



- (51) **International Patent Classification:**
C12N 15/81 (2006.01) *C12N 9/00* (2006.01)
- (21) **International Application Number:**
PCT/US2012/031909
- (22) **International Filing Date:**
2 April 2012 (02.04.2012)
- (25) **Filing Language:** English
- (26) **Publication Language:** English
- (30) **Priority Data:**
61/471,736 5 April 2011 (05.04.2011) US
- (71) **Applicant (for all designated States except US):** **E. I. DU PONT DE NEMOURS AND COMPANY** [US/US];
1007 Market Street, Wilmington, Delaware 19898 (US).
- (72) **Inventor; and**
- (75) **Inventor/Applicant (for US only):** **ZHU, Quinn Qun** [US/US]; 544 Revere Road, West Chester, Pennsylvania 19382 (US).
- (74) **Agent:** **CHRISTENBURY, Lynne, M.**; E. I. du Pont de Nemours and Company, Legal Patent Records Center, 4417 Lancaster Pike, Wilmington, Delaware 19805 (US).

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

(84) **Designated States** (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

- with international search report (Art. 21(3))
- with sequence listing part of description (Rule 5.2(a))

(54) **Title:** YARROWIA N-ALKANE-HYDROXYLATING CYTOCHROME P450 PROMOTER REGIONS FOR GENE EXPRESSION IN YEAST

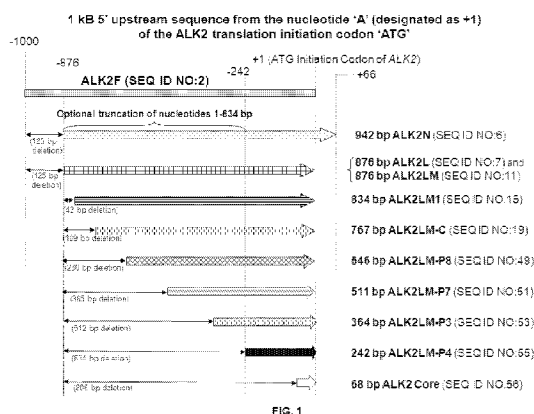


FIG. 1

(57) **Abstract:** Promoter regions associated with the *Yarrowia lipolytica* n-alkane-hydroxylating cytochrome P450 (*ALK2*) gene are disclosed and have been found to be particularly effective for the expression of heterologous genes in yeast. These promoter regions will be useful for driving high-level expression of genes involved in the production of omega-3 and omega-6 fatty acids.

YARROWIA N-ALKANE-HYDROXYLATING CYTOCHROME P450 PROMOTER REGIONS FOR GENE EXPRESSION IN YEAST

This application claims the benefit of U.S. Provisional Application No.
5 61/471,736, filed April 5, 2011, which is incorporated herein by reference in its
entirety.

FIELD OF THE INVENTION

This invention is in the field of biotechnology. More specifically, this
invention pertains to n-alkane-hydroxylating cytochrome P450 ["ALK2"] promoter
10 regions derived from *Yarrowia lipolytica* that are useful for gene expression in
yeast.

BACKGROUND OF THE INVENTION

Oleaginous yeast are defined as those organisms that are naturally
capable of oil synthesis and accumulation, wherein oil accumulation ranges from
15 at least about 25% up to about 80% of the dry cell weight. The technology for
growing oleaginous yeast with high oil content is well developed (for example,
see EP 0 005 277B1; Ratledge, C., *Prog. Ind. Microbiol.*, 16:119-206 (1982)).
And, these organisms have been commercially used for a variety of purposes in
the past.

20 Recently, the natural abilities of oleaginous yeast have been enhanced by
advances in genetic engineering, resulting in organisms capable of producing
polyunsaturated fatty acids ["PUFAs"], carotenoids, resveratrol and sterols. For
example, significant efforts by Applicants' Assignee have demonstrated that
Yarrowia lipolytica can be engineered for production of omega-3 and omega-6
25 fatty acids, by introducing and expressing genes encoding the omega-3/omega-6
biosynthetic pathway (U.S. Patent 7,238,482; U.S. Patent 7,465,564; U.S. Patent
7,550,286; U.S. Patent 7,588,931; and U.S. Patent No. 7,932,077; U.S. Pat.
Appl. Publ. Nos. 2009-0093543-A1 and 2010-0317072-A1).

Recombinant production of any heterologous protein is generally
30 accomplished by constructing an expression cassette in which the DNA coding
for the protein of interest is placed under the control of a promoter suitable for
the host cell. The expression cassette is then introduced into the host cell (i.e.,

usually by plasmid-mediated transformation or targeted integration into the host genome) and production of the heterologous protein is achieved by culturing the transformed host cell under conditions necessary for the proper function of the promoter contained within the expression cassette. Thus, the development of new host cells (e.g., transformed yeast) for recombinant production of proteins generally requires the availability of promoters that are suitable for controlling the expression of a protein of interest in the host cell.

A variety of strong promoters have been isolated from *Yarrowia lipolytica* that are useful for heterologous gene expression in yeast, as shown in the Table below.

Table 1. Characterized *Yarrowia lipolytica* Promoters

Promoter Name	Native Gene	Reference
XPR2	alkaline extracellular protease	U.S. Pat. No. 4,937,189; EP220864
TEF	translation elongation factor EF1- α (<i>tef</i>)	U.S. Pat. No. 6,265,185
GPD, GPM	glyceraldehyde-3-phosphate-dehydrogenase (<i>gpd</i>), phosphoglycerate mutase (<i>gpm</i>)	U.S. Pat. Nos. 7,259,255 and 7,459,546; U.S. Pat. Appl. Publ. No. 2011-0059496-A1
GPDIN	glyceraldehyde-3-phosphate-dehydrogenase (<i>gpd</i>)	U.S. Pat. No. 7,459,546
GPM/FBAIN	chimeric phosphoglycerate mutase (<i>gpm</i>)/ fructose-bisphosphate aldolase (<i>fba1</i>)	U.S. Pat. No. 7,202,356
FBA, FBAIN, FBAINm	fructose-bisphosphate aldolase (<i>fba1</i>)	U.S. Pat. No. 7,202,356
GPAT	glycerol-3-phosphate O-acyltransferase (<i>gpat</i>)	U.S. Pat. No. 7,264,949
YAT1	ammonium transporter enzyme (<i>yat1</i>)	U.S. Pat. Appl. Publ. Nos. 2006-0094102-A1 and 2010-0068789-A1
EXP1	export protein	U.S. Pat. No. 7,932,077

Additionally, Juretzek et al. (*Biotech. Bioprocess Eng.*, 5:320-326 (2000)) compares the glycerol-3-phosphate dehydrogenase ["G3P"], isocitrate lyase ["ICL1"], 3-oxo-acyl-CoA thiolase ["POT1"] and acyl-CoA oxidase ["POX1", "POX2" and "POX5"] promoters with respect to their regulation and activities during growth on different carbon sources.

Despite the utility of these known promoters, however, there is a need for new improved yeast promoters for metabolic engineering of yeast (i.e., oleaginous and non-oleaginous) and for controlling the expression of heterologous genes in yeast. Furthermore, possession of a suite of promoters that can be regulated under a variety of natural growth and induction conditions in yeast will play an important role in industrial settings, wherein economical production of heterologous and/or homologous polypeptides in commercial quantities is desirable.

It is believed that promoter regions derived from the *Yarrowia lipolytica* gene encoding n-alkane-hydroxylating cytochrome P450 ["ALK2"] will be useful in expressing heterologous and/or homologous genes in transformed yeast, including *Yarrowia*.

SUMMARY OF THE INVENTION

In a first embodiment, the invention concerns a method for expressing a coding region of interest in a transformed yeast cell comprising:

a) providing a transformed yeast cell having a recombinant construct, wherein the recombinant construct comprises:

(1) a promoter region of an *ALK2 Yarrowia* gene; and

(2) a coding region of interest which is expressible in the yeast cell;

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

b) growing the transformed yeast cell of step (a) under conditions whereby the recombinant construct of step (a) is expressed.

In a second embodiment, the invention concerns a method for the production of an omega-3 fatty acid or omega-6 fatty acid comprising:

a) providing a transformed oleaginous yeast cell comprising a recombinant construct, wherein the recombinant construct comprises:

i) a promoter region of an *ALK2 Yarrowia* gene; and

ii) a coding region encoding at least one omega-3 fatty acid or omega-6 fatty acid biosynthetic pathway enzyme;

wherein the promoter region and the coding region are operably linked;

- b) growing the transformed oleaginous yeast of step (a) under conditions whereby the at least one omega-3 fatty acid or omega-6 fatty acid biosynthetic pathway enzyme is expressed and the omega-3 fatty acid or the omega-6 fatty acid is produced; and
- c) optionally recovering the omega-3 fatty acid or the omega-6 fatty acid.

In another aspect, the promoter region of a *Yarrowia* gene may comprise SEQ ID NO:56.

- 10 In yet another aspect, the promoter region of the ALK2 *Yarrowia* gene further comprises an enhancer region set forth in SEQ ID NO:57, the enhancer region being operably linked to a functional yeast promoter.

- In some embodiments, the promoter region of an ALK2 *Yarrowia* gene may be as set forth in SEQ ID NO:6, wherein said promoter optionally comprises
- 15 at least one modification selected from the group consisting of:

- a) a deletion at the 5'-terminus of 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177, 178, 179, 180, 181, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 197, 198, 199, 200, 201, 202, 203, 204, 205, 206, 207, 208, 209, 210, 211, 212, 213, 214, 215, 216, 217, 218, 219, 220, 221, 222, 223, 224, 225, 226, 227, 228, 229, 230, 231, 232, 233, 234, 235, 236, 237, 238, 239, 240, 241, 242, 243, 244, 245, 246, 247, 248, 249, 250, 251, 252, 253, 254, 255, 256,

- 257, 258, 259, 260, 261, 262, 263, 264, 265, 266, 267, 268, 269, 270, 271, 272, 273, 274, 275, 276, 277, 278, 279, 280, 281, 282, 283, 284, 285, 286, 287, 288, 289, 290, 291, 292, 293, 294, 295, 296, 297, 298, 299, 300, 301, 302, 303, 304, 305, 306, 307, 308, 309, 310, 311, 312, 313, 314, 315, 316, 317, 318, 319, 320, 321, 322, 323, 324, 325, 326, 327, 328, 329, 330, 331, 332, 333, 334, 335, 336, 337, 338, 339, 340, 341, 342, 343, 344, 345, 346, 347, 348, 349, 350, 351, 352, 353, 354, 355, 356, 357, 358, 359, 360, 361, 362, 363, 364, 365, 366, 367, 368, 369, 370, 371, 372, 373, 374, 375, 376, 377, 378, 379, 380, 381, 382, 383, 384, 385, 386, 387, 388, 389, 390, 391, 392, 393, 394, 395, 396, 397, 398, 399, 400, 401, 402, 403, 404, 405, 406, 407, 408, 409, 410, 411, 412, 413, 414, 415, 416, 417, 418, 419, 420, 421, 422, 423, 424, 425, 426, 427, 428, 429, 430, 431, 432, 433, 434, 435, 436, 437, 438, 439, 440, 441, 442, 443, 444, 445, 446, 447, 448, 449, 450, 451, 452, 453, 454, 455, 456, 457, 458, 459, 460, 461, 462, 463, 464, 465, 466, 467, 468, 469, 470, 471, 472, 473, 474, 475, 476, 477, 478, 479, 480, 481, 482, 483, 484, 485, 486, 487, 488, 489, 490, 491, 492, 493, 494, 495, 496, 497, 498, 499, 500, 501, 502, 503, 504, 505, 506, 507, 508, 509, 510, 511, 512, 513, 514, 515, 516, 517, 518, 519, 520, 521, 522, 523, 524, 525, 526, 527, 528, 529, 530, 531, 532, 533, 534, 535, 536, 537, 538, 539, 540, 541, 542, 543, 544, 545, 546, 547, 548, 549, 550, 551, 552, 553, 554, 555, 556, 557, 558, 559, 560, 561, 562, 563, 564, 565, 566, 567, 568, 569, 570, 571, 572, 573, 574, 575, 576, 577, 578, 579, 580, 581, 582, 583, 584, 585, 586, 587, 588, 589, 590, 591, 592, 593, 594, 595, 596, 597, 598, 599, 600, 601, 602, 603, 604, 605, 606, 607, 608, 609, 610, 611, 612, 613, 614, 615, 616, 617, 618, 619, 620, 621, 622, 623, 624, 625, 626, 627, 628, 629, 630, 631, 632, 633, or 634 consecutive nucleotides, wherein the first nucleotide deleted is the adenine nucleotide ['A'] at position 1 of SEQ ID NO:6;
- b) a deletion at the 3'-terminus of 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49,

50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, or 66 consecutive nucleotides, wherein the first nucleotide deleted is the cytosine ['C'] nucleotide at position 942 of SEQ ID NO:6;

- c) substitution of a guanine ['G'] nucleotide for the cytosine ['C'] nucleotide at position 753 of SEQ ID NO:6;
- d) substitution of an adenine ['A'] nucleotide or a thymine ['T'] nucleotide for the cytosine ['C'] nucleotide at position 753 of SEQ ID NO:6;
- e) substitution of the nucleotide sequence 'AA' for the nucleotide sequence 'TT' at positions 41 to 42 of SEQ ID NO:6;
- f) insertion of the nucleotide sequence 'TCG' between position 109 and 110 of SEQ ID NO:6;
- g) insertion of the nucleotide sequence 'AAAT' between position 131 and position 132 of SEQ ID NO:6;
- h) substitution of the nucleotide sequence 'AAA' for the nucleotide sequence 'TT' at positions 76 to 77 of SEQ ID NO:6;
- i) substitution of the nucleotide sequence 'GTTT' for the nucleotide sequence 'AAAA' at positions 631 to 634 of SEQ ID NO:6;
- j) insertion of the nucleotide sequence 'TA' between position 512 and position 513 of SEQ ID NO:6;
- k) insertion of the nucleotide sequence 'TTTA' between position 365 and position 366 of SEQ ID NO:6;
- l) insertion of the nucleotide sequence 'TTAAA' between position 230 and position 231 of SEQ ID NO:6; and
- m) any combination of part a), part b), part c), part d), part e), part f), part g), part h), part i), part j), part k), and part l) above.

More preferably, the promoter region of an *ALK2 Yarrowia* gene may be as set forth in SEQ ID NO:55, wherein said promoter comprises at least one modification selected from the group consisting of:

- a) a deletion at the 5'-terminus of 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66,

- 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84,
85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101,
102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114,
115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127,
128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140,
141, 142, 143, 144, 145, 146, 147, 148, 149, 150, 151, 152, 153,
154, 155, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165, 166,
167, 168, 169, 170, 171, 172, 173, or 174 consecutive nucleotides,
wherein the first nucleotide deleted is the adenine nucleotide ['A'] at
position 1 of SEQ ID NO:55; and
- b) a deletion of part (a) in combination with a substitution of a thymine
['T'] nucleotide, an adenine ['A'] nucleotide, or a cytosine ['C']
nucleotide for the guanine ['G'] nucleotide at position 119 of SEQ
ID NO:55.

The promoter region of an *ALK2 Yarrowia* gene may be selected from the
group consisting of SEQ ID NO:7, SEQ ID NO:11, SEQ ID NO:27, SEQ ID
NO:15, SEQ ID NO:47, SEQ ID NO:43, SEQ ID NO:39, SEQ ID NO:35, SEQ ID
NO:31, SEQ ID NO:19, SEQ ID NO:49, SEQ ID NO:51, SEQ ID NO:53, and
SEQ ID NO:55.

In various embodiments of the methods of the invention, the transformed
yeast cell is an oleaginous yeast. This oleaginous yeast may be a member of a
genus selected from the group consisting of *Yarrowia*, *Candida*, *Rhodotorula*,
Rhodospiridium, *Cryptococcus*, *Trichosporon* and *Lipomyces*.

Additionally, provided herein is an isolated nucleic acid molecule
comprising a promoter region of an *ALK2 Yarrowia* selected from the group
consisting of:

- (a) SEQ ID NO:7;
- (b) SEQ ID NO:11;
- (c) SEQ ID NO:27;
- (d) SEQ ID NO:15;
- (e) SEQ ID NO:47;
- (f) SEQ ID NO:43;

- (g) SEQ ID NO:39;
(h) SEQ ID NO:35;
(i) SEQ ID NO:31;
(j) SEQ ID NO:19;
5 (k) SEQ ID NO:49;
(l) SEQ ID NO:51;
(m) SEQ ID NO:53;
(n) SEQ ID NO:6, wherein said promoter optionally comprises at least one
modification selected from the group consisting of:
- 10 (i) a deletion at the 5'-terminus of 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13,
14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31,
32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49,
50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67,
68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85,
15 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102,
103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115,
116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128,
129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141,
142, 143, 144, 145, 146, 147, 148, 149, 150, 151, 152, 153, 154,
20 155, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167,
168, 169, 170, 171, 172, 173, 174, 175, 176, 177, 178, 179, 180,
181, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193,
194, 195, 196, 197, 198, 199, 200, 201, 202, 203, 204, 205, 206,
207, 208, 209, 210, 211, 212, 213, 214, 215, 216, 217, 218, 219,
25 220, 221, 222, 223, 224, 225, 226, 227, 228, 229, 230, 231, 232,
233, 234, 235, 236, 237, 238, 239, 240, 241, 242, 243, 244, 245,
246, 247, 248, 249, 250, 251, 252, 253, 254, 255, 256, 257, 258,
259, 260, 261, 262, 263, 264, 265, 266, 267, 268, 269, 270, 271,
272, 273, 274, 275, 276, 277, 278, 279, 280, 281, 282, 283, 284,
30 285, 286, 287, 288, 289, 290, 291, 292, 293, 294, 295, 296, 297,
298, 299, 300, 301, 302, 303, 304, 305, 306, 307, 308, 309, 310,
311, 312, 313, 314, 315, 316, 317, 318, 319, 320, 321, 322, 323,

324, 325, 326, 327, 328, 329, 330, 331, 332, 333, 334, 335, 336,
 337, 338, 339, 340, 341, 342, 343, 344, 345, 346, 347, 348, 349,
 350, 351, 352, 353, 354, 355, 356, 357, 358, 359, 360, 361, 362,
 363, 364, 365, 366, 367, 368, 369, 370, 371, 372, 373, 374, 375,
 5 376, 377, 378, 379, 380, 381, 382, 383, 384, 385, 386, 387, 388,
 389, 390, 391, 392, 393, 394, 395, 396, 397, 398, 399, 400, 401,
 402, 403, 404, 405, 406, 407, 408, 409, 410, 411, 412, 413, 414,
 415, 416, 417, 418, 419, 420, 421, 422, 423, 424, 425, 426, 427,
 428, 429, 430, 431, 432, 433, 434, 435, 436, 437, 438, 439, 440,
 10 441, 442, 443, 444, 445, 446, 447, 448, 449, 450, 451, 452, 453,
 454, 455, 456, 457, 458, 459, 460, 461, 462, 463, 464, 465, 466,
 467, 468, 469, 470, 471, 472, 473, 474, 475, 476, 477, 478, 479,
 480, 481, 482, 483, 484, 485, 486, 487, 488, 489, 490, 491, 492,
 493, 494, 495, 496, 497, 498, 499, 500, 501, 502, 503, 504, 505,
 15 506, 507, 508, 509, 510, 511, 512, 513, 514, 515, 516, 517, 518,
 519, 520, 521, 522, 523, 524, 525, 526, 527, 528, 529, 530, 531,
 532, 533, 534, 535, 536, 537, 538, 539, 540, 541, 542, 543, 544,
 545, 546, 547, 548, 549, 550, 551, 552, 553, 554, 555, 556, 557,
 558, 559, 560, 561, 562, 563, 564, 565, 566, 567, 568, 569, 570,
 20 571, 572, 573, 574, 575, 576, 577, 578, 579, 580, 581, 582, 583,
 584, 585, 586, 587, 588, 589, 590, 591, 592, 593, 594, 595, 596,
 597, 598, 599, 600, 601, 602, 603, 604, 605, 606, 607, 608, 609,
 610, 611, 612, 613, 614, 615, 616, 617, 618, 619, 620, 621, 622,
 623, 624, 625, 626, 627, 628, 629, 630, 631, 632, 633, or 634
 25 consecutive nucleotides, wherein the first nucleotide deleted is the
 adenine nucleotide ['A'] at position 1 of SEQ ID NO:6;
 (ii) a deletion at the 3'-terminus of 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12,
 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30,
 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48,
 30 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, or
 66 consecutive nucleotides, wherein the first nucleotide deleted is
 the cytosine ['C'] nucleotide at position 942 of SEQ ID NO:6;

- (iii) substitution of a guanine ['G'] nucleotide for the cytosine ['C'] nucleotide at position 753 of SEQ ID NO:6;
- (iv) substitution of an adenine ['A'] nucleotide or a thymine ['T'] nucleotide for the cytosine ['C'] nucleotide at position 753 of SEQ ID NO:6;
- (v) substitution of the nucleotide sequence 'AA' for the nucleotide sequence 'TT' at position 41 and 42 of SEQ ID NO:6;
- (vi) insertion of the nucleotide sequence 'TCG' between position 109 and 110 of SEQ ID NO:6;
- (vii) insertion of the nucleotide sequence 'AAAT' between position 131 and position 132 of SEQ ID NO:6;
- (viii) substitution of the nucleotide sequence 'AAA' for the nucleotide sequence 'TT' at positions 76 to 77 of SEQ ID NO:6;
- (ix) substitution of the nucleotide sequence 'GTTT' for the nucleotide sequence 'AAAA' at positions 631 to 634 of SEQ ID NO:6;
- (x) insertion of the nucleotide sequence 'TA' between position 512 and position 513 of SEQ ID NO:6;
- (xi) insertion of the nucleotide sequence 'TTTA' between position 365 and position 366 of SEQ ID NO:6;
- (xii) insertion of the nucleotide sequence 'TTAAA' between position 230 and position 231 of SEQ ID NO:6; and
- (xiii) any combination of part (i), part (ii), part (iii), part (iv), part (v), part (vi), part (vii), part (viii), part (ix), part (x), part (xi), and part (xii) above; and
- (o) a promoter region comprising SEQ ID NO:55.

Also provided herein is an isolated nucleic acid molecule comprising a promoter region of an *ALK2 Yarrowia* gene, wherein the promoter region of the *ALK2 Yarrowia* gene further comprises an enhancer region set forth in SEQ ID NO:57, the enhancer region being operably linked to a functional yeast promoter.

BRIEF DESCRIPTION OF THE DRAWINGS AND SEQUENCE DESCRIPTIONS

FIG. 1 graphically represents the relationship between SEQ ID NOs:2, 6, 7, 11, 15, 19, 49, 51, 53, 55, and 56, each of which relates to promoter regions derived from the 5' upstream region of the n-alkane-hydroxylating cytochrome P450 ["ALK2"] gene in *Yarrowia lipolytica*.

5 FIGs. 2A, 2B, 2C, 2D, 2E, 2F, 2G, 2H, 2I, and 2J (which should be viewed together as FIG. 2) provide an alignment of:

- 10 (a) the *Y. lipolytica* ALK2 (SEQ ID NO:2) promoter region, which is the 1000 bp 5' upstream sequence (i.e., the -1000 to -1 region) of the n-alkane-hydroxylating cytochrome P450 ["ALK2"] gene in *Y. lipolytica*, wherein the nucleotide 'A' of the ALK2 translation initiation codon 'ATG' was designated as +1 (note that the ATG codon is shown in SEQ ID NO:6);
- 15 (b) the 942 bp ALK2N (SEQ ID NO:6) promoter region, which comprises the 876 bp 5' upstream sequence (i.e., the -876 to -1 region) of the n-alkane-hydroxylating cytochrome P450 ["ALK2"] gene in *Y. lipolytica* and the first 66 bp of the coding region, wherein the nucleotide 'A' of the ALK2 translation initiation codon 'ATG' was designated as +1;
- 20 (c) the 876 bp ALK2L (SEQ ID NO:7) promoter region;
- (d) the 876 bp ALK2LM (SEQ ID NO:11) promoter region;
- (e) the 838 bp ALK2LM1-S (SEQ ID NO:27) promoter region;
- (f) the 834 bp ALK2LM1 (SEQ ID NO:15) promoter region;
- (g) the 840 bp ALK2LM1-PP6 (SEQ ID NO:47) promoter region;
- (h) the 839 bp ALK2LM1-PP5 (SEQ ID NO:43) promoter region;
- (i) the 837 bp ALK2LM1-PP2 (SEQ ID NO:39) promoter region;
- 25 (j) the 835 bp ALK2LM1-PP1 (SEQ ID NO:35) promoter region;
- (k) the 835 bp ALK2LM1-P (SEQ ID NO:31) promoter region;
- (l) the 767 bp ALK2LM-C (SEQ ID NO:19) promoter region;
- (m) the 646 bp ALK2LM-P8 (SEQ ID NO:49) promoter region;
- (n) the 511 bp ALK2LM-P7 (SEQ ID NO:51) promoter region;
- 30 (o) the 364 bp ALK2LM-P3 (SEQ ID NO:53) promoter region;
- (p) the 242 bp ALK2LM-P4 (SEQ ID NO:55) promoter region; and
- (q) the 68 bp minimal (Core) ALK2 (SEQ ID NO:56) promoter region.

Base pair differences are highlighted with an arrow and box.

FIG. 3 illustrates the omega-3/omega-6 fatty acid biosynthetic pathway.

FIG. 4 provides plasmid maps for the following: (A) pDMW212 and (B) pAKL2GUS.

5 FIG. 5 is an image of cell cultures comparing the promoter activity of 834 bp AKL2LM1 (SEQ ID NO:15) and FBA in *Yarrowia lipolytica* as determined by histochemical staining.

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

SEQ ID NOs:1-58 are promoters, ORFs encoding genes (or portions thereof), primers, or plasmids, as identified in Table 2.

