

US009771602B2

# (12) United States Patent

# Anthony et al.

# (54) COMPETITIVE GROWTH AND/OR PRODUCTION ADVANTAGE FOR BUTANOLOGEN MICROORGANISM

- (71) Applicant: BUTAMAX ADVANCED BIOFUELS LLC, Wilmington, DE (US)
- (72) Inventors: Larry Cameron Anthony, Aston, PA (US); Michael Dauner, Wilmington, DE (US)
- (73) Assignee: Butamax Advanced Biofuels LLC, Wilmington, DE (US)
- (\*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.
- (21) Appl. No.: 14/213,274
- (22) Filed: Mar. 14, 2014

## (65) **Prior Publication Data**

US 2014/0273130 A1 Sep. 18, 2014

#### **Related U.S. Application Data**

- (60) Provisional application No. 61/801,239, filed on Mar. 15, 2013.
- (51) Int. Cl.

C12P 7/16	(2006.01)
C12N 1/00	(2006.01)
C12P 7/10	(2006.01)

- (52) U.S. Cl. CPC C12P 7/16 (2013.01); C12P 7/10 (2013.01); Y02E 50/10 (2013.01); Y02E 50/16 (2013.01)

#### (56) References Cited

#### U.S. PATENT DOCUMENTS

4,034,091	Α	7/1977	Powell et al.
4,626,505	A *	12/1986	Falco C12N 15/81
			435/254.21
5,686,276	Α	11/1997	Laffend et al.
6,432,688	B1	8/2002	Ito et al.
7,541,173	B2	6/2009	Bramucci et al.
7,574,601	B2	8/2009	Jahromi et al.
7,659,104	B2	2/2010	Bramucci et al.
7,666,644	B2	2/2010	Castle et al.
7,851,188	B2	12/2010	Donaldson et al.
7,863,503	B2	1/2011	Castle et al.
7,910,342	B2	3/2011	Liao et al.
7,993,889	B1	8/2011	Donaldson et al.
8,017,364	B2	9/2011	Bramucci et al.
8,017,376	B2 *	9/2011	Dundon C12N 9/88
			435/254.2
8,101,808	B2	1/2012	Evanko et al.
8,129,162	B2	3/2012	Li et al.
8,178,328	B2	5/2012	Donaldson et al.
8,188,250	B2	5/2012	Bramucci et al.
8,206,970	B2	6/2012	Eliot et al.

# (10) Patent No.: US 9,771,602 B2

# (45) **Date of Patent:** Sep. 26, 2017

8,222,017	B2	7/2012	Li et al.
8,241,878	B2	8/2012	Anthony et al.
8,273,558	B2	9/2012	Donaldson et al.
8,283,144	B2	10/2012	Donaldson et al.
8,372,612	B2	2/2013	Larossa et al.
8,389,252	B2	3/2013	Larossa
8,409,834	B2	4/2013	Burlew et al.
8,426,173	B2	4/2013	Bramucci et al.
8,426,174	B2	4/2013	Bramucci et al.
8,455,224	B2	6/2013	Paul
8,455,225	B2	6/2013	Bramucci et al.
8,465,964	B2	6/2013	Anthony et al.
8,476,047	B2	7/2013	Burlew et al.
8,518,678	B2	8/2013	Flint et al.
8,530,226	B2 *	9/2013	Festel C12N 9/0006
			435/160
8,557,562	B2	10/2013	Bramucci et al.
8,614,085	B2	12/2013	Van Dyk et al.
8,617,861	B2	12/2013	Grady et al.
8,637,281	B2	1/2014	Paul et al.
8,637,289	B2	1/2014	Anthony et al.
8,652,823	B2	2/2014	Flint et al.
8,669,094	B2	3/2014	Anthony et al.
8,691,540	B2	4/2014	Bramucci et al.
8,697,404	B2	4/2014	Anton et al.
8,735,114	B2	5/2014	Donaldson et al.
8,759,044	B2	6/2014	DiCosimo et al.
8,765,425	B2	7/2014	DiCosimo et al.
8,765,433	B2	7/2014	Satagopan et al.
8,785,166	B2	7/2014	Anthony et al.
8,795,992	B2	8/2014	Bramucci et al.
8,828,694	B2	9/2014	Anthony et al.
8,828,695	B2	9/2014	Grady et al.

(Continued)

### FOREIGN PATENT DOCUMENTS

BR	8903744	1/1991
EP	0147198	7/1987
	(Cor	tinued)

#### OTHER PUBLICATIONS

Akada et al., Genetically modified industrial yeast ready for application. J. Biosci. Bioeng., 2002, vol. 94 (6): 536-544.\*

Broun et al., Catalytic plasticity of fatty acid modification enzymes underlying chemical diversity of plant lipids. Science, 1998, vol. 282: 1315-1317.\*

Chica et al., Semi-rational approaches to engineering enzyme activity: combining the benefits of directed evolution and rational design. Curr. Opi. Biotechnol., 2005, vol. 16: 378-384.\*

Devos et al., Practical limits of function prediction. Proteins: Structure, Function, and Genetics. 2000, vol. 41: 98-107.\*

Kisselev L., Polypeptide release factors in prokaryotes and eukaryotes: same function, different structure. Structure, 2002, vol. 10: 8-9.\*

(Continued)

Primary Examiner — Ganapathirama Raghu

#### (57) **ABSTRACT**

Provided herein are recombinant yeast host cells and methods for their use for production of fermentation products. Host cells provided herein comprise a pyruvate-utilizing pathway and a competitive growth advantage over other microorganisms in solution.

#### 10 Claims, 19 Drawing Sheets

#### 2(56) **References Cited**

#### U.S. PATENT DOCUMENTS

0 0 70 704	DO	0/2014	Danaldaan at al
8,828,704	DZ D2	9/2014	Donaldson et al.
8,865,443	B2	10/2014	Burlew et al.
8,871,488	B2	10/2014	Dauner et al.
8,889,385	B2	11/2014	Donaldson et al.
8.895.307	B2	11/2014	Li et al.
8,906,666	B2	12/2014	Alsaker
8 911 981	B2	12/2014	Lietal
8 040 511	B2	1/2015	Larossa
0,940,911	D2 D2	2/2015	Devel 1 event of
8,945,859	DZ D2	2/2015	Donaldson et al.
8,945,899	B2	2/2015	Li et al.
8,951,774	B2	2/2015	Donaldson et al.
8,951,937	B2	2/2015	Flint et al.
8,956,850	B2	2/2015	Anthony et al.
8,962,298	B2	2/2015	Donaldson et al.
8 969 055	B2	3/2015	Grady et al
8,969,065	B2	3/2015	Anthony et al
8,909,003	D2 *	2/2015	Annony et al.
8,975,049	B2 *	3/2015	Liao Ci2N 15/52
			435/157
8,980,612	B2	3/2015	Donaldson et al.
9,012,190	B2	4/2015	Dauner et al.
9.040.263	B2	5/2015	Anton et al.
9 267 157	B2 *	2/2016	Anthony C12N 9/0006
2007/0031018	A 1	2/2007	Dunson et al
000/010031318	A 1	7/2008	Danaldean et al
2008/0182508	AI	11/2008	Donaldson et al.
2008/02/4526	AI	11/2008	Bramucci et al.
2009/0305370	Al	12/2009	Grady et al.
2010/0081154	A1	4/2010	Flint et al.
2010/0081179	A1	4/2010	Anthony et al.
2010/0081182	A1	4/2010	Paul et al.
2010/0093020	A1	4/2010	Bramucci et al
2010/0120105	A 1	5/2010	Anthony et al
011/0007772	A 1	4/2011	Creder et al
2011/009/773	AI	4/2011	Grady et al.
2011/0124060	AI	5/2011	Anthony et al.
2011/0136192	Al	6/2011	Paul et al.
2011/0136193	A1	6/2011	Grady et al.
2011/0195505	A1	8/2011	Euler et al.
2011/0244536	A1	10/2011	Nagarajan et al.
2011/0250610	A1	10/2011	Liao et al
2011/0212044	A 1	12/2011	Anton of al
011/0312044	A1	12/2011	Division et al.
2011/0312055	AI	12/2011	Burlew et al.
2012/0058541	AI	3/2012	Alsaker et al.
2012/0064561	A1	3/2012	Flint et al.
2012/0149080	A1	6/2012	Nagarajan et al.
2012/0196341	A1	8/2012	Donaldson et al.
2012/0237988	A1	9/2012	Anthony et al.
2012/0258873	A 1	10/2012	Gibson et al
2012/0235515	A 1	2/2013	Dobson et al
012/0071909	A 1	2/2013	Anthony at al
2013/00/1898	AI	5/2015	Annony et al.
2013/01/1/06	AI	7/2013	Donaldson et al.
2013/0203138	Al	8/2013	McElvain et al.
2013/0252296	A1	9/2013	Maggio-Hall et al.
2013/0316414	A1	11/2013	Paul et al.
2014/0004526	A1	1/2014	Dauner et al.
2014/0030782	A1	1/2014	Anthony et al.
2014/0030783	A 1	1/2014	Anthony et al
014/0028262	A 1	2/2014	Flint of al
2014/0038203	AI	2/2014	
2014/0038268	AI	2/2014	Flint et al.
2014/0051133	AI	2/2014	Govindarajan et al.
2014/0051137	A1	2/2014	Flint et al.
2014/0057329	A1	2/2014	Li et al.
2014/0093930	A1	4/2014	Li et al.
2014/0093931	A1	4/2014	Dauner et al
2014/0093531	A 1	4/2014	Anton et al
014/0094030	A1	4/2014	Deserves at al
014/0090439	AI	4/2014	Brannucci et al.
2014/01414/9	AI	5/2014	Antnony et al.
2014/0170732	Al	6/2014	Bramucci et al.
2014/0186910	A1	7/2014	Rothman et al.
2014/0186911	A1	7/2014	Anthony et al.
2014/0273116	A1	9/2014	Kelly et al.
2014/0273129	Al	9/2014	Bhalla et al.
014/0308725	A 1	10/2014	Anthony et al
014/0225592	A1	11/2014	Donaldson et al
014/0333382	AI	11/2014	Donaldson et al.
2014/0349349	AI	11/2014	Dauner et al.
2014/0377824	A1	12/2014	Satagopan et al.
2015/0037855	A1	2/2015	Bhadra et al.

2015/0111269	A1	4/2015	Li et al.
2015/0119608	A1	4/2015	Donaldson et al.
2015/0125920	A1	5/2015	Anthony et al.

#### FOREIGN PATENT DOCUMENTS

WO	2006013736	2/2006
WO	2009086423	7/2009

## OTHER PUBLICATIONS

Seffernick et al., Melamine deaminase and Atrazine chlorohydrolase: 98 percent identical but functionally different. J. Bacteriol., 2001, vol. 183 (8): 2405-2410.\*

Sen et al., Developments in directed evolution for improving enzyme functions. Appl. Biochem. Biotechnol., 2007, vol. 143: 212-223.\*

Whisstock et al., Prediction of protein function from protein sequence. Q. Rev. Biophysics., 2003, vol. 36 (3): 307-340.\*

Wishart et al., A single mutation converts a novel phosphotyrosine binding domain into a dual-specificity phosphatase. J. Biol. Chem., 1995, vol. 270(45): 26782-26785.\*

Witkowski et al., Conversion of b-ketoacyl synthase to a Malonyl Decarboxylase by replacement of the active cysteine with glutamine. Biochemistry, 1999, vol. 38: 11643-11650.\*

Bajaj et al., "Construction of Killer Industrial Yeast Saccharomyces cerevisiae HAU-1 and its Fermentation Performance", Brazilian Journal of Microbiology, vol. 41(2):477-485 (2010).

Brachmann et al., "Designer deletion strains derived from *Sac-charomyces cerevisiae* S288C: a useful set of strains plasmids for PCR-mediated gene disruption and other applications," Yeast 14:115-132 (1998).

Cao et al., "A novel 5-enolpyruvylshikimate-3-phosphate synthase shows high glyphosphate tolerance in *Escherichia coli* and Tobacco plants," PLoS ONE 7(6): e38718 (2012).

Chianelli et al., "Isolation of a trifluoroleucine-resistant mutant of *Saccharomyces cerevisiae* deficient in both high- and low-affinity L-leucine transport," Cell. Mol. Biol. 42(6):847-57 (1996) (Abstract).

Dickinson et al., "An investigation of the metabolism of valine to isobutyl alcohol in *Saccharomyces cerevisiae*," J. Biol. Chem. 273:25752-6 (1998).

Domenico et al., "Resistance to bismuth among gram-negative bacteria is dependent upon iron and its uptake", Journal of Antimicrobial Chemotherapy, vol. 38(6):1031-1040 (1996).

Fukuda et al., "Altered regulation of aromatic amino acid biosynthesis in B-phenylmethyl-alcohol-overproducing mutants of Sake yeast *Saccharomyces cerevisiae*," Agric. Biol. Chem. 54:3151-6 (1990).

Gharieb et al., "Role of glutathione in detoxification of metal(loid)s by *Saccharomyces cerevisiae*", Biometals, vol. 17 (2):183-8 (2004). LaRossa and Smul, "ilvB-encoded acetolactate synthase is resistant to the herbicide sulfometuron methyl," J. Bacteriol. 160(1):391-4 (1984).

Meuris, "Feedback-insensitive mutants of the gene for the tyrosineinhibited DAHP synthetase in yeast," Genetics 76:735-44 (1974). Oba et al., "Properties of a trifluoroleucine-resistant mutant of

Saccharomyces cerevisiae," Biosci. Biotechnol. Biochem. 70(7):1776-9 (2006).

Park et al., "Use of sulfite resistance in *Saccharomyces cerevisiae* as a dominant selectable marker," Curr Genet 36(6): 339-44 (1999). Penninckx, "An overview on glutathione in *Saccharomyces* versus non-conventional yeasts", FEMS Yeast Research, Wiley-Blackwell Publishing Ltd. GB, NL, vol. 2(3):295-305 (2002).

Shimura et al., "Genetic transformation of industrial yeasts using an amino acid analog resistance gene as a directly selectable marker," Enzyme Microbiol. Technol. 15:874-6 (1993).

Siehl et al., "Evolution of a microbial acetyltransferase for modification of glyphosphate: a novel tolerance strategy," Pest Manag Sci. 61(3):235-40 (2005).

# (56) **References Cited**

# OTHER PUBLICATIONS

Stalker et al., "A single amino acid substitution in the enzyme 5-enolpyruvylshikimate-3-phosphate synthase confers resistance to the herbicide glyphosphate," J Biol Chem 260(8): 4724-8 (1985). Vande Berg et al., "Characterization and plant expression of a glyphospate-tolerant enolpyruvylshikimate phosphate synthase," Pest Manag Sci. 64(4):340-5 (2008).

Vido et al., "A Proteome Analysis of the Cadmium Response in *Saccharomyces cerevisiae*," J Biol Chem. 276(11): 8469-74 (2001). Yadav et al., "Single amino acid substitutions in the enzyme acetolactate synthase confer resistance to the herbicide sulfometuron methyl," Proc Natl Acad Sci USA 83(12):4418-22 (1986).

Falco and Dumas, "Genetic Analysis of mutants of *Saccharomyces cerevisiae* resistant to the herbicide sulfometuron methyl," Genetics 109:21-35 (1985).

Xiao and Rank, "The construction of recombinant industrial yeasts free of bacterial sequences by directed gene replacement into a nonessential region of the genome," Gene 76:99-107 (1989).

International Search Report and Written Opinion, mailed on Oct. 15, 2014, in International Patent Application No. PCT/US2014/028519, filed on Mar. 14, 2014.

\* cited by examiner





**FIG. 2** 





[l/l] xomum





[4/]] xomum



[4/]] xownw



[4/]] xownw



[4/]] xownw













Sheet 14 of 19













Sheet 17 of 19





# COMPETITIVE GROWTH AND/OR PRODUCTION ADVANTAGE FOR BUTANOLOGEN MICROORGANISM

### CROSS-REFERENCE TO RELATED APPLICATIONS

This application claims benefit of priority from U.S. Provisional Application No. 61/801,239, filed Mar. 15, 2013, which is hereby incorporated by reference in its <sup>10</sup> entirety.

## REFERENCE TO A SEQUENCE LISTING SUBMITTED ELECTRONICALLY VIA EFS-WEB

The content of the electronically submitted sequence listing (Name: 20140314\_CL5075USNP\_SequenceListing.txt; Size: 498, 298 bytes; and Date of Creation: Mar. 14, 2014) is herein <sup>20</sup> incorporated by reference in its entirety.

### FIELD OF THE INVENTION

The invention relates to the fields of industrial microbi-<sup>25</sup> ology and alcohol production. The invention also relates to the development of an industrial microorganism capable of producing fermentation products via an engineered pyruvate-utilizing pathway in the microorganism. The invention also relates to the development and use of a butanologen. <sup>30</sup> The invention also relates to the use of inhibitors, antibiotics, and mixtures thereof to give the butanologen a competitive growth and/or production advantage over other organisms in culture in order to increase the yield of fermentation products. <sup>35</sup>

# BACKGROUND OF THE INVENTION

Butanol is an important industrial chemical, useful as a fuel additive, as a feedstock chemical in the plastics industry, 40 and as a food grade extractant in the food and flavor industry. Each year 10 to 12 billion pounds of butanol are produced by petrochemical means.

Methods for the chemical synthesis of the butanol isomer, isobutanol, are known, such as oxo synthesis, catalytic 45 hydrogenation of carbon monoxide (Ullmann's Encyclopedia of Industrial Chemistry, 6th edition, 2003, Wiley-VCH Verlag GmbH and Co., Weinheim, Germany, Vol. 5, pp. 716-719) and Guerbet condensation of methanol with n-propanol (Carlini et al., *J. Molec. Catal. A. Chem.* 220:215-220, 50 2004). These processes use starting materials derived from petrochemicals. The production of isobutanol from plantderived raw materials could minimize the use of fossil fuels and would represent an advance in the art. Microorganisms capable of fermentative production of isobutanol have been 55 described (for example, in U.S. Pat. Nos. 7,851,188 and 7,993,889).

Isobutanol is produced biologically as a by-product of yeast fermentation. It is a component of "fusel oil" that forms as a result of the incomplete metabolism of amino 60 acids by this group of fungi. Isobutanol may be produced from catabolism of L-valine. After the amine group of L-valine is harvested as a nitrogen source, the resulting  $\alpha$ -keto acid is decarboxylated and reduced to isobutanol by enzymes of the so-called Ehrlich pathway (Dickinson et al., 65 *J. Biol. Chem.* 273:25752-25756, 1998). Microorganisms expressing engineered biosynthetic pathways for producing

butanol isomers, including isobutanol, are also described (see U.S. Pat. Nos. 7,851,188 and 7,993,889, which are incorporated herein by reference).

# SUMMARY OF THE INVENTION

In some embodiments, the invention is directed to a method for production of a fermentation product in a fermentation process comprising contacting a fermentation mix comprising a recombinant production microorganism which comprises a pyruvate-utilizing pathway with at least one compound which preferentially inhibits at least one contaminant yeast microorganism.

In some embodiments, the specific growth rate of the at 15 least one contaminant microorganism is reduced more than the specific growth rate of the recombinant production microorganism.

In some embodiments, production of the fermentation product of the at least one contaminant microorganism is reduced more than production of the fermentation product of the recombinant production microorganism.

In some embodiments, both the production microorganism and the at least one contaminant microorganism are yeast microorganisms. In some embodiments, the contaminant yeast microorganism is *S. cerevisiae*.

In some embodiments, the pyruvate utilizing pathway is a butanol biosynthetic pathway. In some embodiments, the pyruvate utilizing pathway is an isobutanol biosynthetic pathway. In some embodiments, the fermentation product of the at least one contaminant microorganism is ethanol.

In some embodiments, the mechanism of action of the compound that inhibits is heavy metal toxicity, inhibition of amino acid biosynthesis, sulfitolysis, cross-linking, inhibition of ethanol dehydrogenase or inhibition of pyruvate 35 decarboxylase.

In some embodiments, the inhibitor is an inhibitor of an ethanol biosynthesis pathway. In some embodiments, the inhibitor inhibits pyruvate decarboxylase and/or ethanol dehydrogenase. In some embodiments, the inhibitor comprises a member of the XC6H4CH=CHCOCOOH class of inhibitors/substrate analogues, cinnamaldehydes, glyoxalic acid, ketomalonate, regulatory site inhibitors, p chloromercuribenzoic acid, 5,5'-dithiobis(2-nitrobenzoic acid), pyrazole, 4-pyrazolecarboxylic acid, 1-H-pyrazole-1-carboxamidine-HCl, 4-methylpyrazole, 1-bromo-2-butanone, pyrazole-3,5-dicarboxylic acid monohydrate and mixtures thereof. In some embodiments, the inhibitor is selected from the group consisting of fluoroacetate, formaldehyde, sulfite, and mixtures thereof. In some embodiments, the inhibitor is an inhibitor of an amino acid biosynthesis pathway. In some embodiments, the inhibitor is inhibiting at least one enzyme selected from the group consisting of 5-enolpyruvoyl-shikimate-3-phosphate synthetase,  $\alpha$ -isopropyl malate synthase, 3-deoxy-D-arabino-heptolusonate-7-phosphate synthase and mixtures thereof. In some embodiments, the inhibitor is selected from the group consisting of imidazolinone, triazolopyrimidine, pyrimidinyl oxybenzoate, sulfonylurea, sulfonylamino carbonyl triazolinone, glyphosate, trifluoroleucine, fluorophenyalanine and mixtures thereof. In some embodiments, the inhibitor is glyphosate. In some embodiments, the inhibitor is selected from a group consisting of nicosulfuron methyl, metsulfuron methyl, chlorimuron ethyl, sulfometuron methyl, chlorsulfuron, thifensulfuron methyl, and mixtures thereof. In some embodiments, the inhibitor is selected from a group consisting of aureobasidin A, bialaphos, cerulenin, chloramphenicol, cyclohexamide, geneticin, hygromycin B, methotrexate, nourseothricin, phleomycin, triazole, and mixtures thereof. In some embodiments, the inhibitor is selected from a group consisting of bismuth (III), copper (II), and mixtures thereof.

In some embodiments, the recombinant production micro- 5 organism is engineered to express a polypeptide that increases tolerance of the host cell to the at least one compound which preferentially inhibits at least one contaminant microorganism. In some embodiments, the polypeptide comprises an amino acid sequence of at least about 10 80% identity to SEQ ID NO:9, or an active variant, fragment or derivative of SEQ ID NO:9. In some embodiments, the polypeptide comprises an amino acid sequence of at least about 80% identity to formaldehyde dehydrogenase (SEQ ID NO:7). In some embodiments, the polypeptide is selected 15 from a group consisting of an amino acid sequence of at least about 80% identity to SEQ ID NO:6, an amino acid sequence of at least about 80% identity to SEQ ID NO:7, and mixtures thereof. In some embodiments, the polypeptide is selected from a group consisting of an amino acid sequence 20 of at least about 80% identity to SEQ ID NO:11, an amino acid sequence of at least about 80% identity to SEQ ID NO:12, and mixtures thereof. In some embodiments, the polypeptide has 3-phosphoshikimate 1-carboxylvinyltransferase activity. In some embodiments, the polypeptide com- 25 prises an amino acid sequence of at least about 80% identity to 3-phosphoshikimate 1-carboxylvinyltransferase. In some embodiments, the polypeptide comprises an amino acid sequence of at least about 80% identity to SEQ ID NO:13. In some embodiments, the polypeptide is selected from a 30 group consisting of a polypeptide that has 5-enolpyruvoylshikimate-3-phosphate synthetase (ESPS) activity and confers resistance to glyphosate, a polypeptide that has glyphosate N-acetyltransferase activity and confers resistance to glyphosate, and mixtures thereof.

In some embodiments, the polypeptide is from a bacteria of the family Enterobacteriaceae. In some embodiments, the polypeptide is from a bacterial genus selected from the group consisting of: Alishewanella, Alterococcus, Aquamonas, Aranicola, Arsenophonus, Azotivirga, Blochmannia, 40 Brenneria, Buchnera, Budvicia, Buttiauxella, Cedecea, Citrobacter, Cronobacter, Dickeya, Edwardsiella, Enterobacter, Erwinia, Escherichia, Ewingella, Grimontella, Haf-Klebsiella, Kluyvera, Leclercia, Leminorella, nia. Moellerella, Morganella, Obesumbacterium, Pantoea, Can- 45 didatus Phlomobacter, Photorhabdus, Poodoomaamaana, Plesiomonas, Pragia, Proteus, Providencia, Rahnella, Raoultella, Salmonella, Samsonia, Serratia, Shigella, Sodalis, Tatumella, Trabulsiella, Wigglesworthia, Xenorhabdus, Yersinia, and Yokenella. In some embodiments, the poly- 50 peptide is from a microorganism of the genus Saccharomyces.

In some embodiments, the polypeptide is selected from a group consisting of: a polypeptide that has 5-enolpyruvoyl-shikimate-3-phosphate synthetase (ESPS) activity and con-55 fers resistance to glyphosate and a polypeptide that has glyphosate N-acetyltransferase activity and confers resistance to glyphosate. In some embodiments, the polypeptide is encoded by a heterologous polynucleotide.

In some embodiments, the invention is directed to a 60 genetically modified recombinant production microorganism comprising an engineered pyruvate-utilizing pathway; and a polypeptide that increases tolerance of the host cell to inhibitors, antibiotics, or a combination thereof, wherein the production microorganism has a growth advantage over 65 contaminant microorganisms that do not produce a desired fermentation product and do not contain said polypeptide. 4

In some embodiments, the recombinant production microorganism is selected from the group consisting of bacteria, cyanobacteria, filamentous fungi and yeasts. In some embodiments, the microorganism is a bacterial or cyanobacterial cell. In some embodiments, the genus of the microorganism is selected from the group consisting of Bacillus, Brevibacterium, Salmonella, Arthrobacter, Clostridium, Corynebacterium, Gluconobacter, Nocardia, Pseudomonas, Rhodococcus, Streptomyces, Zymomonas, Escherichia, Lactobacillus, Lactococcus, Enterococcus, Alcaligenes, Klebsiella, Paenibacillus, Xanthomonas, Saccharomyces, Pichia, Hansenula, Yarrowia, Aspergillus, Kluyveromyces, Pachysolen, Rhodotorula, Zygosaccharomyces, Galactomyces, Schizosaccharomyces, Torulaspora, Debayomyces, Williopsis, Dekkera. Kloeckera, Metschnikowia, and Candida.

In some embodiments, the recombinant production microorganism further comprises one or more polynucleotides that encode one or more enzymes having the following Enzyme Commission Numbers: EC 2.2.1.6, EC 1.1.1.86, EC 4.2.1.9, EC 4.1.1.72, EC 1.1.1.1, EC 1.1.1.265, EC 1.1.1.2, EC 1.2.4.4, EC 1.3.99.2, EC 1.2.1.57, EC 1.2.1.10, EC 2.6.1.66, EC 2.6.1.42, EC 1.4.1.9, EC 1.4.1.8, EC 4.1.1.14, EC 2.6.1.18, EC 2.3.1.9, EC 2.3.1.16, EC 1.1.130, EC 1.1.1.35, EC 1.1.1.157, EC 1.1.1.36, EC 4.2.1.17, EC 4.2.1.55, EC 1.3.1.44, EC 1.3.1.38, EC 5.4.99.13, EC 4.1.1.5, EC 2.7.1.29, EC 1.1.1.76, EC 1.2.1.57, and EC 4.2.1.28.

In some embodiments, the recombinant production microorganism has reduced expression of an enzyme having the following Enzyme Commission Number: EC 4.1.1.1 (pyruvate decarboxylase). In some embodiments, microorganism has reduced expression of an enzyme having the following Enzyme Commission Number: EC 1.1.1.1 (ethanol dehy-35 drogenase).

Some embodiments are directed to a method for the production of a C3-C6 alcohol comprising the recombinant production microorganisms described herein, wherein said engineered pyruvate-utilizing pathway is a C3-C6 alcohol biosynthetic pathway; contacting said recombinant microorganism with a fermentable carbon substrate in a fermentation medium under conditions whereby a C3-C6 alcohol is produced; and recovering said C3-C6 alcohol.

In some embodiments, the C3-C6 alcohol is produced at a titer from about 5 g/L to about 100 g/L. In some embodiments, the C3-C6 alcohol is produced at a titer of at least 20 g/L. In some embodiments, the C3-C6 alcohol is selected from the group consisting of butanol, isobutanol, propanol, isopropanol, and mixtures thereof.

Some embodiments are directed to a method for the production of ethanol comprising: providing any recombinant microorganism described herein, wherein said pyruvate-utilizing pathway is an ethanol producing pathway; contacting said recombinant microorganism with a fermentable carbon substrate in a fermentation medium under conditions whereby the ethanol is produced; and recovering said ethanol.

In some embodiments, the fermentation medium comprises one or more inhibitors, antibiotics, or combinations thereof.

In some embodiments, the ethanol is produced at a titer from about 80 g/L to about 120 g/L. In some embodiments, the ethanol is produced at a titer of about 120 g/L.

Some embodiments are directed to a composition comprising any genetically modified recombinant microorganism of the invention, a fermentation medium, and one or more inhibitors, antibiotics or combinations thereof.

65

Some embodiments are directed to a method for reducing microbial contamination in a fermentation mix, wherein said method comprises contacting any genetically modified recombinant microorganism of the invention and a fermentation medium with one or more inhibitors, antibiotics, or mixtures thereof, and wherein the addition of more inhibitors, antibiotics, or mixtures thereof results in from about 1 log to about 10 log reduction in contamination. In some embodiments, the fermentation mix is in a propagation tank. In some embodiments, the fermentation mix is in a fermenter.

In some embodiments, reduction in contamination is measured by standard plating assays, qPCR/RT-PCR, or by measuring improved fermentation yields of desired product.

Some embodiments are directed to a method for reducing microbial contamination in a fermentation mix, wherein said <sup>15</sup> method comprises contacting any genetically modified recombinant microorganism of the invention and a fermentation medium with one or more inhibitors, antibiotics, or combinations thereof, and wherein the addition of inhibitors, antibiotics, or combinations thereof results in from about 1 <sup>20</sup> log to about 10 log reduction in contamination.

In some embodiments, the addition of inhibitors, antibiotics, or combinations thereof results in the death of between about 10% and about 100% of the microbial contaminants in the fermentation mix.

Some embodiments of the invention are directed to a method for reducing microbial contamination in a fermentation mix, wherein said method comprises contacting any genetically modified recombinant microorganism of the invention and a fermentation medium comprising one or more inhibitors, antibiotics, or combinations thereof, and wherein the reduction in contamination is associated with a decrease in ethanol production. Some embodiments are directed to any composition of the invention, wherein the ethanol titer is less than about 5 g/L, or less than about 1 g/L.

Some embodiments of the invention are directed to a <sup>35</sup> method for reducing microbial contamination in a fermentation mix, wherein said method comprises contacting any genetically modified recombinant microorganism of the invention and a fermentation medium comprising one or more inhibitors, antibiotics, or combinations thereof, and <sup>40</sup> wherein the reduction in contamination is associated with an increase in ethanol production.

Some embodiments are directed to a method for reducing microbial contamination in a fermentation mix, wherein said method comprises contacting any genetically modified <sup>45</sup> recombinant microorganism of the invention and a fermentation medium comprising one or more inhibitors, antibiotics, or combinations thereof, and wherein the addition of said one or more inhibitors, antibiotics, or combinations thereof results in less than an about 20% loss in the yield of <sup>50</sup> a lower alkyl alcohol produced by said host cell due to the presence of microbial contaminants.

In some embodiments, the addition of said one or more inhibitors, antibiotics, or combinations thereof results in less than an about 10% loss in the yield of a lower alkyl alcohol <sup>55</sup> produced by said host cell due to the presence of microbial contaminants. In some embodiments, the C3-C6 alcohol or ethanol produced is a gasoline fuel component.

Some embodiments are directed to a gasoline blend comprising about 5 to about 20% of the C3-C6 alcohol <sup>60</sup> produced by the recombinant microorganisms described herein.

#### BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 depicts different isobutanol biosynthetic pathways. The steps labeled "a", "b", "c", "d", "e", "f", "g", "h", "i", 6

"j", and "k" represent substrate to product conversions described below. "a" may be catalyzed, for example, by acetolactate synthase. "b" may be catalyzed, for example, by ketol-acid reductoisomerase. "c" may be catalyzed, for example, by acetohydroxy acid dehydratase. "d" may be catalyzed, for example, by branched-chain keto acid decarboxylase. "e" may be catalyzed, for example, by branched chain alcohol dehydrogenase. "f" may be catalyzed, for example, by branched chain keto acid dehydrogenase. "g" may be catalyzed, for example, by acetylating aldehyde dehydrogenase. "h" may be catalyzed, for example, by transaminase or valine dehydrogenase. "j" may be catalyzed, for example, by valine decarboxylase. "j" may be catalyzed, for example, by omega transaminase. "k" may be catalyzed, for example, by isobutyryl-CoA mutase.

FIG. **2** depicts a growth inhibition assay measuring the ability of *S. cerevisiae* PNY0860-1A, PNY 827, and CEN.PK113-7D to grow in the presence of AHAS inhibitors.

FIG. **3** depicts the production of isobutanol and isobutyric acid as a function of time for the strain NYLA84.

FIG. 4 depicts the  $\mu$ max of PNY 827 in dependence on concentration of copper(2+) in the medium.

FIG. **5** depicts the µmax of PNY 827 in dependence on concentration of sulfometuron methyl in the medium.

FIG. 6 depicts the  $\mu$ max of PNY 827 in dependence on concentration of sulfite in the medium.

FIG. 7 depicts the µmax of PNY 827 in dependence on concentration of formaldehyde in the medium.

FIG. 8 depicts µmax of PNY 827 in dependence on concentration of 4-pyrazolecarboxylic acid in the medium.

FIG. 9 depicts  $\mu$ max of PNY 827 in dependence on concentration of 4-methylpyrazole hydrochloride in the medium.

FIG. **10** depicts µmax of PNY 827 in dependence on concentration of glyoxylic acid in the medium.

FIG. **11** depicts µmax of PNY 827 in dependence on concentration of pyrazole in the medium.

FIG. **12** depicts µmax of PNY 827 in dependence on concentration of cinnamaldehyde in the medium.

FIG. 13 depicts µmax of PNY 827 in dependence on concentration of 1-bromo-2-butanone in the medium.

FIG. 14 depicts the ratio of the produced molar butanol to ethanol concentration at EPT=8 h in mixed cultures inoculated in an OD600 ratio of 1 ethanologen strain PNY 827 to 11 butanologen strain PNY 2129 in cultures without addition of an inhibitor ("Ctrl") and trans-cinnamaldehyde concentrations of 250  $\mu$ M (A) and 25 mM (B), 1-bromo-2-butanone concentrations of 2  $\mu$ M (A) and 200  $\mu$ M (B), and pyrazole concentrations of 3 mM (A) and 30 mM (B).

FIG. **15** depicts the ratio of the produced molar butanol to ethanol concentration at EPT=8 h in mixed cultures inoculated in an OD600 ratio of 1 ethanologen strain PNY 827 to 1 butanologen strain PNY 2129 in cultures without addition of an inhibitor ("Ctrl") and trans-cinnamaldehyde at concentrations of 250  $\mu$ M (A) and 25 mM (B), 1-bromo-2butanone at concentrations of 2  $\mu$ M (A) and 200  $\mu$ M (B), and pyrazole at concentrations of 3 mM (A) and 30 mM (B).

FIG. 16 depicts the ratio of the produced molar butanol to ethanol concentration at EPT=48 h in mixed cultures inoculated in an OD600 ratio of 1 ethanologen strain PNY 827 to 11 butanologen strain PNY 2129 in cultures without addition of an inhibitor ("Ctrl") and trans-cinnamaldehyde at concentrations of 250  $\mu$ M (A) and 25 mM (B), 1-bromo-2butanone at concentrations of 2  $\mu$ M (A) and 200  $\mu$ M (B), and pyrazole at concentrations of 3 mM (A) and 30 mM (B). FIG. **17** depicts the ratio of the produced molar butanol to ethanol concentration at EPT=48 h in mixed cultures inoculated in an OD600 ratio of 1 ethanologen strain PNY 827 to 1 butanologen strain PNY 2129 in cultures without addition of an inhibitor ("Ctrl") and trans-cinnamaldehyde at con-5 centrations of 250  $\mu$ M (A) and 25 mM (B), 1-bromo-2butanone at concentrations of 2  $\mu$ M (A) and 200  $\mu$ M (B), and pyrazole at concentrations of 3 mM (A) and 30 mM (B).

FIG. **18** depicts simulated growth curves of strains A and B growing in a mixed culture at a maximum specific growth <sup>10</sup> rate of 0.16 l/h and 0.61 l/h, respectively. The ratio of the biomass of strains A vs. strain B is continuously decreasing during the cultivation and is below 3% at the end of the run.

FIG. **19** depicts the predicted effect of an inhibitor c (compound) on the maximum specific growth rate of a <sup>15</sup> hypothetical strain with a mumax without inhibitor addition  $(\mu^{o}_{max})$  of 1.00 l/h, a K<sub>r</sub>-value of 5 mM, and its behavior according a squared inhibition kinetics as described by equation (2).

#### DETAILED DESCRIPTION

Competition for carbon substrates in a butanologen fermentation process between the butanologen and contaminant microorganisms, such as, for example ethanol-produc- 25 ing yeast strains. A competitive advantage and/or selective pressure in favor of the butanologen could thus favor high yields of butanol. Such an advantage for a butanologen system may be extended to any organisms competing for the carbon substrate. The same competitive advantage may be 30 desirable for any other recombinant production microorganism, particularly yeast competing with wildtype, ethanologen yeast and/or other microbial communities.

This invention is directed to methods employing engineered microorganisms that produce fermentation products 35 for industrial uses, and to optimizations for producing such fermentation products at high rates and titers with advantaged economic process conditions.

Contamination by ethanologen yeast and other microbes can be problematic and can quickly lead to takeover of the 40 fermentation, particularly when the butanologen has a slower growth rate or is otherwise less fit than the ethanologen yeast or microbe.

Applicants have solved the problem of microbial contamination by ethanologen yeast and other microbes through 45 the use of inhibitors, antibiotics, and mixtures thereof. Butanologen yeasts either have resistance to the inhibitors, antibiotics and mixtures thereof employed, or are engineered to have resistance to the inhibitors, antibiotics, and mixtures thereof employed. The yield of the butanol process when 50 contacted with a carbon substrate may be increased without a buildup of microbial contamination.

Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this 55 invention belongs. In case of conflict, the present application including the definitions will control. Also, unless otherwise required by context, singular terms shall include pluralities and plural terms shall include the singular. All publications, patents and other references mentioned herein are incorpo-60 rated by reference in their entireties for all purposes.

In order to further define this invention, the following terms and definitions are herein provided.

As used herein, the terms "comprises," "comprising," "includes," "including," "has," "having," "contains" or 65 "containing," or any other variation thereof, will be understood to imply the inclusion of a stated integer or group of 8

integers but not the exclusion of any other integer or group of integers. For example, a composition, a mixture, a process, a method, an article, or an apparatus that comprises a list of elements is not necessarily limited to only those elements but may include other elements not expressly listed or inherent to such composition, mixture, process, method, article, or apparatus. Further, unless expressly stated to the contrary, "or" refers to an inclusive or and not to an exclusive or. For example, a condition A or B is satisfied by any one of the following: A is true (or present) and B is false (or not present), A is false (or not present) and B is true (or present), and both A and B are true (or present).

As used herein, the term "consists of," or variations such as "consist of" or "consisting of," as used throughout the specification and claims, indicate the inclusion of any recited integer or group of integers, but that no additional integer or group of integers may be added to the specified method, structure, or composition.

As used herein, the term "consists essentially of," or variations such as "consist essentially of" or "consisting essentially of," as used throughout the specification and claims, indicate the inclusion of any recited integer or group of integers, and the optional inclusion of any recited integer or group of integers that do not materially change the basic or novel properties of the specified method, structure or composition. See M.P.E.P. §2111.03.

Also, the indefinite articles "a" and "an" preceding an element or component of the invention are intended to be nonrestrictive regarding the number of instances, i.e., occurrences of the element or component. Therefore "a" or "an" should be read to include one or at least one, and the singular word form of the element or component also includes the plural unless the number is obviously meant to be singular.

The term "invention" or "present invention" as used herein is a non-limiting term and is not intended to refer to any single embodiment of the particular invention but encompasses all possible embodiments as described in the application.

As used herein, the term "about" modifying the quantity of an ingredient or reactant of the invention employed refers to variation in the numerical quantity that can occur, for example, through typical measuring and liquid handling procedures used for making concentrates or solutions in the real world; through inadvertent error in these procedures; through differences in the manufacture, source, or purity of the ingredients employed to make the compositions or to carry out the methods; and the like. The term "about" also encompasses amounts that differ due to different equilibrium conditions for a composition resulting from a particular initial mixture. Whether or not modified by the term "about", the claims include equivalents to the quantities. In embodiments, the term "about" means within 10% of the reported numerical value, preferably within 5% of the reported numerical value.

In some instances, "biomass" as used herein refers to the cell biomass of the fermentation product-producing microorganism, typically provided in units g/L dry cell weight (dcw).

The term "fermentation product" includes any desired product of interest, including, but not limited to lactic acid, 3-hydroxy-propionic acid, acrylic acid, acetic acid, succinic acid, citric acid, fumaric acid, malic acid, itaconic acid, 1,3-propane-diol, ethylene, glycerol, isobutyrate, butanol and other lower alkyl alcohols etc.

The term "fermentation process" refers to any process by which a desired fermentation product is produced.

15

25

The term "specific growth rate", often also referred to as " $\mu$ " or "mu", is defined as  $\mu = 1/cx * dcx/dt$ , representing the change of the biomass concentration cx in an infinitesimal short time interval dt, divided by the biomass concentration at this time.

The term "maximum specific growth rate", often also referred to as " $\mu_{max}$ " or "mumax", refers to the "specific growth rate" ("mu") during the exponential growth phase of a culture. Usually during the exponential growth phase, mu is approximately constant as the substrates are not limiting as well as the produced by-products are still not exerting a significant inhibition on growth.

The term "lower alkyl alcohol" refers to any straightchain or branched, saturated or unsaturated, alcohol molecule with 3-6 carbon atoms.

The term "butanol" refers to 1-butanol, 2-butanol, 2-butanone, isobutanol, or mixtures thereof. Isobutanol is also known as 2-methyl-1-propanol.

The term "C3-C6 alcohol" refers to any alcohol with 3, 4, 5 or 6 carbons.

The term "butanol biosynthetic pathway" as used herein refers to an enzyme pathway to produce 1-butanol, 2-butanol, 2-butanone or isobutanol. For example, isobutanol biosynthetic pathways are disclosed in U.S. Pat. No. 7,851, 188, which is incorporated by reference herein.

The term "isobutanol biosynthetic pathway" refers to the enzymatic pathway to produce isobutanol. From time to time "isobutanol biosynthetic pathway" is used synonymously with "isobutanol production pathway" (see U.S. Pat. Nos. 7,851,188 and 7,993,889, which are herein incorpo- 30 rated herein by reference).

The term "1-butanol biosynthetic pathway" refers to an enzymatic pathway to produce 1-butanol. A "1-butanol biosynthetic pathway" can refer to an enzyme pathway to produce 1-butanol from acetyl-coenzyme A (acetyl-CoA). 35 For example, 1-butanol biosynthetic pathways are disclosed in U.S. Patent Application Publication No. 2008/0182308 and International Publication No. WO 2007/041269, which are herein incorporated by reference in their entireties.

The term "2-butanol biosynthetic pathway" refers to an 40 enzymatic pathway to produce 2-butanol. A "2-butanol biosynthetic pathway" can refer to an enzyme pathway to produce 2-butanol from pyruvate. For example, 2-butanol biosynthetic pathways are disclosed in U.S. Pat. No. 8,206, 970, U.S. Patent Application Publication No. 2007/0292927, 45 International Publication Nos. WO 2007/130518 and WO 2007/130521, which are herein incorporated by reference in their entireties.

The term "2-butanone biosynthetic pathway" as used herein refers to an enzymatic pathway to produce 2-bu- 50 tanone (see U.S. Appl. Pub. No. 2007/0259410 and U.S. Appl. Pub. No. 2009/0155870, which are incorporated herein by reference).

The term "engineered" as used herein refers to an enzymatic pathway that is not present endogenously in a micro- 55 organism and is deliberately constructed to produce a fermentation product from a starting substrate through a series of specific substrate to product conversions.

A "recombinant microbial host cell" or a "recombinant microorganism" is defined as a host cell that has been 60 genetically manipulated to express a biosynthetic production pathway, wherein the host cell either produces a biosynthetic product in greater quantities relative to an unmodified host cell or produces a biosynthetic product that is not ordinarily produced by an unmodified host cell. A "production micro- 65 organism" is any microorganism that produces a desired fermentation product. A "contaminant microorganism" is

10

any microorganism that either does not produce a desired fermentation product or does produce a desired fermentation product, but at lower efficiency (for example, with lower specific productivity, rate, titer or yield) than a production microorganism. It will be appreciated that microorganisms may produce measureable amounts of more than one product, however, for the purposes herein, "product" typically refers to the major product produced by a microorganism.

The term "fermentable carbon substrate" refers to a carbon source capable of being metabolized by the microorganisms such as those disclosed herein. Suitable fermentable carbon substrates include, but are not limited to, monosaccharides, such as glucose or fructose; disaccharides, such as lactose or sucrose; oligosaccharides; polysaccharides, such as starch, cellulose, or lignocellulose, hemicellulose; one-carbon substrates, fatty acids; and any combination of these.

"Fermentation medium" as used herein means the mixture of water, sugars (fermentable carbon substrates), dissolved 20 solids, fermentation product and all other constituents of the material in which the fermentation product is being made by the reaction of fermentable carbon substrates to fermentation products, water and carbon dioxide (CO<sub>2</sub>) by the microorganisms present. From time to time, as used herein the term "fermentation broth", "fermentation mix" and "fermentation mixture" can be used synonymously with "fermentation medium."

The term "aerobic conditions" as used herein means growth conditions in the presence of oxygen.

The term "microaerobic conditions" as used herein means growth conditions with low levels of dissolved oxygen. For example, the oxygen level may be less than about 1% of air-saturation.

The term "anaerobic conditions" as used herein means growth conditions in the absence of oxygen.

The term "carbon substrate" refers to a carbon source capable of being metabolized by the recombinant host cells disclosed herein. Non-limiting examples of carbon substrates are provided herein and include, but are not limited to, monosaccharides, oligosaccharides, polysaccharides, ethanol, lactate, succinate, glycerol, carbon dioxide, methanol, glucose, fructose, sucrose, xylose, arabinose, dextrose, amino acids, and mixtures thereof.

The term "sucrose utilizing butanologen" as used herein refers to a microorganism capable of producing butanol from sucrose. Such microorganisms are typically recombinant microorganisms comprising an engineered butanol biosynthetic pathway. "Sucrose utilizing isobutanologen" as used herein refers to a microorganism capable of producing isobutanol from sucrose. Such microorganisms are typically recombinant microorganisms comprising an engineered isobutanol biosynthetic pathway.

As used herein, the term "yield" refers to the amount of product per amount of carbon source in g/g. The yield may be exemplified for glucose as the carbon source. It is understood unless otherwise noted that yield is expressed as a percentage of the theoretical yield. In reference to a microorganism or metabolic pathway, "theoretical yield" is defined as the maximum amount of product that can be generated per total amount of substrate as dictated by the stoichiometry of the metabolic pathway used to make the product. For example, the theoretical yield for one typical conversion of glucose to isopropanol is 0.33 g/g. As such, a yield of isopropanol from glucose of 29.7 g/g would be expressed as 90% of theoretical or 90% theoretical yield. It is understood that while in the present disclosure the yield is exemplified for glucose as a carbon source, the invention can

be applied to other carbon sources and the yield may vary depending on the carbon source used. One skilled in the art can calculate yields on various carbon sources.

The term "effective titer" as used herein, refers to the total amount of C3-C6 alcohol produced by fermentation per liter <sup>5</sup> of fermentation medium. The total amount of C3-C6 alcohol includes: (i) the amount of C3-C6 alcohol in the fermentation medium; (ii) the amount of C3-C6 alcohol recovered from the organic extractant; and (iii) the amount of C3-C6 alcohol recovered from the gas phase, if gas stripping is <sup>10</sup> used.

The term "effective rate" as used herein, refers to the total amount of C3-C6 alcohol produced by fermentation per liter of fermentation medium per hour of fermentation.

The term "specific productivity" as used herein, refers to the g of C3-C6 alcohol produced per g of dry cell weight of cells per unit time.

As used herein the term "coding sequence" refers to a DNA sequence that encodes for a specific amino acid <sub>20</sub> sequence. "Suitable regulatory sequences" refer to nucleotide sequences located upstream (5' non-coding sequences), within, or downstream (3' non-coding sequences) of a coding sequence, and which influence the transcription, RNA processing or stability, or translation of the associated cod- <sup>25</sup> ing sequence. Regulatory sequences may include promoters, translation leader sequences, introns, polyadenylation recognition sequences, RNA processing site, effector binding site and stem-loop structure.

The terms "derivative" and "analog" refer to a polypeptide differing from the enzymes of the invention, but retaining essential properties thereof. The term "derivative" may also refer to a host cells differing from the host cells of the invention, but retaining essential properties thereof. Generally, derivatives and analogs are overall closely similar, and, in many regions, identical to the enzymes of the invention. The terms "derived-from", "derivative" and "analog" when referring to enzymes of the invention include any polypeptides which retain at least some of the activity of the 40 corresponding native polypeptide or the activity of its catalytic domain.

Derivatives of enzymes disclosed herein are polypeptides which may have been altered so as to exhibit features not found on the native polypeptide. Derivatives can be cova- 45 lently modified by substitution (e.g. amino acid substitution), chemical, enzymatic, or other appropriate means with a moiety other than a naturally occurring amino acid (e.g., a detectable moiety such as an enzyme or radioisotope). Examples of derivatives include fusion proteins, or proteins 50 which are based on a naturally occurring protein sequence, but which have been altered. For example, proteins can be designed by knowledge of a particular amino acid sequence, and/or a particular secondary, tertiary, and/or quaternary structure. Derivatives include proteins that are modified 55 based on the knowledge of a previous sequence, natural or synthetic, which is then optionally modified, often, but not necessarily to confer some improved function. These sequences, or proteins, are then said to be derived from a particular protein or amino acid sequence. In some embodi- 60 ments of the invention, a derivative must retain at least 50% identity, at least 60% identity, at least 70% identity, at least 80% identity, at least 90% identity, at least 95% identity, at least 97% identity, or at least 99% identity to the sequence the derivative is "derived-from." In some embodiments of 65 the invention, an enzyme is said to be derived-from an enzyme naturally found in a particular species if, using

molecular genetic techniques, the DNA sequence for part or all of the enzyme is amplified and placed into a new host cell.

Polypeptides and Polynucleotides for Use in the Invention As used herein, the term "polypeptide" is intended to encompass a singular "polypeptide" as well as plural "polypeptides," and refers to a molecule composed of monomers (amino acids) linearly linked by amide bonds (also known as peptide bonds). The term "polypeptide" refers to any chain or chains of two or more amino acids, and does not refer to a specific length of the product. Thus, peptides, dipeptides, tripeptides, oligopeptides, "protein," "amino acid chain," or any other term used to refer to a chain or chains of two or more amino acids, are included within the definition of "polypeptide," and the term "polypeptide" may be used instead of, or interchangeably with any of these terms. A polypeptide may be derived from a natural biological source or produced by recombinant technology, but is not necessarily translated from a designated nucleic acid sequence. It may be generated in any manner, including by chemical synthesis. The polypeptides used in this invention comprise full-length polypeptides and fragments thereof.

By an "isolated" polypeptide or a fragment, variant, or derivative thereof is intended a polypeptide that is not in its natural milieu. No particular level of purification is required. For example, an isolated polypeptide can be removed from its native or natural environment. Recombinantly produced polypeptides and proteins expressed in host cells are considered isolated for the purposes of the invention, as are native or recombinant polypeptides which have been separated, fractionated, or partially or substantially purified by any suitable technique.

A polypeptide of the invention may be of a size of about 10 or more, 20 or more, 25 or more, 50 or more, 75 or more, 100 or more, 200 or more, 500 or more, 1,000 or more, or 2,000 or more amino acids. Polypeptides may have a defined three-dimensional structure, although they do not necessarily have such structure. Polypeptides with a defined threedimensional structure are referred to as folded, and polypeptides which do not possess a defined three-dimensional structure, but rather can adopt a large number of different conformations, and are referred to as unfolded.

Also included as polypeptides of the present invention are derivatives, analogs, or variants of the foregoing polypeptides, and any combination thereof. The terms "active variant," "active fragment," "active derivative," and "analog" refer to polypeptides of the present invention and include any polypeptides that are capable of catalyzing the reduction of a lower alkyl aldehyde. Variants of polypeptides of the present invention include polypeptides with altered amino acid sequences due to amino acid substitutions, deletions, and/or insertions. Variants may occur naturally or be nonnaturally occurring. Non-naturally occurring variants may be produced using art-known mutagenesis techniques. Variant polypeptides may comprise conservative or non-conservative amino acid substitutions, deletions and/or additions. Derivatives of polypeptides of the present invention are polypeptides which have been altered so as to exhibit additional features not found on the native polypeptide. Examples include fusion proteins. Variant polypeptides may also be referred to herein as "polypeptide analogs." As used herein a "derivative" of a polypeptide refers to a subject polypeptide having one or more residues chemically derivatized by reaction of a functional side group. Also included as "derivatives" are those peptides which contain one or more naturally occurring amino acid derivatives of the twenty standard amino acids. For example, 4-hydroxyproline may

be substituted for proline; 5-hydroxylysine may be substituted for lysine; 3-methylhistidine may be substituted for histidine; homoserine may be substituted for serine; and ornithine may be substituted for lysine.

