6 and 16 was greater than double that of the control strain, MU-D2 + pZP2I7+Ura. These results provide further confirmation that the *Yarrowia* DGAT1 encodes a functional DAG AT involved in oil biosynthesis.

Transformant MU-D2 + pYDA1, clone #7 did not show an increase in lipid content comparable to clones #5, 6 and 16. However, since these chromosomal integrations are generally random, such variation is to be expected.

EXAMPLE 13

Construction And Sequencing Of A *Mortierella alpina* cDNA Library

The present Example describes the construction of a cDNA library of *Mortierella alpina* and subsequent sequencing of the library.

5 Synthesis of *M. alpina* cDNA

M. alpina cDNA was synthesized using the BD-Clontech Creator Smart[®] cDNA library kit (Mississauga, ON, Canada), according to the manufacturer's protocol.

Specifically, *M. alpina* strain ATCC #16266 was grown in 60 mL
10 YPD medium (2% Bacto-yeast extract, 3% Bactor-peptone, 2% glucose) for 3 days at 23 °C. Cells were pelleted by centrifugation at 3750 rpm in a Beckman GH3.8 rotor for 10 min and resuspended in 6X 0.6 mL Trizole reagent (Invitrogen). Resuspended cells were transferred to six 2 mL screw cap tubes each containing 0.6 mL of 0.5 mm glass beads. The cells
15 were homogenized at the HOMOGENIZE setting on a Biospec (Bartlesville, OK) mini bead beater for 2 min. The tubes were briefly spun to settle the beads. Liquid was transferred to 4 fresh 1.5 mL microfuge tubes and 0.2 mL chloroform/isoamyl alcohol (24:1) was added to each tube. The tubes were shaken by hand for 1 min and let stand for 3 min.
20 The tubes were then spun at 14,000 rpm for 10 min at 4 °C. The upper layer was transferred to 4 new tubes. Isopropyl alcohol (0.5 mL) was added to each tube. Tubes were incubated at room temperature for 15 min, followed by centrifugation at 14,000 rpm and 4 °C for 10 min. The pellets were washed with 1 mL each of 75% ethanol, made with RNase
25 free water and air-dried. The total RNA sample was then redissolved in 500 µl of water, and the amount of RNA was measured by A260 nm using 1:50 diluted RNA sample. A total of 3.14 mg RNA was obtained.

This total RNA sample was further purified with the Qiagen RNeasy total RNA Midi kit following the manufacturer's protocol. Thus, the total
30 RNA sample was diluted to 2 mL and mixed with 8 mL of buffer RLT with 80 µl of β-mercaptoethanol and 5.6 mL 100% ethanol. The sample was divided into 4 portions and loaded onto 4 RNeasy midid columns. The columns were then centrifuged for 5 min at 4500Xg. To wash the columns, 2 mL of buffer RPE was loaded and the columns centrifuged for
35 2 min at 4500Xg. The washing step was repeated once, except that the centrifugation time was extended to 5 min. Total RNA was eluted by

applying 250 µl of RNase free water to each column, waiting for 1 min and centrifuging at 4500Xg for 3 min.

PolyA(+)RNA was then isolated from the above total RNA sample, following Pharmacia's kit protocol. Briefly, 2 oligo-dT-cellulose columns
5 were used. The columns were washed twice with 1 mL each of high salt buffer. The total RNA sample from the previous step was diluted to 2 mL total volume and adjusted to 10 mM Tris/HCl, pH 8.0, 1 mM EDTA. The sample was heated at 65 °C for 5 min, then placed on ice. Sample buffer (0.4 mL) was added and the sample was then loaded onto the two oligo-
10 dT-cellulose columns under gravity feed. The columns were centrifuged at 350Xg for 2 min, washed 2X with 0.25 mL each of high salt buffer, each time followed by centrifugation at 350Xg for 2 min. The columns were further washed 3 times with low salt buffer, following the same centrifugation routine. Poly(A)+RNA was eluted by washing the column 4
15 times with 0.25 mL each of elution buffer preheated to 65 °C, followed by the same centrifugation procedure. The entire purification process was repeated once. Purified poly(A)+RNA was obtained with a concentration of 30.4 ng/µl.

cDNA was generated, using the LD-PCR method specified by BD-
20 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:145) and 1 µl of CDSIII/3' PCR primer (SEQ ID NO:146). 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 first
25 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.

The 1st strand cDNA synthesis mixture was used as template for the PCR reaction. Specifically, the reaction mixture contained the
30 following: 2 µl of the 1st strand cDNA mixture, 2 µl 5'-PCR primer (SEQ ID NO:147), 2 µl CDSIII/3'-PCR primer (SEQ ID NO:146), 80 µl water, 10 µl 10X Advantage 2 PCR buffer, 2 µl 50X dNTP mix and 2 µl 50X 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
35 GenAmp 9600 instrument. PCR product was quantitated by agarose gel electrophoresis and ethidium bromide staining.

Seventy-five μ l of the above PCR products (cDNA) were mixed with 3 μ l of 20 μ g/ μ l proteinase K supplied with the kit. The mixture was incubated at 45 °C for 20 min, then 75 μ l of water was added and the mixture was extracted with 150 μ l phenol:chloroform:isoamyl alcohol mixture (25:24:1). The aqueous phase was further extracted with 150 μ l chloroform:isoamyl alcohol (25:1). The aqueous phase was then mixed with 15 μ l of 3 M sodium acetate, 2 μ l of 20 μ g/ μ l glycogen and 400 μ l 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 μ l of 80% ethanol, air dried and dissolved in 79 μ l of water.

Dissolved cDNA was subsequently digested with *Sfi*I (79 μ l of the cDNA was mixed with 10 μ l of 10X *Sfi*I buffer, 10 μ l of *Sfi*I enzyme and 1 μ l of 100X BSA and the mixture was incubated at 50 °C for 2 hrs). Xylene cyanol dye (2 μ l 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 μ l of water, and ligated into kit-supplied pDNR-LIB.

Library Sequencing

The ligation products were used to transform *E. coli* XL-1 Blue electroporation competent cells (Stratagene). An estimated total of 2×10^6 colonies was obtained. Sequencing of the cDNA library was carried out by Agencourt Bioscience Corporation (Beverly, MA), using an M13 forward primer (SEQ ID NO:148).

EXAMPLE 14

Identification And Cloning Of A *Mortierella alpina* Diacylglycerol

Acyltransferase (DGAT1) Gene

The present Example describes the identification of a putative *M. alpina* DGAT1 within one of 9,984 cDNA sequences. Specifically, the *Y. lipolytica* DGAT1 protein sequence (Example 7, SEQ ID NO:14) was used as a query sequence against each of the *M. alpina* cDNA sequences using BLAST program analysis (Basic Local Alignment Search Tool; Altschul, S. F. et al., *J. Mol. Biol.* 215:403-410 (1993)). One cDNA fragment bore significant homology to the *Y. lipolytica* DGAT1 and thus was tentatively

identified as the *M. alpina* DGAT1 (SEQ ID NO:175). Subsequent BLAST analyses with SEQ ID NO:175 as the query against publically available sequence databases confirmed the cDNA's significant degree of similarity with the DGAT1s from several other species. Rapid amplification of cDNA ends (RACE) technology and genome walking were then used to isolate the entire *Mortierella alpina* coding sequence thereof.

Isolation Of Genomic DNA

Genomic DNA was isolated from *Mortierella alpina* (ATCC #16266) using a QiaPrep Spin Miniprep Kit (Qiagen, Catalog #627106). Cells grown on a YPD agar plate were scraped off and resuspended in 1.2 mL kit buffer P1. The resuspended cells were placed into two 2.0 mL screw cap tubes, each containing 0.6 mL glass beads (0.5 mm diameter). The cells were homogenized at the HOMOGENIZE setting on a Biospec (Bartlesville, OK) mini bead beater for 2 min. The tubes were then centrifuged at 14,000 rpm in an Eppendorf microfuge for 2 min. The supernatant (0.75 mL) was transferred to three 1.5 mL microfuge tubes. Equal volumes of kit buffer P2 were added to each tube. After mixing the tubes three times by inversion, 0.35 mL of buffer N3 was added to each tube. The contents was mixed again by inverting the tubes 5 times. The mixture was centrifuged at 14,000 rpm in an Eppendorf microfuge for 5 min. The supernatant from each tube was transferred into a separate kit spin column. The columns were centrifuged for 1 min at 14,000 rpm, washed once with buffer PE and centrifuged at 14,000 rpm for 1 min again, followed by a final centrifugation at 14,000 rpm for 1 min. Buffer EB (50 µl) was added to each column and let stand for 1 min. The genomic DNA was then eluted by centrifugation at 14,000 rpm for 1 min.

Cloning Of The 5'-End Region Of The Putative DGAT1 Gene

A Clontech Universal GenomeWalker™ kit (Palo Alto, CA) was utilized to obtain a piece of genomic DNA corresponding to the 5'-end region of the *M. alpina* DGAT1. Based on the partial DGAT1 gene sequence available (SEQ ID NO:149), the following primers were synthesized for use in the cloning: MARE2-N1 and MARE2-N2 (SEQ ID NOs:150 and 151).

Briefly, 2.5 µg each of *M. alpina* genomic DNA was digested with *DraI*, *EcoRV*, *PvuII* or *StuI* individually, the digested DNA samples were purified using Qiagen Qiaquick PCR purification kits and eluted with 30 µl each of kit buffer EB,

5



15

20

25

30

35

35

103

polyA(+)RNA sample in 11 μ l of water were mixed with 1 μ l of 10 μ M Adaptor primer (AP, SEQ ID NO:159) solution. The mixture was heated at 70 °C for 10 min and cooled on ice for 2 min. Then, 2 μ l each of 10X PCR buffer, 25 mM MgCl₂ and 0.1 M DTT, and 1 μ l of 10 mM dNTP mix were added. The reaction mixture was heated to 42 °C for 3 min and then kit-supplied Superscript II reverse transcriptase (1 μ l) was added. The reaction was allowed to proceed for 50 min at 42 °C. Afterward, the reaction mixture was heated to 70 °C for 15 min and cooled on ice for 2 min. RNaseH (1 μ l) from the kit was added and the entire mixture was incubated at 37 °C for 20 min.

The reaction mixture (2 μ l) was used directly as PCR template. The PCR reaction mixture contained 1 μ l of 20 μ M ARE-N3-1 (SEQ ID NO:157), 2 μ l of 10 μ M kit primer UAP (SEQ ID NO:160), 25 μ l of ExTaq premix Taq 2X PCR solution (TaKaRa) and 20 μ l of water. PCR amplification was performed as previously described above. Diluted PCR reaction mixture (1 μ l of a 1:10 dilution) was then used as template for a second round of PCR using the same conditions, except with primer ARE-N3-2 (SEQ ID NO:158) replacing primer AREN3-1.

A ca. 300 bp fragment was obtained from the PCR. After purification with Qiagen's QiaQuick PCR purification kit, the fragment was cloned into pCR2.1-TOPO and sequenced. Sequence analysis verified that the sequence encoded the 3'-end of the DGAT1 cDNA, including the polyA tail (SEQ ID NO:161).

Complete Assembly Of The Nucleotide Sequence Encoding *M. alpina*'s DGAT1

Assembly of the sequence of the 5'-region (SEQ ID NO:156), the original cDNA fragment (SEQ ID NO:149) and the 3'-region (SEQ ID NO:161) yielded the entire *M. alpina* DGAT1 coding sequence (SEQ ID NO:17). The 5'-region genomic sequence included an intron (nucleotide bases 449 to 845 within SEQ ID NO:17).

EXAMPLE 15

Expression of *Mortierella alpina* DGAT1 In *Yarrowia lipolytica* Strain Y2067U Engineered To Produce Polyunsaturated Fatty Acids

The present Example describes the expression of the *M. alpina* DGAT1 in *Y. lipolytica* strain Y2067U (*supra*, Example 10), and the effect of MDGAT1 expression on the final concentration of EPA and other PUFAs produced.

The *M. alpina* DGAT1 ORF was cloned as follows. First, to aid the cloning of the cDNA, the sequence of the second codon of the DGAT1 was changed from 'ACA' to 'GCA', resulting in an amino acid change of threonine to alanine. This was accomplished by amplifying the complete coding region of the *M. alpina* DGAT1 ORF with primers MACAT-F1 and MACAT-R (SEQ ID NOs:162 and 163, respectively). Specifically, the PCR reaction mixture contained 1 µl each of a 20 µM solution of primers MACAT-F1 and MACAT-R, 1 µl of *M. alpina* cDNA (*supra*, Example 13), 22 µl water and 25 µl ExTaq premix 2X Taq PCR solution (TaKaRa Bio Inc., Otsu, Shiga, 520-2193, Japan). Amplification was carried out as follows: initial denaturation at 94 °C for 150 sec, followed by 30 cycles of denaturation at 94 °C for 30 sec, annealing at 55 °C for 30 sec and elongation at 72 °C for 90 sec. A final elongation cycle at 72 °C for 10 min was carried out, followed by reaction termination at 4 °C. A ~1600 bp DNA fragment was obtained from the PCR reaction. It was purified using Qiagen's PCR purification kit according to the manufacturer's protocol.

The *M. alpina* DGAT1 ORF was to be inserted into *Nco* I- and *Not* I-digested plasmid pZUF17 (SEQ ID NO:164; Figure 7E), such that the ORF was cloned under the control of the FBAIN promoter (SEQ ID NO:111) and the PEX20-3' terminator region. However, since the DGAT1 ORF contained an internal *Nco* I site, it was necessary to perform two separate restriction enzyme digestions for cloning. First, ~2 µg of the purified PCR product was digested with *Bam*HI and *Nco* I. The reaction mixture contained 20 U of each enzyme (Promega) and 6 µl of restriction buffer D in a total volume of 60 µl. The mixture was incubated for 2 hrs at 37 °C. A ~320 bp fragment was separated by agarose gel electrophoresis and purified using a Qiagen Qiaex II gel purification kit. Separately, ~2 µg of the purified PCR product was digested with *Bam*HI and *Not* I using identical reaction conditions to those above, except *Nco* I was replaced by *Not* I. A ~1280 bp fragment was isolated and purified as above. Finally, ~3 µg of pZUF17 was digested with *Nco* I and *Not* I and purified as described above, generating a ~7 kB fragment.

The ~7 kB *Nco* I/*Not* I pZUF17 fragment, the ~320 bp *Nco* I/*Bam*HI DGAT1 fragment and the ~1280 bp *Bam*HI/*Not* I DGAT1 fragment were ligated together in a three-way ligation incubated at room temperature overnight. The ligation mixture contained 100 ng of the 7 kB fragment and 200 ng each of the 320 bp and 1280 bp fragments, 2 µl ligase buffer, and

2 U T4 DNA ligase (Promega) in a total volume of 20 μ l. The ligation products were used to transform *E. coli* Top10 chemical competent cells (Invitrogen) according to the manufacturer's protocol.

Individual colonies (12 total) from the transformation were used to inoculate cultures for miniprep analysis. Restriction mapping and sequencing showed that 5 out of the 12 colonies harbored the desired plasmid, which was named "pMDGAT1-17" (Figure 4C; SEQ ID NO:165).

"Control" vector pZUF-MOD-1 (SEQ ID NO:168) was prepared as follows. First, primers pzuf-mod1 (SEQ ID NO:166) and pzuf-mod2 (SEQ ID NO:167) were used to amplify a 252 bp "stuffer" DNA fragment using pDNR-LIB (ClonTech, Palo Alto, CA) as template. The amplified fragment was purified with a Qiagen QiaQuick PCR purification kit, digested with *Nco*I and *Not*I using standard conditions, and then purified again with a QiaQuick PCR purification kit. This fragment was ligated into similarly digested *Nco*I-/ *Not*I-cut pZUF17 vector (SEQ ID NO:164; Figure 10E) and the resulting ligation mixture was used to transform *E. coli* Top10 cells (Invitrogen). Plasmid DNA was purified from 4 resulting colonies, using a Qiagen QiaPrep Spin Miniprep kit. The purified plasmids were digested with *Nco*I and *Not*I to confirm the presence of the ~250 bp fragment. The resulting plasmid was named "pZUF-MOD-1" (Figure 4D; SEQ ID NO:168).

Y. lipolytica strain Y2067U (from Example 10, producing 14% EPA of total lipids) was transformed with pMDGAT1-17 and pZUF-MOD-1, respectively, according to the General Methods. Transformants were grown for 2 days in synthetic MM supplemented with amino acids, followed by 4 days in HGM. The fatty acid profile of two transformants containing pMDGAT1-17 and two transformants containing pZUF-MOD-1 are shown below in Table 18, based on GC analysis (as described in the General Methods). Fatty acids are identified as 18:0, 18:1 (oleic acid), 18:2 (LA), GLA, DGLA, ARA, ETA and EPA; and the composition of each is presented as a % of the total fatty acids.

Table 18

Lipid Content In *Yarrowia* Strain Y2067U Engineered To Overexpress *M. alpina* DGAT1

	Total Fatty Acids
--	-------------------

Strain	18:0	18:1	18:2	GLA	DGLA	ARA	ETA	EPA
Y2067U + pZUF-MOD-1 #1	1.31	6.92	12.03	23.11	5.72	1.05	3.80	13.20
Y2067U + pZUF-MOD-1 #2	1.39	6.83	12.15	21.99	5.83	1.07	3.82	13.47
Y2067U + pMDGAT1-17 #1	0.89	7.13	10.87	24.88	5.82	1.19	3.97	14.09
Y2067U + pMDGAT1-17 #2	0.86	7.20	10.25	22.42	6.35	1.26	4.38	15.07

As demonstrated above, expression of the *M. alpina* DGAT1 from plasmid pMDGAT1-17 increased the EPA concentration from ~13.3% in the “control” strains to ~14.1% (“Y2067U + pMDGAT1-17 #1”) and ~15.1%
 5 (“Y2067U + pMDGAT1-17 #2”), respectively.

EXAMPLE 16

Identification Of DGAT1 Fungal Homologs

The present Example describes the use of the *Yarrowia lipolytica* and *Mortierella alpina* DGAT1 sequences (SEQ ID NOs:13 and 17,
 10 respectively) to identify orthologous proteins in other fungi.

Orthologous DGAT1 fungal proteins were identified 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
 15 Data Bank, the SWISS-PROT protein sequence database, EMBL and DDBJ databases). The *Yarrowia lipolytica* and *Mortierella alpina* DGAT1 sequences (SEQ ID NOs:13 and 17) were analyzed for similarity to all publicly available DNA sequences contained in the “nr” database using the BLASTN algorithm provided by the National Center for Biotechnology
 20 Information (NCBI). The DNA sequences were 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. These searches resulted in the identification of 4
 25 orthologous proteins (as shown below in Table 19). Table 19 additionally shows the results of sequence comparisons between the *Yarrowia lipolytica* DGAT1 sequence (SEQ ID NO:13) with each of the DGAT1 proteins disclosed herein, in terms of the observed “% Ident.” (defined as the percentage of amino acids that are identical between the two proteins).

30

Table 19
Comparison Of *Yarrowia lipolytica* DGAT1 To Orthologous DGAT1s From
Fungi

Organism	Abbreviation	% Ident	GenBank Accession No., Reference and Annotation	SEQ ID NO
<i>Mortierella alpina</i>	Ma DGAT1	32.4	---	18
<i>Neurospora crassa</i> strain OR74A	Nc DAGAT1	37.0	XP_322121; gi 32403016 ref XP_322121.1 hypothetical protein; gi 28918105 gb EAA27786.1 hypothetical protein	19
<i>Gibberella zeae</i> PH-1	Fm DAGAT1	38.1	EAA77624; gi 42554781 gb EAA77624.1 hypothetical protein FG06688.1	20
<i>Magnaporthe grisea</i> 70-15	Mg DAGAT1	36.2	EAA52634; gi 38106308 gb EAA52634.1 hypothetical protein MG05326.4	21
<i>Aspergillus nidulans</i> FGSC A4	An DAGAT1	41.7	EAA57945; gi 40738755 gb EAA57945.1 hypothetical protein AN6159.2	22

5

EXAMPLE 17

Identification Of Universal And Fungal DGAT1 Motifs

The present Example describes the use of the DGAT1 sequences of the present invention, in conjunction with other known DGAT1 sequences, to identify fungal and universal DGAT1 motifs.