Table 2. Summary Of Nucleic Acid SEQ ID Numbers

Description	Nucleic acid SEQ ID NO.
<i>Yarrowia</i> ALK2 gene ("YALI0F01320g")	1 (2872 bp)
1000 bp ALK2F <i>Yarrowia</i> promoter region	2 (1000 bp)
Primer Y1198	3 (32 bp)
Primer Y1199	4 (33 bp)
Plasmid pT-ALK2Pro	5 (4906 bp)
942 bp ALK2N <i>Yarrowia</i> promoter region	6 (942 bp)
876 bp ALK2L <i>Yarrowia</i> promoter region	7 (876 bp)
Plasmid pT-ALK2Pro-(S)	8 (4906 bp)
Primer Y1202	9 (35 bp)
Primer Y1203	10 (35 bp)
876 bp ALK2LM <i>Yarrowia</i> promoter region	11 (876 bp)
Plasmid pT-ALK2Pro-(S)S	12 (4906 bp)
Primer Y1208	13 (36 bp)
Primer Y1209	14 (36 bp)
834 bp ALK2LM1 <i>Yarrowia</i> promoter region	15 (834 bp)
Plasmid pT-ALK2Pro-(S)C	16 (4909 bp)
Primer Y1262	17 (37 bp)
Primer Y1263	18 (37 bp)
767 bp ALK2LM-C <i>Yarrowia</i> promoter region	19 (767 bp)
Plasmid pALK2GUS	20 (9415 bp)
Plasmid pDMW212	21 (9348 bp)
Plasmid pDG2LGUS (U.S. Appl. No. 61/469,933, filed March 31, 2011)	22 (9499 bp)
Plasmid pALK2GUS-C	23 (9364 bp)
Primer Y2148	24 (38 bp)

Primer Y2149	25 (38 bp)
Plasmid pALK2GUS-S	26 (9419 bp)
838 bp ALK2LM1-S <i>Yarrowia</i> promoter region	27 (838 bp)
Primer Y1244	28 (38 bp)
Primer Y1245	29 (38 bp)
Plasmid pALK2GUS-P	30 (9414 bp)
835 bp ALK2LM1-P <i>Yarrowia</i> promoter region	31 (835 bp)
Primer Y1250	32 (38 bp)
Primer Y1251	33 (38 bp)
Plasmid pALK2GUS-P1	34 (9414 bp)
835 bp ALK2LM1-PP1 <i>Yarrowia</i> promoter region	35 (835 bp)
Primer Y2158	36 (36 bp)
Primer Y2159	37 (36 bp)
Plasmid pALK2GUS-P2	38 (9416 bp)
837 bp ALK2LM1-PP2 <i>Yarrowia</i> promoter region	39 (837 bp)
Primer Y2170	40 (43 bp)
Primer Y2171	41 (43 bp)
Plasmid pALK2GUS-P5	42 (9418 bp)
839 bp ALK2LM1-PP5 <i>Yarrowia</i> promoter region	43 (839 bp)
Primer Y2172	44 (42 bp)
Primer Y2173	45 (42 bp)
Plasmid pALK2GUS-P6	46 (9419 bp)
840 bp ALK2LM1-PP6 <i>Yarrowia</i> promoter region	47 (840 bp)
Plasmid pALK2GUS-P8	48 (9261 bp)
646 bp ALK2LM-P8 <i>Yarrowia</i> promoter region	49 (646 bp)
Plasmid pALK2GUS-P7	50 (9124 bp)
511 bp ALK2LM-P7 <i>Yarrowia</i> promoter region	51 (511 bp)
Plasmid pALK2GUS-P3	52 (8977 bp)
364 bp ALK2LM-P3 <i>Yarrowia</i> promoter region	53 (364 bp)
Plasmid pALK2GUS-P4	54 (8854 bp)
242 bp ALK2LM-P4 <i>Yarrowia</i> promoter region	55 (242 bp)
68 bp minimal ALK2 <i>Yarrowia</i> promoter	56 (68 bp)
147 bp ALK2 <i>Yarrowia</i> enhancer region	57 (147 bp)
Consensus sequence [A(A _{rich}) ₅ NYA(A/T)NN(A _{rich}) ₆] located in promoter sequences of <i>S. cerevisiae</i> genes	58 (18 bp)

DETAILED DESCRIPTION OF THE INVENTION

All patents, patent applications, and publications cited herein are hereby incorporated by reference in their entirety.

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

“Open reading frame” is abbreviated “ORF”.

“Polymerase chain reaction” is abbreviated “PCR”.

“American Type Culture Collection” is abbreviated “ATCC”.

“Polyunsaturated fatty acid(s)” is abbreviated “PUFA(s)”.

“Triacylglycerols” are abbreviated “TAGs”.

- 5 The term “yeast” refers to a phylogenetically diverse grouping of single-celled fungi. Yeast do not form a specific taxonomic or phylogenetic grouping, but instead comprise a diverse assemblage of unicellular organisms that occur in the Ascomycotina and Basidiomycotina. Collectively, about 100 genera of yeast have been identified, comprising approximately 1,500 species (Kurtzman and
- 10 Fell, Yeast Systematics And Phylogeny: Implications Of Molecular Identification Methods For Studies In Ecology. In C. A. Rosa and G. Peter, eds., *The Yeast Handbook*. Germany: Springer-Verlag Berlin Herdelberg, 2006). Yeast reproduce principally by budding (or fission) and derive energy from fermentation, via conversion of carbohydrates to ethanol and carbon dioxide.
- 15 Examples of some yeast genera include, but are not limited to: *Agaricostilbum*, *Ambrosiozyma*, *Arthroascus*, *Arxula*, *Ashbya*, *Babjevia*, *Bensingtonia*, *Botryozyma*, *Brettanomyces*, *Bullera*, *Candida*, *Clavispora*, *Cryptococcus*, *Cystofilobasidium*, *Debaryomyces*, *Dekkera*, *Dipodascus*, *Endomyces*, *Endomycopsella*, *Erythrobasidium*, *Fellomyces*, *Filobasidium*, *Galactomyces*,
- 20 *Geotrichum*, *Guilliermondella*, *Hansenula*, *Hanseniaspora*, *Kazachstania*, *Kloeckera*, *Kluyveromyces*, *Kockovaella*, *Kodamaea*, *Komagataella*, *Kondoa*, *Lachancea*, *Leucosporidium*, *Leucosporidiella*, *Lipomyces*, *Lodderomyces*, *Issatchenkia*, *Magnusiomyces*, *Mastigobasidium*, *Metschnikowia*, *Monosporella*, *Myxozyma*, *Nadsonia*, *Nematospora*, *Oosporidium*, *Pachysolen*, *Pichia*, *Phaffia*,
- 25 *Pseudozyma*, *Reniforma*, *Rhodospiridium*, *Rhodotorula*, *Saccharomyces*, *Saccharomycodes*, *Saccharomycopsis*, *Saturnispora*, *Schizoblastosporion*, *Schizosaccharomyces*, *Sirobasidium*, *Smithiozyma*, *Sporobolomyces*, *Sporopachydermia*, *Starmerella*, *Sympodiomyces*, *Sympodiomyces*, *Torulaspora*, *Tremella*, *Trichosporon*, *Trichosporiella*, *Trigonopsis*, *Udeniomyces*,
- 30 *Wickerhamomyces*, *Williopsis*, *Xanthophyllomyces*, *Yarrowia*, *Zygosaccharomyces*, *Zygotorulaspora*, *Zymoxenogloea* and *Zygozyma*.

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

10 The term "oleaginous yeast" refers to those microorganisms classified as yeasts that can make oil. Examples of oleaginous yeast include, but are not limited to, the following genera: *Yarrowia*, *Candida*, *Rhodotorula*, *Rhodospiridium*, *Cryptococcus*, *Trichosporon* and *Lipomyces*. Alternatively, organisms classified as yeasts that are genetically modified to become
15 oleaginous such that they can produce more than 25% of their dry cell weight as oil are also "oleaginous", e.g., yeast such as *Saccharomyces cerevisiae* (Int'l. App. Pub. No. WO 2006/102342).

 The term "fermentable carbon source" will refer to a carbon source that a microorganism will metabolize to derive energy. Typical carbon sources for use
20 in the methods herein include, but are not limited to: monosaccharides, disaccharides, oligosaccharides, polysaccharides, alkanes, fatty acids, esters of fatty acids, glycerol, monoglycerides, diglycerides, triglycerides, carbon dioxide, methanol, formaldehyde, formate and carbon-containing amines. Most preferred is glucose, sucrose, invert sucrose, fructose, glycerol and/or fatty acids
25 containing between 10-22 carbons. The term "invert sucrose" (or "invert sugar") refers to a mixture comprising equal parts of fructose and glucose resulting from the hydrolysis of sucrose. Invert sucrose may be a mixture comprising 25 to 50% glucose and 25 to 50% fructose. Invert sucrose may also comprise sucrose, the amount of which depends on the degree of hydrolysis.

30 The term "ALK2" refers to an n-alkane-hydroxylating cytochrome P450 ("YALI0F01320g", Dujon, B. et al., *Nature*, 430(6995):35-44 (2004)) encoded by

the *ALK2* gene. *ALK2* transcription was increased in *Yarrowia* strains with their native *snf1* gene knocked-out (U.S. Pat. Appl. Publ. No. 2010-0062502-A1). This enzyme catalyzes the terminal hydroxylation of n-alkanes.

An “*ALK2 Yarrowia* gene” refers to a gene encoding *ALK2* from a yeast of the genus *Yarrowia*. For example, a 2872 bp DNA sequence (“YALI0F01320g”) that encodes an n-alkane-hydroxylating cytochromes P450 is provided as SEQ ID NO:1 (Dujon, B. et al., *Nature*, 430(6995):35-44 (2004)).

The term “promoter region of an *ALK2 Yarrowia* gene” or “*Yarrowia* *ALK2* promoter region” refers to the 5' upstream untranslated region in front of the ‘ATG’ translation initiation codon of a *Yarrowia* *ALK2* gene, or sequences derived therefrom, and that is necessary for expression. Thus, it is believed that promoter regions of an *ALK2 Yarrowia* gene will comprise a portion of the ~1000 bp 5' upstream of an *ALK2 Yarrowia* gene. The sequence of the *Yarrowia* *ALK2* promoter region may correspond exactly to native sequence upstream of the *ALK2 Yarrowia* gene (i.e., a “wildtype” or “native” *Yarrowia* *ALK2* promoter); alternately, the sequence of the *Yarrowia* *ALK2* promoter region may be “modified” or “mutated”, thereby comprising various substitutions, deletions, and/or insertions of one or more nucleotides relative to a wildtype or native *Yarrowia* *ALK2* promoter. These modifications can result in a modified *Yarrowia* *ALK2* promoter having increased, decreased or equivalent promoter activity, when compared to the promoter activity of the corresponding wildtype or native *Yarrowia* *ALK2* promoter. The term “mutant promoter” or “modified promoter” will encompass natural variants and *in vitro* generated variants obtained using methods well known in the art (e.g., classical mutagenesis, site-directed mutagenesis and “DNA shuffling”).

Described herein is a wildtype *Yarrowia* *ALK2* promoter region (SEQ ID NO:2 comprising the -1000 to -1 upstream region of the *ALK2* gene SEQ ID NO:1) based on nucleotide numbering such that the ‘A’ position of the ‘ATG’ translation initiation codon is designated as +1. The ATG translation initiation codon is located at nucleotide positions 1001-1003 in SEQ ID NO:1. Alternately, and yet by no means limiting in nature, a wildtype *Yarrowia* *ALK2* promoter region may comprise the -876 to -1 region of SEQ ID NO:1, the -840 to -1 region

of SEQ ID NO:1, the -839 to -1 region of SEQ ID NO:1, the -838 to -1 region of SEQ ID NO:1, the -837 to -1 region of SEQ ID NO:1, the -835 to -1 region of SEQ ID NO:1, the -834 to -1 region of SEQ ID NO:1; the -767 to -1 region of SEQ ID NO:1; the -646 to -1 region of SEQ ID NO:1; the -511 to -1 region of SEQ ID NO:1; the -364 to -1 region of SEQ ID NO:1; the -242 to -1 region of SEQ ID NO:1, or the -68 to -1 region of SEQ ID NO:1 (where the “-1” position in SEQ ID NO:1 is the nucleotide that is 5'-adjacent to the ATG translation initiation codon). Similarly, a modified *Yarrowia* ALK2 promoter region may comprise the promoter region of an *ALK2 Yarrowia* gene as set forth in SEQ ID NO:6, wherein said promoter optionally comprises at least one modification selected from the group consisting of:

- a) a deletion at the 5'-terminus of 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177, 178, 179, 180, 181, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 197, 198, 199, 200, 201, 202, 203, 204, 205, 206, 207, 208, 209, 210, 211, 212, 213, 214, 215, 216, 217, 218, 219, 220, 221, 222, 223, 224, 225, 226, 227, 228, 229, 230, 231, 232, 233, 234, 235, 236, 237, 238, 239, 240, 241, 242, 243, 244, 245, 246, 247, 248, 249, 250, 251, 252, 253, 254, 255, 256, 257, 258, 259, 260, 261, 262, 263, 264, 265, 266, 267, 268, 269, 270, 271, 272, 273, 274, 275, 276, 277, 278, 279, 280, 281, 282, 283, 284, 285, 286, 287, 288, 289, 290, 291, 292, 293, 294, 295, 296, 297, 298, 299, 300, 301, 302, 303, 304, 305, 306, 307, 308, 309, 310, 311, 312,

- 313, 314, 315, 316, 317, 318, 319, 320, 321, 322, 323, 324, 325, 326,
327, 328, 329, 330, 331, 332, 333, 334, 335, 336, 337, 338, 339, 340,
341, 342, 343, 344, 345, 346, 347, 348, 349, 350, 351, 352, 353, 354,
355, 356, 357, 358, 359, 360, 361, 362, 363, 364, 365, 366, 367, 368,
5 369, 370, 371, 372, 373, 374, 375, 376, 377, 378, 379, 380, 381, 382,
383, 384, 385, 386, 387, 388, 389, 390, 391, 392, 393, 394, 395, 396,
397, 398, 399, 400, 401, 402, 403, 404, 405, 406, 407, 408, 409, 410,
411, 412, 413, 414, 415, 416, 417, 418, 419, 420, 421, 422, 423, 424,
425, 426, 427, 428, 429, 430, 431, 432, 433, 434, 435, 436, 437, 438,
10 439, 440, 441, 442, 443, 444, 445, 446, 447, 448, 449, 450, 451, 452,
453, 454, 455, 456, 457, 458, 459, 460, 461, 462, 463, 464, 465, 466,
467, 468, 469, 470, 471, 472, 473, 474, 475, 476, 477, 478, 479, 480,
481, 482, 483, 484, 485, 486, 487, 488, 489, 490, 491, 492, 493, 494,
495, 496, 497, 498, 499, 500, 501, 502, 503, 504, 505, 506, 507, 508,
15 509, 510, 511, 512, 513, 514, 515, 516, 517, 518, 519, 520, 521, 522,
523, 524, 525, 526, 527, 528, 529, 530, 531, 532, 533, 534, 535, 536,
537, 538, 539, 540, 541, 542, 543, 544, 545, 546, 547, 548, 549, 550,
551, 552, 553, 554, 555, 556, 557, 558, 559, 560, 561, 562, 563, 564,
565, 566, 567, 568, 569, 570, 571, 572, 573, 574, 575, 576, 577, 578,
20 579, 580, 581, 582, 583, 584, 585, 586, 587, 588, 589, 590, 591, 592,
593, 594, 595, 596, 597, 598, 599, 600, 601, 602, 603, 604, 605, 606,
607, 608, 609, 610, 611, 612, 613, 614, 615, 616, 617, 618, 619, 620,
621, 622, 623, 624, 625, 626, 627, 628, 629, 630, 631, 632, 633, or
634 consecutive nucleotides, wherein the first nucleotide deleted is the
25 adenine nucleotide ['A'] at position 1 of SEQ ID NO:6;
- b) a deletion at the 3'-terminus of 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13,
14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31,
32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49,
50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, or 66
30 consecutive nucleotides, wherein the first nucleotide deleted is the
cytosine ['C'] nucleotide at position 942 of SEQ ID NO:6;

- c) substitution of a guanine ['G'] nucleotide for the cytosine ['C'] nucleotide at position 753 of SEQ ID NO:6;
- d) substitution of an adenine ['A'] nucleotide or a thymine ['T'] nucleotide for the cytosine ['C'] nucleotide at position 753 of SEQ ID NO:6;
- 5 e) substitution of the nucleotide sequence 'AA' for the nucleotide sequence 'TT' at positions 41 to 42 of SEQ ID NO:6;
- f) insertion of the nucleotide sequence 'TCG' between position 109 and 110 of SEQ ID NO:6;
- 10 g) insertion of the nucleotide sequence 'AAAT' between position 131 and position 132 of SEQ ID NO:6;
- h) substitution of the nucleotide sequence 'AAA' for the nucleotide sequence 'TT' at positions 76 to 77 of SEQ ID NO:6;
- i) substitution of the nucleotide sequence 'GTTT' for the nucleotide sequence 'AAAA' at positions 631 to 634 of SEQ ID NO:6;
- 15 j) insertion of the nucleotide sequence 'TA' between position 512 and position 513 of SEQ ID NO:6;
- k) insertion of the nucleotide sequence 'TTTA' between position 365 and position 366 of SEQ ID NO:6;
- 20 l) insertion of the nucleotide sequence 'TTAAA' between position 230 and position 231 of SEQ ID NO:6; and
- m) any combination of part a), part b), part c), part d), part e), part f), part g), part h), part i), part j), part k), and part l) above.

These examples are not intended to be limiting in nature and will be elaborated below. FIG. 1 graphically illustrates various *Yarrowia* ALK2 promoter regions

25 (i.e., SEQ ID NO:6 [942 bp ALK2LN, comprising an 876 bp promoter region plus a 66 bp N-terminal protein coding region], SEQ ID NO:7 [876 bp ALK2L], SEQ ID NO:11 [876 bp ALK2LM], SEQ ID NO:15 [834 bp ALK2LM1], SEQ ID NO:19 [767 bp ALK2LM-C], SEQ ID NO:49 [646 bp ALK2LM-P8], SEQ ID NO:51 [511 bp ALK2LM-P7], SEQ ID NO:53 [364 bp ALK2LM-P3], SEQ ID NO:55 [242 bp

30 ALK2LM-P4], and SEQ ID NO:56 [68 bp minimal/core promoter]), with the 1000 bp 5' upstream region (SEQ ID NO:2) of the ALK2 initiation codon of the *Yarrowia* ALK2 gene as a reference.

The term “promoter activity” will refer to an assessment of the transcriptional efficiency of a promoter. This may, for instance, be determined directly by measurement of the amount of mRNA transcription from the promoter (e.g., by quantitative PCR or Northern blotting or primer extension methods) or
5 indirectly by measuring the amount of gene product expressed from the promoter.

The terms “polynucleotide”, “polynucleotide sequence”, “nucleic acid sequence”, “nucleic acid fragment” and “isolated nucleic acid fragment” are used interchangeably herein. These terms encompass nucleotide sequences and the
10 like. A polynucleotide may be a polymer of RNA or DNA that is single- or double-stranded that optionally contains synthetic, non-natural or altered nucleotide bases. A polynucleotide in the form of a polymer of DNA may be comprised of one or more segments of cDNA, genomic DNA, synthetic DNA, or mixtures thereof. Nucleotides (usually found in their 5'-monophosphate form)
15 are referred to by a single letter designation as follows: “A” for adenylate or deoxyadenylate (for RNA or DNA, respectively), “C” for cytidylate or deoxycytidylate, “G” for guanylate or deoxyguanylate, “U” for uridylate, “T” for deoxythymidylate, “R” for purines (A or G), “Y” for pyrimidines (C or T), “K” for G or T, “H” for A or C or T, “I” for inosine, and “N” for any nucleotide.

20 A “substantial portion” of an amino acid sequence or nucleotide sequence is that portion comprising enough of the amino acid sequence of a polypeptide or the nucleotide sequence of a gene to putatively identify that polypeptide or gene, either by manual evaluation of the sequence by one skilled in the art, or by computer-automated sequence comparison and identification using algorithms
25 such as BLAST (Basic Local Alignment Search Tool; Altschul, S. F., et al., *J. Mol. Biol.* 215:403-410 (1993)). In general, a sequence of ten or more contiguous amino acids or thirty or more nucleotides is necessary in order to identify putatively a polypeptide or nucleic acid sequence as homologous to a known protein or gene. Moreover, with respect to nucleotide sequences, gene-
30 specific oligonucleotide probes comprising 20-30 contiguous nucleotides may be used in sequence-dependent methods of gene identification (e.g., Southern hybridization) and isolation (e.g., *in situ* hybridization of bacterial colonies or

bacteriophage plaques). In addition, short oligonucleotides of 12-15 bases may be used as amplification primers in PCR in order to obtain a particular nucleic acid molecule comprising the primers. Accordingly, a "substantial portion" of a nucleotide sequence comprises enough of the sequence to specifically identify
5 and/or isolate a nucleic acid molecule comprising the sequence.

The disclosure herein teaches partial or complete nucleotide sequences containing one or more particular yeast promoters. The skilled artisan, having the benefit of the sequences as reported herein, may now use all or a substantial portion of the disclosed sequences for purposes known to those skilled in this
10 art. Accordingly, the complete sequences as reported in the accompanying Sequence Listing, as well as substantial portions of those sequences as defined above, are encompassed in the present disclosure.

The term "complementary" is used to describe the relationship between nucleotide bases that are capable of hybridizing to one another. For example,
15 with respect to DNA, adenosine is complementary to thymine and cytosine is complementary to guanine. Accordingly, isolated nucleic acid fragments that are complementary to the complete sequences as reported in the accompanying Sequence Listing, as well as those substantially similar nucleic acid sequences, are encompassed in the present disclosure.

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

30 "Sequence identity" or "identity" in the context of nucleic acid or polypeptide sequences refers to the nucleic acid bases or amino acid residues in two

sequences that are the same when aligned for maximum correspondence over a specified comparison window.

Thus, “percentage of sequence identity” or “percent identity” refers to the value determined by comparing two optimally aligned sequences over a comparison window, wherein the portion of the polynucleotide or polypeptide sequence in the comparison window may comprise additions or deletions (i.e., gaps) as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. The percentage is calculated by determining the number of positions at which the identical nucleic acid base or amino acid residue occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison and multiplying the results by 100 to yield the percentage of sequence identity.

Methods to determine “percent identity” and “percent similarity” are codified in publicly available computer programs. Percent identity and percent similarity can be readily calculated by known methods, including but not limited to those described in: 1) Computational Molecular Biology (Lesk, A. M., Ed.) Oxford University: NY (1988); 2) Biocomputing: Informatics and Genome Projects (Smith, D. W., Ed.) Academic: NY (1993); 3) Computer Analysis of Sequence Data, Part I (Griffin, A. M., and Griffin, H. G., Eds.) 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).

Sequence alignments and percent identity or similarity calculations may be determined using a variety of comparison methods designed to detect homologous sequences including, but not limited to, the MegAlign™ program of the LASERGENE bioinformatics computing suite (DNASTAR Inc., Madison, WI). Multiple alignment of the sequences is performed using the “Clustal method of alignment” which encompasses several varieties of the algorithm including the “Clustal V method of alignment” and the “Clustal W method of alignment” (described by Higgins and Sharp, *CABIOS*, 5:151-153 (1989); Higgins, D.G. et al., *Comput. Appl. Biosci.*, 8:189-191(1992)) and found in the

MegAlign™ (version 8.0.2) program of the LASERGENE bioinformatics computing suite (DNASTAR Inc.). After alignment of the sequences using either Clustal program, it is possible to obtain a “percent identity” by viewing the “sequence distances” table in the program.

- 5 For multiple alignments using the Clustal V method of alignment, the default values correspond to GAP PENALTY=10 and GAP LENGTH PENALTY=10. Default parameters for pairwise alignments and calculation of percent identity of protein sequences using the Clustal V method are KTUPLE=1, GAP PENALTY=3, WINDOW=5 and DIAGONALS SAVED=5. For
10 nucleic acids these parameters are KTUPLE=2, GAP PENALTY=5, WINDOW=4 and DIAGONALS SAVED=4.

- Default parameters for multiple alignment using the Clustal W method of alignment correspond to GAP PENALTY=10, GAP LENGTH PENALTY=0.2, Delay Divergent Seqs(%)=30, DNA Transition Weight=0.5, Protein Weight
15 Matrix=Gonnet Series, DNA Weight Matrix=IUB.

- The “BLASTN method of alignment” is an algorithm provided by the National Center for Biotechnology Information [“NCBI”] to compare nucleotide sequences using default parameters, while the “BLASTP method of alignment” is an algorithm provided by the NCBI to compare protein sequences using default
20 parameters.

- It is well understood by one skilled in the art that many levels of sequence identity are useful in identifying polypeptides from other species, wherein such polypeptides have the same or similar function or activity. Likewise, suitable promoter regions (isolated polynucleotides of the present invention) are at least
25 about 70-85% identical, and more preferably at least about 85-95% identical to the nucleotide sequences reported herein. Although preferred ranges are described above, useful examples of percent identities include any integer percentage from 70% to 100%, such as 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%,
30 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99%. Suitable *Yarrowia* ALK2 promoter regions not only have the above homologies but typically are at least 50 nucleotides in length, more preferably at least 100 nucleotides in length, more

preferably at least 250 nucleotides in length, and more preferably at least 500 nucleotides in length.

“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.

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

“Gene” refers to a nucleic acid fragment that expresses a specific protein, and that may refer to the coding region alone or may include regulatory sequences preceding (5' non-coding sequences) and following (3' non-coding sequences) the coding sequence. “Native gene” refers to a gene as found in nature with its own regulatory sequences. “Chimeric gene” refers to any gene that is not a native gene, comprising regulatory and coding sequences that are not found together in nature. Accordingly, a chimeric gene may comprise regulatory sequences and coding sequences that are derived from different sources, or regulatory sequences and coding sequences derived from the same source, but arranged in a manner different than that found in nature. Chimeric genes herein will typically comprise a promoter region of an *ALK2 Yarrowia* gene operably linked to a coding region of interest. “Endogenous gene” refers to a

native gene in its natural location in the genome of an organism. A “foreign” gene refers to a gene that is introduced into the host organism by gene transfer. Foreign genes can comprise native genes inserted into a non-native organism, native genes introduced into a new location within the native host, or chimeric genes. A “transgene” is a gene that has been introduced into the genome by a transformation procedure. A “codon-optimized gene” is a gene having its frequency of codon usage designed to mimic the frequency of preferred codon usage of the host cell.

“Coding sequence” refers to a DNA sequence which codes for a specific amino acid sequence. The terms “coding sequence” and “coding region” are used interchangeably herein. A “coding region of interest” is a coding region which is desired to be expressed. Such coding regions are discussed more fully hereinbelow. “Regulatory sequences” refer to nucleotide sequences located upstream (5’ non-coding sequences), within, or downstream (3’ non-coding sequences) of a coding sequence, and which influence the transcription, RNA processing or stability, or translation of the associated coding sequence. Regulatory sequences may include, but are not limited to: promoters, enhancers, silencers, 5’ untranslated leader sequence (e.g., between the transcription start site and translation initiation codon), introns, polyadenylation recognition sequences, RNA processing sites, effector binding sites and stem-loop structures.

“Promoter” refers to a DNA sequence that facilitates transcription of a coding sequence, thereby enabling gene expression. In general, a promoter is typically located on the same strand and upstream of the coding sequence (i.e., 5’ of the coding sequence). Promoters may be derived in their entirety from a native gene, or be composed of different elements derived from different promoters found in nature, or even comprise synthetic DNA segments. It is understood by those skilled in the art that different promoters may direct the expression of a gene in different tissues or cell types, or at different stages of development, or in response to different environmental or physiological conditions. Promoters that cause a gene to be expressed at almost all stages of development are commonly referred to as “constitutive promoters”. It is further

recognized that since in most cases the exact boundaries of regulatory sequences (especially at their 5' end) have not been completely defined, DNA fragments of some variation may have identical promoter activity.