A "fragment" is a unique portion of a polypeptide or other 5 enzyme used in the invention which is identical in sequence to but shorter in length than the parent full-length sequence. A fragment may comprise up to the entire length of the defined sequence, minus one amino acid residue. For example, a fragment may comprise from 5 to 1000 contigu- 10 ous amino acid residues. A fragment may be at least 5, 10, 15, 20, 25, 30, 40, 50, 60, 75, 100, 150, 250 or at least 500 contiguous amino acid residues in length. Fragments may be preferentially selected from certain regions of a molecule. For example, a polypeptide fragment may comprise a certain 15 length of contiguous amino acids selected from the first 100 or 200 amino acids of a polypeptide as shown in a certain defined sequence. Clearly these lengths are exemplary, and any length that is supported by the specification, including the Sequence Listing, tables, and figures, may be encom- 20 passed by the present embodiments.

Alternatively, recombinant variants encoding these same or similar polypeptides can be synthesized or selected by making use of the "redundancy" in the genetic code. Various codon substitutions, such as the silent changes which pro- 25 duce various restriction sites, may be introduced to optimize cloning into a plasmid or viral vector or expression in a host cell system.

Preferably, amino acid "substitutions" are the result of replacing one amino acid with another amino acid having 30 similar structural and/or chemical properties, i.e., conservative amino acid replacements, or they can be result of replacing one amino acid with an amino acid having different structural and/or chemical properties, i.e., non-conservative amino acid replacements. "Conservative" amino acid 35 substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues involved. For example, nonpolar (hydrophobic) amino acids include alanine, leucine, isoleucine, valine, proline, phenylalanine, 40 tryptophan, and methionine; polar neutral amino acids include glycine, serine, threonine, cysteine, tyrosine, asparagine, and glutamine; positively charged (basic) amino acids include arginine, lysine, and histidine; and negatively charged (acidic) amino acids include aspartic acid and 45 glutamic acid. Alternatively, "non-conservative" amino acid substitutions can be made by selecting the differences in polarity, charge, solubility, hydrophobicity, hydrophilicity, or the amphipathic nature of any of these amino acids. "Insertions" or "deletions" are preferably in the range of 50 about 1 to about 20 amino acids, more preferably 1 to 10 amino acids. The variation allowed may be experimentally determined by systematically making insertions, deletions, or substitutions of amino acids in a polypeptide molecule using recombinant DNA techniques and assaying the result- 55 ing recombinant variants for activity.

By a polypeptide having an amino acid or polypeptide sequence at least, for example, 95% "identical" to a query amino acid sequence of the present invention, it is intended that the amino acid sequence of the subject polypeptide is 60 identical to the query sequence except that the subject polypeptide sequence may include up to five amino acid alterations per each 100 amino acids of the query amino acid sequence. In other words, to obtain a polypeptide having an amino acid sequence at least 95% identical to a query amino 65 acid sequence, up to 5% of the amino acid residues in the subject sequence may be inserted, deleted, or substituted

with another amino acid. These alterations of the reference sequence may occur at the amino or carboxy terminal positions of the reference amino acid sequence or anywhere between those terminal positions, interspersed either individually among residues in the reference sequence or in one or more contiguous groups within the references sequence.

As a practical matter, whether any particular polypeptide is at least 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identical to a reference polypeptide can be determined conventionally using known computer programs. A preferred method for determining the best overall match between a query sequence (a sequence of the present invention) and a subject sequence, also referred to as a global sequence alignment, can be determined using the FASTDB computer program based on the algorithm of Brutlag et al., Comp. Appi. Biosci. 6:237-245 (1990). In a sequence alignment, the query and subject sequences are either both nucleotide sequences or both amino acid sequences. The result of the global sequence alignment is in percent identity. Preferred parameters used in a FASTDB amino acid alignment are: Matrix=PAM 0, k-tuple=2, Mismatch Penalty=1, Joining Penalty=20, Randomization Group Length=0, Cutoff Score=1, Window Size=sequence length, Gap Penalty=5, Gap Size Penalty-0.05, Window Size=500 or the length of the subject amino acid sequence, whichever is shorter.

If the subject sequence is shorter than the query sequence due to N- or C-terminal deletions, not because of internal deletions, a manual correction must be made to the results. This is because the FASTDB program does not account for N- and C-terminal truncations of the subject sequence when calculating global percent identity. For subject sequences truncated at the N- and C-termini, relative to the query sequence, the percent identity is corrected by calculating the number of residues of the query sequence that are N- and C-terminal of the subject sequence, which are not matched/ aligned with a corresponding subject residue, as a percent of the total bases of the query sequence. Whether a residue is matched/aligned is determined by results of the FASTDB sequence alignment. This percentage is then subtracted from the percent identity, calculated by the above FASTDB program using the specified parameters, to arrive at a final percent identity score. This final percent identity score is what is used for the purposes of the present invention. Only residues to the N- and C-termini of the subject sequence, which are not matched/aligned with the query sequence, are considered for the purposes of manually adjusting the percent identity score. That is, only query residue positions outside the farthest N- and C-terminal residues of the subject sequence.

For example, a 90 amino acid residue subject sequence is aligned with a 100 residue query sequence to determine percent identity. The deletion occurs at the N-terminus of the subject sequence and therefore, the FASTDB alignment does not show a matching/alignment of the first 10 residues at the N-terminus. The 10 unpaired residues represent 10% of the sequence (number of residues at the N- and C-termini not matched/total number of residues in the query sequence) so 10% is subtracted from the percent identity score calculated by the FASTDB program. If the remaining 90 residues were perfectly matched the final percent identity would be 90%. In another example, a 90 residue subject sequence is compared with a 100 residue query sequence. This time the deletions are internal deletions so there are no residues at the N- or C-termini of the subject sequence which are not matched/aligned with the query. In this case, the percent identity calculated by FASTDB is not manually corrected. Once again, only residue positions outside the N- and

1

C-terminal ends of the subject sequence, as displayed in the FASTDB alignment, which are not matched/aligned with the query sequence are manually corrected for. No other manual corrections are to be made for the purposes of the present invention.

Polypeptides and other enzymes suitable for use in the present invention and fragments thereof are encoded by polynucleotides. The term "polynucleotide" is intended to encompass a singular nucleic acid as well as plural nucleic acids, and refers to an isolated nucleic acid molecule or construct, e.g., messenger RNA (mRNA), virally-derived RNA, or plasmid DNA (pDNA). A polynucleotide may comprise a conventional phosphodiester bond or a nonconventional bond (e.g., an amide bond, such as found in peptide nucleic acids (PNA)). The term "nucleic acid" refers to any one or more nucleic acid segments, e.g., DNA or RNA fragments, present in a polynucleotide. Polynucleotides according to the present invention further include such molecules produced synthetically. Polynucleotides of 20 the invention may be native to the host cell or heterologous. In addition, a polynucleotide or a nucleic acid may be or may include a regulatory element such as a promoter, ribosome binding site, or a transcription terminator.

In certain embodiments, the polynucleotide or nucleic 25 acid is DNA. In the case of DNA, a polynucleotide comprising a nucleic acid, which encodes a polypeptide normally may include a promoter and/or other transcription or translation control elements operably associated with one or more coding regions. An operable association is when a 30 coding region for a gene product, e.g., a polypeptide, is associated with one or more regulatory sequences in such a way as to place expression of the gene product under the influence or control of the regulatory sequence(s). Two DNA fragments (such as a polypeptide coding region and a 35 promoter associated therewith) are "operably associated" if induction of promoter function results in the transcription of mRNA encoding the desired gene product and if the nature of the linkage between the two DNA fragments does not interfere with the ability of the expression regulatory 40 sequences to direct the expression of the gene product or interfere with the ability of the DNA template to be transcribed. Thus, a promoter region would be operably associated with a nucleic acid encoding a polypeptide if the promoter was capable of effecting transcription of that 45 nucleic acid. Other transcription control elements, besides a promoter, for example enhancers, operators, repressors, and transcription termination signals, can be operably associated with the polynucleotide. Suitable promoters and other transcription control regions are disclosed herein. 50

A polynucleotide sequence can be referred to as "isolated," in which it has been removed from its native environment. For example, a heterologous polynucleotide encoding a polypeptide or polypeptide fragment having enzymatic activity (e.g., the ability to convert a substrate to 55 xylulose) contained in a vector is considered isolated for the purposes of the present invention. Further examples of an isolated polynucleotide include recombinant polynucleotides maintained in heterologous host cells or purified (partially or substantially) polynucleotides in solution. Iso- 60 lated polynucleotides or nucleic acids according to the present invention further include such molecules produced synthetically. An isolated polynucleotide fragment in the form of a polymer of DNA can be comprised of one or more segments of cDNA, genomic DNA, or synthetic DNA. 65

The term "gene" refers to a nucleic acid fragment that is capable of being expressed as a specific protein, optionally including regulatory sequences preceding (5' non-coding sequences) and following (3' non-coding sequences) the coding sequence.

As used herein, a "coding region" or "ORF" is a portion of nucleic acid which consists of codons translated into amino acids. Although a "stop codon" (TAG, TGA, or TAA) is not translated into an amino acid, it may be considered to be part of a coding region, if present, but any flanking sequences, for example promoters, ribosome binding sites, transcriptional terminators, introns, 5' and 3' non-translated regions, and the like, are not part of a coding region. "Suitable regulatory sequences" refer to nucleotide sequences located upstream (5' non-coding sequences), within, or downstream (3' non-coding sequences) of a coding sequence that influence the transcription, RNA processing or stability, or translation of the associated coding sequence. Regulatory sequences can include promoters, translation leader sequences, introns, polyadenylation recognition sequences, RNA processing sites, effector binding sites and stem-loop structures.

A variety of translation control elements are known to those of ordinary skill in the art. These include, but are not limited to ribosome binding sites, translation initiation and termination codons, and elements derived from viral systems (particularly an internal ribosome entry site, or IRES). In other embodiments, a polynucleotide of the present invention is RNA, for example, in the form of messenger RNA (mRNA). RNA of the present invention may be single stranded or double stranded.

Polynucleotide and nucleic acid coding regions of the present invention may be associated with additional coding regions which encode secretory or signal peptides, which direct the secretion of a polypeptide encoded by a polynucleotide of the present invention.

As used herein, the term "transformation" refers to the transfer of a nucleic acid fragment into the genome of a host organism, resulting in genetically stable inheritance. Host organisms containing the transformed nucleic acid fragments are referred to as "recombinant" or "transformed" organisms.

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

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

The term "artificial" refers to a synthetic, or non-host cell derived composition, e.g., a chemically-synthesized oligo-nucleotide.

As used herein, "native" refers to the form of a polynucleotide, gene, or polypeptide as found in nature with its own regulatory sequences, if present.

The term "endogenous," when used in reference to a polynucleotide, a gene, or a polypeptide refers to a native 5 polynucleotide or gene in its natural location in the genome of an organism, or for a native polypeptide, is transcribed and translated from this location in the genome.

The term "heterologous" when used in reference to a polynucleotide, a gene, or a polypeptide refers to a poly- 10 nucleotide, gene, or polypeptide not normally found in the host organism. "Heterologous" also includes a native coding region, or portion thereof, that is reintroduced into the source organism in a form that is different from the corresponding native gene, e.g., not in its natural location in the organism's genome. The heterologous polynucleotide or gene may be introduced into the host organism by, e.g., gene transfer. A heterologous gene may include a native coding region with non-native regulatory regions that is reintroduced into the native host. A "transgene" is a gene that has been introduced into the genome by a transformation procedure.

The term "recombinant genetic expression element" refers to a nucleic acid fragment that expresses one or more specific proteins, including regulatory sequences preceding (5' non-coding sequences) and following (3' termination <sup>25</sup> sequences) coding sequences for the proteins. A chimeric gene is a recombinant genetic expression element. The coding regions of an operon may form a recombinant genetic expression element, along with an operably linked promoter and termination region.

'Regulatory sequences'' refers 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 35 sequence. Regulatory sequences may include promoters, enhancers, operators, repressors, transcription termination signals, translation leader sequences, introns, polyadenylation recognition sequences, RNA processing site, effector binding site and stem-loop structure.

The term "promoter" refers to a nucleic acid sequence 40 capable of controlling the expression of a coding sequence or functional RNA. In general, a coding sequence is located 3' to a promoter sequence. Promoters may be derived in their entirety from a native gene, or be composed of different elements derived from different promoters found in nature, 45 or even comprise synthetic nucleic acid 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. Pro- 50 moters which cause a gene to be expressed in most cell types at most times are commonly referred to as "constitutive promoters". "Inducible promoters," on the other hand, cause a gene to be expressed when the promoter is induced or turned on by a promoter-specific signal or molecule. It is 55 further recognized that since in most cases the exact boundaries of regulatory sequences have not been completely defined, DNA fragments of different lengths may have identical promoter activity. For example, it will be understood that "FBA1 promoter" can be used to refer to a fragment derived from the promoter region of the FBA1 gene.

The term "terminator" as used herein refers to DNA sequences located downstream of a coding sequence. This includes polyadenylation recognition sequences and other sequences encoding regulatory signals capable of affecting 65 codons to code for insertion of a particular amino acid in a 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. It is recognized that since in most cases the exact boundaries of regulatory sequences have not been completely defined, DNA fragments of different lengths may have identical terminator activity. For example, it will be understood that "CYC1 terminator" can be used to refer to a fragment derived from the terminator region of the CYC1 gene.

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

The term "codon-optimized" as it refers to genes or 20 coding regions of nucleic acid molecules for transformation of various hosts, refers to the alteration of codons in the gene or coding regions of the nucleic acid molecules to reflect the typical codon usage of the host organism without altering the polypeptide encoded by the DNA. Such optimization includes replacing at least one, or more than one, or a significant number, of codons with one or more codons that are more frequently used in the genes of that organism.

Deviations in the nucleotide sequence that comprise the codons encoding the amino acids of any polypeptide chain allow for variations in the sequence coding for the gene. Since each codon consists of three nucleotides, and the nucleotides comprising DNA are restricted to four specific bases, there are 64 possible combinations of nucleotides, 61 of which encode amino acids (the remaining three codons encode signals ending translation). The "genetic code" which shows which codons encode which amino acids is reproduced herein as Table 1. As a result, many amino acids are designated by more than one codon. For example, the amino acids alanine and proline are coded for by four triplets, serine and arginine by six, whereas tryptophan and methionine are coded by just one triplet. This degeneracy allows for DNA base composition to vary over a wide range without altering the amino acid sequence of the proteins encoded by the DNA.

TABLE 1

		The Standard G	enetic Code	
	Т	С	А	G
) T	TTT Phe (F) TTC Phe (F) TTA Leu (L)	TCT Ser (S) TCC Ser (S) TCA Ser (S)	TAT Tyr (Y) TAC Tyr (Y) TAA Ter TAG Ter	TGT Cys (C) TGC TGA Ter TGG Trp (W)
С	CTT Leu (L) CTC Leu (L) CTC Leu (L) CTA Leu (L)	CCT Pro (P) CCC Pro (P) CCA Pro (P)	CAT His (H) CAC His (H) CAA Gln (Q)	CGT Arg (R) CGC Arg (R) CGA Arg (R)
A	ATT IIe (I) ATC IIe (I) ATA IIe (I) ATG Met (M)	ACT Thr (T) ACC Thr (T) ACC Thr (T) ACA Thr (T) ACG Thr (T)	AAT Asn (N) AAC Asn (N) AAC Asn (N) AAA Lys (K) AAG Lys (K)	AGT Ser (S) AGC Ser (S) AGA Arg (R) AGG Arg (R)
G )	GTT Val (V) GTC Val (V) GTA Val (V) GTG Val (V)	GCT Ala (A) GCC Ala (A) GCA Ala (A) GCG Ala (A)	GAT Asp (D) GAC Asp (D) GAA Glu (E) GAG Glu (E)	GGT Gly (G) GGC Gly (G) GGA Gly (G) GGG Gly (G)

Many organisms display a bias for use of particular growing peptide chain. Codon preference or codon bias, differences in codon usage between organisms, is afforded

25

by degeneracy of the genetic code, and is well documented among many organisms. Codon bias often correlates with the efficiency of translation of messenger RNA (mRNA), which is in turn believed to be dependent on, inter alia, the properties of the codons being translated and the availability 5 of particular transfer RNA (tRNA) molecules. The predominance of selected tRNAs in a cell is generally a reflection of the codons used most frequently in peptide synthesis. Accordingly, genes can be tailored for optimal gene expression in a given organism based on codon optimization.

Given the large number of gene sequences available for a wide variety of animal, plant and microbial species, it is possible to calculate the relative frequencies of codon usage. Codon usage tables are readily available, for example, at the "Codon Usage Database" available at the Kazusa DNA 15 Research Institute, Japan, and these tables can be adapted in a number of ways. See Nakamura, Y., et al. Nucl. Acids Res. 28:292(2000). Codon usage tables for yeast, calculated from GenBank Release 128.0 [15 Feb. 2002], are reproduced below as Table 2. This table uses mRNA nomenclature, and 20 so instead of thymine (T) which is found in DNA, the tables use uracil (U) which is found in RNA. The Table has been adapted so that frequencies are calculated for each amino acid, rather than for all 64 codons.

TABLE 2

Codon	Usage Table for	Saccharomyces ce	erevisiae Genes
Amino Acid	Codon	Number	Frequency per thousand
Phe	UUU	170666	26.1
Phe	UUC	120510	18.4
Leu	UUA	170884	26.2
Leu	UUG	177573	27.2
Leu	CUU	80076	12.3
Leu	CUC	35545	5.4
Leu	CUA	87619	13.4
Leu	CUG	68494	10.5
Ile	AUU	196893	30.1
Ile	AUC	112176	17.2
Ile	AUA	116254	17.8
Met	AUG	136805	20.9
Val	GUU	144243	22.1
Val	GUC	76947	11.8
Val	GUA	76927	11.8
Val	GUG	70337	10.8
Ser	UCU	153557	23.5
Ser	UCC	92923	14.2
Ser	UCA	122028	18.7
Ser	UCG	55951	8.6
Ser	AGU	92466	14.2
Ser	AGC	63726	9.8
Pro	CCU	88263	13.5
Pro	CCC	44309	6.8
Pro	CCA	119641	18.3
Pro	CCG	34597	5.3
Thr	ACU	132522	20.3
Thr	ACC	83207	12.7
Thr	ACA	116084	17.8
Thr	ACG	52045	8.0
Ala	GCU	138358	21.2
Ala	GCC	82357	12.6
Ala	GCA	105910	16.2
Ala	GCG	40358	6.2
Tyr	UAU	122728	18.8
Tyr	UAC	96596	14.8
His	CAU	89007	13.6
His	CAC	50785	7.8
Gln	CAA	178251	27.3
Gln	CAG	79121	12.1
Asn	AAU	233124	35.7
Asn	AAC	162199	24.8
Lys	AAA	273618	41.9
Lys	AAG	201361	30.8

20 TABLE 2-continued

Amino Acid	Codon	Number	Frequency per thousand
Asp	GAU	245641	37.6
Asp	GAC	132048	20.2
Glu	GAA	297944	45.6
Glu	GAG	125717	19.2
Cys	UGU	52903	8.1
Cys	UGC	31095	4.8
Trp	UGG	67789	10.4
Arg	CGU	41791	6.4
Arg	CGC	16993	2.6
Arg	CGA	19562	3.0
Arg	CGG	11351	1.7
Arg	AGA	139081	21.3
Arg	AGG	60289	9.2
Gly	GGU	156109	23.9
Gly	GGC	63903	9.8
Gly	GGA	71216	10.9
Gly	GGG	39359	6.0
Stop	UAA	6913	1.1
Stop	UAG	3312	0.5

By utilizing this or similar tables, one of ordinary skill in the art can apply the frequencies to any given polypeptide sequence, and produce a nucleic acid fragment of a codonoptimized coding region which encodes the polypeptide, but which uses codons optimal for a given species.

Randomly assigning codons at an optimized frequency to 30 encode a given polypeptide sequence, can be done manually by calculating codon frequencies for each amino acid, and then assigning the codons to the polypeptide sequence randomly. Additionally, various algorithms and computer 35 software programs are readily available to those of ordinary skill in the art. For example, the "EditSeq" function in the Lasergene Package, available from DNAstar, Inc., Madison, WI, the backtranslation function in the VectorNTl Suite, available from InforMax, Inc., Bethesda, MD, and the 40 "backtranslate" function in the GCG--Wisconsin Package, available from Accelrys, Inc., San Diego, CA. In addition, various resources are publicly available to codon-optimize coding region sequences, e.g., the "JAVA Codon Adaptation Tool" (Grote, et al., Nucl. Acids Res. 33:W526-W531, 2005) 45 and the "Codon optimization tool" available at Entelechon GmbH, Regensburg, Germany.

By a nucleic acid or polynucleotide having a nucleotide sequence at least, for example, 95% "identical" to a reference nucleotide sequence of the present invention, it is 50 intended that the nucleotide sequence of the polynucleotide is identical to the reference sequence except that the polynucleotide sequence may include up to five point mutations per each 100 nucleotides of the reference nucleotide sequence. In other words, to obtain a polynucleotide having 55 a nucleotide sequence at least 95% identical to a reference nucleotide sequence, up to 5% of the nucleotides in the reference sequence may be deleted or substituted with another nucleotide, or a number of nucleotides up to 5% of the total nucleotides in the reference sequence may be 60 inserted into the reference sequence.

As a practical matter, whether any particular nucleic acid molecule or polynucleotide is at least 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identical to a nucleotide sequence or polypeptide sequence of the present invention can be 65 determined conventionally using known computer programs. A preferred method for determining the best overall match between a query sequence (a sequence of the present

invention) and a subject sequence, also referred to as a global sequence alignment, can be determined using the FASTDB computer program based on the algorithm of Brutlag et al., *Comp. Appl. Biosci.* 6:237-245 (1990). In a sequence alignment the query and subject sequences are 5 both DNA sequences. An RNA sequence can be compared by converting U's to T's. The result of the global sequence alignment is in percent identity. Preferred parameters used in a FASTDB alignment of DNA sequences to calculate percent identity are: Matrix=Unitary, k-tuple=4, Mismatch Pen-10 alty=1, Joining Penalty-30, Randomization Group Length=0, Cutoff Score=1, Gap Penalty=5, Gap Size Penalty=0.05, Window Size=500 or the length of the subject nucleotide sequences, whichever is shorter.

If the subject sequence is shorter than the query sequence 15 because of 5' or 3' deletions, not because of internal deletions, a manual correction must be made to the results. This is because the FASTDB program does not account for 5' and 3' truncations of the subject sequence when calculating percent identity. For subject sequences truncated at the 5' or 20 3' ends, relative to the query sequence, the percent identity is corrected by calculating the number of bases of the query sequence that are 5' and 3' of the subject sequence, which are not matched/aligned, as a percent of the total bases of the query sequence. Whether a nucleotide is matched/aligned is 25 determined by results of the FASTDB sequence alignment. This percentage is then subtracted from the percent identity, calculated by the above FASTDB program using the specified parameters, to arrive at a final percent identity score. This corrected score is what is used for the purposes of the 30 present invention. Only bases outside the 5' and 3' bases of the subject sequence, as displayed by the FASTDB alignment, which are not matched/aligned with the query sequence, are calculated for the purposes of manually adjusting the percent identity score.

For example, a 90 base subject sequence is aligned to a 100 base query sequence to determine percent identity. The deletions occur at the 5' end of the subject sequence and therefore, the FASTDB alignment does not show a matched/ alignment of the first 10 bases at 5' end. The 10 unpaired 40 bases represent 10% of the sequence (number of bases at the 5' and 3' ends not matched/total number of bases in the query sequence) so 10% is subtracted from the percent identity score calculated by the FASTDB program. If the remaining 90 bases were perfectly matched the final percent identity 45 would be 90%. In another example, a 90 base subject sequence is compared with a 100 base query sequence. This time the deletions are internal deletions so that there are no bases on the 5' or 3' of the subject sequence which are not matched/aligned with the query. In this case the percent 50 identity calculated by FASTDB is not manually corrected. Once again, only bases 5' and 3' of the subject sequence which are not matched/aligned with the query sequence are manually corrected for. No other manual corrections are to be made for the purposes of the present invention.

Standard recombinant DNA and molecular cloning techniques are well known in the art and are described by Sambrook, J., Fritsch, E. F. and Maniatis, T., *Molecular Cloning: A Laboratory Manual*, Second Edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. 60 (1989) (hereinafter "Maniatis"); and by Silhavy, T. J., Bennan, M. L. and Enquist, L. W., *Experiments with Gene Fusions*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1984); and by Ausubel, F. M. et al., *Current Protocols in Molecular Biology*, published by Greene Pub-65 lishing Assoc. and Wiley-Interscience (1987). Additional methods used here are in *Methods in Enzymology*, Volume

194, Guide to Yeast Genetics and Molecular and Cell Biology (Part A, 2004, Christine Guthrie and Gerald R. Fink (Eds.), Elsevier Academic Press, San Diego, Calif.). Other molecular tools and techniques are known in the art and include splicing by overlapping extension polymerase chain reaction (PCR) (Yu, et al. (2004) Fungal Genet. Biol. 41:973-981), positive selection for mutations at the URA3 locus of Saccharomyces cerevisiae (Boeke, J. D. et al. (1984) Mol. Gen. Genet. 197, 345-346; MA Romanos, et al. Nucleic Acids Res. 1991 Jan. 11; 19(1): 187), the cre-lox site-specific recombination system as well as mutant lox sites and FLP substrate mutations (Sauer, B. (1987) Mol Cell Biol 7: 2087-2096; Senecoff, et al. (1988) Journal of Molecular Biology, Volume 201, Issue 2, Pages 405-421; Albert, et al. (1995) The Plant Journal. Volume 7, Issue 4, pages 649-659), "seamless" gene deletion (Akada, et al. (2006) Yeast; 23(5):399-405), and gap repair methodology (Ma et al., Genetics 58:201-216; 1981).

The genetic manipulations of a recombinant host cell disclosed herein can be performed using standard genetic techniques and screening and can be made in any host cell that is suitable to genetic manipulation (*Methods in Yeast Genetics*, 2005, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., pp. 201-202). Construction of butanologens is described herein and in the art, for example in in PCT Pub. No. WO/2012/129555, incorporated herein by reference.

"qPCR" or "RT-PCR" is a PCT-based laboratory technique that simultaneously amplifies and quantifies a target gene.

**Biosynthetic Pathways** 

Biosynthetic pathways for the production of isobutanol that may be used include those described in U.S. Pat. Nos. 7,851,188 and 7,993,889, which are incorporated herein by 35 reference. Isobutanol pathways are referred to with their lettering in FIG. 1. In one embodiment, the isobutanol biosynthetic pathway comprises the following substrate to product conversions:

- a) pyruvate to acetolactate, which may be catalyzed, for example, by acetolactate synthase;
- b) acetolactate to 2,3-dihydroxyisovalerate, which may be catalyzed, for example, by ketol-acid reductoisomerase;
- c) 2,3-dihydroxyisovalerate to α-ketoisovalerate, which may be catalyzed, for example, by dihydroxyacid dehydratase;
- d)  $\alpha$ -ketoisovalerate to isobutyraldehyde, which may be catalyzed, for example, by a branched-chain keto acid decarboxylase; and,
- e) isobutyraldehyde to isobutanol, which may be catalyzed, for example, by a branched-chain alcohol dehydrogenase.

In another embodiment, the isobutanol biosynthetic pathway comprises the following substrate to product conver-55 sions:

- a) pyruvate to acetolactate, which may be catalyzed, for example, by acetolactate synthase;
- b) acetolactate to 2,3-dihydroxyisovalerate, which may be catalyzed, for example, by ketol-acid reductoisomerase;
- c) 2,3-dihydroxyisovalerate to α-ketoisovalerate, which may be catalyzed, for example, by dihydroxyacid dehydratase;
- h) α-ketoisovalerate to valine, which may be catalyzed, for example, by transaminase or valine dehydrogenase;
- i) valine to isobutylamine, which may be catalyzed, for example, by valine decarboxylase;

10

- j) isobutylamine to isobutyraldehyde, which may be catalyzed by, for example, omega transaminase; and,
- e) isobutyraldehyde to isobutanol, which may be catalyzed, for example, by a branched-chain alcohol dehydrogenase.

In another embodiment, the isobutanol biosynthetic pathway comprises the following substrate to product conversions:

- a) pyruvate to acetolactate, which may be catalyzed, for example, by acetolactate synthase;
- b) acetolactate to 2,3-dihydroxyisovalerate, which may be catalyzed, for example, by ketol-acid reductoisomerase;
- c) 2,3-dihydroxyisovalerate to  $\alpha$ -ketoisovalerate, which may be catalyzed, for example, by dihydroxyacid dehydratase;
- f)  $\alpha$ -ketoisovalerate to isobutyryl-CoA, which may be catalyzed, for example, by branched-chain keto acid dehydrogenase;
- g) isobutyryl-CoA to isobutyraldehyde, which may be catalyzed, for example, by acetylating aldehyde dehydrogenase; and,
- e) isobutyraldehyde to isobutanol, which may be catalyzed, for example, by a branched-chain alcohol dehy- 25 drogenase.

In another embodiment, the isobutanol biosynthetic pathway comprises the substrate to product conversions shown as steps k, g, and e in FIG. **1**.

Biosynthetic pathways for the production of 1-butanol 30 that may be used include those described in U.S. Appl. Pub. No. 2008/0182308, which is incorporated herein by reference. In one embodiment, the 1-butanol biosynthetic pathway comprises the following substrate to product conversions: 35

- a) acetyl-CoA to acetoacetyl-CoA, which may be catalyzed, for example, by acetyl-CoA acetyl transferase;
- b) acetoacetyl-CoA to 3-hydroxybutyryl-CoA, which may be catalyzed, for example, by 3-hydroxybutyryl-CoA dehydrogenase; 40
- c) 3-hydroxybutyryl-CoA to crotonyl-CoA, which may be catalyzed, for example, by crotonase;
- d) crotonyl-CoA to butyryl-CoA, which may be catalyzed, for example, by butyryl-CoA dehydrogenase;
- e) butyryl-CoA to butyraldehyde, which may be cata- 45 lyzed, for example, by butyraldehyde dehydrogenase; and,
- f) butyraldehyde to 1-butanol, which may be catalyzed, for example, by butanol dehydrogenase.

Biosynthetic pathways for the production of 2-butanol 50 that may be used include those described in U.S. Appl. Pub. No. 2007/0259410 and U.S. Appl. Pub. No. 2009/0155870, which are incorporated herein by reference. In one embodiment, the 2-butanol biosynthetic pathway comprises the following substrate to product conversions: 55

- a) pyruvate to alpha-acetolactate, which may be catalyzed, for example, by acetolactate synthase;
- b) alpha-acetolactate to acetoin, which may be catalyzed, for example, by acetolactate decarboxylase;
- c) acetoin to 3-amino-2-butanol, which may be catalyzed, 60 for example, acetoin aminase;
- d) 3-amino-2-butanol to 3-amino-2-butanol phosphate, which may be catalyzed, for example, by aminobutanol kinase;
- e) 3-amino-2-butanol phosphate to 2-butanone, which 65 may be catalyzed, for example, by aminobutanol phosphate phosphorylase; and,

f) 2-butanone to 2-butanol, which may be catalyzed, for example, by butanol dehydrogenase.

In another embodiment, the 2-butanol biosynthetic pathway comprises the following substrate to product conversions:

- a) pyruvate to alpha-acetolactate, which may be catalyzed, for example, by acetolactate synthase;
- b) alpha-acetolactate to acetoin, which may be catalyzed, for example, by acetolactate decarboxylase;
- c) acetoin to 2,3-butanediol, which may be catalyzed, for example, by butanediol dehydrogenase;
- d) 2,3-butanediol to 2-butanone, which may be catalyzed, for example, by dial dehydratase; and,
- e) 2-butanone to 2-butanol, which may be catalyzed, for example, by butanol dehydrogenase.

Biosynthetic pathways for the production of 2-butanone that may be used include those described in U.S. Appl. Pub. No. 2007/0259410 and U.S. Appl. Pub. No. 2009/0155870, which are incorporated herein by reference. In one embodi-20 ment, the 2-butanone biosynthetic pathway comprises the

following substrate to product conversions:

- a) pyruvate to alpha-acetolactate, which may be catalyzed, for example, by acetolactate synthase;
- b) alpha-acetolactate to acetoin, which may be catalyzed, for example, by acetolactate decarboxylase;
- c) acetoin to 3-amino-2-butanol, which may be catalyzed, for example, acetoin aminase;
- d) 3-amino-2-butanol to 3-amino-2-butanol phosphate, which may be catalyzed, for example, by aminobutanol kinase; and,
- e) 3-amino-2-butanol phosphate to 2-butanone, which may be catalyzed, for example, by aminobutanol phosphate phosphorylase.

In another embodiment, the 2-butanone biosynthetic path-35 way comprises the following substrate to product conversions:

- a) pyruvate to alpha-acetolactate, which may be catalyzed, for example, by acetolactate synthase;
- b) alpha-acetolactate to acetoin which may be catalyzed, for example, by acetolactate decarboxylase;
- c) acetoin to 2,3-butanediol, which may be catalyzed, for example, by butanediol dehydrogenase;
- d) 2,3-butanediol to 2-butanone, which may be catalyzed, for example, by dial dehydratase.

In one embodiment, the invention produces butanol from plant derived carbon sources, avoiding the negative environmental impact associated with standard petrochemical processes for butanol production. In one embodiment, the invention provides a method for the production of butanol using recombinant industrial host cells comprising a butanol pathway.

In some embodiments, the isobutanol biosynthetic pathway comprises at least one polynucleotide, at least two polynucleotides, at least three polynucleotides, or at least four polynucleotides that is/are heterologous to the host cell. In embodiments, each substrate to product conversion of an isobutanol biosynthetic pathway in a recombinant host cell is catalyzed by a heterologous polypeptide. In embodiments, the polypeptide catalyzing the substrate to product conversions of acetolactate to 2,3-dihydroxyisovalerate and/or the polypeptide catalyzing the substrate to product conversion of isobutyraldehyde to isobutanol are capable of utilizing NADH as a cofactor.

The terms "acetohydroxyacid synthase," "acetolactate synthase" and "acetolactate synthetase" (abbreviated "ALS") are used interchangeably herein to refer to an enzyme that catalyzes the conversion of pyruvate to acetolactate and CO<sub>2</sub>. Example acetolactate synthases are known by the EC number 2.2.1.6 (Enzyme Nomenclature 1992, Academic Press, San Diego). These unmodified enzymes are available from a number of sources, including, but not limited to, *Bacillus subtilis* (GenBank Nos: CAB07802.1, <sup>5</sup> Z99122 (SEQ ID NO:16), NCBI (National Center for Biotechnology Information)), CAB15618), *Klebsiella pneumoniae* (GenBank Nos: AAA25079, M73842, *Lactococcus lactis* (GenBank Nos: AAA25161, L16975), *S. cerevisiae* (SEQ ID NOs:130 and 131), *E. coli* K12 (SEQ ID NOs:132<sup>10</sup> and 133).

The term "ketol-acid reductoisomerase" ("KARI"), "acetohydroxy acid reductoisomerase" and "acetohydroxy acid isomeroreductase" will be used interchangeably and refer to enzymes capable of catalyzing the reaction of (S)-acetolactate to 2,3-dihydroxyisovalerate. Example KARI enzymes may be classified as EC number EC 1.1.1.86 (Enzyme Nomenclature 1992, Academic Press, San Diego), and are available from a vast array of microorganisms, 20 including, but not limited to, Escherichia coli (GenBank Nos: NP\_418222, NC\_000913), Saccharomyces cerevisiae (GenBank Nos: NP\_013459, NC\_001144), Methanococcus maripaludis (GenBank Nos: CAF30210, BX957220), and Bacillus subtilis (GenBank Nos: CAB14789, Z99118). 25 KARIs include Anaerostipes caccae KARI variants "K9G9," "K9D3" and "K9JB4P" (SEQ ID NOs:37, 38, and 182 respectively). Ketol-acid reductoisomerase (KARI) enzymes are described in U.S. Patent Appl. Pub. Nos. 20080261230 A1, 20090163376 A1, 20100197519 A1, PCT 30 Appl. Pub. Nos. WO/2011/041415, and WO/2012/129555, which are incorporated herein by reference. Examples of KARIs disclosed therein are those from Lactococcus lactis, Vibrio cholera, Pseudomonas aeruginosa PAO1, and Pseudomonas fluorescens PF5 mutants. Pseudomonas fluo- 35 rescens KARIs include SEQ ID NO:134. In some embodiments, the KARI utilizes NADH. In some embodiments, the KARI utilizes NADPH. In some embodiments, the KARI utilizes NADH or NADPH.

The term "acetohydroxy acid dehydratase" and "dihy- 40 droxyacid dehydratase" ("DHAD") refers to an enzyme that catalyzes the conversion of 2,3-dihydroxyisovalerate to  $\alpha$ -ketoisovalerate. Example acetohydroxy acid dehydratases are known by the EC number 4.2.1.9. Such enzymes are available from a vast array of microorganisms, including, 45 but not limited to, E. coli (GenBank Nos: YP\_026248, NC 000913), S. cerevisiae (GenBank Nos: NP 012550, NC 001142), M. maripaludis (GenBank Nos: CAF29874, BX957219), B. subtilis (GenBank Nos: CAB14105, Z99115), L. lactis (SEQ ID NO:108), and N. crassa. US 50 Appl. Pub. No. 20100081154 A1, and U.S. Pat. No. 7,851, 188, which are incorporated herein by reference, describe dihydroxyacid dehydratases (DHADs), including a DHAD from Streptococcus mutans (SEQ ID NO:135). Example DHADs include variants of S. mutans DHAD, for example 55 "L2V4" (SEQ ID NO:183).

The term "branched-chain  $\alpha$ -keto acid decarboxylase" or " $\alpha$ -ketoacid decarboxylase" or " $\alpha$ -ketoisovalerate decarboxylase" or "2-ketoisovalerate decarboxylase" ("KIVD") refers to an enzyme that catalyzes the conversion of  $\alpha$ -ketoisovalerate to isobutyraldehyde and CO<sub>2</sub>. Example branched-chain  $\alpha$ -keto acid decarboxylases are known by the EC number 4.1.1.72 and are available from a number of sources, including, but not limited to, *Lactococcus lactis* (GenBank Nos: AAS49166 (SEQ ID NO:141), AY548760; 65 CAG34226, AJ746364, *Salmonella typhimurium* (GenBank Nos: NP\_461346, NC\_003197), *Clostridium acetobutyli*-

*cum* (GenBank Nos: NP\_149189, NC\_001988), *M. caseolyticus* (SEQ ID NOs:118, 137), and *L. grayi* (SEQ ID NO:136).

The term "branched-chain alcohol dehydrogenase" ("ADH") refers to an enzyme that catalyzes the conversion of isobutyraldehyde to isobutanol. Example branched-chain alcohol dehydrogenases are known by the EC number 1.1.1.265, but may also be classified under other alcohol dehydrogenases (specifically, EC 1.1.1.1 or 1.1.1.2). Alcohol dehydrogenases may be NADPH dependent or NADH dependent. Such enzymes are available from a number of sources, including, but not limited to, S. cerevisiae (Gen-Bank Nos: NP\_010656, NC\_001136; NP\_014051, NC\_001145), E. coli (GenBank Nos: NP\_417484, NC 000913),  $C_{\cdot}$ acetobutvlicum (GenBank Nos: NP\_349892, NC\_003030; NP\_349891, NC\_003030). U.S. Pat. No. 8,188,250, which is incorporated herein by reference, describes SadB, an alcohol dehydrogenase (ADH) from Achromobacter xylosoxidans (SEQ ID NO:139). Alcohol dehvdrogenases also include horse liver ADH (SEO ID NO:142) and Beijerinkia indica ADH (SEQ ID NO:138) (as described by U.S. Appl. Publ. No. 20110269199, which is incorporated herein by reference).

The term "butanol dehydrogenase" refers to a polypeptide (or polypeptides) having an enzyme activity that catalyzes the conversion of isobutyraldehyde to isobutanol or the conversion of 2-butanone and 2-butanol. Butanol dehydrogenases are a subset of a broad family of alcohol dehydrogenases. Butanol dehydrogenase may be NAD- or NADPdependent. The NAD-dependent enzymes are known as EC 1.1.1.1 and are available, for example, from Rhodococcus ruber (GenBank Nos: CAD36475, AJ491307). The NADP dependent enzymes are known as EC 1.1.1.2 and are available, for example, from Pyrococcus furiosus (GenBank Nos: AAC25556, AF013169). Additionally, a butanol dehydrogenase is available from Escherichia coli (GenBank Nos: NP\_417484, NC\_000913) and a cyclohexanol dehydrogenase is available from Acinetobacter sp. (GenBank Nos: AAG10026, AF282240). The term "butanol dehydrogenase" also refers to an enzyme that catalyzes the conversion of butyraldehyde to 1-butanol, using either NADH or NADPH as cofactor. Butanol dehydrogenases are available from, for example, C. acetobutylicum (GenBank NOs: NP\_149325, NC\_001988; note: this enzyme possesses both aldehyde and alcohol dehydrogenase activity); NP\_349891, NC\_003030; and NP\_349892, NC\_003030), E. coli (GenBank NOs: NP 417-484, NC 000913), and A. xvlosoxidans (SEO ID NOs:47 and 48, as described in U.S. Pat. No. 8,188,250, which is incorporated herein by reference in its entirety.

The term "branched-chain keto acid dehydrogenase" refers to an enzyme that catalyzes the conversion of  $\alpha$ -ketoisovalerate to isobutyryl-CoA (isobutyryl-coenzyme A), typically using NAD<sup>+</sup> (nicotinamide adenine dinucleotide) as an electron acceptor. Example branched-chain keto acid dehydrogenases are known by the EC number 1.2.4.4. Such branched-chain keto acid dehydrogenases are comprised of four subunits and sequences from all subunits are available from a vast array of microorganisms, including, but not limited to, *B. subtilis* (GenBank Nos: CAB14336, Z99116; CAB14335, Z99116; CAB14334, Z99116; and CAB14337, Z99116) and *Pseudomonas putida* (GenBank Nos: AAA65614, M57613; AAA65615, M57613; AAA65617, M57613; and AAA65618, M57613).

The term "acylating aldehyde dehydrogenase" refers to an enzyme that catalyzes the conversion of isobutyryl-CoA to isobutyraldehyde, typically using either NADH or NADPH as an electron donor. Example acylating aldehyde dehydro-

genases are known by the EC numbers 1.2.1.10 and 1.2.1.57. Such enzymes are available from multiple sources, including, but not limited to, Clostridium beijerinckii (GenBank Nos: AAD31841, AF157306), C. acetobutylicum (GenBank Nos: NP\_149325, NC\_001988; NP\_149199, NC\_001988), 5 P. putida (GenBank Nos: AAA89106, U13232), and Ther-(GenBank Nos: YP\_145486, mus thermophilus NC 006461).

The term "transaminase" refers to an enzyme that catalyzes the conversion of  $\alpha$ -ketoisovalerate to L-valine, using 10 either alanine or glutamate as an amine donor. Example transaminases are known by the EC numbers 2.6.1.42 and 2.6.1.66. Such enzymes are available from a number of sources. Examples of sources for alanine-dependent enzymes include, but are not limited to, E. coli (GenBank 15 Nos: YP\_026231, NC\_000913) and Bacillus licheniformis (GenBank Nos: YP\_093743, NC\_006322). Examples of sources for glutamate-dependent enzymes include, but are not limited to, E. coli (GenBank Nos: YP\_026247, NC 000913), S. cerevisiae (GenBank Nos: NP 012682, 20 NC\_001142) and Methanobacterium thermoautotrophicum (GenBank Nos: NP\_276546, NC\_000916).

The term "valine dehydrogenase" refers to an enzyme that catalyzes the conversion of  $\alpha$ -ketoisovalerate to L-valine, typically using NAD(P)H as an electron donor and ammonia 25 as an amine donor. Example valine dehydrogenases are known by the EC numbers 1.4.1.8 and 1.4.1.9 and such enzymes are available from a number of sources, including, but not limited to, Streptomyces coelicolor (GenBank Nos: NP\_628270, NC\_003888) and *B. subtilis* (GenBank Nos: 30 CAB14339, Z99116).

The term "valine decarboxylase" refers to an enzyme that catalyzes the conversion of L-valine to isobutylamine and CO<sub>2</sub>. Example valine decarboxylases are known by the EC number 4.1.1.14. Such enzymes are found in Streptomyces, 35 such as for example, Streptomyces viridifaciens (GenBank Nos: AAN10242, AY116644).

The term "omega transaminase" refers to an enzyme that catalyzes the conversion of isobutylamine to isobutyraldehyde using a suitable amino acid as an amine donor. 40 Bank NOs: CAA22721, AL939127). Example omega transaminases are known by the EC number 2.6.1.18 and are available from a number of sources, including, but not limited to, Alcaligenes denitrificans (AAP92672, AY330220), Ralstonia eutropha (GenBank Nos: YP\_294474, NC\_007347), Shewanella oneidensis 45 (GenBank Nos: NP\_719046, NC\_004347), and P. putida (GenBank Nos: AAN66223, AE016776).

The term "acetyl-CoA acetyltransferase" refers to an enzyme that catalyzes the conversion of two molecules of acetyl-CoA to acetoacetyl-CoA and coenzyme A (CoA). 50 Example acetyl-CoA acetyltransferases are acetyl-CoA acetyltransferases with substrate preferences (reaction in the forward direction) for a short chain acyl-CoA and acetyl-CoA and are classified as E.C. 2.3.1.9 [Enzyme Nomenclature 1992, Academic Press, San Diego]; although, enzymes 55 with a broader substrate range (E.C. 2.3.1.16) will be functional as well. Acetyl-CoA acetyltransferases are available from a number of sources, for example, Escherichia coli (GenBank Nos: NP\_416728, NC\_000913; NCBI (National Center for Biotechnology Information) amino acid 60 sequence, NCBI nucleotide sequence), Clostridium acetobutylicum (GenBank Nos: NP\_349476.1, NC\_003030; NP\_149242, NC\_001988, Bacillus subtilis (GenBank Nos: NP 390297, NC 000964), and Saccharomyces cerevisiae (GenBank Nos: NP 015297, NC 001148). 65

The term "3-hydroxybutyryl-CoA dehydrogenase" refers to an enzyme that catalyzes the conversion of acetoacetylCoA to 3-hydroxybutyryl-CoA. Example hydroxybutyryl-CoA dehydrogenases may be reduced nicotinamide adenine dinucleotide (NADH)-dependent, with a substrate preference for (S)-3-hydroxybutyryl-CoA or (R)-3-hydroxybutyryl-CoA. Examples may be classified as E.C. 1.1.1.35 and E.C. 1.1.1.30, respectively. Additionally, 3-hydroxybutyryl-CoA dehydrogenases may be reduced nicotinamide adenine dinucleotide phosphate (NADPH)-dependent, with a substrate preference for (S)-3-hydroxybutyryl-CoA or (R)-3hydroxybutyryl-CoA and are classified as E.C. 1.1.1.157 and E.C. 1.1.1.36, respectively. 3-Hydroxybutyryl-CoA dehydrogenases are available from a number of sources, for example, C. acetobutylicum (GenBank NOs: NP\_349314, NC\_003030), B. subtilis (GenBank NOs: AAB09614, U29084), Ralstonia eutropha (GenBank NOs: YP 294481, NC\_007347), and Alcaligenes eutrophus (GenBank NOs: AAA21973, J04987).

The term "crotonase" refers to an enzyme that catalyzes the conversion of 3-hydroxybutyryl-CoA to crotonyl-CoA and H<sub>2</sub>O. Example crotonases may have a substrate preference for (S)-3-hydroxybutyryl-CoA or (R)-3-hydroxybutyryl-CoA and may be classified as E.C. 4.2.1.17 and E.C. 4.2.1.55, respectively. Crotonases are available from a number of sources, for example, E. coli (GenBank NOs: NP\_415911, NC\_000913), C. acetobutylicum (GenBank NOs: NP\_349318, NC\_003030), B. subtilis (GenBank NOs: CAB13705, Z99113), and Aeromonas caviae (GenBank NOs: BAA21816, D88825).

The term "butyryl-CoA dehydrogenase" refers to an enzyme that catalyzes the conversion of crotonyl-CoA to butyryl-CoA. Example butyryl-CoA dehydrogenases may be NADH-dependent, NADPH-dependent, or flavin-dependent and may be classified as E.C. 1.3.1.44, E.C. 1.3.1.38, and E.C. 1.3.99.2, respectively. Butyryl-CoA dehydrogenases are available from a number of sources, for example, acetobutylicum (GenBank NOs: NP\_347102, С. NC 003030), Euglena gracilis (GenBank NOs: Q5EU90), AY741582), Streptomyces collinus (GenBank NOs: AAA92890, U37135), and Streptomyces coelicolor (Gen-

The term "butyraldehyde dehydrogenase" refers to an enzyme that catalyzes the conversion of butyryl-CoA to butyraldehyde, using NADH or NADPH as cofactor. Butyraldehyde dehydrogenases with a preference for NADH are known as E.C. 1.2.1.57 and are available from, for example, Clostridium beijerinckii (GenBank NOs: AAD31841, AF157306) and C. acetobutylicum (GenBank NOs: NP\_149325, NC\_001988).

The term "isobutyryl-CoA mutase" refers to an enzyme that catalyzes the conversion of butyryl-CoA to isobutyryl-CoA. This enzyme uses coenzyme  $B_{12}$  as cofactor. Example isobutyryl-CoA mutases are known by the EC number 5.4.99.13. These enzymes are found in a number of Streptomyces, including, but not limited to, Streptomyces cinnamonensis (GenBank Nos: AAC08713, U67612; CAB59633, AJ246005), S. coelicolor (GenBank Nos: CAB70645, AL939123; CAB92663, AL939121), and Streptomyces avermitilis (GenBank Nos: NP 824008, NC 003155; NP\_824637, NC\_003155).

The term "acetolactate decarboxylase" refers to a polypeptide (or polypeptides) having an enzyme activity that catalyzes the conversion of alpha-acetolactate to acetoin. Example acetolactate decarboxylases are known as EC 4.1.1.5 and are available, for example, from Bacillus subtilis (GenBank Nos: AAA22223, L04470), Klebsiella terrigena (GenBank Nos: AAA25054, L04507) and Klebsiella pneumoniae (GenBank Nos: AAU43774, AY722056).
The term "acetoin aminase" or "acetoin transaminase" refers to a polypeptide (or polypeptides) having an enzyme activity that catalyzes the conversion of acetoin to 3-amino-2-butanol. Acetoin aminase may utilize the cofactor pyridoxal 5'-phosphate or NADH (reduced nicotinamide 5 adenine dinucleotide) or NADPH (reduced nicotinamide adenine dinucleotide phosphate). The resulting product may have (R) or (S) stereochemistry at the 3-position. The pyridoxal phosphate-dependent enzyme may use an amino acid such as alanine or glutamate as the amino donor. The 10 NADH- and NADPH-dependent enzymes may use ammonia as a second substrate. A suitable example of an NADH dependent acetoin aminase, also known as amino alcohol dehydrogenase, is described by Ito et al. (U.S. Pat. No. 6,432,688). An example of a pyridoxal-dependent acetoin 15 aminase is the amine:pyruvate aminotransferase (also called amine:pyruvate transaminase) described by Shin and Kim (J. Org. Chem. 67:2848-2853 (2002)).

The term "acetoin kinase" refers to a polypeptide (or polypeptides) having an enzyme activity that catalyzes the 20 conversion of acetoin to phosphoacetoin. Acetoin kinase may utilize ATP (adenosine triphosphate) or phosphoenolpyruvate as the phosphate donor in the reaction. Enzymes that catalyze the analogous reaction on the similar substrate dihydroxyacetone, for example, include enzymes 25 known as EC 2.7.1.29 (Garcia-Alles et al. (2004) Biochemistry 43:13037-13046).

The term "acetoin phosphate aminase" refers to a polypeptide (or polypeptides) having an enzyme activity that catalyzes the conversion of phosphoacetoin to 3-amino-2-butanol O-phosphate. Acetoin phosphate aminase may use the cofactor pyridoxal 5'-phosphate, NADH or NADPH. The resulting product may have (R) or (S) stereochemistry at the 3-position. The pyridoxal phosphate-dependent enzyme may use an amino acid such as alanine or glutamate. The NADH and NADPH-dependent enzymes may use ammonia as a second substrate. Although there are no reports of enzymes catalyzing this reaction on phosphoacetoin, there is a pyridoxal phosphate-dependent enzyme that is proposed to carry out the analogous reaction on the similar substrate serinol phosphate (Yasuta et al. (2001) Appl. Environ. Microbial. 67:4999-5009.

The term "aminobutanol phosphate phospholyase", also called "amino alcohol O-phosphate lyase", refers to a polypeptide (or polypeptides) having an enzyme activity that 45 catalyzes the conversion of 3-amino-2-butanol O-phosphate to 2-butanone. Amino butanol phosphate phospho-lyase may utilize the cofactor pyridoxal 5'-phosphate. There are reports of enzymes that catalyze the analogous reaction on the similar substrate 1-amino-2-propanol phosphate (Jones et al. 50 (1973) Biochem J. 134:167-182). U.S. Appl. Pub. No. 2007/0259410 describes an aminobutanol phosphate phospho-lyase from the organism *Erwinia carotovora*.

The term "aminobutanol kinase" refers to a polypeptide (or polypeptides) having an enzyme activity that catalyzes 55 the conversion of 3-amino-2-butanol to 3-amino-2butanol O-phosphate. Amino butanol kinase may utilize ATP as the phosphate donor. Although there are no reports of enzymes catalyzing this reaction on 3-amino-2-butanol, there are reports of enzymes that catalyze the analogous reaction on 60 the similar substrates ethanolamine and 1-amino-2-propanol (Jones et al., supra). U.S. Appl. Pub. No. 2009/0155870 describes, in Example 14, an amino alcohol kinase of *Erwinia carotovora* subsp. *Atroseptica*.