To identify motifs (i.e., a set of amino acids conserved at specific positions along an aligned sequence of evolutionarily related proteins) that are indicative of a DGAT1 protein, it was first necessary to generate an alignment of DGAT1 sequences. For this, the following fungal sequences were used: SEQ ID NOs:14, 18, 19, 20, 21 and 22. Additionally, DGAT1 orthologs from 6 non-fungal sources were also included in the comparative alignment: mouse (Mm DGAT1; GenBank Accession No. AF384160, corresponding to SEQ ID NO:169 herein); soy (Gm DGAT1; SEQ ID NO:16 of US20040088759A1, corresponding to SEQ ID NO:170 herein);

Arabidopsis (At DGAT1; SEQ ID NO:2 of US20040088759A1, corresponding to SEQ ID NO:171 herein); rice (Os DGAT1; SEQ ID NO:14 of US20040088759A1, corresponding to SEQ ID NO:172 herein); wheat (Ta DGAT1; SEQ ID NO:22 of US20040088759A1, corresponding to SEQ ID NO:174 herein); and *Perilla frutescens* (Pf; GenBank Accession No. AF298815, corresponding to SEQ ID NO:173 herein).

Alignment was done using the Megalign program of DNASTAR using Clustal W with the following parameters: gap penalty =10, gap length penalty =0.2, delay divergent seqs (%) =30, DNA transition weight =0.5 and protein weight matrix by Gonnet series. The results of this alignment are shown in Figure 8a, 8b, 8c, 8d, 8e, 8f, 8g and 8h. Based on analysis of the alignment, 8 motifs were identified as unique to fungal DGAT1 sequences. Additionally, 7 motifs that were universally present in DGAT1 sequences from plants, animals and fungi were also deduced.

Table 20

Fungal and Universal DGAT1 Motifs

Motif	Align- ment Position*	Fungal Motif Sequence and SEQ ID NO	Universal Motif Sequence and SEQ ID NO
# 1	97-104	<u>(F/Y)</u> xGFxN(L/I) <u>(M/G)</u> (SEQ ID NO:23)	xxGxxNxx (SEQ ID NO:31)
# 2	278-284	<u>(P/Q)</u> YPxN(I/V)T (SEQ ID NO:24)	n/a
# 3	334-340	QY <u>A</u> xPx <u>(L/M)</u> (SEQ ID NO:25)	Q(Y/W)xxPxx (SEQ ID NO:32)
# 4	364-374	KL(A/S) <u>(T/S)</u> x ₁ <u>S</u> XX(I/V)WL (wherein x ₁ can not be P) (SEQ ID NO:26)	KL(A/S)xxxxxxWL (SEQ ID NO:33)
# 5	418-424	PV <u>(Y/N)</u> <u>(Q/T/I)</u> <u>(Y/F)</u> <u>(F/M)</u> <u>(K/R)</u> (SEQ ID NO:27)	PVxxxxx (SEQ ID NO:34)
# 6	415-424	WN(K/R/S)PV <u>(Y/N)</u> x ₁ <u>(Y/F)</u> <u>(F/M)</u> <u>(K/R)</u> (wherein x ₁ can not be K) (SEQ ID NO:28)	WNxPVxxxxx (SEQ ID NO:35)

# 7	456-466	<u>L</u>x<u>G</u>xPT<u>H</u>xx<u>(I/Y)</u><u>G</u>	xxxxPxxxxxx (SEQ ID NO:36)
# 8	513-519	A (L/F)(L/M)Y <u>(F/Y)</u> X <u>(A/H)</u> (SEQ ID NO:30)	x(L/F)(L/M)Yxxx (SEQ ID NO:37)

*[Note: Alignment positions are with respect to that of the *Yarrowia lipolytica* DGAT1, herein identified as SEQ ID NO:14. Those residues shown in bold-type face and underlined are conserved only in fungal DGAT1 sequences.]

5

These motifs, located at positions 97-104, 278-284, 334-340, 364-374, 418-424, 415-424, 456-466 and 513-519 (wherein the alignment positions are with respect to SEQ ID NO:14) in a sequence alignment of a family of protein homologues, have a high degree of conservation among DGAT1 proteins; as such, it is expected that the amino acids residues located therein are essential in the structure, the stability, or the activity of the protein. Based on the sequence conservation observed, one skilled in the art will know how to use the motifs provided as SEQ ID NOs:23-37 as an identifier, or "signature", to determine if a protein with a newly determined sequence belongs to the DGAT1 protein family described herein.

10

15

DEMANDE OU BREVET VOLUMINEUX

LA PRÉSENTE PARTIE DE CETTE DEMANDE OU CE BREVET COMPREND PLUS D'UN TOME.

CECI EST LE TOME 1 DE 2
CONTENANT LES PAGES 1 À 110

NOTE : Pour les tomes additionels, veuillez contacter le Bureau canadien des brevets

JUMBO APPLICATIONS/PATENTS

THIS SECTION OF THE APPLICATION/PATENT CONTAINS MORE THAN ONE VOLUME

THIS IS VOLUME 1 OF 2
CONTAINING PAGES 1 TO 110

NOTE: For additional volumes, please contact the Canadian Patent Office

NOM DU FICHER / FILE NAME :

NOTE POUR LE TOME / VOLUME NOTE:

CLAIMS

What is claimed is:

- 5 1. An isolated nucleic acid molecule encoding a diacylglycerol acyltransferase-1 enzyme, selected from the group consisting of:
 - (a) an isolated nucleic acid molecule encoding the amino acid sequence selected from the group consisting of SEQ ID NOs:14 and 18;
 - 10 (b) an isolated nucleic acid molecule that hybridizes with (a) under the following hybridization conditions: 0.1X SSC, 0.1% SDS, 65 °C and washed with 2X SSC, 0.1% SDS followed by 0.1X SSC, 0.1% SDS; or
 - 15 (c) an isolated nucleic acid molecule that is completely complementary to (a) or (b).
2. The isolated nucleic acid molecule of Claim 1 selected from the group consisting of SEQ ID NOs:13 and 17.
3. A polypeptide encoded by the isolated nucleic acid molecule of Claim 1.
- 20 4. A polypeptide encoded by the isolated nucleic acid molecule of Claim 2.
5. An isolated nucleic acid molecule comprising a first nucleotide sequence encoding a diacylglycerol acyltransferase-1 enzyme of at least 526 amino acids that has at least 70% identity based on the BLASTP method of alignment when compared to a polypeptide having the sequence as set forth in SEQ ID NO:14;
 - 25 or a second nucleotide sequence comprising the complement of the first nucleotide sequence.
6. An isolated nucleic acid molecule comprising a first nucleotide sequence encoding a diacylglycerol acyltransferase-1 enzyme of at least 525 amino acids that has at least 70% identity based on the BLASTP method of alignment when compared to a polypeptide having the sequence as set forth in SEQ ID NO:18;
 - 30 or a second nucleotide sequence comprising the complement of the first nucleotide sequence.
 - 35
7. An isolated nucleic acid molecule encoding an acyl-CoA:sterol-acyltransferase, selected from the group consisting of:

- (a) an isolated nucleic acid molecule encoding the amino acid sequence as set forth in SEQ ID NO:16;
- (b) an isolated nucleic acid molecule that hybridizes with (a) under the following hybridization conditions: 0.1X SSC, 0.1% SDS, 65 °C and washed with 2X SSC, 0.1% SDS followed by 0.1X SSC, 0.1% SDS; or
- (c) an isolated nucleic acid molecule that is completely complementary to (a) or (b).
8. The isolated nucleic acid molecule of Claim 7 selected from the group consisting of SEQ ID NO:15.
9. A polypeptide encoded by the isolated nucleic acid molecule of Claim 7.
10. A polypeptide encoded by the isolated nucleic acid molecule of Claim 8.
11. An isolated nucleic acid molecule comprising a first nucleotide sequence encoding an acyl-CoA:sterol-acyltransferase enzyme of at least 543 amino acids that has at least 70% identity based on the BLASTP method of alignment when compared to a polypeptide having the sequence as set forth in SEQ ID NO:16;
- or a second nucleotide sequence comprising the complement of the first nucleotide sequence.
12. A chimeric gene comprising the isolated nucleic acid molecule of either Claim 1 or Claim 7 operably linked to suitable regulatory sequences.
13. A transformed host cell comprising the chimeric gene of Claim 12.
14. The transformed host cell of Claim 13, selected from the group consisting of algae, bacteria, molds, fungi and yeasts.
15. The transformed host cell of Claim 14, wherein the yeast is an oleaginous yeast.
16. The transformed host cell of Claim 15, wherein the oleaginous yeast cell is selected from the group consisting of *Yarrowia*, *Candida*, *Rhodotorula*, *Rhodospiridium*, *Cryptococcus*, *Trichosporon* and *Lipomyces*.
17. The transformed host cell of Claim 16, wherein the host cell is *Yarrowia lipolytica*.

18. The transformed host cell of Claim 17, 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, *Yarrowia lipolytica* ATCC #90812 and *Yarrowia lipolytica* LGAM S(7)1

19. A method of obtaining a nucleic acid molecule encoding a diacylglycerol acyltransferase-1 enzyme comprising:

- a) probing a genomic library with the nucleic acid molecule of Claim 1;
- b) identifying a DNA clone that hybridizes with the nucleic acid molecule of Claim 1; and
- c) sequencing the genomic fragment that comprises the clone identified in step (b);

wherein the sequenced genomic fragment encodes a diacylglycerol acyltransferase-1 enzyme.

20. A method of obtaining a nucleic acid molecule encoding a diacylglycerol acyltransferase-1 enzyme comprising:

- a) synthesizing at least one oligonucleotide primer corresponding to a portion of the sequence selected from the group consisting of SEQ ID NOs:13 and 17; and
- b) amplifying an insert present in a cloning vector using the oligonucleotide primer of step (a);

wherein the amplified insert encodes a portion of an amino acid sequence encoding a diacylglycerol acyltransferase 1.

21. The product of the method of Claims 19 or 20.

22. An isolated nucleic acid molecule encoding an amino acid motif selected from the group consisting of:

- a) SEQ ID NO:31;
- b) SEQ ID NO:32;
- c) SEQ ID NO:33;
- d) SEQ ID NO:34;
- e) SEQ ID NO:35;
- f) SEQ ID NO:36;
- g) SEQ ID NO:37;
- h) SEQ ID NO:23;
- i) SEQ ID NO:24;
- j) SEQ ID NO:25;

- 5 k) SEQ ID NO:26;
 l) SEQ ID NO:27;
 m) SEQ ID NO:28;
 n) SEQ ID NO:29; and
 o) SEQ ID NO:30.

23. An amino acid motif sequence selected from the group consisting of:

- 10 a) SEQ ID NO:31;
 b) SEQ ID NO:32;
 c) SEQ ID NO:33;
 d) SEQ ID NO:34;
 e) SEQ ID NO:35;
 f) SEQ ID NO:36;
 g) SEQ ID NO:37;
 15 h) SEQ ID NO:23;
 i) SEQ ID NO:24;
 j) SEQ ID NO:25;
 k) SEQ ID NO:26;
 l) SEQ ID NO:27;
 20 m) SEQ ID NO:28;
 n) SEQ ID NO:29; and
 o) SEQ ID NO:30.

24. A method of increasing triacylglycerol content in a transformed host cell comprising:

- 25 (a) providing a transformed host cell comprising:
 (i) at least one gene encoding a diacylglycerol
 acyltransferase-1 enzyme having the amino acid
 sequence selected from the group consisting of SEQ ID
 30 NOs:14, 18, 19, 20, 21 and 22 under the control of
 suitable regulatory sequences; and,
 (ii) a source of fatty acids;
 (b) growing the cell of step (a) under conditions whereby the at
 least one gene encoding a diacylglycerol acyltransferase
 1 enzyme is expressed, resulting in the transfer of the fatty acids
 35 to triacylglycerol; and
 (c) optionally recovering the triacylglycerol of step (b).

25. A method of increasing the ω -3 or ω -6 fatty acid content of triacylglycerols in a transformed host cell comprising:

- (a) providing a transformed host cell comprising:
 - (i) genes encoding a functional ω -3/ ω -6 fatty acid biosynthetic pathway;
 - (ii) at least one gene encoding a diacylglycerol acyltransferase 1 enzyme having the amino acid sequence selected from the group consisting of SEQ ID NOs:14, 18, 19, 20, 21 and 22 under the control of suitable regulatory sequences;
- (b) growing the cell of step (a) under conditions whereby the genes of (i) and (ii) are expressed, resulting in the production of at least one ω -3 or ω -6 fatty acid and its transfer to triacylglycerol; and
- (c) optionally recovering the triacylglycerol of step (b).

26. A method of increasing triacylglycerol content in a transformed host cell comprising:

- (a) providing a transformed host cell comprising:
 - (i) at least one gene encoding a diacylglycerol acyltransferase-1 enzyme comprising all of the amino acid motifs as set forth in:
 - 1) SEQ ID NO:31;
 - 2) SEQ ID NO:32;
 - 3) SEQ ID NO:33;
 - 4) SEQ ID NO:34;
 - 5) SEQ ID NO:35;
 - 6) SEQ ID NO:36; and
 - 7) SEQ ID NO:37;
 - under the control of suitable regulatory sequences; and,
 - (ii) a source of fatty acids;
- (b) growing the cell of step (a) under conditions whereby the at least one gene encoding a diacylglycerol acyltransferase 1 enzyme is expressed, resulting in the transfer of the fatty acids to triacylglycerol; and
- (c) optionally recovering the triacylglycerol of step (b).

27. A method of increasing triacylglycerol content in a transformed host cell comprising:

- (a) providing a transformed host cell comprising:
- (i) at least one gene encoding a diacylglycerol acyltransferase 1 enzyme comprising all of the amino acid motifs as set forth in:
- 5 1) SEQ ID NO:23;
2) SEQ ID NO:24;
3) SEQ ID NO:25;
4) SEQ ID NO:26;
5) SEQ ID NO:27;
10 6) SEQ ID NO:28;
7) SEQ ID NO:29; and
8) SEQ ID NO:30;
- under the control of suitable regulatory sequences; and,
- (ii) a source of fatty acids;
- 15 (b) growing the cell of step (a) under conditions whereby the at least one gene encoding a diacylglycerol acyltransferase-1 enzyme is expressed, resulting in the transfer of the fatty acids to triacylglycerol; and
- (c) optionally recovering the triacylglycerol of step (b).
- 20 28. A method of increasing the ω -3 or ω -6 fatty acid content of triacylglycerols in a transformed host cell comprising:
- (a) providing a transformed host cell comprising:
- (i) genes encoding a functional ω -3/ ω -6 fatty acid biosynthetic pathway; and
- 25 (ii) at least one gene encoding a diacylglycerol acyltransferase 1 enzyme comprising all of the amino acid motifs as set forth in:
- 30 1) SEQ ID NO:31;
2) SEQ ID NO:32;
3) SEQ ID NO:33;
4) SEQ ID NO:34;
5) SEQ ID NO:35;
6) SEQ ID NO:36; and
7) SEQ ID NO:37;
- 35 under the control of suitable regulatory sequences;
- (b) growing the cell of step (a) under conditions whereby the genes of (i) and (ii) are expressed, resulting in the production of at

least one ω -3 or ω -6 fatty acid and its transfer to triacylglycerol;
and

(c) optionally recovering the triacylglycerol of step (b).

29. A method of increasing the ω -3 or ω -6 fatty acid content of
5 triacylglycerols in a transformed host cell comprising:

(a) providing a transformed host cell comprising:

(i) genes encoding a functional ω -3/ ω -6 fatty acid
biosynthetic pathway; and

(ii) at least one gene encoding a diacylglycerol
10 acyltransferase 1 enzyme comprising all of the amino acid
motifs as set forth in:

1) SEQ ID NO:23;

2) SEQ ID NO:24;

3) SEQ ID NO:25;

15 4) SEQ ID NO:26;

5) SEQ ID NO:27;

6) SEQ ID NO:28;

7) SEQ ID NO:29; and

8) SEQ ID NO:30;

20 under the control of suitable regulatory sequences;

(b) growing the cell of step (a) under conditions whereby the genes
of (i) and (ii) are expressed, resulting in the production of at
least one ω -3 or ω -6 fatty acid and its transfer to triacylglycerol;
and

25 (c) optionally recovering the triacylglycerol of step (b).

30. A method according to Claim 25, 28 or 29, wherein the genes
encoding a functional ω -3/ ω -6 fatty acid biosynthetic pathway are
selected from the group consisting of desaturases and elongases.

31. A method according to Claim 30, wherein the desaturase is
30 selected from the group consisting of: Δ 9 desaturase, Δ 12 desaturase, Δ 6
desaturase, Δ 5 desaturase, Δ 17 desaturase, a Δ 8 desaturase, Δ 15
desaturase and Δ 4 desaturase.

32. A method according to any of Claims 24 - 29, wherein the
35 host cell is selected from the group consisting of algae, bacteria, molds,
fungi and yeasts.

33. A method according to Claim 32, wherein the host cell is an oleaginous yeast.

34. A method according to Claim 33 wherein the oleaginous yeast is a member of a genus selected from the group consisting of *Yarrowia*, *Candida*, *Rhodotorula*, *Rhodospiridium*, *Cryptococcus*, *Trichosporon* and *Lipomyces*.

35. A method according to Claim 34, wherein the oleaginous yeast is *Yarrowia lipolytica*.

36. A method according to Claim 35, 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, *Yarrowia lipolytica* ATCC #90812 and *Yarrowia lipolytica* LGAM S(7)1.

37. A method according to any one of Claims 24, 25, 26, 27, 28 or 29, wherein the fatty acid is selected from the group consisting of: stearate, oleic acid, linoleic acid, γ -linoleic acid, dihomogamma-linoleic acid, arachidonic acid, α -linoleic acid, stearidonic acid, eicosatetraenoic acid, eicosapentaenoic acid, docosapentaenoic acid, eicosadienoic acid and eicosatrienoic acid.

38. A method for the identification of a polypeptide having diacylglycerol acyltransferase-1 activity comprising:

- obtaining the amino acid sequence of a polypeptide suspected of having diacylglycerol acyltransferase-1 activity; and,
- identifying, in the amino acid sequence of the polypeptide of step (a), the presence of all of the amino acid motif sequences as set forth in:

- 1) SEQ ID NO:31;
- 2) SEQ ID NO:32;
- 3) SEQ ID NO:33;
- 4) SEQ ID NO:34;
- 5) SEQ ID NO:35;
- 6) SEQ ID NO:36; and
- 7) SEQ ID NO:37;

wherein the presence of all of the motif sequences of step (a) in the polypeptide is indicative of diacylglycerol acyltransferase-1 activity.

39. A method for the identification of a fungal polypeptide having diacylglycerol acyltransferase-1 activity comprising:

- 5 a) obtaining the amino acid sequence of a fungal polypeptide suspected of having diacylglycerol acyltransferase-1 activity; and,
b) identifying, in the amino acid sequence of the polypeptide of step (a), the presence of all of the amino acid motif sequences as set forth in:

- 10 1) SEQ ID NO:23;
2) SEQ ID NO:24;
3) SEQ ID NO:25;
4) SEQ ID NO:26;
5) SEQ ID NO:27;
6) SEQ ID NO:28;
7) SEQ ID NO:29; and
15 8) SEQ ID NO:30;

wherein the presence of all of the motif sequences of step (a) in the polypeptide is indicative of diacylglycerol acyltransferase-1 activity.