“Minimal promoter” refers to the minimal length of DNA sequence that is necessary to initiate basal level transcription of an operably linked coding sequence. The “minimal promoter” usually does not include the untranslated region located between transcription start site and translation start site. Although promoters often interact with the TATA binding protein [“TBP”] to create a transcription initiation complex from which RNA polymerase II transcribes the DNA coding sequence, only some promoters contain a TATA box to which TBP binds directly. In yeast, the TATA-box is usually located about 20 to 130 bp upstream of the transcription start site. For those TATA-less promoters, it is thought that transcription factor TFIID coordinates delivery of TBP and functions largely to stabilize TBP binding in lieu of a TATA box (Basehoar *et al.*, *Cell*, 116:699–709 (2004)). Some TATA-less promoters contain an “initiator” element [A(A_{rich})₅NYA(A/T)NN(A_{rich})₆ (SEQ ID NO:58), Zhang, Z. and Dietrich, F. S., *Nucleic Acids Res.*, 33:2838–2851 (2005), incorporated herein by reference] located around the transcription start site, which can direct basal level transcription.

Thus, the minimal promoter region for the *ALK2* TATA-containing promoters is herein defined as the -68 to -1 region upstream of the *ALK2* gene (i.e., as set forth in SEQ ID NO:56), which contains a TATA box sufficient to initiate basal level transcription of an operably linked coding sequence.

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

The term “enhancer” refers to a *cis*-regulatory sequence that can elevate levels of transcription from an adjacent eukaryotic promoter, thereby increasing transcription of the gene. Enhancers can act on promoters over many kilobases of DNA and can be 5' or 3' to the promoter they regulate. Enhancers can also be
5 located within introns (Giacopelli F. et al., *Gene Expr.*, 11:95-104 (2003)).

“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
10 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
15 the cell. “Antisense RNA” refers to an RNA transcript that is complementary to all or part of a target primary transcript or mRNA, and that blocks the expression of a target gene (U.S. Pat. No. 5,107,065).

The term “operably linked” refers to the association of nucleic acid sequences on a single nucleic acid molecule so that the function of one is
20 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.

25 The term “recombinant” refers to an artificial combination of two otherwise separated segments of sequence, e.g., by chemical synthesis or by the manipulation of isolated segments of nucleic acids by genetic engineering techniques.

The term “expression”, as used herein, refers to the transcription and
30 stable accumulation of sense (mRNA) or antisense RNA. Expression may also refer to translation of mRNA into a protein (either precursor or mature).

“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.

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

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

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

- The terms “recombinant construct”, “expression construct”, “chimeric
- 30 construct”, “construct”, and “recombinant DNA construct” are used interchangeably herein. A recombinant construct comprises an artificial combination of nucleic acid fragments, e.g., regulatory and coding sequences

that are not found together in nature. For example, a recombinant construct may comprise one or more expression cassettes. In another example, a recombinant DNA construct may comprise regulatory sequences and coding sequences that are derived from different sources, or regulatory sequences and coding
5 sequences derived from the same source, but arranged in a manner different than that found in nature. Such a construct may be used by itself or may be used in conjunction with a vector. If a vector is used, then the choice of vector is dependent upon the method that will be used to transform host cells as is well known to those skilled in the art. For example, a plasmid vector can be used.

10 The skilled artisan is well aware of the genetic elements that must be present on the vector in order to successfully transform, select and propagate host cells comprising any of the isolated nucleic acid fragments described herein. The skilled artisan will also recognize that different independent transformation events will result in different levels and patterns of expression (Jones et al.,
15 *EMBO J.*, 4:2411-2418 (1985); De Almeida et al., *Mol. Gen. Genetics*, 218:78-86 (1989)), and thus that multiple events must be screened in order to obtain strains displaying the desired expression level and pattern. Such screening may be accomplished by Southern analysis of DNA, Northern analysis of mRNA expression, Western and/or Elisa analyses of protein expression, formation of a
20 specific product, phenotypic analysis or GC analysis of the PUFA products, among others.

The term "sequence analysis software" refers to any computer algorithm or software program that is useful for the analysis of nucleotide or
25 amino acid sequences. "Sequence analysis software" may be commercially available or independently developed. Typical sequence analysis software will include, but is not limited to: 1) the GCG suite of programs (Wisconsin Package Version 9.0, Genetics Computer Group (GCG), Madison, WI); 2) BLASTP, BLASTN, BLASTX (Altschul et al., *J. Mol. Biol.*, 215:403-410
30 (1990)); 3) DNASTAR (DNASTAR, Inc. Madison, WI); 4) Sequencher (Gene Codes Corporation, Ann Arbor, MI); and, 5) the FASTA program incorporating the Smith-Waterman algorithm (W. R. Pearson, *Comput. Methods Genome*

Res., [Proc. Int. Symp.] (1994), Meeting Date 1992, 111-20. Editor(s): Suhai, Sandor. Plenum: New York, NY). Within this description, whenever sequence analysis software is used for analysis, the analytical results are based on the “default values” of the program referenced, unless otherwise specified. As
5 used herein “default values” will mean any set of values or parameters that originally load with the software when first initialized.

Standard recombinant DNA and molecular cloning techniques used herein are well known in the art and are described more fully in Sambrook, J., Fritsch, E.F. and Maniatis, T. *Molecular Cloning: A Laboratory Manual*; Cold Spring Harbor Laboratory: Cold Spring Harbor, NY (1989); by Silhavy, T. J., Bennis, M. L. and Enquist, L. W., Experiments with Gene Fusions, Cold Spring Harbor Laboratory: Cold Spring Harbor, NY (1984); and by Ausubel, F. M. et al., Current Protocols in Molecular Biology, published by Greene Publishing Assoc. and Wiley-Interscience, Hoboken, NJ (1987).
10

15 N-alkane-hydroxylating cytochrome P450 enzymes include those enzymes that catalyze the terminal hydroxylation of n-alkanes. Within *Yarrowia lipolytica*, a gene encoding an n-alkane-hydroxylating cytochrome P450 enzyme has been identified (SEQ ID NO:1, “YALI0F01320g” locus, Dujon, B. et al., *Nature*, 430(6995):35-44 (2004)).

20 *Yarrowia* mutants having their native *snf1* gene knocked-out can constitutively accumulate high levels of oil, even in nitrogenous growth media, when compared to the wild-type strains (U.S. Pat. Appl. Publ. No. 2010-0062502-A1). The *snf1* gene encodes the alpha subunit of the SNF1 protein kinase, a heterotrimeric serine/threonine protein kinase that appears to function
25 as a global regulator of gene expression. Particularly, SNF1 protein kinase regulates the transcription of numerous glucose-repressed genes, with a significant portion of those genes functioning in transcription and signal transduction. In general, when the heterotrimeric kinase is activated by phosphorylation, for example, in response to glucose limitation, ATP-producing
30 catabolic pathways increase.

Based on microarray analysis in *Y. lipolytica*, it has been determined that over 200 genes are differentially expressed by more than 1.3-fold in *snf1* knock-

out strains, when compared to their expression in control strains (U.S. Pat. Appl. Publ. No. 2010-0062502-A1, Example 11 therein). Interestingly, the transcription of ALK2 in these *snf1* knock-out strains was increased as much as three times that of the wild-type strains.

5 Based on the above, the *ALK2* gene was identified as a potential source of new and improved yeast promoters for metabolic engineering of yeast and for controlling heterologous genes in yeast. In order to understand the means by which ALK2 expression is regulated in *Yarrowia*, the ALK2 promoter was isolated and its functional structure was mechanistically analyzed.

10 In general, a promoter useful for controlling the expression of heterologous genes in yeast should preferably meet criteria with respect to strength, activities, pH tolerance and inducibility, as described in U.S. Patent No. 7,259,255. Additionally, today's complex metabolic engineering utilized for construction of yeast having the capability to produce a variety of heterologous
15 polypeptides in commercial quantities requires a suite of promoters that are regulatable under a variety of natural growth and induction conditions.

 Thus, described herein are a suite of promoter regions of an *ALK2 Yarrowia* gene, useful for driving expression of any suitable coding region of interest in a transformed yeast cell. More specifically, described herein is an
20 isolated nucleic acid molecule comprising a promoter region of an *ALK2 Yarrowia* gene, wherein said promoter region of an *ALK2 Yarrowia* gene is set forth in SEQ ID NO:6 (corresponding to the 5' upstream -876 to -1 region plus the first 66 bp of the N-terminal protein coding region of the *Yarrowia ALK2* gene), and wherein said promoter optionally comprises at least one modification
25 selected from the group consisting of:

 a) a deletion at the 5'-terminus of 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116,

117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130,
131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144,
145, 146, 147, 148, 149, 150, 151, 152, 153, 154, 155, 156, 157, 158,
159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172,
5 173, 174, 175, 176, 177, 178, 179, 180, 181, 182, 183, 184, 185, 186,
187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 197, 198, 199, 200,
201, 202, 203, 204, 205, 206, 207, 208, 209, 210, 211, 212, 213, 214,
215, 216, 217, 218, 219, 220, 221, 222, 223, 224, 225, 226, 227, 228,
10 229, 230, 231, 232, 233, 234, 235, 236, 237, 238, 239, 240, 241, 242,
243, 244, 245, 246, 247, 248, 249, 250, 251, 252, 253, 254, 255, 256,
257, 258, 259, 260, 261, 262, 263, 264, 265, 266, 267, 268, 269, 270,
271, 272, 273, 274, 275, 276, 277, 278, 279, 280, 281, 282, 283, 284,
285, 286, 287, 288, 289, 290, 291, 292, 293, 294, 295, 296, 297, 298,
15 299, 300, 301, 302, 303, 304, 305, 306, 307, 308, 309, 310, 311, 312,
313, 314, 315, 316, 317, 318, 319, 320, 321, 322, 323, 324, 325, 326,
327, 328, 329, 330, 331, 332, 333, 334, 335, 336, 337, 338, 339, 340,
341, 342, 343, 344, 345, 346, 347, 348, 349, 350, 351, 352, 353, 354,
355, 356, 357, 358, 359, 360, 361, 362, 363, 364, 365, 366, 367, 368,
369, 370, 371, 372, 373, 374, 375, 376, 377, 378, 379, 380, 381, 382,
20 383, 384, 385, 386, 387, 388, 389, 390, 391, 392, 393, 394, 395, 396,
397, 398, 399, 400, 401, 402, 403, 404, 405, 406, 407, 408, 409, 410,
411, 412, 413, 414, 415, 416, 417, 418, 419, 420, 421, 422, 423, 424,
425, 426, 427, 428, 429, 430, 431, 432, 433, 434, 435, 436, 437, 438,
439, 440, 441, 442, 443, 444, 445, 446, 447, 448, 449, 450, 451, 452,
25 453, 454, 455, 456, 457, 458, 459, 460, 461, 462, 463, 464, 465, 466,
467, 468, 469, 470, 471, 472, 473, 474, 475, 476, 477, 478, 479, 480,
481, 482, 483, 484, 485, 486, 487, 488, 489, 490, 491, 492, 493, 494,
495, 496, 497, 498, 499, 500, 501, 502, 503, 504, 505, 506, 507, 508,
509, 510, 511, 512, 513, 514, 515, 516, 517, 518, 519, 520, 521, 522,
30 523, 524, 525, 526, 527, 528, 529, 530, 531, 532, 533, 534, 535, 536,
537, 538, 539, 540, 541, 542, 543, 544, 545, 546, 547, 548, 549, 550,
551, 552, 553, 554, 555, 556, 557, 558, 559, 560, 561, 562, 563, 564,

- 565, 566, 567, 568, 569, 570, 571, 572, 573, 574, 575, 576, 577, 578, 579, 580, 581, 582, 583, 584, 585, 586, 587, 588, 589, 590, 591, 592, 593, 594, 595, 596, 597, 598, 599, 600, 601, 602, 603, 604, 605, 606, 607, 608, 609, 610, 611, 612, 613, 614, 615, 616, 617, 618, 619, 620, 621, 622, 623, 624, 625, 626, 627, 628, 629, 630, 631, 632, 633, or 634 consecutive nucleotides, wherein the first nucleotide deleted is the adenine nucleotide ['A'] at position 1 of SEQ ID NO:6;
- 5 b) a deletion at the 3'-terminus of 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, or 66 consecutive nucleotides, wherein the first nucleotide deleted is the cytosine ['C'] nucleotide at position 942 of SEQ ID NO:6;
- 10 c) substitution of a guanine ['G'] nucleotide for the cytosine ['C'] nucleotide at position 753 of SEQ ID NO:6;
- 15 d) substitution of an adenine ['A'] nucleotide or a thymine ['T'] nucleotide for the cytosine ['C'] nucleotide at position 753 of SEQ ID NO:6;
- e) substitution of the nucleotide sequence 'AA' for the nucleotide sequence "TT" at positions 41 to 42 of SEQ ID NO:6;
- 20 f) insertion of the nucleotide sequence 'TCG' between position 109 and 110 of SEQ ID NO:6;
- g) insertion of the nucleotide sequence 'AAAT' between position 131 and position 132 of SEQ ID NO:6;
- 25 h) substitution of the nucleotide sequence 'AAA' for the nucleotide sequence 'TT' at positions 76 to 77 of SEQ ID NO:6;
- i) substitution of the nucleotide sequence 'GTTT' for the nucleotide sequence 'AAAA' at positions 631 to 634 of SEQ ID NO:6;
- j) insertion of the nucleotide sequence 'TA' between position 512 and position 513 of SEQ ID NO:6;
- 30 k) insertion of the nucleotide sequence 'TTTA' between position 365 and position 366 of SEQ ID NO:6;

- l) insertion of the nucleotide sequence 'TTAAA' between position 230 and position 231 of SEQ ID NO:6; and
- m) any combination of part a), part b), part c), part d), part e), part f), part g), part h), part i), part j), part k), and part l) above.

5 In some embodiments, the promoter region of an *ALK2 Yarrowia* gene is selected from the group consisting of SEQ ID NOs:7, 11, 15, 19, 27, 31, 35, 39, 43, 47, 49, 51, 53 and 55. These promoter regions are preferred to provide relatively high levels of inducible promoter activity when operably linked to a coding region of interest.

10 The relationship between the promoter regions of a *Yarrowia ALK2* gene selected from the group consisting of SEQ ID NOs:2, 6, 7, 11, 27, 15, 47, 43, 39, 35, 31, 19, 49, 51, 53, 55, 56, *supra*, is readily observed upon alignment of the individual promoter sequences. Specifically, FIG. 2 (comprising FIG. 2A, 2B, 2C, 2D, 2E, 2F, 2G, 2H, 2I, and 2J) provides an alignment of:

- 15 (a) the 1000 bp promoter region ALK2F (SEQ ID NO:2);
- (b) the 942 bp promoter region ALK2N (SEQ ID NO:6);
- (c) the 876 bp promoter region ALK2L (SEQ ID NO:7);
- (d) the 876 bp promoter region ALK2LM (SEQ ID NO:11);
- (e) the 838 bp promoter region ALK2LM1-S (SEQ ID NO:27);
- 20 (f) the 834 bp promoter region ALK2LM1 (SEQ ID NO:15);
- (g) the 840 bp promoter region ALK2LM1-PP6 (SEQ ID NO:47);
- (h) the 839 bp promoter region ALK2LM1-PP5 (SEQ ID NO:43);
- (i) the 837 bp promoter region ALK2LM1-PP2 (SEQ ID NO:39);
- (j) the 835 bp promoter region ALK2LM1-PP1 (SEQ ID NO:35);
- 25 (k) the 835 bp promoter region ALK2LM1-P (SEQ ID NO:31);
- (l) the 767 bp promoter region ALK2LM-C (SEQ ID NO:19);
- (m) the 646 bp promoter region ALK2LM-P8 (SEQ ID NO:49);
- (n) the 511 bp promoter region ALK2LM-P7 (SEQ ID NO:51);
- (o) the 364 bp promoter region ALK2LM-P3 (SEQ ID NO:53);
- 30 (p) the 242 bp promoter region ALK2LM-P4 (SEQ ID NO:55); and,
- (q) the 68 bp *Yarrowia* promoter region ALK2 minimal/core promoter (SEQ ID NO:56).

Nucleotide differences are highlighted with a box and an arrow.

As will be obvious to one of skill in the art, the above discussion is by no means limiting to the description of suitable promoter regions of an *ALK2 Yarrowia* gene. For example, alternate *Yarrowia* *ALK2* promoter regions may be
5 longer than the 1000 bp sequence 5' upstream of the nucleotide 'A' (designated as +1) of the translation initiation codon 'ATG' of SEQ ID NO:1, thereby encompassing additional nucleotides.

Similarly, it should be recognized that promoter fragments of various diminishing lengths may have identical promoter activity, since the exact
10 boundaries of the regulatory sequences have not been completely defined. Thus, for example, it is also contemplated that a suitable promoter region of an *ALK2 Yarrowia* gene could also include a promoter region of SEQ ID NO:6, wherein the 5'-terminus deletion was greater than 634 consecutive nucleotides.

More specifically, based on sequence analysis of the promoter region set
15 forth in SEQ ID NO:55, and identification of a TATA box spanning positions -68 to -62 5' upstream of the ATG translation initiation codon, it is hypothesized herein that the minimal promoter region that could function for basal level transcription initiation of an operably linked coding region of interest encompasses (at least) the 68 bp 5' upstream untranslated region from the 'ATG'
20 translation initiation codon of an *ALK2 Yarrowia* gene comprising the -68 to -1 region of SEQ ID NO:1; this 68 bp region is set forth independently as SEQ ID NO:56.

In alternate embodiments, SEQ ID NO:56 could be utilized as a minimal promoter to fuse with enhancers to drive increased transcription from a coding
25 region of interest. One of skill in the art would readily be able to conduct appropriate deletion studies to determine the appropriate length of a promoter region of an *ALK2 Yarrowia* gene required to enable the desired level of promoter activity.

In alternate embodiments, SEQ ID NO:57 could be used as an enhancer
30 to elevate levels of transcription from an adjacent eukaryotic coding region of interest, or fused with a minimal promoter to form a recombinant promoter to drive expression of a coding region of interest. One of skill in the art would

readily be able to conduct appropriate deletion studies to determine the minimal length of the enhancer region (SEQ ID NO:57) to enable the desired level of promoter activity.

Thus, in alternate embodiments, described herein is an isolated nucleic acid molecule comprising a promoter region of an *ALK2 Yarrowia* gene, wherein
5 said isolated nucleic acid molecule is selected from the group consisting of SEQ ID NO:7, SEQ ID NO:11, SEQ ID NO:27, SEQ ID NO:15, SEQ ID NO:47, SEQ ID NO:43, SEQ ID NO:39, SEQ ID NO:35, SEQ ID NO:31, SEQ ID NO:19, SEQ ID NO:49, SEQ ID NO:51, SEQ ID NO:53, and SEQ ID NO:55

10 More specifically, additional variant *Yarrowia* ALK2 promoter regions may be constructed, wherein the DNA sequence of the promoter has one or more nucleotide substitutions (i.e., deletions, insertions, substitutions, or addition of one or more nucleotides in the sequence) which do not affect (in particular, impair) the yeast promoter activity. Regions that can be modified without
15 significantly affecting the yeast promoter activity can be identified by deletion studies. A variant promoter of the present invention has at least about 10%, more preferably at least about 20%, more preferably at least about 40%, more preferably at least about 60%, more preferably at least about 80%, more preferably at least about 90%, more preferably at least about 100%, more
20 preferably at least about 200%, more preferably at least about 300% and most preferably at least about 500% of the promoter activity of any of the *Yarrowia* ALK2 promoter regions described herein as SEQ ID NO:7, SEQ ID NO:11, SEQ ID NO:27, SEQ ID NO:15, SEQ ID NO:47, SEQ ID NO:43, SEQ ID NO:39, SEQ ID NO:35, SEQ ID NO:31, SEQ ID NO:19, SEQ ID NO:49, SEQ ID NO:51, SEQ
25 ID NO:53, and SEQ ID NO:55.

U.S. Patent No. 7,259,255 describes a variety of methods for mutagenesis, suitable for the generation of mutant promoters. This would permit production of a putative promoter having, for example, a more desirable level of promoter activity in the host cell or a more desirable sequence for purposes of
30 cloning (e.g., removal of a restriction enzyme site within the native promoter region). Similarly, the cited reference also discusses means to examine regions of a nucleotide of interest important for promoter activity (i.e., functional analysis

via deletion mutagenesis to determine the minimum portion of the putative promoter necessary for activity).

All variant promoter regions of an *ALK2 Yarrowia* gene, derived from the promoter regions described herein, are within the scope of the present

5 disclosure.

Similarly, it should be noted that one could isolate regions upstream of the *ALK2* translation initiation codon in various *Yarrowia* species and strains, other than the region isolated herein from *Yarrowia lipolytica* ATCC #20362, and thereby identify alternate promoter regions of an *ALK2 Yarrowia* gene. As is well

10 known in the art, isolation of homologous promoter regions or genes using sequence-dependent protocols is readily possible using various techniques (see, U.S. Pat. No. 7,259,255). Examples of sequence-dependent protocols useful to isolate homologous promoter regions include, but are not limited to: 1) methods of nucleic acid hybridization; 2) methods of DNA and RNA amplification, as

15 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. U.S.A.*, 82:1074 (1985); or strand displacement amplification (SDA), Walker, et al., *Proc. Natl. Acad. Sci. U.S.A.*, 89:392 (1992)]; 3) methods of library construction and screening by

20 complementation; and, 4) methods of genome sequencing. Based on sequence conservation between related organisms, one would expect that the promoter regions would likely share significant homology (i.e., at least about 70-85% identity, more preferably at least about 85-90% identity and more preferably at least about 90-95% identity); however, one or more differences in nucleotide

25 sequence could be observed when aligned with promoter regions of comparable length derived from the upstream region of SEQ ID NO:2. For example, one of skill in the art could readily isolate the *Yarrowia* *ALK2* promoter region from any of the various *Y. lipolytica* strains available through the American Type Culture Collection ["ATCC"], including, for example #8661, #8662, #9773, #15586,

30 #16617, #16618, #18942, #18943, #18944, #18945, #20114, #20177, #20182, #20225, #20226, #20228, #20327, #20255, #20287, #20297, #20315, #20320, #20324, #20336, #20341, #20346, #20348, #20363, #20364, #20372, #20373,

- #20383, #20390, #20400, #20460, #20461, #20462, #20496, #20510, #20628, #20688, #20774, #20775, #20776, #20777, #20778, #20779, #20780, #20781, #20794, #20795, #20875, #20241, #20422, #20423, #32338, #32339, #32340, #32341, #34342, #32343, #32935, #34017, #34018, #34088, #34922, #34922, #38295, #42281, #44601, #46025, #46026, #46027, #46028, #46067, #46068, #46069, #46070, #46330, #46482, #46483, #46484, #46436, #60594, #62385, #64042, #74234, #76598, #76861, #76862, #76982, #90716, #90811, #90812, #90813, #90814, #90903, #90904, #90905, #96028, #201241, #201242, #201243, #201244, #201245, #201246, #201247, #201249, or #201847.
- 10 Similarly, the following strains of *Yarrowia lipolytica* could be obtained from the Herman J. Phaff Yeast Culture Collection, University of California Davis (Davis, CA): *Y. lipolytica* 49-14, *Y. lipolytica* 49-49, *Y. lipolytica* 50-140, *Y. lipolytica* 50-46, *Y. lipolytica* 50-47, *Y. lipolytica* 51-30, *Y. lipolytica* 60-26, *Y. lipolytica* 70-17, *Y. lipolytica* 70-18, *Y. lipolytica* 70-19, *Y. lipolytica* 70-20, *Y. lipolytica* 74-78, *Y.*
- 15 *lipolytica* 74-87, *Y. lipolytica* 74-88, *Y. lipolytica* 74-89, *Y. lipolytica* 76-72, *Y. lipolytica* 76-93, *Y. lipolytica* 77-12T and *Y. lipolytica* 77-17. Or, strains could be obtained from the Laboratoire de Microbiologie et Génétique Moléculaire of Dr. Jean-Marc Nicaud, INRA Centre de Grignon, France, including for example, *Yarrowia lipolytica* JMY798 (Mlíčková, K. et al., *Appl Environ Microbiol.*
- 20 70(7):3918-24 (2004)), *Y. lipolytica* JMY399 (Barth, G., and C. Gaillardin. In, *Nonconventional Yeasts In Biotechnology*; Wolf, W.K., Ed.; Springer-Verlag: Berlin, Germany, 1996; pp 313-388) and *Y. lipolytica* JMY154 (Wang, H.J., et al., *J. Bacteriol.* 181(17):5140-8 (1999)).

In general, microbial expression systems and expression vectors

25 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, which could then be introduced into appropriate microorganisms via transformation to provide high-level expression of the encoded enzymes.

30 Vectors (e.g., constructs, plasmids) and DNA expression cassettes useful for the transformation of suitable microbial host cells are well known in the art. The specific choice of sequences present in the construct is dependent upon the

desired expression products, the nature of the host cell and the proposed means of separating transformed cells versus non-transformed cells. Typically, however, the vector contains at least one expression cassette, a selectable marker and sequences allowing autonomous replication or chromosomal
5 integration. Suitable expression cassettes comprise a region 5' of the gene that controls transcription (e.g., a promoter), the gene coding sequence, and a region 3' of the DNA fragment that controls transcriptional termination, i.e., a terminator. It is most preferred when both control regions are derived from genes from the transformed yeast cell, although they need not be derived from
10 genes native to the host.

Herein, transcriptional control regions (also initiation control regions or promoters) that are useful to drive expression of a coding gene of interest in the desired yeast cell are those promoter regions of an *ALK2 Yarrowia* gene, as described *supra*. Once the promoter regions are identified and isolated, they
15 may be operably linked to a coding region of interest to create a chimeric gene. The chimeric gene may then be expressed in a suitable expression vector in transformed yeast cells, particularly in the cells of oleaginous yeast (e.g., *Yarrowia lipolytica*).

Coding regions of interest to be expressed in transformed yeast cells may
20 be either endogenous to the host or heterologous. Genes encoding proteins of commercial value are particularly suitable for expression. For example, suitable coding regions of interest may include (but are not limited to) those encoding viral, bacterial, fungal, plant, insect, or vertebrate coding regions of interest, including mammalian polypeptides. Further, these coding regions of interest
25 may be, for example, structural proteins, signal transduction proteins, transcription factors, enzymes (e.g., oxidoreductases, transferases, hydrolyases, lyases, isomerases, ligases), or peptides. A non-limiting list includes genes encoding enzymes such as acyltransferases, aminopeptidases, amylases, carbohydrases, carboxypeptidases, catalyases, cellulases, chitinases, cutinases,
30 cyclodextrin glycosyltransferases, deoxyribonucleases, esterases, alpha-galactosidases, beta-glucanases, beta-galactosidases, glucoamylases, alpha-glucosidases, beta-glucosidases, invertases, laccases, lipases, mannosidases,

mutanases, oxidases, pectinolytic enzymes, peroxidases, phospholipases, phosphatases, phytases, polyphenoloxidases, proteolytic enzymes, ribonucleases, transglutaminases or xylanases.