The term "butanediol dehydrogenase" also known as 65 "acetoin reductase" refers to a polypeptide (or polypeptides) having an enzyme activity that catalyzes the conversion of

acetoin to 2,3-butanediol. Butanediol dehydrogenases are a subset of the broad family of alcohol dehydrogenases. Butanediol dehydrogenase enzymes may have specificity for production of (R)- or (S)-stereochemistry in the alcohol product. (S)-specific butanediol dehydrogenases are known as EC 1.1.1.76 and are available, for example, from *Klebsiella pneumoniae* (GenBank Nos: BBA13085, D86412). (R)-specific butanediol dehydrogenases are known as EC 1.1.1.4 and are available, for example, from *Bacillus cereus* (GenBank Nos. NP 830481, NC\_004722; AAP07682, AE017000), and *Lactococcus lactis* (GenBank Nos. AAK04995, AE006323).

The term "butanediol dehydratase", also known as "dial dehydratase" or "propanediol dehydratase" refers to a polypeptide (or polypeptides) having an enzyme activity that catalyzes the conversion of 2,3-butanediol to 2-butanone. Butanediol dehydratase may utilize the cofactor adenosyl cobalamin (also known as coenzyme Bw or vitamin B12; although vitamin B12 may refer also to other forms of cobalamin that are not coenzyme B12). Adenosyl cobalamin-dependent enzymes are known as EC 4.2.1.28 and are available, for example, from Klebsiella oxytoca (GenBank Nos: AA08099 (alpha subunit), D45071; BAA08100 (beta subunit), D45071; and BBA08101 (gamma subunit), D45071 (Note all three subunits are required for activity), and Klebsiella pneumonia (GenBank Nos: AAC98384 (alpha subunit), AF102064; GenBank Nos: AAC98385 (beta subunit), AF102064, GenBank Nos: AAC98386 (gamma subunit), AF102064). Other suitable dial dehydratases include, but are not limited to, B12-dependent dial dehydratases available from Salmonella typhimurium (GenBank Nos: AAB84102 (large subunit), AF026270; GenBank Nos: AAB84103 (medium subunit), AF026270; GenBank Nos: AAB84104 (small subunit), AF026270); and Lactobacillus AJ297723; GenBank Nos: CAC82542 (medium subunit); AJ297723; GenBank Nos: CAD01091 (small subunit), AJ297723); and enzymes from Lactobacillus brevis (particularly strains CNRZ 734 and CNRZ 735, Speranza et al., J. Agric. Food Chem. (1997) 45:3476-3480), and nucleotide sequences that encode the corresponding enzymes. Methods of dial dehydratase gene isolation are well known in the art (e.g., U.S. Pat. No. 5,686,276).

The term "pyruvate decarboxylase" refers to an enzyme that catalyzes the decarboxylation of pyruvic acid to acetaldehyde and carbon dioxide. Pyruvate decarboxylases are known by the EC number 4.1.1.1. These enzymes are found in a number of yeast, including *Saccharomyces cerevisiae* (GenBank Nos: CAA97575 (SEQ ID NO:1), CAA97705 (SEQ ID NO:2), CAA97091 (SEQ ID NO:3)).

The term "ethanol dehydrogenase" or "alcohol dehydrogenase" refers to an enzyme that catalyze the interconversion between aldehydes or ketones and alcohols, frequently using either NADH and/or NADPH as cofactors. Ethanol dehydrogenases comprise the EC numbers 1.1.1.1, 1.1.99.8., 1.1.1.244., 1.1.2.B1., 1.1.2.B2., 1.1.2.B3.

It will be appreciated that host cells comprising an isobutanol biosynthetic pathway as provided herein may further comprise one or more additional modifications. U.S. Appl. Pub. No. 20090305363 (incorporated by reference) discloses increased conversion of pyruvate to acetolactate by engineering yeast for expression of a cytosol-localized acetolactate synthase and substantial elimination of pyruvate decarboxylase activity. In some embodiments, the host cells comprise modifications to reduce glycerol-3-phosphate dehydrogenase activity and/or disruption in at least one gene encoding a polypeptide having pyruvate decarboxylase activity or a disruption in at least one gene encoding a regulatory element controlling pyruvate decarboxylase gene expression as described in U.S. Patent Appl. Pub. No. 20090305363 (incorporated herein by reference), modifications to a host cell that provide for increased carbon flux 5 through an Entner-Doudoroff Pathway or reducing equivalents balance as described in U.S. Patent Appl. Pub. No. 20100120105 (incorporated herein by reference). Other modifications are described in PCT Pub. No. WO/2012/ 129555, incorporated herein by reference, and include inte- 10 gration of at least one polynucleotide encoding a polypeptide that catalyzes a step in a pyruvate-utilizing biosynthetic pathway. Other modifications include at least one deletion, mutation, and/or substitution in an endogenous polynucleotide encoding a polypeptide having acetolactate reductase 15 activity. In embodiments, the polypeptide having acetolactate reductase activity is YMR226C (SEQ ID NOs:4, 5) of Saccharomyces cerevisiae or a homolog thereof. Additional modifications include a deletion, mutation, and/or substitution in an endogenous polynucleotide encoding a polypep- 20 tide having aldehyde dehydrogenase and/or aldehyde oxidase activity. In embodiments, the polypeptide having aldehyde dehydrogenase activity is ALD6 from Saccharomyces cerevisiae or a homolog thereof. A genetic modification which has the effect of reducing glucose repression 25 wherein the yeast production host cell is pdc- is described in U.S. Appl. Publication No. 20110124060, incorporated herein by reference. In some embodiments, the pyruvate decarboxylase that is deleted or downregulated is selected from the group consisting of: PDC1, PDC5, PDC6, and 30 combinations thereof. In some embodiments, the pyruvate decarboxylase is selected from those enzymes described in U.S. Patent Appl. Pub. No. 20090305363. In some embodiments, host cells contain a deletion or downregulation of a polynucleotide encoding a polypeptide that catalyzes the 35 conversion of glyceraldehyde-3-phosphate to glycerate 1,3, bisphosphate. In some embodiments, the enzyme that catalyzes this reaction is glyceraldehyde-3-phosphate dehydrogenase.

Yeasts may have one or more genes encoding pyruvate 40 decarboxylase. For example, there is one gene encoding pyruvate decarboxylase in *Candida glabrata* and *Schizosac-charomyces pombe*, while there are three isozymes of pyruvate decarboxylase encoded by the PDC1, PCD5, and PDC6 genes in *Saccharomyces*. In some embodiments, in the 45 present yeast cells at least one PDC gene is inactivated. If the yeast cell used has more than one expressed (active) PDC gene, then each of the active PDC genes may be modified or inactivated thereby producing a pdc– cell. For example, in *S. cerevisiae* the PDC1, PDC5, and PDC6 genes may be 50 modified or inactivated. If a PDC gene is not active under the fermentation conditions to be used then such a gene would not need to be modified or inactivated.

Other target genes, such as those encoding pyruvate decarboxylase proteins having at least 70-75%, at least 55 75-80%, at least 80-85%, at least 85%-90%, at least 90%-95%, or at least 96%, at least 97%, at least 98%, or at least 99% sequence identity to the pyruvate decarboxylases described in U.S. Patent Appl. Pub. No. 20090305363 may be identified in the literature and in bioinformatics databases 60 well known to the skilled person.

Recombinant host cells may further comprise (a) at least one heterologous polynucleotide encoding a polypeptide having dihydroxy-acid dehydratase activity; and (b)(i) at least one deletion, mutation, and/or substitution in an endog-65 enous gene encoding a polypeptide affecting Fe—S cluster biosynthesis; and/or (ii) at least one heterologous polynucle-

otide encoding a polypeptide affecting Fe—S cluster biosynthesis described in U.S. Patent Appl. Pub. No. US20120064561, incorporated herein by reference. In embodiments, the polypeptide affecting Fe—S cluster biosynthesis is encoded by AFT1, AFT2, FRA2, GRX3 or CCC1. AFT1 and AFT2 are described by WO/2001/103300, which is incorporated herein by reference. In embodiments, the polypeptide affecting Fe—S cluster biosynthesis is constitutive mutant AFT1 L99A, AFT1 L102A, AFT1 C291F, or AFT1 C293F.

In some embodiments, the host cell further comprises one or more polynucleotides that encode one or more enzymes having the following Enzyme Commission Numbers: EC 4.1.1.1 (PDC1, 5, and 6) (SEQ ID NOs:1, 2, and 3) and EC 1.1.1.1 (alcohol dehydrogenase).

In some embodiments of the invention, there are one or more inhibitors, antibiotics, or combinations thereof in the fermentation medium.

In some embodiments, the inhibitor is an inhibitor of an ethanol biosynthesis pathway. In some embodiments, the inhibitor inhibits pyruvate decarboxylase and/or alcohol dehydrogenase. In some embodiments, the inhibitor is selected from the group consisting of: the XC6H4CH=CHCOCOOH class of inhibitors/substrate analogues, cinnamaldehydes, glyoxalic acid, ketomalonate, regulatory site inhibitors, p-chloromercuribenzoic acid (pCMB), 5,5'-dithiobis(2-nitrobenzoic acid) (DNTB), pyrazole, 4-pyrazolecarboxylic acid, 1-H -pyrazole-1-carboxamidine-HC1, 4-methylpyrazole, 1-bromo-2-butanone, pyrazole-3,5-dicarboxylic acid monohydrate, and mixtures thereof. In some embodiments, the XC<sub>6</sub>H<sub>4</sub>CH=CHCOCOOH inhibitors/substrate analogue is CPB((E)-4-(4-chlorophenyl)-2-oxo-3-butenoic acid. In some embodiments the cinnamaldehyde is p-nitrocinnamaldehyde (NA). In some embodiments, the regulatory site inhibitors are iodoacetate, 1,3-dibromoacetone, 1-bromo-2butanone. "Cinnamaldehyde" includes both trans-cinnamaldehydes and 4-nitrocinnamaldehydes. In some embodiments, copper (II) is added at a concentration of at least about 1.1 mM, at least about 11 mM, at least about 33 mM. In some embodiments, sulfometuron methyl is added at a concentration of at least about 0.001 mM, at least about 0.01 mM, at least about 0.1 mM. In some embodiments, sulfite is added at a concentration of at least about 0.6 mM, at least about 6.2 mM, at least about 62 mM. In some embodiments, formaldehyde is added at a concentration of at least about 0.09 mM, at least about 0.9 mM, at least about 2.7 mM. In some embodiments, pyrazole is added at a concentration of at least about 0.3 mM, at least about 3 mM, at least about 30 mM. In some embodiments, 4-methylpyrazole hydrochloride is added at a concentration of at least about 4.1 mM, at least about 41mM, at least about 123 mM. In some embodiments, 4-pyrazolecarboxylic acid is added at a concentration of at least about 10 mM, at least about 100 mM, at least about 300 mM. In some embodiments, 1-bromo-2-butanone is added at a concentration of at least about 0.0002 mM, at least about 0.002 mM, at least about 0.006 mM. In some embodiments, trans-cinnamaldehyde is added at a concentration of at least about 0.025 mM, at least about 0.25 mM, at least about 0.75 mM. In some embodiments, glyoxylic acid is added at a concentration of at least about 16.8 mM, at least about 168 mM, at least about 504 mM.

In some embodiments, the inhibitor is a chemical. In some embodiments, the chemical is selected from the group consisting of: fluoroacetate (dehH1), fluorophenyalanine, formaldehyde (SFA1), sulfite (FZF1-4), and trifluoroleucine (LEU4-1).

In some embodiments, the inhibitor is an inhibitor of an amino acid biosynthesis pathway. In some embodiments, the inhibitor is an acetohydroxy acid synthase (AHAS) inhibitor. In some embodiments, the inhibitor is a sulfonvlurea herbicide. In some embodiments, the sulfonylurea herbicide is selected from the group consisting of: imidazolinones, triazolopyrimidines, pyrimidinyl oxybenzoates, sulfonylureas, sulfonylamino carbonyl triazolinones, and mixtures thereof. In some embodiments, the inhibitor is selected from the group consisting of: nicosulfuron methyl, metsulfuron methyl, chlorimuron ethyl, sulfometuron methyl, chlorsulfuron, thifensulfuron methyl, and mixtures thereof. In some embodiments, the sulfonylurea herbicide is an acetohydroxyacid synthase (AHAS) inhibitor).

In some embodiments, resistance to the sulfonyl urea is conferred by a polypeptide encoded by a heterologous polynucleotide. In some embodiments, the heterologous polynucleotide provides resistance to AHAS inhibitors and comprises a sequence having at least 80% identity to a 20 sequence selected from the group consisting of: SEQ ID NO:130 (ILV2 gene from S. cerevisiae BY4700) and SEQ ID NO:132 (ALS I gene from E. coli K12). In some embodiments, the heterologous polypeptide provides resistance to AHAS inhibitors and comprises an amino acid 25 sequence having at least 80% identity to an amino acid sequence selected from the group consisting of: SEQ ID NO:131 (ILV2 from S. cerevisiae BY4700) and SEQ ID NO:133 (ALS I from E. coli K12). In some embodiments, the polypeptide provides resistance to AHAS inhibitors and 30 comprises an amino acid sequence selected from the group consisting of: SEQ ID NO:131 (ILV2 from S. cerevisiae BY4700) and SEQ ID NO:133 (ALS I gene from E. coli K12) or an active variant, fragment or derivative thereof. In some embodiments, the polypeptide is from a bacteria of the 35 family Enterobacteriaceae. In some embodiments, the polypeptide is from a bacterial genus selected from the group consisting of: Alishewanella, Alterococcus, Aquamonas, Aranicola, Arsenophonus, Azotivirga, Blochmannia, Brenneria, Buchnera, Budvicia, Buttiauxella, Cedecea, Cit- 40 robacter, Cronobacter, Dickeya, Edwardsiella, Enterobacter, Erwinia, Escherichia, Ewingella, Grimontella, Haf-Kluyvera, Leclercia, Leminorella, nia. Klebsiella, Moellerella, Morganella, Obesumbacterium, Pantoea, Candidatus Phlomobacter, Photorhabdus, Poodoomaamaana, 45 Plesiomonas, Pragia, Proteus, Providencia, Rahnella, Raoultella, Salmonella, Samsonia, Serratia, Shigella, Sodalis, Tatumella, Trabulsiella, Wigglesworthia, Xenorhabdus, Yersinia, and Yokenella. In some embodiments, the polypeptide is from a microorganism of the genus Saccharomy- 50 ces. In some embodiments, the AHAS enzymes can be mutated to confer sulfonyl urea resistance. In some embodiments, the B. subtilis AlsS enzyme is mutated to increase its sulfonyl urea resistance.

thesis is glyphosate. In some embodiments, resistance to the glyphosate is conferred by a polypeptide that has 5-enolpyruvoyl-shikimate-3-phosphate synthetase (EPSPS) activity. In some embodiments, the polypeptide is encoded by a heterologous polynucleotide. In some embodiments, the 60 inhibitor is a 5-enolpyruvoyl-shikimate-3-phosphate synthetase (EPSPS) inhibitor. In some embodiments, the inhibitor is a glyphosate derivative. In some embodiments, resistance to the glyphosate is conferred by a polypeptide that has glyphosate N-acetyltransferase activity. In some embodi- 65 ments, the polypeptide is encoded by a heterologous polynucleotide. Sequences describing polypeptides with glypho-

sate N-acetyltransferase activity are described in, for example, U.S. Pat. No. 7,863,503, which is incorporated herein by reference.

In some embodiments, the antibiotic is selected from the group consisting of: aureobasidin A, bialaphos, cerulenin, chloramphenicol, cyclohexamide, geneticin/G418, hygromycin B, methotrexate, nourseothricin, phleomycin, triazole, and mixtures thereof. In some embodiments, a polypeptide confers resistance to one or more antibiotics. In some embodiments, the polypeptide is encoded by a heterologous polynucleotide.

In some embodiments a polypeptide confers resistance to the inhibitor or antibiotic. In some embodiments, the polypeptide is encoded by a polynucleotide. In some embodiments, the polypeptide conferring resistance to the inhibitor or antibiotic has one or more amino acid deletions when compared with the amino acid sequence of the corresponding native polypeptide. In some embodiments, the amino acid sequence of the polypeptide has one or more amino acid substitutions when compared with the amino acid sequence of the corresponding native polypeptide.

In some embodiments, the inhibitor is an  $\alpha$ -isopropyl malate (a-IPM) synthase inhibitor. In some embodiments, the inhibitor is trifluoroleucine or a trifluoroleucine derivative. In some embodiments, the inhibitor is a 3-deoxy-Darabino-heptolusonate-7-phosphate synthase (DAHPS) inhibitor. In some embodiments, the inhibitor is fluorophenyalanine or a fluorophenyalanine derivative. In some embodiments, the inhibitor is bismuth (III) or copper (II).

In some embodiments, the polypeptide confers tolerance to fluoroacetate. In some embodiments, the polypeptide confers tolerance to formaldehyde. In some embodiments, the polypeptide confers tolerance to sulfite.

In some embodiments, the polypeptide confers tolerance to an  $\alpha$ -isopropyl malate (a-IPM) synthase inhibitor. In some embodiments, the polypeptide confers tolerance to trifluoroleucine or a trifluoroleucine derivative (isopropyl malate resistance). In some embodiments, the polynucleotide sequence encoding the polypeptide providing resistance to trifluoroleucine comprises a sequence having at least 80% identity to a sequence disclosed by: Chianelli, M. S., et al., Cell. Mol. Biol. 42(6):847-57 (1996) or Oba, T., et al., Biosci. Biotechnol. Biochem. 70(7):1776-9 (2006) and incorporated by reference. In some embodiments, the polypeptide confers tolerance to a 3-deoxy-D-arabino-heptolusonate-7-phosphate synthase (DAHPS) inhibitor. In some embodiments, the polynucleotide sequence encoding the polypeptide providing resistance to DAHPS comprises a sequence having at least 80% identity to a sequence disclosed by: Fukada, K., et al., Agric. Biol. Chem. 54:3151-3156 (1990); Meuris, P. 1974. Genetics 76:735-744 (1974); Shimura, K., et al., 1993. Enzyme Microbiol. Technol. 15:874-876 (1993) and incorporated by reference.

In some embodiments, the polypeptide confers tolerance In some embodiments, the inhibitor of amino acid syn- 55 to an antibiotic. In some embodiments, the polypeptide confers tolerance to an antibiotic selected from the group consisting of: aureobasidin A, bialaphos, cerulenin, chloramphenicol, cyclohexamide, geneticin, hygromycin B, methotrexate, nourseothricin, phleomycin, triazole, and mixtures thereof. In some embodiments, the polynucleotide sequence encoding the polypeptide comprises a sequence having at least 80% identity to a sequence selected from the group consisting of: SEQ ID NOs: 92 and 143-157 (Aureobasidin A resistance (AUR1-C) (SEQ ID NOs:143 and 144); bialiphos resistance protein (SEQ ID NOs:145 and 146); cerulenin resistance YML007W Chr 13 (SEQ ID NOs:147 and 148); Geneticin resistance (kanMX) (SEQ ID

NOs:149 and 150); Hygromycin B resistance (HygR) (SEQ ID NOs:151 and 152); *Streptomyces noursei* nourseothricin resistance (natl) (SEQ ID NOs:153 and 154); phleomycin/ zeocin binding protein (SEQ ID NOs:155 and 156); and Triazole resistance (cyp51A) (SEQ ID NOs:157 and 92).

In some embodiments, the inhibitor is inhibiting at least one enzyme selected from the group consisting of: 5-enolpyruvoyl-shikimate-3-phosphate synthetase,  $\alpha$ -isopropyl malate synthase, 3-deoxy-D-arabino-heptolusonate-7-phosphate synthase and mixtures thereof. In some embodiments, the polynucleotide sequence encoding the polypeptide providing resistance to 5-enolpyruvoyl -shikimate-3-phosphate synthetase comprises a sequence having at least 80% identity to a sequence disclosed by: Cao G, et al., (2012) PLoS ONE 7(6): e38718 (2012) incorporated by 15 reference. In some embodiments, the polynucleotide sequence encoding the polypeptide with glyphosate N-acetyltransferase activity comprises a sequence having at least 80% identity to a sequence disclosed by U.S. Pat. No. 7,666,644, which is incorporated herein by reference in its 20 entirety and Siehl D.L., et al., Pest Manag Sci. 61(3):235-40 (2005) incorporated by reference. In some embodiments, the polynucleotide sequence encoding the polypeptide providing resistance to 3-phosphoshikimate 1-carboxylvinyltransferase comprises a sequence having at least 80% identity to 25 a sequence disclosed by: Vande Berg B.J., et al., Pest Manag Sci. 64(4):340-5 (2008) incorporated by reference. In some embodiments, the polypeptide that provides resistance to the inhibitor is a formaldehyde dehydrogenase. In some embodiments, the polypeptide comprises an amino acid sequence of at least about 80% identity to SEQ ID NO:6 or 7.

In some embodiments, the polypeptide that confers resistance comprises an amino acid sequence of at least about 80% identity to SEQ ID NO:11 or SEQ ID NO:12. In some 35 embodiments, the polypeptide that confers resistance is a 3-phosphoshikimate 1-carboxylvinyltransferase. In some embodiments, the polypeptide comprises an amino acid sequence of at least about 80% identity to SEQ ID NO:13.

In some embodiments, one or more AHAS inhibitors is 40 present at a concentration from about 0.1 g/mL to about 2 g/mL, about 1.0 g/mL to about 0.1 g/mL, about 1 mg/mL to about 0.1 g/mL, or about 10 mg/mL to about 100 mg/mL. In some embodiments, one or more AHAS inhibitors is present at a concentration of 0.0125 mg/mL. In some embodiments, 45 one or more AHAS inhibitors is present at a concentration of 1 mg/mL. In some embodiments, one or more AHAS inhibitors is present at a concentration of 2 mg/mL.

In some embodiments, glyphosate is at a concentration from about 0.1 µg/mL to about 2 g/mL, for example about 10 µg/mL, about 100 µg/mL, about 1 mg/mL, about 10 mg/mL, about 100 mg/mL, about 1 g/mL, or about 2 g/mL.

In some embodiments, the antibiotic is present at a concentration from about 2 ppm to about 500 ppm, for example about 5 ppm, about 20 ppm, about 50 ppm, about 55 100 ppm, about 150 ppm, about 200 ppm, about 300 ppm, about 400 ppm, or about 500 ppm.

In some embodiments, the addition inhibitors of amino acid synthesis, antibiotics, or combinations thereof results in death of at least 10%, 20%, 30%, 40%, 50%, 60%, 70%, 60 80%, 90% or 100% of the microbial contaminants in the fermentation mix.

In embodiments, the effective concentration of inhibitor for use in methods provided herein can be determined empirically for a given production strain, contaminant strain, 65 and production process. Alternatively, minimal data may be obtained for a given system and used to determine appro-

priate concentrations for inhibitors. Such determination is disclosed and demonstrated herein (see Examples), and is readily available to one of skill in the art, equipped with this disclosure.

Briefly, as described herein, growth and/or production competitiveness may be improved by, for example, i) adding a genetic trait that provides growth and/or production competitiveness, or ii) by providing growth conditions that increase growth and/or production competitiveness, e.g. through the addition of selective inhibitors. While not wishing to be bound by theory, in the first case, increased growth competitiveness of a strain A carrying a genetic trait that provides growth competitiveness will have a higher biomass ratio after a growth phase compared to a competing strain B than an strain C isogenic to strain A without the genetic trait, i.e.  $cx_{A_mod}(t)/cx_{B_mod}(t) > cx_C(t)/cx_B(t)$ . In the later case, strain A will have a higher biomass ratio after a growth phase compared to a competing strain B under the conditions promoting growth competitiveness, i.e.  $cx_{A}$  (t,c(inhibitor)>0  $g/L)/cx_B$  (t,c(inhibitor)>0 g/L)> $cx_A$ (t,c(inhibitor)=0 g/L)/ $cx_4(t, c(inhibitor)=0 g/L)$ . After a growth phase indicates that the biomass of strain A had to increase during the respective time interval, i.e.  $cx_A(t) > cx_A(t_{start})$ . While not wishing to be bound by theory, in the first case, increased production competitiveness of a strain A carrying a genetic trait that provides production competitiveness will have a higher product ratio compared to a product of a competing strain B than an strain C isogenic to strain A without the genetic trait, i.e.  $cp_A(t)/cp_B(t)>cp_C(t)/cp_B(t)$ . In the later case, strain A will have a higher product ratio compared to a product of a competing strain B, i.e.  $cp_A$  (t,c(inhibitor)>0 g/L/ $cp_B$  (t,c(inhibitor)>0 g/L)> $cp_A$ (t,c(inhibitor)=0 g/L)/  $cp_B(t, c(inhibitor)=0 g/L).$ 

Under situations where substrates are not limiting, e.g. under glucose excess conditions, maximum specific growth rate of the strains under the given cultivation conditions (medium, temperature, etc.) will be a component for determining growth competitiveness. Changes in the given conditions (e.g. changing concentrations of products, substrates, signaling molecules, etc.) may result in different values of maximum specific growth rate, and the maximum specific growth rate of strains may be different in a given condition. Considering such factors and assuming a constant  $\mu_{max}$  for an exponential growth phase, the biomass concentration during the exponential growth phase that started at  $t_{lag}$  can be approximately described according to

$$c_X(t) = c_X(t) e^{\mu_{max}(t-t) \log t}$$
Eq. (1)

Under aerobic, glucose-excess conditions and acetic acid plary ethanologen S. cerevisiae strain (PNY 827) was determined to be 0.61 l/h. In contrast, an isobutanologen S. cerevisiae strain (PNY 2129, constructed using PNY827) exhibited a maximum specific growth rate of 0.16 l/h. If an aerobic batch cultivation with a mixed culture consisting of both strains with a biomass ratio of 1:1, i.e. with a cell dry weight concentration of 1 g/L each, would be started, and both strains would be growing for 8 hours at  $\mu_{max}$  without any lag phase, at the end of the process PNY 827 would account for approximately 131.6 g/L cell dry weight in the mixed culture, while PNY 2129 would account for only about 3.6 g/L. The ratio of biomass PNY 2129/PNY 827 would be below 3%. This phenomenon is illustrated in FIG. 18 where PNY 2129 is represented by strain A and PNY 827 by strain B.

In order to describe growth performance of a strain according to Eq. (1) under the influence of different inhibi-

10

15

30

tor/mixture of inhibitors concentrations in a batch experiment,  $\mu_{max}$  was determined in a way to incorporate effect of the inhibitor/mixture of inhibitors. According to the observed inhibition kinetics, usually a "squared" inhibition term according to Eq. (2) was applied,

$$\mu_{max} = \frac{\mu_{max}^{o}}{\left(1 + \frac{c(I)^2}{K_I^2}\right)},$$
 Eq. (2)

with  $\mu_{max}$  denoting a strain characteristic maximum specific growth rate at the inhibitor concentration c(I), and  $\mu^{\circ}_{max}$  the maximum specific growth rate of the strain under the same conditions, but without inhibitor (c(inhibitor)=0 g/L). Finally K<sub>I</sub> represents an inhibitory constant K<sub>I</sub>.

In some occasions, inhibition of a cellular process significantly reduces but not completely abolishes growth of the organism. This effect can sometimes be explained by the action of alternative cellular processes available to the <sup>20</sup> organism. In this particular situation, the "squared" inhibition term according to Eq. (2) is insufficient to describe growth of the strain, and a hybrid modeling approach according to Eq. (3) was used instead according to 25

$$\mu_{max} = \frac{\mu_{max}^o 1}{\left(1 + \frac{c(I)^2}{K_I^2}\right)} + \mu_{max 2}^o, \qquad \text{Eq. (3)}$$

with  $\mu_{max}$  denoting a strain characteristic maximum specific growth rate at the inhibitor concentration c(I), and the sum of  $\mu^{o}{}_{max1}$  and  $\mu^{o}{}_{max2}$  the maximum specific growth rate of the strain under the same conditions, but without inhibitor 35 (c(inhibitor)=0 g/L). Finally  $K_I$  represents an inhibitory constant K<sub>1</sub>. Using such equations to fit minimal data collected for a given system allows for determination of strain-specific parameters, i.e. of a maximum specific growth rate without inhibitor  $(\mu^{\circ}_{max} \text{ or sum of } \mu^{\circ}_{max1} \text{ and } 40$  $\mu^{\circ}_{max^2}$ ) and an inhibitory constant K<sub>I</sub>. Based on these parameters, effect of inhibitor concentrations on maximum specific growth rate  $\mu_{max}$  of a given production or contaminant strain can be made, as well as the IC50 value of the inhibitor on their growth be estimated. Equipped with this 45 disclosure, one of skill in the art will be able to utilize parameters such as the IC50 to determine suitable concentrations of compounds for methods provided herein.

One embodiment is directed to a method for improving production competitiveness of a butanologen in a fermen- 50 tation mix, wherein the method comprises contacting a genetically modified host cell and a fermentation medium comprising one or more inhibitors, antibiotics, or combinations thereof, as well as a contaminating organism, and wherein the improved production competitiveness is asso-55 ciated with a higher butanol yield on the consumed substrate.

One embodiment is directed to a method for improving production competitiveness of a butanologen in a fermentation mix, wherein the method comprises contacting a 60 genetically modified host cell and a fermentation medium comprising one or more inhibitors, antibiotics, or combinations thereof, as well as an ethanologen yeast, and wherein the improved production competitiveness is associated with a higher butanol-to-ethanol ratio as compared to a cultivation without addition of one or more inhibitors, antibiotics, or combinations thereof.

One embodiment is directed to a method for improved production competitiveness of a butanologen in a fermentation mix, wherein the method comprises contacting a genetically modified host cell and a fermentation medium comprising one or more inhibitors, antibiotics, or combinations thereof, and wherein the addition of the one or more inhibitors, antibiotics, or combinations thereof results in less than a 20% loss in the yield of a lower alkyl alcohol produced by the host cell due to the presence of microbial contaminants. In some embodiments, the addition of the one or more inhibitors of amino acid synthesis, antibiotics, or combinations thereof results in less than a 10% loss in the yield of a lower alkyl alcohol produced by the host cell due to the presence of microbial contaminants.

It will be appreciated that compounds such as an inhibitor, antibiotic, or combinations thereof can be incorporated into a fermentation mix using any method known in the art. In embodiments, compounds are introduced by incorporation into a fermentation feed. In embodiments, compounds are introduced as a bolus or over the course of a fermentation process or a portion of the process as suitable for the compound and production process.

Alcohol Production

Disclosed herein are processes suitable for production of fermentation products from a carbon substrate. In one embodiment a lower alcohol is produced. In one embodiment, butanol is produced, and a butanologen is employed. In another embodiment, isobutanol is produced, and an isobutanologen is employed. In some embodiments, isobutanologens may comprise an isobutanol biosynthetic pathway, such as, but not limited to isobutanol biosynthetic pathways disclosed elsewhere herein. The ability to utilize carbon substrates to produce isobutanol can be confirmed using methods known in the art, including, but not limited to those described in U.S. Pat. Nos. 7,851,188 and 7,993,889 which are incorporated herein by reference. For example, to confirm isobutanol production, the concentration of isobutanol in the culture media can be determined by a number of methods known in the art. For example, a specific high performance liquid chromatography (HPLC) method utilized a Shodex SH-1011 column with a Shodex SH-G guard column, both purchased from Waters Corporation (Milford, Mass.), with refractive index (RI) detection. Chromatographic separation was achieved using 0.01  $\rm M\,H_2SO_4$  as the mobile phase with a flow rate of 0.5 mL/min and a column temperature of 50° C. Isobutanol had a retention time of 46.6 min under the conditions used. Alternatively, gas chromatography (GC) methods are available. For example, a specific GC method utilized an HP-INNOWax column (30 m×0.53 mm id, 1 µm film thickness, Agilent Technologies, Wilmington, Del.), with a flame ionization detector (FID). The carrier gas was helium at a flow rate of 4.5 mL/min, measured at 150° C. with constant head pressure; injector split was 1:25 at 200° C.; oven temperature was 45° C. for 1 min, 45 to 220° C. at 10° C./min, and 220° C. for 5 min; and FID detection was employed at 240° C. with 26 mL/min helium makeup gas. The retention time of isobutanol was 4.5 min.

In some embodiments, the butanologen comprises an engineered butanol pathway. In some embodiments, the butanologen is an isobutanologen. In some embodiments, the butanologen is a yeast. In some embodiments, the butanologen is a member of a genus of *Saccharomyces*, *Schizosaccharomyces*, *Hansenula*, *Candida*, *Kluyveromyces*, *Yarrowia*, *Issatchenkia*, or *Pichia*. In some embodiments, the butanologen is *Saccharomyces* cerevisiae.

In some embodiments, the engineered isobutanologen contains one or more polypeptides selected from a group of enzymes having the following Enzyme Commission Numbers: EC 2.2.1.6, EC 1.1.1.86, EC 4.2.1.9, EC 4.1.1.72, EC 1.1.1.1, EC 1.1.1.265, EC 1.1.1.2, EC 1.2.4.4, EC 1.3.99.2, 5 EC 1.2.1.57, EC 1.2.1.10, EC 2.6.1.66, EC 2.6.1.42, EC 1.4.1.9, EC 1.4.1.8, EC 4.1.1.14, EC 2.6.1.18, EC 2.3.1.9, EC 2.3.1.16, EC 1.1.130, EC 1.1.1.35, EC 1.1.1.157, EC 1.1.1.36, EC 4.2.1.17, EC 4.2.1.55, EC 1.3.1.44, EC 1.3.1.38, EC 5.4.99.13, EC 4.1.1.5, EC 2.7.1.29, EC 10 1.1.1.76, EC 1.2.1.57, and EC 4.2.1.28.

In some embodiments, the engineered isobutanologen contains one or more polypeptides selected from acetolactate synthase, acetohydroxy acid isomeroreductase, acetohydroxy acid dehydratase, branched-chain alpha-keto acid 15 decarboxylase, branched-chain alcohol dehydrogenase, acylating aldehyde dehydrogenase, branched-chain keto acid dehydrogenase, butyryl-CoA dehydrogenase, transaminase, valine dehydrogenase, valine decarboxylase, omega transaminase, acetyl-CoA acetyltransferase, 3-hydroxybu- 20 tyryl-CoA dehydrogenase, crotonase, butyryl-CoA dehydrogenase, isobutyryl-CoA mutase, acetolactate decarboxylase, acetonin aminase, butanol dehydrogenase, butyraldehyde dehydrogenase, acetoin kinase, acetoin phosphate aminase, aminobutanol phosphate phospholyase, aminobutanol 25 kinase, butanediol dehydrogenase, and butanediol dehydratase.

In some embodiments, the carbon substrate is selected from the group consisting of: oligosaccharides, polysaccharides, monosaccharides, and mixtures thereof. In some 30 embodiments, the carbon substrate is selected from the group consisting of: fructose, glucose, lactose, maltose, galactose, sucrose, starch, cellulose, feedstocks, ethanol, lactate, succinate, glycerol, corn mash, sugar cane, biomass, a C5 sugar, such as xylose and arabinose, and mixtures 35 thereof.

In some embodiments, the engineered isobutanol pathway comprises the following substrate to product conversions: a. pyruvate to acetolactate

- b. acetolactate to 2,3-dihydroxyisovalerate
- c. 2,3-dihydroxyisovalerate to  $\alpha$ -ketoisovalerate
- d.  $\alpha$ -ketoisovalerate to isobutyraldehyde, and
- e. isobutyraldehyde to isobutanol.

In some embodiments, one or more of the substrate to product conversions utilizes NADH or NADPH as a cofac- 45 tor.

In some embodiments, enzymes from the biosynthetic pathway are localized to the cytosol. In some embodiments, enzymes from the biosynthetic pathway that are usually localized to the mitochondria are localized to the cytosol. In 50 some embodiments, an enzyme from the biosynthetic pathway is localized to the cytosol by removing the mitochondrial targeting sequence. In some embodiments, mitochondrial targeting is eliminated by generating new start codons as described in e.g., U.S. Pat. Nos. 7,851,188 and 7,993,889, 55 which are incorporated herein by reference in its entirety. In some embodiments, the enzyme from the biosynthetic pathway that is localized to the cytosol is DHAD. In some embodiments, the enzyme from the biosynthetic pathway that is localized to the cytosol is KARI. 60

In some embodiments, the butanologen produces butanol at least 90% of effective yield, at least 91% of effective yield, at least 92% of effective yield, at least 93% of effective yield, at least 94% of effective yield, at least 95% of effective yield, at least 96% of effective yield, at least 97% of effective yield, at least 98% of effective yield, or at least 99% of effective yield. In some embodiments, the butanologen produces

butanol at least 55% to at least 75% of effective yield, at least 50% to at least 80% of effective yield, at least 45% to at least 85% of effective yield, at least 40% to at least 90% of effective yield, at least 35% to at least 95% of effective yield, at least 25% to at least 90% of effective yield, at least 25% to at least 99% of effective yield or at least 10% to at least 100% of effective yield.

In some embodiments, the host cell produces ethanol at least 90% of effective yield, at least 91% of effective yield, at least 92% of effective yield, at least 93% of effective yield, at least 94% of effective yield, at least 95% of effective yield, at least 96% of effective yield, at least 97% of effective yield, at least 98% of effective yield, or at least 99% of effective yield, at least 55% to at least 75% of effective yield, at least 55% to at least 75% of effective yield, at least 80% of effective yield, at least 45% to at least 85% of effective yield, at least 40% to at least 85% of effective yield, at least 95% of effective yield, at least 30% of effective yield, at least 95% of effective yield, at least 90% of effective yield, at least 95% of effective yield, at least 90% of effective yield, at least 95% of effective yield, at least 90% of effective yield, at least 95% of effective yield, at least 90% of effective yield.

In some embodiments, the host cell produces a C3-C6 alcohol at least 90% of effective yield, at least 91% of effective yield, at least 92% of effective yield, at least 93% of effective yield, at least 92% of effective yield, at least 95% of effective yield, at least 96% of effective yield, at least 97%
of effective yield, at least 96% of effective yield, at least 97%
of effective yield, at least 98% of effective yield, at least 97%
of effective yield, at least 98% of effective yield, at least 97%
of effective yield. In some embodiments, the host cell produces a C3-C6 alcohol at least 55% to at least 75% of effective yield, at least 50% to at least 80% of effective yield, at least 45% to at least 85% of effective yield, at least 40% to at least 90% of effective yield, at least 30% to at least 95% of effective yield, at least 25% to at least 99% of effective yield, at least 90% of effective yield, at least 10% to at least 99% of effective yield.

One embodiment of this invention is directed to a method for the production of a C3-C6 alcohol comprising:

- a. providing a host cell with an engineered pyruvate-utilizing pathway and a polypeptide conferring resistance to one or more inhibitors, antibiotics or combinations thereof, wherein the engineered pyruvate-utilizing pathway is a C3-C6 alcohol biosynthetic pathway;
- b. contacting the host cell with a fermentable carbon substrate in a fermentation medium under conditions whereby the C3-C6 alcohol is produced; and
- c. recovering the C3-C6 alcohol.

In some embodiments, the fermentation medium comprises one or more inhibitors, antibiotics or combinations thereof.

In some embodiments, the C3-C6 alcohol is produced at a titer from about 5 g/L to about 100 g/L. In some embodiments, the C3-C6 alcohol is produced at a titer of at least 20 g/L. In some embodiments, the C3-C6 alcohol is selected from the group consisting of: butanol, isobutanol, propanol,
and isopropanol.

One embodiment is a method for the production of ethanol comprising:

 a. providing a host cell with a pyruvate-utilizing pathway and a polypeptide conferring resistance to one or more inhibitors, antibiotics or combinations thereof, wherein the pyruvate-utilizing pathway is an ethanol producing pathway; b. contacting the host cell with a fermentable carbon substrate in a fermentation medium under conditions whereby the ethanol is produced; and

c. recovering the ethanol.

In some embodiments, the invention provides a method 5 for production of a fermentation product in a fermentation process comprising contacting a fermentation mix comprising a recombinant production microorganism which comprises a pyruvate-utilizing pathway with at least one compound which preferentially inhibits at least one contaminant 10 microorganism. In some embodiments the inhibition is measured through a reduction in the specific growth rate. In some embodiments the inhibition is measured through a reduced specific product formation rate of the contaminant. In some embodiments, the specific growth rate of the at least 15 one contaminant microorganism is reduced more than the specific growth rate of the recombinant production microorganism. In some embodiments, the production of the fermentation product of the at least one contaminant microorganism is reduced more than production of the fermenta- 20 tion product of the recombinant production microorganism.

In some embodiments, the major product of a production microorganism is ethanol. In some embodiments, the titer of ethanol that is produced may be at least about 80 g/L to at least about 120 g/L. In some embodiments, the titer of 25 ethanol that is produced is least about 50 g/L, at least about 60 g/L, at least about 70 g/L, at least about 80 g/L, at least about 90 g/L, at least about 95 g/L, at least about 100 g/L, at least about 105 g/L, at least about 110 g/L, at least about 115 g/L, or at least about 120 g/L. 30

In some embodiments, the major product of a production microorganism is butanol. In some embodiments, the titer of butanol that is produced may be at least about 80 g/L to at least about 120 g/L. In some embodiments, the titer of butanol that is produced is least about 50 g/L, at least about 35 60 g/L, at least about 70 g/L, at least about 80 g/L, at least about 90 g/L, at least about 95 g/L, at least about 100 g/L, at least about 105 g/L, at least about 110 g/L, at least about 115 g/L, or at least about 120 g/L.

In some embodiments, the major product of a production 40 microorganism is butanol, for example, isobutanol. In some embodiments, the major product of a contaminant microorganism is ethanol. The titer of ethanol may be less than that of butanol. In some embodiments, the titer of ethanol is less than about 20 g/L, 10 g/L, less than about 5 g/L, or less than 45 about 2 g/L.

In embodiments, the major product of a production microorganism is butanol, for example isobutanol, and the major product of a contaminant microorganism is ethanol. In embodiments, the percentage of ethanol produced as a 50 fraction of the amount of butanol produced is less than about 25%, less than about 20%, less than about 10%, less than about 5%, less than about 2%, or less than about 1%.

In some embodiments, butanol is contacted with a fatty acid and a lipase producing a fatty acid butyl ester 55 CAT-1, CBS7959, CBS7960, and CBS7961. ("FABE"), which may be used as a biodiesel fuel. Saccharomyces cerevisiae PNY860 (o

The reduction in contamination can be measured through any assay known in the art, including, but not limited to, standard plating assays, qPCR/RT-PCR, or by measuring fermentation titer, yield, or specific growth rate of a production microorganism in relation to a contaminant microorganism.

In some embodiments, reduction in contamination and increased production competitiveness of the butanologen is observed through measurement of the ratio of the desired 65 fermentation product to the contaminant fermentation production (e.g. butanol to ethanol). As the contaminant micro-

organism is inhibited to a greater degree than the production microorganism is inhibited, by either specific inhibitors or different concentrations of those inhibitors, the ratio of the desired fermentation product to the contaminant fermentation product will increase. The production of fermentation product in the aqueous phase can be quantified by HPLC, as described in the General Methods Section.

In some embodiments, the reduction in contamination is observed through measurement of the specific growth rate of samples treated with varying concentrations of inhibitors in a cell suspension, and measuring the OD of the samples at designated time points.

In some embodiments, the reduction in contamination is seen through the use of plating assays. In some embodiments, early stationary phase cultures are used to inoculate top agar media which is poured onto petri plates. Filter disks containing different concentrations of inhibitor can be applied to the plate surface, and, after a period of incubation, zones of growth inhibition can be observed. Host Cells and Microorganisms

The terms "host cell" and "microorganism" are synonymous and used interchangeably throughout. In embodiments, suitable host cells include any yeast host useful for genetic modification and recombinant gene expression. In some embodiments, the host cell is a butanologen. In some embodiments, the host cell is an isobutanologen. In some embodiments, the isobutanologen host cell can be a member of the genera Schizosaccharomyces, Issatchenkia, Kluyveromyces, Yarrowia, Pichia, Candida, Hansenula, Aspergillus, Pachysolen, Rhodotorula, Zygosaccharomyces, Galactomyces, Torulaspora, Debayomyces, Williopsis, Dekkera, Kloeckera, Metschnikowia or Saccharomyces. In other embodiments, the host cell can be Saccharomyces cerevisiae, Schizosaccharomyces pombe, Kluyveromyces lactis, Kluyveromyces thermotolerans, Kluyveromyces marxianus, Candida glabrata, Candida albicans, Pichia stipitis, Yarrowia lipolytica, E. coli, or L. plantarum. In still other embodiments, the host cell is a yeast host cell. In some embodiments, the host cell is a member of the genera Saccharomyces. In some embodiments, the host cell is Kluyveromyces lactis, Candida glabrata or Schizosaccharomvces pombe. In some embodiments, the host cell is Saccharomyces cerevisiae. S. cerevisiae yeast are known in the art and are available from a variety of sources, including, but not limited to, American Type Culture Collection (Rockville, Md.), Centraalbureau voor Schimmelcultures (CBS) Fungal Biodiversity Centre, LeSaffre, Gert Strand AB, Ferm Solutions, North American Bioproducts, Martrex, and Lallemand. S. cerevisiae include, but are not limited to, BY4741, CEN.PK 113-7D, Ethanol Red® yeast, Ferm Pro<sup>™</sup> yeast, Bio-Ferm<sup>®</sup> XR yeast, Gert Strand Prestige

Batch Turbo alcohol yeast, Gert Strand Pot Distillers yeast, Gert Strand Distillers Turbo yeast, FerMax<sup>™</sup> Green yeast, FerMax<sup>™</sup> Gold yeast, Thermosacc® yeast, BG-1, PE-2, CAT-1, CBS7959, CBS7960, and CBS7961. Saccharomyces cerevisiae PNY860 (or PNY0860).

Saccharomyces cerevisiae PNY860 (or PNY0860), described in Example 4, was deposited with the American Type Culture Collection (ATCC), 10801 University Boulevard, Manassas, Va., 20110 on Jul. 21, 2011, and assigned ATCC Accession No. PTA-12007.

*Saccharomyces cerevisiae* PNY827, described in Examples 3 and 13, was deposited with the American Type Culture Collection (ATCC), 10801 University Boulevard, Manassas, Va., 20110 on Sep. 22, 2011, and assigned ATCC Accession No. PTA-12105.

In some embodiments, the host cell expresses an engineered butanol biosynthetic pathway. From time to time, such a host cell is referred to as a "butanologen". In some embodiments, the butanologen is an isobutanologen expressing an engineered isobutanol biosynthetic pathway. In some embodiments, the butanologen is a bacteria, cyanobacteria or filamentous fungi. In some embodiments, the 5 genus of the host cell is selected from the group consisting of: Salmonella, Arthrobacter, Bacillus, Brevibacterium, Clostridium, Corynebacterium, Gluconobacter, Nocardia, Pseudomonas, Rhodococcus, Streptomyces, Zymomonas, Escherichia, Lactobacillus, Lactococcus, Enterococcus, 10 Alcaligenes, Klebsiella, Paenibacillus, and Xanthomonas.

Some embodiments comprise a genetically modified host cell comprising:

a. an engineered C3-C6 alcohol biosynthetic pathway; and,

b. a polypeptide that is resistant to inhibitors, antibiotics, or 15 a combination thereof.

Carbon Substrates

Suitable carbon substrates may include, but are not limited to, monosaccharides such as fructose or glucose, oligosaccharides such as lactose, maltose, galactose, or 20 sucrose, polysaccharides such as starch or cellulose or mixtures thereof and unpurified mixtures from renewable feedstocks such as cheese whey permeate, cornsteep liquor, sugar beet molasses, and barley malt. Other carbon substrates may include ethanol, lactate, succinate, or glycerol. 25

"Sugar" includes monosaccharides such as fructose or glucose, oligosaccharides such as lactose, maltose, galactose, or sucrose, polysaccharides such as starch or cellulose, C5 sugars such as xylose and arabinose, and mixtures thereof.

Additionally the carbon substrate may also be one-carbon substrates such as carbon dioxide, or methanol for which metabolic conversion into key biochemical intermediates has been demonstrated. In addition to one and two carbon substrates, methylotrophic organisms are also known to 35 utilize a number of other carbon containing compounds such as methylamine, glucosamine and a variety of amino acids for metabolic activity. For example, methylotrophic yeasts are known to utilize the carbon from methylamine to form trehalose or glycerol (Bellion et al., Microb. Growth C1 40 Compd., [Int. Symp.], 7th (1993), 415-32, Editor(s): Murrell, J. Collin; Kelly, Don P. Publisher: Intercept, Andover, UK). Similarly, various species of Candida will metabolize alanine or oleic acid (Sulter et al., Arch. Microbiol. 153:485-489 (1990)). Hence it is contemplated that the source of 45 carbon utilized in the present invention may encompass a wide variety of carbon containing substrates and will only be limited by the choice of organism.

Although it is contemplated that all of the above mentioned carbon substrates and mixtures thereof are suitable in 50 the present invention, in some embodiments, the carbon substrates are glucose, fructose, and sucrose, or mixtures of these with C5 sugars such as xylose and arabinose for yeasts cells modified to use C5 sugars. Sucrose may be derived from renewable sugar sources such as sugar cane, sugar 55 beets, cassava, sweet sorghum, and mixtures thereof. Glucose and dextrose may be derived from renewable grain sources through saccharification of starch based feedstocks including grains such as corn, wheat, rye, barley, oats, and mixtures thereof. In addition, fermentable sugars may be 60 derived from renewable cellulosic or lignocellulosic biomass through processes of pretreatment and saccharification, as described, for example, in U.S. Patent Application Publication No. 20070031918 A1, which is incorporated herein by reference. Biomass includes materials comprising cellu- 65 lose, and optionally further comprising hemicellulose, lignin, starch, oligosaccharides and/or monosaccharides.

44

Biomass may also comprise additional components, such as protein and/or lipid. Biomass may be derived from a single source, or biomass can comprise a mixture derived from more than one source; for example, biomass may comprise a mixture of corn cobs and corn stover, or a mixture of grass and leaves. Biomass includes, but is not limited to, bioenergy crops, agricultural residues, municipal solid waste, industrial solid waste, sludge from paper manufacture, yard waste, wood and forestry waste. Examples of biomass include, but are not limited to, corn grain, corn cobs, crop residues such as corn husks, corn stover, grasses, wheat, wheat straw, barley, barley straw, hay, rice straw, switchgrass, waste paper, sugar cane bagasse, sorghum, soy, components obtained from milling of grains, trees, branches, roots, leaves, wood chips, sawdust, shrubs and bushes, vegetables, fruits, flowers, animal manure, and mixtures thereof.

In some embodiments, the carbon substrate is glucose derived from corn. In some embodiments, the carbon substrate is glucose derived from wheat. In some embodiments, the carbon substrate is sucrose derived from sugar cane.

In addition to an appropriate carbon source, fermentation media must contain suitable minerals, salts, cofactors, buffers and other components, known to those skilled in the art, suitable for the growth of the cultures and promotion of an enzymatic pathway described herein.

Fermentation Conditions

Typically cells are grown at a temperature in the range of about 20° C. to about 40° C. in an appropriate medium. Suitable growth media in the present invention include common commercially prepared media such as Sabouraud Dextrose (SD) broth, Yeast Medium (YM) broth, or broth that includes yeast nitrogen base, ammonium sulfate, and dextrose (as the carbon/energy source) or YPD Medium, a blend of peptone, yeast extract, and dextrose in optimal proportions for growing most Saccharomyces cerevisiae strains. Other defined or synthetic growth media may also be used, and the appropriate medium for growth of the particular microorganism will be known by one skilled in the art of microbiology or fermentation science. The use of agents known to modulate catabolite repression directly or indirectly, e.g., cyclic adenosine 2':3'-monophosphate, may also be incorporated into the fermentation medium.

Suitable pH ranges for the fermentation are from about pH 5.0 to about pH 9.0. In one embodiment, about pH 6.0 to about pH 8.0 is used for the initial condition. Suitable pH ranges for the fermentation of yeast are typically from about pH 3.0 to about pH 9.0. In one embodiment, about pH 5.0 to about pH 8.0 is used for the initial condition. Suitable pH ranges for the fermentation of other microorganisms are from about pH 3.0 to about pH 7.5. In one embodiment, about pH 4.5 to about pH 6.5 is used for the initial condition.

Fermentations may be performed under aerobic or anaerobic conditions. In one embodiment, anaerobic or microaerobic conditions are used for fermentations.

Industrial Batch and Continuous Fermentations

Isobutanol, or other products, may be produced using a batch method of fermentation. A classical batch fermentation is a closed system where the composition of the medium is set at the beginning of the fermentation and not subject to artificial alterations during the fermentation. A variation on the standard batch system is the fed-batch system. Fed-batch fermentation processes are also suitable in the present invention and comprise a typical batch system with the exception that the substrate is added in increments as the fermentation progresses. Fed-batch systems are useful when catabolite repression is apt to inhibit the metabolism of the cells and where it is desirable to have limited amounts of substrate in the media. Batch and fed-batch fermentations are common and well known in the art and examples may be found in Thomas D. Brock in *Biotechnology: A Textbook of Industrial Microbiology*, Second Edition (1989) Sinauer Associates, <sup>5</sup> Inc., Sunderland, Mass., or Deshpande, Mukund V., *Appl. Biochem. Biotechnol.*, 36:227, (1992).

Isobutanol, or other products, may also be produced using continuous fermentation methods. Continuous fermentation is an open system where a defined fermentation medium is added continuously to a bioreactor and an equal amount of conditioned media is removed simultaneously for processing. Continuous fermentation generally maintains the cultures at a constant high density where cells are primarily in log phase growth. Continuous fermentation allows for the modulation of one factor or any number of factors that affect cell growth or end product concentration. Methods of modulating nutrients and growth factors for continuous fermentation processes as well as techniques for maximizing the product formation are well known in the art of industrial microbiology and a variety of methods are detailed by Brock, supra.