1/28

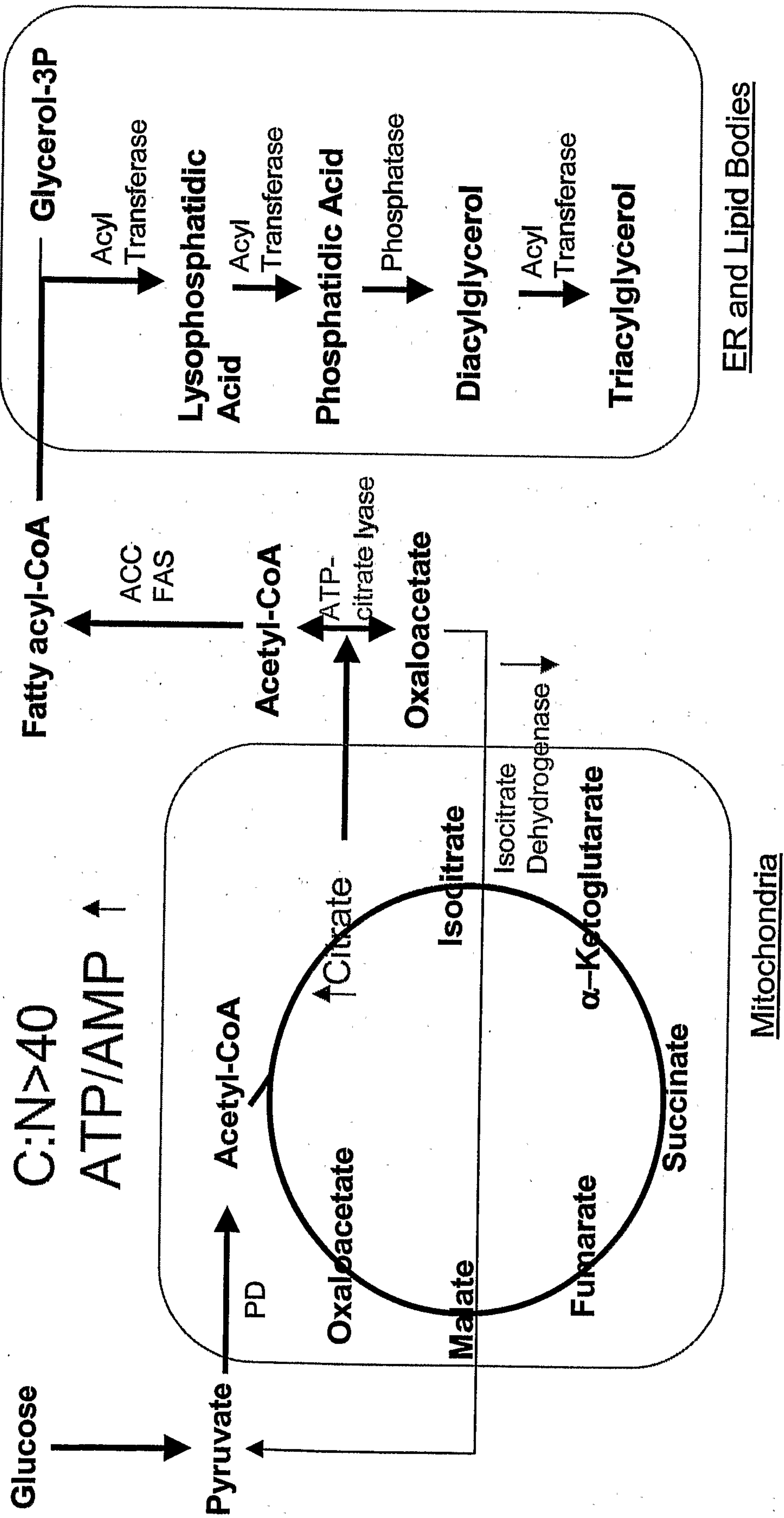


FIG. 1

2/28

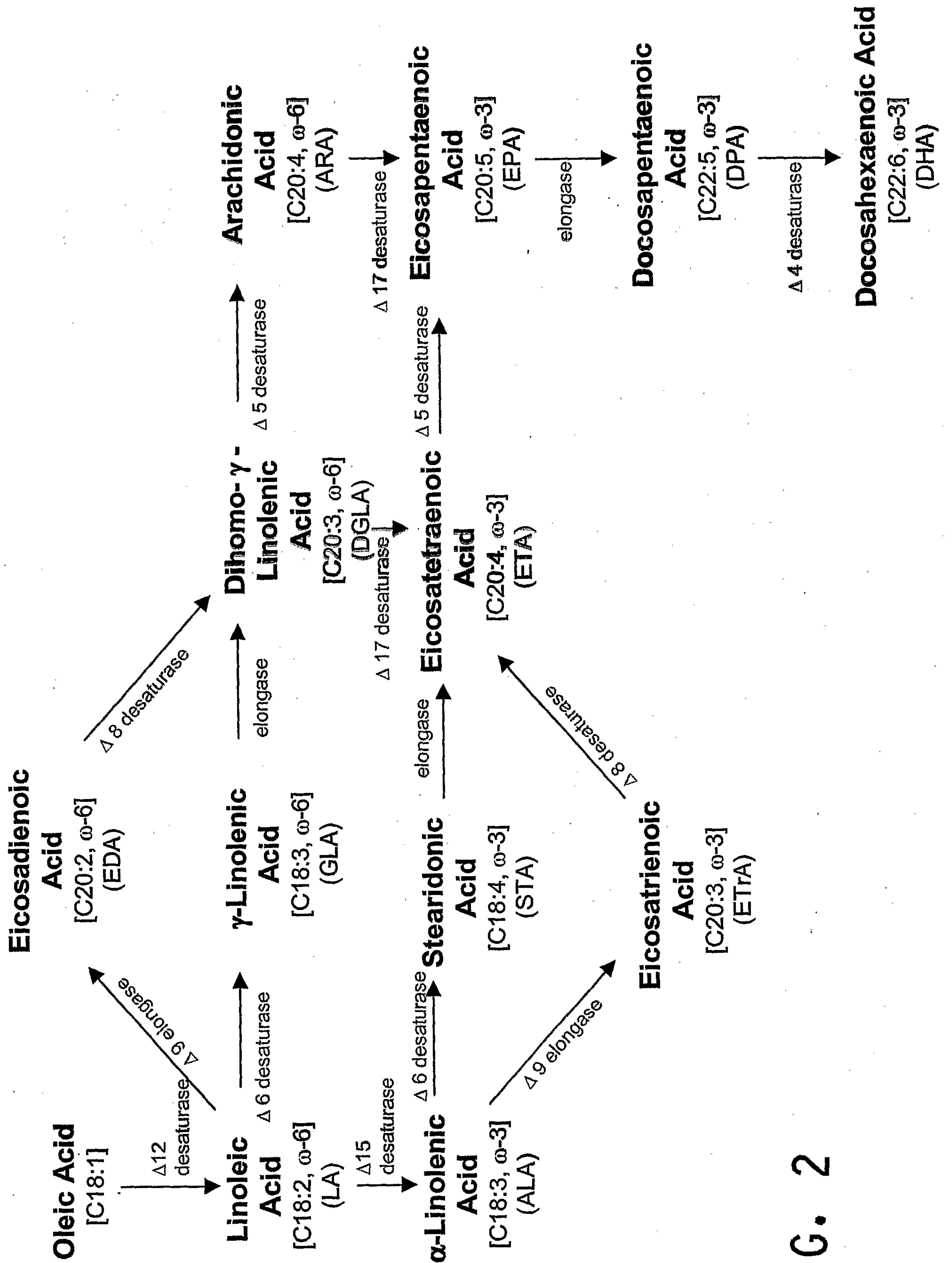


FIG. 2

3/28

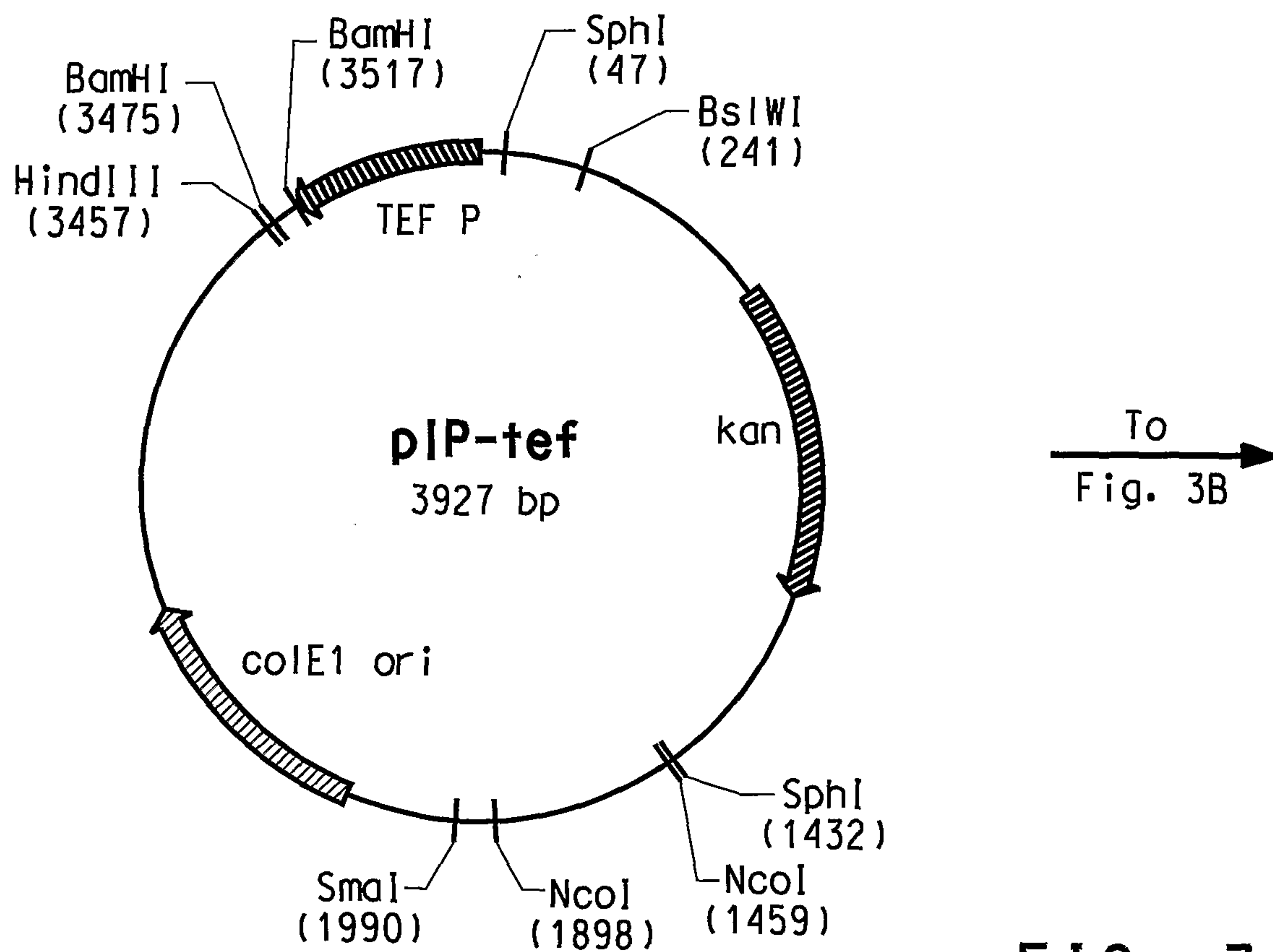
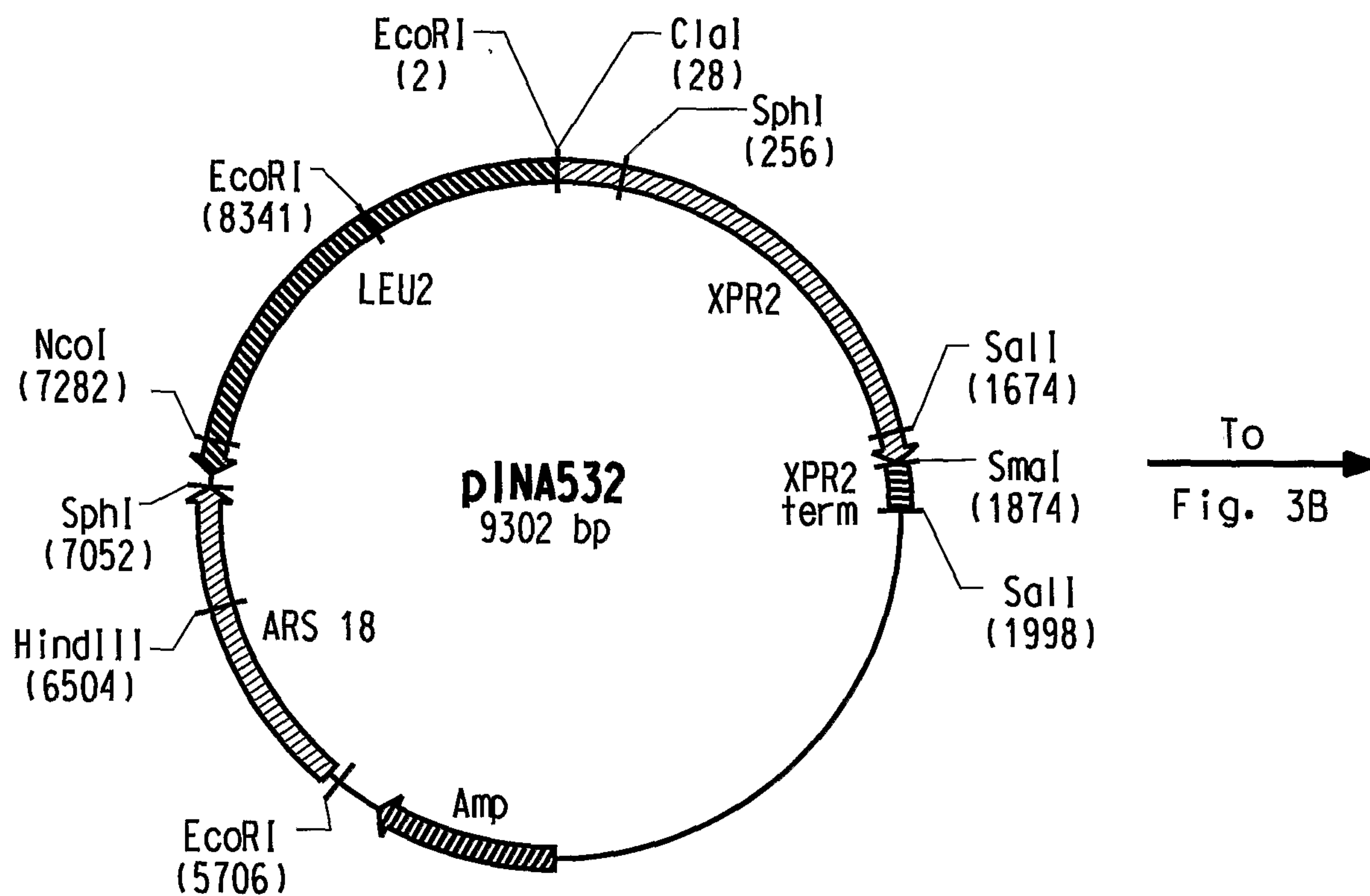
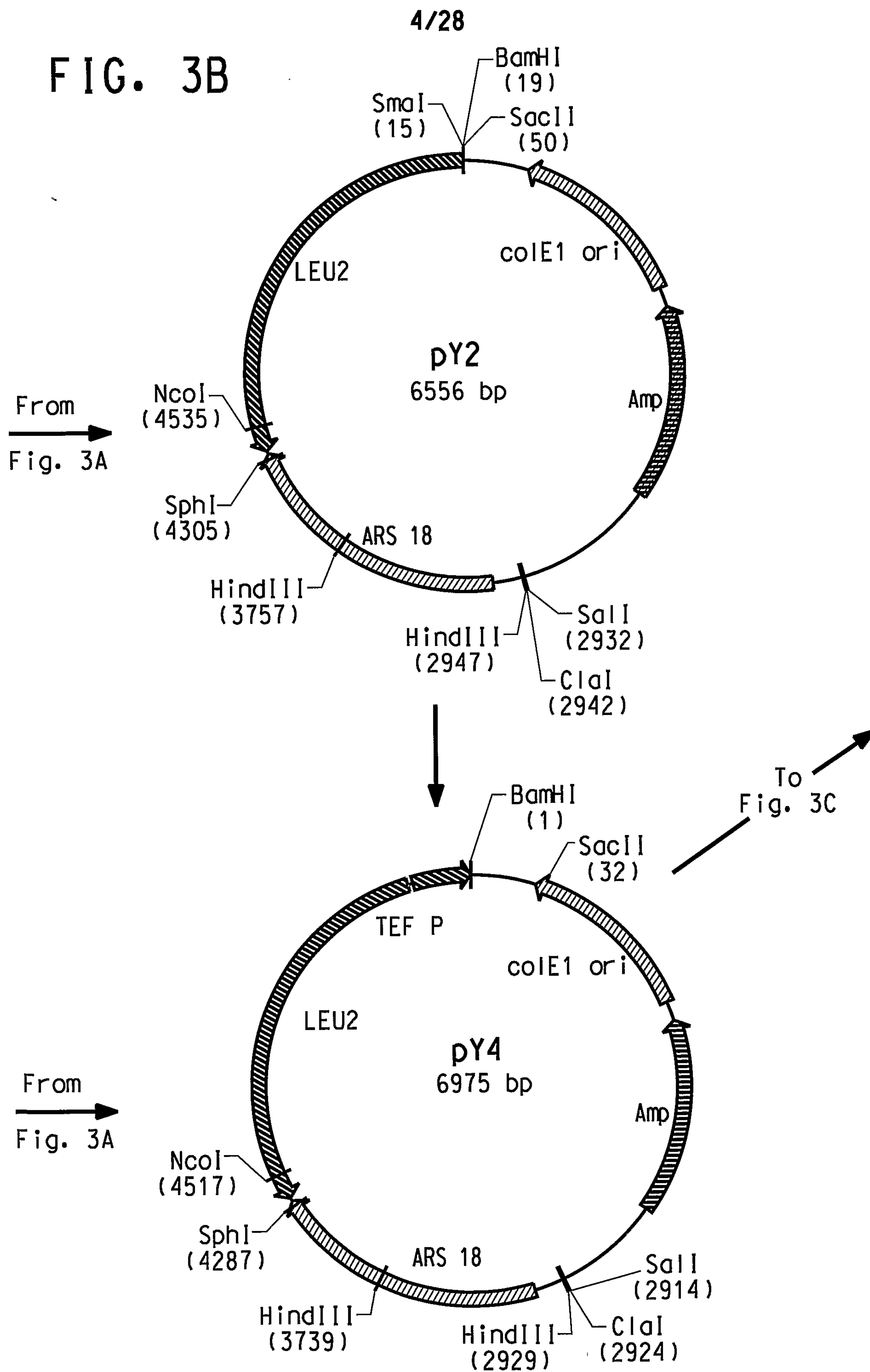