Thus, one aspect of the present disclosure provides a recombinant
 5 construct comprising a *Yarrowia ALK2* promoter region, as well as recombinant expression vectors comprising the recombinant construct. The ALK2 promoter may also be comprised within a chimeric gene.

Also provided herein is a method for the expression of a coding region of interest in a transformed yeast cell comprising:

- 10 a) providing a transformed yeast cell having a recombinant construct, wherein the recombinant construct comprises:
 - (1) a promoter region of an *ALK2 Yarrowia* gene; and
 - (2) a coding region of interest which is expressible in the yeast cell;
- 15 wherein the promoter region is operably linked to the coding region of interest; and
- b) growing the transformed yeast cell of step (a) under conditions whereby the recombinant construct is expressed.

The polypeptide so produced by expression of the recombinant construct may
 20 optionally be recovered from the culture.

In some embodiments herein, preferred coding regions of interest are those encoding enzymes involved in the production of microbial oils, including omega-6 and omega-3 fatty acids (i.e., omega-6 and omega-3 fatty acid biosynthetic pathway enzymes). Thus, preferred coding regions include those
 25 encoding desaturases (e.g., delta-8 desaturases, delta-5 desaturases, delta-17 desaturases, delta-12 desaturases, delta-4 desaturases, delta-6 desaturases, delta-15 desaturases and delta-9 desaturases) and elongases (e.g., C_{14/16} elongases, C_{16/18} elongases, C_{18/20} elongases, C_{20/22} elongases, delta-6 elongases and delta-9 elongases).

30 More specifically, the omega-3/ omega-6 fatty acid biosynthetic pathway is illustrated in FIG. 3. All pathways require the initial conversion of oleic acid [18:1] to linoleic acid ["LA"; 18:2], the first of the omega-6 fatty acids, by a delta-

12 desaturase. Then, using the “delta-9 elongase/ delta-8 desaturase pathway” and LA as substrate, long-chain omega-6 fatty acids are formed as follows:

- 1) LA is converted to eicosadienoic acid [“EDA”; 20:2] by a delta-9 elongase; 2) EDA is converted to dihomo-gamma-linolenic acid [“DGLA”; 20:3] by a delta-8 desaturase; 3) DGLA is converted to arachidonic acid [“ARA”; 20:4] by a delta-5 desaturase; 4) ARA is converted to docosatetraenoic acid [“DTA”; 22:4] by a C_{20/22} elongase; and, 5) DTA is converted to docosapentaenoic acid [“DPA”; 22:5] by a delta-4 desaturase. To clarify, “omega-6 fatty acids” are polyunsaturated fatty acids having the first unsaturated double bond six carbon atoms from the omega (methyl) end of the molecule and additionally having a total of two or more double bonds, with each subsequent unsaturation occurring 3 additional carbon atoms toward the carboxyl end of the molecule.

- The “delta-9 elongase/ delta-8 desaturase pathway” can also use alpha-linolenic acid [“ALA”; 18:3] as substrate to produce long-chain omega-3 fatty acids as follows:
- 1) LA is converted to ALA, the first of the omega-3 fatty acids, by a delta-15 desaturase; 2) ALA is converted to eicosatrienoic acid [“ETra”; 20:3] by a delta-9 elongase; 3) ETra is converted to eicosatetraenoic acid [“ETA”; 20:4] by a delta-8 desaturase; 4) ETA is converted to eicosapentaenoic acid [“EPA”; 20:5] by a delta-5 desaturase; 5) EPA is converted to docosapentaenoic acid [“DPA”; 22:5] by a C_{20/22} elongase; and, 6) DPA is converted to docosahexaenoic acid [“DHA”; 22:6] by a delta-4 desaturase. Optionally, omega-6 fatty acids may be converted to omega-3 fatty acids. For example, ETA and EPA are produced from DGLA and ARA, respectively, by delta-17 desaturase activity. To clarify, “omega-3 fatty acids” are polyunsaturated fatty acids having the first unsaturated double bond three carbon atoms away from the omega end of the molecule and additionally having a total of three or more double bonds, with each subsequent unsaturation occurring 3 additional carbon atoms toward the carboxyl end of the molecule.

- Alternate pathways for the biosynthesis of omega-3/ omega-6 fatty acids utilize a delta-6 desaturase and C_{18/20} elongase, that is, the “delta-6 desaturase/ delta-6 elongase pathway”. More specifically, LA and ALA may be converted to GLA and stearidonic acid [“STA”; 18:4], respectively, by a delta-6 desaturase;

then, a C_{18/20} elongase converts GLA to DGLA and/or STA to ETA. Downstream PUFAs are subsequently formed as described above.

One of skill in the art will appreciate that the disclosure herein also provides a method for the production of an omega-3 fatty acid or omega-6 fatty acid comprising:

- a) providing a transformed oleaginous yeast comprising a recombinant construct, wherein the recombinant construct comprises:
 - i) a promoter region of an *ALK2 Yarrowia* gene; and
 - 10 ii) a coding region encoding at least one omega-3 fatty acid or omega-6 fatty acid biosynthetic pathway enzyme;wherein the promoter region and the coding region are operably linked; and
- b) growing the transformed oleaginous yeast of step (a) under conditions
15 whereby the at least one omega-3 fatty acid or omega-6 fatty acid biosynthetic pathway enzyme is expressed and the omega-3 fatty acid or the omega-6 fatty acid is produced; and
- c) optionally recovering the omega-3 fatty acid or the omega-6 fatty acid.

The omega-3 fatty acid or the omega-6 fatty acid may be selected from the
20 group consisting of: LA, GLA, EDA, DGLA, ARA, DTA, DPAn-6, ALA, STA, ETrA, ETA, EPA, DPAn-3 and DHA.

Once a DNA cassette (e.g., comprising a chimeric gene comprising a promoter region of an *ALK2 Yarrowia* gene, ORF and terminator) suitable for expression in a yeast cell has been obtained, it is placed in a plasmid vector
25 capable of autonomous replication in the yeast cell, or it is directly integrated into the genome of the yeast cell. Integration of expression cassettes can occur randomly within the yeast genome or can be targeted through the use of constructs containing regions of homology with the yeast genome sufficient to target recombination to a specific locus. All or some of the transcriptional and
30 translational regulatory regions can be provided by the endogenous locus where constructs are targeted to an 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 chimeric genes are expressed at the necessary levels to provide for synthesis of the desired products.

U.S. Patent No. 7,259,255 describes means to increase expression of a particular coding region of interest.

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

For convenience, a yeast cell that has been manipulated by any method to take up a DNA sequence, for example, in an expression cassette, is referred to herein as "transformed", "transformant" or "recombinant" (as these terms will be used interchangeably herein). The transformed yeast will have at least one copy of the expression construct and may have two or more, depending upon whether the expression cassette is integrated into the genome or is present on an extrachromosomal element having multiple copy numbers.

The transformed yeast cell can be identified by various selection techniques, as described in U.S. Patent 7,238,482, U.S. Patent 7,259,255 and U.S. Pat. No. 7,932,077.

Following transformation, substrates upon which the translated products of the chimeric genes act may be produced by the yeast either naturally or transgenically, or they may be provided exogenously.

Yeast cells for expression of the instant chimeric genes comprising a promoter region of an *ALK2 Yarrowia* gene may include yeast that grow on a variety of feedstocks, including simple or complex carbohydrates, fatty acids,

organic acids, oils, glycerol and alcohols, and/or hydrocarbons over a wide range of temperature and pH values. It is contemplated that because transcription, translation and the protein biosynthetic apparatus are highly conserved, any yeast will be a suitable host for expression of the present chimeric genes.

- 5 As previously noted, yeast do not form a specific taxonomic or phylogenetic grouping, but instead comprise a diverse assemblage of unicellular organisms that occur in the Ascomycotina and Basidiomycotina, most of which reproduce by budding (or fission) and derive energy via fermentation processes. Examples of some yeast genera include, but are not limited to: *Agaricostilbum*,
 10 *Ambrosiozyma*, *Arthroascus*, *Arxula*, *Ashbya*, *Babjevia*, *Bensingtonia*, *Botryozyma*, *Brettanomyces*, *Bullera*, *Candida*, *Clavispora*, *Cryptococcus*, *Cystofilobasidium*, *Debaryomyces*, *Dekkera*, *Dipodascus*, *Endomyces*, *Endomycopsella*, *Erythrobasidium*, *Fellomyces*, *Filobasidium*, *Galactomyces*, *Geotrichum*, *Guilliermondella*, *Hansenula*, *Hanseniaspora*, *Kazachstania*,
 15 *Kloeckera*, *Kluyveromyces*, *Kockovaella*, *Kodamaea*, *Komagataella*, *Kondoa*, *Lachancea*, *Leucosporidium*, *Leucosporidiella*, *Lipomyces*, *Lodderomyces*, *Issatchenkia*, *Magnusiomyces*, *Mastigobasidium*, *Metschnikowia*, *Monosporella*, *Myxozyma*, *Nadsonia*, *Nematospora*, *Oosporidium*, *Pachysolen*, *Pichia*, *Phaffia*, *Pseudozyma*, *Reniforma*, *Rhodosporidium*, *Rhodotorula*, *Saccharomyces*,
 20 *Saccharomycodes*, *Saccharomycopsis*, *Saturnispora*, *Schizoblastosporion*, *Schizosaccharomyces*, *Sirobasidium*, *Smithiozyma*, *Sporobolomyces*, *Sporopachydermia*, *Starmerella*, *Sympodiomyces*, *Sympodiomyces*, *Torulaspora*, *Tremella*, *Trichosporon*, *Trichosporiella*, *Trigonopsis*, *Udeniomyces*, *Wickerhamomyces*, *Williopsis*, *Xanthophyllomyces*, *Yarrowia*,
 25 *Zygosaccharomyces*, *Zygotorulaspora*, *Zymoxenogloea* and *Zygozyma*.

In preferred embodiments, the transformed yeast is an oleaginous yeast. These organisms are naturally capable of oil synthesis and accumulation, wherein the oil can comprise greater than about 25% of the dry cell weight, more preferably greater than about 30% of the dry cell weight, more preferably greater
 30 than about 40% of the dry cell weight, more preferably greater than about 50% of the dry cell weight, and most preferably greater than about 60% of the dry cell weight. Genera typically identified as oleaginous yeast include, but are not

limited to: *Yarrowia*, *Candida*, *Rhodotorula*, *Rhodospiridium*, *Cryptococcus*, *Trichosporon* and *Lipomyces*. More specifically, illustrative oil-synthesizing 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*). Alternately, oil biosynthesis may be genetically engineered such that the transformed yeast can produce more than 25% oil of the dry cell weight, and thereby be considered oleaginous.

Most preferred is the oleaginous yeast *Yarrowia lipolytica*. In a further embodiment, most preferred are the *Y. lipolytica* strains designated as ATCC #20362, ATCC #8862, ATCC #18944, ATCC #76982 and/or LGAM S(7)1 (Papanikolaou S., and Aggelis G., *Bioresour. Technol.*, 82(1):43-9 (2002)). The *Y. lipolytica* strain designated as ATCC #20362 was the particular strain from which the *ALK2 Yarrowia* gene and promoter regions encompassed within SEQ ID NO:1 were isolated.

Specific teachings applicable for transformation of oleaginous yeasts (i.e., *Yarrowia lipolytica*) via integration techniques based on linearized fragments of DNA include U.S. Patent 4,880,741 and U.S. Patent 5,071,764 and Chen, D. C. et al. (*Appl. Microbiol. Biotechnol.*, 48(2):232-235 (1997)). Specific teachings applicable for expression of omega-3 fatty acid or omega-6 fatty acid biosynthetic pathway enzymes in the oleaginous yeast *Y. lipolytica* are described in U.S. Patent No. 7,238,482, U.S. Patent No. 7,550,286, U.S. Patent No. 7,588,931, U.S. Pat. No. 7,932,077, U.S. Pat. Appl. Publ. No. 2009-0093543-A1, and U.S. Pat. Appl. Publ. No. 2010-0317072-A1, each incorporated herein by reference in their entirety.

The transformed yeast cell is grown under conditions that optimize expression of the chimeric gene(s). In general, media conditions may be optimized by modifying the type and amount of carbon source, the type and amount of nitrogen source, the carbon-to-nitrogen ratio, the amount of different mineral ions, the oxygen level, growth temperature, pH, length of the biomass production phase, length of the oil accumulation phase and the time and method of cell harvest. Microorganisms of interest, such as oleaginous yeast (e.g.,

Yarrowia lipolytica) are generally grown in a complex medium such as yeast extract-peptone-dextrose broth ["YPD"] or a defined minimal media that lacks a component necessary for growth and thereby forces selection of the desired expression cassettes (e.g., Yeast Nitrogen Base (DIFCO Laboratories, Detroit, MI)).

Fermentation media suitable for the transformed yeast described herein should contain a suitable carbon source. Suitable carbon sources may include, but are not limited to: monosaccharides, disaccharides, oligosaccharides, polysaccharides, sugar alcohols, mixtures from renewable feedstocks, alkanes, fatty acids, esters of fatty acids, glycerol, monoglycerides, diglycerides, triglycerides, phospholipids, various commercial sources of fatty acids, and one-carbon sources, such as are described in U.S. Patent 7,259,255. Hence it is contemplated that the source of carbon utilized may encompass a wide variety of carbon-containing sources and will only be limited by the choice of the yeast species. Although all of the above mentioned carbon sources and mixtures thereof are expected to be suitable herein, preferred carbon sources are sugars (e.g., glucose, invert sucrose, sucrose, fructose and combinations thereof), glycerols, and/or fatty acids (see U.S. Pat. Appl. Publ. No. 2011-0059204 A1).

Nitrogen may be supplied from an inorganic (e.g., $(\text{NH}_4)_2\text{SO}_4$) or organic (e.g., urea or glutamate) source. In addition to appropriate carbon and nitrogen sources, the fermentation media must also contain suitable minerals, salts, cofactors, buffers, vitamins and other components known to those skilled in the art suitable for the growth of the transformed yeast (and optionally, promotion of the enzymatic pathways necessary for omega-3/omega-6 fatty acid production). Particular attention is given to several metal ions, such as Fe^{+2} , Cu^{+2} , Mn^{+2} , Co^{+2} , Zn^{+2} and 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 for the methods and transformed yeast cells described herein are common commercially prepared media, such as Yeast Nitrogen Base (DIFCO Laboratories, Detroit, MI). Other defined or synthetic

growth media may also be used and the appropriate medium for growth of the transformant host cells will be known by one skilled in the art of microbiology or fermentation science. A suitable pH range for the fermentation is typically between about pH 4.0 to pH 8.0, wherein pH 5.5 to pH 7.5 is preferred as the
5 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 omega-3/omega-6 fatty acids in oleaginous yeast cells requires a two-stage process, since the metabolic state
10 must be "balanced" between growth and synthesis/storage of fats. Thus, most preferably, a two-stage fermentation process is necessary for the production of omega-3/omega-6 fatty acids in oleaginous yeast (e.g., *Yarrowia lipolytica*). This approach is described in U.S. Patent No. 7,238,482.

Host cells comprising a suitable coding region of interest operably linked
15 to promoter regions of an *ALK2 Yarrowia* gene may be cultured using methods known in the art. For example, the cell may be cultivated by shake flask cultivation or small-/large-scale fermentation in laboratory or industrial fermentors performed in a suitable medium and under conditions allowing expression of the coding region of interest. Similarly, where commercial production of a product
20 that relies on the instant genetic chimera is desired, a variety of culture methodologies may be applied. For example, large-scale production of a specific gene product over-expressed from a recombinant host may be produced by a batch, fed-batch or continuous fermentation process (see U.S. Patent No. 7,259,255).

25

EXAMPLES

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

30 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. Bennis, 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).

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), New England Biolabs (Ipswich, MA), or Sigma Chemical Company (St. Louis, MO), unless otherwise specified. *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*). DNA sequence was generated on an ABI Automatic sequencer using dye terminator technology (U.S. Pat. No. 5,366,860; EP 272,007) using a combination of vector and insert-specific primers. Sequence editing was performed in Sequencher (Gene Codes Corporation, Ann Arbor, MI). All sequences represent coverage at least two times in both directions. Comparisons of genetic sequences were accomplished using DNASTAR software (DNASTAR Inc., Madison, WI).

The meaning of abbreviations is as follows: "sec" means second(s), "min" means minute(s), "h" means hour(s), "d" means day(s), "μL" means microliter(s), "mL" means milliliter(s), "L" means liter(s), "μM" means micromolar, "mM" means millimolar, "M" means molar, "mmol" means millimole(s), "μmole" means micromole(s), "g" means gram(s), "μg" means microgram(s), "ng" means

nanogram(s), “U” means unit(s), “bp” means base pair(s) and “kB” means kilobase(s).

Nomenclature For Expression Cassettes: The structure of an expression cassette will be represented by a simple notation system of “X::Y::Z”, wherein X describes the promoter fragment, Y describes the gene fragment, and Z describes the terminator fragment, which are all operably linked to one another.

Transformation And Cultivation Of *Yarrowia lipolytica*: *Y. lipolytica* strains with ATCC Accession Nos. #20362, #76982 and #90812 were purchased from the American Type Culture Collection (Rockville, MD). *Y. lipolytica* strains were typically grown at 28-30 °C. Agar plates were prepared as required by addition of 20 g/L agar to the liquid media, according to standard methodology.

EXAMPLE 1

Isolation of the 5' Upstream Regions of the *ALK2* Gene from *Yarrowia lipolytica*

According to the DNA sequence of the *Yarrowia ALK2* locus (ORF “YALI0F01320g”, Dujon, B. et al., *Nature*, 430(6995):35-44 (2004), SEQ ID NO:1 herein), the 1 kB length 5' upstream sequence from the nucleotide ‘A’ (designated as +1) of the translation initiation codon ‘ATG’ was assumed to encode the promoter region (designated herein as ALK2F, SEQ ID NO:2).

In order to study the promoter region upstream of the *ALK2* gene, oligonucleotides Y1198 (SEQ ID NO:3) and Y1199 (SEQ ID NO:4) were designed as primers to amplify a 942 bp fragment, comprising an 876 bp 5' upstream sequence from the nucleotide ‘A’ of the translation initiation codon ‘ATG’ and the first 66 bp of the coding region of the *ALK2* gene. A *Clal* site was included at the 5' portion of oligonucleotide Y1198 (SEQ ID NO:3). In order to incorporate an *NcoI* site at the 3' end of the 942 bp fragment, a cytosine [‘C’] nucleotide was substituted for the guanine [‘G’] nucleotide at position +66 of the *ALK2* gene in oligonucleotide Y1199 (SEQ ID NO:4). The guanine [‘G’] to cytosine [‘C’] mutation did not affect the resulting translated amino acid sequence (i.e., both the wild-type ACG and mutated ACC codons encode a threonine residue).

The 942 bp DNA sequence (comprising the 876 bp 5' upstream sequence and the first 66 bp of the *ALK2* coding region) of the *ALK2* gene was amplified using *Y. lipolytica* strain ATCC #20362 genomic DNA as template and oligonucleotides Y1198 and Y1199 as primers. The PCR amplification was
5 carried out in a 50 µl total volume comprising: PCR buffer (containing 10 mM KCl, 10 mM (NH₄)₂SO₄, 20 mM Tris-HCl (pH 8.75), 2 mM MgSO₄, 0.1% Triton X-100), 100 µg/mL BSA (final concentration), 200 µM each deoxyribonucleotide triphosphate, 10 pmole of each primer and 1 µl of Pfu DNA polymerase (Stratagene, San Diego, CA)). The thermocycler conditions were set for 35
10 cycles at 95 °C for 1 min, 56 °C for 30 sec, and 72 °C for 1 min, followed by a final extension at 72 °C for 10 min.

The PCR product comprising the 942 bp fragment of the *ALK2* gene was purified using a Qiagen PCR purification kit, followed by gel electrophoresis in 1% (weight/volume) agarose. The product was then cloned into the pCR4TOPO
15 vector (Invitrogen, San Diego, CA). The ligated DNA sample was used to transform *E. coli* DH5α cells individually, and transformants were selected on LB agar containing ampicillin (100 µg/mL).

Analyses of the plasmid DNA from transformants confirmed the presence of a 942 bp fragment. The plasmid containing the 942 bp DNA fragment was
20 designated pT-ALK2Pro (SEQ ID NO:5). Sequence analyses showed that pT-ALK2Pro contained a fragment of 942 bp (designated as ALK2N; SEQ ID NO:6), comprising the 876 bp 5' upstream sequence from the nucleotide 'A' of the translation initiation codon 'ATG' of the *ALK2* gene and the first 66 bp of the coding region of the *ALK2* gene. The 876 bp 5' upstream sequence of the *ALK2*
25 gene was designated as ALK2L (wherein the "L" is for "long"; SEQ ID NO:7).

EXAMPLE 2

Modification to ALK2L: Synthesis of Promoters 876 bp ALK2LM, 834 bp ALK2LM1, and 767 bp ALK2LM-C

30 The present Example describes the synthesis of plasmids pT-ALK2Pro-(S), pT-ALK2Pro-(S)S and pT-ALK2Pro-(S)C, each comprising a modified ALK2L

promoter based on removal or insertion of a specific restriction enzyme site, respectively.

Specifically, plasmid pT-ALK2Pro-(S) (SEQ ID NO:8) was generated by site-directed mutagenesis using plasmid pT-ALK2Pro (Example 1) as template, and oligonucleotides Y1202 (SEQ ID NO:9) and Y1203 (SEQ ID NO:10) as primers. The internal *SphI* site (i.e., GCATGC at nucleotides 752-757 of SEQ ID NO:7) of the ALK2L promoter was mutated into GGATGC in plasmid pT-ALK2Pro-(S) (SEQ ID NO:8). The modified ALK2L promoter lacking the internal *SphI* site within plasmid pT-ALK2Pro-(S) was designated as 876 bp ALK2LM (wherein the "LM" is for "long, modified"; SEQ ID NO:11).

Plasmid pT-ALK2Pro-(S)S was generated by site-directed mutagenesis using plasmid pT-ALK2Pro-(S) plasmid as template, and oligonucleotides Y1208 (SEQ ID NO:13) and Y1209 (SEQ ID NO:14) as primers. A *SwaI* site was added by substituting an 'AA' nucleotide sequence for the 'TT' nucleotide sequence at positions 41 and 42 of the ALK2LM (SEQ ID NO:11) promoter in plasmid pT-ALK2Pro-(S)S (SEQ ID NO:12). The 834 bp 5' upstream sequence of the *ALK2* gene in plasmid pT-ALK2Pro-(S)S was designated as 834 bp ALK2LM1 (SEQ ID NO:15).

Plasmid pT-ALK2Pro-(S)C was generated by site-directed mutagenesis using plasmid pT-ALK2Pro-(S) as template, and oligonucleotides Y1262 (SEQ ID NO:17) and Y1263 (SEQ ID NO:18) as primers. A *ClaI* site was generated by insertion of the nucleotide sequence 'TCG' between positions 109 and 110 of the ALK2LM promoter in plasmid pT-ALK2Pro-(S)C (SEQ ID NO:16). The 767 bp 5' upstream sequence of the *ALK2* gene in plasmid pT-ALK2Pro-(S)C was designated as 767 bp ALK2LM-C (SEQ ID NO:19).

EXAMPLE 3

Synthesis and Transformation of an Expression Plasmid Comprising the 834 bp ALK2LM1 Promoter

To perform comparative studies investigating the activity of the 834 bp ALK2LM1 promoter, an expression plasmid was created such that the ALK2LM1 promoter and the first 66 bp of N-terminal protein coding region of the *ALK2* gene

were operably linked to a reporter gene (i.e., the *E. coli* gene encoding β -glucuronidase ("GUS"; Jefferson, R.A., *Nature*, 342(6251):837-838 (1989)).

U.S. Patent No. 7,202,356 describes the synthesis of pDMW212 (FIG. 4A and SEQ ID NO:21 herein), comprising a recombinant FBA::GUS::XPR2
 5 construct. More specifically, this expression cassette comprises an FBA promoter fragment (i.e., 5' upstream untranslated region in front of the 'ATG' translation initiation codon of a fructose-bisphosphate aldolase enzyme [E.C. 4.1.2.13] encoded by the *fbpA* gene and that is necessary for expression), a GUS reporter gene fragment and an XPR2 terminator fragment (comprising ~100 bp of
 10 the 3' region of the *Yarrowia Xpr* gene (GenBank Accession No. M17741)), which are all operably linked to one another.

The *PmeI/NcoI* fragment of pDMW212 (comprising the FBA promoter within the recombinant FBA::GUS::XPR2 construct) was replaced with the 834 bp ALK2LM1 promoter and the 66 bp N-terminal protein coding region of the
 15 *ALK2* gene. Specifically, the *SwaI/NcoI* fragment of pT-ALK2Pro-(S)S (Example 2), comprising the 834 bp ALK2LM1 promoter, and the 66 bp N-terminal protein coding region ("N-22 AA") of the *ALK2* gene were ligated with the *PmeI/NcoI* linearized pDMW212 fragment, thereby creating plasmid pALK2GUS (FIG. 4B; SEQ ID NO:20) comprising a recombinant 834 bp ALK2LM1 promoter plus 66 bp
 20 N-terminal *ALK2* gene coding region::GUS::XPR construct. Thus, plasmid pALK2GUS contains the following components:

Table 3. Description of Plasmid pALK2GUS

RE Sites and Nucleotide position in SEQ ID NO:20	Description of Fragment and Recombinant Construct Components
'GTTTAAAT' (product of ligating <i>SwaI</i> - and <i>PmeI</i> -cut blunt ends) / <i>SacI</i> (8514-2023)	ALK2LM1/N-22 AA::GUS::XPR, comprising: • 834 bp <i>Y. lipolytica</i> ALK2LM1 promoter plus 66 bp N-terminal protein coding sequence ("N-22 AA") of the <i>ALK2</i> gene (SEQ ID NO:15); • GUS: <i>E. coli</i> beta-D-glucuronidase (GenBank Accession No. AAA68923); • XPR2: ~100 bp of the 3' region of <i>Xpr</i> gene of <i>Y. lipolytica</i> (GenBank Accession No. M17741)
3163-2283	ColE1 plasmid origin of replication
4093-3233	Ampicillin-resistance gene (<i>Amp^R</i>) for selection in <i>E. coli</i>
<i>EcoRI</i> / <i>SphI</i>	ARS18: <i>Y. lipolytica</i> centromere and autonomously replicatin

(4923-6269)	sequence 18 (GenBank Accession No. M91600)
<i>Sall</i> / <i>SphI</i> (8505-6269)	Leu2: beta-isopropylmalate dehydrogenase gene of <i>Y. lipolytica</i> (GenBank Accession No. M37309)

Y. lipolytica strain Y4001 has been described in U.S. Patent No. 7,709,239 (Example 3 therein). Strain Y4001, derived from *Yarrowia lipolytica* ATCC #20362, was capable of producing about 17% eicosadienoic acid ["EDA"; 20:2 omega-6] relative to the total lipids. The final genotype of strain Y4001 with respect to wild type *Y. lipolytica* ATCC #20362 was: *Leu*-, *GPD*::*FmD12*::*Pex20*, *EXP1*::*EgD9e*::*Lip1*, *FBAINm*::*EgD9eS*::*Lip2* and *YAT1*::*ME3S*::*Pex16*. Abbreviations are as follows: *FmD12* is a *Fusarium moniliforme* delta-12 desaturase gene [U.S. Pat. No. 7,504,259]; *ME3S* is a codon-optimized *C_{16/18}* elongase gene, derived from *Mortierella alpina* [U.S. Pat. No. 7,470,532]; *EgD9e* is a *Euglena gracilis* delta-9 elongase gene [U.S. Pat. No. 7,645,604]; and, *EgD9eS* is a codon-optimized delta-9 elongase gene, derived from *E. gracilis* [U.S. Pat. No. 7,645,604].