It is contemplated that the production of isobutanol, or other products, may be practiced using batch, fed-batch or 25 continuous processes and that any known mode of fermentation would be suitable. Additionally, it is contemplated that cells may be immobilized on a substrate as whole cell catalysts or encapsulated within porous material (e.g. alginate beads) and subjected to fermentation conditions for 30 isobutanol production.

Methods for Isobutanol Isolation from the Fermentation Medium

Bioproduced isobutanol may be isolated from the fermentation medium using methods known in the art for ABE 35 fermentations (see, e.g., Durre, *Appl. Microbiol. Biotechnol.* 49:639-648 (1998), Groot et al., *Process. Biochem.* 27:61-75 (1992), and references therein). For example, solids may be removed from the fermentation medium by centrifugation, filtration, decantation, or the like. Then, the isobutanol may 40 be isolated from the fermentation medium using methods such as distillation, azeotropic distillation, liquid-liquid extraction, adsorption, gas stripping, membrane evaporation, or pervaporation.

Because isobutanol forms a low boiling point, azeotropic 45 mixture with water, distillation can be used to separate the mixture up to its azeotropic composition. Distillation may be used in combination with another separation method to obtain separation around the azeotrope. Methods that may be used in combination with distillation to isolate and purify 50 butanol include, but are not limited to, decantation, liquid-liquid extraction, adsorption, and membrane-based techniques. Additionally, butanol may be isolated using azeotropic distillation using an entrainer (see, e.g., Doherty and Malone, *Conceptual Design of Distillation Systems*, 55 McGraw Hill, New York, 2001).

The butanol-water mixture forms a heterogeneous azeotrope so that distillation may be used in combination with decantation to isolate and purify the isobutanol. In this method, the isobutanol containing fermentation broth is 60 distilled to near the azeotropic composition. Then, the azeotropic mixture is condensed, and the isobutanol is separated from the fermentation medium by decantation. The decanted aqueous phase may be returned to the first distillation column as reflux. The isobutanol-rich decanted organic 65 phase may be further purified by distillation in a second distillation column.

The isobutanol can also be isolated from the fermentation medium using liquid-liquid extraction in combination with distillation. In this method, the isobutanol is extracted from the fermentation broth using liquid-liquid extraction with a suitable solvent. The isobutanol-containing organic phase is then distilled to separate the butanol from the solvent.

Distillation in combination with adsorption can also be used to isolate isobutanol from the fermentation medium. In this method, the fermentation broth containing the isobutanol is distilled to near the azeotropic composition and then the remaining water is removed by use of an adsorbent, such as molecular sieves (Aden et al., *Lignocellulosic Biomass to Ethanol Process Design and Economics Utilizing Co-Current Dilute Acid Prehydrolysis and Enzymatic Hydrolysis for Corn Stover*, Report NREL/TP-510-32438, National Renewable Energy Laboratory, June 2002).

Additionally, distillation in combination with pervaporation may be used to isolate and purify the isobutanol from the fermentation medium. In this method, the fermentation broth containing the isobutanol is distilled to near the azeotropic composition, and then the remaining water is removed by pervaporation through a hydrophilic membrane (Guo et al., *J. Membr. Sci.* 245, 199-210 (2004)).

In situ product removal (ISPR) (also referred to as extractive fermentation) can be used to remove butanol (or other fermentative alcohol) from the fermentation vessel as it is produced, thereby allowing the microorganism to produce butanol at high yields. One method for ISPR for removing fermentative alcohol that has been described in the art is liquid-liquid extraction. In general, with regard to butanol fermentation, for example, the fermentation medium, which includes the microorganism, is contacted with an organic extractant at a time before the butanol concentration reaches a toxic level. The organic extractant and the fermentation medium form a biphasic mixture. The butanol partitions into the organic extractant phase, decreasing the concentration in the aqueous phase containing the microorganism, thereby limiting the exposure of the microorganism to the inhibitory butanol.

Liquid-liquid extraction can be performed, for example, according to the processes described in U.S. Patent Appl. Pub. No. 2009/0305370, the disclosure of which is hereby incorporated in its entirety. U.S. Patent Appl. Pub. No. 2009/0305370 describes methods for producing and recovering butanol from a fermentation broth using liquid-liquid extraction, the methods comprising the step of contacting the fermentation broth with a water immiscible extractant to form a two-phase mixture comprising an aqueous phase and an organic phase. Typically, the extractant can be an organic extractant selected from the group consisting of saturated, mono-unsaturated, poly-unsaturated (and mixtures thereof)  $\rm C_{12}$  to  $\rm C_{22}$  fatty alcohols,  $\rm C_{12}$  to  $\rm C_{22}$  fatty acids, esters of  $\rm C_{12}$ to  $C_{22}$  fatty acids,  $C_{12}$  to  $C_{22}$  fatty aldehydes, and mixtures thereof. The extractant(s) for ISPR can be non-alcohol extractants. The ISPR extractant can be an exogenous organic extractant such as olevl alcohol, behenvl alcohol, cetyl alcohol, lauryl alcohol, myristyl alcohol, stearyl alcohol, 1-undecanol, oleic acid, lauric acid, myristic acid, stearic acid, methyl myristate, methyl oleate, undecanal, lauric aldehyde, 20-methylundecanal, and mixtures thereof.

In some embodiments, an ester can be formed by contacting the alcohol in a fermentation medium with an organic acid (e.g., fatty acids) and a catalyst capable of esterfying the alcohol with the organic acid. In such embodiments, the organic acid can serve as an ISPR extractant into which the alcohol esters partition. The organic acid can be supplied to the fermentation vessel and/or derived from the biomass

supplying fermentable carbon fed to the fermentation vessel. Lipids present in the feedstock can be catalytically hydrolyzed to organic acid, and the same catalyst (e.g., enzymes) can esterify the organic acid with the alcohol. The catalyst can be supplied to the feedstock prior to fermentation, or can be supplied to the fermentation vessel before or contemporaneously with the supplying of the feedstock. When the catalyst is supplied to the fermentation vessel, alcohol esters can be obtained by hydrolysis of the lipids into organic acid and substantially simultaneous esterification of the organic acid with butanol present in the fermentation vessel. Organic acid and/or native oil not derived from the feedstock can also be fed to the fermentation vessel, with the native oil being hydrolyzed into organic acid. Any organic acid not esterified with the alcohol can serve as part of the ISPR extractant. The extractant containing alcohol esters can be separated from the fermentation medium, and the alcohol can be recovered from the extractant. The extractant can be recycled to the fermentation vessel. Thus, in the case of butanol production, for example, the conversion of the butanol to an ester reduces the free butanol concentration in the fermentation 20 medium, shielding the microorganism from the toxic effect of increasing butanol concentration. In addition, unfractionated grain can be used as feedstock without separation of lipids therein, since the lipids can be catalytically hydrolyzed to organic acid, thereby decreasing the rate of build-up of lipids in the ISPR extractant. Other butanol product recovery and/or ISPR methods may be employed, including those described in U.S. Pat. No. 8,101,808, incorporated herein by reference.

In situ product removal can be carried out in a batch mode or a continuous mode. In a continuous mode of in situ product removal, product is continually removed from the reactor. In a batchwise mode of in situ product removal, a volume of organic extractant is added to the fermentation vessel and the extractant is not removed during the process. For in situ product removal, the organic extractant can 35 contact the fermentation medium at the start of the fermentation forming a biphasic fermentation medium. Alternatively, the organic extractant can contact the fermentation medium after the microorganism has achieved a desired amount of growth, which can be determined by measuring 40 the optical density of the culture. Further, the organic extractant can contact the fermentation medium at a time at which the product alcohol level in the fermentation medium reaches a preselected level. In the case of butanol production according to some embodiments of the present invention, 45 the organic acid extractant can contact the fermentation medium at a time before the butanol concentration reaches a toxic level, so as to esterify the butanol with the organic acid to produce butanol esters and consequently reduce the concentration of butanol in the fermentation vessel. The ester-containing organic phase can then be removed from the fermentation vessel (and separated from the fermentation broth which constitutes the aqueous phase) after a desired effective titer of the butanol esters is achieved. In some embodiments, the ester-containing organic phase is separated from the aqueous phase after fermentation of the 55 available fermentable sugar in the fermentation vessel is substantially complete.

Isobutanol titer in any phase can be determined by methods known in the art, such as via high performance liquid chromatography (HPLC) or gas chromatography, as <sup>60</sup> described, for example in U.S. Patent Appl. Pub. No. US20090305370, which is incorporated herein by reference.

# EXAMPLES

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

General Methods

Standard recombinant DNA and molecular cloning techniques used in the Examples are well known in the art and are described by Sambrook et al. (Sambrook, J., Fritsch, E. F. and Maniatis, T. (Molecular Cloning: A Laboratory Manual; Cold Spring Harbor Laboratory Press, Cold Spring Harbor, 1989, here in referred to as Maniatis) and by Ausubel et al. (Ausubel et al., Current Protocols in Molecular Biology, pub. by Greene Publishing Assoc. and Wiley-Interscience, 1987).

Materials and methods suitable for the maintenance and growth of bacterial 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 et al., eds., American Society for Microbiology, Washington, D.C., 1994) or by Thomas D. Brock in (Brock, Biotechnology: A Textbook of Industrial Microbiology, Second Edition, Sinauer Associates, Inc., Sunderland, Mass. (1989). All reagents, restriction enzymes and materials used for the growth and maintenance of bacterial cells were obtained from Sigma-Aldrich Chemicals (St. Louis, Mo.), BD Diagnostic Systems (Sparks, Md.), Invitrogen (Carlsbad, Calif.), HiMedia (Mumbai, India), SD Fine chemicals (India), or Takara Bio Inc. (Shiga, Japan), unless otherwise specified.

The meaning of abbreviations is as follows: "sec" means second(s), "min" means minute(s), "h" means hour(s), "nm" means nanometers, "uL" or "µl" means microliter(s), "mL" means milliliter(s), "mg/mL" means milligram per milliliter, "L" means liter(s), "nm" means nanometers, "mM" means millimolar, "M" means molar, "mmol" means millimole(s), "umole" means micromole(s), "kg" means kilogram, "g" means gram(s), "µg" means microgram(s) and "ng" means nanogram(s), "PCR" means polymerase chain reaction, "OD" means optical density, "OD600" means the optical density measured at a wavelength of 600 nm, "kDa" means kilodaltons, "g" can also mean the gravitation constant, "bp" means base pair(s), "kbp" means kilobase pair(s), "kb" means kilobase, "%" means percent, "% w/v" means weight/ volume percent, "% v/v" means volume/volume percent, "HPLC" means high performance liquid chromatography, "g/L" means gram per liter, "µg/L" means microgram per liter, "ng/µL" means nanogram per microliter, "pmol/µL" means picomol per microliter, "RPM" means rotation per minute, "µmol/min/mg" means micromole per minute per milligram, "w/v" means weight per volume, "v/v" means volume per volume.

#### Example 1

Construction of Expression Vectors for Isobutanol Pathway Gene Expression in *S. cerevisiae* 

# pLH475-JEA1 Construction

65

The pLH475-JEA1 plasmid (SEQ ID NO:23) was constructed for expression of ALS and KARI in yeast. pLH475-JEA1 is a pHR81 vector (ATCC #87541) containing the following chimeric genes: (1) the CUP1 promoter (SEQ ID NO:24), acetolactate synthase coding region from *Bacillus subtilis* (AlsS; SEQ ID NO:25; protein SEQ ID NO:36) and CYC1 terminator2 (SEQ ID NO:27); (2) an ILV5 promoter (SEQ ID NO:28), Pf5.IIvC-JEA1 coding region (SEQ ID NO:29; protein SEQ ID NO:30 and ILV5 terminator (SEQ ID NO:31); and (3) the FBA1 promoter (SEQ ID NO:32), *S. cerevisiae* KARI coding region (ILV5; SEQ ID NO:33; 5 protein SEQ ID NO:34) and CYC1 terminator (SEQ ID NO:35). The Pf5.IIvC-JEA1 coding region is a sequence encoding KARI derived from *Pseudomonas fluorescens* but containing mutations, that was described in commonly owned and co-pending US Patent Application Publication 10 US20100197519A1, which is herein incorporated by reference (Pf5.IIvC-JEA1 encoded KARI: SEQ ID NO:29; protein SEQ ID NO:30)

#### Expression Vector pLH468

The pLH468 plasmid (SEQ ID NO:39) was constructed 15 for expression of DHAD, KivD and HADH in yeast. Coding regions for L. lactis ketoisovalerate decarboxylase (KivD) (SEQ ID NO:141) and Horse liver alcohol dehydrogenase (HADH) (SEQ ID NO:40 and 142) were synthesized by DNA2.0 based on codons that were optimized for expres- 20 sion in Saccharomyces cerevisiae and provided in plasmids pKivDy-DNA2.0 and pHadhy-DNA2.0. Individual expression vectors for KivD and HADH were constructed. To assemble pLH467 (pRS426::P<sub>TDH3</sub>-kivDy-TDH3t), vector pNY8 (SEQ ID NO:14; also named pRS426.GPD-ald- 25 GPDt, described in commonly owned and co-pending US Patent App. Pub. US2008/0182308, Example 17, which is herein incorporated by reference) was digested with AscI and SfiI enzymes, thus excising the GPD promoter and the ald coding region. A TDH3 promoter fragment (SEQ ID NO:41) from pNY8 was PCR amplified to add an AscI site at the 5' end, and an SpeI site at the 3' end, using 5' primer OT1068 and 3' primer OT1067 (SEQ ID NOs:42 and 43). The AscI/SfiI digested pNY8 vector fragment was ligated with the TDH3 promoter PCR product digested with AscI 35 and SpeI, and the Spa-SfiI fragment containing the codon optimized kivD coding region isolated from the vector pKivD-DNA2.0. The triple ligation generated vector pLH467 (pRS426::PTDH3+kivDy-TDH3t). pLH467 (SEQ ID NO:44) was verified by restriction mapping and sequencing. 40

pLH435 (pRS425::P<sub>GPM1</sub>-Hadhy-ADH1t) (SEQ ID NO:52) was derived from vector pRS425::GPM-sadB (SEQ ID NO:45) which is described in commonly owned and co-pending US Patent App. Pub No. US20090305363 A1, Example 3, which is herein incorporated by reference in its 45 entirety. pRS425::GPM-sadB is the pRS425 vector (ATCC #77106) with a chimeric gene containing the GPM1 promoter (SEQ ID NO: 46), coding region from a butanol dehydrogenase of Achromobacter xylosoxidans (sadB; DNA SEQ ID NO:47; protein SEQ ID NO:48: disclosed in U.S. 50 Pat. No. 8,188,250, which is herein incorporated by reference in its entirety), and ADH1 terminator (SEQ ID NO:49). pRS425::GPMp-sadB contains BbvI and PacI sites at the 5' and 3' ends of the sadB coding region, respectively. A Nha site was added at the 5' end of the sadB coding region by 55 site-directed mutagenesis using primers OT1074 and OT1075 (SEQ ID NO:50 and 51) to generate vector pRS425-GPMp-sadB-NheI, which was verified by sequencing. pRS425::PGPM1-sadB-NheI was digested with NheI and PacI to drop out the sadB coding region, and ligated with the 60 Nhel-Pacl fragment containing the codon optimized HADH coding region from vector pHadhy-DNA2.0 to create pLH435.

To combine KivD and HADH expression cassettes in a single vector, yeast vector pRS411 (ATCC #87474) was digested with Sad and Not I, and ligated with the SacI-SaII fragment from pLH467 that contains the  $P_{TDH3}$ -kivDy-

TDH3t cassette together with the SalI-NotI fragment from pLH435 that contains the  $P_{GPM1}$ -Hadhy-ADH1t cassette in a triple ligation reaction. This yielded the vector pRS411::  $P_{TDH3}$ -kivDy- $P_{GPM1}$ -Hadhy (pLH441), which was verified by restriction mapping.

In order to generate a co-expression vector for all three genes in the lower isobutanol pathway: ilvD, kivDy and Hadhy, we used pRS423 FBA ilvD(Strep) (SEQ ID NO:53, which is described in commonly owned and co-pending US Patent App. Pub. US 20100081154A1, which is herein incorporated by reference in its entirety, as the source of the IlvD gene. This shuttle vector contains an F1 origin of replication (nt 1423 to 1879) for maintenance in E. coli and a 2 micron origin (nt 8082 to 9426) for replication in yeast. The vector has an FBA1 promoter (nt 2111 to 3108; SEQ ID NO:32) and FBA terminator (nt 4861 to 5860; SEQ ID NO:54). In addition, it carries the His marker (nt 504 to 1163) for selection in yeast and ampicillin resistance marker (nt 7092 to 7949) for selection in E. coli. The ilvD coding region (nt 3116 to 4828; SEQ ID NO:55; protein SEQ ID NO:56) from Streptococcus mutans UA159 (ATCC #700610) is between the FBA promoter and FBA terminator forming a chimeric gene for expression. In addition there is a lumio tag fused to the ilvD coding region (nt 4829-4849).

The first step was to linearize pRS423 FBA ilvD(Strep) (also called pRS423-FBA(SpeI)-IlvD(*Streptococcus mutans*)-Lumio) with SacI and SacII (with SacII site blunt ended using T4 DNA polymerase), to give a vector with total length of 9,482 bp. The second step was to isolate the kivDy-hADHy cassette from pLH441 with SacI and KpaI (with KpaI site blunt ended using T4 DNA polymerase), which gives a 6,063 bp fragment. This fragment was ligated with the 9,482 bp vector fragment from pRS423-FBA (SpeI)-IlvD(*Streptococcus mutans*)-Lumio. This generated vector pLH468 (pRS423::P<sub>FBA1</sub>-ilvD(Strep)Lumio-FBA1t-P<sub>TDH3</sub>-kivDy-TDH3t-P<sub>GPM1</sub>-hadhy-ADH1t), which was confirmed by restriction mapping and sequencing.

#### Example 2

# Construction of *S. cerevisiae* Host Strain Containing Disruptions in Pyruvate Decarboxylase and Hexokinase II

This example describes insertion-inactivation of endogenous PDC1, PDC5, and PDC6 genes of *S. cerevisiae*. PDC1, PDC5, and PDC6 genes encode the three major isozymes of pyruvate decarboxylase. Hexokinase II, which is responsible for phosphorylation of glucose and transcriptional repression, is also inactivated. The resulting PDC/ HXK2 inactivation strain (U.S. Publication No: 2011/ 0124060, which is incorporated herein by reference) was used as a host for expression vectors pLH475-JEA1 and pLH468 that were described in Example 2.

Construction of pdc6:: $P_{GPM1}$ -sadB Integration Cassette and PDC6 Deletion:

A pdc6::P<sub>*GPM*1</sub>-sadB-ADH1t-URA3r integration cassette was made by joining the GPM-sadB-ADHt segment from pRS425::GPM-sadB (described above) to the URA3r gene from pUC19-URA3r. pUC19-URA3r (SEQ ID NO:57) contains the URA3 marker from pRS426 (ATCC #77107) flanked by 75 bp homologous repeat sequences to allow homologous recombination in vivo and removal of the URA3 marker. The two DNA segments were joined by SOE PCR (as described by Horton et al. (1989) *Gene* 77:61-68) using as template pRS425::GPM-sadB and pUC19-URA3r plasmid DNAs, with Phusion DNA polymerase (New England Biolabs Inc., Beverly, Mass.; catalog no. F-5405) and primers 114117-11A through 114117-11D (SEQ ID NOs:58, 59, 60 and 61), and 114117-13A and 114117-13B (SEQ ID NOs:62 and 63).

The outer primers for the SOE PCR (114117-13A and 5 114117-13B) contained 5' and 3' ~50 bp regions homologous to regions upstream and downstream of the PDC6 promoter and terminator, respectively. The completed cassette PCR fragment was transformed into BY4700 (ATCC #200866) and transformants were maintained on synthetic complete media lacking uracil and supplemented with 2% glucose at 30° C. using standard genetic techniques (Methods in Yeast Genetics, 2005, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., pp. 201-202). Transformants were screened by PCR using primers 112590-34G and 112590-15 34H (SEQ ID NOs:64 and 65), and 112590-34F and 112590-49E (SEQ ID NOs:66 and 67) to verify integration at the PDC6 locus with deletion of the PDC6 coding region. The URA3r marker was recycled by plating on synthetic complete media supplemented with 2% glucose and 5-FOA at 20 30° C. following standard protocols. Marker removal was confirmed by patching colonies from the 5-FOA plates onto SD -URA media to verify the absence of growth. The resulting identified strain has the genotype: BY4700 pdc6:: P<sub>GPM1</sub>-sadB-ADH1t. 25

Construction of pdc1::P<sub>PDC1</sub>-ilvD Integration Cassette and PDC1 Deletion:

A pdc1::P<sub>PDC1</sub>-ilvD-FBA1t-URA3r integration cassette was made by joining the ilvD-FBA1t segment from pLH468 (described above) to the URA3r gene from pUC19-URA3r 30 by SOE PCR (as described by Horton et al. (1989) Gene 77:61-68) using as template pLH468 and pUC19-URA3r plasmid DNAs, with Phusion DNA polymerase (New England Biolabs Inc., Beverly, Mass.; catalog no. F-540S) and primers 114117-27A through 114117-27D (SEQ ID NOs:68, 35 69, 70 and 71).

The outer primers for the SOE PCR (114117-27A and 114117-27D) contained 5' and 3' ~50 bp regions homologous to regions downstream of the PDC1 promoter and downstream of the PDC1 coding sequence. The completed 40 cassette PCR fragment was transformed into BY4700 pdc6:: P<sub>GPM1</sub>-sadB-ADH1t and transformants were maintained on synthetic complete media lacking uracil and supplemented with 2% glucose at 30° C. using standard genetic techniques (Methods in Yeast Genetics, 2005, Cold Spring Harbor 45 Laboratory Press, Cold Spring Harbor, N.Y., pp. 201-202). Transformants were screened by PCR using primers 114117-36D and 135 (SEQ ID NOs:72 and 73), and primers 112590-49E and 112590-30F (SEQ ID NOs:67 and 74) to verify integration at the PDC1 locus with deletion of the PDC1 50 coding sequence. The URA3r marker was recycled by plating on synthetic complete media supplemented with 2% glucose and 5-FOA at 30° C. following standard protocols. Marker removal was confirmed by patching colonies from the 5-FOA plates onto SD-URA media to verify the absence 55 a ~2.2 kb PCR product. The PDC5 portion of each primer of growth. The resulting identified strain "NYLA67" has the genotype: BY4700 pdc6::P<sub>GPM1</sub>-sadB-ADH1t pdc1::P<sub>PDC1</sub>ilvD-FBA1t.

HIS3 Deletion

To delete the endogenous HIS3 coding region, a his3:: 60 URA3r2 cassette was PCR-amplified from URA3r2 template DNA (SEQ ID NO:75). URA3r2 contains the URA3 marker from pRS426 (ATCC #77107) flanked by 500 bp homologous repeat sequences to allow homologous recombination in vivo and removal of the URA3 marker. PCR was 65 done using Phusion DNA polymerase and primers 114117-45A and 114117-45B (SEQ ID NOs:76 and 77) which

generated a ~2.3 kb PCR product. The HIS3 portion of each primer was derived from the 5' region upstream of the HIS3 promoter and 3' region downstream of the coding region such that integration of the URA3r2 marker results in replacement of the HIS3 coding region. The PCR product was transformed into NYLA67 using standard genetic techniques (Methods in Yeast Genetics, 2005, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., pp. 201-202) and transformants were selected on synthetic complete media lacking uracil and supplemented with 2% glucose at 30° C. Transformants were screened to verify correct integration by replica plating of transformants onto synthetic complete media lacking histidine and supplemented with 2% glucose at 30° C. The URA3r marker was recycled by plating on synthetic complete media supplemented with 2% glucose and 5-FOA at 30° C. following standard protocols. Marker removal was confirmed by patching colonies from the 5-FOA plates onto SD –URA media to verify the absence of growth. The resulting identified strain, called NYLA73, has the genotype: BY4700 pdc6::P<sub>GPM1</sub>-sadB-ADH1t pdc1::Pp<sub>PDC1</sub>-ilvD-FBA1t dhis3.

Deletion of HXK2 (Hexokinase II):

A hxk2::URA3r cassette was PCR-amplified from URA3r2 template (described above) using Phusion DNA polymerase and primers 384 and 385 (SEQ ID NOs:78 and 79) which generated a ~2.3 kb PCR product. The HXK2 portion of each primer was derived from the 5' region upstream of the HXK2 promoter and 3' region downstream of the coding region such that integration of the URA3r2 marker results in replacement of the HXK2 coding region. The PCR product was transformed into NYLA73 using standard genetic techniques (Methods in Yeast Genetics, 2005, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., pp. 201-202) and transformants were selected on synthetic complete media lacking uracil and supplemented with 2% glucose at 30° C. Transformants were screened by PCR to verify correct integration at the HXK2 locus with replacement of the HXK2 coding region using primers N869 and N871 (SEQ ID NOs:80 and 81). The URA3r2 marker was recycled by plating on synthetic complete media supplemented with 2% glucose and 5-FOA at 30° C. following standard protocols. Marker removal was confirmed by patching colonies from the 5-FOA plates onto SD -URA media to verify the absence of growth, and by PCR to verify correct marker removal using primers N946 and N947 (SEQ ID NOs:82 and 83). The resulting identified strain named NYLA83 has the genotype: BY4700 pdc6:: P<sub>GPM1</sub>-sadB-ADH1t pdc1::P<sub>PDC1</sub>-ilvD-FBA1t dhis3 dhxk2. Construction of pdc5::kanMX Integration Cassette and PDC5 Deletion

A pdc5::kanMX4 cassette was PCR-amplified from strain YLR134W chromosomal DNA (ATCC No. 4034091) using Phusion DNA polymerase and primers PDC5::KanMXF and PDC5::KanMXR (SEQ ID NOs:84 and 85) which generated was derived from the 5' region upstream of the PDC5 promoter and 3' region downstream of the coding region such that integration of the kanMX4 marker results in replacement of the PDC5 coding region. The PCR product was transformed into NYLA83, and transformants were selected and screened as described above. The identified correct transformants named NYLA84 have the genotype: pdc6::P<sub>GPM1</sub>-sadB-ADH1t pdc1::P<sub>PDC1</sub>-ilvD-BY4700 FBA1t dhis3 dhxk2 pdc5::kanMX4.

Plasmid vectors pLH468 and pLH475-JEA1 were simultaneously transformed into strain NYLA84 (BY4700 pdc6:: P<sub>GPM1</sub>-sadB-ADH1t pdc1::P<sub>PDC1</sub>-ilvD-FBA1t dhis3 dhxk2

pdc5::kanMX4) using standard genetic techniques (Methods in Yeast Genetics, 2005, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.) and the resulting strain was maintained on synthetic complete media lacking histidine and uracil, and supplemented with 1% ethanol at 30° C.

#### Example 3

#### Construction of S. cerevisiae Host Strain Containing Disruptions in URA3, HIS3, and Insertion of Sulfonylurea-Resistant ILV2

This example describes inactivation of the URA3 and HIS3 genes of S. cerevisiae, and replacement of the native ILV2 gene with a variant that is resistant to sulfonylurea 15 herbicides. The resulting strain will be used as a host for expression vectors pLH475-JEA1 and pLH468 that were described in Example 1.

**URA3** Deletion

To delete the endogenous URA3 coding region, a deletion 20 cassette was PCR-amplified from pLA54 (SEQ ID NO:100) which contains a P<sub>TEF1</sub>-kanMX-TEF1t cassette flanked by loxP sites to allow homologous recombination in vivo and subsequent removal of the KanMX marker. PCR was performed using Phusion DNA polymerase and primers BK505 25 and BK506 (SEQ ID NOs:101 and 102). The URA3 portion of each primer was derived from the 5' region 180 bp upstream of the URA3 ATG and 3' region 78 bp downstream of the coding region such that integration of the KanMX cassette results in replacement of the URA3 coding region. 30 The PCR product was transformed into PNY827 (ATCC # PTA-12105), using standard genetic techniques (Methods in Yeast Genetics, 2005, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., pp. 201-202) and transformants were selected on rich media supplemented 2% glucose and 35 100 µg/ml Geneticin at 30° C. Transformants were screened by colony PCR with primers BK468 and LA492 (SEQ ID NOs:103 and 104) to verify presence of the integration cassette. A heterozygous URA3 mutant was obtained; NYLA98 MATa/α URA3/ura3::loxP-kanMX-loxP. To 40 obtain haploids, NYLA98 was sporulated using standard methods (Codón AC, Gasent-Ramírez J M, Benítez T., Appl Environ Microbiol. 1995 PMID: 7574601). Tetrads were dissected using a micromanipulator and grown on rich media supplemented with 2% glucose. Tetrads containing 45 four viable spores were patched onto synthetic complete media lacking uracil and supplemented with 2% glucose, and the mating type was verified by multiplex colony PCR using primers AK109-1, AK109-2, and AK109-3 (SEQ ID NOs:105, 106, and 107). The resulting identified haploid 50 strain called NYLA103 has the genotype MAT $\alpha$  ura3 $\Delta$ :: loxP-kanMX-loxP, and NYLA106 has the genotype MATa ura $3\Delta$ ::loxP-kanMX-loxP.

HIS3 Deletion

The four fragments for the PCR cassette for the scarless 55 HIS3 deletion were amplified using Phusion High Fidelity PCR Master Mix (New England BioLabs; Ipswich, Mass.) and CEN.PK 113-7D genomic DNA as template, prepared with a Gentra Puregene Yeast/Bact kit (Qiagen; Valencia, Calif.). HIS3 Fragment A was amplified with primer oBP452 60 (SEQ ID NO: 89) and primer oBP453 (SEQ ID NO:109), containing a 5' tail with homology to the 5' end of HIS3 Fragment B. HIS3 Fragment B was amplified with primer oBP454 (SEQ ID NO:110), containing a 5' tail with homology to the 3' end of HIS3 Fragment A, and primer oBP455 (SEQ ID NO:90), containing a 5' tail with homology to the 5' end of HIS3 Fragment U. HIS3 Fragment U was amplified

with primer oBP456 (SEQ ID NO:91, containing a 5' tail with homology to the 3' end of HIS3 Fragment B, and primer oBP457 (SEQ ID NO:86), containing a 5' tail with homology to the 5' end of HIS3 Fragment C. HIS3 Fragment C was amplified with primer oBP458 (SEQ ID NO:87), containing a 5' tail with homology to the 3' end of HIS3 Fragment U, and primer oBP459 (SEQ ID NO:88). PCR products were purified with a PCR Purification kit (Qiagen). HIS3 Fragment AB was created by overlapping PCR by mixing HIS3 10 Fragment A and HIS3 Fragment B and amplifying with primers oBP452 (SEQ ID NO:89) and oBP455 (SEQ ID NO:90). HIS3 Fragment UC was created by overlapping PCR by mixing HIS3 Fragment U and HIS3 Fragment C and amplifying with primers oBP456 (SEQ ID NO:91) and oBP459 (SEQ ID NO:88). The resulting PCR products were purified on an agarose gel followed by a Gel Extraction kit (Qiagen). The HIS3 ABUC cassette was created by overlapping PCR by mixing HIS3 Fragment AB and HIS3 Fragment UC and amplifying with primers oBP452 (SEQ ID NO:89) and oBP459 (SEQ ID NO:88). The final PCR product was purified with a PCR Purification kit (Qiagen).

To delete the endogenous HIS3 coding region, the "scarless" deletion cassette was transformed into NYLA106 using standard techniques and plated on synthetic complete media lacking uracil and supplemented with 2% glucose. Transformants were screened to verify correct integration by replica plating onto synthetic complete media lacking histidine and supplemented with 2% glucose at 30° C. Genomic DNA preps were made to verify the integration by PCR using primers BP460 and LA135 (SEQ ID NOs:93 and 94) for the 5' end and primers BP461 and LA92 (SEQ ID NOs:95 and 96) for the 3' end. The URA3 marker was recycled by plating on synthetic complete media supplemented with 2% glucose and 5-FOA at 30° C. following standard protocols. Marker removal was confirmed by patching colonies from the 5-FOA plates onto SD -URA media to verify the absence of growth. The resulting identified strain, called PNY2003, has the genotype MATa ura $3\Delta$ ::loxP-kanMX-loxP his $3\Delta$ .

Deletion of PDC1:

To delete the endogenous PDC1 coding region, a deletion cassette was PCR-amplified from pLA59 (SEQ ID NO:97), which contains a URA3 marker flanked by degenerate loxP sites to allow homologous recombination in vivo and subsequent removal of the URA3 marker. PCR was done by using Phusion DNA polymerase and primers LA678 and LA679 (SEQ ID NOs:98 and 99). The PDC1 portion of each primer was derived from the 5' region 50 bp downstream of the PDC1 start codon and 3' region 50 bp upstream of the stop codon such that integration of the URA3 cassette results in replacement of the PDC1 coding region but leaves the first 50 bp and the last 50 bp of the coding region. The PCR product was transformed into strain PNY2003 using standard genetic techniques with selection on synthetic complete media lacking uracil and supplemented with 2% glucose at 30° C. Transformants were screened to verify correct integration by colony PCR using primers LA337 (SEQ ID NO:111), external to the 5' coding region and LA135 (SEQ ID NO:94), an internal to URA3. Positive transformants were then screened by colony PCR using primers LA692 and LA693 (SEQ ID NOs:112 and 113), which were internal to the PDC1 coding region. The URA3 marker was recycled by transforming with pRS423::P<sub>GAL1</sub>-cre (SEQ ID NO:121) and plated on synthetic complete media lacking histidine and supplemented with 2% glucose at 30° C. Transformants were plated on YP supplemented with 0.5% galactose to induce expression of Cre recombinase. Marker removal was

confirmed by patching colonies to synthetic complete media lacking uracil and supplemented with 2% glucose to verify absence of growth. The resulting identified strain, called PNY2008, has the genotype MATa ura3 $\Delta$ ::loxP-kanMXloxP his3 $\Delta$  pdc1 $\Delta$ ::loxP71/66.

Construction of ILV2-410 Integration Vector:

A fragment of the native ILV2 gene from S. cerevisiae BY4700 was PCR-amplified using Phusion DNA polymerase and primers LA684 and LA685 (SEQ ID NOs: 114 and 115). The ~2 kb PCR product was digested with BamHI 10and SphI and cloned into pUC19, and the resulting vector was named pUC19::ILV2 (SEQ ID No:17). Site-directed mutagenesis (QuickChange XL, Stratagene, CA) was used to introduce a C to T transition at base pair 574, resulting in a proline-to-serine substitution (Yadav et al. 1986 PNAS. 15 83:4418-4422). PfuUltra DNA polymerase (Stratagene, CA), primers LA682 and LA683 (SEQ ID NOs:116 and 117), and pUC19::ILV2 template were used to introduce the mutation following standard protocol. The PCR reaction was digested with DpnI to remove parental DNA, and the 20 reaction was transformed into DH5a competent cells on LB-Amp (100 µg/ml). The presence of DNA containing the ILV2-410 allele was confirmed by DNA sequencing of plasmid DNA isolated from transformants. The resulting vector was named pUC19::ILV2-410.

The ILV2-410 fragment was digested from pUC19::ILV2-410 by BamHI SphI digest and subcloned into pLA59. pLA59 (SEQ ID No:97) is a pUC19 cloning vector that contains a loxP71-URA3-loxP66 cassette. The resulting vector, pLA59::ILV2-410 (SEQ ID NO:18), was used as <sup>30</sup> template for PCR of the full integration cassette. The ILV2-410-loxP71-URA3-loxP66 integration cassette was PCR amplified from pLA59::ILV2-410 template using Phusion DNA polymerase and primers LA686 and LA687 (SEQ ID NOs:119 and 120), producing a ~3 kb product. The ILV2 <sup>35</sup> portion of each primer was derived from the 5' region downstream of the ILV2 start codon and 3' region downstream of the stop codon such that integration of the URA3 cassette results in replacement of the native ILV2 coding region. 40

The PCR product was transformed into strain PNY2008 and plated on synthetic complete media lacking uracil and supplemented with 2% glucose at 30° C. The loxP71-URA3loxP66 marker was recycled by transformation with pRS423::P<sub>GAL1</sub>-cre (SEQ ID NO:121) and plating on syn- 45 thetic complete media lacking histidine supplemented with 3% glucose at 30° C. Colonies were patched onto YP (1% galactose) plates at 30° C. to induce URA3 marker excision and were transferred onto YP (2% glucose) plates at 30° C. for recovery. Removal of the URA3 marker were confirmed 50 by patching colonies from the YP (2% glucose) plates onto synthetic complete media lacking uracil supplemented with 2% glucose to verify the absence of growth. The resulting identified strain, called PNY2010, has the genotype MATa ura3A::loxP-kanMX-loxP his3A pdc1A::loxP71/66 ILV2- 55 410::loxP71/66.

#### Example 4

# Susceptibility of Wildtype *S. cerevisiae* Strains to Sulfonylurea Herbicides

60

This example describes experiments that demonstrate yeast strains, expressing wildtype acetolactate synthase, are resistant to many AHAS-inhibiting sulfonylurea herbicides. 65 Strains tested in this experiment included: *S. cerevisiae* yeast strain PNY0860-1A), a haploid derived from sporulation of

the yeast strain deposited with ATCC (ATCC #PTA-12007; *S. cerevisiae* yeast strain PNY827 (ATCC # PTA-12105); and *S. cerevisiae* strain CEN.PK 113-7D (Centraalbureau voor Schimmelcultures (CBS) Fungal Biodiversity Centre #8340).

The following AHAS inhibitors were resuspended in 10 mM KOH to give final concentrations of 2 mg/ml (w/v).

Accent TM	nicosulfuron methyl (V9360)
W4189-128	research sample
Ally TM	metsulfuron methyl (T6376)
Classic TM	chlorimuron ethyl (F6025)
Oust TM	sulfometuron methyl (SM)
Harmony TM	thifensulfuron methyl

The yeast strains were initially grown on synthetic complete liquid media lacking all amino acids and supplemented with 2% glucose at 30° C. Early stationary phase cultures (OD600 nm of ~5.0) were used to inoculate 40 ml of top agar media (synthetic complete lacking all amino acids with 0.7% agarose), which were poured into petri plates. Filter disks containing 20 µg AHAS inhibitor (20 p. 1 of a 1 mg/ml stock) or 10 µg AHAS inhibitor (20 µl of a 0.2 mg/ml stock) were added to the plate surface. Plates were incubated for 72 hours at 30° C. before visualization of zones of growth inhibition. Clear zones surrounding the AHAS-laced filter disks indicate inhibition of yeast growth. These results suggested that Classic, Ally, and (just herbicides inhibit growth of the yeast strains. Accent, Harmony, and W4189 did not inhibit the industrial yeast strains at the concentrations used in this experiment. (FIG. 2)

#### Example 5

# Resistance of Engineered *S. cerevisiae* Strains Containing an ILV2 Variant Gene to Sulfonylurea Herbicides

This example describes experiments that demonstrate 40 yeast strains, expressing a resistant variant of acetolactate synthase, are resistant to the AHAS inhibitor sulfometuron methyl.

*S. cerevisiae* yeast strains PNY2008 and PNY2010 are described in Example 3. PNY2010 contains the ILV2-410 allele that confers resistance to sulfonylureas.

The yeast strains were initially grown on synthetic complete media supplemented with 2% glucose at  $30^{\circ}$  C. The strains were patched onto either synthetic complete media supplemented with 2% glucose at  $30^{\circ}$  C. or synthetic complete media supplemented with 2% glucose and 12.5 µg/ml sulfometuron methyl (prepared in 10 mM KOH as in Example 4). Plates were incubated for 48 hours at  $30^{\circ}$  C. before visualization. Strain PNY2008 was unable to grow on plates containing sulfometuron methyl, whereas strain PNY2010 grew normally due to the presence of the ILV2-410 allele.

#### Example 6

# Production of Isobutanol in Recombinant S. cerevisiae NYLA84

The purpose of this example is to describe the production of isobutanol in the yeast strain NYLA84. The yeast strain comprises deletions of PDC1, PDC5, and PDC6, genes encoding three isozymes of pyruvate decarboxylase, and deletion of HXK2 encoding hexokinase II. The strain also

contains constructs for heterologous expression of AlsS (acetolactate synthase), KARI (keto acid reductoisomerase), DHAD (dihydroxy acid dehydratase), KivD (ketoisovalerate decarboxylase), and SadB (secondary alcohol dehydrogenase).

Strain Construction

Plasmids pLH468 and pLH475-JEA1 were introduced into NYLA84, described in Example 3, by standard PEG/ lithium acetate-mediated transformation methods. Transformants were selected on synthetic complete medium lacking <sup>10</sup> glucose, histidine and uracil. Ethanol (1% v/v) was used as the carbon source. After three days, transformants were patched to synthetic complete medium lacking histidine and uracil supplemented with both 2% glucose and 0.5% ethanol as carbon sources. Freezer vials were made by dilution of <sup>15</sup> log-phase cultures with 45% glycerol to a final glycerol concentration of 15% (w/v).

Production of Isobutanol

80 ml of synthetic complete medium lacking histidine and uracil supplemented with both 2% glucose and 0.5% ethanol <sup>20</sup> as carbon sources was inoculated with a yeast strain.

Fermentation Conditions:

Medium (final concentration): 6.7 g/L, Yeast Nitrogen Base w/o amino acids (Difco); 2.8 g/L, Yeast Synthetic Drop-out Medium Supplement Without Histidine, Leucine, <sup>25</sup> Tryptophan and Uracil (Sigma Y2001); 20 mL/L of 1% (w/v) L-Leucine; 4 mL/L of 1% (w/v) L-Tryptophan; 1 mL/L ergosterol/tween/ethanol solution; 0.2 mL/L Sigma DF204; 10 g/L glucose

The fermenter was set to control at pH 5.5 with KOH,  $^{30}$  30% dO, temperature 30° C., and airflow of 0.2 SLPM (or, 0.25 vvm). At inoculation, the airflow was set to 0.01 SLPM initially, then increased to 0.2 SLPM once growth was established. Glucose was maintained at 5-15 g/L throughout by manual addition.  $^{35}$ 

Because efficient production of isobutanol with NYLA84 pLH486/pLH475 requires microaerobic conditions to enable redox balance in the biosynthetic pathway, air was continuously supplied to the fermenter at 0.25 vvm. Continuous aeration led to significant stripping of isobutanol from the <sup>40</sup> aqueous phase of the fermenter. To quantify the loss of isobutanol due to stripping, the off-gas from the fermenter was directly sent to a mass spectrometer (Prima dB mass spectrometer, Thermo Electron Corp., Madison, Wis.) to quantify the amount of isobutanol in the gas stream. The <sup>45</sup> isobutanol peaks at mass to charge ratios of 74 or 42 were monitored continuously to quantify the amount of isobutanol in the gas stream.

Glucose and organic acids in the aqueous phase were monitored during the fermentation using HPLC. Glucose <sup>50</sup> was also monitored quickly using a glucose analyzer (YSI, Inc., Yellow Springs, Ohio). Isobutanol in the aqueous phase was quantified by HPLC as described in the General Methods Section herein above after the aqueous phase was removed periodically from the fermenter. The effective titer, <sup>55</sup> corrected for the isobutanol lost due to stripping, was 7.5 g/L. The titer of isobutyric acid was 1.28 g/L. (FIG. **3**)

#### Example 7 (Prophetic):

# Resistance of Engineered *S. cerevisiae* Isobutanologens Containing an IL V2 Variant Gene to Sulfonylurea Herbicides

This example describes experiments that demonstrate 65 yeast strains that contain an engineered isobutanol production pathway which also express a resistant variant of

acetolactate synthase, are resistant to the AHAS inhibitor sulfometuron methyl. Construction of strain NYLA84 is shown in Example 2.

The ILV2-410-loxP71-URA3-loxP66 integration cassette (described in Example 3) is PCR amplified from pLA59:: ILV2-410 template using Phusion DNA polymerase and primers LA686 and LA687 (SEQ ID NOs:119 and 120), producing a ~3 kb product. The PCR product is transformed into strain NYLA84 and plated on synthetic complete media lacking uracil and supplemented with 1% ethanol at 30° C. The loxP71-URA3-loxP66 marker is recycled by transformation with pRS423::PGAL1-cre (SEQ ID NO:121) and plating on synthetic complete media lacking histidine supplemented with 1% ethanol at 30° C. Colonies are patched onto YP (1% galactose) plates at 30° C. to induce URA3 marker excision and are transferred onto YP (1% ethanol) plates at 30° C. for recovery. Removal of the URA3 marker is confirmed by patching colonies from the YP (1% ethanol) plates onto synthetic complete media lacking uracil supplemented with 1% ethanol to verify the absence of growth. The resulting identified strain has the genotype NYLA84 ILV2-410::loxP71/66.

Plasmid vectors pLH468 and pLH475-JEA1 were simultaneously transformed into strain NYLA84 ILV2-410:: loxP71/66 using standard genetic techniques (*Methods in Yeast Genetics*, 2005, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.) and the resulting strain was maintained on synthetic complete media lacking histidine and uracil, and supplemented with 1% ethanol at 30° C.

The yeast strains are initially grown on synthetic complete media lacking histidine and uracil, and supplemented with 1% ethanol at 30° C. After three days, transformants are patched to synthetic complete medium lacking histidine and uracil supplemented with both 2% glucose and 0.5% ethanol as carbon sources.

20 ml of synthetic complete medium lacking histidine and uracil supplemented with both 0.2% glucose and 0.5% ethanol as carbon sources at 30° C. is inoculated with the yeast strain. Each strain is diluted to an initial OD of 0.2 in tubes containing fresh synthetic complete medium lacking histidine and uracil supplemented with both 2% glucose and 0.5% ethanol as carbon sources. The AHAS inhibitor sulfometuron methyl is added to the tubes at concentrations ranging from 0 µg/ml to 50 µg/ml. The tubes are incubated overnight at 30° C. shaking at 220 rpm and are scored the following day for growth. Strains expressing cytosolic acetolactate synthase demonstrate higher resistance to sulfometuron methyl. Isobutanol in the aqueous phase is quantified by HPLC as described in the General Methods Section.

#### Example 8 (Prophetic):

#### Production of Isobutanol in Recombinant *S. cerevisiae* NYLA84 in the Presence of Wildtype Yeast Competitor and Sulfometuron Methyl Herbicide

The purpose of this example is to describe the production of isobutanol in the yeast strain NYLA84 ILV2-410:: 0 loxP71/66 pLH468/pLH475-JEA1 when challenged with a wildtype yeast strain. The AHAS inhibitor sulfometuron methyl is added to prevent or retard growth of the wildtype yeast strain.

Strain Construction

Plasmids pLH468 and pLH475-JEA1 are introduced into NYLA84 ILV2-410::loxP71/66 pLH468/pLH475-JEA1, described in Example 7, by standard PEG/lithium acetate-

mediated transformation methods. Transformants are selected on synthetic complete medium lacking glucose, histidine and uracil. Ethanol (1% v/v) is used as the carbon source. After three days, transformants are patched to synthetic complete medium lacking histidine and uracil supple- $^5$  mented with both 2% glucose and 0.5% ethanol as carbon sources.

Wildtype competitor strain Ethanol Red (Fermentis) is grown is synthetic complete medium supplemented with 2% glucose as carbon source.

Production of Isobutanol

80 ml of synthetic complete medium lacking histidine and uracil supplemented with both 2% glucose and 0.5% ethanol as carbon sources is inoculated with the yeast strain. Fermentation Conditions:

Medium (final concentration): 6.7 g/L, Yeast Nitrogen Base w/o amino acids (Difco); 2.8 g/L, Yeast Synthetic Drop-out Medium Supplement Without Histidine, Leucine Tryptophan and Uracil (Sigma Y2001); 20 mL/L of 1% 20 (w/v) L-Leucine; 4 mL/L of 1% (w/v) L-Tryptophan; 1 mL/L ergosterol/tween/ethanol solution; 0.2 mL/L Sigma DF204; 10 g/L glucose.

Both fermenters are inoculated with NYLA84 pLH486/ pLH475 and Ethanol Red (at 0.5× number of cells as the <sup>25</sup> NYLA84 strain). Sulfometuron methyl is added to one fermenter at a concentration found to be inhibitory (see Example 4). The fermenters are set to control at pH 5.5 with KOH, 30% dO, temperature 30° C., and airflow of 0.2 SLPM (or, 0.25 vvm). At inoculation, the airflow is set to <sup>30</sup> 0.01 SLPM initially, then increased to 0.2 SLPM once growth is established. Glucose is maintained at 5-15 g/L throughout by manual addition.

Because efficient production of isobutanol with NYLA84 pLH486/pLH475 requires microaerobic conditions to enable <sup>35</sup> redox balance in the biosynthetic pathway, air is continuously supplied to the fermenter at 0.25 vvm. Continuous aeration leads to significant stripping of isobutanol from the aqueous phase of the fermenter. To quantify the loss of isobutanol due to stripping, the off-gas from the fermenter is <sup>40</sup> directly sent to a mass spectrometer (Prima dB mass spectrometer, Thermo Electron Corp., Madison, Wis.) to quantify the amount of isobutanol in the gas stream. The isobutanol peaks at mass to charge ratios of 74 or 42 are monitored continuously to quantify the amount of isobuta- <sup>45</sup> nol in the gas stream.

Glucose and organic acids in the aqueous phase are monitored during the fermentation using HPLC. Glucose is also monitored quickly using a glucose analyzer (YSI, Inc., Yellow Springs, Ohio). Isobutanol in the aqueous phase is <sup>50</sup> quantified by HPLC as described in the General Methods Section herein above, after the aqueous phase is removed periodically from the fermenter.

# Example 9 (Prophetic):

Isobutanol Production in an Engineered *S. cerevisiae* Isobutanologens Containing a Heterologous Acetolactate Synthase that is Resistant to Sulfonylurea Herbicides

This example describes experiments that demonstrate yeast strains, which contain an engineered isobutanol production pathway and express a heterologous acetolactate synthase that is resistant to sulfonylurea herbicides, are 65 resistant to the AHAS inhibitor sulfometuron methyl. Construction of strain NYLA84 is shown in Example 2.

The enzyme ALS I (encoded by ilvB) from the enterobacteria Escherichia coli K12, which is intrinsically resistant to sulfonylurea herbicides, is PCR-amplified from E. coli K12 genomic DNA using Phusion DNA polymerase and primers T001 and T002 (SEQ ID NOs:122 and 123). The FBA1 promoter is PCR amplified from BY4700 genomic DNA using Phusion DNA polymerase and primers T003 and T004 (SEQ ID NOs:124 and 125). The FBA1 terminator is PCR amplified from BY4700 genomic DNA using Phusion DNA polymerase and primers T005 and T006 (SEQ ID NOs:126 and 127). The FBA1 promoter is digested with SphI KpnI, the ilvB gene is digested with KpnI NotI, and the FBA1 terminator is digested with NotI BamHI. The three fragments are ligated together and subcloned into vector pLA59 (described in Example 3) via SphI BamHI sites, creating vector pLA59::ilvB (SEQ ID NO:19)

bdh1A::P<sub>FBA1</sub>-ilvB-FBA1t-loxP71-URA3-loxP66 The integration cassette is PCR amplified from pLA59::ilvB template (SEQ ID NO:19) using Phusion DNA polymerase and primers T007 and T008 (SEO ID NOs:128 and 129). The BDH1 portion of each primer was derived from the 5' region 50 bp downstream of the BDH1 start codon and 3' region 50 bp upstream of the stop codon such that integration of the URA3 cassette results in replacement of the BDH1 coding region but leaves the first ~50 bp and the last ~50 bp of the coding region. The PCR product is transformed into strain NYLA84 and plated on synthetic complete media lacking uracil and supplemented with 1% ethanol at 30° C. The loxP71-URA3-loxP66 marker is recycled by transformation with  $pRS423::P_{GAL1}$ -cre (SEQ ID NO:121) and plating on synthetic complete media lacking histidine supplemented with 1% ethanol at 30° C. Colonies are patched onto YP (1% galactose) plates at 30° C. to induce URA3 marker excision and are transferred onto YP (1% ethanol) plates at 30° C. for recovery. Removal of the URA3 marker is confirmed by patching colonies from the YP (1% ethanol) plates onto synthetic complete media lacking uracil supplemented with 1% ethanol to verify the absence of growth. The resulting identified strain has the genotype NYLA84 bdh1::ilvB-loxP71/66.

Plasmid vectors pLH468 and pLH475-JEA1 are simultaneously transformed into strain NYLA84 bdh1::ilvBloxP71/66 using standard genetic techniques (*Methods in Yeast Genetics*, 2005, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.) and the resulting strain is maintained on synthetic complete media lacking histidine and uracil, and supplemented with 1% ethanol at 30° C.

The yeast strains are initially grown on synthetic complete media lacking histidine and uracil, and supplemented with 1% ethanol at 30° C. After three days, transformants are patched to synthetic complete medium lacking histidine and uracil supplemented with both 2% glucose and 0.5% ethanol as carbon sources.

20 ml of synthetic complete medium lacking histidine and
55 uracil supplemented with both 0.2% glucose and 0.5% ethanol as carbon sources at 30° C. is inoculated with the yeast strain. Each strain is diluted to an initial OD of 0.2 in tubes containing fresh synthetic complete medium lacking histidine and uracil supplemented with both 2% glucose and
60 0.5% ethanol as carbon sources. The AHAS inhibitor sulfometuron methyl is added to the tubes at concentrations ranging from 0 µg/ml to 50 µg/ml. The tubes are incubated overnight at 30° C. shaking at 220 rpm and are scored the following day for growth. Strains expressing cytosolic
65 acetolactate synthase demonstrate higher resistance to sulfometuron methyl. Isobutanol in the aqueous phase is quantified by HPLC as described in the General Methods Section.