FIG. 3A

FIG. 3B



5/28

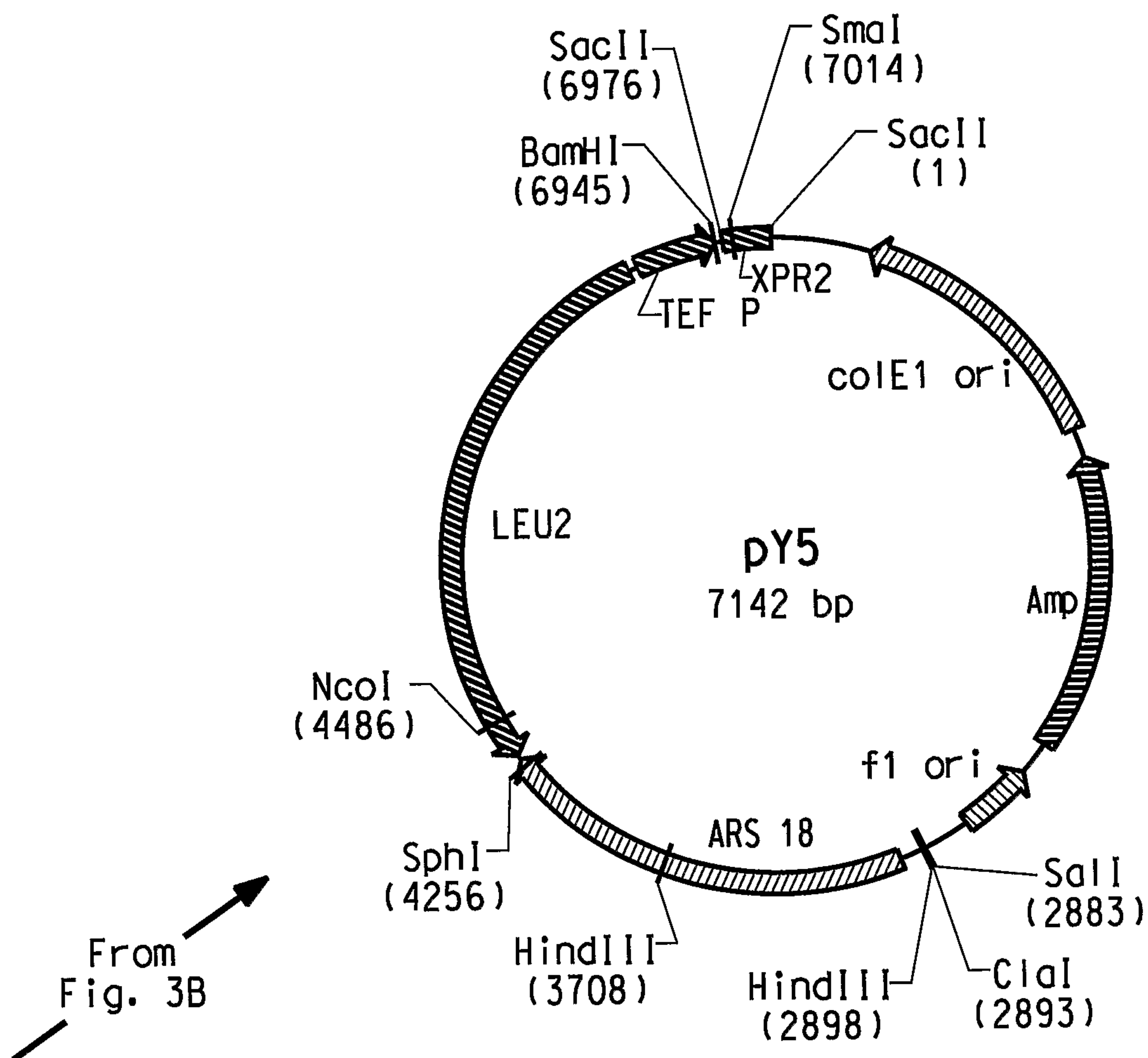


FIG. 3C

6/28

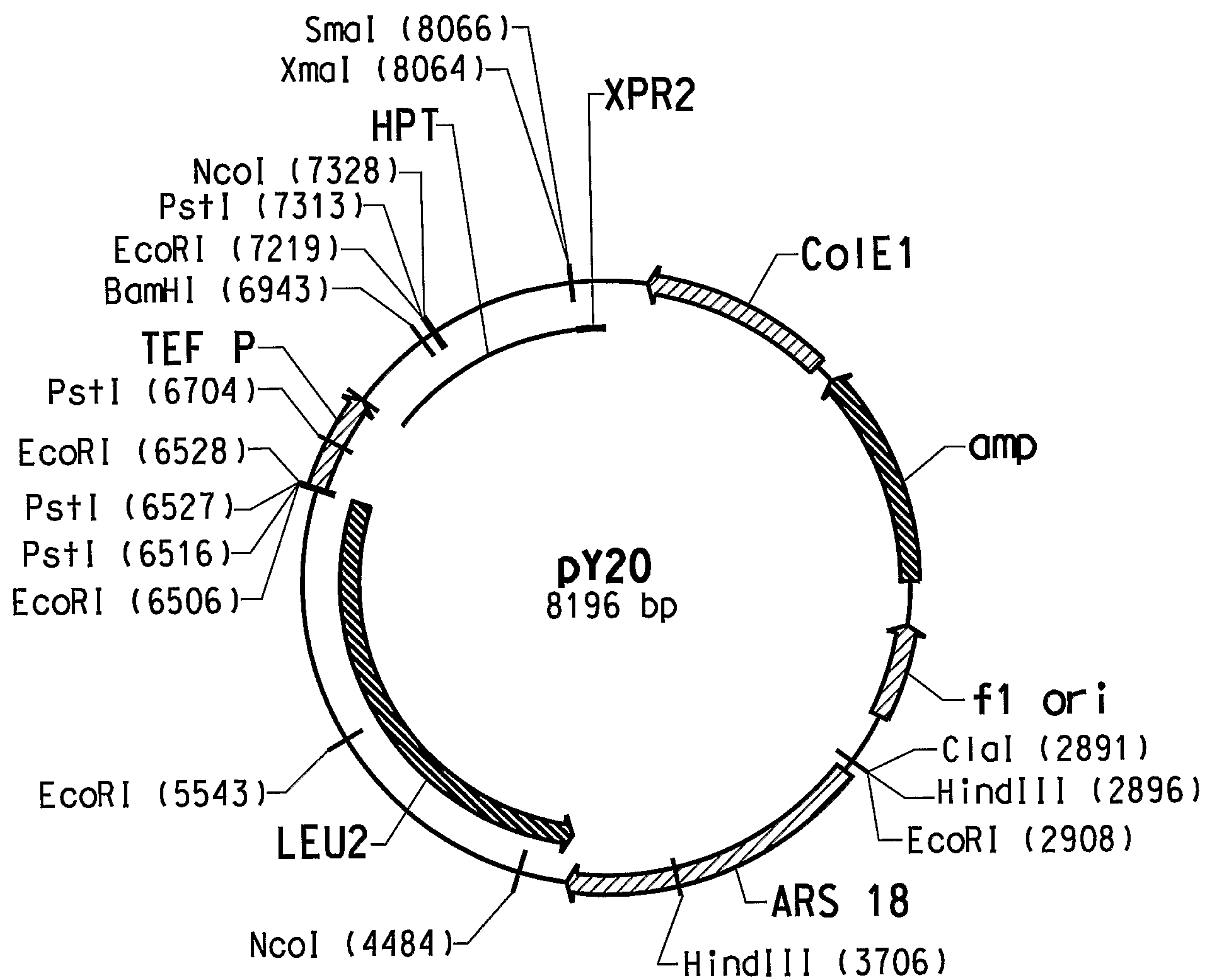


FIG. 4A

7/28

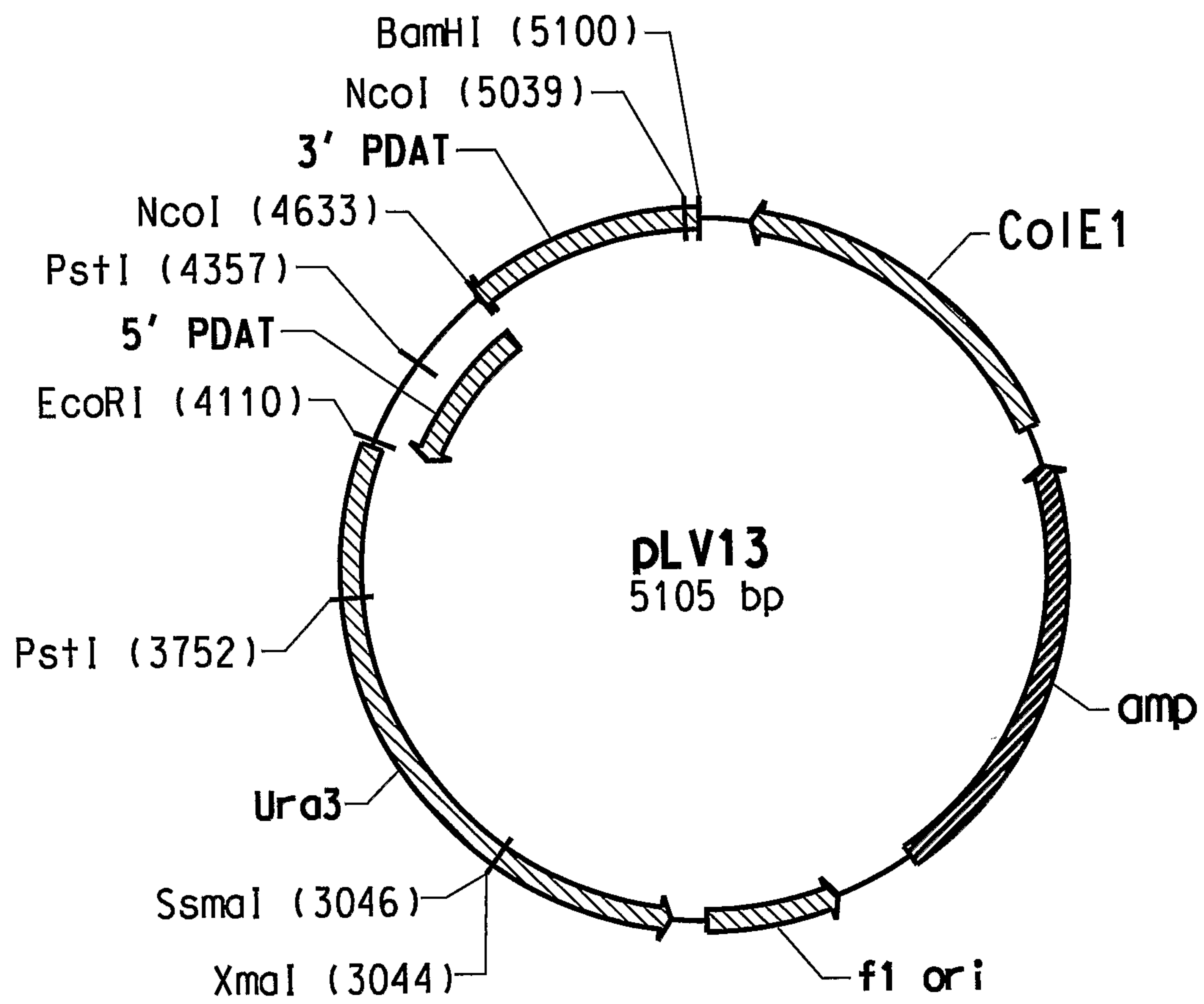


FIG. 4B

8/28

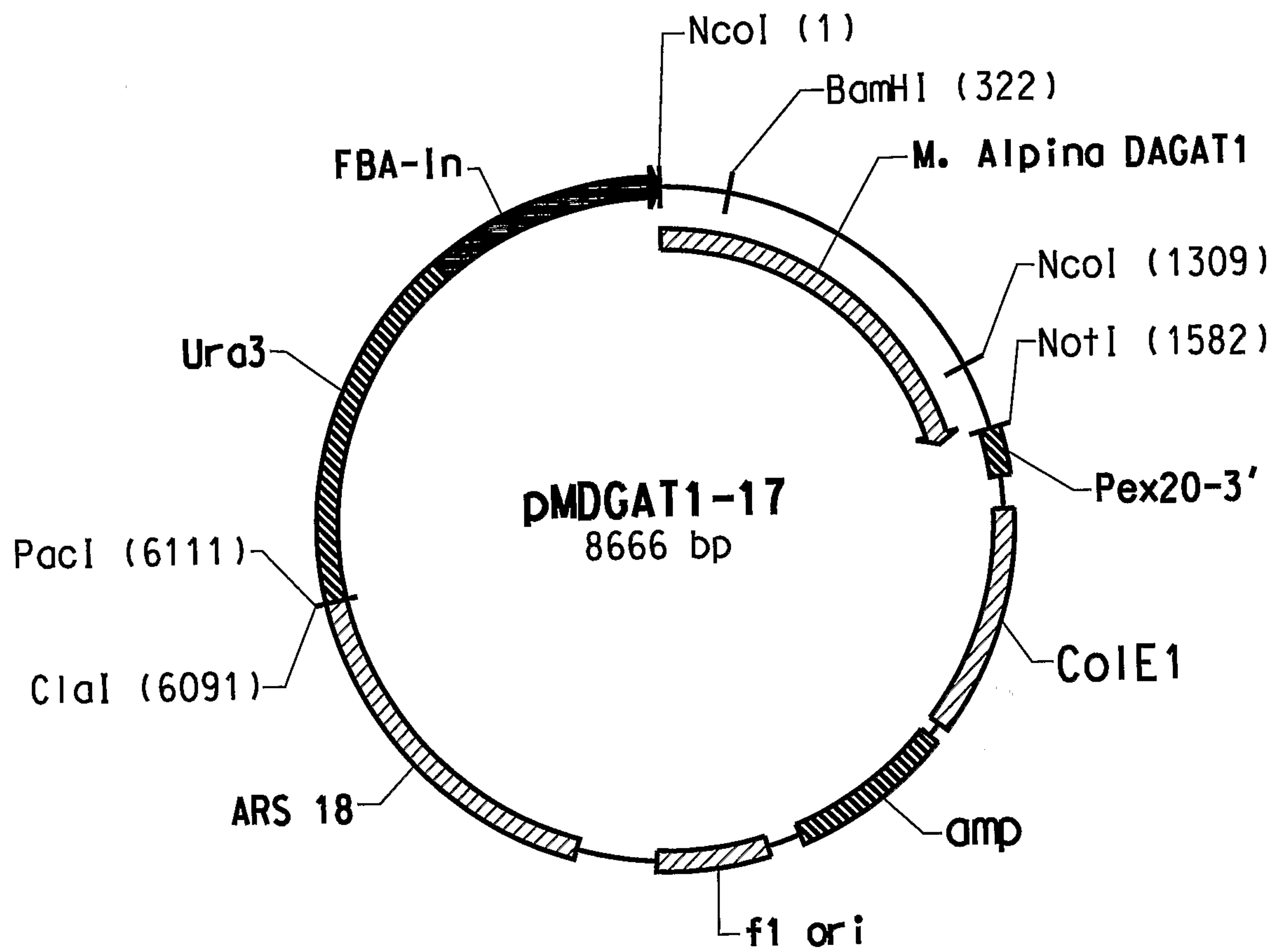


FIG. 4C

9/28

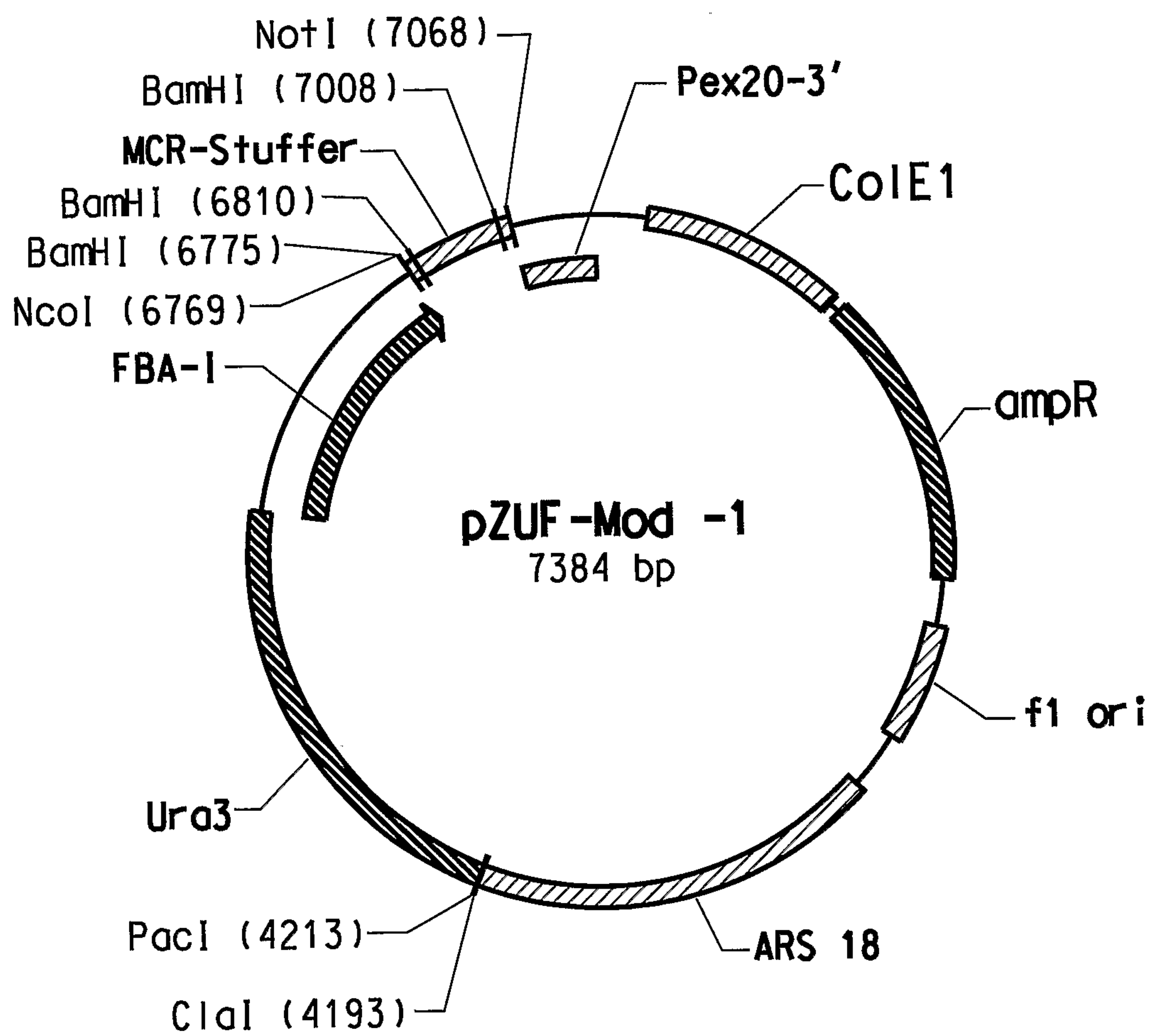


FIG. 4D

10/28

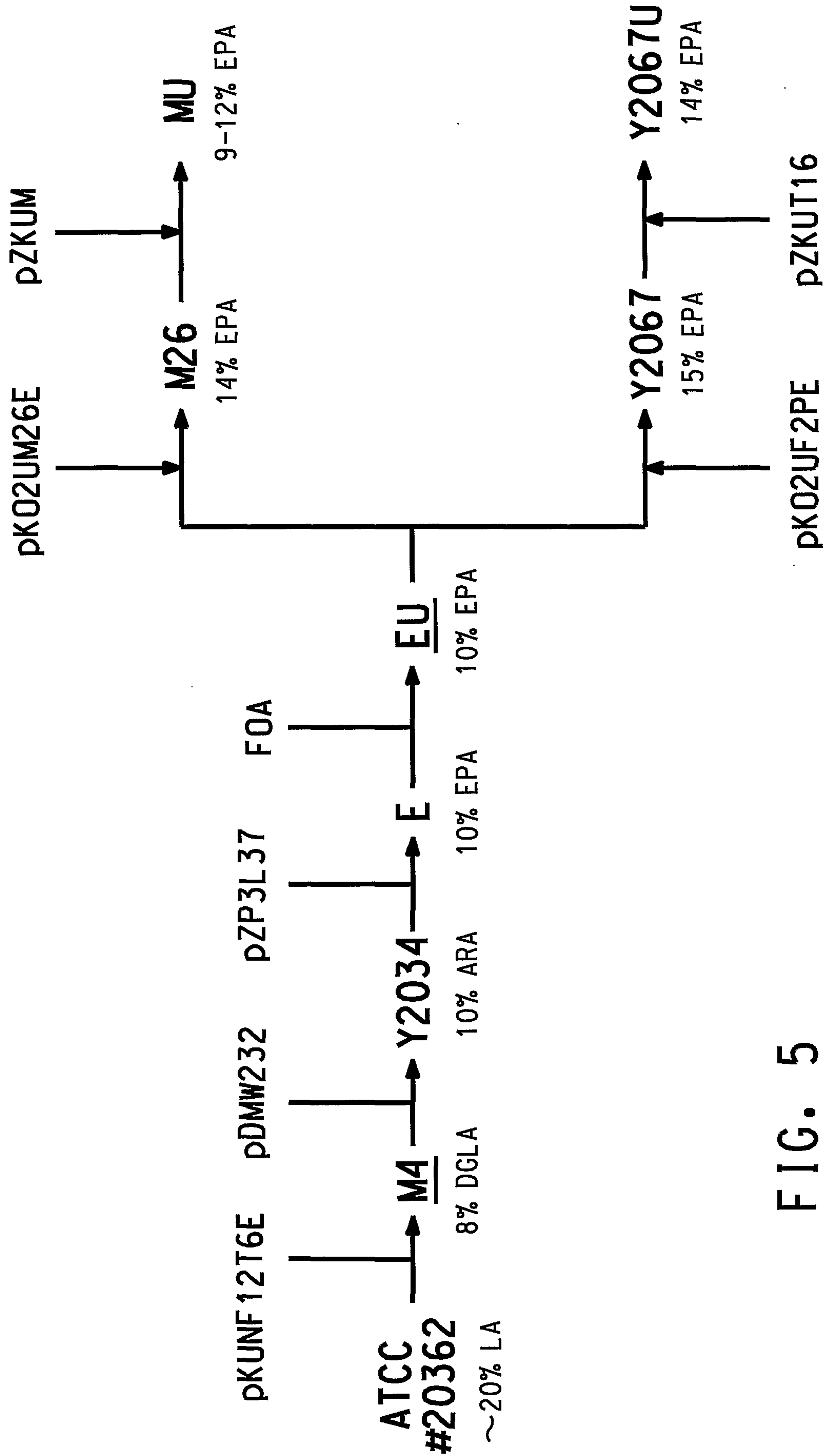


FIG. 5

11/28

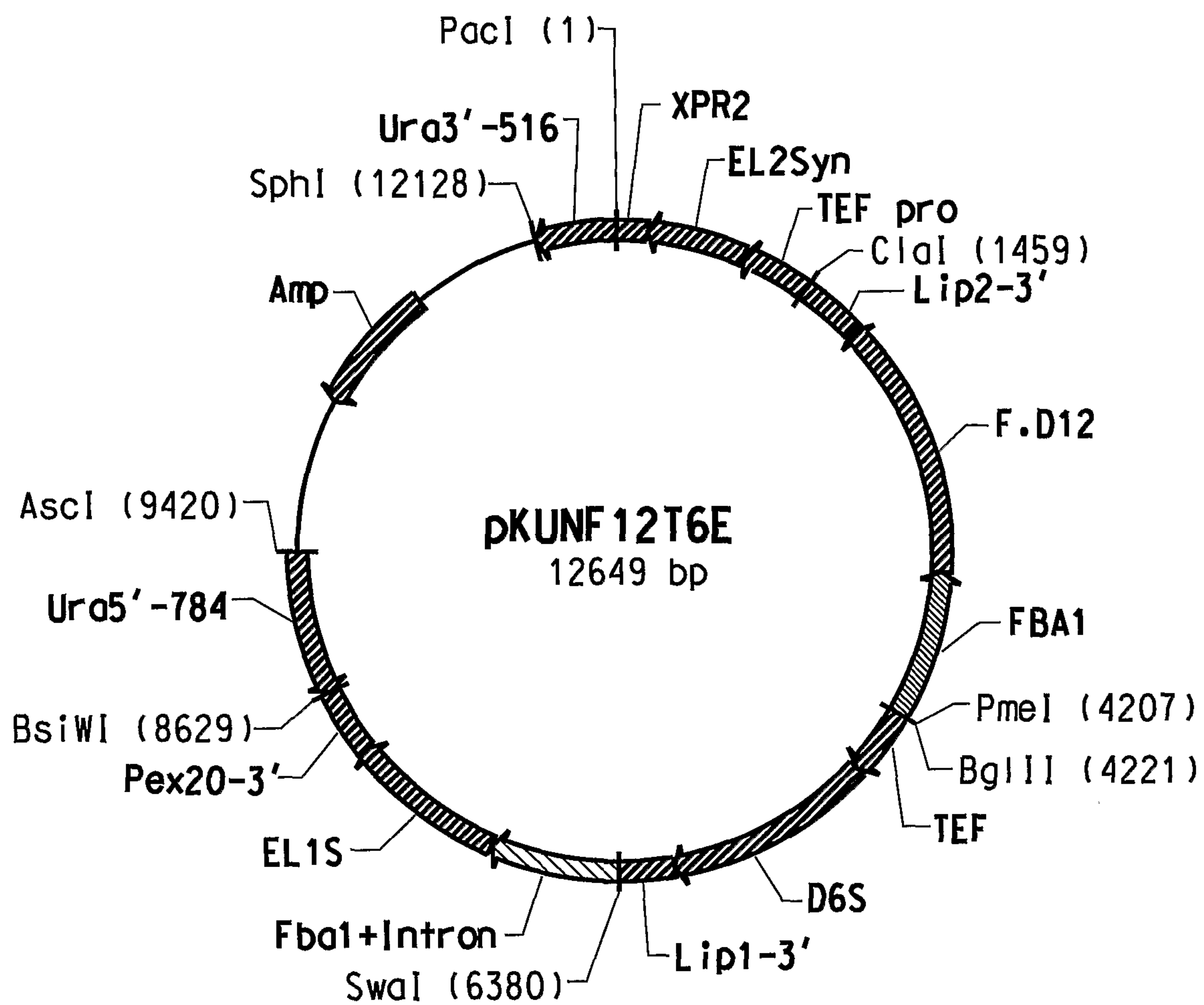


FIG. 6A

12/28

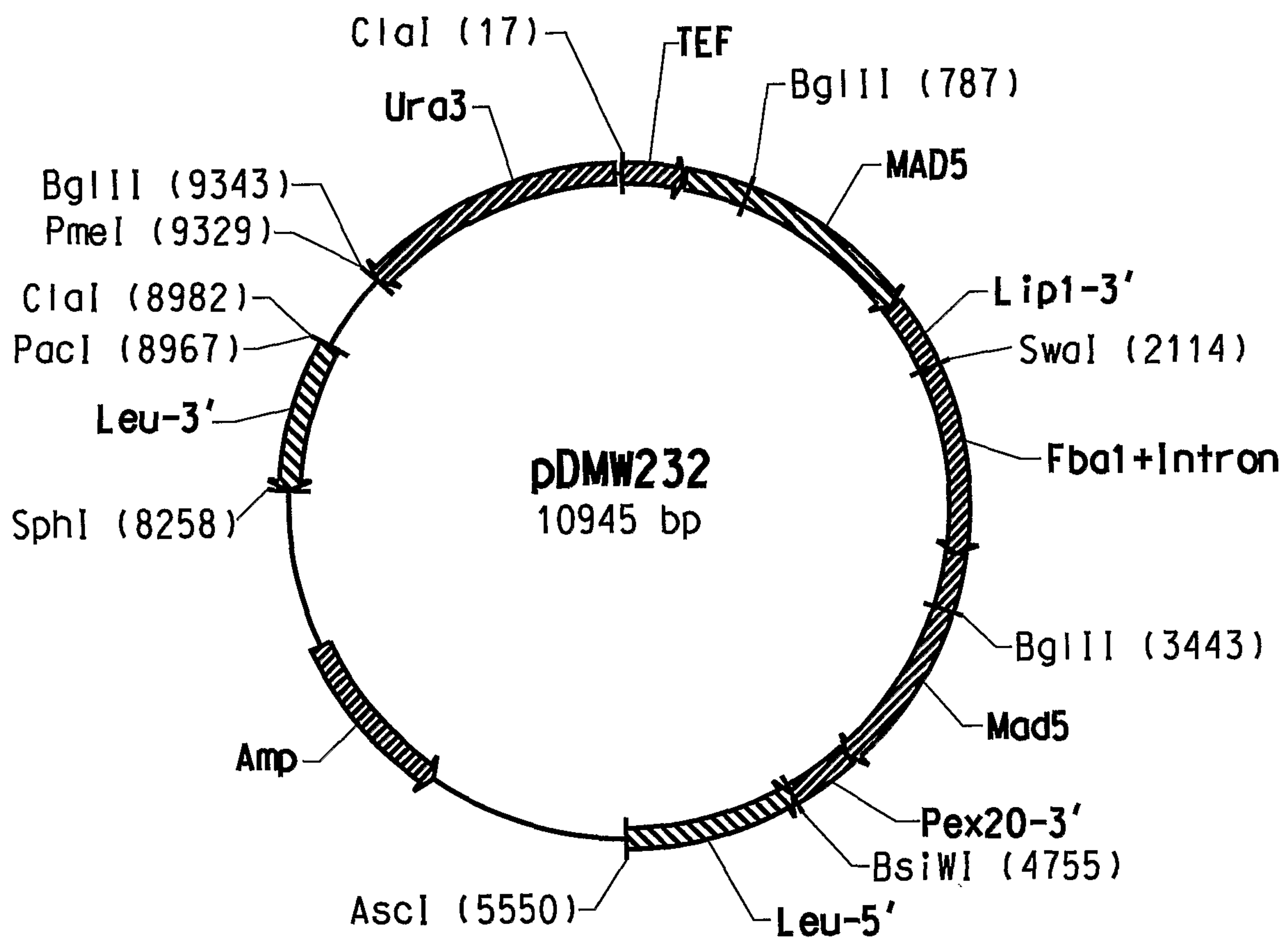


FIG. 6B

13/28

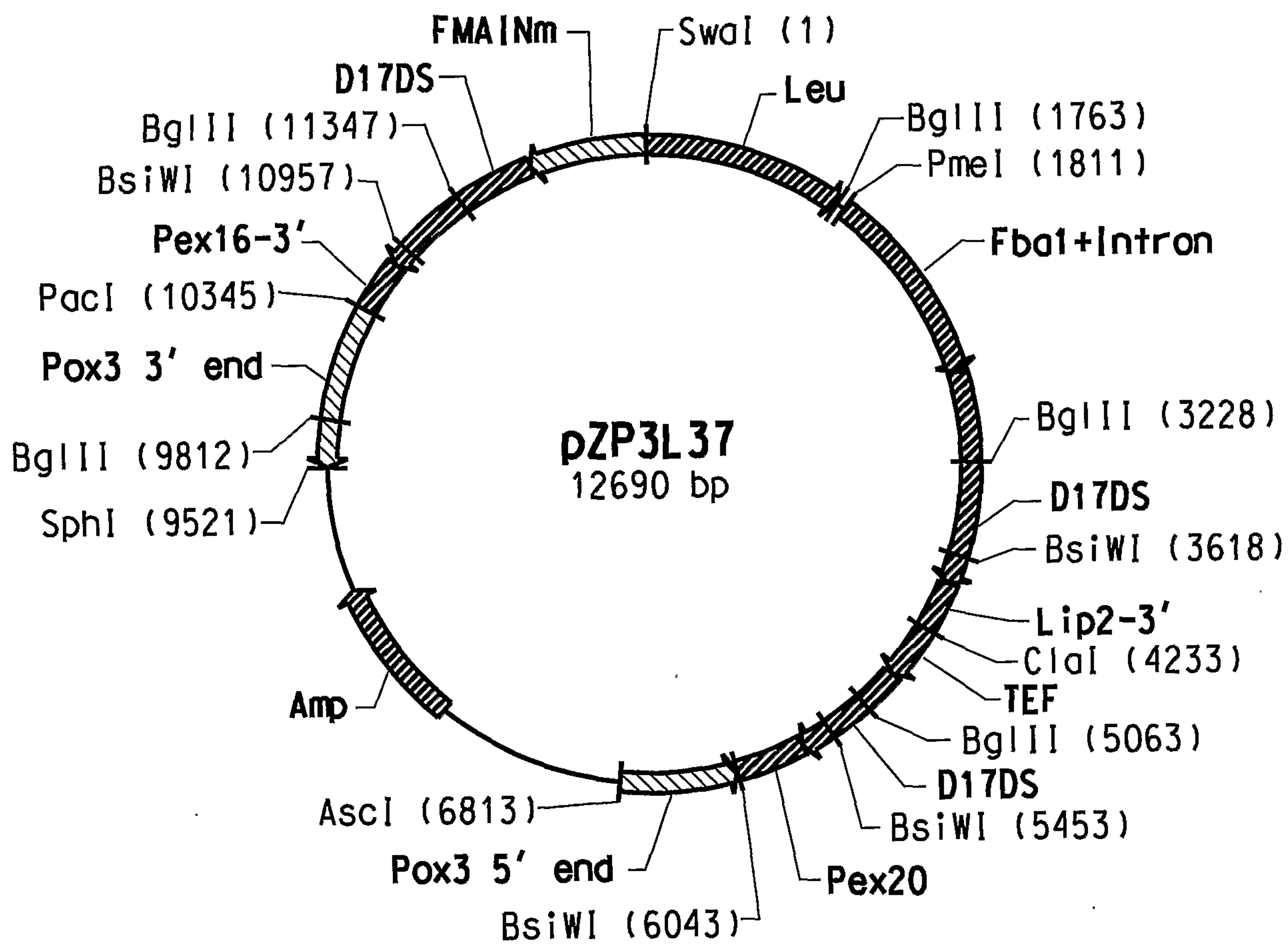


FIG. 6C

14/28

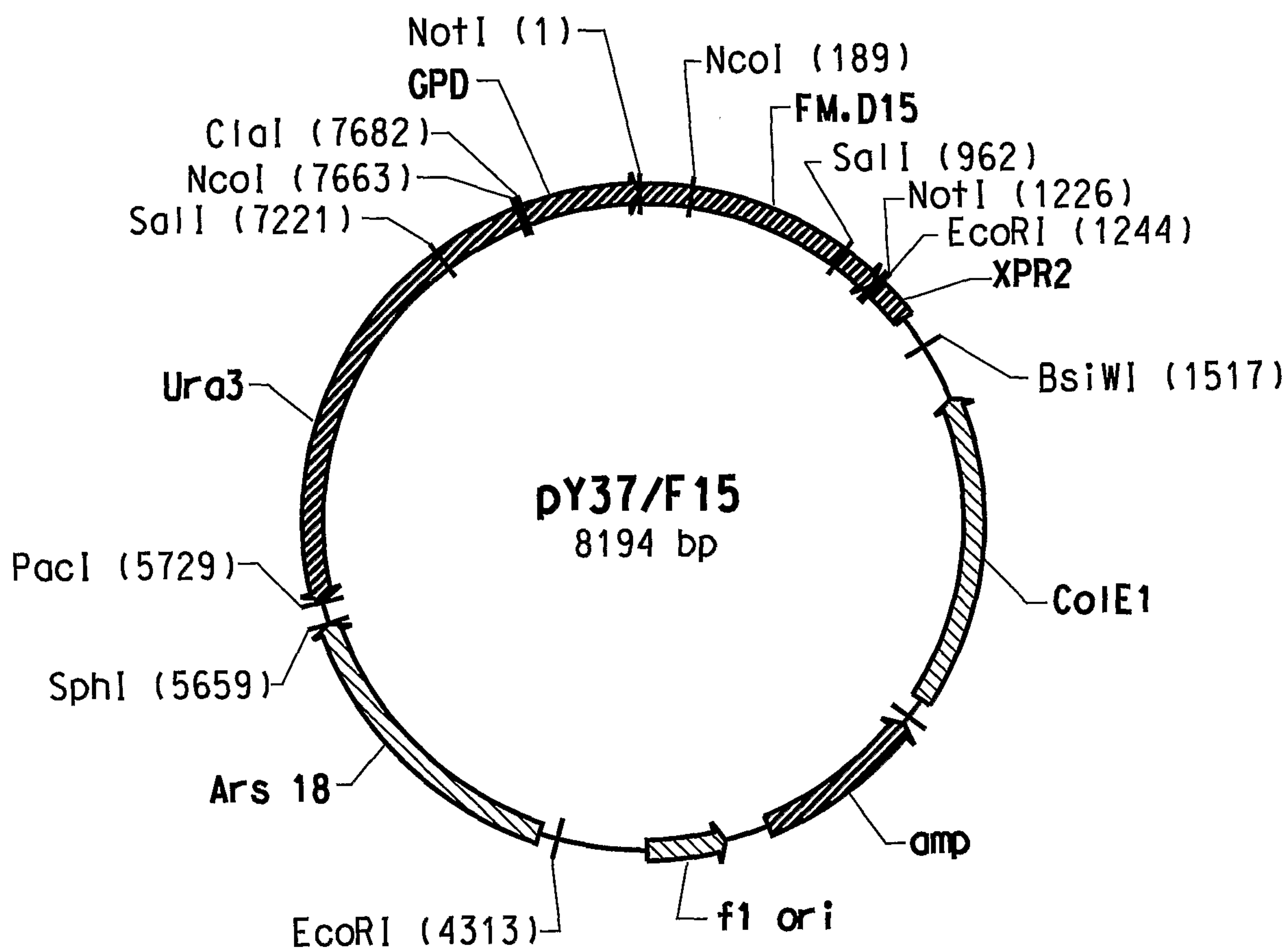


FIG. 6D

15/28

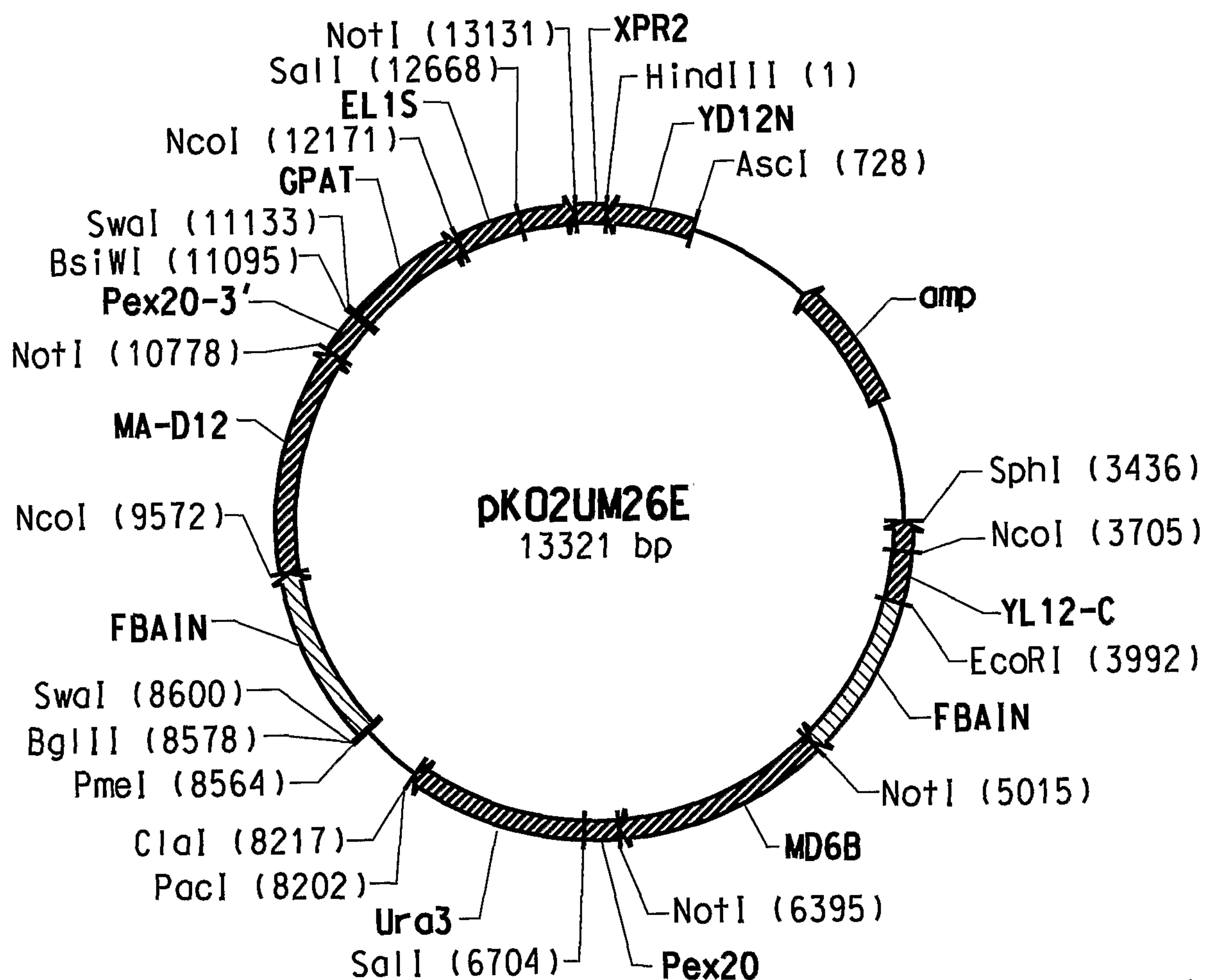


FIG. 6E

16/28

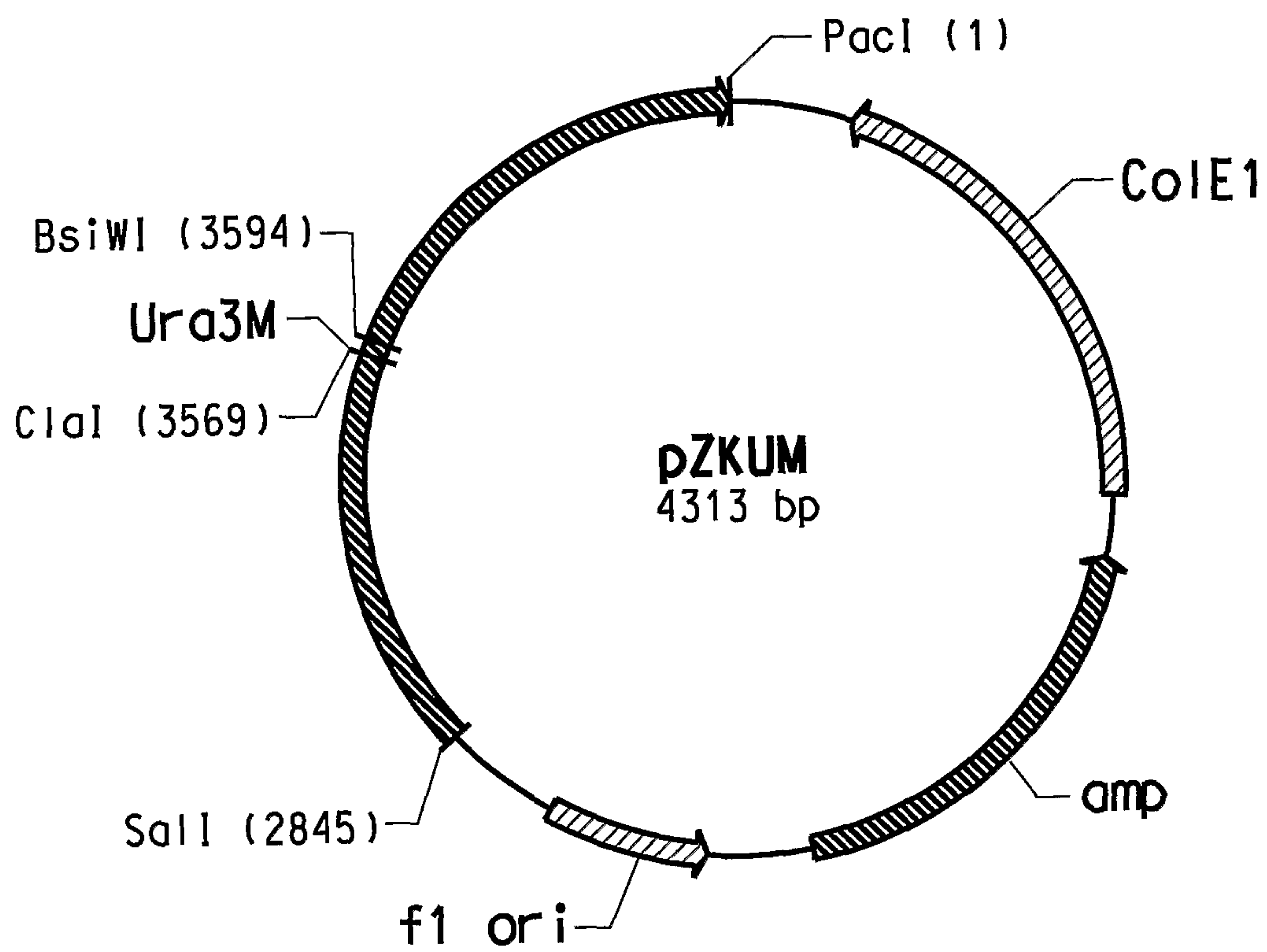


FIG. 7A

17/28

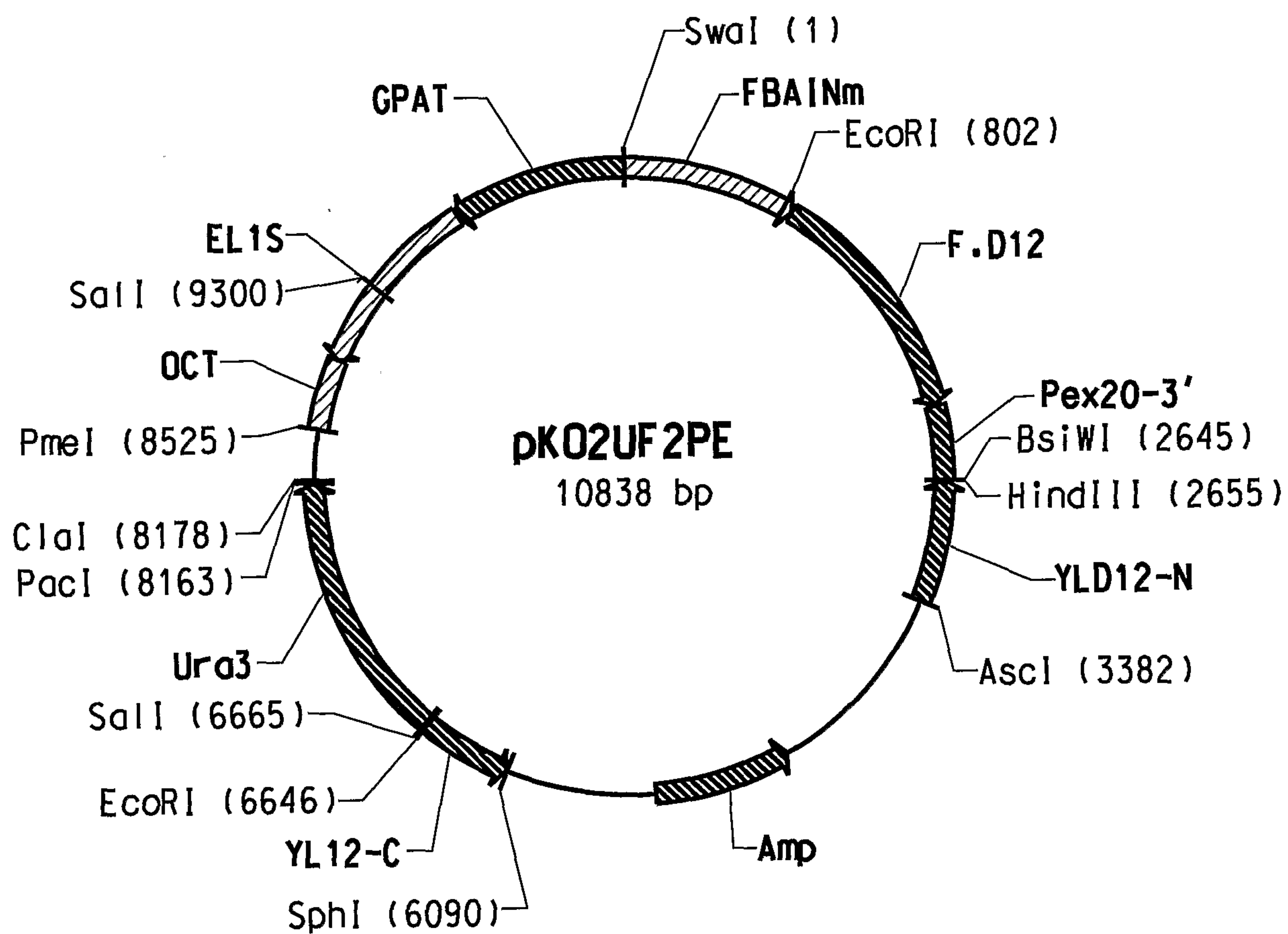


FIG. 7B

18/28

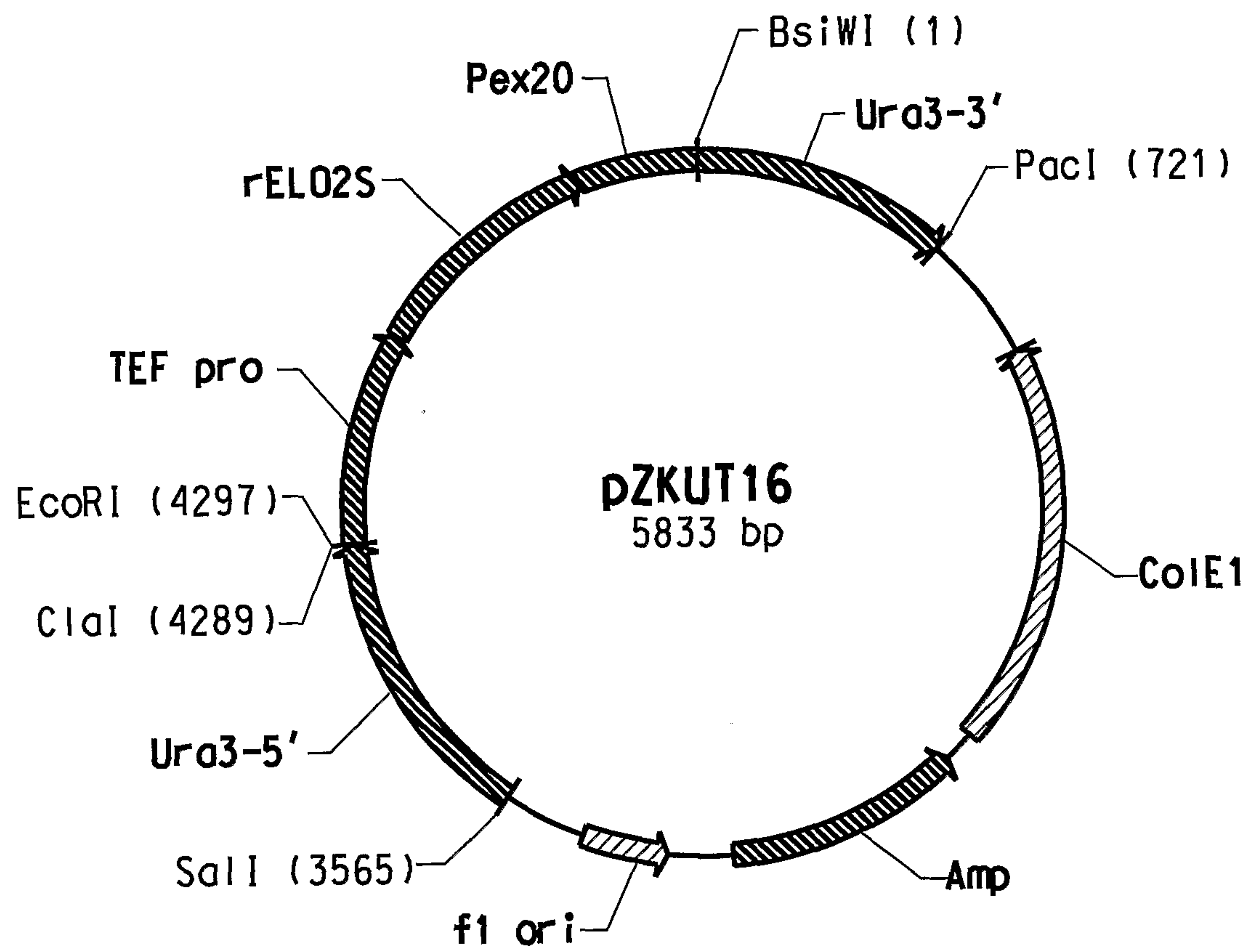


FIG. 7C

19/28

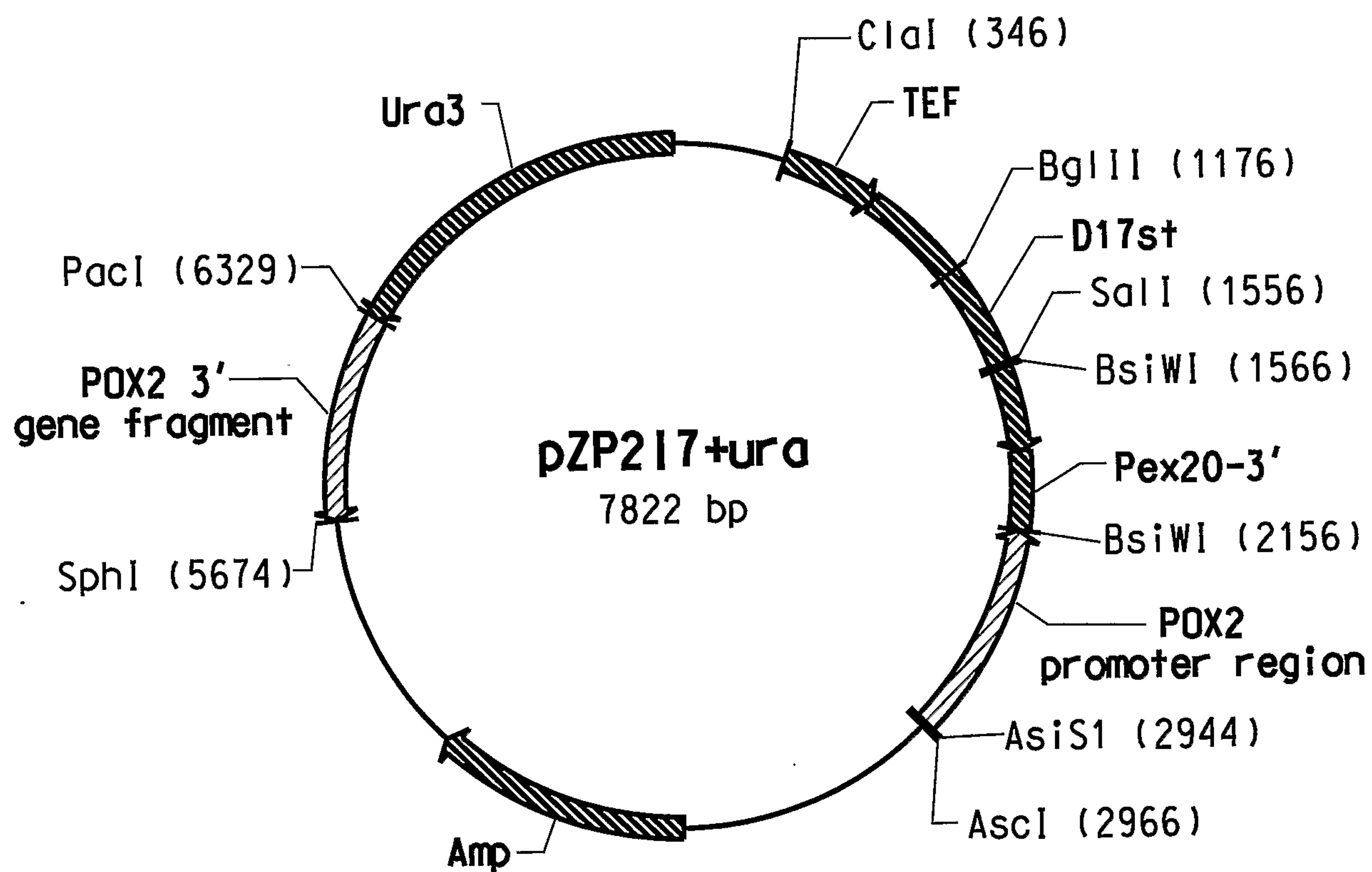


FIG. 7D

20/28

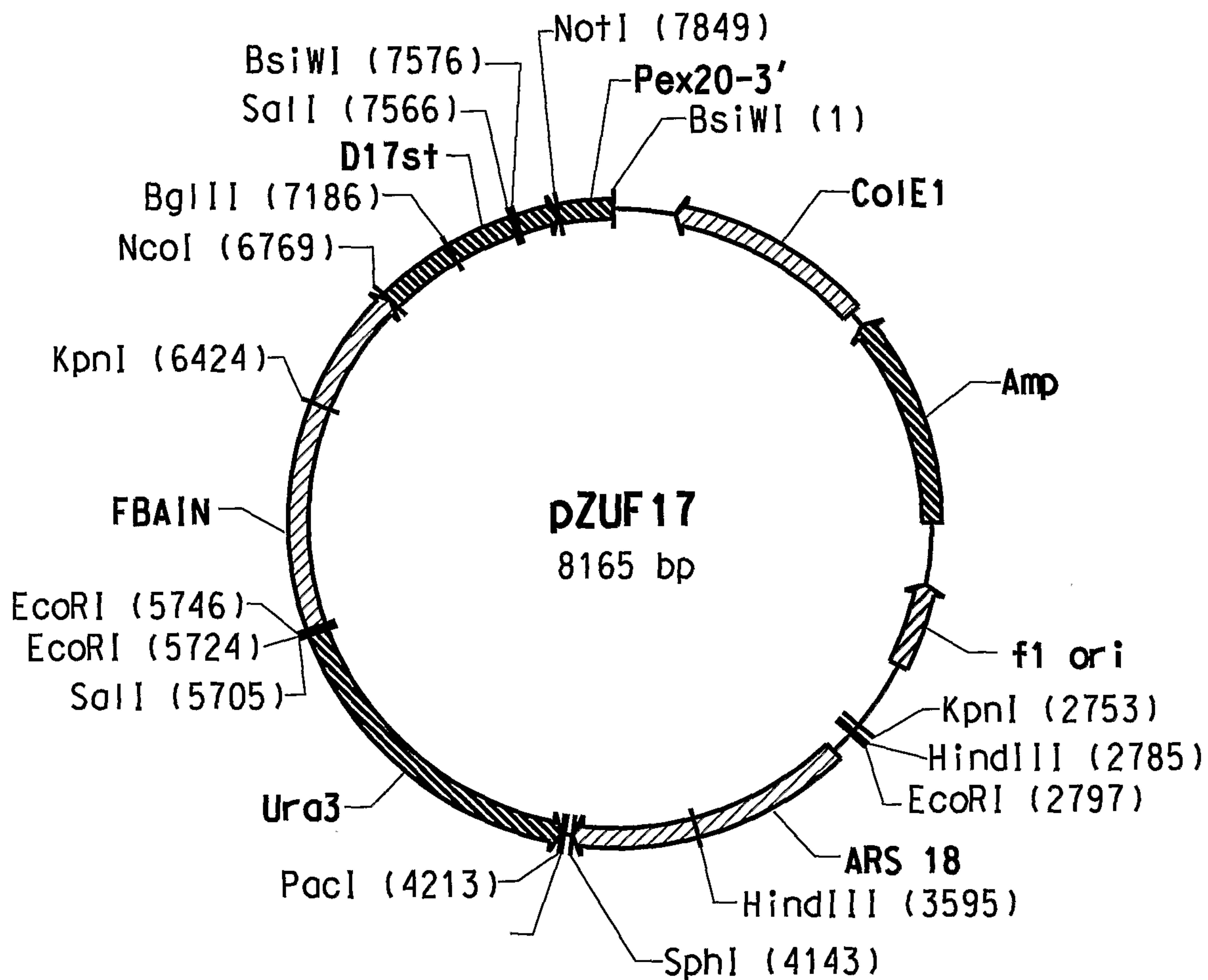


FIG. 7E

21/28

		Consensus	<u>SEQ</u> ID NO:
1	- - - - - M S S S T - - - - -	Nc DAGAT:	19
1	- - - - - M N S A T - - - - -	Fm DAGAT:	20
1	- - - - - M T E S T - - - - -	Ma DGAT1	18
1	M A A A T A T G L D L A A Q E G A Q Q R R S T - - - - -	Mg DAGAT:	21
1	- -	An DAGAT:	22
1	- - - - - [M] E V R R R K I D V L K - - - - -	Yl DGAT1	14
1	- - - - - - - M G D R G [G] A [G] S S - - - - -	Mm DGAT1	129
1	- - - - - [M] A I S D E F E S V A T - - - - -	Gm DGAT1	130
1	- - - - - [M] A I L D S A G V T T V T E N G G G E - - - - -	At DGAT1	131
1	- - - - - [M] V G S D G D G D G G G G - - - - -	Os DGAT1	132
1	- - - - - [M] A I L D S P E I L D T T S S S A D N - - - - -	Pfdgat1	133
1	- - - - - [M] S K G N P D P H L P G S F L - - - - -	Ta DGAT1	134
		Consensus	
6	- - - - - A T T [T] G L D P A [V] H T S N [D] - - - - -	Nc DAGAT:	
6	- - - - - T T S [T] E T [S] N G S T S - - - - -	Fm DAGAT:	
6	- - - - - T T T C A K E E G I A N - - - - -	Ma DGAT1	
24	- - - - - A T N Q [S] A D D D [V] T T N A [D] G - - - - -	Mg DAGAT:	A
1	- -	An DAGAT:	
13	- - - - - A Q K N G Y E S [G] P P S R Q S - - - - -	Yl DGAT1	