Plasmids pALK2GUS and pDMW212 were transformed separately into *Y. lipolytica* strain Y4001 according to the method of Chen, D. C. et al. (*Appl. Microbiol Biotechnol.*, 48(2):232-235 (1997)) and as described in U.S. Patent No. 7,709,239.

Transformed cells were plated onto Minimal Media ["MM"] plates lacking leucine and maintained at 30 °C for 2 to 3 days (MM comprises per liter: 20 g glucose, 1.7 g yeast nitrogen base without amino acids, 1.0 g proline, and pH 6.1 (do not need to adjust)). Thus, transformants were obtained comprising pALK2GUS and pDMW212 plasmids, respectively.

EXAMPLE 4

Comparative Analyses Of 834 bp ALK2LM1 And FBA Promoter Activities

In *Yarrowia lipolytica* Strain Y4001

The promoter activities of the 834 bp ALK2LM1 promoter (SEQ ID NO:15) and the FBA promoter were determined in *Yarrowia* transformants containing plasmids pALK2GUS and pDMW212, respectively, based on expression of the GUS reporter gene as measured by histochemical assays (Jefferson, R. A. *Plant Mol. Biol. Reporter* 5:387-405 (1987)).

Specifically, *Y. lipolytica* transformants containing plasmids pALK2GUS and pDMW212 respectively were grown from single colonies in 3 mL MM at 30 °C for 2 days. Then, 1 mL of cells was collected by centrifugation. The remaining cultures were centrifuged and washed 2 times with High Glucose Media ["HGM"], resuspended in 3 mL each of HGM and allowed to grow at 30 °C for another 5 days (HGM comprises per liter: 80 g glucose, 2.58 g KH₂PO₄ and 5.36 g K₂HPO₄, pH 7.5 (do not need to adjust)). Cell samples from cultures grown 2 days in MM, as well as cultures grown 2 days in MM and 5 days in HGM were collected by centrifugation, resuspended in 100 ml of histochemical staining buffer, and incubated at 30 °C. Staining buffer was prepared by dissolving 5 mg of 5-bromo-4-chloro-3-indolyl glucuronide ["X-Gluc"] in 50 ml dimethyl formamide, followed by the addition of 5 mL 50 mM NaPO₄, pH 7.0.

The results (Fig. 5) of histochemical staining showed that the 834 bp ALK2LM1 promoter in construct pALK2GUS had almost no activity when the transformed *Yarrowia* cells were grown in MM media. By contrast, very strong expression was observed in *Yarrowia* strains growing in nitrogen-limited HGM media. As expected, the constitutive FBA promoter in construct pDMW212 demonstrated strong activity when pDMW212-transformed cells were grown in either MM or nitrogen-limited HGM media.

Based on the above results, one of skill in the art will therefore recognize that the ALK2LM1 promoter set forth in SEQ ID NO:15 is a strong inducible promoter useful for expression of heterologous and/or homologous genes in transformed yeast, including *Yarrowia*.

EXAMPLE 5

Synthesis and Transformation of Expression Plasmids pALK2GUS-C, pALK2GUS-S, pALK2GUS-P, pALK2GUS-P1, pALK2GUS-P2, pALK2GUS-P5, and pALK2GUS-P6 Comprising 767 bp ALK2LM-C, 838 bp ALK2LM1-S, 835 bp ALK2LM-P, 835 bp ALK2LM1-PP1, 837 bp ALK2LM1-PP2, 839 bp ALK2LM1-PP5, and 840 bp ALK2LM1-PP6 Promoters

Comparative studies were performed to investigate the promoter activity of modified ALK2 promoters having lengths of 767 bp, 838 bp, 835 bp, 835 bp, 837

bp, 839 bp, or 840 bp. Expression plasmids pALK2GUS-C, pALK2GUS-S, pALK2GUS-P, pALK2GUS-P1, pALK2GUS-P2, pALK2GUS-P5, and pALK2GUS-P6 were created, respectively, each comprising a modified ALK2 promoter operably linked to the GUS reporter gene.

5 First, the *Cla*I/*Nco*I fragment of pDG2LGUS (Example 3 and Fig. 4B of U.S. Appl. No. 61/469,933, filed March 31, 2011; SEQ ID NO:22 herein) was replaced with the 767 bp *Cla*I/*Nco*I fragment of pT-ALK2Pro-(S)C (designated ALK2LM-C, SEQ ID NO:19). Specifically, the *Cla*I/*Nco*I fragment of pT-ALK2Pro-(S)C (Example 2), comprising the 767 bp ALK2LM-C promoter and the 66 bp N-
10 terminal protein coding region of the ALK2 gene, was ligated with the *Cla*I/*Nco*I linearized pDG2GUS fragment, thereby creating plasmid pALK2GUS-C (SEQ ID NO:23) comprising a recombinant 767 bp ALK2LM-C plus 66 bp N-terminal protein coding sequence of the ALK2 gene + ::GUS::XPR2 construct.

Second, site-directed mutagenesis was performed using pALK2GUS as
15 template, and oligonucleotides Y2148 (SEQ ID NO:24) and Y2149 (SEQ ID NO:25) as primers. Specifically, primer Y2148 was designed to insert four nucleotides (i.e., 'AAAT') between positions 89 and 90 of SEQ ID NO:15, thereby resulting in creation of a *Swa*I site in the 834 bp ALK2LM1 promoter in the resultant plasmid, pALK2GUS-S (SEQ ID NO:26). The modified ALK2LM1
20 promoter in pALK2GUS-S was 838 bp in length and was designated as 838 bp ALK2LM1-S (SEQ ID NO:27).

Similarly, site-directed mutagenesis was performed using pALK2GUS as template, and oligonucleotides Y1244 (SEQ ID NO:28) and Y1245 (SEQ ID NO:29) as primers. Specifically, the nucleotide sequence 'AAA' was substituted
25 for the nucleotide sequence 'TT' at positions 34 to 35 of the 834 bp ALK2LM1 promoter, thereby introducing a *Pme*I site at position +30 to +37 of ALK2LM1 in the resultant plasmid, pALK2GUS-P (SEQ ID NO:30). The modified ALK2LM1 promoter in pALK2GUS-P was 835 bp in length, and was designated as 835 bp ALK2LM1-P (SEQ ID NO:31).

30 Site-directed mutagenesis was also performed using pALK2GUS-P as template and oligonucleotides Y1250 (SEQ ID NO:32) and Y1251 (SEQ ID NO:33) as primers. Specifically, primer Y1250 was designed to mutate four

oligonucleotides (i.e., 'AAAA' to 'GTTT') at positions 590 to 593 of the ALK2LM1-P promoter, thereby resulting in creation of a *PmeI* site in the 835 bp ALK2LM1-P promoter in the resultant plasmid, pALK2GUS-P1 (SEQ ID NO:34). The modified ALK2LM1-P promoter in pALK2GUS-P1 was 835 bp in length, and was
5 designated as 835 bp ALK2LM1-PP1 (SEQ ID NO:35).

Site-directed mutagenesis was next performed using pALK2GUS-P as template and oligonucleotides Y2158 (SEQ ID NO:36) and Y2159 (SEQ ID NO:37) as primers. Specifically, a *PmeI* site was introduced by the insertion of the nucleotide sequence 'TA' between positions 471 and 472 of the 835 bp
10 ALK2LM1-P promoter in the resulting plasmid pALK2GUS-P2 (SEQ ID NO:38). The modified ALK2LM1-P promoter in pALK2GUS-P2 was 837 bp in length, and was designated as 837 bp ALK2LM1-PP2 (SEQ ID NO:39).

Site-directed mutagenesis was further performed using pALK2GUS-P as template, and oligonucleotides Y2170 (SEQ ID NO:40) and Y2171 (SEQ ID
15 NO:41) as primers. A *PmeI* site was introduced by the insertion of the nucleotide sequence 'TTTA' between positions 324 and 325 of the 835 bp ALK2LM1-P promoter in the resulting plasmid, pALK2GUS-P5 (SEQ ID NO:42). The modified ALK2LM1-P promoter in pALK2GUS-P5 was 839 bp in length, and was designated 839 bp ALK2LM1-PP5 (SEQ ID NO:43).

20 Site-directed mutagenesis was again performed using pALK2GUS-P as template, and oligonucleotides Y2172 (SEQ ID NO:44) and Y2173 (SEQ ID NO:45) as primers. A *PmeI* site was introduced by insertion of the nucleotide sequence 'TTAAA' between positions 189 and 190 of the 835 bp ALK2LM1-P promoter in the resulting plasmid pALK2GUS-P6 (SEQ ID NO:46). The modified
25 ALK2LM1-P promoter in pALK2GUS-P6 was 840 bp in length, and was designated as 840 bp ALK2LM1-PP6 (SEQ ID NO:47).

Thus, pALK2GUS-C (comprising the 767 bp ALK2LM-C promoter (SEQ ID NO:23) plus 66 bp N-terminal protein coding region of the ALK2 gene::GUS::XPR recombinant construct), pALK2GUS-S (comprising the 838 bp ALK2LM1-S
30 promoter (SEQ ID NO:26) plus 66 bp N-terminal protein coding region of the ALK2 gene::GUS::XPR recombinant construct), pALK2GUS-P (comprising the 835 bp ALK2LM1 promoter (SEQ ID NO:30) plus 66 bp N-terminal protein coding

region of the ALK2 gene::GUS::XPR recombinant construct), pALK2GUS-P1 (comprising the 835 bp ALK2LM1-PP1 promoter (SEQ ID NO:34) plus 66 bp N-terminal protein coding region of the ALK2 gene::GUS::XPR recombinant construct), pALK2GUS-P2 (comprising the 837 bp ALK2LM1-PP2 promoter (SEQ ID NO:38) plus 66 bp N-terminal protein coding region of the ALK2 gene::GUS::XPR recombinant construct), pALK2GUS-P5 (comprising the 839 bp ALK2LM1-PP5 promoter (SEQ ID NO:42) plus 66 bp N-terminal protein coding region of the ALK2 gene::GUS::XPR recombinant construct), and pALK2GUS-P6 (comprising the 840 bp ALK2LM1-PP6 promoter (SEQ ID NO:46) plus 66 bp N-terminal protein coding region of the ALK2 gene::GUS::XPR recombinant construct) are identical expression constructs, with the exception that either a 767 bp ALK2LM-C (SEQ ID NO:19), 838 bp ALK2LM1-S (SEQ ID NO:27), 835 bp ALK2LM-P (SEQ ID NO:31), 835 bp ALK2LM1-PP1 (SEQ ID NO:35), 837 bp ALK2LM1-PP2 (SEQ ID NO:39), 839 bp ALK2LM1-PP5 (SEQ ID NO:43), or 840 bp ALK2LM1-PP6 (SEQ ID NO:47) promoter derived from the 5' upstream region and the 66 bp ALK2 N-terminal protein coding region of the *Y. lipolytica* ALK2 gene was operably linked to the GUS::XPR2 recombinant construct.

Plasmids pALK2GUS, pALK2GUS-C, pALK2GUS-S, pALK2GUS-P, pALK2GUS-P1, pALK2GUS-P2, pALK2GUS-P5, and pALK2GUS-P6 were transformed separately into *Y. lipolytica* strain Y4001 as described in Example 3. Transformant cells were plated onto MM plates lacking leucine and maintained at 30 °C for 2 to 3 days. Thus, transformants were obtained comprising the pALK2GUS-C, pALK2GUS-S, pALK2GUS-P, pALK2GUS-P1, pALK2GUS-P2, pALK2GUS-P5, and pALK2GUS-P6 plasmids, respectively.

EXAMPLE 6

Comparative Analysis of 834 bp ALK2LM1, 767 bp ALK2LM-C, 838 bp ALK2LM1-S, 835 bp ALK2LM1-P, 835 bp ALK2LM1-PP1, 837 bp ALK2LM1-PP2, 839 bp ALK2LM1-PP5, and 840 bp ALK2LM1-PP6 Promoter Activities in

Yarrowia lipolytica Strain Y4001

The promoter activities of the 834 bp ALK2LM1 (SEQ ID NO:15), 767 bp ALK2LM-C (SEQ ID NO:19), 838 bp ALK2LM1-S (SEQ ID NO:27), 835 bp

ALK2LM1-P (SEQ ID NO:31), 835 bp ALK2LM1-PP1 (SEQ ID NO:35), 837 bp ALK2LM1-PP2 (SEQ ID NO:39), 839 bp ALK2LM1-PP5 (SEQ ID NO:43), and 840 bp ALK2LM1-PP6 (SEQ ID NO:47) promoters were determined in *Yarrowia* transformants containing pALK2GUS, pALK2GUS-C, pALK2GUS-S, pALK2GUS-P, pALK2GUS-P1, pALK2GUS-P2, pALK2GUS-P5, and pALK2GUS-P6, respectively. GUS activity in each expression construct was measured by histochemical assays as described in Example 4.

The results of histochemical staining showed that the promoter activities of 767 bp ALK2LM-C (SEQ ID NO:19), 838 bp ALK2LM1-S (SEQ ID NO:27), 835 bp ALK2LM1-P (SEQ ID NO:31), 835 bp ALK2LM1-PP1 (SEQ ID NO:35), 837 bp ALK2LM1-PP2 (SEQ ID NO:39), 839 bp ALK2LM1-PP5 (SEQ ID NO:43), and 840 bp ALK2LM1-PP6 (SEQ ID NO:47), were equivalent to the 834 bp ALK2LM1 (SEQ ID NO:15, Example 4) promoter. Specifically, each of the promoters was almost completely inactive when their respective, transformed *Yarrowia* cells were grown in MM, but were very strongly active when their constructs were expressed by cells grown in nitrogen-limited HGM media. These results indicated that the promoter activities of the 767 bp ALK2LM-C, 838 bp ALK2LM1-S, 835 bp ALK2LM1-P, 835 bp ALK2LM1-PP1, 837 bp ALK2LM1-PP2, 839 bp ALK2LM1-PP5, and 840 bp ALK2LM1-PP6 promoters were not affected by the various modifications performed in Example 5.

EXAMPLE 7

Synthesis and Transformation of Expression Plasmids pALK2GUS-P8, pALK2GUS-P7, pALK2GUS-P3, and pALK2GUS-P4 Comprising 646 bp ALK2LM-P8, 511 bp ALK2GUS-P7, 364 bp ALK2GUS-P3, and 242 bp ALK2GUS-P4 Promoters

Comparative studies were performed to investigate the promoter activity of truncated ALK2 promoters having lengths of 646 bp, 511 bp, 364 bp, and 242 bp, using expression plasmids pALK2GUS-P8, pALK2GUS-P7, pALK2GUS-P3, and pALK2GUS-P4, respectively.

Plasmid pALK2GUS-P6 (Example 5, SEQ ID NO:46) was digested with *PmeI*, and the large *PmeI* fragment of pALK2GUS-P6 was self-ligated to

generate pALK2GUS-P8 (SEQ ID NO:48). The ALK2 promoter fragment in pALK2GUS-P8 was 646 bp in length and was designated as 646 bp ALK2LM-P8 (SEQ ID NO:49).

5 Plasmid pALK2GUS-P5 (Example 5, SEQ ID NO:42) was digested with *PmeI*, and the large *PmeI* fragment of pALK2GUS-P5 was self-ligated to generate pALK2GUS-P7 (SEQ ID NO:50). The ALK2 promoter fragment in pALK2GUS-P7 was 511 bp in length and was designated as 511 bp ALK2LM-P7 (SEQ ID NO: 51).

10 Plasmid pALK2GUS-P2 (Example 5, SEQ ID NO:38) was digested with *PmeI*, and the large *PmeI* fragment of pALK2GUS-P2 was self-ligated to generate pALK2GUS-P3 (SEQ ID NO:52). The ALK2 promoter fragment in pALK2GUS-P3 was 364 bp in length and was designated as 364 bp ALK2LM-P3 (SEQ ID NO:53)

15 Plasmid pALK2GUS-P1 (Example 5, SEQ ID NO:34) was digested with *PmeI*, and the large *PmeI* fragment of pALK2GUS-P1 was self-ligated to generate pALK2GUS-P4 (SEQ ID NO:54). The ALK2 promoter fragment in pALK2GUS-P4 was 242 bp in length and was designated as 242 bp ALK2LM-P4 (SEQ ID NO:55).

20 Thus, plasmids pALK2GUS (Example 3, comprising the 834 bp ALK2LM1 promoter (SEQ ID NO:20) plus 66 bp N-terminal protein coding region of the ALK2 gene::GUS::XPR recombinant construct), pALK2GUS-P8 (comprising the 646 bp ALK2LM-P8 promoter (SEQ ID NO:48) plus 66 bp N-terminal protein coding region of the ALK2 gene::GUS::XPR recombinant construct), pALK2GUS-P7 (comprising the 511 bp ALK2LM-P7 promoter (SEQ ID NO:50) plus 66 bp N-
25 terminal protein coding region of the ALK2 gene::GUS::XPR recombinant construct), pALK2GUS-P3 (comprising the 364 bp ALK2LM-P3 promoter (SEQ ID NO:52) plus 66 bp N-terminal protein coding region of the ALK2 gene::GUS::XPR recombinant construct), and pALK2GUS-P4 (comprising the 242 bp ALK2LM-P4 promoter (SEQ ID NO:54) plus 66 bp N-terminal protein
30 coding region of the ALK2 gene::GUS::XPR recombinant construct) are identical expression constructs, with the exception that either an 834 bp ALK2LM1, 646 bp ALK2LM-P8, 511 bp ALK2GUS-P7, 364 bp ALK2GUS-P3, or 242 bp ALK2GUS-

P4 promoter derived from the 5' upstream region and the 66 bp ALK2 N-terminal protein coding region of the *Y. lipolytica* ALK2 gene was operably linked to the GUS::XPR2 recombinant construct.

Plasmids pALK2GUS, pALK2GUS-P8, pALK2GUS-P7, pALK2GUS-P3, and pALK2GUS-P4 were transformed separately into *Y. lipolytica* strain Y4001 as described in Example 3. Transformant cells were plated onto MM plates lacking leucine and maintained at 30 °C for 2 to 3 days. Thus, transformants were obtained comprising the pALK2GUS, pALK2GUS-P8, pALK2GUS-P7, pALK2GUS-P3, and pALK2GUS-P4 plasmids, respectively.

10

EXAMPLE 8

Comparative Analyses of 834 bp ALK2LM1, 646 bp ALK2LM-P8, 511 bp ALK2LM-P7, 364 bp ALK2LM-P3, and 242 bp ALK2LM-P4 Promoter Activities in *Yarrowia lipolytica* Strain Y4001

The promoter activities of the 834 bp ALK2LM1 (SEQ ID NO:15), 646 bp ALK2LM-P8 (SEQ ID NO:49), 511 bp ALK2LM-P7 (SEQ ID NO:51), 364 bp ALK2LM-P3 (SEQ ID NO:53), and 242 bp ALK2LM-P4 (SEQ ID NO:55) promoters were determined in *Yarrowia* transformants containing pALK2GUS, pALK2GUS-P8, pALK2GUS-P7, pALK2GUS-P3, and pALK2GUS-P4, respectively. GUS activity in each expressed construct was measured by histochemical assays as described in Example 4.

The results of histochemical staining showed that the activities of the 646 bp ALK2LM-P8 and 511 bp ALK2LM-P7 promoters were equivalent to the 834 bp ALK2LM1 promoter (SEQ ID NO:15, Example 4). Specifically, the 646 bp ALK2LM-P8 and 511 bp ALK2LM-P7 promoters were almost completely inactive when their respective transformed *Yarrowia* cells were grown in MM, but were very strongly active when their constructs were expressed by cells grown in nitrogen-limited HGM media. The 346 bp ALK2LM-P3 and 242 bp ALK2LM-P4 promoters were similarly inactive when their respective transformed *Yarrowia* cells were grown in MM, but were weakly active when their constructs were expressed by cells grown in nitrogen-limited HGM media. These results indicated that the DNA sequence responsible for nitrogen-limited induction was

located between positions 324 and 470 of the ALK2LM1 promoter (SEQ ID NO:15) (corresponding to positions -511 and -364 respectively, in SEQ ID NO:1, where the 'A' of the ATG start codon is designated as +1). This particular sequence is set forth herein as SEQ ID NO:57.

5 Based on the above results, one of skill in the art will recognize that the ALK2LM promoter set forth as SEQ ID NO:11 can be truncated and retain promoter activity. Specifically, deleting the region defined by nucleotide positions 1 to 42 bp of SEQ ID NO:11 resulted in the active mutant promoter described herein as 834 bp ALK2LM1; deleting the region defined by nucleotide positions 1 to 109 bp of SEQ ID NO:11 resulted in the active mutant promoter described herein as 767 bp ALK2LM-C; deleting the region defined by nucleotide positions 1 to 230 bp of SEQ ID NO:11 resulted in the active mutant promoter described herein as 646 bp ALK2LM-P8; deleting the region defined by nucleotide positions 1 to 365 bp of SEQ ID NO:11 resulted in the active mutant promoter described herein as 511 bp ALK2LM-P7; deleting the region defined by nucleotide positions 1 to 512 bp of SEQ ID NO:11 resulted in the weakly active mutant promoter described herein as 364 bp ALK2LM-P3; and deleting the region defined by nucleotide positions 1 to 634 bp of SEQ ID NO:11 resulted in the weakly active mutant promoter described herein as 242 bp ALK2LM-P4 (FIG. 1). It is therefore
20 assumed that a variety of modified ALK2 promoters could be utilized for expression of a coding region of interest in a *Yarrowia* host cell, wherein the promoter optionally comprises at least one modification selected from the group consisting of: a deletion at the 5'-terminus of 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177,

178, 179, 180, 181, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193,
194, 195, 196, 197, 198, 199, 200, 201, 202, 203, 204, 205, 206, 207, 208, 209,
210, 211, 212, 213, 214, 215, 216, 217, 218, 219, 220, 221, 222, 223, 224, 225,
226, 227, 228, 229, 230, 231, 232, 233, 234, 235, 236, 237, 238, 239, 240, 241,
5 242, 243, 244, 245, 246, 247, 248, 249, 250, 251, 252, 253, 254, 255, 256, 257,
258, 259, 260, 261, 262, 263, 264, 265, 266, 267, 268, 269, 270, 271, 272, 273,
274, 275, 276, 277, 278, 279, 280, 281, 282, 283, 284, 285, 286, 287, 288, 289,
290, 291, 292, 293, 294, 295, 296, 297, 298, 299, 300, 301, 302, 303, 304, 305,
306, 307, 308, 309, 310, 311, 312, 313, 314, 315, 316, 317, 318, 319, 320, 321,
10 322, 323, 324, 325, 326, 327, 328, 329, 330, 331, 332, 333, 334, 335, 336, 337,
338, 339, 340, 341, 342, 343, 344, 345, 346, 347, 348, 349, 350, 351, 352, 353,
354, 355, 356, 357, 358, 359, 360, 361, 362, 363, 364, 365, 366, 367, 368, 369,
370, 371, 372, 373, 374, 375, 376, 377, 378, 379, 380, 381, 382, 383, 384, 385,
386, 387, 388, 389, 390, 391, 392, 393, 394, 395, 396, 397, 398, 399, 400, 401,
15 402, 403, 404, 405, 406, 407, 408, 409, 410, 411, 412, 413, 414, 415, 416, 417,
418, 419, 420, 421, 422, 423, 424, 425, 426, 427, 428, 429, 430, 431, 432, 433,
434, 435, 436, 437, 438, 439, 440, 441, 442, 443, 444, 445, 446, 447, 448, 449,
450, 451, 452, 453, 454, 455, 456, 457, 458, 459, 460, 461, 462, 463, 464, 465,
466, 467, 468, 469, 470, 471, 472, 473, 474, 475, 476, 477, 478, 479, 480, 481,
20 482, 483, 484, 485, 486, 487, 488, 489, 490, 491, 492, 493, 494, 495, 496, 497,
498, 499, 500, 501, 502, 503, 504, 505, 506, 507, 508, 509, 510, 511, 512, 513,
514, 515, 516, 517, 518, 519, 520, 521, 522, 523, 524, 525, 526, 527, 528, 529,
530, 531, 532, 533, 534, 535, 536, 537, 538, 539, 540, 541, 542, 543, 544, 545,
546, 547, 548, 549, 550, 551, 552, 553, 554, 555, 556, 557, 558, 559, 560, 561,
25 562, 563, 564, 565, 566, 567, 568, 569, 570, 571, 572, 573, 574, 575, 576, 577,
578, 579, 580, 581, 582, 583, 584, 585, 586, 587, 588, 589, 590, 591, 592, 593,
594, 595, 596, 597, 598, 599, 600, 601, 602, 603, 604, 605, 606, 607, 608, 609,
610, 611, 612, 613, 614, 615, 616, 617, 618, 619, 620, 621, 622, 623, 624, 625,
626, 627, 628, 629, 630, 631, 632, 633, or 634 consecutive nucleotides, wherein
30 the first nucleotide deleted is the adenine nucleotide ['A'] at position 1 of SEQ ID
NO:7. It is further assumed that SEQ ID NO:57 could be used as an enhancer to

elevate levels of transcription of an operably linked eukaryotic promoter, thereby increasing transcription of an adjacent gene.