Enterobacterial ALS enzymes are described in LaRossa and Smul, *J. Bacteriol.* 160(1):391-394 (1984). LaRossa describes ALSI enzymes from *S. typhimurium* and *E. coli* that are resistant to sulfonylurea herbicides.

Materials and Methods for Examples 10 to 20

Yeast synthetic medium w/o amino acids, w/o glucose, w/o ethanol/Tween (2x) 13.4 g/l, Yeast Nitrogen Base w/o amino acids (Difco 0919-15-3); 40 mg/L thiamine; 40 mg/L niacin; 200 ml/L 1M MES buffer, pH=5.5

Supplement as sol. without histidine and uracil (SAAS-1 10 10×): 18.5 g/L, Synthetic complete amino acid dropout (Kaiser)-His, -Ura (Formedium).

Na-acetate stock solution: 3 M sodium acetate solution Glucose stock solution: 250 g/L glucose solution

Inhibitor stock solutions: (1) copper (II) sulfate pentahy- 15 drate: CuSO4.5H2O (MW 249.6 g/mol, CAS Number 7758-99-8): 150 mM; (2) formaldehyde solution (SIGMA F8775, 36.5-38% in H2O, MW 30.03 g/mol, CAS Number 50-00-0): 12.15 M; (3) sodium sulfite (Na2SO3, SIGMA-AL-DRICH 50505, CAS Number 7757-83-7, MW 126.04 20 g/mol): 100 mM in SF11, 500 mM in SF12; (4) bismuth(III) citrate (CAS Number 813-93-4, [O2CCH2C(OH)(CO2) CH2CO2] Bi, MW 398.08 g/mol): saturated solution; (5) sulfometuron methyl (Fluka #34224, CAS Number 74222-97-2, C15H16N4O5S, MW 364.38 g/mol): 10 g/L in 25 DMSO; (6) 4-pyrazolecarboxylic acid (Sigma-Aldrich, #300713, C4H4N2O2, MW: 112.09 g/mol, CAS Number: 37718-11-9: 1.0 M in SF12 (=112 mg/ml (DMSO)); (7) 4-methylpyrazole hydrochloride (Sigma, # M1387, C4H6N2.HCl, MW: 118.56 g/mol, CAS: 56010-88-9): 1.0M 30 in SF12 (=119 mg/ml (DMSO)); (8) pyrazole (Aldrich, # P56607, C3H4N2, MW: 68.08 g/mol, CAS Number: 288-13-1): 0.5 M in SF12 (=34 mg/ml); (9) glyoxylic acid sodium salt monohydrate (HC(O)COONa.H2O, MW: 114.03, CAS Number: 918149-31-2): 0.5 M in SF12 (=57 35 mg/ml); (10) pyrazole (Aldrich, # P56607, C3H4N2, MW: 68.08 g/mol, CAS Number: 288-13-1): 0.5 M (=34 mg/ml) in SF13; (11) trans-cinnamaldehyde (Aldrich #239968, C6H5CH=CHCHO, MW: 132.16 g/mol, CAS: 14371-10-9, d=1.050 g/ml): SF12 and SF13=pure liquid, SF14=20 40 mM in DMSO; (12) 1-bromo-2-butanone (Sigma-Aldrich #243299, C2H5COCH2Br, MW: 151.00 g/mol, CAS: 816-40-0, d=1.479 g/l): SF12 and SF13=pure liquid, SF14=10 mM in DMSO; (13) 4-nitrocinnamaldehyde (predominantly "trans" form, Aldrich #281670, O2NC6H4CH=CHCHO, 45 MW: 177.16 g/mol, CAS: 49678-08-2): SF12=pure substance was weighted and added to the culture

SEED medium: 10,000 mL Yeast synthetic medium w/o aa, w/o glucose, w/o ethanol/Tween (2×); 2.000 mL Supplement aa sol. without histidine and uracil (SAAS-1 10×); 3.200 mL 250 g/L glucose solution (resulting in 40 g/l glucose); 0.046 mL Na-acetate stock solution; 4.754 mL H2O.

# Example 10

# Inhibition of Ethanologen Yeast PNY 827 by Copper(II)

The inhibitory effect of copper (II) on ethanologen yeast PNY 827 was investigated. Therefore a 125 ml aerobic shake flask was prepared with 20 ml SEED medium and inoculated with 1 ml of frozen glycerol stock culture of PNY 827. The culture was inoculated over night at 30° C. and 250 rpm in an Innova Laboratory Shaker (New Brunswick Scientific, Edison, N.J.). Subsequently, a sufficient amount of the seed culture was transferred into shake flasks containing 20 ml of production medium without copper or addition of copper at concentrations of 5 mM, 10 mM and 25 mM, to give an initial OD of approximately 0.1. The cultures were incubated at 250 rpm for 24 h in an Innova Laboratory Shaker (New Brunswick Scientific, Edison, N.J.) and samples of about 1 ml for OD determination withdrawn at designated hours. Optical density was measured with an Ultrospec 3000 spectrophotometer (Pharmacia Biotech) at  $\lambda$ =600 nm. In case cell dry weight concentrations were needed, an OD-DW-correlation of 0.33 gDW/OD was applied. Maximum specific growth rates  $\mu_{max}$  were determined by applying the exponential regression function of Microsoft Excel (Microsoft Office Excel 2003, SP 3). Outliers were discarded until good fit of the regression curve with measurements was confirmed by visual inspection. Parameters of the inhibition kinetics were determined by least square minimization of the differences between measured and predicted  $\mu_{max}$  values. Employed search algorithm was a quasi-Newton method with linear extrapolation from a tangent vector, as implemented in the solver routine of Microsoft Excel (Microsoft Office Excel 2003, SP 3). The solution with 25 mM showed precipitation and was not analyzed. At 5 mM of copper  $\mu_{max}$  was determined to be 0.46 1/h. In the medium containing 10 mM of copper, maximum specific growth rate of  $\mu_{max}$ =0.33 l/h was found. Fitting the data to the "squared inhibition" kinetics yielded parameters of  $\mu^{\circ}_{max}=0.58$  l/h and a K<sub>I</sub> value of K<sub>I</sub>=11 mM (FIG. 4). Decrease in  $\mu_{max}$  with increasing copper (II) concentrations in the medium indicates inhibition of ethanologen veast PNY 827 Inhibition kinetics were used and fitted parameters predict an IC<sub>50</sub> value of 11 mM. Data from the samples is seen in Table 3 below.

TABLE	3
	~

Data for control samples and copper-inhibited experiments.												
sample	time	time time [min] time [h] OD600 dilution [1:x] OD600corr [ ]										
	F1-ctrl-a											
0	8:25	0	0.00	0.154	1	0.087						
1	9:25	60	1.00	0.157	1	0.090						
2	10:40	135	2.25	0.188	1	0.121	0.121					
3	11:40	195	3.25	0.255	1	0.188	0.188					
4	12:40	255	4.25	0.372	1	0.305	0.305					
5	14:25	360	6.00	0.285	5	1.102	1.102					
6	8:05	1420	23.67	0.672	20	12.157						
				F2-ctrl-b	)							
0	8:25	0	0.00	0.154	1	0.087						
1	9:25	60	1.00	0.153	1	0.086						

Data for control samples and copper-inhibited experiments.										
sample	time	time [min]	time [h]	<b>OD</b> 600	dilution [1:x]	OD600corr [ ]	OD600corr [ ]			
2	10:40	135	2.25	0.187	1	0.120	0.120			
3	11:40	195	3.25	0.251	1	0.184	0.184			
4	12:40	255	4.25	0.371	1	0.304	0.304			
5	14:25	360	6.00	0.281	5	1.082	1.082			
6	8:05	1420	23.67	0.647	20	11.657				
				F5-cu-1						
0	8:25	0	0.00	0.336	1	0.097				
1	9:25	60	1.00	0.346	1	0.107	0.107			
2	10:40	135	2.25	0.467	1	0.228	0.228			
3	11:40	195	3.25	0.538	1	0.299	0.299			
4	12:40	255	4.25	0.543	1	0.304				
5	14:25	360	6.00	0.152	5	0.265				
6	18:05	580	9.67	0.163	5	0.320				
6	8:05	1420	23.67	0.171	5	0.360				
				F6-cu-2						
0	8:25	0	0.00	0.389	1	0.115				
1	9:25	60	1.00	0.399	1	0.125				
2	10:40	135	2.25	0.401	1	0.127	0.127			
3	11:40	195	3.25	0.437	1	0.163	0.163			
4	12:40	255	4.25	0.521	1	0.247	0.247			
5	14:25	360	6.00	0.158	5	0.260				
6	18:05	580	9.67	0.166	5	0.300				
6	8:05	1420	23.67	0.202	5	0.480				
				F7-cu-3						

TABLE 3-continued

precipitation

Copper concentrations in the experiments were:

F1-ctrl-a: 0 mM:

F2-ctrl-b: 0 mM:

F5-cu-1: 5 mM: E6-cu-2: 10 mM-

F7-cu-3: 25 mM.

Example 11

# Inhibition of Ethanologen Yeast PNY 827 by Sulfometuron Methyl

The inhibitory effect of the sulfonylurea sulfometuron methyl on ethanologen yeast PNY 827 was investigated. Therefore a 125 ml aerobic shake flask was prepared with 20 ml SEED medium and inoculated with 1 ml of frozen glycerol stock culture of PNY 827. The culture was inocu- 45  $\mu^{\circ}_{max}=0.59$  1/h, respectively (5). lated over night at 30° C. and 250 rpm in an Innova Laboratory Shaker (New Brunswick Scientific, Edison, N.J.). Subsequently, a sufficient amount of the seed culture was transferred into shake flasks containing 20 ml of production medium without sulfometuron methyl or addition of 50 sulfometuron methyl at concentrations of 0.11 mM, 0.16 mM and 0.27, to give an initial OD of approximately 0.1. The cultures were incubated at 250 rpm for 24 h in an Innova Laboratory Shaker (New Brunswick Scientific, Edison, N.J.) and samples of about 1 ml for OD determination withdrawn 55 at designated hours. Optical density was measured with an Ultrospec 3000 spectrophotometer (Pharmacia Biotech) at  $\lambda$ =600 nm. In case cell dry weight concentrations were needed, an OD-DW-correlation of 0.33 gDW/OD was applied. Maximum specific growth rates  $\mu_{max}$  were deter- 60 mined by applying the exponential regression function of Microsoft Excel (Microsoft Office Excel 2003, SP 3). Outliers were discarded until good fit of the regression curve with measurements was confirmed by visual inspection. Parameters of the inhibition kinetics were determined by 65 least square minimization of the differences between measured and predicted  $\mu_{max}$  values. Employed search algorithm

was a quasi-Newton method with linear extrapolation from a tangent vector, as implemented in the solver routine of Microsoft Excel (Microsoft Office Excel 2003, SP 3).

At all three applied concentrations of 0.11 mM, 0.16 mM and 0.27 mM of sulfometuron methyl a significant reduction in specific maximum growth rate was found, yielding  $\mu_{max}$ values of 0.04 l/h, 0.05 l/h and 0.04 l/h, down from uninhibited maximum specific growth in the experiment of

Sulfometuron methyl is poorly water soluble, consequently the compound was administered dissolved in DMSO. In order to make sure the observed inhibition did not result from DMSO, DMSO was added only ad 0.14 mM to a culture and a maximum specific growth rate of 0.56 l/h was found. Follow up experiments with DMSO point to a "squared inhibition" with a  $K_{\tau}$  value of about 16 mM (data not shown). So while DMSO alone seems to have an inhibitory effect on growth, its inhibitory effects at concentrations of 0.06 mM, 0.08 mM and 0.14 mM, as used in the experiments with sulfometuron methyl, can be neglected. Fitting a "hybrid" inhibition kinetics model to the measurements yields values of  $\mu^{\circ}_{max 1}$ =0.55 l/h and  $\mu^{\circ}_{max 2}$ =0.04 l/h. Not sufficient data are available for accurate determination of  $K_{r}$ , but from the curve shape it can be concluded that the K<sub>I</sub> value is significantly below 0.1 mM. The "hybrid" inhibition kinetics model predicts an overall observable  $\mu^{\circ}_{max}$  of 0.59 l/h. Due to the underdetermined K<sub>1</sub> value, the IC<sub>50</sub> value cannot reliably be determined, but it can be concluded that it is significantly lower than 0.1 mM. Data from the samples is seen in Table 4 below.

TABLE 4

Data for control samples and sulfometuron methyl-inhibited experiments. Sulfometuron methyl concentrations in the experiments were: F1-ctrl-a: 0 mM; F2-ctrl-b: 0 mM; F12-sm-1: 0.11 mM; F13-sm-2: 0.16 mM; F14-sm-3: 0.27 mM.

sam- ple	time	time [min]	time [h]	<b>OD</b> 600	dilu- tion [1:x]	OD600corr []	OD600corr []			
F1-ctrl-a 10										
0	8:25	0	0.00	0.154	1	0.087				
1	9:25	60	1.00	0.157	1	0.090				
2	10:40	135	2.25	0.188	1	0.121	0.121			
3	11:40	195	3.25	0.255	1	0.188	0.188			
4	12:40	255	4.25	0.372	1	0.305	0.305	1.5		
5	14:25	360	6.00	0.285	5	1.102	1.102	15		
6	8:05	1420	23.67	0.672	20	12.157				
				F2-ct	rl-b					
0	8:25	0	0.00	0.154	1	0.087				
1	9:25	60	1.00	0.153	1	0.086				
2	10:40	135	2.25	0.187	1	0.120	0.120	20		
3	11:40	195	3.25	0.251	1	0.184	0.184			
4	12:40	255	4.25	0.371	1	0.304	0.304			
5	14:25	360	6.00	0.281	5	1.082	1.082			
6	8:05	1420	23.67	0.647	20	11.657				
				F12-s	m-1			_		
								25		
0	8:25	0	0.00	0.155	1	0.088	0.000			
1	9:25	60	1.00	0.160	1	0.093	0.093			
2	10:40	135	2.25	0.165	1	0.098	0.098			
3	11:40	195	3.25	0.170	1	0.103	0.103			
4	12:40	255	4.25	0.175	1	0.108	0.108			
2	14:25	360	6.00	0.182	1	0.115	0.115	30		
0	17:50	202	9.42	0.222	1	0.155				
/	8:05	1420	23.07	0.272 E13-s		1.037				
				115-8	III-2			•		
0	8:25	0	0.00	0.158	1	0.091				
1	9:25	60	1.00	0.160	1	0.093	0.093	25		
2	10:40	135	2.25	0.166	1	0.099	0.099	35		
3	11:40	195	3.25	0.171	1	0.104	0.104			
4	12:40	255	4.25	0.175	1	0.108	0.108			
5	14:25	360	6.00	0.184	1	0.117	0.117			
6	17:50	565	9.42	0.219	1	0.152				
7	8:05	1420	23.67	0.467	1	0.400		10		
				F14-s	m-3			. 40		
0	8:25	0	0.00	0.154	1	0.087	0.000			
1	9:25	60	1.00	0.157	1	0.090	0.090			
2	10:40	135	2.25	0.164	1	0.097	0.097			
3	11:40	195	3.25	0.164	1	0.097	0.097	15		
4	12:40	200	4.25	0.100	1	0.099	0.099	43		
5	14:20	565	0.00	0.177	1	0.110	0.110			
7	8:05	1420	9.42 23.67	0.201	1	0.134	0.134			
/	0.05	1420	25.07	0.500	1	0.239	0.235			

#### Example 12

# Inhibition of Ethanologen Yeast PNY 827 by Sulfite

The inhibitory effect of sulfite on ethanologen yeast PNY 827 was investigated. Therefore a 125 ml aerobic shake flask was prepared with 20 ml SEED medium and inoculated with 1 ml of frozen glycerol stock culture of PNY 827. The culture was inoculated over night at 30° C. and 250 rpm in an Innova Laboratory Shaker (New Brunswick Scientific, Edison, N.J.). Subsequently, a sufficient amount of the seed culture was transferred into shake flasks containing 20 ml of production medium without sulfite or addition of sulfite at concentrations of 1 mM, 2 mM, 5 mM, 10 mM, 20 mM and 50 mM, to give an initial OD of approximately 0.1. The cultures were incubated at 250 rpm for 24 h in an Innova

Laboratory Shaker (New Brunswick Scientific, Edison, N.J.) and samples of about 1 ml for OD determination withdrawn at designated hours. Optical density was measured with an Ultrospec 3000 spectrophotometer (Pharmacia Biotech) at  $\lambda$ =600 nm. In case cell dry weight concentrations were needed, an OD-DW-correlation of 0.33 gDW/OD was applied. Maximum specific growth rates  $\mu_{max}$  were determined by applying the exponential regression function of Microsoft Excel (Microsoft Office Excel 2003, SP 3). Outliers were discarded until good fit of the regression curve with measurements was confirmed by visual inspection. Parameters of the inhibition kinetics were determined by least square minimization of the differences between measured and predicted  $\mu_{max}$  values. Employed search algorithm was a quasi-Newton method with linear extrapolation from a tangent vector, as implemented in the solver routine of Microsoft Excel (Microsoft Office Excel 2003, SP 3).

Sulfite concentrations of 1 mM, 2 mM, 5 mM, 10 mM, 20 mM and 50 mM resulted in maximum specific growth rates of 0.59 l/h, 0.54 l/h, 0.33 l/h, 0.23 l/h, 0.18 l/h and 0.14 l/h, respectively, indicating significant inhibitory effect of sulfite on ethanologen yeast PNY 807. Fitting the measured data to the "hybrid" inhibition kinetics model, values of  $\mu^{\circ}_{max}$  1=0.46 l/h,  $\mu^{\circ}_{max}$  2=0.15 l/h and KI=4.4 mM were determined. The "hybrid" inhibition kinetics model predicts an overall observable  $\mu^{\circ}_{max}$ =0.61 l/h and an IC50 value of 6.2 mM. Measured  $\mu_{max}$  values and fitted dependency of  $\mu_{max}$  on the concentration of sulfite in the medium is depicted in FIG. 6. Data from the samples is seen in Table 5 below.

TABLE 5

Data for control samples and sulfite-inhibited experiments. Sulfite concentrations in the experiments were: F1-ctrl-a: 0 mM; F2-ctrl-b: 0 mM; F8-su-1: 1 mM; F9-su-2: 2 mM; F10- su-3: 5 mM; F11-su-4: 10 mM; SF12-F1-CtrlA: 0 mM; SF12- F2-ctrlB: 0M: F3-su-1: 20 mM: F4-su-2: 50 mM.
dilu-

40	sam- ple	time	time [min]	time [h]	OD600	tion [1:x]	OD600corr []	OD600corr []
					F1-cts	l-a		
	0	8:25	0	0.00	0.154	1	0.087	
	1	9:25	60	1.00	0.157	1	0.090	
	2	10:40	135	2.25	0.188	1	0.121	0.121
45	3	11:40	195	3.25	0.255	1	0.188	0.188
	4	12:40	255	4.25	0.372	1	0.305	0.305
	5	14:25	360	6.00	0.285	5	1.102	1.102
	6	8:05	1420	23.67	0.672	20	12.157	
					F2-cti	l-b		
50	0	8.75	0	0.00	0.154	1	0.087	
50	1	0.25	60	1.00	0.154	1	0.087	
	2	9.23	135	2.00	0.133	1	0.080	0.120
	2	11.40	105	2.25	0.167	1	0.120	0.120
	3	12:40	255	1.25	0.251	1	0.184	0.184
	5	14.25	360	6.00	0.371	5	1.082	1.082
	6	8:05	1420	23.67	0.201	20	11.657	1.062
22	0	0.05	1-120	25.07	F8-su	-1	11.057	
	0	8:25	0	0.00	0.148	1	0.081	
	1	9:25	60	1.00	0.149	1	0.082	
	2	10:40	135	2.25	0.188	1	0.121	0.121
60	3	11:40	195	3.25	0.252	1	0.185	0.185
00	4	12:40	255	4.25	0.374	1	0.307	0.307
	5	14:25	360	6.00	0.285	5	1.102	1.102
	6	8:05	1420	23.67	0.622	20	11.157	
					F9-su	-2		
	Ο	8.25	0	0.00	0.154	1	0.087	
65	1	9.25	60	1.00	0.153	1	0.086	
	2	10:40	135	2.25	0.135	1	0.120	0.120
	~	20110	100		0.107		5.1E0	0.120

#### TABLE 5-continued

Data for control samples and sulfite-inhibited experiments. Sulfite concentrations in the experiments were: F1-ctrl-a: 0 mM; F2-ctrl-b: 0 mM; F8-su-1: 1 mM; F9-su-2: 2 mM; F10su-3: 5 mM; F11-su-4: 10 mM; SF12-F1-CtrlA: 0 mM; SF12-F2-ctrlB: 0M; F3-su-1: 20 mM; F4-su-2: 50 mM.

sam- ple	time	time [min]	time [h]	<b>OD</b> 600	dilu- tion [1:x]	OD600corr []	OD600corr []	
3	11:40	195	3.25	0.240	1	0.173	0.173	10
4	12:40	255	4.25	0.332	1	0.265	0.265	
5	14:25	360	6.00	0.242	5	0.887	0.887	
6	8:05	1420	23.67	0.675 E10 c	20	12.217		
				F10-8	u-3			
0	8:25	0	0.00	0.155	1	0.088		15
1	9:25	60	1.00	0.157	1	0.090		
2	10:40	135	2.25	0.183	1	0.116	0.116	
3	11:40	195	3.25	0.220	1	0.153	0.153	
4	12:40	255	4.25	0.275	1	0.208	0.208	
5	14:25	575	6.00	0.459	1	0.392	0.392	20
0	18:00	1/20	9.58	0.579	20	2.572		
/	0.05	1420	23.07	6.650 F11-s	u-4	11./1/		
0	8:25	0	0.00	0.159	1	0.092		
1	9:25	60	1.00	0.152	1	0.085	0.085	25
2	10:40	135	2.25	0.178	1	0.111	0.111	25
3	11:40	195	3.25	0.207	1	0.140	0.140	
4	12:40	233	4.25	0.242	1	0.175	0.175	
5	14:23	575	0.00	0.333	5	1 102	0.208	
6	8:05	1420	23.67	0.285	20	11 437		
0	0.05	1720	25.07	SF12-F1	-Ctrl-A	11.457		30
								50
0	8:55	0	0.00	0.156	1	0.089		
1	10:25	90	1.50	0.170	1	0.103		
2	11:35	160	2.67	0.222	1	0.155	0.155	
3	12:55	240	4.00	0.350	1	0.283	0.283	
4	8:30	1305	2.22	0.220	20	11 677	0.777	35
0	0.50	1595	23.23	SF12-F2	-ctrl-B	11.077		
0	8:55	0	0.00	0.162	1	0.095		
1	10:25	90	1.50	0.172	1	0.105		
2	11:35	160	2.67	0.223	1	0.156	0.156	40
3	12:55	240	4.00	0.354	1	0.287	0.287	-10
4	14:15	320	5.33	0.228	20	0.817	0.817	
3	8:50	1393	23.23	0.007 F3-si	1-1	12.057		
				10 0				
0	8:55	0	0.00	0.154	1	0.087		
1	10:25	90	1.50	0.166	1	0.099		45
2	11:35	160	2.67	0.182	1	0.115	0.115	
3	12:55	240	4.00	0.217	1	0.150	0.150	
4	14:15	320	5.33	0.249	1	0.182	0.182	
2	15:40	405	6.75	0.293	1	0.226	0.226	
07	18:20	205	9.42	0.403	20	0.396	0.396	50
/	0.50	1393	23.23	0.349 F4-si	1-2	9.097		50
				2.0				
0	8:55	0	0.00	0.155	1	0.088		
1	10:25	90	1.50	0.162	1	0.095		
2	11:35	160	2.67	0.183	1	0.116	0.116	
3	12:55	240	4.00	0.202	1	0.135	0.135	55
4	14:15	320 405	5.55	0.228	1	0.101	0.161	
5	18.20	400 565	0.75	0.239	1	0.192	0.192	
7	8:30	1395	23.25	0.600	20	10.717	0.293	
'	0.50	1000	20.20	0.000	20	10./1/		

# Example 13

#### Inhibition of Ethanologen Yeast PNY 827 by Formaldehyde

The inhibitory effect of formaldehyde on ethanologen yeast PNY 827 was investigated. Therefore a 125 ml aerobic

shake flask was prepared with 20 ml SEED medium and inoculated with 1 ml of frozen glycerol stock culture of PNY 827. The culture was inoculated over night at 30° C. and 250 rpm in an Innova Laboratory Shaker (New Brunswick Scientific, Edison, N.J.). Subsequently, a sufficient amount of the seed culture was transferred into shake flasks containing 20 ml of production medium without formaldehyde or addition of formaldehyde at concentrations of 1 mM, 2 mM, 5 mM and 10 mM, to give an initial OD of approximately 0.1. The cultures were incubated at 250 rpm for 24 h in an Innova Laboratory Shaker (New Brunswick Scientific, Edison, N.J.) and samples of about 1 ml for OD determination withdrawn at designated hours. Optical density was measured with an Ultrospec 3000 spectrophotometer (Pharmacia Biotech) at  $\lambda$ =600 nm. In case cell dry weight concentrations were needed, an OD-DW-correlation of 0.33 gDW/OD was applied. Maximum specific growth rates  $\mu_{max}$  were determined by applying the exponential regression function of Microsoft Excel (Microsoft Office Excel 2003, SP 3). Outliers were discarded until good fit of the regression curve with measurements was confirmed by visual inspection. Parameters of the inhibition kinetics were determined by least square minimization of the differences between measured and predicted  $\mu_{max}$  values. Employed search algorithm was a quasi-Newton method with linear extrapolation from a tangent vector, as implemented in the solver routine of Microsoft Excel (Microsoft Office Excel 2003, SP 3).

<sup>30</sup> With the investigated formaldehyde concentrations of 1 mM, 2 mM, 5 mM and 10 mM, corresponding maximum specific growth rates of PNY827 were 0.28 l/h, 0.08 l/h, 0.00 l/h (no growth), and 0.00 l/h (no growth), respectively. Mumax values determined without inhibitor addition were 0.59 l/h and 0.59 l/h. Fitting the measured data to the "squared inhibition" kinetics model, a K<sub>I</sub> value of K<sub>I</sub>=904  $\mu$ M ( $\mu^{o}_{max}$ =0.59 l/h) was found, indicating a very strong inhibition of *S. cerevisiae* by formaldehyde. The derived ICSO value is 0.9 mM. Measured  $\mu_{max}$  values and fitted dependency of  $\mu_{max}$  on the concentration of formaldehyde in the medium is depicted in FIG. 7. Data from the samples is seen in Table 6 below.

TABLE 6
Data for control samples and formaldehyde-inhibited experiments.

	Formaldehyde concentrations in the experiments were: F1- ctrl-a: 0 mM; F2-ctrl-b: 0 mM; F16-fa-1: 1 mM; F17-fa-2: 2 mM; F18-fa-3: 5 mM; F19-fa-4: 10 mM.											
50	sam- ple	time	time [min]	time [h]	<b>OD</b> 600	dilu- tion [1:x]	OD600corr []	OD600corr []				
					F1-ct	rl-a						
	0	8:25	0	0.00	0.154	1	0.087					
55	1	9:25	60	1.00	0.157	1	0.090					
55	2	10:40	135	2.25	0.188	1	0.121	0.121				
	3	11:40	195	3.25	0.255	1	0.188	0.188				
	4	12:40	255	4.25	0.372	1	0.305	0.305				
	5	14:25	360	6.00	0.285	5	1.102	1.102				
	6	8:05	1420	23.67	0.672	20	12.157					
60					F2-cti	:l-b						
00												
	0	8:25	0	0.00	0.154	1	0.087					
	1	9:25	60	1.00	0.153	1	0.086					
	2	10:40	135	2.25	0.187	1	0.120	0.120				
	3	11:40	195	3.25	0.251	1	0.184	0.184				
	4	12:40	255	4.25	0.371	1	0.304	0.304				
65	5	14:25	360	6.00	0.281	5	1.082	1.082				
	6	8:05	1420	23.67	0.647	20	11.657					

TABLE 6-continued

Data for control samples and formaldehyde-inhibited experiments.
Formaldehyde concentrations in the experiments were: F1-
ctrl-a: 0 mM; F2-ctrl-b: 0 mM; F16-fa-1: 1 mM; F17-fa-2:
2 mM; F18-fa-3: 5 mM; F19-fa-4: 10 mM.

dilu-

sam- ple	time	time [min]	time [h]	<b>OD6</b> 00	tion [1:x]	OD600corr []	OD600corr [ ]	_	
F16-fa-1									
0	8:35	0	0.00	0.157	1	0.090			
1	9.35	60	1.00	0.157	1	0.090			
2	10:50	135	2.25	0.165	1	0.098			
3	11:50	195	3.25	0.182	1	0.115	0.115		
4	12:50	255	4.25	0.206	1	0.139	0.139	15	
5	14:35	360	6.00	0.285	1	0.218	0.218	15	
6	17:40	545	9.08	0.655	1	0.588	0.588		
7	8:20	1425	23.75	0.657	20	11.857			
				F17-f	à-2			-	
0	8.35	0	0.00	0.158	1	0.091			
ĩ	9:35	60	1.00	0.158	î	0.091		20	
2	10:50	135	2.25	0.159	î	0.092			
3	11:50	195	3.25	0.160	1	0.093	0.093		
4	12:50	255	4.25	0.165	1	0.098	0.098		
5	14:35	360	6.00	0.177	1	0.110	0.110		
6	17:40	545	9.08	0.211	1	0.144	0.144		
7	8:20	1425	23.75	0.382	20	6.357		25	
				F18-f	à-3			_	
<u> </u>	0.25	0	0.00	0.157	1	0.000			
1	8:35	0	0.00	0.157	1	0.090			
1	9:33	125	1.00	0.162	1	0.093			
2	11:50	105	3.25	0.101	1	0.094	0.000	20	
1	12:50	255	1.25	0.155	1	0.090	0.090	30	
5	14.35	360	6.00	0.155	1	0.088	0.088		
6	17:40	545	9.08	0.155	1	0.089	0.089		
7	8.20	1425	23.75	0.160	1	0.093	0.005		
,	0.20	1 125	25175	F19-f	à-4	01035		_	
								35	
0	8:35	0	0.00	0.162	1	0.095			
1	9:35	60	1.00	0.168	1	0.101			
2	10:50	135	2.25	0.164	1	0.097			
3	11:50	195	3.25	0.163	1	0.096	0.096		
4	12:50	255	4.25	0.160	1	0.093	0.093		
5	14:35	360	6.00	0.161	1	0.094	0.094	40	
6	17:40	545	9.08	0.162	1	0.095	0.095	-10	
7	8:20	1425	23.75	0.168	1	0.101			

# Example 14

# Inhibition of Ethanologen Yeast PNY 827 by 4-pyrazolecarboxylic acid

The inhibitory effect of 4-nitrocinnamaldehyde (predominantly trans) on ethanologen yeast PNY 827 was investigated. Therefore a 125 ml aerobic shake flask was prepared with 20 ml SEED medium and inoculated with 1 ml of frozen glycerol stock culture of PNY 827. The culture was inoculated over night at 30° C. and 250 rpm in an Innova Laboratory Shaker (New Brunswick Scientific, Edison, N.J.). Subsequently, a sufficient amount of the seed culture was transferred into shake flasks containing 20 ml of production medium without 4-pyrazolecarboxylic acid or addition of 4-nitrocinnamaldehyde at concentrations of 1 mM and 50 mM, to give an initial OD of approximately 0.1. The cultures were incubated at 250 rpm for 24 h in an Innova Laboratory Shaker (New Brunswick Scientific, Edison, N.J.) and samples of about 1 ml for OD determination withdrawn at designated hours. Optical density was measured with an Ultrospec 3000 spectrophotometer (Pharmacia Biotech) at  $\lambda$ =600 nm. In case cell dry weight concentrations were needed, an OD-DW-correlation of 0.33 gDW/OD was

applied. Maximum specific growth rates  $\mu_{max}$  were determined by applying the exponential regression function of Microsoft Excel (Microsoft Office Excel 2003, SP 3). Outliers were discarded until good fit of the regression curve with measurements was confirmed by visual inspection. Parameters of the inhibition kinetics were determined by least square minimization of the differences between measured and predicted  $\mu_{max}$  values. Employed search algorithm was a quasi-Newton method with linear extrapolation from 0 a tangent vector, as implemented in the solver routine of Microsoft Excel (Microsoft Office Excel 2003, SP 3).

The inhibitory effect of 4-pyrazolecarboxylic acid (PA) was investigated at 1 mM and 50 mM. 4-pyrazolecarboxylic acid was administered as a DMSO solution, resulting in <sup>5</sup> DMSO concentrations in the cell suspension of 14 mM and 704 mM, respectively. Assuming an additive effect of 4-pyrazolecarboxylic acid and DMSO inhibition, observed maximum specific growth rate of the two cultures was corrected by 0.00 l/h and 0.27 l/h due to the effect of DMSO, resulting in 4-pyrazolecarboxylic acid-based mumax values of 0.59 l/h and 0.50 l/h derived from the observed values of 0.59 l/h and 0.23 l/h, respectively. Fitting the data to the "squared inhibition" kinetics (observed maximum specific growth rates without inhibitor addition were 0.59 l/h, 0.59 l/h, 0.60 l/h and 0.62 l/h) yielded parameters of  $\mu^{\circ}_{max}$ =0.60 1/h and a KI value of KI=100 mM (FIG. 8), indicating only weak inhibitory effects of 4-pyrazolecarboxylic acid. Data from the samples is seen in Table 7 below.

TABLE 7

	inhibited experiments. 4-pyrazolecarboxylic acid										
35		concer F2-c	itration trl-B: (	s in the ) mM; I	experime 12-pa-1:	nts wer 1 mM;	e: F1-Ctrl-A: ( F13-pa-2: 50 r	) mM; nM.			
	sam- ple	time	time [min]	time [h]	<b>OD</b> 600	dilu- tion [1:x]	OD600corr []	OD600corr []			
40					F1-Ct	l-A					
40											
	0	8:55	0	0.00	0.156	1	0.089				
	1	10:25	90	1.50	0.170	1	0.103				
	2	11:35	160	2.67	0.222	1	0.155	0.155			
	3	12:55	240	4.00	0.350	1	0.283	0.283			
45	4	14:15	320	5.33	0.220	5	0.777	0.777			
45	6	8:30	1395	23.25	0.648	20	11.677				
					F2-ctr	п-В					
	~	0.55	0	0.00	0.1.02	1	0.005				
	1	8:55	0	1.50	0.102	1	0.095				
	1	10:25	160	1.50	0.172	1	0.105	0.156			
50	2	11:55	240	2.07	0.225	1	0.130	0.130			
30	2	12:55	240	4.00	0.334	5	0.287	0.287			
	5	8.30	1305	2.25	0.228	20	12.057	0.817			
	5	8.50	1595	23.25	E12-n	a_1	12.057				
					112-p	a-1					
	0	8.55	0	0.00	0.156	1	0.089				
= =	ĩ	10:25	90	1.50	0.172	1	0.105				
33	2	11.35	160	2 67	0.227	1	0.160	0.160			
	3	12:55	240	4.00	0.361	1	0.294	0.294			
	4	14:15	320	5.33	0.217	5	0.762	0.762			
	5	8:30	1395	23.25	0.659	20	11.897				
					F13-p	a-2					
60											
00	0	8:55	0	0.00	0.157	1	0.090				
	1	10:25	90	1.50	0.167	1	0.100	0.100			
	2	11:35	160	2.67	0.192	1	0.125	0.125			
	3	12:55	240	4.00	0.234	1	0.167	0.167			
	4	14:15	320	5.33	0.305	1	0.238	0.238			
65	5	3:40	405	6.75	0.405	1	0.338	0.338			
65	6	8:30	1395	23.25	0.655	20	11.817				

#### Example 15

#### Inhibition of Ethanologen Yeast PNY 827 by 4-methylpyrazole hydrochloride

The inhibitory effect of 4-methylpyrazole hydrochloride on ethanologen yeast PNY 827 was investigated. Therefore a 125 ml aerobic shake flask was prepared with 20 ml SEED medium and inoculated with 1 ml of frozen glycerol stock culture of PNY 827. The culture was inoculated over night at 30° C. and 250 rpm in an Innova Laboratory Shaker (New Brunswick Scientific, Edison, N.J.). Subsequently, a sufficient amount of the seed culture was transferred into shake flasks containing 20 ml of production medium without 4-methylpyrazole hydrochloride or addition of 4-methyl- 1 pyrazole hydrochloride at concentrations of 1 mM and 30 mM, to give an initial OD of approximately 0.1. The cultures were incubated at 250 rpm for 24 h in an Innova Laboratory Shaker (New Brunswick Scientific, Edison, N.J.) and samples of about 1 ml for OD determination withdrawn at 2 designated hours. Optical density was measured with an Ultrospec 3000 spectrophotometer (Pharmacia Biotech) at  $\lambda$ =600 nm. In case cell dry weight concentrations were needed, an OD-DW-correlation of 0.33 gDW/OD was applied. Maximum specific growth rates  $\mu_{max}$  were deter- 2 mined by applying the exponential regression function of Microsoft Excel (Microsoft Office Excel 2003, SP 3). Outliers were discarded until good fit of the regression curve with measurements was confirmed by visual inspection. Parameters of the inhibition kinetics were determined by 3 least square minimization of the differences between measured and predicted  $\mu_{max}$  values. Employed search algorithm was a quasi-Newton method with linear extrapolation from a tangent vector, as implemented in the solver routine of Microsoft Excel (Microsoft Office Excel 2003, SP 3). 35

The inhibitory effect of 4-methylpyrazole hydrochloride was investigated at 1 mM and 30 mM. 4-methylpyrazole hydrochloride was administered as a DMSO solution, resulting in DMSO concentrations in the cell suspension of 14 mM and 423 mM, respectively. Assuming an additive effect 40 of 4-methylpyrazole hydrochloride and DMSO inhibition, observed maximum specific growth rate of the two cultures was corrected by 0.00 1/h and 0.14 1/h due to the effect of DMSO, resulting in 4-methylpyrazole hydrochloride-based mumax values of 0.48 1/h and 0.38 1/h derived from the 45 observed values of 0.48 l/h and 0.24 l/h, respectively. Fitting the data to the "squared inhibition" kinetics (observed maximum specific growth rates without inhibitor addition were 0.59 l/h, 0.59 l/h, 0.60 l/h and 0.62 l/h) yielded parameters of  $\mu^{o}_{max}$ =0.58 l/h and a K<sub>I</sub> value of K<sub>I</sub>=41 mM, 50 indicating inhibitory effects of 4-methylpyrazole hydrochloride (9). Data from the samples is seen in Table 8 below.

TABLE 8

	Data fo. acid-in acid c mM; F2	r contro hibited oncentr -ctrl-B:	experin experin ations 0 mM	les and 4- nents. 4-n in the exp ; F14-mp-	methylp nethylpy eriment 1:1 ml	yrazole hydrod vrazole hydrocl s were: F1-Ctr M; F15-mp-2:	zhloride hloride l-A: 0 30 mM.
sam- ple	time	time [min]	time [h]	OD600	dilu- tion [1:x]	OD600corr []	OD600corr []
				F1-Ct	rl-A		
0	8:55	0	0.00	0.156	1	0.089	
1	10:25	90	1.50	0.170	1	0.103	
2	11:35	160	2.67	0.222	1	0.155	0.155

72

TABLE 8-continued	
Data for control samples and 4-methylpyrazole hydrochl	oride

5		acid-in acid c mM; F2	nhibited concenti 2-ctrl-B	experin rations i : 0 mM	nents. 4-n in the exp ; F14-mp-	nethylpy eriment 1: 1 ml	vrazole hydroc s were: F1-Ctr M; F15-mp-2:	hloride 1-A: 0 30 mM.
	sam- ple	time	time [min]	time [h]	<b>OD</b> 600	dilu- tion [1:x]	OD600corr []	OD600corr []
0	3 4 6	12:55 14:15 8:30	240 320 1395	4.00 5.33 23.25	0.350 0.220 0.648 F2-ctr	1 5 20	0.283 0.777 11.677	0.283 0.777
5	0 1 2 3 4 5	8:55 10:25 11:35 12:55 14:15 8:30	0 90 160 240 320 1395	0.00 1.50 2.67 4.00 5.33 23.25	0.162 0.172 0.223 0.354 0.228 0.667 F14-m	1 1 1 5 20 1p-1	0.095 0.105 0.156 0.287 0.817 12.057	0.156 0.287 0.817
5	0 1 2 3 4 5	8:55 10:25 11:35 12:55 14:15 8:30	0 90 160 240 320 1395	0.00 1.50 2.67 4.00 5.33 23.25	0.163 0.174 0.221 0.342 0.203 0.626 F15-m	1 1 1 5 20 1p-2	$\begin{array}{c} 0.096 \\ 0.107 \\ 0.154 \\ 0.275 \\ 0.692 \\ 11.237 \end{array}$	0.107 0.154 0.275 0.692
0	0 1 2 3 4 5 6	8:55 10:25 11:35 12:55 14:15 3:40 8:30	0 90 160 240 320 405 1395	$\begin{array}{c} 0.00\\ 1.50\\ 2.67\\ 4.00\\ 5.33\\ 6.75\\ 23.25\end{array}$	0.161 0.172 0.189 0.237 0.305 0.434 0.729	1 1 1 1 1 1 20	$\begin{array}{c} 0.094 \\ 0.105 \\ 0.122 \\ 0.170 \\ 0.238 \\ 0.367 \\ 13.297 \end{array}$	0.105 0.122 0.170 0.238 0.367

#### Example 16

#### Inhibition of Ethanologen Yeast PNY 827 by Glyoxylic Acid

The inhibitory effect of glyoxylic acid on ethanologen yeast PNY 827 was investigated. Therefore a 125 ml aerobic shake flask was prepared with 20 ml SEED medium and inoculated with 1 ml of frozen glycerol stock culture of PNY 827. The culture was inoculated over night at 30° C. and 250 rpm in an Innova Laboratory Shaker (New Brunswick Scientific, Edison, N.J.). Subsequently, a sufficient amount of the seed culture was transferred into shake flasks containing 20 ml of production medium without glyoxylic acid or addition of glyoxylic acid at concentrations of 10 mM and 50 mM, to give an initial OD of approximately 0.1. The cultures were incubated at 250 rpm for 24 h in an Innova Laboratory Shaker (New Brunswick Scientific, Edison, N.J.) and samples of about 1 ml for OD determination withdrawn at designated hours. Optical density was measured with an Ultrospec 3000 spectrophotometer (Pharmacia Biotech) at  $\lambda$ =600 nm. In case cell dry weight concentrations were needed, an OD-DW-correlation of 0.33 gDW/OD was applied. Maximum specific growth rates  $\mu_{max}$  were determined by applying the exponential regression function of Microsoft Excel (Microsoft Office Excel 2003, SP 3). Outliers were discarded until good fit of the regression curve with measurements was confirmed by visual inspection. Parameters of the inhibition kinetics were determined by 55 least square minimization of the differences between measured and predicted  $\mu_{max}$  values. Employed search algorithm was a quasi-Newton method with linear extrapolation from

a tangent vector, as implemented in the solver routine of Microsoft Excel (Microsoft Office Excel 2003, SP 3).

The inhibitory effect of glyoxylic acid was investigated at 10 mM and 50 mM. At the two concentrations, mumax values of 0.59 l/h and 0.55 l/h were found, respectively. Fitting the data to the "squared inhibition" kinetics (observed maximum specific growth rates without inhibitor addition were 0.59 l/h, 0.59 l/h, 0.60 l/h and 0.62 l/h) yielded parameters of  $\mu^{o}_{max}$ =0.60 l/h and a K<sub>*I*</sub> value of K<sub>*I*</sub>=168 mM, indicating a weak inhibitory effect of extracellular glyoxylic acid on growth of ethanologen yeast (FIG. **10**.). Data from the samples is seen in Table 9 below.

TABLE 9

Gl	yoxylic mM; F	acid co <u>2-ctrl-1</u>	oncentra 3: 0 mN	tions in the first the first the first the first the first second	ryne ac 1e expei - <u>1: 10 n</u>	riments were: 1 nM; F19-ga-2:	F1-Ctrl-A: 50 mM.	
sam- ple	time	time [min]	time [h]	<b>OD</b> 600	dilu- tion [1:x]	OD600corr []	OD600corr []	20
				F1-Ct	rl-A			
0 1 2 3 4 6	8:55 10:25 11:35 12:55 14:15 8:30	0 90 160 240 320 1395	0.00 1.50 2.67 4.00 5.33 23.25	0.156 0.170 0.222 0.350 0.220 0.648 F2-ctr	1 1 1 5 20 1-B	$\begin{array}{c} 0.089 \\ 0.103 \\ 0.155 \\ 0.283 \\ 0.777 \\ 11.677 \end{array}$	0.155 0.283 0.777	25
0 1 2 3 4 5	8:55 10:25 11:35 12:55 14:15 8:30	0 90 160 240 320 1395	0.00 1.50 2.67 4.00 5.33 23.25	0.162 0.172 0.223 0.354 0.228 0.667 F18-g	1 1 1 5 20 (a-1	0.095 0.105 0.156 0.287 0.817 12.057	0.156 0.287 0.817	30
0 1 2 3 4 5	8:55 10:25 11:35 12:55 14:15 8:30	0 90 160 240 320 1395	0.00 1.50 2.67 4.00 5.33 23.25	0.161 0.175 0.226 0.361 0.217 0.634 F19-g	1 1 1 5 20 (a-2	0.094 0.108 0.159 0.294 0.762 11.397	0.159 0.294 0.762	40
0 1 2 3 4 5	8:55 10:25 11:35 12:55 14:15 8:30	0 90 160 240 320 1395	0.00 1.50 2.67 4.00 5.33 23.25	0.157 0.173 0.222 0.337 0.199 0.655	1 1 1 5 20	0.090 0.106 0.155 0.270 0.672 11.817	0.155 0.270 0.672	45

# Example 17

# Inhibition of Ethanologen Yeast PNY 827 by Pyrazole

The inhibitory effect of pyrazole on ethanologen yeast PNY 827 was investigated. Therefore a 125 ml aerobic shake flask was prepared with 20 ml SEED medium and inoculated with 1 ml of frozen glycerol stock culture of PNY 827. The culture was inoculated over night at 30° C. and 250 rpm in an Innova Laboratory Shaker (New Brunswick Scientific, Edison, N.J.). Subsequently, a sufficient amount of the seed culture was transferred into shake flasks containing 20 ml of production medium without pyrazole or addition of pyrazole at concentrations of 1 mM, 5 mM, 10 mM, 25 mM and 50 mM, to give an initial OD of approximately 0.1. The cultures were incubated at 250 rpm for 24 h in an Innova Laboratory Shaker (New Brunswick Scientific, Edison, N.J.) and samples of about 1 ml for OD determination withdrawn at designated hours. Optical density was measured with an Ultrospec 3000 spectrophotometer (Pharmacia Biotech) at  $\lambda$ =600 nm. In case cell dry weight concentrations were needed, an OD-DW-correlation of 0.33 gDW/OD was applied. Maximum specific growth rates  $\mu_{max}$  were determined by applying the exponential regression function of Microsoft Excel (Microsoft Office Excel 2003, SP 3). Outliers were discarded until good fit of the regression curve with measurements was confirmed by visual inspection. Parameters of the inhibition kinetics were determined by least square minimization of the differences between measured and predicted  $\mu_{max}$  values. Employed search algorithm was a quasi-Newton method with linear extrapolation from a tangent vector, as implemented in the solver routine of Microsoft Excel (Microsoft Office Excel 2003, SP 3).

<sup>20</sup> Pyrazole concentrations of 1 mM, 5 mM, 10 mM, 25 mM and 50 mM were tested, resulting in maximum specific growth rates of 0.54 l/h, 0.21 l/h, 0.12 l/h, 0.09 l/h and 0.08 l/h. Mumax values determined without inhibitor addition were 0.59 l/h, 0.59 l/h, 0.60 l/h, 0.62 l/h, 0.61 l/h and 0.62
<sup>25</sup> l/h, respectively. Inhibitory effect of pyrazole on growth was best described by the hybrid growth model. If fitted to the "hybrid" inhibition kinetics model, values of µ°<sub>max 1</sub>=0.52 l/h, µ°<sub>max 2</sub>=0.08 l/h and K<sub>1</sub>=2.8 mM were determined. The "hybrid" inhibition kinetics model predicts an overall observable µ°<sub>max</sub>=0.60 l/h and an IC<sub>50</sub> (inhibitor concentration with a specific growth rate of 50% µ°<sub>max</sub>) value of 3.3 mM. Measured µ<sub>max</sub> values and fitted dependency of µ<sub>max</sub> on the concentration of pyrazole in the medium is depicted in FIG. 11. Data from the samples is seen in Table 10 below.

 TABLE 10

 Data for control samples and pyrazole-inhibited experiments.

40		F1-Ctrl-A: 0 mM; SF12-F2-ctrl-B: 0 mM; F16-py-1: 1 mM; F17-py-2: 50 mM; F1-ctrl-A: 0 mM; F2-ctrl-B: 0 mM; F8- Py-5: 5 mM; F9-Py-10: 10 mM; F10-Py-25: 25 mM.											
45	sam- ple	time	time [min]	time [h]	OD600	dilu- tion [1:x]	OD600corr []	OD600corr []					
					SF12-F1-	Ctrl-A							
50	0 1 2 3 4 6	8:55 10:25 11:35 12:55 14:15 8:30	0 90 160 240 320 1395	0.00 1.50 2.67 4.00 5.33 23.25	0.156 0.170 0.222 0.350 0.220 0.648 SF12-F2-	1 1 1 5 20 -ctrl-B	0.089 0.103 0.155 0.283 0.777 11.677	0.155 0.283 0.777					
55	0 1 2 3 4 5	8:55 10:25 11:35 12:55 14:15 8:30	0 90 160 240 320 1395	0.00 1.50 2.67 4.00 5.33 23.25	0.162 0.172 0.223 0.354 0.228 0.667 F16-p	1 1 1 5 20 y-1	0.095 0.105 0.156 0.287 0.817 12.057	0.156 0.287 0.817					
60 65	0 1 2 3 4 5 6	8:55 10:25 11:35 12:55 14:15 3:40 8:30	0 90 160 240 320 405 1395	0.00 1.50 2.67 4.00 5.33 6.75 23.25	0.155 0.167 0.199 0.277 0.444 0.262 0.645	1 1 1 1 1 5 20	0.088 0.100 0.132 0.210 0.377 0.987 11.617	0.132 0.210 0.377 0.987					

TABLE 10-continued

Data for control samples and pyrazole-inhibited experiments. Pyrazole concentrations in the experiments were: SF12-F1-Ctrl-A: 0 mM; SF12-F2-ctrl-B: 0 mM; F16-py-1: 1 mM; F17-py-2: 50 mM; F1-ctrl-A: 0 mM; F2-ctrl-B: 0 mM; F8-Py-5: 5 mM; F9-Py-10: 10 mM; F10-Py-25: 25 mM.

					dilu-		
sam-		time	time	00(00)	tion	OD600corr	OD600corr
ple	time	[min]	[n]	OD600	[1:x]	IJ	[]
				F17-p	y-2		
~	0.55	0	0.00	0.164	1	0.007	
1	8:55	0	1.50	0.164	1	0.097	0.000
2	11.35	160	2.67	0.160	1	0.099	0.099
3	12.55	240	4.00	0.109	1	0.102	0.102
3	14.15	240	5.22	0.175	1	0.110	0.108
4	14.15	405	5.55	0.100	1	0.119	0.119
5	18.20	565	0.75	0.190	1	0.123	0.123
7	8.30	1305	23.42	0.209	1	0.142	0.142
ó	8.55	1393	23.23	0.302	1	0.235	0.255
1	10.25		1.50	0.166	1	0.097	0.000
2	11.35	160	2.67	0.160	1	0.102	0.000
3	12.55	240	4.00	0.175	1	0.102	0.102
1	14.15	320	5 33	0.186	1	0.110	0.100
5	15.40	405	6.75	0.100	1	0.123	0.112
6	18.20	565	9.42	0.120	1	0.123	0.123
7	8.30	1395	23.75	0.209	1	0.235	0.235
ó	8.55	1575	0.00	0.164	1	0.097	0.235
1	10.25	00	1.50	0.166	1	0.027	0 000
2	11.35	160	2.67	0.160	1	0.102	0.022
ź	12.55	240	4.00	0.175	1	0.102	0.108
4	14.15	320	533	0.186	1	0.110	0.110
5	15.40	405	6.75	0.190	1	0.123	0.123
6	18.70	565	0.75	0.120	1	0.123	0.123
7	8.30	1395	23.75	0.202	1	0.235	0.235
'	0.50	1575	23.23	F1-ct	rl-A	0.235	0.235
~	0.15	0	0.00	0.1.00	1	0.002	
1	8:15	0	0.00	0.160	1	0.093	
1	9:45	170	1.50	0.109	1	0.102	0.1.02
2	12.20	170	2.83	0.229	1	0.162	0.162
3	12:20	245	4.08	0.308	1	0.301	0.301
4	13:33	1455	2.22	0.215	20	11 507	0.752
0	8.50	1455	24.23	E2-ct	rl-R	11.397	
				12.00			
0	8:15	0	0.00	0.159	1	0.092	
1	9:45	90	1.50	0.168	1	0.101	
2	11:05	170	2.83	0.228	1	0.161	0.161
3	12:20	245	4.08	0.372	1	0.305	0.305
4	13:35	320	5.33	0.215	5	0.752	0.752
5	8:30	1455	24.25	0.652	20	11.757	
				F8-P	y-5		
0	8:15	0	0.00	0.163	1	0.096	
1	9:45	90	1.50	0.162	1	0.095	0.095
2	11:05	170	2.83	0.177	1	0.110	0.110
3	12:20	245	4.08	0.210	1	0.143	0.143
4	13:35	320	5.33	0.256	1	0.189	0.189
5	3:00	405	6.75	0.362	1	0.295	0.295
6	8:30	1455	24.25	0.551	20	9.737	
				F9-Py	-10		
0	8.15	0	0.00	0.164	1	0.007	
1	0.15	0	1.50	0.104	1	0.097	0 000
2	11.05	170	2.50	0.181	1	0.099	0.099
2	12.00	2/5	2.65	0.101	1	0.114	0.114
4	13.35	320	533	0.238	1	0.171	0.171
+ 5	3.00	405	6.75	0.230	1	0.171	0.222
5	8.30	1455	24.25	0.209	20	8 217	0.222
	0.50	1400	27.23	F10-P	y-25	0.217	
0	0.15	~	0.00	0.163	1	0.005	
0	8:15	0	0.00	0.162	1	0.095	0.000
1	9:45	170	1.50	0.159	1	0.092	0.092
2	11:05	170	2.83	0.167	1	0.100	0.100
3	12:20	245	4.08	0.181	1	0.114	0.114
4	13:35	320	5.33	0.194	1	0.127	0.127
2	3:00	405	0.75	0.211	1	0.144	0.144
	8:30	1455	24.25	0.374	20	6.197	

# 76

# Example 18

# Inhibition of Ethanologen Yeast PNY 827 by Cinnamaldehyde

The inhibitory effect of cinnamaldehyde on ethanologen yeast PNY 827 was investigated. Therefore a 125 ml aerobic shake flask was prepared with 20 ml SEED medium and inoculated with 1 ml of frozen glycerol stock culture of PNY 827. The culture was inoculated over night at 30° C. and 250 rpm in an Innova Laboratory Shaker (New Brunswick Scientific, Edison, N.J.). Subsequently, a sufficient amount <sup>15</sup> of the seed culture was transferred into shake flasks containing 20 ml of production medium without cinnamaldehyde or addition of cinnamaldehyde at concentrations of 200 mM, 100 mM, 50 mM, 25 mM, 10 mM, 1 mM, 0.1 mM, 20 0.01 mM and 0.001 mM, to give an initial OD of approximately 0.1. The cultures were incubated at 250 rpm for 24 h in an Innova Laboratory Shaker (New Brunswick Scientific, Edison, N.J.) and samples of about 1 ml for OD determination withdrawn at designated hours. Optical density was measured with an Ultrospec 3000 spectrophotometer (Pharmacia Biotech) at  $\lambda$ =600 nm. In case cell dry weight concentrations were needed, an OD-DW-correlation of 0.33 gDW/OD was applied. Maximum specific growth <sup>30</sup> rates  $\mu_{max}$  were determined by applying the exponential regression function of Microsoft Excel (Microsoft Office Excel 2003, SP 3). Outliers were discarded until good fit of the regression curve with measurements was confirmed by visual inspection. Parameters of the inhibition kinetics were determined by least square minimization of the differences between measured and predicted  $\mu_{max}$  values. Employed search algorithm was a quasi-Newton method with linear extrapolation from a tangent vector, as implemented in the solver routine of Microsoft Excel (Microsoft Office Excel 2003, SP 3).