11	- - - - - [R R R R] R T G [S R] - [V] S V Q G G [S] - - - - -	Mm DGAT1	
13	- A L N H S S L R R R R P [S] - A T S T A G L F N S - - - - -	Gm DGAT1	
20	- F V D L D R L R R R R K [S] R [S] D S S N G L L L S [S] G S D N N S - - - - -	At DGAT1	
14	- E A H A G G P R R R R A G Q L R - - G R L R D E - - - - -	Os DGAT1	
20	G A A H H T T L R R R R Q [S] - A R S [V] P P L L D S [S] D S N S L E - - - - -	Pfdgat1	
16	- P S H G G P P P K P K T P P R T F R N L P S [S] - - - - -	Ta DGAT1	
		Consensus	
21	- - - N -	Nc DAGAT:	
18	- -	Fm DAGAT:	[V] S K - -
18	- -	Ma DGAT1	
41	A A A P -	Mg DAGAT:	S L K G T T A D
1	- -	An DAGAT:	
28	- S Q P -	Yl DGAT1	[S] [S] R A [S] [S] R T
26	- -	Mm DGAT1	[G] P K [V] E E D E
35	P E T T T D S S -	Gm DGAT1	[G] D D L A K D S
49	P S D D V G A P A D V R -	At DGAT1	D R I D [S] V [V] N D D A
35	- A A P -	Os DGAT1	[G] S P P R P R P
49	A E S A I N D S E N V R N D A N L I E N L R -	Pfdgat1	[G] G A [V] E [S] E N
39	- S T H -	Ta DGAT1	[G] P A P [S] V A A