FIGs. 2A, 2B, 2C, 2D, 2E, 2F, 2G, 2H, 2I, and 2J is an alignment of the following *Y. lipolytica* ALK2 promoter regions described herein: the ALK2 (SEQ ID NO:2) promoter region, which is the 1000 bp 5' upstream sequence (i.e., the -1000 to -1 region) of the n-alkane-hydroxylating cytochrome P450 ["ALK2"] gene in *Y. lipolytica*, wherein the nucleotide 'A' of the ALK2 translation initiation codon 'ATG' was designated as +1; the 942 bp ALK2N (SEQ ID NO:6) promoter region, which comprises the 876 bp 5' upstream sequence (i.e., the -876 to -1 region) of the n-alkane-hydroxylating cytochrome P450 ["ALK2"] gene and the first 66 bp of the protein coding region, wherein the nucleotide 'A' of the ALK2 translation initiation codon 'ATG' was designated as +1; the 876 bp ALK2L (SEQ ID NO:7) promoter region; the 876 bp ALK2LM (SEQ ID NO:11) promoter region; the 838 bp ALK2LM1-S (SEQ ID NO:27) promoter region; the 834 bp ALK2LM1 (SEQ ID NO:15) promoter region; the 840 bp ALK2LM1-PP6 (SEQ ID NO:47) promoter region; the 839 bp ALK2LM1-PP5 (SEQ ID NO:43) promoter region; the 837 bp ALK2LM1-PP2 (SEQ ID NO:39) promoter region; the 835 bp ALK2LM1-PP1 (SEQ ID NO:35) promoter region; the 835 bp ALK2LM1-P (SEQ ID NO:31) promoter region; the 767 bp ALK2LM-C (SEQ ID NO:19) promoter region; the 646 bp ALK2LM-P8 (SEQ ID NO:49) promoter region; the 511 bp ALK2LM-P7 (SEQ ID NO:51) promoter region; the 364 bp ALK2LM-P3 (SEQ ID NO:53) promoter region; the 242 bp ALK2LM-P4 (SEQ ID NO:55) promoter region; and the 68 bp minimal/core ALK2 (SEQ ID NO:56) promoter region. Sequence differences are noted with an arrow over the alignment and a box.

All of the modified promoters derived from the ALK2L promoter set forth as SEQ ID NO:7 (e.g., 876 bp ALK2LM, 646 bp ALK2LM-P8, 511 bp ALK2LM-P7, 364 bp ALK2LM-P3, and 242 bp ALK2LM-P4) comprise a cytosine ['C'] to guanine ['G'] substitution at the nucleotide corresponding to position 753 of SEQ ID NO:7. Therefore, one of skill in the art will appreciate that a suitable promoter region of an *ALK2 Yarrowia* gene may optionally comprise a cytosine ['C'] or a guanine ['G'] nucleotide at the nucleotide corresponding to position 753 of SEQ ID NO:7; alternately, the modified ALK2 promoter may also tolerate substitution

of an adenine ['A'] nucleotide or a thymine ['T'] nucleotide for the wildtype cytosine ['C'] at this position.

EXAMPLE 9

5 Sequence Analysis of Promoter Regions of an ALK2 *Yarrowia* Gene

This Example describes the analysis and definition of the ALK2 minimal promoter region.

Although promoters interact with the TATA binding protein ["TBP"] to create a transcription initiation complex from which RNA polymerase II transcribes the DNA coding sequence, only some promoters contain a TATA box to which TBP binds directly while other promoters are TATA-less promoters. The "TATA box" or "Goldberg-Hogness box" is a DNA sequence (i.e., cis-regulatory element) found in the promoter region of some genes in archaea and eukaryotes. For example, approximately 24% of human genes contain a TATA box within the core promoter (Yang C, et al., *Gene*, 389:52-65 (2007)); phylogenetic analysis of six *Saccharomyces* species revealed that about 20% of the 5,700 yeast genes contained a TATA-box element (Basehoar *et al.*, *Cell*, 116:699–709 (2004)). The TATA box has a core DNA sequence of 5'-TATAAA-3' or a variant thereof and is usually located ~200 to 25 base pairs upstream of the transcriptional start site.

20 The transcription initiation complex forms at the site of the TATA box (Smale, and Kadonaga, T., *Ann. Rev. Biochem.* 72:449–479 (2003)). This complex comprises the TATA binding protein, RNA polymerase II, and various transcription factors (i.e., TFIID, TFIIA, TFIIB, TFIIF, TFIIE and TFIIH). Both the TATA box itself and the distance between the TATA box and transcription start site affect activity of TATA box-containing promoters in eukaryotes (Zhu et al.,

25 *Plant Cell* 7:1681-1689 (1995)).

The genes within *Yarrowia* can be largely classified into three classes according to their promoter sequences. Specifically, the first class of genes includes those comprising a TATA box, usually, ~130 to 20 base pairs upstream of the gene's transcription start site. The second class of genes includes those

30 comprising an initiator element(s) around the gene's transcription start site. And,

the third class of genes lacks both a TATA box and initiator element in the gene's promoter region.

Analysis of the sequence of the 942 bp ALK2N promoter region (Example 1, SEQ ID NO:6) revealed that it contains a typical TATA-box (TATATAA) at the nucleotides corresponding to positions -68 to -62 of SEQ ID NO:1, wherein the nucleotide 'A' of the ALK2 translation initiation codon 'ATG' was designated as +1. The 68 bp fragment 5' upstream spanning positions -68 to -1 of the *ALK2* gene is believed to be necessary for promoter activity. Based on identification of this 68 bp fragment, it is believed that a suitable minimal *ALK2* promoter region for basal level transcription initiation would comprise this fragment, set forth herein as SEQ ID NO:56.

EXAMPLE 10

Comparison of Various *Yarrowia* ALK2 Promoter Regions

The present Example summarizes the relative activity of various *ALK2* promoter regions exemplified in Examples 4, 6, and 8, above.

It was concluded that the ALK2 promoter is a very strong inducible promoter that can be used to drive high level expression of various genes in engineered *Yarrowia* strains under nitrogen-limited conditions, as shown in Table 5 below. The results also demonstrated the presence of an enhancer responsible for induced expression under nitrogen-limited conditions located between positions +511 to +365 of SEQ ID. NO:6.

Table 4. Summary of Relative Activity of Various ALK2 Promoter Regions

Construct Comprising GUS Reporter	Promoter Operably Linked to GUS Reporter	Promo ter Length	Promoter Activity	
			Cultured in MM*	Cultured in MM+HGM**
pALK2GUS (SEQ ID NO:20)	ALK2LM1 (SEQ ID NO:15)	834 bp	-	+++
pT-ALK2Pro-(S)C (SEQ ID NO:16)	ALK2LM-C (SEQ ID NO:19)	767 bp	-	+++
pALK2GUS-S (SEQ ID NO:26)	ALK2LM1-S (SEQ ID NO:27)	838 bp	-	+++
pALK2GUS-P (SEQ ID NO:30)	ALK2LM1-P (SEQ ID NO:31)	835 bp	-	+++
pALK2GUS-P1	ALK2LM1-PP1	835 bp	-	+++

(SEQ ID NO:34)	(SEQ ID NO:35)			
pALK2GUS-P2 (SEQ ID NO:38)	ALK2LM1-PP2 (SEQ ID NO:39)	837 bp	-	+++
pALK2GUS-P5 (SEQ ID NO:42)	ALK2LM1-PP5 (SEQ ID NO:43)	839 bp	-	+++
pALK2GUS-P6 (SEQ ID NO:46)	ALK2LM1-PP6 (SEQ ID NO:47)	840 bp	-	+++
pALK2GUS-P8 (SEQ ID NO:48)	ALK2LM-P8 (SEQ ID NO:49)	646 bp	-	+++
pALK2GUS-P7 (SEQ ID NO:50)	ALK2LM-P7 (SEQ ID NO:51)	511 bp	-	+++
pALK2GUS-P3 (SEQ ID NO:52)	ALK2LM-P3 (SEQ ID NO:53)	364 bp	-	-/+
pALK2GUS-P4 (SEQ ID NO:55)	ALK2LM-P4 (SEQ ID NO:55)	242 bp	-	-/+

* Cultured In MM refers to 2 days growth in MM.

** Cultured in MM+ HGM refers to 2 days growth in MM, followed by 3 days growth in HGM.

CLAIMS

What is claimed is:

1. A method for the expression of a coding region of interest in a transformed yeast cell comprising:
 - a) providing a transformed yeast cell having a recombinant construct, wherein the recombinant construct comprises:
 - (1) a promoter region of an *ALK2 Yarrowia* gene; and
 - (2) a coding region of interest which is expressible in the yeast cell;
 wherein the promoter region is operably linked to the coding region of interest; and
 - b) growing the transformed yeast cell of step (a) under conditions whereby the recombinant construct of step (a) is expressed.
2. The method according to Claim 1, wherein the promoter region of the *ALK2 Yarrowia* gene comprises SEQ ID NO:56.
3. The method according to Claim 1, wherein the promoter region of the *ALK2 Yarrowia* gene comprises SEQ ID NO:6, wherein said promoter optionally comprises at least one modification selected from the group consisting of:
 - a) a deletion at the 5'-terminus of 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172,

173, 174, 175, 176, 177, 178, 179, 180, 181, 182, 183, 184, 185, 186,
187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 197, 198, 199, 200,
201, 202, 203, 204, 205, 206, 207, 208, 209, 210, 211, 212, 213, 214,
215, 216, 217, 218, 219, 220, 221, 222, 223, 224, 225, 226, 227, 228,
229, 230, 231, 232, 233, 234, 235, 236, 237, 238, 239, 240, 241, 242,
243, 244, 245, 246, 247, 248, 249, 250, 251, 252, 253, 254, 255, 256,
257, 258, 259, 260, 261, 262, 263, 264, 265, 266, 267, 268, 269, 270,
271, 272, 273, 274, 275, 276, 277, 278, 279, 280, 281, 282, 283, 284,
285, 286, 287, 288, 289, 290, 291, 292, 293, 294, 295, 296, 297, 298,
299, 300, 301, 302, 303, 304, 305, 306, 307, 308, 309, 310, 311, 312,
313, 314, 315, 316, 317, 318, 319, 320, 321, 322, 323, 324, 325, 326,
327, 328, 329, 330, 331, 332, 333, 334, 335, 336, 337, 338, 339, 340,
341, 342, 343, 344, 345, 346, 347, 348, 349, 350, 351, 352, 353, 354,
355, 356, 357, 358, 359, 360, 361, 362, 363, 364, 365, 366, 367, 368,
369, 370, 371, 372, 373, 374, 375, 376, 377, 378, 379, 380, 381, 382,
383, 384, 385, 386, 387, 388, 389, 390, 391, 392, 393, 394, 395, 396,
397, 398, 399, 400, 401, 402, 403, 404, 405, 406, 407, 408, 409, 410,
411, 412, 413, 414, 415, 416, 417, 418, 419, 420, 421, 422, 423, 424,
425, 426, 427, 428, 429, 430, 431, 432, 433, 434, 435, 436, 437, 438,
439, 440, 441, 442, 443, 444, 445, 446, 447, 448, 449, 450, 451, 452,
453, 454, 455, 456, 457, 458, 459, 460, 461, 462, 463, 464, 465, 466,
467, 468, 469, 470, 471, 472, 473, 474, 475, 476, 477, 478, 479, 480,
481, 482, 483, 484, 485, 486, 487, 488, 489, 490, 491, 492, 493, 494,
495, 496, 497, 498, 499, 500, 501, 502, 503, 504, 505, 506, 507, 508,
509, 510, 511, 512, 513, 514, 515, 516, 517, 518, 519, 520, 521, 522,
523, 524, 525, 526, 527, 528, 529, 530, 531, 532, 533, 534, 535, 536,
537, 538, 539, 540, 541, 542, 543, 544, 545, 546, 547, 548, 549, 550,
551, 552, 553, 554, 555, 556, 557, 558, 559, 560, 561, 562, 563, 564,
565, 566, 567, 568, 569, 570, 571, 572, 573, 574, 575, 576, 577, 578,
579, 580, 581, 582, 583, 584, 585, 586, 587, 588, 589, 590, 591, 592,
593, 594, 595, 596, 597, 598, 599, 600, 601, 602, 603, 604, 605, 606,
607, 608, 609, 610, 611, 612, 613, 614, 615, 616, 617, 618, 619, 620,

621, 622, 623, 624, 625, 626, 627, 628, 629, 630, 631, 632, 633, or 634 consecutive nucleotides, wherein the first nucleotide deleted is the adenine nucleotide ['A'] at position 1 of SEQ ID NO:6;

- b) a deletion at the 3'-terminus of 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, or 66 consecutive nucleotides, wherein the first nucleotide deleted is the cytosine ['C'] nucleotide at position 942 of SEQ ID NO:6;
- c) substitution of a guanine ['G'] nucleotide for the cytosine ['C'] nucleotide at position 753 of SEQ ID NO:6;
- d) substitution of an adenine ['A'] nucleotide or a thymine ['T'] nucleotide for the cytosine ['C'] nucleotide at position 753 of SEQ ID NO:6;
- e) substitution of the nucleotide sequence 'AA' for the nucleotide sequence 'TT' at positions 41 and 42 of SEQ ID NO:6;
- f) insertion of the nucleotide sequence 'TCG' between positions 109 and 110 of SEQ ID NO:6;
- g) insertion of the nucleotide sequence 'AAAT' between position 131 and position 132 of SEQ ID NO:6;
- h) substitution of the nucleotide sequence 'AAA' for the nucleotide sequence 'TT' at positions 76 to 77 of SEQ ID NO:6;
- i) substitution of the nucleotide sequence 'GTTT' for the nucleotide sequence 'AAAA' at positions 631 to 634 of SEQ ID NO:6;
- j) insertion of the nucleotide sequence 'TA' between position 512 and position 513 of SEQ ID NO:6;
- k) insertion of the nucleotide sequence 'TTTA' between position 365 and position 366 of SEQ ID NO:6;
- l) insertion of the nucleotide sequence 'TTAAA' between position 230 and position 231 of SEQ ID NO:6; and
- m) any combination of part a), part b), part c), part d), part e), part f), part g), part h), part i), part j), part k), and part l).

4. The method according to Claim 3, wherein the promoter region of the *ALK2 Yarrowia* gene comprises a sequence selected from the group consisting of SEQ ID NO:7, SEQ ID NO:11, SEQ ID NO:27, SEQ ID NO:15, SEQ ID NO:47, SEQ ID NO:43, SEQ ID NO:39, SEQ ID NO:35, SEQ ID NO:31, SEQ ID NO:19, SEQ ID NO:49, SEQ ID NO:51, SEQ ID NO:53 and SEQ ID NO:55.
5. The method according to Claim 4, wherein the promoter region of the *ALK2 Yarrowia* gene comprises SEQ ID NO:55, and wherein said promoter region optionally comprises at least one modification selected from the group consisting of:
 - a) a deletion at the 5'-terminus of 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172, 173, or 174 consecutive nucleotides, wherein the first nucleotide deleted is the adenine nucleotide ['A'] at position 1 of SEQ ID NO:55; and
 - b) a deletion of part (a) in combination with a substitution of a thymine ['T'] nucleotide, adenine ['A'] nucleotide or a cytosine ['C'] nucleotide for the guanine ['G'] nucleotide at position 119 of SEQ ID NO:55.
6. The method according to Claim 1, wherein the transformed yeast cell is an oleaginous yeast cell.

7. The method of Claim 6, wherein the oleaginous yeast cell is a member of a genus selected from the group consisting of *Yarrowia*, *Candida*, *Rhodotorula*, *Rhodospiridium*, *Cryptococcus*, *Trichosporon* and *Lipomyces*.
8. The method according to Claim 1, wherein the coding region of interest encodes a polypeptide, wherein the polypeptide is selected from the group consisting of: desaturases, elongases, acyltransferases, aminopeptidases, amylases, carbohydrases, carboxypeptidases, catalases, cellulases, chitinases, cutinases, cyclodextrin glycosyltransferases, deoxyribonucleases, esterases, alpha-galactosidases, beta-galactosidases, glucoamylases, alpha-glucosidases, beta-glucanases, beta-glucosidases, invertases, laccases, lipases, mannosidases, mutanases, oxidases, pectinolytic enzymes, peroxidases, phospholipases, phosphatases, phytases, polyphenoloxidases, proteolytic enzymes, ribonucleases, transglutaminases and xylanases.
9. The method according to Claim 1, wherein the promoter region of the ALK2 *Yarrowia* gene further (i) comprises an enhancer region set forth in SEQ ID NO:57, or (ii) is operably linked to said enhancer region.
10. The method according to Claim 1, wherein
the coding region encodes at least one omega-3 fatty acid or omega-6 fatty acid biosynthetic pathway enzyme, and
wherein an omega-3 fatty acid or omega-6 fatty acid is produced in step b).
11. The method according to Claim 10, wherein the omega-3 fatty acid or omega-6 fatty acid biosynthetic pathway enzyme is selected from the group consisting of desaturases and elongases.

12. An isolated nucleic acid molecule comprising a promoter region of an *ALK2* *Yarrowia* gene as set forth in SEQ ID NO:6, wherein said promoter region optionally comprises at least one modification selected from the group consisting of:
- a) a deletion at the 5'-terminus of 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177, 178, 179, 180, 181, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 197, 198, 199, 200, 201, 202, 203, 204, 205, 206, 207, 208, 209, 210, 211, 212, 213, 214, 215, 216, 217, 218, 219, 220, 221, 222, 223, 224, 225, 226, 227, 228, 229, 230, 231, 232, 233, 234, 235, 236, 237, 238, 239, 240, 241, 242, 243, 244, 245, 246, 247, 248, 249, 250, 251, 252, 253, 254, 255, 256, 257, 258, 259, 260, 261, 262, 263, 264, 265, 266, 267, 268, 269, 270, 271, 272, 273, 274, 275, 276, 277, 278, 279, 280, 281, 282, 283, 284, 285, 286, 287, 288, 289, 290, 291, 292, 293, 294, 295, 296, 297, 298, 299, 300, 301, 302, 303, 304, 305, 306, 307, 308, 309, 310, 311, 312, 313, 314, 315, 316, 317, 318, 319, 320, 321, 322, 323, 324, 325, 326, 327, 328, 329, 330, 331, 332, 333, 334, 335, 336, 337, 338, 339, 340, 341, 342, 343, 344, 345, 346, 347, 348, 349, 350, 351, 352, 353, 354, 355, 356, 357, 358, 359, 360, 361, 362, 363, 364, 365, 366, 367, 368, 369, 370, 371, 372, 373, 374, 375, 376, 377, 378, 379, 380, 381, 382, 383, 384, 385, 386, 387, 388, 389, 390, 391, 392, 393, 394, 395, 396, 397, 398, 399, 400, 401, 402, 403, 404, 405, 406, 407, 408, 409, 410,

411, 412, 413, 414, 415, 416, 417, 418, 419, 420, 421, 422, 423, 424, 425, 426, 427, 428, 429, 430, 431, 432, 433, 434, 435, 436, 437, 438, 439, 440, 441, 442, 443, 444, 445, 446, 447, 448, 449, 450, 451, 452, 453, 454, 455, 456, 457, 458, 459, 460, 461, 462, 463, 464, 465, 466, 467, 468, 469, 470, 471, 472, 473, 474, 475, 476, 477, 478, 479, 480, 481, 482, 483, 484, 485, 486, 487, 488, 489, 490, 491, 492, 493, 494, 495, 496, 497, 498, 499, 500, 501, 502, 503, 504, 505, 506, 507, 508, 509, 510, 511, 512, 513, 514, 515, 516, 517, 518, 519, 520, 521, 522, 523, 524, 525, 526, 527, 528, 529, 530, 531, 532, 533, 534, 535, 536, 537, 538, 539, 540, 541, 542, 543, 544, 545, 546, 547, 548, 549, 550, 551, 552, 553, 554, 555, 556, 557, 558, 559, 560, 561, 562, 563, 564, 565, 566, 567, 568, 569, 570, 571, 572, 573, 574, 575, 576, 577, 578, 579, 580, 581, 582, 583, 584, 585, 586, 587, 588, 589, 590, 591, 592, 593, 594, 595, 596, 597, 598, 599, 600, 601, 602, 603, 604, 605, 606, 607, 608, 609, 610, 611, 612, 613, 614, 615, 616, 617, 618, 619, 620, 621, 622, 623, 624, 625, 626, 627, 628, 629, 630, 631, 632, 633, or 634 consecutive nucleotides, wherein the first nucleotide deleted is the adenine nucleotide ['A'] at position 1 of SEQ ID NO:6;

- b) a deletion at the 3'-terminus of 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, or 66 consecutive nucleotides, wherein the first nucleotide deleted is the cytosine ['C'] nucleotide at position 942 of SEQ ID NO:6;
- c) substitution of a guanine ['G'] nucleotide for the cytosine ['C'] nucleotide at position 753 of SEQ ID NO:6;
- d) substitution of an adenine ['A'] nucleotide or a thymine ['T'] nucleotide for the cytosine ['C'] nucleotide at position 753 of SEQ ID NO:6;
- e) substitution of the nucleotide sequence 'AA' for the nucleotide sequence 'TT' at positions 41 and 42 of SEQ ID NO:6;
- f) insertion of the nucleotide sequence 'TCG' between positions 109 and 110 of SEQ ID NO:6;

- g) insertion of the nucleotide sequence 'AAAT' between position 131 and position 132 of SEQ ID NO:6;
 - h) substitution of the nucleotide sequence 'AAA' for the nucleotide sequence 'TT' at positions 76 to 77 of SEQ ID NO:6;
 - i) substitution of the nucleotide sequence 'GTTT' for the nucleotide sequence 'AAAA' at positions 631 to 634 of SEQ ID NO:6;
 - j) insertion of the nucleotide sequence 'TA' between position 512 and position 513 of SEQ ID NO:6;
 - k) insertion of the nucleotide sequence 'TTTA' between position 365 and position 366 of SEQ ID NO:6;
 - l) insertion of the nucleotide sequence 'TTAAA' between position 230 and position 231 of SEQ ID NO:6; and
 - m) any combination of part a), part b), part c), part d), part e), part f), part g), part h), part i), part j), part k), and part l).
13. The isolated nucleic acid molecule of Claim 12, wherein the promoter region comprises a sequence selected from the group consisting of SEQ ID NO:7, SEQ ID NO:11, SEQ ID NO:27, SEQ ID NO:15, SEQ ID NO:47, SEQ ID NO:43, SEQ ID NO:39, SEQ ID NO:35, SEQ ID NO:31, SEQ ID NO:19, SEQ ID NO:49, SEQ ID NO:51, SEQ ID NO:53 and SEQ ID NO:55.
14. The isolated nucleic acid molecule of Claim 12, wherein the promoter region of the *ALK2 Yarrowia* gene further (i) comprises an enhancer region set forth in SEQ ID NO:57, or (ii) is operably linked to said enhancer region.
15. An isolated nucleic acid molecule comprising a promoter region of an *ALK2 Yarrowia* gene comprising SEQ ID NO:56.

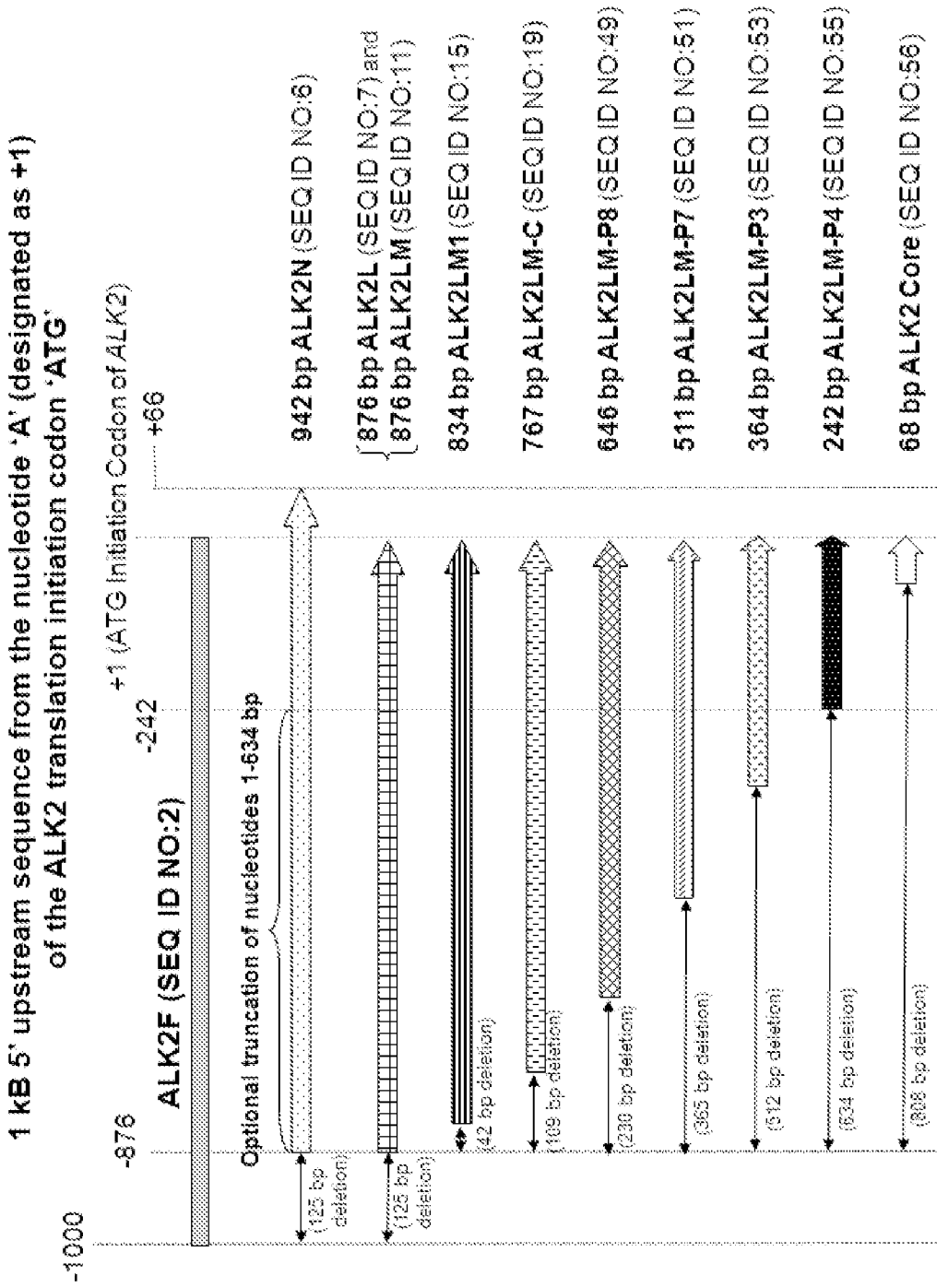


FIG. 1

2/14

FIG. 2A

			1	GTATAGACGTCCGGGATGCGATTGAGTTGGCTTAGCCCCAGCTTTCGAAATCAAGCTTTC	60
ALK2F (SEQ ID NO:2)	(1)				
942 bp ALK2N (SEQ ID NO:6)	(1)				
876 bp ALK2L (SEQ ID NO:7)	(1)				
876 bp ALK2LM (SEQ ID NO:11)	(1)				
838 bp ALK2LM1-S (SEQ ID NO:27)	(1)				
834 bp ALK2LM1 (SEQ ID NO:15)	(1)				
840 bp ALK2LM1-PP6 (SEQ ID NO:47)	(1)				
839 bp ALK2LM1-PP5 (SEQ ID NO:43)	(1)				
837 bp ALK2LM1-PP2 (SEQ ID NO:39)	(1)				
835 bp ALK2LM1-PP1 (SEQ ID NO:35)	(1)				
835 bp ALK2LM1-P (SEQ ID NO:31)	(1)				
767 bp ALK2LM-C (SEQ ID NO:19)	(1)				
646 bp ALK2LM-P8 (SEQ ID NO:49)	(1)				
511 bp ALK2LM-P7 (SEQ ID NO:51)	(1)				
364 bp ALK2LM-P3 (SEQ ID NO:53)	(1)				
242 bp ALK2LM-P4 (SEQ ID NO:55)	(1)				
68 bp ALK2 Core (SEQ ID NO:56)	(1)				
			61	CATTCTGAGCCTGAGGAATGGGACGGGGGGTGAATAATGGGGTGGGAAGTCCCTTGA	120
ALK2F (SEQ ID NO:2)	(61)				
942 bp ALK2N (SEQ ID NO:6)	(1)				
876 bp ALK2L (SEQ ID NO:7)	(1)				
876 bp ALK2LM (SEQ ID NO:11)	(1)				
838 bp ALK2LM1-S (SEQ ID NO:27)	(1)				
834 bp ALK2LM1 (SEQ ID NO:15)	(1)				
840 bp ALK2LM1-PP6 (SEQ ID NO:47)	(1)				
839 bp ALK2LM1-PP5 (SEQ ID NO:43)	(1)				
837 bp ALK2LM1-PP2 (SEQ ID NO:39)	(1)				
835 bp ALK2LM1-PP1 (SEQ ID NO:35)	(1)				
835 bp ALK2LM1-P (SEQ ID NO:31)	(1)				
767 bp ALK2LM-C (SEQ ID NO:19)	(1)				
646 bp ALK2LM-P8 (SEQ ID NO:49)	(1)				
511 bp ALK2LM-P7 (SEQ ID NO:51)	(1)				
364 bp ALK2LM-P3 (SEQ ID NO:53)	(1)				
242 bp ALK2LM-P4 (SEQ ID NO:55)	(1)				
68 bp ALK2 Core (SEQ ID NO:56)	(1)				