The inhibitory effect of cinnamaldehyde (CA) was investigated at 200 mM, 100 mM, 50 mM, 25 mM, 10 mM, 1 <sup>45</sup> mM, 0.1 mM, 0.01 mM and 0.001 mM. For generating the concentrations of 0.1 mM, 0.01 mM and 0.001 mM, cinnamaldehyde was diluted with DMSO, resulting in DMSO concentrations in the cell suspension of 0.7 mM, 7 mM and 70 mM of DMSO, respectively. Assuming an additive effect of cinnamaldehvde and DMSO inhibition, observed maximum specific growth rates of the two cultures were corrected by 0.000 1/h, 0.000 1/h and 0.005 1/h due to the effect of DMSO, resulting in cinnamaldehyde-based mumax val-5 ues of 0.64 1/h, 0.63 1/h and 0.55 1/h derived from the observed values of 0.64 l/h, 0.63 l/h and 0.55 l/h, respectively. At all the other concentrations, no DMSO was used in the stock solution of cinnamaldehyde. However, in none of the non-DMSO experiments any cell growth was 50 observed. Fitting the data to the "squared inhibition" kinetics (observed maximum specific growth rates without inhibitor addition were 0.59 1/h, 0.59 1/h, 0.60 1/h, 0.62 1/h, 0.62 1/h and 0.62 l/h) yielded parameters of  $\mu^{\circ}_{max}=0.62$  l/h and a K<sub>1</sub> value of  $K_r=0.25$  mM (FIG. 12). These findings indicate 5 strong inhibition of growth of PNY 827 by trans-cinnamaldehyde with a derived IC50 of 0.25 mM. Data from the samples is seen in Table 11 below.

Data for control samples and cinnamaldehyde-inhibited experiments. Cinnamaldehyde concentrations in the experiments were: SF12-F1-Ctrl-A: 0 mM; SF12-F2-ctrl-B: 0 mM; F7-ca-1: 50 mM; F8ca-2: 100 mM; F9-ca-3: 200 mM; SF13-F1-ctrl-A: 0 mM; SF13f2-ctrl-B: 0 mM; F3-Ca-1: 1 mM; F4-Ca-10: 10 mM; F5-Ca-25: 25 mM; F1-Ctrl-A: 0 mM; F2-ctrl-B: 0 mM; F3-Ca-a: 0.001 mM; F4-Ca-b: 0.01 mM; F5-Ca-c: 0.1 mM.

TABLE 11-continued

Data for control samples and cinnamaldehyde-inhibited experiments. Cinnamaldehyde concentrations in the experiments were: SF12-F1-Ctrl-A: 0 mM; SF12-F2-ctrl-B: 0 mM; F7-ca-1: 50 mM; F8ca-2: 100 mM; F9-ca-3: 200 mM; SF13-F1-ctrl-A: 0 mM; SF13-F2-ctrl-B: 0 mM; F3-Ca-1: 1 mM; F4-Ca-10: 10 mM; F5-Ca-25: 25 mM; F1-Ctrl-A: 0 mM; F2-ctrl-B: 0 mM; F3-Ca-a: 0.001 mM; F4-Ca-b: 0.01 mM; F5-Ca-c: 0.1 mM.

sam- ple	time	time [min]	time [h]	OD600	dilu- tion [1:x]	OD600corr []	OD600corr []	10	sam- ple	time	time [min]	time [h]	OD600
				SF12-F1-	Ctrl-A				2	11:05	170	2.83	0.170
0 1 2	8:55 10:25 11:35	0 90 160	0.00 1.50 2.67	0.156 0.170 0.222	1 1 1	0.089 0.103 0.155	0.155	15	3 4 5	12:20 13:35 8:30	245 320 1455	4.08 5.33 24.25	0.174 0.172 0.141 F5-0
3 4 6	12:55 14:15 8:30	240 320 1395	4.00 5.33 23.25	0.350 0.220 0.648 SF12-F2	1 5 20 -ctrl-B	0.283 0.777 11.677	0.283 0.777	_	0 1 2	8:15 9:45 11:05	0 90 170	0.00 1.50 2.83	0.171 0.173 0.172
0 1 2	8:55 10:25 11:35	0 90 160	0.00 1.50 2.67	0.162 0.172 0.223	1 1 1	0.095 0.105 0.156	0.156	20	3 4 5	12:20 13:35 8:30	245 320 1455	4.08 5.33 24.25	0.174 0.168 0.140 F1-0
3 4 5	12:55 14:15 8:30	320 1395	5.33 23.25	0.334 0.228 0.667 F7-ca	5 20 1-1	0.287 0.817 12.057	0.287	. 25	0 1 2 3	8:25 9:55 11:15	0 90 170 250	0.00 1.50 2.83	0.167 0.190 0.264
0 1 2	8:55 10:25 11:35	0 90 160 240	0.00 1.50 2.67	0.167 0.168 0.172	1 1 1	0.100 0.101 0.105 0.104	0.101 0.105		4	12:55	315	5.25	0.432 0.246 F2-4
4 5	12:55 14:15 8:30	320 1395	5.33 23.25	0.171 0.171 0.144 F8-ca	1 1 1 1-2	0.104 0.104 0.077	0.104	30	1 2 3	9:55 11:15 12:35	90 170 250	0.00 1.50 2.83 4.17	0.103 0.190 0.264 0.460
0 1 2 3	8:55 10:25 11:35 12:55	0 90 160 240	0.00 1.50 2.67 4.00	0.178 0.179 0.175 0.174	1 1 1 1	0.111 0.112 0.108 0.107	0.112 0.108 0.107	35	4 0 1	8:25 9:55	0 90	0.00 1.50	0.248 F3- 0.166 0.192
4 5	14:15 8:30	320 1395	5.33 23.25	0.170 0.136 F9-ca	1 1 1-3	0.103 0.069	0.103		2 3 4	11:15 12:35 13:40	170 250 315	2.83 4.17 5.25	0.266 0.461 0.256 F4-
1 2 3 4 5	8:55 10:25 11:35 12:55 14:15 8:30	90 160 240 320 1395	1.50 2.67 4.00 5.33 23.25	0.173 0.179 0.173 0.169 0.160 0.122 SF13-F1-	1 1 1 1 1 -ctrl-A	0.108 0.112 0.106 0.102 0.093 0.055	0.112 0.106 0.102 0.093	40	0 1 2 3 4	8:25 9:55 11:15 12:35 13:40	0 90 170 250 315	0.00 1.50 2.83 4.17 5.25	0.169 0.190 0.262 0.443 0.245 F5-
0 1 2 3 4 6	8:15 9:45 11:05 12:20 13:35 8:30	0 90 170 245 320 1455	0.00 1.50 2.83 4.08 5.33 24.25	0.160 0.169 0.229 0.368 0.215 0.644 SE13 - E2	1 1 1 5 20	0.093 0.102 0.162 0.301 0.752 11.597	0.162 0.301 0.752	50 50 <sup>50</sup>	0 1 2 3 4 5	8:25 9:55 11:15 12:35 13:40 14:55	0 90 170 250 315 390	0.00 1.50 2.83 4.17 5.25 6.50	0.164 0.186 0.235 0.364 0.575 0.321
0 1 2 3 4 5	8:15 9:45 11:05 12:20 13:35 8:30	0 90 170 245 320 1455	0.00 1.50 2.83 4.08 5.33 24.25	0.159 0.168 0.228 0.372 0.215 0.652 F3-C:	1 1 1 1 5 20 a-1	0.092 0.101 0.161 0.305 0.752 11.757	0.161 0.305 0.752	55		Inhil	oition	of Etl 1-bi	Exam nanolo romo-2
0 1 2 3 4 5	8:15 9:45 11:05 12:20 13:35 8:30	0 90 170 245 320 1455	0.00 1.50 2.83 4.08 5.33 24.25	0.165 0.164 0.165 0.167 0.167 0.167 F4-Ca	1 1 1 1 1 10	0.098 0.097 0.098 0.100 0.100 0.100	0.097 0.098 0.100 0.100 0.100	60	Th gen y aerob and in PNY and 2	e inhil veast H ic sha nocula 827. 250 rpi	NY 8 NY 8 ke flas ted w The c n in a	effect 327 was sk was ith 1 n ulture in Inno	of 1-b as inv prepa nl of f was i ova La
0	8:15 9:45	0 90	0.00 1.50	0.167 0.169	1	0.100 0.102	0.102	65	wick amou conta	Scien Int of t ining	tific, he see 20 ml	Ediso ed cult of pr	n, N.J ure wa oducti

10	sam- ple	time	time [min]	time [h]	<b>OD</b> 600	dilu- tion [1:x]	OD600corr []	OD600corr []
	2	11:05	170	2.83	0.170	1	0.103	0.103
	3	12:20	245	4.08	0.174	1	0.107	0.107
	4	13:35	320	5.33	0.172	1	0.105	0.105
1.5	5	8:30	1455	24.25	0.141	1	0.074	0.074
15					F5-Ca	-25		
		0.15	0	0.00	0.171		0.104	
	0	8:15	0	0.00	0.171	1	0.104	0.107
	1	9:45	90	1.50	0.173	1	0.106	0.106
	2	11:05	170	2.83	0.172	1	0.105	0.105
20	3	12:20	245	4.08	0.174	1	0.107	0.107
20	4	13:35	320	5.33	0.168	1	0.101	0.101
	5	8:30	1455	24.25	0.140	1	0.073	0.073
					F1-Ct	rl-A		
	0	8:25	0	0.00	0.167	1	0.100	
	1	9:55	90	1.50	0.190	1	0.123	
25	2	11:15	170	2.83	0.264	1	0.197	0.197
	3	12:35	250	4.17	0.452	1	0.385	0.385
	4	13:40	315	5.25	0.246	5	0.907	0.907
					F2-cti	·l-B		
	0	8.75	0	0.00	0.165	1	0.008	
•	1	0:25	0	1.50	0.105	1	0.098	
30	2	9:55	170	1.50	0.190	1	0.125	0.107
	2	11:15	170	2.85	0.204	1	0.197	0.197
	3	12:35	250	4.17	0.460	1	0.393	0.393
	4	13:40	315	5.25	0.248	3	0.917	0.917
					F3-C	a-a		
35	0	8:25	0	0.00	0.166	1	0.099	
55	1	9:55	90	1.50	0.192	1	0.125	
	2	11:15	170	2.83	0.266	1	0.199	0.199
	3	12:35	250	4.17	0.461	1	0.394	0.394
	4	13:40	315	5.25	0.256	5	0.957	0.957
					F4-C	a-b		
40	0	8.75	0	0.00	0.160	1	0.102	
	1	0.25		1.50	0.109	1	0.102	
	2	11.15	170	2.20	0.150	1	0.125	0.105
	3	12.35	250	4.05	0.202	1	0.155	0.155
	1	12.55	315	5.25	0.245	5	0.370	0.370
	-	15.40	515	5.25	0.245 E5-C	9-C	0.902	0.902
45					15-0			
	0	8:25	0	0.00	0.164	1	0.097	
	1	9:55	90	1.50	0.186	1	0.119	

# mple 19

1

1

1

-5

0.168

0.297

0.508

1.282

0.168

0.297

0.508

1.282

# ogen Yeast PNY 827 by -2-butanone

bromo-2-butanone on ethanolovestigated. Therefore a 125 ml bared with 20 ml SEED medium frozen glycerol stock culture of inoculated overnight at 30° C. aboratory Shaker (New Bruns-.J.). Subsequently, a sufficient vas transferred into shake flasks tion medium without 1-bromo-2-butanone or addition of 1-bromo-2-butanone at concen-

5

0

1

8:25

9:55

0 0.00

90

1.50

0.165

0.180

1

trations of 50 mM, 5 mM, 1 mM, 0.5 mM, 0.1 mM, 0.01 mM and 0.001 mM, to give an initial OD of approximately 0.1. The cultures were incubated at 250 rpm for 24 h in an Innova Laboratory Shaker (New Brunswick Scientific, Edison, N.J.) and samples of about 1 ml for OD determination withdrawn 5 at designated hours. Optical density was measured with an Ultrospec 3000 spectrophotometer (Pharmacia Biotech) at  $\lambda$ =600 nm. In case cell dry weight concentrations were needed, an OD-DW-correlation of 0.33 gDW/OD was applied. Maximum specific growth rates  $\mu_{max}$  were determined by applying the exponential regression function of Microsoft Excel (Microsoft Office Excel 2003, SP 3). Outliers were discarded until good fit of the regression curve with measurements was confirmed by visual inspection. Parameters of the inhibition kinetics were determined by 1 least square minimization of the differences between measured and predicted  $\mu_{max}$  values. Employed search algorithm was a quasi-Newton method with linear extrapolation from a tangent vector, as implemented in the solver routine of Microsoft Excel (Microsoft Office Excel 2003, SP 3). 2

The inhibitory effect of 1-bromo-2-butanone was investigated at 50 mM, 5 mM, 1 mM, 0.5 mM, 0.1 mM, 0.01 mM and 0.001 mM. For generating the concentrations of 0.1 mM, 0.01 mM and 0.001 mM, 1-bromo-2-butanone was diluted with DMSO, resulting in DMSO concentrations in 2 the cell suspension of 0.7 mM, 7 mM and 70 mM of DMSO, respectively. Assuming an additive effect of 1-bromo-2butanone and DMSO inhibition, observed maximum specific growth rates of the two cultures were corrected by 0.000 l/h, 0.000 l/h and 0.005 l/h due to the effect of DMSO, 3 resulting in 1-bromo-2-butanone-based mumax values of 0.54 1/h, 0.00 1/h and 0.00 1/h derived from the observed values of 0.54 l/h, 0.00 l/h and 0.00 l/h, respectively. At all the other concentrations, no DMSO was used for dilution of 1-bromo-2-butanone. However, in all of the non-DMSO 3 experiments no cell growth was observed. Fitting the data to the "squared inhibition" kinetics (observed maximum specific growth rates without inhibitor addition were 0.59 l/h, 0.59 l/h, 0.60 l/h, 0.62 l/h, 0.62 l/h and 0.62 l/h) yielded parameters of  $\mu^{o}_{max}$ =0.61 l/h and a K<sub>I</sub> value of K<sub>I</sub>=0.002 4 mM (FIG. 13). This corresponds to an IC50 value of 1-bromo-2-butanone on growth of 0.002 mM, indication of strong inhibition of ethanologen yeast by 1-bromo-2-butanone. Data from the samples is seen in Table 12 below.

TABLE 12

Data for control samples and 1-bromo-2-butanone-inhibited experiments, 1-bromo-2-butanone concentrations in the experiments were: SF12-
F1-Ctrl-A: 0 mM; SF12-F2-ctrl-B: 0 mM; F10-bb-1: 5 mM; F11-bb-2:
50 mM; SF13-F1-ctrl-A: 0 mM; SF13-F2-ctrl-B: 0 mM; F6-Bb-0.5: 0.5 mM; F7-Bb-1: 1 mM; F1-Ctrl-A: 0 mM; F2-ctrl-B: 0 mM; F6-Bb-a:
0.001 mM; F7-Bb-b: 0.01 mM; F8-Bb-c: 0.1 mM.

sam- ple	time	time [min]	time [h]	<b>OD6</b> 00	dilu- tion [1:x]	OD600corr []	OD600corr []	55
				SF12-F1-	Ctrl-A			
0	0.55	0	0.00	0.150	1	0.080		
U	8:55	0	0.00	0.156	1	0.089		_
1	10:25	90	1.50	0.170	1	0.103		
2	11:35	160	2.67	0.222	1	0.155	0.155	60
3	12:55	240	4.00	0.350	1	0.283	0.283	60
4	14:15	320	5.33	0.220	5	0.777	0.777	
6	8:30	1395	23.25	0.648	20	11.677		
				SF12-F2	-ctrl-B			
0	8:55	0	0.00	0.162	1	0.095		_
1	10:25	90	1.50	0.172	1	0.105		65
2	11.35	160	2 67	0.223	1	0.156	0.156	

TABLE 12-continued

Data for control samples and 1-bromo-2-butanone-inhibited experiments, 1-bromo-2-butanone concentrations in the experiments were: SF12-F1-Ctrl-A: 0 mM; SF12-F2-ctrl-B: 0 mM; F10-bb-1: 5 mM; F11-bb-2: 50 mM; SF13-F1-ctrl-A: 0 mM; SF13-F2-ctrl-B: 0 mM; F6-Bb-0.5:

0	sam- ple	time	time [min]	time [h]	<b>OD</b> 600	dilu- tion [1:x]	OD600corr []	OD600corr []
	3	12:55	240	4.00	0.354	1	0.287	0.287
	4	14:15	320	5.33	0.228	5	0.817	0.817
	5	8:30	1395	23.25	0.667	20	12.057	
_					F10-6	0-1		
5	0	8:55	0	0.00	0.161	1	0.094	
	1	10:25	90	1.50	0.176	1	0.109	0.109
	2	11:35	160	2.67	0.170	1	0.103	0.103
	3	12:55	240	4.00	0.168	1	0.101	0.101
	5	8:30	1395	23.25	0.170	1	0.103	0.103
0	U U	0.00	1070	20120	F11-b	b-2		
		0.55	0	0.00	0.2(1		0.104	
	1	8:55	90	1.50	0.261	1	0.194	0.171
	2	11:35	160	2.67	0.275	1	0.208	0.208
	3	12:55	240	4.00	0.266	1	0.199	0.199
5	4	14:15	320	5.33	0.264	1	0.197	0.197
	5	8:30	1395	23.25	0.161 SE12 E1	1 atul A	0.094	0.094
					SF15-F1	-ctrl-A		
	0	8:15	0	0.00	0.160	1	0.093	
	1	9:45	90	1.50	0.169	1	0.102	
0	2	11:05	170	2.83	0.229	1	0.162	0.162
	3	12:20	245	4.08	0.368	1	0.301	0.301
	4	8.30	1455	24 25	0.213	20	11 597	0.752
	0	0.50	1 100	21125	SF13 - F2	2-ctrl-B	11.000	
		0.15	0	0.00	0.150		0.002	
5	0	8:15	0	0.00	0.159	1	0.092	
	2	11:05	170	2.83	0.228	1	0.161	0.161
	3	12:20	245	4.08	0.372	1	0.305	0.305
	4	13:35	320	5.33	0.215	5	0.752	0.752
	5	8:30	1455	24.25	0.652	20	11.757	
0					Fo-Bb	-0.5		
	0	8:15	0	0.00	0.165	1	0.098	
	1	9:45	90	1.50	0.166	1	0.099	0.099
	2	11:05	170	2.83	0.169	1	0.102	0.102
	3	12:20	245	4.08	0.169	1	0.102	0.102
5	5	8:30	1455	24.25	0.169	1	0.102	0.102
		0.00	1.00	220	F7-B	b-1		0.1102
	0	0.15	0	0.00	0.157	1	0.000	
	1	9:45	90	1.50	0.161	1	0.090	0.094
	2	11:05	170	2.83	0.160	1	0.093	0.093
0	3	12:20	245	4.08	0.160	1	0.093	0.093
	4	13:35	320	5.33	0.158	1	0.091	0.091
	3	8:50	1455	24.23	0.161 F1-Ct	rl-A	0.094	0.094
	0	8:25	0	0.00	0.167	1	0.100	
5	1	9:55	90 170	1.50	0.190	1	0.123	0.107
	3	12:35	250	4.17	0.452	1	0.385	0.385
	4	13:40	315	5.25	0.246	5	0.907	0.907
					F2-cti	l-B		
	0	8:25	0	0.00	0.165	1	0.098	
0	1	9:55	90	1.50	0.190	1	0.123	
	2	11:15	170	2.83	0.264	1	0.197	0.197
	3	12:35	250	4.17	0.460	1	0.393	0.393
	4	13:40	315	5.25	0.248 E6-D	5 h-9	0.917	0.917
					1.0-P	0-a		

0.098

0.113

#### TABLE 12-continued

Data for control samples and 1-bromo-2-butanone-inhibited experiments, 1-bromo-2-butanone concentrations in the experiments were: SF12-F1-Ctrl-A: 0 mM; SF12-F2-ctrl-B: 0 mM; F10-bb-1: 5 mM; F11-bb-2: 50 mM; SF13-F1-ctrl-A: 0 mM; SF13-F2-ctrl-B: 0 mM; F6-Bb-0.5: 0.5 mM; F7-Bb-1: 1 mM; F1-Ctrl-A: 0 mM; F2-Bb-2: 0.1 mM; F6-Bb-a: 0.001 mM; F7-Bb-b: 0.01 mM; F8-Bb-c: 0.1 mM.

sam- ple	time	time [min]	time [h]	OD600	dilu- tion [1:x]	OD600corr []	OD600corr []	1
2	11:15	170	2.83	0.219	1	0.152	0.152	
3	12:35	250	4.17	0.314	1	0.247	0.247	
4	13:40	315	5.25	0.500	1	0.433	0.433	
5	14:55	390	6.50	0.287	5	1.112	1.112	
				F7-B	b-b			1
								1
0	8:25	0	0.00	0.166	1	0.099		
1	9:55	90	1.50	0.184	1	0.117	0.117	
2	11:15	170	2.83	0.183	1	0.116	0.116	
3	12:35	250	4.17	0.189	1	0.122	0.122	
4	13:40	315	5.25	0.186	1	0.119	0.119	
5	14:55	390	6.50	0.191	1	0.124	0.124	2
				F8-B	b-c			
0	8:25	0	0.00	0.164	1	0.097		
1	9:55	90	1.50	0.168	1	0.101	0.101	
2	11:15	170	2.83	0.166	1	0.099	0.099	
3	12:35	250	4.17	0.170	1	0.103	0.103	2
4	13:40	315	5.25	0.170	1	0.103	0.103	
5	14:55	390	6.50	0.170	1	0.103		

#### Example 20

# Effect of Ethanol Dehydrogenase and Pyruvate Decarboxylase Inhibitors on Growth and Product Formation of Mixed Cultures of Ethanologen and Butanologen Yeast

Effects of addition of ethanol dehydrogenase and pyruvate decarboxylase inhibitors on mixed cultures of ethanologen 40 S. cerevisiae PNY 827 and the butanologen yeast S. cerevisiae PNY 2129 were investigated. Therefore two 125 ml aerobic shake flask were prepared with 20 ml SEED medium and each inoculated with 1 ml of frozen glycerol stock culture of PNY 2129 in the morning. Another 125 ml aerobic 45 shake flask was prepared with 20 ml SEED medium and inoculated with 1 ml of frozen glycerol stock culture of PNY 827 in the afternoon. All cultures were incubated at 30° C. and 250 rpm overnight in an Innova Laboratory Shaker 50 (New Brunswick Scientific, Edison, N.J.). In the morning, sufficient seed culture volume of each strain to give OD600 of 1.000 in the resuspended solution was separately transferred into 50 mL sterile centrifuge tubes and spun down at 9500 rpm for 20 min in an Eppendorf Centrifuge 5804R 55 (Eppendorf, Hamburg, Germany). Supernatants were discarded and the cell pellets resuspended in 20 ml of 0.9% NaCl solution. Optical density was measured with an Ultrospec 3000 spectrophotometer (Pharmacia Biotech) at  $\lambda$ =600 nm. Subsequently "production" cultures were prepared in 25 60 ml Balch tubes by adding into each tube 6 ml Yeast synthetic w/o aa, w/o glucose, w/o ethanol, w/o Tween (2x), 1.2 ml supplement amino acid solution without histidine and uracil (SAAS-2 10×), 1.92 ml of 250 g/l glucose (ca. 40 g/l glucose) and 2.3 µl of 3M sodium acetate, as well as a 65 specific amount of inoculum solutions, inhibitor solution and water according to the schema in Table 13:

TABLE	13
	_

	Schei	ma showing st	rains and inl	ubitors s	olution	5.	
5		inoculun	n solution	stock:	300 mM inhib	2 mM (200 mM) pitor sol	250 mM (2500 mM) ution
10		PNY2129 [ul]	PNY827 [ul]	H2O [ul]	Py [ul]	Bb [ul]	Ca [ul]
15	T1-ctrl1 B T2-ctrl2 B T3-ctrl1 E T4-ctrl2 E T5-1:1-PyA T6-1:1-PyB T7-11:1-PyA T8-11:1-PyB T9-1:1-BbA T10-1:1-BbB	1200 1200 600 1100 1100 600 600	1200 1200 600 100 100 600 600 100	1200 1200 1200 1200 1080 1080 1188 1188	120 1200 120 1200	12 12	
20 25	111-11:1-BbA T12-11:1-BbB T13-11-CaA T14-1:1-CaB T15-11:1-CaB T15-11:1-CaB T17-1:1-ctrl1 T18-11:1-ctrl1	1100     1100     600     600     1100     1100     600     1100	100     100     600     100     100     600     100	1188 1188 1188 1080 1188 1080 1200 1200		12	12 120 12 120

Inhibitor solutions were trans-cinnamaldehyde (Aldrich, #239968, CAS: 14371-10-9) dissolved in water either ad 250 mM or 2500 mM, 1-bromo-2-butanone (Sigma-Aldrich, 30 #243299, CAS: 816-40-0) dissolved in water either at 2 mM or 200 mM, and pyrazole (Aldrich, # P56607, CAS Number: 288-13-1), dissolved in water at 300 mM. Resulting inhibitor concentrations in the Balch tube cultures were pyrazole 35 (PY): 3 mM (A) and 30 mM (B), 1-bromo-2-butanone (BB): 2 µM (A) and 200 µM (B), and trans-cinnamaldehyde (CA): 250 µM (A) and 25 mM (B), respectively. Each Balch tube was fitted with a butyl rubber septum and crimped to the tube with a sheet metal with circular opening to allow samples withdrawal by syringes. The cultures were mixed by a vial/tube rotator (Glas-Col, Terre-Haute, Ind.) that was placed in an Innova Laboratory Shaker (New Brunswick Scientific, Edison, N.J.) for keeping the temperature at 30° C. Samples of about 1 ml for OD determination and extracellular compound analysis were withdrawn at designated hours. Extracellular compound analysis in supernatant was accomplished by HPLC. An Aminex® HPX-87H column (Bio-Rad, Hercules, Calif.) was used in an isocratic method with 0.01N sulfuric acid as eluent on an Alliance® 2695 Separations Module (Waters Corp., Milford, Mass.). Flow rate was 0.60 mL/min, column temperature 40° C., injection volume 10 µL and run time 58 min. Detection was carried out with a refractive index detector (Waters 2414 RI, Waters Corp., Milford, Mass.) operated at 40° C. and an UV detector (Waters 2996 PDA, Waters Corp., Milford, Mass.) at 210 nm. Determined optical densities as well as concentrations of extracellular compounds at selected sampling time points can be found in Table 14.

Butanol to ethanol formed in the mixed cultures with inhibitors was compared to the ratio of butanol to ethanol formed in the mixed cultures without inhibitor (Ctrl) at 8 hours (EPT=8 h, FIG. 14 and FIG. 15) and at 48 hours (EPT=48 h, FIG. 16 and FIG. 17) of the experiments inoculated with a butanologen-to-ethanologen ratio of 11:1 (b:e=11:1, FIG. 14 and FIG. 16) or 1:1 (b:e=1:1, FIG. 15 and FIG. 17).

No growth in both mixed cultures and at both time points was observed at the high concentration of trans-cinnamaldehyde of 25 mM (FIG. 14-FIG. 17). At the lower concentration of 250  $\mu$ M, both mixed cultures grew and produced alcohols. However, at both sampling times (EPT=8 h and 5 EPT=48 h) as well as at both inoculum ratios (1:1 and 1:11), the ratio of butanol vs. ethanol produced was lower with addition of trans-cinnamaldehyde than without addition (FIG. 14-FIG. 17).

With 1-bromo-2-butanone, no growth was observed in the 10 1:1 culture at the high concentration of 200  $\mu$ M until EPT=8 h, only at EPT=48 h. At low concentration (2  $\mu$ M), cultures

with both inoculum ratios showed increased butanol-toethanol ratios at EPT=8 h, but not at EPT=48 h. The same findings apply to the culture with 1:11 ratio at the high concentration (FIG. 14-FIG. 17).

With pyrazole addition at both concentrations, 3 mM and 30 mM, cultures with inoculum ratios 1:1 as well as 1:11 showed dramatically increased butanol-to-ethanol ratios at EPT=8 h (FIG. 14 and FIG. 15). However, at EPT=48 h cultures with both inoculum ratios maintained significantly increased butanol:ethanol ratios only at the higher pyrazole concentration of 30 mM, but not at the lower concentration of 3 mM (FIG. 16 and FIG. 17).

Optical density (OD) and extracellular compound concentrations at the different sampling time points (EPT = elapsed process time) of different pure and mixed cultures. Abbreviations used were: EtOH = ethanol, PYR = pyruvate, KTV = ketoisovalerate, DHIV = dihydroisovalerate, DHMB = 2,3-dihydroxy-2-metylbutyrate, GLY = glycerol, ACE = acetate, IBOOH = isobutyric acid, IBOH = isobutanol, m-BDO = meso-butanediol, d/l-BDO = d/l-butanediol, LAC = lactate, SUC = succinate.

							DHIV +					m-	d/l-		
Sample	EPT	OD	GLC	EtOH	PYR	KIV	DHMB	GLY	ACE	IBOOH	IBOH	BDO	BDO	LAC	SUC
[]	[h]	[]	[mM]	[mM]	[mM]	[mM]	[mM]	[mM]	[mM]	[mM]	[mM]	[mM]	[mM]	[mM]	[mM]
SF-17-T1-ctrl-1-B-0	0.00	0.115	226.0	0.0	0.0	0.0	0.0	0.1	6.8	0.0	0.0	0.0	0.0	0.0	0.0
SF-17-T1-ctrl-1-B-3	3.00	0.159	223.4	0.0	0.0	0.2	0.0	0.2	6.7	0.0	0.6	0.0	0.0	0.0	0.0
SF-17-T1-ctrl-1-B-6	6.00	0.210	221.3	0.0	0.1	0.2	0.0	0.3	6.7	0.3	1.4	0.0	0.0	0.0	0.0
SF-17-T1-ctrl-1-B-8	8.00	0.247	222.4	0.0	0.1	0.4	0.0	0.4	6.5	0.4	2.4	0.0	0.0	0.0	0.0
SF-17-T1-ctrl-1-B-24	24.00	1.382	178.0	1.2	1.8	3.1	0.6	2.8	4.4	2.5	24.8	0.0	0.0	0.1	0.3
SF-17-T1-ctrl-1-B-31	31.00	1.747	147.9	2.0	2.2	3.8	1.2	5.2	4.1	3.1	39.7	0.0	0.5	0.1	0.5
SF-17-T1-ctrl-1-B-48	48.00	1.917	106.9	3.6	2.9	4.4	1.9	11.9	3.0	3.3	69.1	0.0	1.3	0.2	0.5
SF-17-T1-ctrl-1-B-0	0.00	0.109	226.0	0.0	0.0	0.0	0.0	0.1	6.8	0.0	0.0	0.0	0.0	0.0	0.0
SF-17-T2-ctrl-2-B-3	3.00	0.157	223.4	0.0	0.0	0.1	0.0	0.2	6.7	0.0	0.6	0.0	0.0	0.0	0.0
SF-17-T2-ctrl-2-B-6	6.00	0.209	222.0	0.0	0.1	0.2	0.0	0.3	6.7	0.2	1.4	0.0	0.0	0.0	0.0
SF-17-T2-ctrl-2-B-8	8.00	0.239	221.1	0.0	0.1	0.4	0.0	0.4	6.5	0.3	2.4	0.0	0.0	0.0	0.0
SF-17-T2-ctrl-2-B-24	24.00	1.197	183.6	0.0	1.6	2.9	0.4	2.6	4.4	2.9	22.2	0.0	0.0	0.0	0.2
SF-17-T2-ctrl-2-B-31	31.00	1.627	152.0	2.1	2.1	3.9	1.1	4.9	4.0	3.3	38.1	0.0	0.5	0.0	0.5
SF-17-T2-ctrl-2-B-48	48.00	1.867	104.7	4.0	3.0	4.6	2.0	11.2	2.9	3.3	69.7	0.0	1.4	0.3	0.5
SF-17-T3-ctrl-1-E-0	0.00	0.088	224.0	0.0	0.0	0.0	0.0	0.0	7.1	0.0	0.0	0.0	0.0	0.0	0.0
SF-17-T3-ctrl-1-E-3	3.00	0.222	221.8	4.5	0.0	0.0	0.0	0.4	6.8	0.0	0.0	0.0	0.0	0.0	0.0
SF-17-T3-ctrl-1-E-6	6.00	1.317	206.1	30.3	0.1	0.0	0.0	1.6	6.6	0.0	0.0	0.0	0.0	0.0	0.0
SF-17-T3-ctrl-1-A-8	8.00	3.637	162.9	103.8	0.5	0.0	0.0	5.7	6.3	0.0	0.0	0.0	0.2	0.0	0.3
SF-17-T3-ctrl-1-E-24	24.00	9.497	0.0	378.5	2.5	0.0	0.0	18.6	9.5	0.0	0.0	0.0	0.5	0.4	0.5
SF-17-T3-ctrl-1-E-31	31.00	11.997	0.0	372.7	2.6	0.0	0.0	18.6	10.8	0.0	0.5	0.0	0.4	0.0	0.5
SF-17-T3-ctrl-1-E-48	48.00	11.897	0.0	382.0	2.3	0.0	0.0	18.6	11.0	0.0	0.0	0.0	0.3	0.3	0.6
SF-17-T3-ctrl-1-E-0	0.00	0.088	224.0	0.0	0.0	0.0	0.0	0.0	7.1	0.0	0.0	0.0	0.0	0.0	0.0
SF-17-T4-ctrl-2-E-3	3.00	0.225	221.7	4.6	0.0	0.0	0.0	0.4	6.8	0.0	0.0	0.0	0.0	0.0	0.0
SF-17-T4-ctrl-2-E-6	6.00	1.352	207.3	30.6	0.1	0.0	0.0	1.6	6.8	0.0	0.0	0.0	0.0	0.0	0.0
SF-17-T4-ctrl-1-B-8	8.00	3.727	161.8	104.8	0.5	0.0	0.0	5.6	6.4	0.0	0.0	0.0	0.0	0.2	0.3
SF-17-T4-ctrl-2-E-24	24.00	11.847	0.0	377.1	2.5	0.0	0.0	17.0	12.1	0.0	0.0	0.0	0.4	0.4	0.5
SF-17-T4-ctrl-2-E-31	31.00	11.547	0.0	367.9	2.4	0.0	0.0	17.1	13.0	0.0	0.6	0.0	0.4	0.4	0.5
SF-17-T4-ctrl-2-E-48	48.00	11.747	0.0	378.7	2.3	0.0	0.0	16.9	13.6	0.0	0.5	0.0	0.3	0.3	0.4
SF-17-T5-1:1-Py-A-0	0.00	0.104	224.0	0.0	0.0	0.0	0.0	0.1	6.8	0.0	0.0	0.0	0.0	0.0	0.0
SF-17-T5-1:1-Py-A-3	3.00	0.141	222.8	1.3	0.0	0.0	0.0	0.7	6.8	0.0	0.0	0.0	0.0	0.0	0.0
SF-17-T5-1:1-Py-A-6	6.00	0.273	218.1	5.4	0.1	0.1	0.0	2.5	6.9	0.0	0.0	0.0	0.0	0.0	0.0
SF-17-T5-1:1-Py-A-8	8.00	0.504	210.2	14.8	0.2	0.1	0.0	6.4	6.8	0.0	0.6	0.4	0.0	0.0	0.0
SF-17-T5-1:1-Py-A-24	24.00	9.777	0.0	346.4	2.9	0.1	0.0	37.0	10.3	0.1	1.6	0.9	1.5	0.4	0.4
SF-17-T5-1:1-Py-A-31	31.00	10.647	0.0	337.3	2.8	0.1	0.0	37.0	11.7	0.3	1.6	0.9	1.6	0.3	0.4
SF-17-T5-1:1-Py-A-48	48.00	10.797	0.0	347.2	2.7	0.1	0.0	37.1	12.7	0.0	1.6	0.9	1.7	0.3	0.5
SF-17-T5-1:1-Py-A-0	0.00	0.103	224.0	0.0	0.0	0.0	0.0	0.1	6.8	0.0	0.0	0.0	0.0	0.0	0.0
SF-17-T6-1:1-Py-B-3	3.00	0.125	222.4	0.0	0.0	0.0	0.0	0.7	6.8	0.0	0.0	0.0	0.0	0.0	0.0
SF-17-T6-1:1-Py-B-6	6.00	0.149	220.7	1.4	0.0	0.1	0.0	2.0	6.8	0.0	0.0	0.2	0.0	0.0	0.0
SF-17-T6-1:1-Py-B-8	8.00	0.173	218.7	2.4	0.1	0.1	0.0	3.9	7.0	0.0	0.2	0.3	0.0	0.0	0.0
SF-17-T6-1:1-Py-B-24	24.00	0.602	175.8	24.6	2.0	0.3	0.0	32.7	7.5	0.0	2.1	2.5	0.0	0.0	0.0
SF-17-T6-1:1-Py-B-31	31.00	0.727	149.4	46.0	3.8	0.4	0.0	48.4	8.2	0.0	2.5	3.2	0.7	0.0	0.0
SF-17-16-1:1-Py-B-48	48.00	1.477	68.8	140.3	9.1	0.5	0.0	88.0	10.0	0.0	3.0	3.8	2.0	0.0	0.6
SF-17-18-11:1-Py-B-0	0.00	0.110	223.6	0.0	0.0	0.0	0.0	0.1	6.8	0.0	0.0	0.0	0.0	0.0	0.0
SF-17-17-11:1-Py-A-3	3.00	0.146	223.6	0.0	0.0	0.1	0.0	0.3	6.8	0.0	0.0	0.0	0.0	0.0	0.0
SF-17-17-11:1-Py-A-6	6.00	0.174	224.2	0.9	0.0	0.1	0.0	0.8	6.7	0.3	0.4	0.0	0.0	0.0	0.0
SF-1/-1/-11:1-Py-A-8	8.00	0.216	221.3	2.5	0.1	0.2	0.0	1.6	6.6	0.4	0.7	0.0	0.0	0.0	0.0
SF-1/-1/-11:1-Py-A-24	24.00	9.517	0.0	345.0	3.0	0.3	0.0	37.3	8.8	0.8	3.4	0.0	1.2	0.4	0.5
SF-17-17-11:1-Py-A-31	31.00	10.047	0.0	336.0	2.9	0.3	0.0	37.4	10.1	0.8	3.4	0.9	1.4	0.4	0.5
SF-17-T7-11:1-Py-A-48	48.00	11.347		344.2	2.8	0.2	0.0	37.2	11.4	0.6	3.5	0.9	1.4	0.3	0.5
SF-17-T8-11:1-Py-B-0	0.00	0.108	223.6	0.0	0.0	0.0	0.0	0.1	6.8	0.0	0.0	0.0	0.0	0.0	0.0
SF-17-T8-11:1-Py-B-3	3.00	0.136	223.2	0.0	0.0	0.1	0.0	0.3	6.8	0.0	0.0	0.0	0.0	0.0	0.0
SF-17-T8-11:1-Py-B-6	6.00	0.150	224.5	0.4	0.0	0.2	0.0	0.7	6.7	0.3	0.2	0.0	0.0	0.0	0.0
SF-17-T8-11:1-Py-B-8	8.00	0.161	222.1	0.7	0.0	0.2	0.0	1.1	6.8	0.3	0.2	0.0	0.0	0.0	0.0
SF-17-T8-11:1-Py-B-24	24.00	0.277	209.8	6.0	0.4	0.6	0.0	6.6	6.6	0.8	1.8	0.0	0.0	0.0	0.0
SF-17-T8-11:1-Pv-B-31	31.00	0.327	205.2	10.2	0.6	0.8	0.0	10.7	6.7	1.1	3.6	0.3	0.0	0.0	0.0

TABLE	14-continued
IADLE	14-conunueu

Optical density (OD) and extracellular compound concentrations at the different sampling time points (EPT = elapsed process time) of different pure and mixed cultures. Abbreviations used were: EtOH = ethanol, PYR = pyruvate, KTV = ketoisovalerate, DHIV = dihydroisovalerate, DHMB = 2,3-dihydroxy-2-metylbutyrate, GLY = glycerol, ACE = acetate, IBOOH = isobutyric acid, IBOH = isobutanol, m-BDO = meso-butanediol, d/l-BDO = d/l-butanediol, LAC = lactate, SUC = succinate.

							DHIV $+$					m-	d/l-		
Sample	EPT	OD	GLC	EtOH	PYR	KIV	DHMB	GLY	ACE	IBOOH	IBOH	BDO	BDO	LAC	SUC
[]	[h]	[]	[mM]	[mM]	[mM]	[mM]	[mM]	[mM]	[mM]	[mM]	[mM]	[mM]	[mM]	[mM]	[mM]
SE-17-T8-11-1-Pv-B-48	48.00	0.667	164.4	31.3	2.0	13	0.0	20.7	73	2.0	12.1	12	0.0	0.0	0.0
SF-17-T9-1:1-Bb-A-0	0.00	0.104	224.2	0.0	0.0	0.0	0.0	0.1	6.9	0.0	0.0	0.0	0.0	0.0	0.0
SF-17-T9-1:1-Bb-A-3	3.00	0.144	223.2	1.4	0.0	0.1	0.0	0.2	6.8	0.0	0.0	0.0	0.0	0.0	0.0
SF-17-T9-1:1-Bb-A-6	6.00	0.330	219.1	7.3	0.0	0.1	0.0	0.7	6.6	0.0	0.4	0.0	0.0	0.0	0.0
SF-17-T9-1:1-Bb-A-8	8.00	1.057	207.4	26.7	0.1	0.1	0.0	2.0	6.6	0.2	0.7	0.0	0.0	0.0	0.0
SF-17-19-1:1-Bb-A-24 SE 17 TO 1-1 Db A 31	24.00	9.877	0.0	3/1.9	2.7	0.2	0.0	20.5	9.9	0.5	2.0	0.0	0.3	0.3	0.5
SF-17-T9-1:1-Bb-A-48	48.00	12 247	0.0	375.3	2.0	0.2	0.0	20.5	11.8	0.7	2.0	0.0	0.5	0.3	0.5
SF-17-T9-1:1-Bb-A-0	0.00	0.106	224.2	0.0	0.0	0.0	0.0	0.1	6.9	0.0	0.0	0.0	0.0	0.0	0.0
SF-17-T10-1:1-Bb-B-3	3.00	0.114	223.7	0.0	0.0	0.0	0.0	0.2	6.8	0.0	0.0	0.0	0.0	0.0	0.0
SF-17-T10-1:1-Bb-B-6	6.00	0.139	223.4	1.4	0.0	0.0	0.0	0.3	6.8	0.0	0.2	0.0	0.0	0.0	0.0
SF-17-T10-1:1-Bb-B-8	8.00	0.189	223.8	3.8	0.0	0.1	0.0	0.4	6.8	0.0	0.3	0.0	0.0	0.0	0.0
SF-17-T10-1:1-Bb-B-31	31.00	11.647	0.0	364.2	2.0	0.1	0.0	19.2	9.5	0.2	1.5	0.0	0.2	0.3	0.6
SF-17-T10-1:1-Bb-B-48	48.00	11.747	0.0	380.5	2.5	0.1	0.0	19.2	12.0	0.3	1.3	0.0	0.3	0.3	0.6
SF-17-T12-11:1-Bb-B-0	0.00	0.110	223.9	0.0	0.0	0.0	0.0	0.1	6.8	0.0	0.0	0.0	0.0	0.0	0.0
SF-17-T11-11:1-Bb-A-3	3.00	0.144	224.0	0.0	0.0	0.1	0.0	0.2	6.8	0.0	0.0	0.0	0.0	0.0	0.0
SF-17-T11-11:1-Bb-A-6	6.00	0.187	222.0	1.0	0.0	0.0	0.0	0.4	6.6	0.0	0.6	0.0	0.0	0.0	0.0
SF-17-T11-11:1-Bb-A-24	24.00	10.239	220.8	363.4 363.4	2.8	0.2	0.0	223	0.5 8.4	0.0	1.0	0.0	0.2	0.0	0.0
SF-17-T11-11:1-Bb-A-31	31.00	11.197	0.0	353.8	2.0	0.5	0.0	22.3	10.0	1.0	4.3	0.6	0.4	0.4	0.5
SF-17-T11-11:1-Bb-A-48	48.00	11.497	0.0	365.6	2.7	0.4	0.0	22.2	11.3	1.4	4.5	0.6	0.5	0.3	0.5
SF-17-T12-11:1-Bb-B-0	0.00	0.107	223.9	0.0	0.0	0.0	0.0	0.1	6.8	0.0	0.0	0.0	0.0	0.0	0.0
SF-17-T12-11:1-Bb-B-3	3.00	0.123	222.4	0.0	0.0	0.0	0.0	0.2	6.7	0.0	0.0	0.0	0.0	0.0	0.0
SF-17-T12-T1:1-Bb-B-6 SE 17 T12 11:1 Ph P 8	6.00 8.00	0.134	222.4	0.0	0.0	0.1	0.0	0.2	6.6	0.0	0.4	0.0	0.0	0.0	0.0
SF-17-T12-11:1-Bb-B-24	24.00	10 317	16.0	344.8	27	0.1	0.0	20.4	6.4	0.0	2.8	0.0	0.0	0.0	0.0
SF-17-T12-11:1-Bb-B-31	31.00	11.447	0.0	360.8	2.7	0.2	0.0	21.4	9.1	0.5	2.8	0.0	0.2	0.3	0.6
SF-17-T12-11:1-Bb-B-48	48.00	11.897	0.0	372.0	2.6	0.2	0.0	21.1	10.8	0.6	2.8	0.0	0.3	0.4	0.6
SF-17-T14-1:1-Ca-B-0	0.00	0.106	233.7	0.0	0.0	0.0	0.0	0.1	7.2	0.0	0.0	0.0	0.0	0.0	0.0
SF-17-T13-1:1-Ca-A-3	3.00	0.142	222.8	1.6	0.0	0.0	0.0	0.3	6.8	0.0	0.0	0.0	0.0	0.0	0.0
SF-17-T13-1:1-Ca-A-8	8.00	1 417	201.3	9.0 36.8	0.0	0.0	0.0	2.5	6.5	0.0	0.5	0.0	0.0	0.0	0.0
SF-17-T13-1:1-Ca-A-24	24.00	11.347	0.0	374.9	2.7	0.1	0.0	18.4	11.2	0.5	1.7	0.0	0.2	0.4	0.5
SF-17-T13-1:1-Ca-A-31	31.00	10.647	0.0	358.7	2.6	0.1	0.0	18.5	11.9	0.7	1.6	0.0	0.3	0.3	0.5
SF-17-T13-1:1-Ca-A-48	48.00	12.097	0.0	369.5	2.5	0.1	0.0	18.4	13.0	0.7	1.6	0.2	0.2	0.3	0.5
SF-17-T14-1:1-Ca-B-0	0.00	n.d.	233.7	0.0	0.0	0.0	0.0	0.1	7.2	0.0	0.0	0.0	0.0	0.0	0.0
SF-17-T14-1:1-Ca-B-3	3.00	n.d.	233.9	0.0	0.0	0.0	0.0	0.1	7.2	0.0	0.0	0.0	0.0	0.0	0.0
SF-17-114-1:1-Ca-B-6 SE 17 T14 1:1 Co B 8	6.00 8.00	n.d.	233.4	0.0	0.0	0.0	0.0	0.0	7.1	0.0	0.0	0.0	0.0	0.0	0.0
SF-17-T14-1:1-Ca-B-0 SF-17-T14-1:1-Ca-B-24	24.00	n.d.	237.4	0.0	0.0	0.0	0.0	0.1	7.2	0.0	0.0	0.0	0.0	0.0	0.0
SF-17-T14-1:1-Ca-B-31	31.00	n.d.	234.0	0.0	0.0	0.0	0.0	0.1	7.1	0.0	0.0	0.0	0.0	0.0	0.0
SF-17-T14-1:1-Ca-B-48	48.00	n.d.	234.7	0.0	0.0	0.0	0.0	0.0	7.5	0.0	0.0	0.0	0.0	0.0	0.0
SF-17-T15-11:1-Ca-A-0	0.00	0.113	232.1	0.0	0.0	0.0	0.0	0.1	7.0	0.0	0.0	0.0	0.0	0.0	0.0
SF-17-T15-11:1-Ca-A-3	3.00	0.141	231.6	0.0	0.0	0.0	0.0	0.2	7.0	0.0	0.0	0.0	0.0	0.0	0.0
SF-17-T15-11:1-Ca-A-6	6.00	0.199	230.2	1.4	0.0	0.1	0.0	0.5	6.8	0.0	0.4	0.0	0.0	0.0	0.0
SF-17-T15-11:1-Ca-A-8	8.00	0.387	228.2	6.5	0.1	0.2	0.0	1.0	6.8	0.6	0.8	0.0	0.0	0.0	0.0
SF-17-115-11:1-Ca-A-24 SE 17 T15-11:1-Ca-A-31	24.00	11.047	0.0	366.1	2.9	0.3	0.0	22.4	9.5	1.1	3.5	0.4	0.3	0.3	0.5
SF-17-T15-11:1-Ca-A-48	48.00	12 147	0.0	382.7	2.9	0.3	0.0	22.0	11.9	1.5	3.7	0.4	0.2	0.4	0.5
SF-17-T15-11:1-Ca-A-0	0.00	n.d.	232.1	0.0	0.0	0.0	0.0	0.1	7.0	0.0	0.0	0.0	0.0	0.0	0.0
SF-17-T16-11:1-Ca-B-3	3.00	n.d.	224.4	0.0	0.0	0.0	0.0	0.1	6.9	0.0	0.0	0.0	0.0	0.0	0.0
SF-17-T16-11:1-Ca-B-6	6.00	n.d.	223.9	0.0	0.0	0.0	0.0	0.1	6.7	0.0	0.0	0.0	0.0	0.0	0.0
SF-17-T16-11:1-Ca-B-8	8.00	n.d.	227.0	0.0	0.0	0.0	0.0	0.1	7.1	0.0	0.0	0.0	0.0	0.0	0.0
SF-17-T16-11:1-Ca-B-24	24.00	n.d.	224.1	0.0	0.0	0.0	0.0	0.0	6.8	0.0	0.0	0.0	0.0	0.0	0.0
SF-17-T16-11:1-Ca-B-31	31.00	n.d.	226.4	0.0	0.0	0.0	0.0	0.1	6.9	0.0	0.0	0.0	0.0	0.0	0.0
SF-1/-110-11:1-Ca-B-48	48.00	n.d.	225.3	0.0	0.0	0.0	0.0	0.1	0.8 4 9	0.0	0.0	0.0	0.0	0.0	0.0
SF-17-T17-1:1-ctrl-1-3	3.00	0.101	223.9	2.5	0.0	0.0	0.0	0.1	6.8	0.0	0.0	0.0	0.0	0.0	0.0
SF-17-T17-1:1-ctrl-1-6	6.00	0.862	211.4	17.9	0.1	0.1	0.0	1.3	6.5	0.0	0.8	0.0	0.0	0.0	0.0
SF-17-T17-1:1-ctrl-1-8	8.00	2.787	181.0	67.6	0.4	0.2	0.0	4.7	6.5	0.2	1.2	0.0	0.0	0.1	0.2
SF-17-T17-1:1-ctrl-1-24	24.00	12.097	0.0	372.5	2.7	0.2	0.0	19.6	9.7	0.3	2.2	0.0	0.2	0.4	0.5
SF-17-T17-1:1-ctrl-1-31	31.00	11.997	0.0	358.7	2.7	0.2	0.0	19.9	10.3	0.0	2.1	0.3	0.4	0.3	0.6
SF-17-T17-1:1-ctrl-1-48	48.00	12.297	0.0	374.6	2.6	0.2	0.0	19.4	11.0	0.3	2.3	0.0	0.4	0.3	0.6
SF-17-T17-1:1-ctrl-1-0	0.00	0.114	223.9	0.0	0.0	0.0	0.0	0.1	6.8	0.0	0.0	0.0	0.0	0.0	0.0
SF-17-T18-11:1-ctrl-1-3	3.00	0.177	222.1	0.0	0.0	0.1	0.0	0.2	6.7	0.0	0.7	0.0	0.0	0.0	0.0
SF-1/-118-11:1-ctrl-1-6	0.00 8.00	0.303	218.8	2.9	0.1	0.3	0.0	0.5	0.4	0.3	1.5	0.0	0.0	0.0	0.0
SF-17-T18-11:1-ctrl-1-24	8.00 24.00	11.397	212.3	354.1	3.1	0.4	0.0	21.6	0.4 8.4	0.5	2.3 7.2	0.0	0.4	0.4	0.0
VULL AT	=	~~··//	v.v	iii		J.J	~ • *		0.7		· • 4	0.0	. T	. T	0.0

TABLE 14-continued

Optical density (OD) and extracellular compound concentrations at the different sampling time points (EPT = elapsed process time) of different pure and mixed cultures. Abbreviations used were: EtOH = ethanol, PYR = pyruvate, KTV = ketoisovalerate, DHIV = dihydroisovalerate, DHMB = 2,3-dihydroxy-2-metylbutyrate, GLY = glycerol, ACE = acetate, IBOOH = isobutyric acid, IBOH = isobutanol, m-BDO = meso-butanediol, d/l-BDO = d/l-butanediol, LAC = lactate, SUC = succinate.