FIG. 8A

22/28

																											Consensus	SEQ ID NO:					
30	E	N	G	S	T	P	N	D	K	A	N	A	G	G	E	P	E	T	E	T	K	R	H	S	K	K	V	-	V	R	Nc	DAGAT:	19
21	R	N	G	H	D	V	T	R	T	N	G	N	G	T	T	T	S	P	P	K	K	A	G	-	-	-	-	-	-	Fm	DAGAT:	20	
18	-	-	-	S	A	A	L	P	D	I	P	P	K	M	E	D	L	K	S	S	R	K	T	G	-	-	-	-	-	Ma	DGAT1	18	
53	T	N	G	T	S	N	G	N	G	N	G	N	G	N	V	D	E	D	E	Q	T	K	A	L	R	K	A	-	F	T	Mg	DAGAT:	21
1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	M	A	T	R	K	T	A	-	-	-	-	An	DAGAT:	22
39	R	N	K	H	S	S	S	T	L	S	L	S	G	L	T	M	K	V	Q	K	K	P	A	G	P	P	A	N	S	K	Yl	DGAT1	14
34	V	R	D	A	A	V	S	P	D	L	G	A	G	G	D	A	P	A	P	A	P	A	P	A	H	T	R	D	K	D	Mm	DGAT1	129
51	G	S	D	D	S	I	N	S	D	D	-	A	A	V	N	S	Q	-	-	Q	Q	N	E	K	Q	D	T	D	F	S	Gm	DGAT1	130
72	Q	G	T	A	N	L	A	G	D	N	N	G	G	G	D	N	N	G	G	G	R	G	G	G	E	G	R	G	N	A	At	DGAT1	131
46	R	P	R	P	R	G	G	D	S	N	G	R	S	V	L	R	P	G	G	G	G	G	R	G	G	G	G	D	F	S	Os	DGAT1	132
79	E	K	Q	E	S	Y	G	K	E	E	-	G	A	K	V	K	E	N	G	E	T	S	N	G	N	G	T	D	V	M	Pfdgat1	133	
50	A	T	I	A	T	T	P	P	S	A	S	A	A	P	L	P	P	T	V	H	G	E	A	A	H	G	A	A	A	A	Ta	DGAT1	134

. S . . S																												Consensus				
59	S	-	-	-	-	K	Y	R	H	V	E	A	V	H	S	Q	S	R	P	S	C	L	S	H	D	T	T	E	S	-	Nc	DAGAT:
45	Q	-	-	-	-	K	Y	R	H	V	A	A	V	H	K	K	T	R	P	S	C	L	S	H	D	S	D	A	A	-	Fm	DAGAT:
39	S	-	-	-	-	S	Y	K	H	T	F	P	V	H	T	K	T	I	P	S	P	L	S	K	E	A	P	P	E	-	Ma	DGAT1
82	R	-	-	-	-	K	Y	R	H	V	A	A	L	H	S	Q	A	R	P	S	T	L	S	H	D	S	E	A	S	-	Mg	DAGAT:
8	-	-	-	-	-	I	Y	R	H	A	V	A	V	H	S	Q	V	Q	H	S	C	L	S	R	D	S	T	K	A	-	An	DAGAT:
69	T	-	-	-	-	P	F	L	H	I	K	P	V	H	T	C	C	S	T	S	M	L	S	R	D	Y	D	G	S	N	Yl	DGAT1
64	G	R	T	S	V	G	D	G	Y	W	D	L	R	C	H	R	L	Q	D	S	L	F	S	S	D	S	G	F	S	-	Mm	DGAT1
78	V	L	K	-	F	A	Y	R	P	S	V	P	A	H	R	K	V	K	E	S	P	L	S	S	D	T	I	F	R	-	Gm	DGAT1
102	D	A	T	-	F	T	Y	R	P	S	V	P	A	H	R	R	A	R	E	S	P	L	S	S	D	A	I	F	K	-	At	DGAT1
76	A	-	-	-	F	T	F	R	A	A	A	P	V	H	R	K	A	K	E	S	P	L	S	S	D	A	I	F	K	-	Os	DGAT1
108	A	V	K	-	F	T	F	R	P	A	A	P	A	H	R	K	N	K	E	S	P	L	S	S	D	A	I	F	K	-	Pfdgat1	
80	A	R	R	D	A	L	L	P	G	V	G	A	A	H	R	R	V	K	E	S	P	L	S	S	D	A	I	F	R	-	Ta	DGAT1

. . . . G . . N L																									Consensus							
84	P	S	F	L	G	F	R	N	L	M	V	I	V	L	A	N	N	S	H	-	-	-	-	-	-	-	Q	Y	G	Nc	DAGAT:	
70	P	S	F	I	G	F	R	N	L	M	V	I	V	L	G	I	Y	H	I	-	-	-	-	-	G	M	S	Q	F	D	Fm	DAGAT:
64	-	S	Y	R	G	F	V	N	L	G	M	L	L	L	F	G	N	N	I	R	L	I	I	E	N	Y	L	K	Y	G	Ma	DGAT1
107	P	S	F	V	G	F	R	N	L	M	V	I	V	L	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Mg	DAGAT:
32	T	S	F	I	G	F	R	N	L	M	V	V	V	L	V	A	M	N	L	R	L	V	I	E	N	F	L	K	Y	G	An	DAGAT:
95	P	S	F	K	G	F	K	N	I	G	M	I	I	L	I	V	G	N	-	-	-	-	-	-	-	-	-	-	-	-	Yl	DGAT1
93	N	-	Y	R	G	I	L	N	W	C	V	V	M	L	I	L	S	N	A	R	L	F	L	E	N	L	I	K	Y	G	Mm	DGAT1
106	Q	S	H	A	G	L	F	N	L	C	I	V	V	L	V	A	V	N	S	R	L	I	I	E	N	L	M	K	Y	G	Gm	DGAT1
130	Q	S	H	A	G	L	F	N	L	C	V	V	V	L	I	A	V	N	S	R	L	I	I	E	N	L	M	K	Y	G	At	DGAT1
102	Q	S	H	A	G	L	F	N	L	C	I	V	V	L	V	A	V	N	S	R	L	I	I	E	N	L	M	K	Y	G	Os	DGAT1
136	Q	S	H	A	G	L	F	N	L	C	I	V	V	L	V	A	V	N	S	R	L	I	I	E	N	L	M	K	Y	G	Pfdgat1	
109	Q	S	H	A	G	L	L	N	L	C	I	V	V	L	I	A	V	N	S	R	L	I	I	E	N	L	M	K	Y	G	Ta	DGAT1

Motif #1

FIG. 8B

23/28

																										Consensus	SEQ ID NO:						
106	V	L	I	C	I	-	-	G	C	H	D	F	R	K	S	D	-	-	I	N	L	G	L	L	L	Y	F	L	I	P	Nc	DAGAT:	19
95	S	E	Q	P	I	-	-	D	T	A	S	Y	R	-	Q	D	-	-	I	F	L	G	L	L	L	Y	F	L	I	P	Fm	DAGAT:	20
93	F	L	L	S	I	-	-	P	G	S	S	V	S	K	Q	D	-	-	W	I	L	A	A	L	T	H	A	I	L	P	Ma	DGAT1	18
121	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Mg	DAGAT:	21
62	V	L	I	C	I	-	-	R	C	H	D	Y	R	K	Q	D	-	-	V	V	I	G	A	I	L	F	A	L	V	P	An	DAGAT:	22
113	-	-	L	R	L	-	-	A	F	E	N	Y	L	K	Y	G	-	-	I	S	N	P	F	F	D	P	K	I	T	P	Yl	DGAT1	14
122	I	L	V	D	P	-	I	Q	V	V	S	L	F	L	K	D	P	Y	S	W	P	A	P	C	V	I	I	A	S	N	Mm	DGAT1	129
136	W	L	I	K	S	G	F	W	F	S	S	K	S	L	R	D	-	-	W	P	L	F	M	C	C	L	S	L	V	Gm	DGAT1	130	
160	W	L	I	R	T	D	F	W	F	S	S	R	S	L	R	D	-	-	W	P	L	F	M	C	W	I	S	L	S	At	DGAT1	131	
132	L	L	I	R	A	G	F	W	F	N	D	K	S	L	R	D	-	-	W	P	L	L	M	C	C	L	S	L	P	Os	DGAT1	132	
166	W	L	I	K	S	G	F	W	F	S	S	T	S	L	R	D	-	-	W	P	L	L	M	C	C	L	S	L	P	Pfdgat1	133		
139	L	L	I	R	A	G	F	W	F	S	A	R	S	L	G	D	-	-	W	P	L	L	M	C	C	L	T	L	P	Ta	DGAT1	134	
																										Consensus							
132	C	H	L	F	I	A	Y	I	I	E	Y	Y	A	A	V	Q	A	R	A	E	R	N	V	S	A	S	E	-	-	-	Nc	DAGAT:	
120	C	H	L	L	A	A	Y	L	I	E	L	A	A	A	Q	Q	A	R	G	-	-	-	-	-	S	-	-	-	-	-	Fm	DAGAT:	
119	V	N	L	I	L	A	Y	K	L	E	S	W	A	K	E	R	A	V	G	Y	R	K	R	R	S	D	E	P	I	A	Ma	DGAT1	
121	-	-	-	-	-	-	-	-	-	E	L	L	A	A	Q	Q	A	R	N	S	R	-	-	-	G	Y	F	-	-	-	Mg	DAGAT:	
88	C	Q	L	L	C	S	Y	F	I	E	L	A	A	S	R	H	A	Q	R	-	-	-	-	-	-	-	-	-	-	-	An	DAGAT:	
137	S	E	W	Q	L	S	G	-	-	L	L	I	V	V	A	Y	A	H	I	-	-	-	-	-	-	-	-	-	-	-	Yl	DGAT1	
151	I	F	V	V	A	A	F	Q	I	E	K	R	L	A	V	G	A	L	T	-	-	-	-	-	-	-	-	-	-	-	Mm	DGAT1	
163	V	F	P	F	A	A	F	I	V	E	K	L	A	Q	R	K	C	I	P	-	-	-	-	-	-	-	-	-	-	-	Gm	DGAT1	
187	I	F	P	L	A	A	F	T	V	E	K	L	V	L	Q	K	Y	I	S	-	-	-	-	-	-	-	-	-	-	-	At	DGAT1	
159	A	F	P	L	G	A	F	A	V	E	K	L	A	F	N	N	V	I	T	-	-	-	-	-	-	-	-	-	-	-	Os	DGAT1	
193	V	F	A	L	A	S	F	L	V	E	K	L	V	K	L	N	Y	I	P	-	-	-	-	-	-	-	-	-	-	-	Pfdgat1		
166	I	F	P	L	A	A	L	M	T	E	K	W	A	Q	R	K	L	I	R	-	-	-	-	-	-	-	-	-	-	-	Ta	DGAT1	
																										Consensus							
159	-	-	-	Q	N	A	K	E	H	Q	H	Q	D	G	T	N	S	P	T	E	E	Q	H	R	K	F	Q	-	-	-	Nc	DAGAT:	
140	-	-	-	-	-	-	L	K	R	Y	N	D	S	A	S	G	G	P	S	D	Q	E	R	K	K	F	H	-	-	-	Fm	DAGAT:	
149	Q	E	S	T	K	A	V	X	A	G	D	N	D	A	I	K	T	T	K	P	A	K	A	Q	D	L	T	P	E	A	Ma	DGAT1	
136	-	-	-	N	R	G	R	T	G	S	S	R	D	G	S	T	S	P	T	E	D	E	S	R	R	F	V	-	-	-	Mg	DAGAT:	
107	-	-	-	-	-	-	V	I	G	R	A	K	K	Q	D	K	D	R	I	L	N	E	S	K	-	-	-	-	-	-	An	DAGAT:	
154	-	-	-	-	-	-	L	M	A	Y	A	I	E	S	A	A	K	L	L	F	L	S	S	K	-	-	-	-	-	-	Yl	DGAT1	
170	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Mm	DGAT1		
182	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Gm	DGAT1		
206	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	At	DGAT1		
178	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Os	DGAT1		
212	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Pfdgat1			
185	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Ta	DGAT1		

FIG. 8C

24/28

. H																											Consensus	SEQ ID NO:					
183	-	-	-	S	T	W	K	L	V	R	L	L	H	A	I	N	V	T	T	A	L	V	L	T	S	Y	V	V	Y	Y	Nc	DAGAT:	19
161	-	-	-	K	T	W	V	I	V	A	W	A	H	L	F	N	I	T	L	A	L	V	L	T	T	W	V	V	Y	F	Fm	DAGAT:	20
179	L	A	R	K	E	Q	S	T	V	G	W	L	H	V	F	N	L	F	T	I	V	A	W	P	S	F	M	S	Y	F	Ma	DGAT1	18
160	-	-	-	S	T	W	K	L	I	A	L	V	H	G	I	N	V	N	S	A	L	L	I	T	T	Y	T	V	Y	F	Mg	DAGAT:	21
125	-	-	-	R	T	W	F	A	I	A	L	L	H	S	I	I	S	F	F	G	L	A	A	T	S	Y	V	I	F	Y	An	DAGAT:	22
172	-	-	-	H	H	Y	M	A	V	G	L	L	H	T	M	N	T	L	S	S	I	S	L	L	S	Y	V	V	Y	Y	Yl	DGAT1	14
170	-	-	-	-	E	Q	M	G	L	L	L	H	V	V	N	L	A	T	I	I	C	F	P	A	A	V	A	L	L	Mm	DGAT1	129	
182	-	-	-	-	E	P	V	V	V	V	L	H	I	I	I	T	S	T	S	L	F	Y	P	V	L	V	I	L	R	Gm	DGAT1	130	
206	-	-	-	-	E	P	V	G	I	F	L	H	I	I	I	T	M	T	E	V	L	Y	P	V	Y	V	T	L	R	At	DGAT1	131	
178	-	-	-	-	D	A	V	A	T	C	L	H	I	F	L	S	T	T	E	I	V	Y	P	V	L	V	I	L	K	Os	DGAT1	132	
212	-	-	-	-	E	W	V	A	V	F	L	H	V	T	I	T	T	V	E	I	L	F	P	V	V	V	I	L	R	Pfdgat1	133		
185	-	-	-	-	D	H	V	S	I	L	L	H	I	I	I	T	T	T	V	L	I	Y	P	V	V	V	I	L	K	Ta	DGAT1	134	
. L K . . S																											Consensus						
210	H	I	H	H	P	L	I	G	T	L	T	E	V	H	A	-	I	V	V	W	L	K	T	A	S	Y	A	F	T	N	Nc	DAGAT:	
188	K	I	H	H	P	L	I	G	T	L	T	E	M	H	A	-	I	A	V	W	L	K	T	A	S	Y	A	F	T	N	Fm	DAGAT:	
209	M	I	Y	H	P	F	V	A	M	S	C	L	M	N	G	-	L	I	L	F	L	K	M	T	S	F	A	L	V	N	Ma	DGAT1	
187	H	I	H	H	P	L	I	G	T	L	T	E	M	H	A	-	V	I	V	W	L	K	T	A	S	Y	A	F	T	N	Mg	DAGAT:	
152	Y	V	N	H	P	G	I	G	T	V	C	E	V	Q	V	-	I	I	V	S	L	K	S	Y	S	Y	A	L	T	N	An	DAGAT:	
199	Y	L	P	N	P	V	A	G	T	I	V	E	F	V	A	-	V	I	L	S	L	K	L	A	S	Y	A	L	T	N	Yl	DGAT1	
195	V	E	S	I	T	P	V	G	S	V	F	A	L	A	S	Y	S	I	M	F	L	K	L	Y	S	Y	R	D	V	N	Mm	DGAT1	
207	C	D	S	A	F	V	S	G	V	T	L	M	L	F	S	-	C	V	V	W	L	K	L	V	S	Y	A	H	T	N	Gm	DGAT1	
231	C	D	S	A	F	L	S	G	V	T	L	M	L	L	T	-	C	I	V	W	L	K	L	V	S	Y	A	H	T	S	At	DGAT1	
203	C	D	S	A	V	L	S	G	F	L	L	I	F	I	A	-	C	I	V	W	L	K	L	V	S	F	A	H	T	N	Os	DGAT1	
237	C	D	S	A	V	L	S	G	V	T	L	M	L	F	A	-	C	T	V	W	L	K	L	V	S	Y	A	H	T	N	Pfdgat1		
210	C	E	S	A	V	L	S	G	F	V	L	M	F	I	A	-	S	I	T	W	L	K	L	V	S	F	A	H	T	N	Ta	DGAT1	
. R																											Consensus						
239	-	-	-	R	D	L	R	H	A	Y	L	H	P	A	R	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Nc	DAGAT:	
217	-	-	-	R	D	L	R	H	A	Y	L	H	P	V	E	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Fm	DAGAT:	
238	-	-	-	Q	E	L	R	A	A	Y	I	F	G	T	P	V	D	T	F	Q	H	M	A	K	V	H	D	I	S	-	Ma	DGAT1	
216	-	-	-	R	D	L	R	H	A	Y	L	H	P	V	K	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Mg	DAGAT:	
181	-	-	-	R	D	L	R	R	A	M	L	G	S	P	S	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	An	DAGAT:	
228	-	-	-	S	D	L	R	K	A	A	I	H	A	Q	K	L	D	K	T	Q	D	D	N	E	K	E	S	T	S	S	Yl	DGAT1	
225	L	W	C	R	Q	R	R	V	K	A	K	A	V	S	T	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Mm	DGAT1	
236	-	-	-	Y	D	M	R	A	L	T	K	L	V	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Gm	DGAT1	
260	-	-	-	Y	D	I	R	S	L	A	N	A	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	At	DGAT1	
232	-	-	-	H	D	I	R	Q	L	T	M	G	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Os	DGAT1	
266	-	-	-	Y	D	L	R	V	L	A	K	S	L	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Pfdgat1		
239	-	-	-	Y	D	I	R	I	L	S	Q	S	I	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Ta	DGAT1	

FIG. 8D

25/28

													Consensus	SEQ ID NO:																		
251	-	-	-	-	-	-	-	-	-	G	E	L	D	A	L	P	G	L	Y	A	E	C	P	Y	P	E	N	I	T	Nc DAGAT:	19	
229	-	-	-	-	-	-	-	-	-	G	E	R	E	L	V	P	E	L	Y	T	Q	C	P	Y	P	Q	N	I	T	Fm DAGAT:	20	
264	-	-	-	-	-	-	G	-	-	K	D	L	T	K	K	E	I	F	Q	Y	D	I	Q	Y	P	D	N	I	T	Ma DGAT1	18	
228	-	-	-	-	-	-	-	-	-	G	E	L	D	A	L	P	E	L	Y	K	Q	C	P	Y	P	N	N	I	T	Mg DAGAT:	21	
193	-	-	-	-	-	-	-	-	-	A	D	S	D	-	I	P	E	L	Y	R	S	C	P	Y	P	R	N	I	T	An DAGAT:	22	
255	S	S	S	S	D	D	A	E	T	L	A	D	I	D	V	I	P	A	Y	Y	A	Q	L	P	Y	P	Q	N	V	T	Yl DGAT1	14
240	-	-	-	-	-	-	-	-	-	G	K	K	V	S	G	A	A	A	Q	Q	A	V	S	Y	P	D	N	L	T	Mm DGAT1	129	
246	-	-	-	-	-	-	-	-	-	E	K	G	E	A	L	L	D	T	L	N	M	D	Y	P	Y	N	V	S	Gm DGAT1	130		
270	-	-	-	-	-	-	-	-	-	D	K	-	-	-	-	-	-	A	N	P	E	V	S	Y	Y	V	S	At DGAT1	131			
242	-	-	-	-	-	-	-	-	-	K	K	V	D	N	E	L	S	T	V	D	M	D	N	L	Q	P	P	T	Os DGAT1	132		
276	-	-	-	-	-	-	-	-	-	D	K	W	E	A	M	S	R	Y	W	N	L	D	Y	A	Y	D	V	S	Pfdgat1	133		
249	-	-	-	-	-	-	-	-	-	E	K	G	A	T	H	G	S	S	I	D	E	E	N	I	K	G	P	T	Ta DGAT1	134		