3/14

FIG. 2B

					120
					130
					140
					150
					160
					170
					180
					190
					200
					210
					220
					230
					240
					250
					260
					270
					280
					290
					300
					310
					320
					330
					340
					350
					360
					370
					380
					390
					400
					410
					420
					430
					440
					450
					460
					470
					480
					490
					500
					510
					520
					530
					540
					550
					560
					570
					580
					590
					600
					610
					620
					630
					640
					650
					660
					670
					680
					690
					700
					710
					720
					730
					740
					750
					760
					770
					780
					790
					800
					810
					820
					830
					840
					850
					860
					870
					880
					890
					900
					910
					920
					930
					940
					950
					960
					970
					980
					990
					1000

FIG. 2C

ALK2F	(SEQ ID NO:2)	241	TCTCGTCTACTGATTT	300	---CACATGGAACCTTTGCTATTTCGGGGATAACCCCTTTGC
942 bp ALK2N	(SEQ ID NO:6)	(240)	---TCTCGTCTACTGATTT	---	---CACATGGAACCTTTGCTATTTCGGGGATAACCCCTTTGC
876 bp ALK2L	(SEQ ID NO:7)	(116)	---TCTCGTCTACTGATTT	---	---CACATGGAACCTTTGCTATTTCGGGGATAACCCCTTTGC
876 bp ALK2LM	(SEQ ID NO:11)	(116)	---TCTCGTCTACTGATTT	---	---CACATGGAACCTTTGCTATTTCGGGGATAACCCCTTTGC
838 bp ALK2LM1-S	(SEQ ID NO:27)	(74)	---TCTCGTCTACTGATTT	---	---CACATGGAACCTTTGCTATTTCGGGGATAACCCCTTTGC
834 bp ALK2LM1	(SEQ ID NO:15)	(74)	---TCTCGTCTACTGATTT	---	---CACATGGAACCTTTGCTATTTCGGGGATAACCCCTTTGC
840 bp ALK2LM1-PP6	(SEQ ID NO:47)	(75)	---TCTCGTCTACTGATTT	---	---CACATGGAACCTTTGCTATTTCGGGGATAACCCCTTTGC
839 bp ALK2LM1-PP5	(SEQ ID NO:43)	(75)	---TCTCGTCTACTGATTT	---	---CACATGGAACCTTTGCTATTTCGGGGATAACCCCTTTGC
837 bp ALK2LM1-PP2	(SEQ ID NO:39)	(75)	---TCTCGTCTACTGATTT	---	---CACATGGAACCTTTGCTATTTCGGGGATAACCCCTTTGC
835 bp ALK2LM1-PP1	(SEQ ID NO:35)	(75)	---TCTCGTCTACTGATTT	---	---CACATGGAACCTTTGCTATTTCGGGGATAACCCCTTTGC
835 bp ALK2LM1-P	(SEQ ID NO:31)	(75)	---TCTCGTCTACTGATTT	---	---CACATGGAACCTTTGCTATTTCGGGGATAACCCCTTTGC
767 bp ALK2LM-C	(SEQ ID NO:19)	(7)	---TCTCGTCTACTGATTT	---	---CACATGGAACCTTTGCTATTTCGGGGATAACCCCTTTGC
646 bp ALK2LM-P8	(SEQ ID NO:49)	(1)	---	---	---
511 bp ALK2LM-P7	(SEQ ID NO:51)	(1)	---	---	---
364 bp ALK2LM-P3	(SEQ ID NO:53)	(1)	---	---	---
242 bp ALK2LM-P4	(SEQ ID NO:55)	(1)	---	---	---
68 bp ALK2 Core	(SEQ ID NO:56)	(1)	---	---	---
ALK2F	(SEQ ID NO:2)	301	CATTGCACGATGACGTGGC	360	---CATTGCACGATGACGTGGC
942 bp ALK2N	(SEQ ID NO:6)	(296)	---CATTGCACGATGACGTGGC	---	---CATTGCACGATGACGTGGC
876 bp ALK2L	(SEQ ID NO:7)	(172)	---CATTGCACGATGACGTGGC	---	---CATTGCACGATGACGTGGC
876 bp ALK2LM	(SEQ ID NO:11)	(172)	---CATTGCACGATGACGTGGC	---	---CATTGCACGATGACGTGGC
838 bp ALK2LM1-S	(SEQ ID NO:27)	(134)	---CATTGCACGATGACGTGGC	---	---CATTGCACGATGACGTGGC
834 bp ALK2LM1	(SEQ ID NO:15)	(130)	---CATTGCACGATGACGTGGC	---	---CATTGCACGATGACGTGGC
840 bp ALK2LM1-PP6	(SEQ ID NO:47)	(131)	---CATTGCACGATGACGTGGC	---	---CATTGCACGATGACGTGGC
839 bp ALK2LM1-PP5	(SEQ ID NO:43)	(131)	---CATTGCACGATGACGTGGC	---	---CATTGCACGATGACGTGGC
837 bp ALK2LM1-PP2	(SEQ ID NO:39)	(131)	---CATTGCACGATGACGTGGC	---	---CATTGCACGATGACGTGGC
835 bp ALK2LM1-PP1	(SEQ ID NO:35)	(131)	---CATTGCACGATGACGTGGC	---	---CATTGCACGATGACGTGGC
835 bp ALK2LM1-P	(SEQ ID NO:31)	(131)	---CATTGCACGATGACGTGGC	---	---CATTGCACGATGACGTGGC
767 bp ALK2LM-C	(SEQ ID NO:19)	(63)	---CATTGCACGATGACGTGGC	---	---CATTGCACGATGACGTGGC
646 bp ALK2LM-P8	(SEQ ID NO:49)	(1)	---	---	---
511 bp ALK2LM-P7	(SEQ ID NO:51)	(1)	---	---	---
364 bp ALK2LM-P3	(SEQ ID NO:53)	(1)	---	---	---
242 bp ALK2LM-P4	(SEQ ID NO:55)	(1)	---	---	---
68 bp ALK2 Core	(SEQ ID NO:56)	(1)	---	---	---

6/14

FIG. 2E

ALK2F (SEQ ID NO: 2)	(471)	GC	GGGGTAGCAGGATACTG	540	ACTACAGTACGATTTGCTTGTCTACTGCTTGTAGCAA
942 bp ALK2N (SEQ ID NO: 6)	(347)	GC	GGGGTAGCAGGATACTG	540	ACTACAGTACGATTTGCTTGTCTACTGCTTGTAGCAA
876 bp ALK2L (SEQ ID NO: 7)	(347)	GC	GGGGTAGCAGGATACTG	540	ACTACAGTACGATTTGCTTGTCTACTGCTTGTAGCAA
876 bp ALK2LM (SEQ ID NO: 11)	(347)	GC	GGGGTAGCAGGATACTG	540	ACTACAGTACGATTTGCTTGTCTACTGCTTGTAGCAA
838 bp ALK2LM1-S (SEQ ID NO: 27)	(309)	GC	GGGGTAGCAGGATACTG	540	ACTACAGTACGATTTGCTTGTCTACTGCTTGTAGCAA
834 bp ALK2LM1 (SEQ ID NO: 15)	(305)	GC	GGGGTAGCAGGATACTG	540	ACTACAGTACGATTTGCTTGTCTACTGCTTGTAGCAA
840 bp ALK2LM1-PP6 (SEQ ID NO: 47)	(311)	GC	GGGGTAGCAGGATACTG	540	ACTACAGTACGATTTGCTTGTCTACTGCTTGTAGCAA
839 bp ALK2LM1-PP5 (SEQ ID NO: 43)	(306)	GC	GGGGTAGCAGGATACTG	540	ACTACAGTACGATTTGCTTGTCTACTGCTTGTAGCAA
837 bp ALK2LM1-PP2 (SEQ ID NO: 39)	(306)	GC	GGGGTAGCAGGATACTG	540	ACTACAGTACGATTTGCTTGTCTACTGCTTGTAGCAA
835 bp ALK2LM1-PP1 (SEQ ID NO: 35)	(306)	GC	GGGGTAGCAGGATACTG	540	ACTACAGTACGATTTGCTTGTCTACTGCTTGTAGCAA
835 bp ALK2LM1-P (SEQ ID NO: 31)	(306)	GC	GGGGTAGCAGGATACTG	540	ACTACAGTACGATTTGCTTGTCTACTGCTTGTAGCAA
767 bp ALK2LM1-C (SEQ ID NO: 19)	(238)	GC	GGGGTAGCAGGATACTG	540	ACTACAGTACGATTTGCTTGTCTACTGCTTGTAGCAA
646 bp ALK2LM-P8 (SEQ ID NO: 49)	(117)	GC	GGGGTAGCAGGATACTG	540	ACTACAGTACGATTTGCTTGTCTACTGCTTGTAGCAA
511 bp ALK2LM-P7 (SEQ ID NO: 51)	(1)	---	---	541	---
364 bp ALK2LM-P3 (SEQ ID NO: 53)	(1)	---	---	541	---
242 bp ALK2LM-P4 (SEQ ID NO: 55)	(1)	---	---	541	---
68 bp ALK2 Core (SEQ ID NO: 56)	(1)	---	---	541	---
ALK2F (SEQ ID NO: 2)	(527)	TT	ACCTTTACTGTAGGGACCAACACCTCTGGTTTCAATGTCCTTCTCGCCTCGACAAAG	600	TTACCTTTACTGTAGGGACCAACACCTCTGGTTTCAATGTCCTTCTCGCCTCGACAAAG
942 bp ALK2N (SEQ ID NO: 6)	(403)	TT	ACCTTTACTGTAGGGACCAACACCTCTGGTTTCAATGTCCTTCTCGCCTCGACAAAG	600	TTACCTTTACTGTAGGGACCAACACCTCTGGTTTCAATGTCCTTCTCGCCTCGACAAAG
876 bp ALK2L (SEQ ID NO: 7)	(403)	TT	ACCTTTACTGTAGGGACCAACACCTCTGGTTTCAATGTCCTTCTCGCCTCGACAAAG	600	TTACCTTTACTGTAGGGACCAACACCTCTGGTTTCAATGTCCTTCTCGCCTCGACAAAG
876 bp ALK2LM (SEQ ID NO: 11)	(403)	TT	ACCTTTACTGTAGGGACCAACACCTCTGGTTTCAATGTCCTTCTCGCCTCGACAAAG	600	TTACCTTTACTGTAGGGACCAACACCTCTGGTTTCAATGTCCTTCTCGCCTCGACAAAG
838 bp ALK2LM1-S (SEQ ID NO: 27)	(365)	TT	ACCTTTACTGTAGGGACCAACACCTCTGGTTTCAATGTCCTTCTCGCCTCGACAAAG	600	TTACCTTTACTGTAGGGACCAACACCTCTGGTTTCAATGTCCTTCTCGCCTCGACAAAG
834 bp ALK2LM1 (SEQ ID NO: 15)	(361)	TT	ACCTTTACTGTAGGGACCAACACCTCTGGTTTCAATGTCCTTCTCGCCTCGACAAAG	600	TTACCTTTACTGTAGGGACCAACACCTCTGGTTTCAATGTCCTTCTCGCCTCGACAAAG
840 bp ALK2LM1-PP6 (SEQ ID NO: 47)	(367)	TT	ACCTTTACTGTAGGGACCAACACCTCTGGTTTCAATGTCCTTCTCGCCTCGACAAAG	600	TTACCTTTACTGTAGGGACCAACACCTCTGGTTTCAATGTCCTTCTCGCCTCGACAAAG
839 bp ALK2LM1-PP5 (SEQ ID NO: 43)	(366)	TT	ACCTTTACTGTAGGGACCAACACCTCTGGTTTCAATGTCCTTCTCGCCTCGACAAAG	600	TTACCTTTACTGTAGGGACCAACACCTCTGGTTTCAATGTCCTTCTCGCCTCGACAAAG
837 bp ALK2LM1-PP2 (SEQ ID NO: 39)	(362)	TT	ACCTTTACTGTAGGGACCAACACCTCTGGTTTCAATGTCCTTCTCGCCTCGACAAAG	600	TTACCTTTACTGTAGGGACCAACACCTCTGGTTTCAATGTCCTTCTCGCCTCGACAAAG
835 bp ALK2LM1-PP1 (SEQ ID NO: 35)	(362)	TT	ACCTTTACTGTAGGGACCAACACCTCTGGTTTCAATGTCCTTCTCGCCTCGACAAAG	600	TTACCTTTACTGTAGGGACCAACACCTCTGGTTTCAATGTCCTTCTCGCCTCGACAAAG
835 bp ALK2LM1-P (SEQ ID NO: 31)	(362)	TT	ACCTTTACTGTAGGGACCAACACCTCTGGTTTCAATGTCCTTCTCGCCTCGACAAAG	600	TTACCTTTACTGTAGGGACCAACACCTCTGGTTTCAATGTCCTTCTCGCCTCGACAAAG
767 bp ALK2LM1-C (SEQ ID NO: 19)	(294)	TT	ACCTTTACTGTAGGGACCAACACCTCTGGTTTCAATGTCCTTCTCGCCTCGACAAAG	600	TTACCTTTACTGTAGGGACCAACACCTCTGGTTTCAATGTCCTTCTCGCCTCGACAAAG
646 bp ALK2LM-P8 (SEQ ID NO: 49)	(173)	TT	ACCTTTACTGTAGGGACCAACACCTCTGGTTTCAATGTCCTTCTCGCCTCGACAAAG	600	TTACCTTTACTGTAGGGACCAACACCTCTGGTTTCAATGTCCTTCTCGCCTCGACAAAG
511 bp ALK2LM-P7 (SEQ ID NO: 51)	(38)	TT	ACCTTTACTGTAGGGACCAACACCTCTGGTTTCAATGTCCTTCTCGCCTCGACAAAG	600	TTACCTTTACTGTAGGGACCAACACCTCTGGTTTCAATGTCCTTCTCGCCTCGACAAAG
364 bp ALK2LM-P3 (SEQ ID NO: 53)	(1)	---	---	601	---
242 bp ALK2LM-P4 (SEQ ID NO: 55)	(1)	---	---	601	---
68 bp ALK2 Core (SEQ ID NO: 56)	(1)	---	---	601	---

7/14

FIG. 2F

ALK2F	(SEQ ID NO:2)	(587)	CAAACTGTCACCCCAATCACACCTTGTTCAATATCATATTAGTGCATCCGTT	660
942 bp ALK2N	(SEQ ID NO:6)	(463)	CAAACTGTCACCCCAATCACACCTTGTTCAATATCATATTAGTGCATCCGTT	ACCTTGA
876 bp ALK2L	(SEQ ID NO:7)	(463)	CAAACTGTCACCCCAATCACACCTTGTTCAATATCATATTAGTGCATCCGTT	ACCTTGA
876 bp ALK2LM	(SEQ ID NO:11)	(463)	CAAACTGTCACCCCAATCACACCTTGTTCAATATCATATTAGTGCATCCGTT	ACCTTGA
838 bp ALK2LM1-S	(SEQ ID NO:27)	(425)	CAAACTGTCACCCCAATCACACCTTGTTCAATATCATATTAGTGCATCCGTT	ACCTTGA
834 bp ALK2LM1	(SEQ ID NO:15)	(421)	CAAACTGTCACCCCAATCACACCTTGTTCAATATCATATTAGTGCATCCGTT	ACCTTGA
840 bp ALK2LM1-PP6	(SEQ ID NO:47)	(427)	CAAACTGTCACCCCAATCACACCTTGTTCAATATCATATTAGTGCATCCGTT	ACCTTGA
839 bp ALK2LM1-PP5	(SEQ ID NO:43)	(426)	CAAACTGTCACCCCAATCACACCTTGTTCAATATCATATTAGTGCATCCGTT	ACCTTGA
837 bp ALK2LM1-PP2	(SEQ ID NO:39)	(422)	CAAACTGTCACCCCAATCACACCTTGTTCAATATCATATTAGTGCATCCGTT	ACCTTGA
835 bp ALK2LM1-PP1	(SEQ ID NO:35)	(422)	CAAACTGTCACCCCAATCACACCTTGTTCAATATCATATTAGTGCATCCGTT	ACCTTGA
835 bp ALK2LM1-P	(SEQ ID NO:31)	(422)	CAAACTGTCACCCCAATCACACCTTGTTCAATATCATATTAGTGCATCCGTT	ACCTTGA
767 bp ALK2LM-C	(SEQ ID NO:19)	(354)	CAAACTGTCACCCCAATCACACCTTGTTCAATATCATATTAGTGCATCCGTT	ACCTTGA
646 bp ALK2LM-P8	(SEQ ID NO:49)	(233)	CAAACTGTCACCCCAATCACACCTTGTTCAATATCATATTAGTGCATCCGTT	ACCTTGA
511 bp ALK2LM-P7	(SEQ ID NO:51)	(98)	CAAACTGTCACCCCAATCACACCTTGTTCAATATCATATTAGTGCATCCGTT	ACCTTGA
364 bp ALK2LM-P3	(SEQ ID NO:53)	(1)	CAAACTGTCACCCCAATCACACCTTGTTCAATATCATATTAGTGCATCCGTT	ACCTTGA
242 bp ALK2LM-P4	(SEQ ID NO:55)	(1)	CAAACTGTCACCCCAATCACACCTTGTTCAATATCATATTAGTGCATCCGTT	ACCTTGA
68 bp ALK2 Core	(SEQ ID NO:56)	(1)	CAAACTGTCACCCCAATCACACCTTGTTCAATATCATATTAGTGCATCCGTT	ACCTTGA
ALK2F	(SEQ ID NO:2)	(645)	CATGACACTTCTCATACTAGTATAGGGCTGTAGTTGAGACAAGTTGATTCACACGGGATA	720
942 bp ALK2N	(SEQ ID NO:6)	(521)	CATGACACTTCTCATACTAGTATAGGGCTGTAGTTGAGACAAGTTGATTCACACGGGATA	
876 bp ALK2L	(SEQ ID NO:7)	(521)	CATGACACTTCTCATACTAGTATAGGGCTGTAGTTGAGACAAGTTGATTCACACGGGATA	
876 bp ALK2LM	(SEQ ID NO:11)	(521)	CATGACACTTCTCATACTAGTATAGGGCTGTAGTTGAGACAAGTTGATTCACACGGGATA	
838 bp ALK2LM1-S	(SEQ ID NO:27)	(483)	CATGACACTTCTCATACTAGTATAGGGCTGTAGTTGAGACAAGTTGATTCACACGGGATA	
834 bp ALK2LM1	(SEQ ID NO:15)	(479)	CATGACACTTCTCATACTAGTATAGGGCTGTAGTTGAGACAAGTTGATTCACACGGGATA	
840 bp ALK2LM1-PP6	(SEQ ID NO:47)	(485)	CATGACACTTCTCATACTAGTATAGGGCTGTAGTTGAGACAAGTTGATTCACACGGGATA	
839 bp ALK2LM1-PP5	(SEQ ID NO:43)	(484)	CATGACACTTCTCATACTAGTATAGGGCTGTAGTTGAGACAAGTTGATTCACACGGGATA	
837 bp ALK2LM1-PP2	(SEQ ID NO:39)	(482)	CATGACACTTCTCATACTAGTATAGGGCTGTAGTTGAGACAAGTTGATTCACACGGGATA	
835 bp ALK2LM1-PP1	(SEQ ID NO:35)	(480)	CATGACACTTCTCATACTAGTATAGGGCTGTAGTTGAGACAAGTTGATTCACACGGGATA	
835 bp ALK2LM1-P	(SEQ ID NO:31)	(480)	CATGACACTTCTCATACTAGTATAGGGCTGTAGTTGAGACAAGTTGATTCACACGGGATA	
767 bp ALK2LM-C	(SEQ ID NO:19)	(412)	CATGACACTTCTCATACTAGTATAGGGCTGTAGTTGAGACAAGTTGATTCACACGGGATA	
646 bp ALK2LM-P8	(SEQ ID NO:49)	(291)	CATGACACTTCTCATACTAGTATAGGGCTGTAGTTGAGACAAGTTGATTCACACGGGATA	
511 bp ALK2LM-P7	(SEQ ID NO:51)	(156)	CATGACACTTCTCATACTAGTATAGGGCTGTAGTTGAGACAAGTTGATTCACACGGGATA	
364 bp ALK2LM-P3	(SEQ ID NO:53)	(9)	CATGACACTTCTCATACTAGTATAGGGCTGTAGTTGAGACAAGTTGATTCACACGGGATA	
242 bp ALK2LM-P4	(SEQ ID NO:55)	(1)	CATGACACTTCTCATACTAGTATAGGGCTGTAGTTGAGACAAGTTGATTCACACGGGATA	
68 bp ALK2 Core	(SEQ ID NO:56)	(1)	CATGACACTTCTCATACTAGTATAGGGCTGTAGTTGAGACAAGTTGATTCACACGGGATA	

8/14

FIG. 2G

ALK2F (SEQ ID NO:2)	(705)	CATACAAAGCCTCAGAGAGAGCAAAATGTTATATATCTCAGGACCGGACCAATCAAAATAACAC	721	780
942 bp ALK2N (SEQ ID NO:6)	(581)	CATACAAAGCCTCAGAGAGAGCAAAATGTTATATATCTCAGGACCGGACCAATCAAAATAACAC		
876 bp ALK2L (SEQ ID NO:7)	(581)	CATACAAAGCCTCAGAGAGAGCAAAATGTTATATATCTCAGGACCGGACCAATCAAAATAACAC		
876 bp ALK2LM (SEQ ID NO:11)	(581)	CATACAAAGCCTCAGAGAGAGCAAAATGTTATATATCTCAGGACCGGACCAATCAAAATAACAC		
838 bp ALK2LM1-S (SEQ ID NO:27)	(543)	CATACAAAGCCTCAGAGAGAGCAAAATGTTATATATCTCAGGACCGGACCAATCAAAATAACAC		
834 bp ALK2LM1 (SEQ ID NO:15)	(539)	CATACAAAGCCTCAGAGAGAGCAAAATGTTATATATCTCAGGACCGGACCAATCAAAATAACAC		
840 bp ALK2LM1-PP6 (SEQ ID NO:47)	(545)	CATACAAAGCCTCAGAGAGAGCAAAATGTTATATATCTCAGGACCGGACCAATCAAAATAACAC		
839 bp ALK2LM1-PP5 (SEQ ID NO:43)	(544)	CATACAAAGCCTCAGAGAGAGCAAAATGTTATATATCTCAGGACCGGACCAATCAAAATAACAC		
837 bp ALK2LM1-PP2 (SEQ ID NO:39)	(542)	CATACAAAGCCTCAGAGAGAGCAAAATGTTATATATCTCAGGACCGGACCAATCAAAATAACAC		
835 bp ALK2LM1-PP1 (SEQ ID NO:35)	(540)	CATACAAAGCCTCAGAGAGAGCAAAATGTTATATATCTCAGGACCGGACCAATCAAAATAACAC		
835 bp ALK2LM1-P (SEQ ID NO:31)	(540)	CATACAAAGCCTCAGAGAGAGCAAAATGTTATATATCTCAGGACCGGACCAATCAAAATAACAC		
767 bp ALK2LM-C (SEQ ID NO:19)	(472)	CATACAAAGCCTCAGAGAGAGCAAAATGTTATATATCTCAGGACCGGACCAATCAAAATAACAC		
646 bp ALK2LM-P8 (SEQ ID NO:49)	(351)	CATACAAAGCCTCAGAGAGAGCAAAATGTTATATATCTCAGGACCGGACCAATCAAAATAACAC		
511 bp ALK2LM-P7 (SEQ ID NO:51)	(216)	CATACAAAGCCTCAGAGAGAGCAAAATGTTATATATCTCAGGACCGGACCAATCAAAATAACAC		
364 bp ALK2LM-P3 (SEQ ID NO:53)	(69)	CATACAAAGCCTCAGAGAGAGCAAAATGTTATATATCTCAGGACCGGACCAATCAAAATAACAC		
242 bp ALK2LM-P4 (SEQ ID NO:55)	(1)	CATACAAAGCCTCAGAGAGAGCAAAATGTTATATATCTCAGGACCGGACCAATCAAAATAACAC		
68 bp ALK2 Core (SEQ ID NO:56)	(1)	CATACAAAGCCTCAGAGAGAGCAAAATGTTATATATCTCAGGACCGGACCAATCAAAATAACAC		
ALK2F (SEQ ID NO:2)	(765)	ACTCCTAATAAACCACCAATTTCCATCTCAGGCGTACCTACTCTGTGAGCTGCCCCACATTGC	781	840
942 bp ALK2N (SEQ ID NO:6)	(641)	ACTCCTAATAAACCACCAATTTCCATCTCAGGCGTACCTACTCTGTGAGCTGCCCCACATTGC		
876 bp ALK2L (SEQ ID NO:7)	(641)	ACTCCTAATAAACCACCAATTTCCATCTCAGGCGTACCTACTCTGTGAGCTGCCCCACATTGC		
876 bp ALK2LM (SEQ ID NO:11)	(641)	ACTCCTAATAAACCACCAATTTCCATCTCAGGCGTACCTACTCTGTGAGCTGCCCCACATTGC		
838 bp ALK2LM1-S (SEQ ID NO:27)	(603)	ACTCCTAATAAACCACCAATTTCCATCTCAGGCGTACCTACTCTGTGAGCTGCCCCACATTGC		
834 bp ALK2LM1 (SEQ ID NO:15)	(599)	ACTCCTAATAAACCACCAATTTCCATCTCAGGCGTACCTACTCTGTGAGCTGCCCCACATTGC		
840 bp ALK2LM1-PP6 (SEQ ID NO:47)	(605)	ACTCCTAATAAACCACCAATTTCCATCTCAGGCGTACCTACTCTGTGAGCTGCCCCACATTGC		
839 bp ALK2LM1-PP5 (SEQ ID NO:43)	(604)	ACTCCTAATAAACCACCAATTTCCATCTCAGGCGTACCTACTCTGTGAGCTGCCCCACATTGC		
837 bp ALK2LM1-PP2 (SEQ ID NO:39)	(602)	ACTCCTAATAAACCACCAATTTCCATCTCAGGCGTACCTACTCTGTGAGCTGCCCCACATTGC		
835 bp ALK2LM1-PP1 (SEQ ID NO:35)	(600)	ACTCCTAATAAACCACCAATTTCCATCTCAGGCGTACCTACTCTGTGAGCTGCCCCACATTGC		
835 bp ALK2LM1-P (SEQ ID NO:31)	(600)	ACTCCTAATAAACCACCAATTTCCATCTCAGGCGTACCTACTCTGTGAGCTGCCCCACATTGC		
767 bp ALK2LM-C (SEQ ID NO:19)	(532)	ACTCCTAATAAACCACCAATTTCCATCTCAGGCGTACCTACTCTGTGAGCTGCCCCACATTGC		
646 bp ALK2LM-P8 (SEQ ID NO:49)	(411)	ACTCCTAATAAACCACCAATTTCCATCTCAGGCGTACCTACTCTGTGAGCTGCCCCACATTGC		
511 bp ALK2LM-P7 (SEQ ID NO:51)	(276)	ACTCCTAATAAACCACCAATTTCCATCTCAGGCGTACCTACTCTGTGAGCTGCCCCACATTGC		
364 bp ALK2LM-P3 (SEQ ID NO:53)	(129)	ACTCCTAATAAACCACCAATTTCCATCTCAGGCGTACCTACTCTGTGAGCTGCCCCACATTGC		
242 bp ALK2LM-P4 (SEQ ID NO:55)	(7)	ACTCCTAATAAACCACCAATTTCCATCTCAGGCGTACCTACTCTGTGAGCTGCCCCACATTGC		
68 bp ALK2 Core (SEQ ID NO:56)	(1)	ACTCCTAATAAACCACCAATTTCCATCTCAGGCGTACCTACTCTGTGAGCTGCCCCACATTGC		