Sample []	EPT [h]	OD []	GLC [mM]	EtOH [mM]	PYR [mM]	KIV [mM]	DHIV + DHMB [mM]	GLY [mM]	ACE [mM]	IBOOH [mM]	IBOH [mM]	m- BDO [mM]	d/l- BDO [mM]	LAC [mM]	SUC [mM]
SF-17-T18-11:1-ctrl-1-31	31.00	10.947	0.0	351.4	3.1	0.6	$\begin{array}{c} 0.1 \\ 0.1 \end{array}$	21.9	9.6	1.0	7.1	0.8	0.5	0.4	0.5
SF-17-T18-11:1-ctrl-1-48	48.00	11.797	0.0	355.5	3.0	0.6		21.4	10.5	0.8	7.3	0.9	0.7	0.3	0.5

Materials & Methods for Examples 21-25

Yeast synthetic medium w/o amino acids, w/o glucose, 15 w/o ethanol/Tween (2×): 13.4 g/l, Yeast Nitrogen Base w/o amino acids (Difco 0919-15-3); 40 mg/L thiamine; 40 mg/L niacin; 200 ml/L 1M MES buffer, pH=5.5; Supplement aa sol. without histidine and uracil (SAAS-1 10×); 18.5 g/L, Synthetic complete amino acid dropout (Kaiser)-His, -Ura <sup>20</sup> (Formedium).

SEED medium: 10.000 mL Yeast synthetic medium w/o aa, w/o glucose, w/o ethanol/Tween (2×); 2.000 mL Supplement aa sol. without histidine and uracil (SAAS-1 10×); 3.200 mL 250 g/L glucose solution (resulting in 40 g/l <sup>25</sup> glucose); 0.046 mL Na-acetate stock solution; 4.754 mL  $H_2O$ 

# Example 21 (Prophetic):

#### Construction of Isobutanologen Strains Expressing a Formaldehyde Dehydrogenase

*P. putida* fdhA (SEQ ID NO:7) (GI:1169603) and *S. cerevisiae* SFA1 (SEQ ID NO:6) (van den Berg et al., *Yeast* 35 13(6): 551-9 (1997)) are used to synthesize genes in vitro using codon-optimization algorithms for *S. cerevisiae* (e.g. DNA 2.0). The gene cassettes are designed to place 5' BamHI and 3' MluI restriction sites for subcloning of the coding sequences into expression plasmid pBTX1 (SEQ ID 40 NO:15). pBTX1 is derived from the pRS413 vector backbone (ATCC #87518) and contains the FBA1 promoter, multiple cloning site (BamHI, MluI), and ADH1 terminator.

An isobutanologen is constructed by transformation of plasmids pBTX1::SFA1 and pLH804::L2V4 into the host 45 strain PNY2145. Plasmid pLH804::L2V4 is derived from the pHR81 vector backbone (ATCC #87541) and contains: the A. caccae K9JB4P KARI driven by the ILV5 promoter and ILV5 terminator, and the S. mutans L2V4 DHAD driven by the TEF1(M7) promoter and FBA1 terminator (SEQ ID 50 NO:22). Plasmids are introduced by lithium acetate transformation method (Methods in Yeast Genetics, 2005, page 113), and transformants are selected on synthetic complete medium, minus histidine and uracil, with 1% ethanol as carbon source. Transformants are then transferred to plates 55 containing synthetic complete medium, minus histidine and uracil, with 2% glucose as carbon source and optionally ethanol (0.05%) or acetate (2 mM) as a C2 supplement. Freezer vials are made by dilution of log-phase cultures with 45% glycerol to a final glycerol concentration of 15% (w/v). 60

An isobutanologen is constructed by transformation of plasmids pBTX1::fdhA and pLH804::L2V4 into the host strain PNY2145. Plasmid pLH804::L2V4 is derived from the pHR81 vector backbone (ATCC #87541) and contains: the *A. caccae* K9JB4P KARI driven by the ILV5 promoter 65 and ILV5 terminator, and the *S. mutans* L2V4 DHAD driven by the TEF1(M7) promoter and FBA1 terminator (SEQ ID

NO:22). Plasmids are introduced by lithium acetate transformation method (Methods in Yeast Genetics, 2005, page 113), and transformants are selected on synthetic complete medium, minus histidine and uracil, with 1% ethanol as carbon source. Transformants are then transferred to plates containing synthetic complete medium, minus histidine and uracil, with 2% glucose as carbon source and optionally ethanol (0.05%) or acetate (2 mM) as a C2 supplement. Freezer vials are made by dilution of log-phase cultures with 45% glycerol to a final glycerol concentration of 15% (w/v).

#### Example 22 (Prophetic):

#### Construction of Isobutanologen Strains Expressing a Sulfonylurea-Resistant ALS (e.g. SMR1-410)

To construct an expression plasmid, the protein coding <sup>30</sup> sequence for *S. cerevisiae* SMR1-410 (SEQ ID NO:9; nucleic acid sequence SEQ ID NO:8) is used to synthesize genes in vitro using codon-optimization algorithms for *S. cerevisiae* (e.g. DNA 2.0). The SMR1-410 gene cassette is designed to place 5' BamHI and 3' MluI restriction sites for <sup>35</sup> subcloning of the coding sequences into expression plasmid pBTX1 (SEQ ID NO:15). pBTX1 is derived from the pRS413 vector backbone (ATCC #87518) and contains the FBA1 promoter, multiple cloning site (BamHI, MluI), and ADH1 terminator.

An isobutanologen is constructed by transformation of plasmids pBTX1::SMR1-410 and pLH804::L2V4 into the host strain PNY2145 (referenced in US Pat. Publ. No. 2014/0004526, which is incorporated herein by reference in its entirety, and described in Example 26). Plasmid pLH804::L2V4 is derived from the pHR81 vector backbone (ATCC #87541) and contains: the A. caccae K<sub>9</sub>JB4P KARI driven by the ILV5 promoter and ILV5 terminator, and the S. mutans L2V4 DHAD driven by the TEF1(M7) promoter and FBA1 terminator (SEQ ID NO:22). Plasmids are introduced by lithium acetate transformation method (Methods in Yeast Genetics, 2005, page 113), and transformants are selected on synthetic complete medium, minus histidine and uracil, with 1% ethanol as carbon source. Transformants are then transferred to plates containing synthetic complete medium, minus histidine and uracil, with 2% glucose as carbon source and optionally ethanol (0.05%) or acetate (2 mM) as a C2 supplement. Freezer vials are made by dilution of log-phase cultures with 45% glycerol to a final glycerol concentration of 15% (w/v).

# Example 23 (Prophetic):

#### Construction of Isobutanologen Strains Expressing Genes Conferring Sulfite Resistance

To construct expression plasmids, the protein coding sequences for *S. cerevisiae* FZF1-4 (SEQ ID NO:11) (Park,

Lopez et al. 1999) and SSU1 (SEQ ID NO:12) are used to synthesize genes in vitro using codon-optimization algorithms for *S. cerevisiae* (e.g. DNA 2.0). SEQ ID NO:10 is the wild type protein sequence for FZF1. The gene cassettes are designed to place 5' BamHI and 3' MluI restriction sites for 5 subcloning of the coding sequences into expression plasmid pBTX1 (SEQ ID NO:15). pBTX1 is derived from the pRS413 vector backbone (ATCC #87518) and contains the FBA1 promoter, multiple cloning site (BamHI, MluI), and ADH1 terminator. 10

An isobutanologen is constructed by transformation of plasmids pBTX1::FZF1-4 and pLH804::L2V4 into the host strain PNY2145 (described herein) that contains a deletion of the chromosomal FZF1 gene. The FZF1 deletion in PNY2145 is made using standard yeast deletions using a 15 kanMX4 cassette (Brachmann, et al. Designer deletion strains derived from Saccharomyces cerevisiae S288C: a useful set of strains and plasmids for PCR-mediated gene disruption and other applications. Yeast. 14, 115-132 (1998). Plasmid pLH804::L2V4 is derived from the pHR81 vector 20 backbone (ATCC #87541) and contains: the A. caccae K9JB4P KARI driven by the ILV5 promoter and ILV5 terminator, and the S. mutans L2V4 DHAD driven by the TEF1(M7) promoter and FBA1 terminator (SEQ ID NO:22). Plasmids are introduced by lithium acetate trans- 25 formation method (Methods in Yeast Genetics, 2005, page 113), and transformants are selected on synthetic complete medium, minus histidine and uracil, with 1% ethanol as carbon source. Transformants are then transferred to plates containing synthetic complete medium, minus histidine and 30 uracil, with 2% glucose as carbon source and optionally ethanol (0.05%) or acetate (2 mM) as a C2 supplement. Freezer vials are made by dilution of log-phase cultures with 45% glycerol to a final glycerol concentration of 15% (w/v).

An isobutanologen is constructed by transformation of 35 plasmids pBTX1::SSU/and pLH804::L2V4 into the host strain PNY2145 (described herein). Plasmid pLH804::L2V4 is derived from the pHR81 vector backbone (ATCC #87541) and contains: the A. caccae K9JB4P KARI driven by the ILV5 promoter and ILV5 terminator, and the S. mutans 40 L2V4 DHAD driven by the TEF1(M7) promoter and FBA1 terminator (SEQ ID NO:22). Plasmids are introduced by lithium acetate transformation method (Methods in Yeast Genetics, 2005, page 113), and transformants are selected on synthetic complete medium, minus histidine and uracil, with 45 1% ethanol as carbon source. Transformants are then transferred to plates containing synthetic complete medium, minus histidine and uracil, with 2% glucose as carbon source and optionally ethanol (0.05%) or acetate (2 mM) as a C2 supplement. Freezer vials are made by dilution of log-phase 50 cultures with 45% glycerol to a final glycerol concentration of 15% (w/v).

#### Example 24 (Prophetic):

# Construction of Isobutanologen Strains Expressing a Glyphosate Resistance 3-phosphoshikimate 1-carboxylvinyltransferase

To construct an expression plasmid, the protein coding 60 sequence for *Salmonella typhi* aro $A^{GLY+}$  (SEQ ID NO:13) (Stalker, et al., *J Biol Chem* 260(8): 4724-8 (1985)) is used to synthesize genes in vitro using codon-optimization algorithms for *S. cerevisiae* (e.g. DNA 2.0). The aro $A^{GLY+}$  gene cassette is designed to place 5' BamHI and 3' MluI restriction sites for subcloning of the coding sequences into expression plasmid pBTX1 (SEQ ID NO:15). pBTX1 is

derived from the pRS413 vector backbone (ATCC #87518) and contains the FBA1 promoter, multiple cloning site (BamHI, MluI), and ADH1 terminator.

An isobutanologen is constructed by transformation of plasmids pBTX1::aroA<sup>GLY+</sup> and pLH804::L2V4 into the host strain PNY2145 (described herein). Plasmid pLH804:: L2V4 is derived from the pHR81 vector backbone (ATCC #87541) and contains: the A. caccae K9JB4P KARI driven by the ILV5 promoter and ILV5 terminator, and the S. mutans L2V4 DHAD driven by the TEFL (M7) promoter and FBA1 terminator (SEQ ID NO:22). Plasmids are introduced by lithium acetate transformation method (Methods in Yeast Genetics, 2005, page 113), and transformants are selected on synthetic complete medium, minus histidine and uracil, with 1% ethanol as carbon source. Transformants are then transferred to plates containing synthetic complete medium, minus histidine and uracil, with 2% glucose as carbon source and optionally ethanol (0.05%) or acetate (2 mM) as a C2 supplement. Freezer vials are made by dilution of log-phase cultures with 45% glycerol to a final glycerol concentration of 15% (w/v).

# Example 25 (Prophetic):

#### Genetic Engineering for Increased Inhibitor Tolerance in Butanologen Yeast

In some embodiments, the butanologen is engineered for increased inhibitor tolerance by expressing or overexpressing a formaldehyde dehydrogenase. The formaldehyde dehydrogenase is selected from one of the following EC groups: EC 1.1.1.284, EC 1.1.1.1, EC 1.2.1.46, EC 1.2.1.66, EC 3.1.2.12, EC 1.2.2.B1 and EC 1.2.2.B2 are no official designators, but are defined by the BRENDA protein database. Especially suited formaldehyde dehydrogenases are:

SFA1 (YDL168W, ADH5): glutathione-dependent formaldehyde dehydrogenase (van den Berg et al., *Yeast* 13(6): 551-9 (1997)) (SEQ ID NO:6) and *Pseudomonas putida* glutathione-independent formaldehyde dehydrogenase (SEQ ID NO:7).

In some embodiments, the butanologen is engineered for increased inhibitor tolerance by expressing or overexpressing a sulfonylurea-resistant ALS (e.g. SMR1-410) (Yadav et al., *Proc Natl Acad Sci USA* 83(12): 4418-22 (1986)) (SEQ ID NO:9).

In some embodiments, the butanologen is engineered for increased inhibitor tolerance by expressing or overexpressing other sulfonylurea-resistant ALS enzymes that qualify for (over)expression.

In some embodiments, the butanologen is engineered for increased inhibitor tolerance by expressing or overexpressing sulfite resistance by convert FZF1 (SEQ ID NO:10) to FZF1-4 (SEQ ID NO:11) (Park et al., *Curr Genet* 36(6): 55 339-44. (1999)) or overexpressing SSU1 (SEQ ID NO:12).

FZF1 (YGL254W, NRC299, RSU1 2, SUL1) is a transcription factor involved in sulfite metabolism, sole identified regulatory target is SSU1, overexpression suppresses sulfite-sensitivity of many unrelated mutants due to hyperactivation of SSU1.

Overexpression of SSU1 (YPL092W, LPG16). SSU1 is a plasma membrane sulfite pump involved in sulfite metabolism and required for efficient sulfite efflux. Homolog enzymes may be considered for overexpression as well to confer increased tolerance/improved competitiveness.

In some embodiments, the butanologen is engineered for increased inhibitor tolerance by expressing or overexpressing a glyphosate resistance 3-phosphoshikimate 1-carboxylvinyltransferase (e.g. aroA<sup>gly+</sup>) (SEQ ID NO:13) (Stalker et al., J Biol Chem 260(8): 4724-8 (1985)).

All other glyphosate resistant 3-phosphoshikimate 1-carboxyvinyltransferases qualify for expression or overexpres- 5 sion.

#### Example 26

#### Strain Construction

#### Construction of Strain PNY2115

Saccharomyces cerevisiae strain PNY0827 is used as the host cell for further genetic manipulation for PNY2115. PNY0827 refers to a strain derived from Saccharomyces 15 cerevisiae which has been deposited at the ATCC under the Budapest Treaty on Sep. 22, 2011 at the American Type Culture Collection, Patent Depository 10801 University Boulevard, Manassas, Va. 20110-2209 and has the patent deposit designation PTA-12105.

Deletion of URA3 and Sporulation into Haploids In order to delete the endogenous URA3 coding region, a deletion cassette was PCR-amplified from pLA54 (SEQ ID NO:158) which contains a  $P_{TEF1}$ -kanMX4-TEF1t cassette flanked by loxP sites to allow homologous recombination in 25 vivo and subsequent removal of the KANMX4 marker. PCR was done by using Phusion High Fidelity PCR Master Mix (New England BioLabs; Ipswich, Mass.) and primers BK505 (SEQ ID NO:101) and BK506 (SEQ ID NO:102). The URA3 portion of each primer was derived from the 5' 30 region 180 bp upstream of the URA3 ATG and 3' region 78 bp downstream of the coding region such that integration of the kanMX4 cassette results in replacement of the URA3 coding region. The PCR product was transformed into PNY0827 using standard genetic techniques (Methods in 35 Yeast Genetics, 2005, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., pp. 201-202) and transformants were selected on YEP medium supplemented 2% glucose and 100 µg/ml Geneticin at 30° C. Transformants were screened by colony PCR with primers LA468 (SEQ ID 40 recycled by plating on synthetic complete medium supple-NO:161) and LA492 (SEQ ID NO:104) to verify presence of the integration cassette. A heterozygous diploid was obtained: NYLA98, which has the genotype MATa/ $\alpha$ URA3/ura3::loxP-kanMX4-loxP. To obtain haploids, NYLA98 was sporulated using standard methods (Codón A 45 C, Gasent-Ramirez J M, Benítez T. Factors which affect the frequency of sporulation and tetrad formation in Saccharomyces cerevisiae baker's yeast. Appl Environ Microbiol. 1995 PMID: 7574601). Tetrads were dissected using a micromanipulator and grown on rich YPE medium supple- 50 mented with 2% glucose. Tetrads containing four viable spores were patched onto synthetic complete medium lacking uracil supplemented with 2% glucose, and the mating type was verified by multiplex colony PCR using primers AK109-1 (SEQ ID NO:105), AK109-2 (SEQ ID NO: 106), 55 and AK109-3 (SEQ ID NO:107). The resulting identified haploid strain called NYLA103, which has the genotype: MAT $\alpha$  ura3 $\Delta$ ::loxP-kanMX4-loxP, and NYLA106, which has the genotype: MATa ura $3\Delta$ ::loxP-kanMX4-loxP. Deletion of His3 60

To delete the endogenous HIS3 coding region, a scarless deletion cassette was used. The four fragments for the PCR cassette for the scarless HIS3 deletion were amplified using Phusion High Fidelity PCR Master Mix (New England BioLabs; Ipswich, Mass.) and CEN.PK 113-7D genomic 65 DNA as template, prepared with a Gentra Puregene Yeast/ Bact kit (Qiagen; Valencia, Calif.). HIS3 Fragment A was

amplified with primer oBP452 (SEQ ID NO:89) and primer oBP453 (SEQ ID NO:109), containing a 5' tail with homology to the 5' end of HIS3 Fragment B. HIS3 Fragment B was amplified with primer oBP454 (SEQ ID NO:110), containing a 5' tail with homology to the 3' end of HIS3 Fragment A, and primer oBP455 (SEQ ID NO:90) containing a 5' tail with homology to the 5' end of HIS3 Fragment U. HIS3 Fragment U was amplified with primer oBP456 (SEQ ID NO:91), containing a 5' tail with homology to the 3' end of 10 HIS3 Fragment B, and primer oBP457 (SEQ ID NO:86), containing a 5' tail with homology to the 5' end of HIS3 Fragment C. HIS3 Fragment C was amplified with primer oBP458 (SEQ ID NO:87), containing a 5' tail with homology to the 3' end of HIS3 Fragment U, and primer oBP459 (SEQ ID NO:88). PCR products were purified with a PCR Purification kit (Qiagen). HIS3 Fragment AB was created by overlapping PCR by mixing HIS3 Fragment A and HIS3 Fragment B and amplifying with primers oBP452 (SEQ ID NO:89) and oBP455 (SEQ ID NO:90). HIS3 Fragment UC 20 was created by overlapping PCR by mixing HIS3 Fragment U and HIS3 Fragment C and amplifying with primers oBP456 (SEQ ID NO:91) and oBP459 (SEQ ID NO:88). The resulting PCR products were purified on an agarose gel followed by a Gel Extraction kit (Qiagen). The HIS3 ABUC cassette was created by overlapping PCR by mixing HIS3 Fragment AB and HIS3 Fragment UC and amplifying with primers oBP452 (SEQ ID NO:89) and oBP459 (SEQ ID NO:88). The PCR product was purified with a PCR Purification kit (Qiagen). Competent cells of NYLA106 were transformed with the HIS3 ABUC PCR cassette and were plated on synthetic complete medium lacking uracil supplemented with 2% glucose at 30° C. Transformants were screened to verify correct integration by replica plating onto synthetic complete medium lacking histidine and supplemented with 2% glucose at 30° C. Genomic DNA preps were made to verify the integration by PCR using primers oBP460 (SEQ ID NO:93) and LA135 (SEQ ID NO:94) for the 5' end and primers oBP461 (SEQ ID NO:95) and LA92 (SEQ ID NO:96) for the 3' end. The URA3 marker was mented with 2% glucose and 5-FOA at 30° C. following standard protocols. Marker removal was confirmed by patching colonies from the 5-FOA plates onto SD -URA medium to verify the absence of growth. The resulting identified strain, called PNY2003 has the genotype: MATa ura3 $\Delta$ ::loxP-kanMX4-loxP his3 $\Delta$ .

Deletion of PDC1

To delete the endogenous PDC1 coding region, a deletion cassette was PCR-amplified from pLA59 (SEQ ID NO:97), which contains a URA3 marker flanked by degenerate loxP sites to allow homologous recombination in vivo and subsequent removal of the URA3 marker. PCR was done by using Phusion High Fidelity PCR Master Mix (New England BioLabs; Ipswich, Mass.) and primers LA678 (SEQ ID NO:98) and LA679 (SEQ ID NO:99). The PDC1 portion of each primer was derived from the 5' region 50 bp downstream of the PDC1 start codon and 3' region 50 bp upstream of the stop codon such that integration of the URA3 cassette results in replacement of the PDC1 coding region but leaves the first 50 bp and the last 50 bp of the coding region. The PCR product was transformed into PNY2003 using standard genetic techniques and transformants were selected on synthetic complete medium lacking uracil and supplemented with 2% glucose at 30° C. Transformants were screened to verify correct integration by colony PCR using primers LA337 (SEQ ID NO:111), external to the 5' coding region and LA135 (SEQ ID NO:94), an internal primer to URA3.

Positive transformants were then screened by colony PCR using primers LA692 (SEQ ID NO: 112) and LA693 (SEQ ID NO:113), internal to the PDC1 coding region. The URA3 marker was recycled by transforming with pLA34 (SEQ ID NO:184) containing the CRE recombinase under the GAL1 5 promoter and plated on synthetic complete medium lacking histidine and supplemented with 2% glucose at 30° C. Transformants were plated on rich medium supplemented with 0.5% galactose to induce the recombinase. Marker removal was confirmed by patching colonies to synthetic 10 complete medium lacking uracil and supplemented with 2% glucose to verify absence of growth. The resulting identified strain, called PNY2008 has the genotype: MATa ura3A:: loxP-kanMX4-loxP his $3\Delta$  pdc $1\Delta$ ::loxP71/66. Deletion of PDC5 15

To delete the endogenous PDC5 coding region, a deletion cassette was PCR-amplified from pLA59 (SEQ ID NO:97), which contains a URA3 marker flanked by degenerate loxP sites to allow homologous recombination in vivo and subsequent removal of the URA3 marker. PCR was done by 20 using Phusion High Fidelity PCR Master Mix (New England BioLabs; Ipswich, Mass.) and primers LA722 (SEQ ID NO:185) and LA733 (SEQ ID NO:186). The PDC5 portion of each primer was derived from the 5' region 50 bp upstream of the PDC5 start codon and 3' region 50 bp 25 downstream of the stop codon such that integration of the URA3 cassette results in replacement of the entire PDC5 coding region. The PCR product was transformed into PNY2008 using standard genetic techniques and transformants were selected on synthetic complete medium lacking 30 uracil and supplemented with 1% ethanol at 30° C. Transformants were screened to verify correct integration by colony PCR using primers LA453 (SEQ ID NO:187), external to the 5' coding region and LA135 (SEQ ID NO:94), an internal primer to URA3. Positive transformants were then 35 screened by colony PCR using primers LA694 (SEQ ID NO:188) and LA695 (SEQ ID NO:189), internal to the PDC5 coding region. The URA3 marker was recycled by transforming with pLA34 (SEQ ID NO:184) containing the CRE recombinase under the GAL1 promoter and plated on 40 synthetic complete medium lacking histidine and supplemented with 1% ethanol at 30° C. Transformants were plated on rich YEP medium supplemented with 1% ethanol and 0.5% galactose to induce the recombinase. Marker removal was confirmed by patching colonies to synthetic complete 45 medium lacking uracil and supplemented with 1% ethanol to verify absence of growth. The resulting identified strain, called PNY2009 has the genotype: MATa ura3A::loxPkanMX4-loxP his3 $\Delta$  pdc1 $\Delta$ ::loxP71/66 pdc5 $\Delta$ ::loxP71/66. Deletion of FRA2

The FRA2 deletion was designed to delete 250 nucleotides from the 3' end of the coding sequence, leaving the first 113 nucleotides of the FRA2 coding sequence intact. An in-frame stop codon was present 7 nucleotides downstream of the deletion. The four fragments for the PCR cassette for 55 the scarless FRA2 deletion were amplified using Phusion High Fidelity PCR Master Mix (New England BioLabs; Ipswich, Mass.) and CEN.PK 113-7D genomic DNA as template, prepared with a Gentra Puregene Yeast/Bact kit (Qiagen; Valencia, Calif.). FRA2 Fragment A was amplified 60 with primer oBP594 (SEQ ID NO:190) and primer oBP595 (SEQ ID NO:191), containing a 5' tail with homology to the 5' end of FRA2 Fragment B. FRA2 Fragment B was amplified with primer oBP596 (SEQ ID NO:192), containing a 5' tail with homology to the 3' end of FRA2 Fragment 65 A, and primer oBP597 (SEQ ID NO:193), containing a 5' tail with homology to the 5' end of FRA2 Fragment U. FRA2

Fragment U was amplified with primer oBP598 (SEQ ID NO:194), containing a 5' tail with homology to the 3' end of FRA2 Fragment B, and primer oBP599 (SEQ ID NO:195), containing a 5' tail with homology to the 5' end of FRA2 Fragment C. FRA2 Fragment C was amplified with primer oBP600 (SEQ ID NO:196), containing a 5' tail with homology to the 3' end of FRA2 Fragment U, and primer oBP601 (SEQ ID NO:197). PCR products were purified with a PCR Purification kit (Qiagen). FRA2 Fragment AB was created by overlapping PCR by mixing FRA2 Fragment A and FRA2 Fragment B and amplifying with primers oBP594 (SEQ ID NO:190) and oBP597 (SEQ ID NO:193). FRA2 Fragment UC was created by overlapping PCR by mixing FRA2 Fragment U and FRA2 Fragment C and amplifying with primers oBP598 (SEQ ID NO:194) and oBP601 (SEQ ID NO:197). The resulting PCR products were purified on an agarose gel followed by a Gel Extraction kit (Qiagen). The FRA2 ABUC cassette was created by overlapping PCR by mixing FRA2 Fragment AB and FRA2 Fragment UC and amplifying with primers oBP594 (SEQ ID NO:190) and oBP601 (SEQ ID NO:197). The PCR product was purified with a PCR Purification kit (Qiagen).

To delete the endogenous FRA2 coding region, the scarless deletion cassette obtained above was transformed into PNY2009 using standard techniques and plated on synthetic complete medium lacking uracil and supplemented with 1% ethanol. Genomic DNA preps were made to verify the integration by PCR using primers oBP602 (SEQ ID NO:198) and LA135 (SEQ ID NO:94) for the 5' end, and primers oBP602 (SEQ ID NO:198) and oBP603 (SEQ ID NO:199) to amplify the whole locus. The URA3 marker was recycled by plating on synthetic complete medium supplemented with 1% ethanol and 5-FOA (5-Fluoroorotic Acid) at 30° C. following standard protocols. Marker removal was confirmed by patching colonies from the 5-FOA plates onto synthetic complete medium lacking uracil and supplemented with 1% ethanol to verify the absence of growth. The resulting identified strain, PNY2037, has the genotype: MATa ura3A::loxP-kanMX4-loxP his3A pdc1A::loxP71/66 pdc5 $\Delta$ ::loxP71/66 fra2 $\Delta$ .

Addition of Native 2 Micron Plasmid

The loxP71-URA3-loxP66 marker was PCR-amplified using Phusion DNA polymerase (New England BioLabs; Ipswich, Mass.) from pLA59 (SEQ ID NO:97), and transformed along with the LA811×LA817 (SEQ ID NOs:200, 201) and LA812×LA818 (SEQ ID NOs:202, 203) 2-micron plasmid fragments (amplified from the native 2-micron plasmid from CEN.PK 113-7D; Centraalbureau voor Schimmelcultures (CBS) Fungal Biodiversity Centre) into strain PNY2037 on SE -URA plates at 30° C. The resulting strain PNY2037 2u::loxP71-URA3-loxP66 was transformed with pLA34 (pRS423::cre) (also called, pLA34) (SEQ ID NO:184) and selected on SE -HIS -URA plates at 30° C. Transformants were patched onto YP-1% galactose plates and allowed to grow for 48 hrs at 30° C. to induce Cre recombinase expression. Individual colonies were then patched onto SE -URA, SE -HIS, and YPE plates to confirm URA3 marker removal. The resulting identified strain, PNY2050, has the genotype: MATa ura3A::loxPkanMX4-loxP, his3A pdc1A::loxP71/66 pdc5A::loxP71/66 fra2 $\Delta$  2-micron.

Construction of PNY2115 from PNY2050

Construction of PNY2115 [MATa ura3A::loxP his3A pdc5A::loxP66/71 fra2A 2-micron plasmid (CEN.PK2) pdc1A::P[PDC1]-ALS|alsS\_Bs-CYC1t-loxP71/66 pdc6A:: (UAS)PGK1-P[FBA1]-KIVD|Lg(y)-TDH3t-loxP71/66 adh1A::P[ADH1]-ADH|Bi(y)-ADHt-loxP71/66 fra2∆::P

 $[ILV5]-ADH|Bi(y)-ADHt-loxP71/66 \qquad gpd2\Delta::loxP71/66] \label{eq:bound} from PNY2050 was as follows.$ 

Pdc1A::P[PDC1]-ALS|alsS\_Bs-CYC1t-loxP71/66

To integrate alsS into the pdc1A::loxP66/71 locus of PNY2050 using the endogenous PDC1 promoter, an inte- 5 gration cassette was PCR-amplified from pLA71 (SEQ ID NO:209), which contains the gene acetolactate synthase from the species Bacillus subtilis with a FBA1 promoter and a CYC1 terminator, and a URA3 marker flanked by degenerate loxP sites to allow homologous recombination in vivo 10 and subsequent removal of the URA3 marker. PCR was done by using KAPA HiFi and primers 895 (SEQ ID NO:212) and 679 (SEQ ID NO:213). The PDC1 portion of each primer was derived from 60 bp of the upstream of the coding sequence and 50 bp that are 53 bp upstream of the 15 stop codon. The PCR product was transformed into PNY2050 using standard genetic techniques and transformants were selected on synthetic complete media lacking uracil and supplemented with 1% ethanol at 30° C. Transformants were screened to verify correct integration by 20 colony PCR using primers 681 (SEQ ID NO:214), external to the 3' coding region and 92 (SEQ ID NO:215), internal to the URA3 gene. Positive transformants were then prepped for genomic DNA and screened by PCR using primers N245 (SEQ ID NO:216) and N246 (SEQ ID NO:217). The URA3 25 marker was recycled by transforming with pLA34 (SEQ ID NO:184) containing the CRE recombinase under the GAL1 promoter and plated on synthetic complete media lacking histidine and supplemented with 1% ethanol at 30° C. Transformants were plated on rich media supplemented with 30 1% ethanol and 0.5% galactose to induce the recombinase. Marker removal was confirmed by patching colonies to synthetic complete media lacking uracil and supplemented with 1% ethanol to verify absence of growth. The resulting identified strain, called PNY2090 has the genotype MATa 35 ura3 $\Delta$ ::loxP, his3 $\Delta$ , pdc1 $\Delta$ ::loxP71/66, pdc5 $\Delta$ ::loxP71/66 fra2A 2-micron pdc1A::P[PDC1]-ALS|alsS\_Bs-CYC1tloxP71/66.

Pdc6A::(UAS)PGK1-P[FBA1]-KIVD|Lg(y)-TDH3tloxP71/66

To delete the endogenous PDC6 coding region, an integration cassette was PCR-amplified from pLA78 (SEQ ID NO:210), which contains the kivD gene from the species Listeria grayi with a hybrid FBA1 promoter and a TDH3 terminator, and a URA3 marker flanked by degenerate loxP 45 sites to allow homologous recombination in vivo and subsequent removal of the URA3 marker. PCR was done by using KAPA HiFi and primers 896 (SEQ ID NO:218) and 897 (SEQ ID NO:219). The PDC6 portion of each primer was derived from 60 bp upstream of the coding sequence 50 and 59 bp downstream of the coding region. The PCR product was transformed into PNY2090 using standard genetic techniques and transformants were selected on synthetic complete media lacking uracil and supplemented with 1% ethanol at 30° C. Transformants were screened to verify 55 correct integration by colony PCR using primers 365 (SEQ ID NO:220) and 366 (SEQ ID NO:221), internal primers to the PDC6 gene. Transformants with an absence of product were then screened by colony PCR N638 (SEQ ID NO:222), external to the 5' end of the gene, and 740 (SEQ ID NO:223), 60 internal to the FBA1 promoter. Positive transformants were than the prepped for genomic DNA and screened by PCR with two external primers to the PDC6 coding sequence. Positive integrants would yield a 4720 bp product, while PDC6 wild type transformants would yield a 2130 bp 65 product. The URA3 marker was recycled by transforming with pLA34 containing the CRE recombinase under the

GAL1 promoter and plated on synthetic complete media lacking histidine and supplemented with 1% ethanol at 30° C. Transformants were plated on rich media supplemented with 1% ethanol and 0.5% galactose to induce the recombinase. Marker removal was confirmed by patching colonies to synthetic complete media lacking uracil and supplemented with 1% ethanol to verify absence of growth. The resulting identified strain is called PNY2093 and has the genotype MATa ura3A::loxP his3A pdc5A::loxP71/66 fra2A 2-micron pdc1A::P[PDC1]-ALSIalsS\_Bs-CYC1t-loxP71/66 pdc6A::(UAS)PGK1-P[FBA1]-KIVDILg(y)-TDH3tloxP71/66.

Adh1A::P[ADH1]-ADH|Bi(y)-ADHt-loxP71/66

To delete the endogenous ADH1 coding region and integrate BiADH using the endogenous ADH1 promoter, an integration cassette was PCR-amplified from pLA65 (SEQ ID NO:211), which contains the alcohol dehydrogenase from the species Beijerinckii incida with an ILV5 promoter and a ADH1 terminator, and a URA3 marker flanked by degenerate loxP sites to allow homologous recombination in vivo and subsequent removal of the URA3 marker. PCR was done by using KAPA HiFi and primers 856 (SEQ ID NO:224) and 857 (SEQ ID NO:225). The ADH1 portion of each primer was derived from the 5' region 50 bp upstream of the ADH1 start codon and the last 50 bp of the coding region. The PCR product was transformed into PNY2093 using standard genetic techniques and transformants were selected on synthetic complete media lacking uracil and supplemented with 1% ethanol at 30° C. Transformants were screened to verify correct integration by colony PCR using primers BK415 (SEQ ID NO:226), external to the 5' coding region and N1092 (SEQ ID NO:227), internal to the BiADH gene. Positive transformants were then screened by colony PCR using primers 413 (SEQ ID NO:160), external to the 3' coding region, and 92 (SEQ ID NO:215), internal to the URA3 marker. The URA3 marker was recycled by transforming with pLA34 (SEQ ID NO:184) containing the CRE recombinase under the GAL1 promoter and plated on synthetic complete media lacking histidine and supplemented with 1% ethanol at 30° C. Transformants were plated on rich media supplemented with 1% ethanol and 0.5% galactose to induce the recombinase. Marker removal was confirmed by patching colonies to synthetic complete media lacking uracil and supplemented with 1% ethanol to verify absence of growth. The resulting identified strain, called PNY2101 has the genotype MATa ura3Δ::loxP his3Δ pdc5Δ::loxP71/66 fra2A 2-micron pdc1A::P[PDC1]-ALS|alsS\_Bs-CYC1tpdc6A::(UAS)PGK1-P[FBA1]-KIVD|Lg(y)loxP71/66 TDH3t-loxP71/66 adh1A::P[ADH1]-ADH|Bi(y)-ADHtloxP71/66.

Fra2A::P[ILV5]-ADH|Bi(y)-ADHt-loxP71/66

To integrate BiADH into the flan locus of PNY2101, an integration cassette was PCR-amplified from pLA65 (SEQ ID NO:211), which contains the alcohol dehydrogenase from the species Beijerinckii indica with an ILV5 promoter and an ADH1 terminator, and a URA3 marker flanked by degenerate loxP sites to allow homologous recombination in vivo and subsequent removal of the URA3 marker. PCR was performed by using KAPA HiFi and primers 906 (SEQ ID NO:228) and 907 (SEQ ID NO:229). The FRA2 portion of each primer was derived from the first 60 bp of the coding sequence starting at the ATG and 56 bp downstream of the stop codon. The PCR product was transformed into PNY2101 using standard genetic techniques and transformants were selected on synthetic complete media lacking uracil and supplemented with 1% ethanol at 30° C. Transformants were screened to verify correct integration by

colony PCR using primers 667 (SEQ ID NO:230), external to the 5' coding region and 749 (SEQ ID NO:159), internal to the ILV5 promoter. The URA3 marker was recycled by transforming with pLA34 (SEQ ID NO:184) containing the CRE recombinase under the GAL1 promoter and plated on synthetic complete media lacking histidine and supplemented with 1% ethanol at 30° C. Transformants were plated on rich media supplemented with 1% ethanol and 0.5% galactose to induce the recombinase. Marker removal was confirmed by patching colonies to synthetic complete media lacking uracil and supplemented with 1% ethanol to verify absence of growth. The resulting identified strain, called PNY2110 has the genotype MATa  $ura3\Delta::loxP$  his3 $\Delta$ pdc5∆::loxP66/71 2-micron pdc1A::P[PDC1]-ALS|alsS\_Bs-CYC1t-loxP71/66 pdc6Δ:(UAS)PGK1-P[FBA1]-KIVD|Lg(y)-TDH3t-loxP71/66 adh1A::P[ADH1]-ADH|Bi  $(\mathbf{y})$ -ADHt-loxP71/66 fra2A::P[ILV5]-ADH|Bi(y)-ADHtloxP71/66.

GPD2 Deletion

To delete the endogenous GPD2 coding region, a deletion cassette was PCR amplified from pLA59 (SEQ ID NO:97), which contains a URA3 marker flanked by degenerate loxP sites to allow homologous recombination in vivo and sub-25 sequent removal of the URA3 marker. PCR was done by using KAPA HiFi and primers LA512 (SEQ ID NO:204) and LA513 (SEQ ID NO:205). The GPD2 portion of each primer was derived from the 5' region 50 bp upstream of the GPD2 start codon and 3' region 50 bp downstream of the stop codon such that integration of the URA3 cassette results in replacement of the entire GPD2 coding region. The PCR product was transformed into PNY2110 using standard genetic techniques and transformants were selected on synthetic complete medium lacking uracil and supplemented with 1% ethanol at 30° C. Transformants were screened to verify correct integration by colony PCR using primers LA516 (SEQ ID NO:206) external to the 5' coding region and LA135 (SEQ ID NO:94), internal to URA3. Positive 40 transformants were then screened by colony PCR using primers LA514 (SEQ ID NO:207) and LA515 (SEQ ID NO:208), internal to the GPD2 coding region. The URA3 marker was recycled by transforming with pLA34 (SEQ ID NO:184) containing the CRE recombinase under the GAL1 promoter and plated on synthetic complete medium lacking histidine and supplemented with 1% ethanol at 30° C. Transformants were plated on rich medium supplemented with 1% ethanol and 0.5% galactose to induce the recom- 50 binase. Marker removal was confirmed by patching colonies to synthetic complete medium lacking uracil and supplemented with 1% ethanol to verify absence of growth. The resulting identified strain, called PNY2115, has the genotype MATa ura3Δ::loxP his3Δ pdc5Δ::loxP66/71 fra2Δ 2-micron  $pdc1\Delta{::}P[PDC1]-ALS|alsS\_Bs-CYC1t-loxP71/66 \quad pdc6\Delta{::}$ (UAS)PGK1-P[FBA1]-KIVD|Lg(y)-TDH3t-loxP71/66 adh1A::P[ADH1]-ADH|Bi(y)-ADHt-loxP71/66 fra2∆::P [ILV5]-ADH|Bi(y)-ADHt-loxP71/66 gpd2A::loxP71/66. 60 Creation of PNY2145 from PNY2115

PNY2145 was constructed from PNY2115 by the additional integration of a phosphoketolase gene cassette at the pdc5 $\Delta$  locus and by replacing the native AMN1 gene with a codon optimized verison of the ortholog from CEN.PK. 65 Integration constructs are further described below. pdc5 $\Delta$ ::FBA(L8)-xpk1-CYC1t-loxP71/66

The TEF(M4)-xpk1-CYC1t gene from pRS423::TEF (M4)-xpk1+ENO1-eutD (SEQ ID NO:162) was PCR amplified using primers N1341 and N1338 (SEQ ID NOs:163 and 164), generating a 3.1 kb product. The loxP-flanked URA3 gene cassette from pLA59 (SEQ ID NO:97) was amplified with primers N1033c and N1342 (SEQ ID NOs:165 and 166), generating a 1.6 kb product. The xpk1 and URA3 PCR products were fused by combining them without primers for an additional 10 cycles of PCR using Phusion DNA polymerase. The resulting reaction mix was then used as a template for a PCR reaction with KAPA Hi Fi and primers N1342 and N1364 (SEQ ID NOs:166 and 167). A 4.2 kb PCR product was recovered by purification from an electrophoresis agarose gel (Zymo kit). FBA promoter variant L8 (SEQ ID NO:168) was amplified using primers N1366 and N1368 (SEQ ID NOs:169 and 170). The xpk1::URA3 PCR product was combined with the FBA promoter by additional rounds of PCR. The resulting product was phosphorylated with polynucleotide kinase and ligated into 20 pBR322 that had been digested with EcoRV and treated with calf intestinal phosphatase. The ligation reaction was transformed into E. coli cells (Stb13 competent cells from Invitrogen). The integration cassette was confirmed by sequencing. To prepare DNA for integration, the plasmid was used as a template in a PCR reaction with Kapa HiFi and primers N1371 and N1372 (SEQ ID NOs:171 and 172). The PCR product was isolated by phenol-chloroform extraction and ethanol precipitation (using standard methods; e.g. Maniatas, et al.). Five micrograms of DNA were used to transform strain PNY2115. Transformants were selected on medium lacking uracil (synthetic complete medium minus uracil with 1% ethanol as the carbon source). Colonies were screened for the integration event using PCR (JumpStart) with primers BK93 and N1114 (SEQ ID NOs:173 and 174). Two clones were selected to carry forward. The URA3 marker was recycled by transforming with pJT254 (SEQ ID NO:175) containing the CRE recombinase under the GAL1 promoter and plating on synthetic complete medium lacking histidine and supplemented with 1% ethanol at 30° C. Transformants were grown in rich medium supplemented with 1% ethanol to derepress the recombinase. Marker removal was confirmed for single colony isolates by patching to synthetic complete medium lacking uracil and supplemented with 1% ethanol to verify absence of growth. Loss of the recombinase plasmid, pJT254, was confirmed by patching the colonies to synthetic complete medium lacking histidine and supplemented with 1% ethanol. Proper marker removal was confirmed by PCR (primers N160SeqF5 (SEQ ID NO:176) and BK380. One resulting clone was designated PNY2293.

 $amn1\Delta::AMN1(y)-loxP71/66$ 

To replace the endogenous copy of AMN1 with a codonoptimized version of the AMN1 gene from CEN.PK2, an integration cassette containing the CEN.PK AMN1 promoter, AMN1(y) gene (nucleic acid SEQ ID NO:177; amino acid SEQ ID NO:178), and CEN.PK AMN1 terminator was assembled by SOE PCR and subcloned into the shuttle vector pLA59. The AMN1(y) gene was ordered from DNA 2.0 with codon-optimization for *S. cerevisiae*. The completed pLA67 plasmid (SEQ ID NO:179) contained: 1) pUC19 vector backbone sequence containing an *E. coli* replication origin and ampicillin resistance gene; 2) URA3 selection marker flanked by loxP71 and loxP66 sites; and 3)  $P_{AMN1(CEN.PK)}$ -AMN1(y)-term<sub>AMN1(CEN.PK)</sub> expression cassette

PCR amplification of the AMN1(y)-loxP71-URA3loxP66 cassette was performed by using KAPA HiFi from Kapa Biosystems, Woburn, Mass. and primers LA712 (SEQ ID NO:180) and LA746 (SEQ ID NO:181). The PCR product was transformed into PNY2293 using standard genetic techniques and transformants were selected on synthetic complete medium lacking uracil and supplemented 5 with 1% ethanol at 30° C. Transformants were observed under magnification for the absence of a clumping phenotype with respect to the control (PNY2293). The URA3 marker was recycled using the pJT254 Cre recombinase plasmid as described above. After marker recycle, clones 10 were again observed under magnification to confirm absence of the clumping phenotype. A resulting identified strain, PNY2145, has the genotype: MATa ura $3\Delta$ ::loxP his $3\Delta$ pdc5A::P[FBA(L8)]-XPK|xpk1\_Lp-CYCt-loxP66/71 fra2A 2-micron plasmid (CEN.PK2) pdc1A::P[PDC1]-ALS|al- 15 sS\_Bs-CYC1t-loxP71/66 pdc6Δ::(UAS)PGK1-P[FBA1]-KIVDILg(y)-TDH3t-loxP71/66 adh14::P[ADH1]-ADHIBi (y)-ADHt-loxP71/66 fra2A::P[ILV5]-ADH|Bi(y)-ADHtloxP71/66 gpd2Δ::loxP71/66 amn1Δ::AMN1(y).

SEQUENCE LISTING

# 100

# INCORPORATION BY REFERENCE

All documents cited herein, including journal articles or abstracts, published or corresponding U.S. or foreign patent applications, issued or foreign patents, or any other documents, are each entirely incorporated by reference herein, including all data, tables, figures, and text presented in the cited documents.

#### **EQUIVALENTS**

Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. Such equivalents are intended to be encompassed by the following claims.