Motif #2

. Y F . . A P T L . Y P R														Consensus																		
271	M	G	N	L	C	Y	F	W	W	A	P	T	L	V	Y	Q	P	V	Y	P	R	T	A	K	I	R	W	S	F	V	Nc	DAGAT:
249	F	S	N	L	A	Y	F	W	W	A	P	T	L	V	Y	Q	P	V	Y	P	R	T	D	K	I	R	W	G	F	V	Fm	DAGAT:
285	L	K	N	I	G	Y	F	W	L	A	P	T	L	C	Y	Q	P	S	Y	P	R	T	T	V	F	R	K	S	F	F	Ma	DGAT1
248	M	K	N	L	C	Y	F	W	W	A	P	T	L	I	Y	Q	P	V	Y	P	R	S	G	R	I	R	W	V	F	F	Mg	DAGAT:
212	L	G	N	L	A	Y	F	L	W	A	P	T	L	V	Y	Q	P	V	Y	P	R	T	P	R	I	R	W	S	F	V	An	DAGAT:
285	L	S	N	L	L	Y	F	W	F	A	P	T	L	V	Y	Q	P	V	Y	P	K	T	E	R	I	R	P	K	H	V	Yl	DGAT1
260	Y	R	D	L	Y	Y	F	I	F	A	P	T	L	C	Y	E	L	N	F	P	R	S	P	R	I	R	K	R	F	L	Mm	DGAT1
265	F	K	S	L	A	Y	F	L	V	A	P	T	L	C	Y	Q	P	S	Y	P	R	T	P	Y	I	R	K	G	W	L	Gm	DGAT1
282	L	K	S	L	A	Y	F	M	V	A	P	T	L	C	Y	Q	P	S	Y	P	R	S	A	C	I	R	K	G	W	V	At	DGAT1
261	L	G	N	L	I	Y	F	M	M	A	P	T	L	C	Y	Q	P	S	Y	P	R	T	S	C	V	R	K	G	W	L	Os	DGAT1
295	F	K	S	L	A	Y	F	M	V	A	P	T	L	C	Y	Q	P	S	Y	P	R	T	A	C	I	R	K	G	W	V	Pfdgat1	
268	I	N	S	V	V	Y	F	M	L	A	P	T	L	C	Y	Q	P	S	Y	P	R	T	A	F	I	R	K	G	W	V	Ta	DGAT1

. Q . . . P														Consensus																		
301	A	K	R	C	G	E	V	I	C	L	S	V	F	I	W	F	L	S	A	Q	Y	A	T	P	V	L	R	N	S	L	Nc	DAGAT:
279	A	K	R	V	G	E	I	F	G	L	S	V	F	I	W	V	A	S	A	Q	Y	A	A	P	V	L	R	N	S	L	Fm	DAGAT:
315	L	K	R	V	A	E	I	V	T	C	L	G	M	M	Y	F	L	V	E	Q	Y	A	T	P	T	L	Q	N	S	V	Ma	DGAT1
278	F	K	R	V	A	E	V	F	C	L	S	V	C	I	W	F	L	S	A	Q	Y	A	T	P	V	L	V	N	S	L	Mg	DAGAT:
242	G	K	R	L	F	E	F	V	C	L	S	V	V	M	W	L	L	S	A	Q	Y	A	A	P	L	L	R	N	A	T	An	DAGAT:
315	I	R	N	L	F	E	L	V	S	L	C	M	L	I	Q	F	L	I	F	Q	Y	A	Y	P	I	M	Q	S	C	L	Yl	DGAT1
290	L	R	R	V	L	E	M	L	F	F	T	Q	L	Q	V	G	L	I	Q	Q	W	M	V	P	T	I	Q	N	S	M	Mm	DGAT1
295	F	R	Q	L	V	K	L	I	I	F	T	G	V	M	G	F	I	I	D	Q	Y	I	N	P	I	V	Q	N	S	Q	Gm	DGAT1
312	A	R	Q	F	A	K	L	V	I	F	T	G	F	M	G	F	I	I	E	Q	Y	I	N	P	I	V	R	N	S	K	At	DGAT1
291	I	R	Q	I	I	L	Y	L	I	F	T	G	L	Q	G	F	I	I	E	Q	Y	I	N	P	I	V	V	N	S	Q	Os	DGAT1
325	V	R	Q	L	I	K	L	V	I	F	T	G	L	M	G	F	I	I	E	Q	Y	I	N	P	I	V	Q	N	S	Q	Pfdgat1	
298	T	R	Q	L	I	K	C	V	V	F	T	G	L	M	G	F	I	I	E	Q	Y	I	N	P	I	V	Q	N	S	K	Ta	DGAT1

Motif #3

FIG. 8E

26/28

. E K L W L .																				Consensus		SEQ ID NO:											
331	D	K	I	A	S	-	-	L	D	I	P	S	I	V	E	R	L	L	K	L	S	T	I	S	L	I	I	W	L	A	Nc	DAGAT:	19
309	D	K	I	A	S	-	-	L	D	L	M	S	I	L	E	R	L	L	K	L	S	T	I	S	L	A	I	W	L	A	Fm	DAGAT:	20
345	R	A	F	D	E	-	-	L	A	F	G	T	I	L	E	R	V	L	K	L	S	T	T	S	V	I	I	W	L	L	Ma	DGAT1	18
308	D	K	I	A	S	-	-	L	D	M	P	A	I	L	E	R	L	L	K	L	S	T	I	S	L	A	I	W	L	A	Mg	DAGAT:	21
272	Q	K	I	A	T	-	-	L	D	I	A	S	I	L	E	R	G	L	K	L	S	T	I	S	L	V	I	W	L	A	An	DAGAT:	22
345	A	L	F	F	Q	P	K	L	D	Y	A	N	I	S	E	R	L	M	K	L	A	S	V	S	M	M	V	W	L	I	Yl	DGAT1	14
320	K	P	F	K	D	M	-	-	D	Y	S	R	I	I	E	R	L	L	K	L	A	V	P	N	H	L	I	W	L	I	Mm	DGAT1	129
325	H	P	L	K	G	-	-	-	N	L	L	Y	A	T	E	R	V	L	K	L	S	V	P	N	L	Y	V	W	L	C	Gm	DGAT1	130
342	H	P	L	K	G	-	-	-	D	L	L	Y	A	I	E	R	V	L	K	L	S	V	P	N	L	Y	V	W	L	C	At	DGAT1	131
321	H	P	L	K	G	-	-	-	G	L	L	N	A	V	E	T	V	L	K	L	S	L	P	N	V	Y	L	W	L	C	Os	DGAT1	132
355	H	P	L	K	G	-	-	-	N	L	L	Y	A	I	E	R	V	L	K	L	S	V	P	N	L	Y	V	W	L	C	Pfdgat1	133	
328	H	P	L	N	G	-	-	-	N	F	L	D	A	I	E	R	V	L	K	L	S	V	P	T	L	Y	V	W	L	C	Ta	DGAT1	134

. F . . . F A E . . . F . . R . F Y . . W W N																											Motif #4	Consensus				
359	G	F	F	A	L	F	Q	S	F	L	N	A	L	A	E	V	T	R	F	A	D	R	S	F	Y	D	E	W	W	N	Nc	DAGAT:
337	G	F	F	A	L	F	Q	S	F	L	N	A	L	A	E	V	L	R	F	G	D	R	S	F	Y	D	D	W	W	N	Fm	DAGAT:
373	M	F	Y	T	F	F	H	S	F	F	N	A	L	A	E	A	L	Y	F	G	D	R	R	F	Y	L	A	W	W	N	Ma	DGAT1
336	G	F	F	A	L	F	Q	S	F	L	N	A	L	A	E	I	T	R	F	G	D	R	S	F	Y	E	A	W	W	N	Mg	DAGAT:
300	G	F	Y	A	L	F	Q	S	L	L	N	G	L	A	E	I	M	R	F	G	D	R	E	F	Y	T	D	W	W	N	An	DAGAT:
375	G	F	Y	A	F	F	Q	N	G	L	N	L	I	A	E	L	T	C	F	G	N	R	T	F	Y	Q	Q	W	W	N	Yl	DGAT1
348	F	F	Y	W	F	F	H	S	C	L	N	A	V	A	E	L	L	Q	F	G	D	R	E	F	Y	R	D	W	W	N	Mm	DGAT1
352	M	F	Y	C	F	F	H	L	W	L	N	I	L	A	E	L	L	R	F	G	D	R	E	F	Y	K	D	W	W	N	Gm	DGAT1
369	M	F	Y	C	F	F	H	L	W	L	N	I	L	A	E	L	L	C	F	G	D	R	E	F	Y	K	D	W	W	N	At	DGAT1
348	M	F	Y	A	F	F	H	L	W	L	S	I	L	A	E	I	L	R	F	G	D	R	E	F	Y	K	D	W	W	N	Os	DGAT1
382	M	F	Y	C	F	F	H	L	W	L	N	I	L	A	E	L	L	C	F	G	D	R	E	F	Y	K	D	W	W	N	Pfdgat1	
355	M	F	Y	S	F	F	H	L	W	L	N	I	L	A	E	L	L	R	F	G	D	R	E	F	Y	K	D	W	W	N	Ta	DGAT1

. W . . W N . P V H . Y . P																											Consensus					
389	S	E	S	L	G	V	Y	W	R	T	W	N	K	P	V	Y	Q	Y	F	K	R	H	V	Y	S	P	M	R	S	R	Nc	DAGAT:
367	S	E	S	L	G	A	Y	W	R	T	W	N	K	P	V	Y	T	Y	F	K	R	H	L	Y	M	P	M	I	G	R	Fm	DAGAT:
403	A	T	G	V	G	M	Y	W	K	T	W	N	S	P	V	Y	T	F	F	K	R	H	V	Y	L	P	L	I	T	S	Ma	DGAT1
366	S	E	S	L	G	V	Y	W	R	T	W	N	K	P	V	Y	Q	Y	F	K	R	H	V	Y	S	P	M	L	G	R	Mg	DAGAT:
330	S	P	S	F	G	V	Y	W	R	S	W	N	R	P	V	Y	I	F	M	K	R	H	V	Y	M	P	L	V	T	R	An	DAGAT:
405	S	R	S	I	G	Q	Y	W	T	L	W	N	K	P	V	N	Q	Y	F	R	H	H	V	Y	V	P	L	L	A	R	Yl	DGAT1
378	A	E	S	V	T	Y	F	W	Q	N	W	N	I	P	V	H	K	W	C	I	R	H	F	Y	K	P	M	L	R	H	Mm	DGAT1
382	A	K	T	V	E	D	Y	W	R	M	W	N	M	P	V	H	K	W	M	I	R	H	L	Y	F	P	C	L	R	H	Gm	DGAT1
399	A	K	S	V	G	D	Y	W	R	M	W	N	M	P	V	H	K	W	M	V	R	H	I	Y	F	P	C	L	R	S	At	DGAT1
378	A	K	T	I	D	E	Y	W	R	K	W	N	M	P	V	H	K	W	V	V	R	H	I	Y	F	P	C	M	R	N	Os	DGAT1
412	A	R	T	V	E	E	Y	W	R	M	W	N	M	P	V	H	K	W	M	V	R	H	I	Y	C	P	C	L	Q	N	Pfdgat1	
385	A	K	T	V	E	E	Y	W	R	M	W	N	M	P	V	H	K	W	I	V	R	H	I	Y	F	P	C	I	R	N	Ta	DGAT1

Motif #6

Motif #5

FIG. 8F

27/28

. A F . . S A . . H E P . .															Consensus		us:	SEQ	ID NO:														
419	G	W	S	N	A	T	A	S	L	A	V	F	F	L	S	A	V	L	H	E	L	L	V	G	V	P	T	H	N	L	Nc	DAGAT:	19
397	G	W	S	P	Q	A	A	S	F	F	V	F	L	V	S	A	I	L	H	E	I	L	V	G	V	P	T	H	N	I	Fm	DAGAT:	20
433	G	T	S	P	M	V	A	S	I	V	I	F	L	I	S	A	V	L	H	E	I	L	I	G	F	P	T	H	M	I	Ma	DGAT1	18
396	G	W	A	P	R	T	A	S	A	S	V	F	L	I	S	A	V	L	H	E	I	L	V	G	V	P	T	H	N	I	Mg	DAGAT:	21
360	G	W	N	P	T	L	A	G	T	V	V	F	A	V	S	A	V	L	H	E	I	L	V	G	V	P	T	H	N	L	An	DAGAT:	22
435	G	M	S	R	F	N	A	S	V	V	V	F	F	F	S	A	V	I	H	E	L	L	V	G	I	P	T	H	N	I	Yl	DGAT1	14
408	G	S	S	K	W	V	A	R	T	G	V	F	L	T	S	A	F	F	H	E	Y	L	V	S	V	P	L	R	M	F	Mm	DGAT1	129
412	G	L	P	K	A	A	A	L	L	I	A	F	L	V	S	A	L	F	H	E	L	C	I	A	V	P	C	H	I	F	Gm	DGAT1	130
429	K	I	P	K	T	L	A	I	I	I	A	F	L	V	S	A	V	F	H	E	L	C	I	A	V	P	C	R	L	F	At	DGAT1	131
408	G	I	S	K	E	V	A	V	L	I	S	F	L	V	S	A	V	L	H	E	I	C	V	A	V	P	C	R	I	L	Os	DGAT1	132
442	G	I	P	K	I	V	A	V	L	I	A	F	L	V	S	A	I	F	H	E	L	C	V	A	V	P	C	Q	I	F	Pfdgat1	133	
415	G	L	S	K	G	C	A	I	L	I	A	F	L	V	S	A	V	F	H	E	L	C	I	A	V	P	C	H	I	F	Ta	DGAT1	134

Motif #7..

. . . A Q . P L																									Motif #7..					Consensus					us
449	I	G	V	A	F	L	G	M	F	L	Q	L	P	L	I	Q	F	T	K	P	L	E	-	K	K	T	S	P	N	G	Nc	DAGAT:			
427	I	G	V	A	F	L	G	M	F	L	Q	L	P	L	I	H	L	T	K	P	L	E	N	M	K	L	G	H	T	G	Fm	DAGAT:			
463	Y	G	Y	A	F	A	G	M	F	L	Q	I	P	L	I	I	L	T	R	P	L	E	K	W	R	G	T	G	S	G	Ma	DGAT1			
426	I	G	V	A	F	M	G	M	F	L	Q	V	P	L	I	I	L	T	A	P	L	E	-	K	R	K	S	P	T	G	Mg	DAGAT:			
390	I	G	V	A	S	I	A	M	M	F	Q	L	P	L	I	L	L	T	A	P	F	E	-	R	F	K	S	P	L	G	An	DAGAT:			
465	I	G	A	A	F	F	G	M	M	S	Q	V	P	L	I	M	A	T	E	N	L	Q	-	H	I	N	S	S	L	G	Yl	DGAT1			
438	R	L	W	A	F	T	A	M	M	A	Q	V	P	L	A	W	I	V	G	-	-	-	-	-	-	-	R	F	F	Q	Mm	DGAT1			
442	K	L	W	A	F	G	G	I	M	F	Q	V	P	L	V	L	I	T	N	Y	L	Q	-	-	-	N	K	F	R	N	Gm	DGAT1			
459	K	L	W	A	F	L	G	I	M	F	Q	V	P	L	V	F	I	T	N	Y	L	Q	-	-	-	E	R	F	G	-	At	DGAT1			
438	K	F	W	A	F	L	G	I	M	L	Q	I	P	L	I	V	L	T	A	Y	L	K	-	-	-	S	K	F	R	D	Os	DGAT1			
472	K	F	W	A	F	S	G	I	M	L	Q	V	P	L	V	I	V	T	N	Y	L	Q	-	-	-	E	K	F	K	N	Pfdgat1				
445	K	L	W	A	F	S	G	I	M	F	Q	I	P	L	L	F	L	T	K	Y	L	Q	-	-	-	D	K	F	K	N	Ta	DGAT1			

..Motif #7

. . . G N . . . W G Q P Y																				Consensus		us:										
478	K	L	L	G	N	I	I	F	W	V	S	F	T	I	F	G	Q	P	F	A	A	L	M	Y	F	Y	A	-	-	W	Nc	DAGAT:
457	K	I	V	G	N	T	I	F	W	V	S	F	T	I	F	G	Q	P	F	A	A	L	M	Y	F	Y	A	-	-	W	Fm	DAGAT:
493	-	-	L	G	N	M	I	F	W	V	S	F	T	I	L	G	Q	P	A	C	A	L	L	Y	Y	Y	H	-	-	W	Ma	DGAT1
455	K	L	I	G	N	S	I	F	W	V	S	F	T	I	F	G	Q	P	L	A	A	L	M	Y	F	Y	A	-	-	W	Mg	DAGAT:
419	K	A	I	G	N	S	F	F	W	V	T	F	C	V	V	G	Q	P	L	G	A	L	L	Y	F	F	A	-	-	W	An	DAGAT:
494	P	F	L	G	N	C	A	F	W	F	T	F	-	F	L	G	Q	P	T	C	A	F	L	Y	Y	L	A	-	-	Y	Yl	DGAT1
461	G	N	Y	G	N	A	A	V	W	V	T	L	-	I	I	G	Q	P	V	A	V	L	M	Y	V	H	D	Y	Y	V	Mm	DGAT1
469	S	M	V	G	N	M	I	F	W	F	I	F	S	I	L	G	Q	P	M	C	V	L	L	Y	Y	H	D	-	-	L	Gm	DGAT1
485	S	T	V	G	N	M	I	F	W	F	I	F	C	I	F	G	Q	P	M	C	V	L	L	Y	Y	H	D	-	-	L	At	DGAT1
465	T	M	V	G	N	M	I	F	W	F	F	F	C	I	Y	G	Q	P	M	C	L	L	L	Y	Y	H	D	-	-	V	Os	DGAT1
499	S	M	V	G	N	M	M	F	W	C	F	F	C	I	F	G	Q	P	M	C	V	L	L	Y	Y	H	D	-	-	L	Pfdgat1	
472	T	M	V	G	N	M	I	F	W	F	F	F	S	I	V	G	Q	P	M	C	V	L	L	Y	Y	H	D	-	-	V	Ta	DGAT1

Motif #8

FIG. 8G

28/28

.....																									Consensus:	<u>SEQ</u> <u>ID NO:</u>					
506	Q	A	K	Y	G	S	V	S	K	M	T	T	S	Q	Q	L	V	Q	Q	G	Q	G	T	C	P	P	L	V	Nc	DAGAT:	19
485	Q	A	K	Y	G	S	V	T	D	S	G	F	S	I	S														Fm	DAGAT:	20
519	T	K	R	H	M	D	V																						Ma	DGAT1	18
483	Q	A	K	Y	G	S	V	S	K	M	G	Y	A	T	S	K	A	A	L	T	N								Mg	DAGAT:	21
447	Q	A	K	Y	G	S	V	S	Q	T	H	P																	An	DAGAT:	22
521	N	Y	K	Q	N	Q																							Yl	DGAT1	14
490	L	N	Y	D	A	P	V	G	V	.																			Mm	DGAT1	129
497	M	N	R	K	G	K	L	D																					Gm	DGAT1	130
513	M	N	R	K	G	S	M	S																					At	DGAT1	131
493	M	N	R	I	E	K	A	R																					Os	DGAT1	132
527	M	N	R	K	A	S	A	R																					Pfdgat1	133	
500	M	N	R	Q	A	Q	T	N	G																				Ta	DGAT1	134

FIG. 8H