9/14

FIG. 2H

900																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																															
-----	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

10/14

FIG. 21

			961				1020
	ALK2F	(SEQ ID NO: 2)	(1945)	GGGAGAGCCCTAAGTCTTTTCGAAACCCCGGACCTTCACCTTCCACAACC			
942 bp	ALK2N	(SEQ ID NO: 6)	(821)	GGGAGAGCCCTAAGTCTTTTCGAAACCCCGGACCTTCACCTTCCACAACC			
876 bp	ALK2L	(SEQ ID NO: 7)	(821)	GGGAGAGCCCTAAGTCTTTTCGAAACCCCGGACCTTCACCTTCCACAACC			
876 bp	ALK2LM	(SEQ ID NO: 11)	(821)	GGGAGAGCCCTAAGTCTTTTCGAAACCCCGGACCTTCACCTTCCACAACC			
838 bp	ALK2LM1-S	(SEQ ID NO: 27)	(783)	GGGAGAGCCCTAAGTCTTTTCGAAACCCCGGACCTTCACCTTCCACAACC			
834 bp	ALK2LM1	(SEQ ID NO: 15)	(779)	GGGAGAGCCCTAAGTCTTTTCGAAACCCCGGACCTTCACCTTCCACAACC			
840 bp	ALK2LM1-PP6	(SEQ ID NO: 47)	(785)	GGGAGAGCCCTAAGTCTTTTCGAAACCCCGGACCTTCACCTTCCACAACC			
839 bp	ALK2LM1-PP5	(SEQ ID NO: 43)	(784)	GGGAGAGCCCTAAGTCTTTTCGAAACCCCGGACCTTCACCTTCCACAACC			
837 bp	ALK2LM1-PP2	(SEQ ID NO: 39)	(782)	GGGAGAGCCCTAAGTCTTTTCGAAACCCCGGACCTTCACCTTCCACAACC			
835 bp	ALK2LM1-PP1	(SEQ ID NO: 35)	(780)	GGGAGAGCCCTAAGTCTTTTCGAAACCCCGGACCTTCACCTTCCACAACC			
835 bp	ALK2LM1-P	(SEQ ID NO: 31)	(780)	GGGAGAGCCCTAAGTCTTTTCGAAACCCCGGACCTTCACCTTCCACAACC			
767 bp	ALK2LM-C	(SEQ ID NO: 19)	(712)	GGGAGAGCCCTAAGTCTTTTCGAAACCCCGGACCTTCACCTTCCACAACC			
646 bp	ALK2LM-P8	(SEQ ID NO: 49)	(591)	GGGAGAGCCCTAAGTCTTTTCGAAACCCCGGACCTTCACCTTCCACAACC			
511 bp	ALK2LM-P7	(SEQ ID NO: 51)	(456)	GGGAGAGCCCTAAGTCTTTTCGAAACCCCGGACCTTCACCTTCCACAACC			
364 bp	ALK2LM-P3	(SEQ ID NO: 53)	(309)	GGGAGAGCCCTAAGTCTTTTCGAAACCCCGGACCTTCACCTTCCACAACC			
242 bp	ALK2LM-P4	(SEQ ID NO: 55)	(187)	GGGAGAGCCCTAAGTCTTTTCGAAACCCCGGACCTTCACCTTCCACAACC			
68 bp	ALK2 Core	(SEQ ID NO: 56)	(13)	GGGAGAGCCCTAAGTCTTTTCGAAACCCCGGACCTTCACCTTCCACAACC			
			1021				1080
	ALK2F	(SEQ ID NO: 2)	(1001)	TCATCTTATAGTCTTTGGGCGTTGGGCTCTCCCTTCATCTTCAGAGAGTCACTTACA			
942 bp	ALK2N	(SEQ ID NO: 6)	(881)	TCATCTTATAGTCTTTGGGCGTTGGGCTCTCCCTTCATCTTCAGAGAGTCACTTACA			
876 bp	ALK2L	(SEQ ID NO: 7)	(877)	TCATCTTATAGTCTTTGGGCGTTGGGCTCTCCCTTCATCTTCAGAGAGTCACTTACA			
876 bp	ALK2LM	(SEQ ID NO: 11)	(877)	TCATCTTATAGTCTTTGGGCGTTGGGCTCTCCCTTCATCTTCAGAGAGTCACTTACA			
838 bp	ALK2LM1-S	(SEQ ID NO: 27)	(839)	TCATCTTATAGTCTTTGGGCGTTGGGCTCTCCCTTCATCTTCAGAGAGTCACTTACA			
834 bp	ALK2LM1	(SEQ ID NO: 15)	(835)	TCATCTTATAGTCTTTGGGCGTTGGGCTCTCCCTTCATCTTCAGAGAGTCACTTACA			
840 bp	ALK2LM1-PP6	(SEQ ID NO: 47)	(841)	TCATCTTATAGTCTTTGGGCGTTGGGCTCTCCCTTCATCTTCAGAGAGTCACTTACA			
839 bp	ALK2LM1-PP5	(SEQ ID NO: 43)	(840)	TCATCTTATAGTCTTTGGGCGTTGGGCTCTCCCTTCATCTTCAGAGAGTCACTTACA			
837 bp	ALK2LM1-PP2	(SEQ ID NO: 39)	(838)	TCATCTTATAGTCTTTGGGCGTTGGGCTCTCCCTTCATCTTCAGAGAGTCACTTACA			
835 bp	ALK2LM1-PP1	(SEQ ID NO: 35)	(836)	TCATCTTATAGTCTTTGGGCGTTGGGCTCTCCCTTCATCTTCAGAGAGTCACTTACA			
835 bp	ALK2LM1-P	(SEQ ID NO: 31)	(836)	TCATCTTATAGTCTTTGGGCGTTGGGCTCTCCCTTCATCTTCAGAGAGTCACTTACA			
767 bp	ALK2LM-C	(SEQ ID NO: 19)	(769)	TCATCTTATAGTCTTTGGGCGTTGGGCTCTCCCTTCATCTTCAGAGAGTCACTTACA			
646 bp	ALK2LM-P8	(SEQ ID NO: 49)	(647)	TCATCTTATAGTCTTTGGGCGTTGGGCTCTCCCTTCATCTTCAGAGAGTCACTTACA			
511 bp	ALK2LM-P7	(SEQ ID NO: 51)	(512)	TCATCTTATAGTCTTTGGGCGTTGGGCTCTCCCTTCATCTTCAGAGAGTCACTTACA			
364 bp	ALK2LM-P3	(SEQ ID NO: 53)	(365)	TCATCTTATAGTCTTTGGGCGTTGGGCTCTCCCTTCATCTTCAGAGAGTCACTTACA			
242 bp	ALK2LM-P4	(SEQ ID NO: 55)	(243)	TCATCTTATAGTCTTTGGGCGTTGGGCTCTCCCTTCATCTTCAGAGAGTCACTTACA			
68 bp	ALK2 Core	(SEQ ID NO: 56)	(69)	TCATCTTATAGTCTTTGGGCGTTGGGCTCTCCCTTCATCTTCAGAGAGTCACTTACA			

11/14

FIG. 2J

	ALK2F (SEQ ID NO:2)	(1001)	--	1091
942 bp	ALK2N (SEQ ID NO:6)	(941)	CC	
876 bp	ALK2L (SEQ ID NO:7)	(877)	--	
876 bp	ALK2LM (SEQ ID NO:11)	(877)	--	
838 bp	ALK2LM1-S (SEQ ID NO:27)	(839)	--	
834 bp	ALK2LM1 (SEQ ID NO:15)	(835)	--	
840 bp	ALK2LM1-PP6 (SEQ ID NO:47)	(841)	--	
839 bp	ALK2LM1-PP5 (SEQ ID NO:43)	(840)	--	
837 bp	ALK2LM1-PP2 (SEQ ID NO:39)	(838)	--	
835 bp	ALK2LM1-PP1 (SEQ ID NO:35)	(836)	--	
835 bp	ALK2LM1-P (SEQ ID NO:31)	(836)	--	
767 bp	ALK2LM-C (SEQ ID NO:19)	(768)	--	
646 bp	ALK2LM-P8 (SEQ ID NO:49)	(647)	--	
511 bp	ALK2LM-P7 (SEQ ID NO:51)	(512)	--	
364 bp	ALK2LM-P3 (SEQ ID NO:53)	(365)	--	
242 bp	ALK2LM-P4 (SEQ ID NO:55)	(243)	--	
68 bp	ALK2 Core (SEQ ID NO:56)	(69)	--	

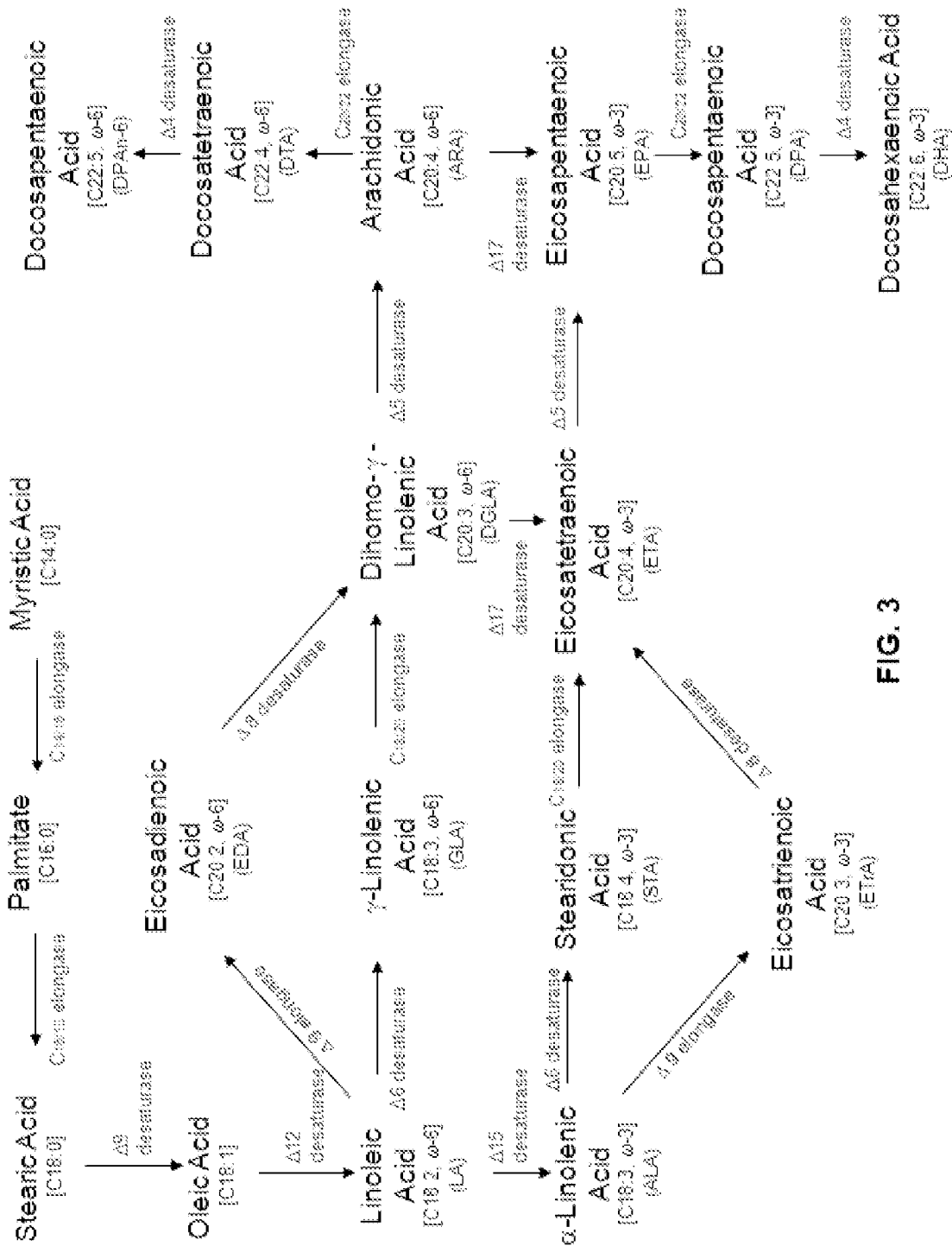


FIG. 3

13/14

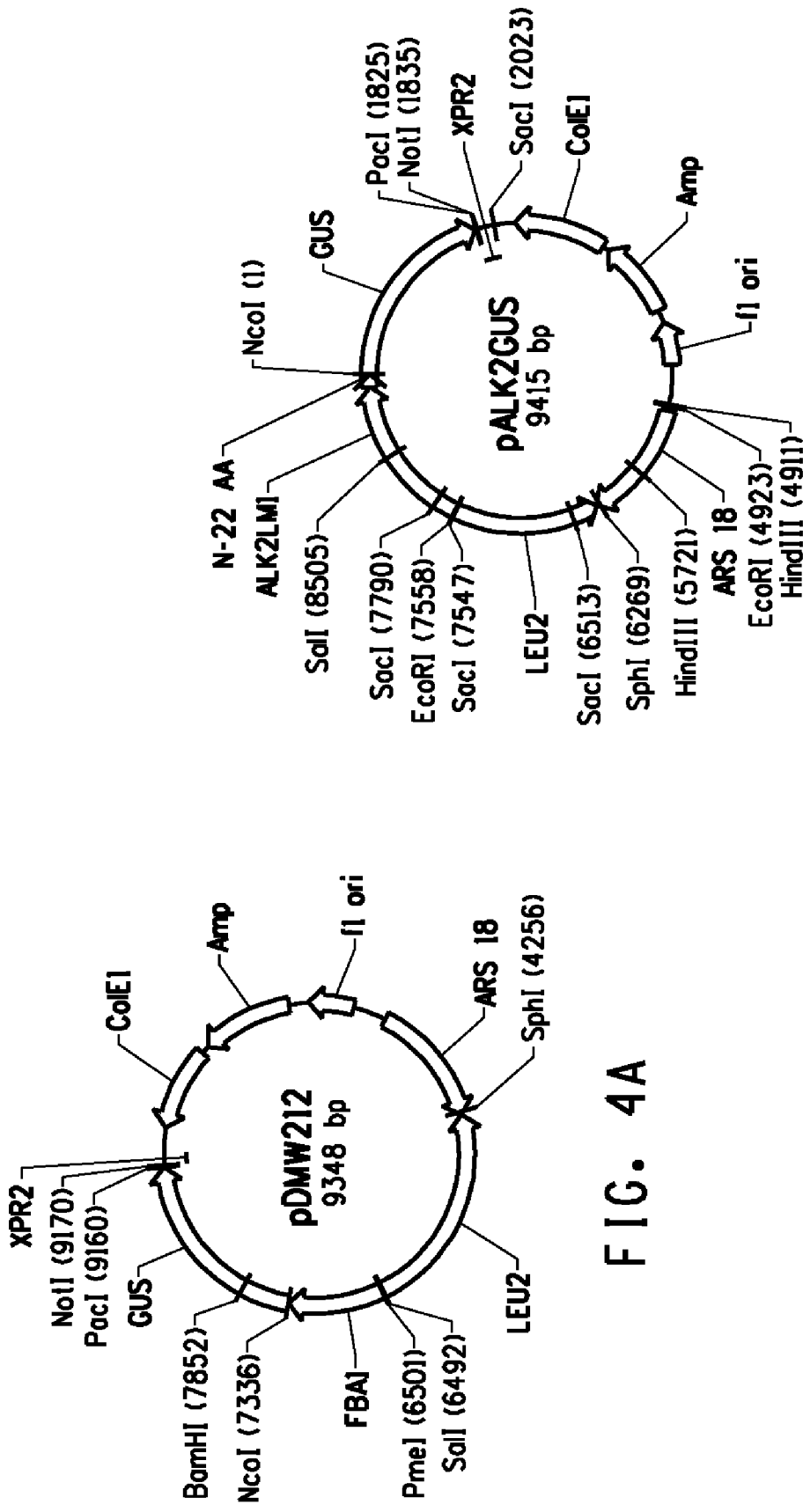


FIG. 4A

FIG. 4B

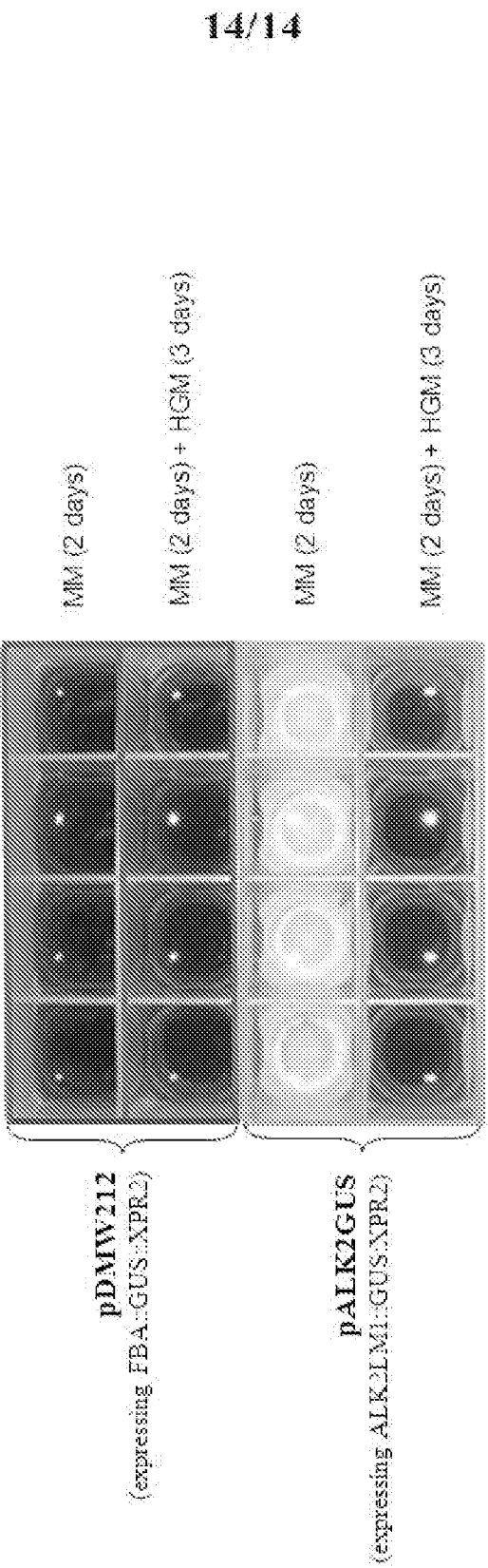


FIG. 5

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2012/031909

A. CLASSIFICATION OF SUBJECT MATTER
INV. C12N15/81 C12N9/00
ADD.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

EP0-Internal, BIOSIS, Sequence Search, EMBASE, FSTA, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 2004/016756 A2 (COGNIS CORP [US]) 26 February 2004 (2004-02-26)	1-15
Y	page 17; claims; examples; sequence 81 the whole document	1-15
X	----- DATABASE Geneseq [Online] 26 April 2012 (2012-04-26), "Yarrowia lipolytica ALK2LM1 promoter DNA, SEQ:93.", XP002677653, retrieved from EBI accession no. GSN:AZT93123 Database accession no. AZT93123	1-15
L	the whole document -/--	

☒ Further documents are listed in the continuation of Box C.

☒ See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

18 June 2012

Date of mailing of the international search report

05/07/2012

Name and mailing address of the ISA/

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040,
Fax: (+31-70) 340-3016

Authorized officer

Madruga, Jaime

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2012/031909

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X,P	-& WO 2012/027689 A1 (DU PONT [US]; HONG SEUNG-PYO [US]; SHARPE PAMELA L [US]; XUE ZHIXIONG) 1 March 2012 (2012-03-01) pages 55-56; claims; examples 6-8; table 5; sequences 93,95,100, 101 the whole document	1-15
X,P	----- DATABASE Geneseq [Online] 1 March 2012 (2012-03-01), "Plasmid pYPS233 related 1019 bp stuffer fragment, SEQ ID 230.", XP002677654, retrieved from EBI accession no. GSN:AZT79773 Database accession no. AZT79773 the whole document	1-15
X,P	-& WO 2012/027698 A1 (DU PONT [US]; BOSTICK MICHAEL W [US]; DAUBERT MELISSA D [US]; HE HONGX) 1 March 2012 (2012-03-01) claims; figures; examples; sequences 230-231 the whole document	1-15
Y	----- US 2005/014270 A1 (PICATAGGIO STEPHEN K [US] ET AL) 20 January 2005 (2005-01-20) cited in the application the whole document	1-15
A	----- US 6 673 613 B2 (CRAFT DAVID L [US] ET AL) 6 January 2004 (2004-01-06) the whole document	1-15
A	----- KOGURE T ET AL: "n-Alkane and clofibrate, a peroxisome proliferator, activate transcription of ALK2 gene encoding cytochrome P450alk2 through distinct cis-acting promoter elements in Candida maltosa", BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS, ACADEMIC PRESS INC. ORLANDO, FL, US, vol. 329, no. 1, 1 April 2005 (2005-04-01) , pages 78-86, XP004757000, ISSN: 0006-291X, DOI: 10.1016/J.BBRC.2005.01.099 the whole document	1-15
	----- -/--	

INTERNATIONAL SEARCH REPORT

International application No

PCT/US2012/031909

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	SEGHEZZI W ET AL: "Characterization of a second alkane-inducible cytochrome P450-encoding gene, CYP52A2, from <i>Candida tropicalis</i>", GENE, ELSEVIER, AMSTERDAM, NL, vol. 106, no. 1, 30 September 1991 (1991-09-30), pages 51-60, XP023541533, ISSN: 0378-1119, DOI: 10.1016/0378-1119(91)90565-S [retrieved on 1991-09-30] the whole document -----	1-15
A	SANGLARD D ET AL: "Characterization of the alkane-inducible cytochrome P450 (P450alk) gene from the yeast <i>Candida tropicalis</i>: identification of a new P450 gene family", GENE, ELSEVIER, AMSTERDAM, NL, vol. 76, no. 1, 15 March 1989 (1989-03-15), pages 121-136, XP025705303, ISSN: 0378-1119, DOI: 10.1016/0378-1119(89)90014-0 [retrieved on 1989-03-15] the whole document -----	1-15
A	STRUHL ET AL: "Yeast HIS3 expression in <i>Escherichia coli</i> depends upon fortuitous homology between eukaryotic and prokaryotic promoter elements", JOURNAL OF MOLECULAR BIOLOGY, ACADEMIC PRESS, UNITED KINGDOM, vol. 191, no. 2, 20 September 1986 (1986-09-20), pages 221-229, XP024014562, ISSN: 0022-2836, DOI: 10.1016/0022-2836(86)90259-7 [retrieved on 1986-09-20] the whole document -----	1-15

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/US2012/031909

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2004016756 A2	26-02-2004	AU 2003258244 A1 US 2005208497 A1 WO 2004016756 A2	03-03-2004 22-09-2005 26-02-2004
WO 2012027689 A1	01-03-2012	US 2012052537 A1 WO 2012027689 A1	01-03-2012 01-03-2012
WO 2012027698 A1	01-03-2012	US 2012052549 A1 WO 2012027698 A1	01-03-2012 01-03-2012
US 2005014270 A1	20-01-2005	AU 2004254361 A1 BR PI0411675 A CA 2526896 A1 CN 1842598 A EP 1636368 A2 JP 2007525969 A US 2005014270 A1 US 2008220474 A1 WO 2005003310 A2	13-01-2005 29-08-2006 13-01-2005 04-10-2006 22-03-2006 13-09-2007 20-01-2005 11-09-2008 13-01-2005
US 6673613 B2	06-01-2004	AT 383420 T AU 7603601 A CN 1444649 A DE 60132332 T2 DK 1303593 T3 EP 1303593 A2 US 2002034788 A1 US 2003175896 A1 WO 0208412 A2	15-01-2008 05-02-2002 24-09-2003 02-01-2009 19-05-2008 23-04-2003 21-03-2002 18-09-2003 31-01-2002