<160	D> NU	JMBEH	R OF	SEQ	ID 1	10S :	230									
<210 <210	)> SH L> LH 2> TY	EQ II ENGTH ZPE :	) NO H: 50 PRT	1 53												
<213	3> OI	RGAN	ISM:	Saccharomyces cerevisiae												
<400	)> SI	EQUEI	NCE :	1												
Met 1	Ser	Glu	Ile	Thr 5	Leu	Gly	ГЛа	Tyr	Leu 10	Phe	Glu	Arg	Leu	Ser 15	Gln	
Val	Asn	Сүз	Asn 20	Thr	Val	Phe	Gly	Leu 25	Pro	Gly	Asp	Phe	Asn 30	Leu	Ser	
Leu	Leu	Asp 35	Lys	Leu	Tyr	Glu	Val 40	ГЛа	Gly	Met	Arg	Trp 45	Ala	Gly	Asn	
Ala	Asn 50	Glu	Leu	Asn	Ala	Ala 55	Tyr	Ala	Ala	Asp	Gly 60	Tyr	Ala	Arg	Ile	
Lys 65	Gly	Met	Ser	Суз	Ile 70	Ile	Thr	Thr	Phe	Gly 75	Val	Gly	Glu	Leu	Ser 80	
Ala	Leu	Asn	Gly	Ile 85	Ala	Gly	Ser	Tyr	Ala 90	Glu	His	Val	Gly	Val 95	Leu	
His	Val	Val	Gly 100	Val	Pro	Ser	Ile	Ser 105	Ser	Gln	Ala	Lys	Gln 110	Leu	Leu	
Leu	His	His 115	Thr	Leu	Gly	Asn	Gly 120	Asp	Phe	Thr	Val	Phe 125	His	Arg	Met	
Ser	Ala 130	Asn	Ile	Ser	Glu	Thr 135	Thr	Ala	Met	Ile	Thr 140	Asp	Ile	Ala	Asn	
Ala 145	Pro	Ala	Glu	Ile	Asp 150	Arg	Суз	Ile	Arg	Thr 155	Thr	Tyr	Thr	Thr	Gln 160	
Arg	Pro	Val	Tyr	Leu 165	Gly	Leu	Pro	Ala	Asn 170	Leu	Val	Asp	Leu	Asn 175	Val	
Pro	Ala	Lys	Leu 180	Leu	Glu	Thr	Pro	Ile 185	Asp	Leu	Ser	Leu	Lys 190	Pro	Asn	
Aap	Ala	Glu 195	Ala	Glu	Ala	Glu	Val 200	Val	Arg	Thr	Val	Val 205	Glu	Leu	Ile	
Lya	Asp 210	Ala	Lys	Asn	Pro	Val 215	Ile	Leu	Ala	Asp	Ala 220	Сүз	Ala	Ser	Arg	
His 225	Asp	Val	Lys	Ala	Glu 230	Thr	Lys	Lys	Leu	Met 235	Asp	Leu	Thr	Gln	Phe 240	
Pro	Val	Tyr	Val	Thr 245	Pro	Met	Gly	Lys	Gly 250	Ala	Ile	Asp	Glu	Gln 255	His	
--------------	----------------	--------------	---------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	
Pro	Arg	Tyr	Gly 260	Gly	Val	Tyr	Val	Gly 265	Thr	Leu	Ser	Arg	Pro 270	Glu	Val	
Lys	Lys	Ala 275	Val	Glu	Ser	Ala	Asp 280	Leu	Ile	Leu	Ser	Ile 285	Gly	Ala	Leu	
Leu	Ser 290	Asp	Phe	Asn	Thr	Gly 295	Ser	Phe	Ser	Tyr	Ser 300	Tyr	Lys	Thr	ГЛа	
Asn 305	Ile	Val	Glu	Phe	His 310	Ser	Asp	His	Ile	Lys 315	Ile	Arg	Asn	Ala	Thr 320	
Phe	Pro	Gly	Val	Gln 325	Met	Lys	Phe	Ala	Leu 330	Gln	ГÀа	Leu	Leu	Asp 335	Ala	
Ile	Pro	Glu	Val 340	Val	Lys	Asp	Tyr	Lys 345	Pro	Val	Ala	Val	Pro 350	Ala	Arg	
Val	Pro	Ile 355	Thr	Lys	Ser	Thr	Pro 360	Ala	Asn	Thr	Pro	Met 365	Lys	Gln	Glu	
Trp	Met 370	Trp	Asn	His	Leu	Gly 375	Asn	Phe	Leu	Arg	Glu 380	Gly	Asp	Ile	Val	
Ile 385	Ala	Glu	Thr	Gly	Thr 390	Ser	Ala	Phe	Gly	Ile 395	Asn	Gln	Thr	Thr	Phe 400	
Pro	Thr	Asp	Val	Tyr 405	Ala	Ile	Val	Gln	Val 410	Leu	Trp	Gly	Ser	Ile 415	Gly	
Phe	Thr	Val	Gly 420	Ala	Leu	Leu	Gly	Ala 425	Thr	Met	Ala	Ala	Glu 430	Glu	Leu	
Asp	Pro	Lys 435	Lys	Arg	Val	Ile	Leu 440	Phe	Ile	Gly	Asp	Gly 445	Ser	Leu	Gln	
Leu	Thr 450	Val	Gln	Glu	Ile	Ser 455	Thr	Met	Ile	Arg	Trp 460	Gly	Leu	Lys	Pro	
Tyr 465	Ile	Phe	Val	Leu	Asn 470	Asn	Asn	Gly	Tyr	Thr 475	Ile	Glu	Lys	Leu	Ile 480	
His	Gly	Pro	His	Ala 485	Glu	Tyr	Asn	Glu	Ile 490	Gln	Gly	Trp	Asp	His 495	Leu	
Ala	Leu	Leu	Pro 500	Thr	Phe	Gly	Ala	Arg 505	Asn	Tyr	Glu	Thr	His 510	Arg	Val	
Ala	Thr	Thr 515	Gly	Glu	Trp	Glu	Lys 520	Leu	Thr	Gln	Asp	Lys 525	Asp	Phe	Gln	
Asp	Asn 530	Ser	Lys	Ile	Arg	Met 535	Ile	Glu	Val	Met	Leu 540	Pro	Val	Phe	Asp	
Ala 545	Pro	Gln	Asn	Leu	Val 550	Lya	Gln	Ala	Gln	Leu 555	Thr	Ala	Ala	Thr	Asn 560	
Ala	Lys	Gln														
<210 <211	)> SH L> LH	Q II NGTH	) NO 1: 56	2 53												
<213	3> 0F	GAN]	ISM:	Saco	charo	omyce	es ce	erevi	lsiae	9						
<400	)> SE	QUEN	ICE :	2												
Met 1	Ser	Glu	Ile	Thr 5	Leu	Gly	Lys	Tyr	Leu 10	Phe	Glu	Arg	Leu	Ser 15	Gln	
Val	Asn	Суз	Asn 20	Thr	Val	Phe	Gly	Leu 25	Pro	Gly	Asp	Phe	Asn 30	Leu	Ser	
Leu	Leu	Asp 35	Lys	Leu	Tyr	Glu	Val 40	Lys	Gly	Met	Arg	Trp 45	Ala	Gly	Asn	

Ala Asn Glu Leu Asn Ala Ala Tyr Ala Ala Asp Gly Tyr Ala Arg Ile Lys Gly Met Ser Cys Ile Ile Thr Thr Phe Gly Val Gly Glu Leu Ser Ala Leu Asn Gly Ile Ala Gly Ser Tyr Ala Glu His Val Gly Val Leu His Val Val Gly Val Pro Ser Ile Ser Ser Gln Ala Lys Gln Leu Leu Leu His His Thr Leu Gly Asn Gly Asp Phe Thr Val Phe His Arg Met Ser Ala Asn Ile Ser Glu Thr Thr Ala Met Ile Thr Asp Ile Ala Asn Ala Pro Ala Glu Ile Asp Arg Cys Ile Arg Thr Thr Tyr Thr Gln Arg Pro Val Tyr Leu Gly Leu Pro Ala Asn Leu Val Asp Leu Asn Val Pro Ala Lys Leu Leu Glu Thr Pro Ile Asp Leu Ser Leu Lys Pro Asn Asp Ala Glu Ala Glu Ala Glu Val Val Arg Thr Val Val Glu Leu Ile Lys Asp Ala Lys Asn Pro Val Ile Leu Ala Asp Ala Cys Ala Ser Arg His Asp Val Lys Ala Glu Thr Lys Lys Leu Met Asp Leu Thr Gln Phe Pro Val Tyr Val Thr Pro Met Gly Lys Gly Ala Ile Asp Glu Gln His Pro Arg Tyr Gly Gly Val Tyr Val Gly Thr Leu Ser Arg Pro Glu Val Lys Lys Ala Val Glu Ser Ala Asp Leu Ile Leu Ser Ile Gly Ala Leu Leu Ser Asp Phe Asn Thr Gly Ser Phe Ser Tyr Ser Tyr Lys Thr Lys Asn Ile Val Glu Phe His Ser Asp His Ile Lys Ile Arg Asn Ala Thr Phe Pro Gly Val Gln Met Lys Phe Ala Leu Gln Lys Leu Leu Asp Ala Ile Pro Glu Val Val Lys Asp Tyr Lys Pro Val Ala Val Pro Ala Arg Val Pro Ile Thr Lys Ser Thr Pro Ala Asn Thr Pro Met Lys Gln Glu Trp Met Trp Asn His Leu Gly Asn Phe Leu Arg Glu Gly Asp Ile Val Ile Ala Glu Thr Gly Thr Ser Ala Phe Gly Ile Asn Gln Thr Thr Phe Pro Thr Asp Val Tyr Ala Ile Val Gln Val Leu Trp Gly Ser Ile Gly Phe Thr Val Gly Ala Leu Leu Gly Ala Thr Met Ala Ala Glu Glu Leu Asp Pro Lys Lys Arg Val Ile Leu Phe Ile Gly Asp Gly Ser Leu Gln Leu Thr Val Gln Glu Ile Ser Thr Met Ile Arg Trp Gly Leu Lys Pro 

#### -continued

			_												
Tyr 465	Ile	Phe	Val	Leu	Asn 470	Asn	Asn	Gly	Tyr	Thr 475	Ile	Glu	Lys	Leu	Ile 480
His	Gly	Pro	His	Ala 485	Glu	Tyr	Asn	Glu	Ile 490	Gln	Gly	Trp	Asp	His 495	Leu
Ala	Leu	Leu	Pro 500	Thr	Phe	Gly	Ala	Arg 505	Asn	Tyr	Glu	Thr	His 510	Arg	Val
Ala	Thr	Thr 515	Gly	Glu	Trp	Glu	Lys 520	Leu	Thr	Gln	Asp	Lys 525	Aab	Phe	Gln
Asp	Asn 530	Ser	Lys	Ile	Arg	Met 535	Ile	Glu	Val	Met	Leu 540	Pro	Val	Phe	Asp
Ala 545	Pro	Gln	Asn	Leu	Val 550	ГЛЗ	Gln	Ala	Gln	Leu 555	Thr	Ala	Ala	Thr	Asn 560
Ala	Lys	Gln													
<210 <211 <212 <212	0> SI L> LI 2> T 3> OI	EQ II ENGTH (PE : RGANI	) NO H: 53 PRT [SM:	3 33 Saco	char	omyc	es C(	erev:	isia	9					
<400	)> SI	EQUEI	ICE :	3											
Met 1	Ser	Glu	Ile	Thr 5	Leu	Gly	Lys	Tyr	Leu 10	Phe	Glu	Arg	Leu	Lys 15	Gln
Val	Asn	Val	Asn 20	Thr	Ile	Phe	Gly	Leu 25	Pro	Gly	Asp	Phe	Asn 30	Leu	Ser
Leu	Leu	Asp 35	Lys	Ile	Tyr	Glu	Val 40	Asp	Gly	Leu	Arg	Trp 45	Ala	Gly	Asn
Ala	Asn 50	Glu	Leu	Asn	Ala	Ala 55	Tyr	Ala	Ala	Asp	Gly 60	Tyr	Ala	Arg	Ile
Lys 65	Gly	Leu	Ser	Val	Leu 70	Val	Thr	Thr	Phe	Gly 75	Val	Gly	Glu	Leu	Ser 80
Ala	Leu	Asn	Gly	Ile 85	Ala	Gly	Ser	Tyr	Ala 90	Glu	His	Val	Gly	Val 95	Leu
His	Val	Val	Gly 100	Val	Pro	Ser	Ile	Ser 105	Ala	Gln	Ala	Lys	Gln 110	Leu	Leu
Leu	His	His 115	Thr	Leu	Gly	Asn	Gly 120	Asp	Phe	Thr	Val	Phe 125	His	Arg	Met
Ser	Ala 130	Asn	Ile	Ser	Glu	Thr 135	Thr	Ser	Met	Ile	Thr 140	Asp	Ile	Ala	Thr
Ala 145	Pro	Ser	Glu	Ile	Asp 150	Arg	Leu	Ile	Arg	Thr 155	Thr	Phe	Ile	Thr	Gln 160
Arg	Pro	Ser	Tyr	Leu 165	Gly	Leu	Pro	Ala	Asn 170	Leu	Val	Asp	Leu	Lys 175	Val
Pro	Gly	Ser	Leu 180	Leu	Glu	ГЛЗ	Pro	Ile 185	Asp	Leu	Ser	Leu	Lys 190	Pro	Asn
Asp	Pro	Glu 195	Ala	Glu	Lys	Glu	Val 200	Ile	Asp	Thr	Val	Leu 205	Glu	Leu	Ile
Gln	Asn 210	Ser	Lys	Asn	Pro	Val 215	Ile	Leu	Ser	Asp	Ala 220	СЛа	Ala	Ser	Arg
His 225	Asn	Val	Lys	ГЛа	Glu 230	Thr	Gln	Гла	Leu	Ile 235	Asp	Leu	Thr	Gln	Phe 240
Pro	Ala	Phe	Val	Thr 245	Pro	Leu	Gly	Lys	Gly 250	Ser	Ile	Asp	Glu	Gln 255	His
				-					-					-	

Lys	Gln	Ala 275	Val	Glu	Ser	Ala	Asp 280	Leu	Ile	Leu	Ser	Val 285	Gly	Ala	Leu			
Leu	Ser 290	Asp	Phe	Asn	Thr	Gly 295	Ser	Phe	Ser	Tyr	Ser 300	Tyr	Lys	Thr	Lys			
Asn 305	Val	Val	Glu	Phe	His 310	Ser	Asp	Tyr	Val	Lys 315	Val	Lys	Asn	Ala	Thr 320			
Phe	Leu	Gly	Val	Gln 325	Met	Lys	Phe	Ala	Leu 330	Gln	Asn	Leu	Leu	Lys 335	Val			
Ile	Pro	Asp	Val 340	Val	Lys	Gly	Tyr	Lys 345	Ser	Val	Pro	Val	Pro 350	Thr	Lys			
Thr	Pro	Ala 355	Asn	Lys	Gly	Val	Pro 360	Ala	Ser	Thr	Pro	Leu 365	Lys	Gln	Glu			
Trp	Leu 370	Trp	Asn	Glu	Leu	Ser 375	Гла	Phe	Leu	Gln	Glu 380	Gly	Asp	Val	Ile			
Ile 385	Ser	Glu	Thr	Gly	Thr 390	Ser	Ala	Phe	Gly	Ile 395	Asn	Gln	Thr	Ile	Phe 400			
Pro	Lys	Aap	Ala	Tyr 405	Gly	Ile	Ser	Gln	Val 410	Leu	Trp	Gly	Ser	Ile 415	Gly			
Phe	Thr	Thr	Gly 420	Ala	Thr	Leu	Gly	Ala 425	Ala	Phe	Ala	Ala	Glu 430	Glu	Ile			
Asp	Pro	Asn 435	Lys	Arg	Val	Ile	Leu 440	Phe	Ile	Gly	Asp	Gly 445	Ser	Leu	Gln			
Leu	Thr 450	Val	Gln	Glu	Ile	Ser 455	Thr	Met	Ile	Arg	Trp 460	Gly	Leu	Lys	Pro			
Tyr 465	Leu	Phe	Val	Leu	Asn 470	Asn	Asp	Gly	Tyr	Thr 475	Ile	Glu	Lys	Leu	Ile 480			
His	Gly	Pro	His	Ala 485	Glu	Tyr	Asn	Glu	Ile 490	Gln	Thr	Trp	Asp	His 495	Leu			
Ala	Leu	Leu	Pro 500	Ala	Phe	Gly	Ala	Lys 505	Lys	Tyr	Glu	Asn	His 510	Lys	Ile			
Ala	Thr	Thr 515	Gly	Glu	Trp	Asp	Ala 520	Leu	Thr	Thr	Asp	Ser 525	Glu	Phe	Gln			
Lys	Asn 530	Ser	Val	Ile														
- 211	1. CI	го ті	ס או כ	4														
<21	1> 51 1> 11	ENGTI	H: 8	4 04														
<21	2> T	YPE :	DNA															
<21	3 > 01	(GAN.	ISM:	Sac	char	omyc	es ce	erev:	ISIA	9								
<40	)> SI	EQUEI	NCE :	4														
atg	cece	aag g	gtag	aaaa	gc t	gcag	aaaga	a ttę	ggeta	aaga	aga	ctgt	eet d	catta	acaggt	60		
gca	cctgo	etg g	gtati	tggt	aa g	gcga	ccgca	a tta	agagi	act	tgga	aggca	atc d	caato	ggtgat	120		
atg	aaact	iga 1	tett	ggct	gc ta	agaa	gatta	a gaa	aaag	ctcg	agga	aatto	gaa q	gaaga	accatt	180		
gat	caaga	agt 1	ttcc	aaac	gc a	aaag	ttcai	t gto	ggcc	cagc	tgga	atato	cac t	caa	gcagaa	240		
aaa	atcaa	agc (	cctt	catt	ga a	aact	tgeea	a caa	agagi	tca	agga	atati	tga d	catto	ctggtg	300		
aaca	aatgo	ccg g	gaaa	ggct	ct t	ggca	gtga	c cgt	gtg	ggcc	agai	tegea	aac ç	ggage	gatatc	360		
cag	gacgt	gt 1	ttga	cacc	aa c	gtca	cggct	t tta	aatca	aata	tca	cacaa	agc t	gta	etgeee	420		
ata	tcca	aag (	ccaa	gaati	tc ag	ggag.	atati	t gta	aaati	tgg	gtto	caat	cgc t	ggca	agagac	480		
gcat	cacco	caa 🤇	cagg	tteta	at ci	tatt	gtgc	c tơ	caagi	ttg	ccgi	tggg	ggc (	gttca	actgat	540		
agti	tga	gaa a	agga	gete	at ca	aaca	ctaaa	a ati	zaga	gtca	ttc	taati	tgc a	acca	gggcta	600		

# 

-continued	
------------	--

gtcgagactg aattttcact agttagatac agaggtaacg aggaacaagc ca	aagaatgtt 660
tacaaggata ctaccccatt gatggctgat gacgtggctg atctgatcgt ct	tatgcaact 720
tccagaaaac aaaatactgt aattgcagac actttaatct ttccaacaaa co	caagegtea 780
cctcatcata tcttccgtgg ataa	804
<210> SEQ ID NO 5 <211> LENGTH: 267 <212> TYPE: PRT <213> ORGANISM: Saccharomyces cerevisiae	
<400> SEQUENCE: 5	
Met Ser Gln Gly Arg Lys Ala Ala Glu Arg Leu Ala Lys Lys T 1 5 10 1	Thr Val 15
Leu Ile Thr Gly Ala Ser Ala Gly Ile Gly Lys Ala Thr Ala L 20 25 30	Leu Glu
Tyr Leu Glu Ala Ser Asn Gly Asp Met Lys Leu Ile Leu Ala A 35 40 45	Ala Arg
Arg Leu Glu Lys Leu Glu Glu Leu Lys Lys Thr Ile Asp Gln G505560	Glu Phe
Pro Asn Ala Lys Val His Val Ala Gln Leu Asp Ile Thr Gln A 65 70 75	Ala Glu 80
Lys Ile Lys Pro Phe Ile Glu Asn Leu Pro Gln Glu Phe Lys A 85 90 9	Asp Ile 95
Asp Ile Leu Val Asn Asn Ala Gly Lys Ala Leu Gly Ser Asp A 100 105 110	Arg Val
Gly Gln Ile Ala Thr Glu Asp Ile Gln Asp Val Phe Asp Thr A 115 120 125	Asn Val
Thr Ala Leu Ile Asn Ile Thr Gln Ala Val Leu Pro Ile Phe G 130 135 140	Gln Ala
Lys Asn Ser Gly Asp Ile Val Asn Leu Gly Ser Ile Ala Gly A 145 150 155	Arg Asp 160
Ala Tyr Pro Thr Gly Ser Ile Tyr Cys Ala Ser Lys Phe Ala V 165 170 1	Val Gly 175
Ala Phe Thr Asp Ser Leu Arg Lys Glu Leu Ile Asn Thr Lys I 180 185 190	Ile Arg
Val Ile Leu Ile Ala Pro Gly Leu Val Glu Thr Glu Phe Ser L 195 200 205	Leu Val
Arg Tyr Arg Gly Asn Glu Glu Gln Ala Lys Asn Val Tyr Lys A 210 215 220	Asp Thr
Thr Pro Leu Met Ala Asp Asp Val Ala Asp Leu Ile Val Tyr A 225 230 235	Ala Thr 240
Ser Arg Lys Gln Asn Thr Val Ile Ala Asp Thr Leu Ile Phe P 245 250 2	Pro Thr 255
Asn Gln Ala Ser Pro His His Ile Phe Arg Gly 260 265	
<210> SEQ ID NO 6 <211> LENGTH: 386 <212> TYPE: PRT <213> ORGANISM: Saccharomyces cerevisiae	
<400> SEQUENCE: 6	
Met Ser Ala Ala Thr Val Gly Lys Pro Ile Lys Cys Ile Ala A 1 5 10 1	Ala Val 15
Ala Tyr Asp Ala Lys Lys Pro Leu Ser Val Glu Glu Ile Thr V	Val Asp

111

			20					25					30		
Ala	Pro	Lys 35	Ala	His	Glu	Val	Arg 40	Ile	Гла	Ile	Glu	Tyr 45	Thr	Ala	Val
Сүз	His 50	Thr	Asp	Ala	Tyr	Thr 55	Leu	Ser	Gly	Ser	Asp 60	Pro	Glu	Gly	Leu
Phe 65	Pro	Сув	Val	Leu	Gly 70	His	Glu	Gly	Ala	Gly 75	Ile	Val	Glu	Ser	Val 80
Gly	Asp	Asp	Val	Ile 85	Thr	Val	Lys	Pro	Gly 90	Asp	His	Val	Ile	Ala 95	Leu
Tyr	Thr	Ala	Glu 100	Суз	Gly	ГЛа	Суз	Lys 105	Phe	Суз	Thr	Ser	Gly 110	Lys	Thr
Asn	Leu	Cys 115	Gly	Ala	Val	Arg	Ala 120	Thr	Gln	Gly	Гла	Gly 125	Val	Met	Pro
Asp	Gly 130	Thr	Thr	Arg	Phe	His 135	Asn	Ala	Гла	Gly	Glu 140	Asp	Ile	Tyr	His
Phe 145	Met	Gly	Суз	Ser	Thr 150	Phe	Ser	Glu	Tyr	Thr 155	Val	Val	Ala	Asp	Val 160
Ser	Val	Val	Ala	Ile 165	Asp	Pro	Lys	Ala	Pro 170	Leu	Asp	Ala	Ala	Cys 175	Leu
Leu	Gly	Cys	Gly 180	Val	Thr	Thr	Gly	Phe 185	Gly	Ala	Ala	Leu	Lys 190	Thr	Ala
Asn	Val	Gln 195	Lys	Gly	Asp	Thr	Val 200	Ala	Val	Phe	Gly	Cys 205	Gly	Thr	Val
Gly	Leu 210	Ser	Val	Ile	Gln	Gly 215	Ala	Lys	Leu	Arg	Gly 220	Ala	Ser	Lys	Ile
Ile 225	Ala	Ile	Asp	Ile	Asn 230	Asn	Lys	Lys	Lys	Gln 235	Tyr	Суз	Ser	Gln	Phe 240
Gly	Ala	Thr	Asp	Phe 245	Val	Asn	Pro	Lys	Glu 250	Asp	Leu	Ala	Lys	Asp 255	Gln
Thr	Ile	Val	Glu 260	Гла	Leu	Ile	Glu	Met 265	Thr	Asp	Gly	Gly	Leu 270	Asp	Phe
Thr	Phe	Asp 275	Суз	Thr	Gly	Asn	Thr 280	Lys	Ile	Met	Arg	Asp 285	Ala	Leu	Glu
Ala	Cys 290	His	Lys	Gly	Trp	Gly 295	Gln	Ser	Ile	Ile	Ile 300	Gly	Val	Ala	Ala
Ala 305	Gly	Glu	Glu	Ile	Ser 310	Thr	Arg	Pro	Phe	Gln 315	Leu	Val	Thr	Gly	Arg 320
Val	Trp	Lys	Gly	Ser 325	Ala	Phe	Gly	Gly	Ile 330	Lys	Gly	Arg	Ser	Glu 335	Met
Gly	Gly	Leu	Ile 340	Lys	Asp	Tyr	Gln	Lys 345	Gly	Ala	Leu	Lys	Val 350	Glu	Glu
Phe	Ile	Thr 355	His	Arg	Arg	Pro	Phe 360	Lys	Glu	Ile	Asn	Gln 365	Ala	Phe	Glu
Asp	Leu 370	His	Asn	Gly	Asp	Сув 375	Leu	Arg	Thr	Val	Leu 380	Lys	Ser	Asp	Glu
Ile 385	Lys														
<210 <211 <212 <213	)> SH L> LH 2> TY 3> OH	EQ II ENGTH (PE : RGANI	D NO H: 39 PRT ISM:	7 99 Psei	udomo	onas	put:	Lda							

<400> SEQUENCE: 7

#### -continued

_															
Met 1	Ser	Gly	Asn	Arg 5	Gly	Val	Val	Tyr	Leu 10	Gly	Ser	Gly	Lys	Val 15	Glu
Val	Gln	Lys	Ile 20	Asp	Tyr	Pro	Гла	Met 25	Gln	Asp	Pro	Arg	Gly 30	Lys	ГЛа
Ile	Glu	His 35	Gly	Val	Ile	Leu	Lys 40	Val	Val	Ser	Thr	Asn 45	Ile	Сүз	Gly
Ser	Asp 50	Gln	His	Met	Val	Arg 55	Gly	Arg	Thr	Thr	Ala 60	Gln	Val	Gly	Leu
Val 65	Leu	Gly	His	Glu	Ile 70	Thr	Gly	Glu	Val	Ile 75	Glu	Lys	Gly	Arg	Asp 80
Val	Glu	Asn	Leu	Gln 85	Ile	Gly	Asp	Leu	Val 90	Ser	Val	Pro	Phe	Asn 95	Val
Ala	Суз	Gly	Arg 100	Суз	Arg	Ser	Cys	Lys 105	Glu	Met	His	Thr	Gly 110	Val	Сув
Leu	Thr	Val 115	Asn	Pro	Ala	Arg	Ala 120	Gly	Gly	Ala	Tyr	Gly 125	Tyr	Val	Asp
Met	Gly 130	Asp	Trp	Thr	Gly	Gly 135	Gln	Ala	Glu	Tyr	Leu 140	Leu	Val	Pro	Tyr
Ala 145	Asp	Phe	Asn	Leu	Leu 150	Гла	Leu	Pro	Asp	Arg 155	Asp	Lys	Ala	Met	Glu 160
Lys	Ile	Arg	Asp	Leu 165	Thr	Суз	Leu	Ser	Asp 170	Ile	Leu	Pro	Thr	Gly 175	Tyr
His	Gly	Ala	Val 180	Thr	Ala	Gly	Val	Gly 185	Pro	Gly	Ser	Thr	Val 190	Tyr	Val
Ala	Gly	Ala 195	Gly	Pro	Val	Gly	Leu 200	Ala	Ala	Ala	Ala	Ser 205	Ala	Arg	Leu
Leu	Gly 210	Ala	Ala	Val	Val	Ile 215	Val	Gly	Asp	Leu	Asn 220	Pro	Ala	Arg	Leu
Ala 225	His	Ala	Lys	Ala	Gln 230	Gly	Phe	Glu	Ile	Ala 235	Asp	Leu	Ser	Leu	Asp 240
Thr	Pro	Leu	His	Glu 245	Gln	Ile	Ala	Ala	Leu 250	Leu	Gly	Glu	Pro	Glu 255	Val
Asp	Суз	Ala	Val 260	Asp	Ala	Val	Gly	Phe 265	Glu	Ala	Arg	Gly	His 270	Gly	His
Glu	Gly	Ala 275	Lys	His	Glu	Ala	Pro 280	Ala	Thr	Val	Leu	Asn 285	Ser	Leu	Met
Gln	Val 290	Thr	Arg	Val	Ala	Gly 295	Lys	Ile	Gly	Ile	Pro 300	Gly	Leu	Tyr	Val
Thr 305	Glu	Asp	Pro	Gly	Ala 310	Val	Asp	Ala	Ala	Ala 315	Lys	Ile	Gly	Ser	Leu 320
Ser	Ile	Arg	Phe	Gly 325	Leu	Gly	Trp	Ala	Lуз 330	Ser	His	Ser	Phe	His 335	Thr
Gly	Gln	Thr	Pro 340	Val	Met	Lys	Tyr	Asn 345	Arg	Ala	Leu	Met	Gln 350	Ala	Ile
Met	Trp	Asp 355	Arg	Ile	Asn	Ile	Ala 360	Glu	Val	Val	Gly	Val 365	Gln	Val	Ile
Ser	Leu 370	Asp	Asp	Ala	Pro	Arg 375	Gly	Tyr	Gly	Glu	Phe 380	Asp	Ala	Gly	Val
Pro 385	Гла	Lys	Phe	Val	Ile 390	Asp	Pro	His	Lys	Thr 395	Phe	Ser	Ala	Ala	
<210 <211 <211	0> SI 1> LI 2> TY	EQ II ENGTH YPE :	) NO 1: 20 DNA	8 064											

-continued

<213> ORGAN	NISM: Saccha	aromyces cei	revisiae				
<400> SEQUI	ENCE: 8						
atgatcagac	aatctacgct	aaaaaacttc	gctattaagc	gttgctttca	acatatagca	60	
taccgcaaca	cacctgccat	gagatcagta	gctctcgcgc	agcgctttta	tagttcgtct	120	
tcccgttatt	acagtgcgtc	tccattacca	gcctctaaaa	ggccagagcc	tgctccaagt	180	
ttcaatgttg	atccattaga	acageceget	gaaccttcaa	aattggctaa	gaaactacgc	240	
gctgagcctg	acatggatac	ctctttcgtc	ggtttaactg	gtggtcaaat	atttaacgaa	300	
atgatgtcca	gacaaaacgt	tgatactgta	tttggttatc	caggtggtgc	tatcctacct	360	
gtttacgatg	ccattcataa	cagtgataaa	ttcaacttcg	ttcttccaaa	acacgaacaa	420	
ggtgccggtc	acatggcaga	aggetaegee	agagettetg	gtaaaccagg	tgttgtcttg	480	
gttacttctg	ggccaggtgc	caccaatgtc	gttactccaa	tggcagatgc	ctttgcagac	540	
gggattccaa	tggttgtctt	tacagggcaa	gtctcaacta	gtgctatcgg	tactgatgct	600	
ttccaagagg	ctgacgtcgt	tggtatttct	agatettgta	cgaaatggaa	tgtcatggtc	660	
aagtccgtgg	aagaattgcc	attgcgtatt	aacgaggctt	ttgaaattgc	cacgagcggt	720	
agaccgggac	cagtcttggt	cgatttacca	aaggatgtta	cagcagctat	cttaagaaat	780	
ccaattccaa	caaaaacaac	tcttccatca	aacgcactaa	accaattaac	cagtcgcgca	840	
caagatgaat	ttgtcatgca	aagtatcaat	aaagcagcag	atttgatcaa	cttggcaaag	900	
aaacctgtct	tatacgtcgg	tgctggtatt	ttaaaccatg	cagatggtcc	aagattacta	960	
aaagaattaa	gtgaccgtgc	tcaaatacct	gtcaccacta	ctttacaagg	tttaggttca	1020	
ttcgaccaag	aagatccaaa	atcattggat	atgcttggta	tgcacggttg	tgctactgcc	1080	
aacctggcag	tgcaaaatgc	cgacttgata	attgcagttg	gtgctagatt	cgacgaccgt	1140	
gtcactggta	atatttctaa	attcgctcca	gaagctcgtc	gtgcagctgc	cgagggtaga	1200	
ggtggtatta	ttcatttcga	ggttagtcca	aaaaacataa	acaaggttgt	tcaaactcaa	1260	
atagcagtgg	aaggtgatgc	tacgaccaat	ctgggcaaaa	tgatgtcaaa	gattttccca	1320	
gttaaggaga	ggtctgaatg	gtttgctcaa	ataaataaat	ggaagaagga	atacccatac	1380	
gcttatatgg	aggagactcc	aggatctaaa	attaaaccac	agacggttat	aaagaaacta	1440	
tccaaggttg	ccaacgacac	aggaagacat	gtcattgtta	caacgggtgt	ggggcaacat	1500	
caaatgtggg	ctgctcaaca	ctggacatgg	agaaatccac	atactttcat	cacatcaggt	1560	
ggtttaggta	cgatgggtta	cggtctccct	gccgccatcg	gtgctcaagt	tgcaaagcca	1620	
gaatctttgg	ttattgacat	tgatggtgac	gcatccttta	acatgactct	aacggaattg	1680	
agttctgccg	ttcaagctgg	tactccagtg	aagattttga	ttttgaacaa	tgaagagcaa	1740	
ggtatggtta	ctcaatggca	atccctgttc	tacgaacatc	gttattccca	cacacatcaa	1800	
ttgaaccctg	atttcataaa	actagcggag	gctatgggtt	taaaaggttt	aagagtcaag	1860	
aagcaagagg	aattggacgc	taagttgaaa	gaattcgttt	ctaccaaggg	cccagttttg	1920	
cttgaagtgg	aagttgataa	aaaagttcct	gttttgccaa	tggtggcagg	tggtagcggt	1980	
ctagacgagt	tcataaattt	tgacccagaa	gttgaaagac	aacagactga	attacgtcat	2040	
aagcgtacag	gcggtaagca	ctga				2064	

<210> SEQ ID NO 9 <211> LENGTH: 687 <212> TYPE: PRT <213> ORGANISM: Saccharomyces cerevisiae

<400	)> SE	EQUEN	ICE :	9											
Met 1	Ile	Arg	Gln	Ser 5	Thr	Leu	Lys	Asn	Phe 10	Ala	Ile	LYa	Arg	Cys 15	Phe
Gln	His	Ile	Ala 20	Tyr	Arg	Asn	Thr	Pro 25	Ala	Met	Arg	Ser	Val 30	Ala	Leu
Ala	Gln	Arg 35	Phe	Tyr	Ser	Ser	Ser 40	Ser	Arg	Tyr	Tyr	Ser 45	Ala	Ser	Pro
Leu	Pro 50	Ala	Ser	Lys	Arg	Pro 55	Glu	Pro	Ala	Pro	Ser 60	Phe	Asn	Val	Asp
Pro 65	Leu	Glu	Gln	Pro	Ala 70	Glu	Pro	Ser	Гла	Leu 75	Ala	ГЛа	Lys	Leu	Arg 80
Ala	Glu	Pro	Asp	Met 85	Asp	Thr	Ser	Phe	Val 90	Gly	Leu	Thr	Gly	Gly 95	Gln
Ile	Phe	Asn	Glu 100	Met	Met	Ser	Arg	Gln 105	Asn	Val	Asp	Thr	Val 110	Phe	Gly
Tyr	Pro	Gly 115	Gly	Ala	Ile	Leu	Pro 120	Val	Tyr	Asp	Ala	Ile 125	His	Asn	Ser
Asp	Lys 130	Phe	Asn	Phe	Val	Leu 135	Pro	Lys	His	Glu	Gln 140	Gly	Ala	Gly	His
Met 145	Ala	Glu	Gly	Tyr	Ala 150	Arg	Ala	Ser	Gly	Lys 155	Pro	Gly	Val	Val	Leu 160
Val	Thr	Ser	Gly	Pro 165	Gly	Ala	Thr	Asn	Val 170	Val	Thr	Pro	Met	Ala 175	Asp
Ala	Phe	Ala	Asp 180	Gly	Ile	Pro	Met	Val 185	Val	Phe	Thr	Gly	Gln 190	Val	Ser
Thr	Ser	Ala 195	Ile	Gly	Thr	Asp	Ala 200	Phe	Gln	Glu	Ala	Asp 205	Val	Val	Gly
Ile	Ser 210	Arg	Ser	Сүз	Thr	Lys 215	Trp	Asn	Val	Met	Val 220	Lys	Ser	Val	Glu
Glu 225	Leu	Pro	Leu	Arg	Ile 230	Asn	Glu	Ala	Phe	Glu 235	Ile	Ala	Thr	Ser	Gly 240
Arg	Pro	Gly	Pro	Val 245	Leu	Val	Asp	Leu	Pro 250	Lys	Asp	Val	Thr	Ala 255	Ala
Ile	Leu	Arg	Asn 260	Pro	Ile	Pro	Thr	Lys 265	Thr	Thr	Leu	Pro	Ser 270	Asn	Ala
Leu	Asn	Gln 275	Leu	Thr	Ser	Arg	Ala 280	Gln	Asp	Glu	Phe	Val 285	Met	Gln	Ser
Ile	Asn 290	Lys	Ala	Ala	Asp	Leu 295	Ile	Asn	Leu	Ala	Lys 300	Lys	Pro	Val	Leu
Tyr 305	Val	Gly	Ala	Gly	Ile 310	Leu	Asn	His	Ala	Asp 315	Gly	Pro	Arg	Leu	Leu 320
Lys	Glu	Leu	Ser	Asp 325	Arg	Ala	Gln	Ile	Pro 330	Val	Thr	Thr	Thr	Leu 335	Gln
Gly	Leu	Gly	Ser 340	Phe	Asp	Gln	Glu	Asp 345	Pro	Lys	Ser	Leu	Asp 350	Met	Leu
Gly	Met	His 355	Gly	Cys	Ala	Thr	Ala 360	Asn	Leu	Ala	Val	Gln 365	Asn	Ala	Asp
Leu	Ile 370	Ile	Ala	Val	Gly	Ala 375	Arg	Phe	Asp	Asp	Arg 380	Val	Thr	Gly	Asn
Ile 385	Ser	Lys	Phe	Ala	Pro 390	Glu	Ala	Arg	Arg	Ala 395	Ala	Ala	Glu	Gly	Arg 400
Gly	Gly	Ile	Ile	His	Phe	Glu	Val	Ser	Pro	Lys	Asn	Ile	Asn	Lys	Val

				405					410					415	
Val	Gln	Thr	Gln 420	Ile	Ala	Val	Glu	Gly 425	Asp	Ala	Thr	Thr	Asn 430	Leu	Gly
Гла	Met	Met 435	Ser	Гла	Ile	Phe	Pro 440	Val	Lys	Glu	Arg	Ser 445	Glu	Trp	Phe
Ala	Gln 450	Ile	Asn	Гла	Trp	Lys 455	Lys	Glu	Tyr	Pro	Tyr 460	Ala	Tyr	Met	Glu
Glu 465	Thr	Pro	Gly	Ser	Lys 470	Ile	Lys	Pro	Gln	Thr 475	Val	Ile	Lys	Lys	Leu 480
Ser	Гла	Val	Ala	Asn 485	Asp	Thr	Gly	Arg	His 490	Val	Ile	Val	Thr	Thr 495	Gly
Val	Gly	Gln	His 500	Gln	Met	Trp	Ala	Ala 505	Gln	His	Trp	Thr	Trp 510	Arg	Asn
Pro	His	Thr 515	Phe	Ile	Thr	Ser	Gly 520	Gly	Leu	Gly	Thr	Met 525	Gly	Tyr	Gly
Leu	Pro 530	Ala	Ala	Ile	Gly	Ala 535	Gln	Val	Ala	Lys	Pro 540	Glu	Ser	Leu	Val
Ile 545	Asp	Ile	Asp	Gly	Asp 550	Ala	Ser	Phe	Asn	Met 555	Thr	Leu	Thr	Glu	Leu 560
Ser	Ser	Ala	Val	Gln 565	Ala	Gly	Thr	Pro	Val 570	Lys	Ile	Leu	Ile	Leu 575	Asn
Asn	Glu	Glu	Gln 580	Gly	Met	Val	Thr	Gln 585	Trp	Gln	Ser	Leu	Phe 590	Tyr	Glu
His	Arg	Tyr 595	Ser	His	Thr	His	Gln 600	Leu	Asn	Pro	Asp	Phe 605	Ile	ГЛа	Leu
Ala	Glu 610	Ala	Met	Gly	Leu	Lys 615	Gly	Leu	Arg	Val	Lys 620	Lys	Gln	Glu	Glu
Leu 625	Asp	Ala	Lys	Leu	Lys 630	Glu	Phe	Val	Ser	Thr 635	ГЛа	Gly	Pro	Val	Leu 640
Leu	Glu	Val	Glu	Val 645	Asp	Lys	Lys	Val	Pro 650	Val	Leu	Pro	Met	Val 655	Ala
Gly	Gly	Ser	Gly 660	Leu	Asp	Glu	Phe	Ile 665	Asn	Phe	Asp	Pro	Glu 670	Val	Glu
Arg	Gln	Gln 675	Thr	Glu	Leu	Arg	His 680	Lys	Arg	Thr	Gly	Gly 685	Lys	His	
<21 <21 <21	0> SI 1> LI 2> TY	EQ II ENGTH YPE :	) NO 1: 29 PRT	10 99											
<21	3 > OI	RGAN	ISM:	Sac	char	omyc	es ce	erev:	isia	Ð					
<40	)> SI	EQUEI	ICE :	10											
Met 1	Thr	Asp	Ile	Gly 5	Arg	Thr	Lys	Ser	Arg 10	Asn	Tyr	Lys	Суз	Ser 15	Phe
Asp	Gly	Cys	Glu 20	Lys	Val	Tyr	Asn	Arg 25	Pro	Ser	Leu	Leu	Gln 30	Gln	His
Gln	Asn	Ser	His	Thr	Asn	Gln	Lys	Pro	Tyr	His	Cys	Asp	Glu	Pro	Gly

35 40 45 Cys Gly Lys Lys Phe Ile Arg Pro Cys His Leu Arg Val His Lys  $\operatorname{Trp}$ 50 55 60 Thr His Ser Gln Ile Lys Pro Lys Ala Cys Thr Leu Cys Gln Lys  $\operatorname{Arg}$ 

70 65 75 80

Phe Val Thr Asn Gln Gln Leu Arg Arg His Leu Asn Ser His Glu Arg 85 90 95

-continued

_															
Lys	Ser	Lys	Leu 100	Ala	Ser	Arg	Ile	Asp 105	Arg	Lys	His	Glu	Gly 110	Val	Asn
Ala	Asn	Val 115	ГЛа	Ala	Glu	Leu	Asn 120	Gly	Lys	Glu	Gly	Gly 125	Phe	Asp	Pro
Lys	Leu 130	Pro	Ser	Gly	Ser	Pro 135	Met	Сүз	Gly	Glu	Glu 140	Phe	Ser	Gln	Gly
His 145	Leu	Pro	Gly	Tyr	Asp 150	Asp	Met	Gln	Val	Leu 155	Gln	Суз	Pro	Tyr	Lys 160
Ser	Суз	Gln	Lys	Val 165	Thr	Ser	Phe	Asn	Asp 170	Asp	Leu	Ile	Asn	His 175	Met
Leu	Gln	His	His 180	Ile	Ala	Ser	Lys	Leu 185	Val	Val	Pro	Ser	Gly 190	Asp	Pro
Ser	Leu	Lys 195	Glu	Ser	Leu	Pro	Thr 200	Ser	Glu	Lys	Ser	Ser 205	Ser	Thr	Asp
Thr	Thr 210	Ser	Ile	Pro	Gln	Leu 215	Ser	Phe	Ser	Thr	Thr 220	Gly	Thr	Ser	Ser
Ser 225	Glu	Ser	Val	Asp	Ser 230	Thr	Thr	Ala	Gln	Thr 235	Pro	Thr	Asp	Pro	Glu 240
Ser	Tyr	Trp	Ser	Asp 245	Asn	Arg	Cys	Lys	His 250	Ser	Aap	Сүз	Gln	Glu 255	Leu
Ser	Pro	Phe	Ala 260	Ser	Val	Phe	Asp	Leu 265	Ile	Asp	His	Tyr	Asp 270	His	Thr
His	Ala	Phe	Ile	Pro	Glu	Thr	Leu 280	Val	Lys	Tyr	Ser	Tyr 285	Ile	His	Leu
Tyr	Lys 290	Pro	Ser	Val	Trp	Asp 295	Leu	Phe	Glu	Tyr		200			
<212 <212 <220 <222	2 > T 3 > OI 0 > FI 3 > O 3 > O	YPE : RGANI EATUH THER EOUEI	PRT ISM: RE: INF NCE:	Art ORMA 11	ific: TION	ial : : FZI	Seque F1-4	ence							
Met	Thr	Asp	Ile	Gly	Arg	Thr	Гла	Ser	Arg	Asn	Tyr	Гла	Суз	Ser	Phe
Aab	Gly	Cys	Glu	ь ГЛа	Val	Tyr	Asn	Arg	Pro	Ser	Leu	Leu	Gln	Gln	His
Gln	Asn	Ser	20 His	Thr	Asn	Gln	Lys	25 Pro	Tyr	His	Суз	Asp	30 Glu	Pro	Gly
Cys	Gly	35 Lys	Lys	Phe	Ile	Arg	40 Pro	Tyr	His	Leu	Arg	45 Val	His	Lys	Trp
Thr	50 His	Ser	Gln	Ile	Lys	55 Pro	Lys	Ala	Суз	Thr	60 Leu	Cys	Gln	Lys	Arg
65 Phe	Val	Thr	Asn	Gln	Gln	Leu	Arg	Arg	His	75 Leu	Asn	Ser	His	Glu	80 Arg
Lys	Ser	Lys	Leu	85 Ala	Ser	Arg	Ile	Asp	90 Arg	Lys	His	Glu	Gly	95 Val	Asn
Ala	Asn	Val	100 Lvs	Ala	Glu	Leu	Asn	105 Glv	Lvs	Glu	Glv	Glv	110 Phe	Asp	Pro
Lare	Lev	115 Prc	Sor	G1	Sor	Dro	120 Mot		2 <sup>2</sup>	G1.,	-1 Glu	125 Pho	Cor	Clr.	C1.7
пЛа	цец 130	F.T.O	ser,	σтλ	ser	135	Met	сув	σтλ	σти	140	rile	ser	GTU	σтλ
His	Leu	Pro	Gly	Tyr	Asp 150	Asp	Met	Gln	Val	Leu 155	Gln	Cys	Pro	Tyr	Lys 160

Ser	Cys	Gln	Lys	Val 165	Thr	Ser	Phe	Asn	Asp 170	Asp	Leu	Ile	Asn	H1s 175	Met
Leu	Gln	His	His 180	Ile	Ala	Ser	Lys	Leu 185	Val	Val	Pro	Ser	Gly 190	Asp	Pro
Ser	Leu	Lys 195	Glu	Ser	Leu	Pro	Thr 200	Ser	Glu	Lys	Ser	Ser 205	Ser	Thr	Asp
Thr	Thr 210	Ser	Ile	Pro	Gln	Leu 215	Ser	Phe	Ser	Thr	Thr 220	Gly	Thr	Ser	Ser
Ser 225	Glu	Ser	Val	Asp	Ser 230	Thr	Thr	Ala	Gln	Thr 235	Pro	Thr	Aab	Pro	Glu 240
Ser	Tyr	Trp	Ser	Asp 245	Asn	Arg	Суз	Lys	His 250	Ser	Asp	Cys	Gln	Glu 255	Leu
Ser	Pro	Phe	Ala 260	Ser	Val	Phe	Asp	Leu 265	Ile	Asp	His	Tyr	Asp 270	His	Thr
His	Ala	Phe 275	Ile	Pro	Glu	Thr	Leu 280	Val	Lys	Tyr	Ser	Tyr 285	Ile	His	Leu
Tyr	Lys 290	Pro	Ser	Val	Trp	Asp 295	Leu	Phe	Glu	Tyr					
<210 <211 <212 <213	)> SH L> LH 2> TY 3> OF	EQ II ENGTH (PE : RGANJ	) NO 1: 45 PRT [SM:	12 58 Saco	charo	omyce	es ce	erev:	isiae	9					
<400	)> SE	QUEN	ICE :	12											
Met 1	Val	Ala	Asn	Trp 5	Val	Leu	Ala	Leu	Thr 10	Arg	Gln	Phe	Asp	Pro 15	Phe
Met	Phe	Met	Met 20	Val	Met	Gly	Val	Gly 25	Ile	Ser	Ser	Asn	Ile 30	Leu	Tyr
Ser	Phe	Pro 35	Tyr	Pro	Ala	Arg	Trp 40	Leu	Arg	Ile	Суз	Ser 45	Tyr	Ile	Met
Phe	Ala 50	Ile	Ala	Суз	Leu	Ile 55	Phe	Ile	Ala	Val	Gln 60	Ala	Leu	Gln	Ile
Leu 65	His	Leu	Ile	Val	Tyr 70	Ile	Гла	Glu	Гла	Ser 75	Phe	Arg	Glu	Tyr	Phe 80
Asn	Asp	Phe	Phe	Arg 85	Asn	Met	Lys	His	Asn 90	Leu	Phe	Trp	Gly	Thr 95	Tyr
Pro	Met	Gly	Leu 100	Val	Thr	Ile	Ile	Asn 105	Phe	Leu	Gly	Ala	Leu 110	Ser	Lys
Ala	Asn	Thr 115	Thr	Lys	Ser	Pro	Thr 120	Asn	Ala	Arg	Asn	Leu 125	Met	Ile	Phe
Val	Tyr 130	Val	Leu	Trp	Trp	Tyr 135	Asp	Leu	Ala	Val	Cys 140	Leu	Val	Ile	Ala
Trp 145	Gly	Ile	Ser	Phe	Leu 150	Ile	Trp	His	Asp	Tyr 155	Tyr	Pro	Leu	Glu	Gly 160
Ile	Gly	Asn	Tyr	Pro 165	Ser	Tyr	Asn	Ile	Lys 170	Met	Ala	Ser	Glu	Asn 175	Met
Lys	Ser	Val	Leu 180	Leu	Leu	Asp	Ile	Ile 185	Pro	Leu	Val	Val	Val 190	Ala	Ser
Ser	Суз	Gly 195	Thr	Phe	Thr	Met	Ser 200	Glu	Ile	Phe	Phe	His 205	Ala	Phe	Asn
Arg	Asn 210	Ile	Gln	Leu	Ile	Thr 215	Leu	Val	Ile	Суз	Ala 220	Leu	Thr	Trp	Leu
His 225	Ala	Ile	Ile	Phe	Val 230	Phe	Ile	Leu	Ile	Ala 235	Ile	Tyr	Phe	Trp	Ser 240

Len															
Deu	Tyr	Ile	Asn	Lys 245	Ile	Pro	Pro	Met	Thr 250	Gln	Val	Phe	Thr	Leu 255	Phe
Leu	Leu	Leu	Gly 260	Pro	Met	Gly	Gln	Gly 265	Ser	Phe	Gly	Val	Leu 270	Leu	Leu
Thr	Asp	Asn 275	Ile	Lys	Lys	Tyr	Ala 280	Gly	Lys	Tyr	Tyr	Pro 285	Thr	Asp	Asn
Ile	Thr 290	Arg	Glu	Gln	Glu	Ile 295	Leu	Thr	Ile	Ala	Val 300	Pro	Trp	Суз	Phe
Lys 305	Ile	Leu	Gly	Met	Val 310	Ser	Ala	Met	Ala	Leu 315	Leu	Ala	Met	Gly	Tyr 320
Phe	Phe	Thr	Val	Ile 325	Ser	Val	Val	Ser	Ile 330	Leu	Ser	Tyr	Tyr	Asn 335	Lys
Lys	Glu	Ile	Glu 340	Asn	Glu	Thr	Gly	Lys 345	Val	ГÀа	Arg	Val	Tyr 350	Thr	Phe
His	Lys	Gly 355	Phe	Trp	Gly	Met	Thr 360	Phe	Pro	Met	Gly	Thr 365	Met	Ser	Leu
Gly	Asn 370	Glu	Glu	Leu	Tyr	Val 375	Gln	Tyr	Asn	Gln	Tyr 380	Val	Pro	Leu	Tyr
Ala 385	Phe	Arg	Val	Leu	Gly 390	Thr	Ile	Tyr	Gly	Gly 395	Val	Cys	Val	Cys	Trp 400
Ser	Ile	Leu	Сүз	Leu 405	Leu	Сүз	Thr	Leu	His 410	Glu	Tyr	Ser	Lys	Lys 415	Met
Leu	His	Ala	Ala 420	Arg	Γλa	Ser	Ser	Leu 425	Phe	Ser	Glu	Ser	Gly 430	Thr	Glu
Lys	Thr	Thr 435	Val	Ser	Pro	Tyr	Asn 440	Ser	Ile	Glu	Ser	Val 445	Glu	Glu	Ser
-	Ser	Ala	Len	Agn	Dho	Thr	Ara	Len	Δla						
Asn	450		Deu	нар	rne	455		Dea	mu						
Asn <210	450 )> SE	20 11	D NO	13	r ne	455		Deu	niu						
<210 <211	450 )> SE L> LE	EQ II ENGTH	) NO H: 42	13 27	FIIC	455		Lea	ni a						
<pre>Asn &lt;210 &lt;211 &lt;212 &lt;213</pre>	450 )> SE )> LE 2> TY 3> OF	EQ II ENGTH PE : RGANI	D NO I: 42 PRT [SM:	13 27 Salt	nonel	455 Lla t	ring		- Tita						
<pre>Asn &lt;210 &lt;211 &lt;212 &lt;213 &lt;400</pre>	450 )> SE L> LE 2> TY 3> OF	EQ II ENGTH (PE : RGAN] EQUEN	) NO H: 42 PRT ISM:	13 27 Salr 13	none]	455	rng	-	- III u						
Asn <210 <211 <212 <213 <400 Met 1	450 )> SE  > LE 2> TY 3> OF )> SE Glu	EQ II ENGTH (PE: RGAN] EQUEN Ser	) NO H: 42 PRT (SM: ICE: Leu	13 27 Salt 13 Thr 5	none] Leu	455 Lla t Gln	ryphi Pro	Ile	Ala 10	Arg	Val	Asp	Gly	Ala 15	Ile
<pre>Asn &lt;210 &lt;211 &lt;212 &lt;213 &lt;400 Met 1 Asn</pre>	450 )> SE  > LE 2> TY 3> OF Glu Leu	EQ II ENGTH (PE : RGAN) EQUEN Ser Pro	D NO H: 42 PRT ISM: JCE: Leu Gly 20	13 27 Salr 13 Thr 5 Ser	none] Leu Lys	455 lla t Gln Ser	Pro Val	Ile Ser 25	Ala 10 Asn	Arg Arg	Val Ala	Asp Leu	Gly Leu 30	Ala 15 Leu	Ile Ala
<pre>&lt;210 &lt;211 &lt;212 &lt;212 &lt;212 &lt;400 Met 1 Asn Ala</pre>	450 )> SE 2> LE 2> TY 3> OF Glu Leu Leu	EQ II ENGTH YPE: CGANJ EQUEN Ser Pro Ala 35	D NO H: 42 PRT ISM: JCE: Leu Gly 20 Cys	13 27 Salr 13 Thr 5 Ser Gly	Leu Lys Lys	455 lla t Gln Ser Thr	Pro Val Val 40	Ile Ser 25 Leu	Ala 10 Asn Thr	Arg Arg Asn	Val Ala Leu	Asp Leu Leu 45	Gly Leu 30 Asp	Ala 15 Leu Ser	Ile Ala Asp
<pre>&lt;210 &lt;211 &lt;212 &lt;213 &lt;400 Met 1 Asn Ala Asp</pre>	<pre>450 /************************************</pre>	EQ II ENGTH PPE: CQUEN Ser Pro Ala 35 Arg	D NO H: 42 PRT ISM: NCE: Leu Cys His	13 27 Salr 13 Thr 5 Ser Gly Met	nonel Leu Lys Lys Leu	lla t Gln Ser Thr Asn 55	Val Val Ala	Ile Ser 25 Leu Leu	Ala 10 Asn Thr Ser	Arg Arg Asn Ala	Val Ala Leu Leu	Asp Leu 45 Gly	Gly Leu 30 Asp Ile	Ala 15 Leu Ser Asn	Ile Ala Asp Tyr
<pre>&lt;210 &lt;211 &lt;212 &lt;213 &lt;400 Met 1 Asn Ala Asp Thr 65</pre>	<pre>450 &gt;&gt; SE &gt;&gt; LE &gt;&gt; LE &gt;&gt; OF Glu Leu Val 50 Leu</pre>	2Q III ENGTH YPE: CGANJ Ser Pro Ala 35 Arg Ser	D NO H: 42 PRT ISM: JCE: Leu Cys His Ala	13 27 Salr 13 Thr 5 Ser Gly Met Asp	nonel Leu Lys Leu Arg 70	455 lla t Gln Ser Thr Asn 55 Thr	Pro Val Val Ala Arg	Ile Ser 25 Leu Leu Cys	Ala 10 Asn Thr Ser Asp	Arg Arg Asn Ala Ile 75	Val Ala Leu 60 Thr	Asp Leu 45 Gly Gly	Gly Leu 30 Asp Ile Asn	Ala 15 Leu Ser Asn Gly	Ile Ala Asp Tyr Gly 80
<pre>Asn &lt;2110 &lt;211 &lt;212 &lt;213 &lt;400 Met 1 Asn Ala Asp Thr 65 Pro</pre>	<pre>// 450 // 5 SH // 2&gt; TY // 3&gt; OF Glu Leu Leu Leu Leu Leu Leu</pre>	EQ II ENGTH PPE: CGANI GQUEN Ser Ala 35 Arg Ser Arg	) NO H: 42 PRT ISM: ICE: Leu Gly 20 Cys His Ala	13 27 Salr 13 Thr 5 Ser Gly Met Asp Ser 85	none] Leu Lys Lys Leu Arg 70 Gly	455 lla t Gln Ser Thr Asn 55 Thr Thr	yphi Pro Val Val Ala Arg Leu	Ile Ser 25 Leu Cys Glu	Ala 10 Asn Thr Ser Asp Leu 90	Arg Arg Asn Ala Ile 75 Phe	Val Ala Leu 60 Thr Leu	Asp Leu 45 Gly Gly Gly	Gly Leu 30 Asp Ile Asn Asn	Ala 15 Leu Ser Asn Gly Ala 95	Ile Ala Asp Tyr Gly 80 Gly
<pre>Asn &lt;2110 &lt;211 &lt;212 &lt;213 &lt;400 Met 1 Asn Ala Asp Thr 65 Pro Thr</pre>	<pre>/450 //450 //2015</pre>	EQ II ENGTI PE: CQUEN Ser Pro Ala 35 Arg Ser Arg Met	O NO H: 42 PRT ISM: ICE: Leu Cys Ala Ala Arg 100	13 27 Salt 13 Thr 5 Ser Gly Met Asp Ser 85 Pro	nonel Leu Lys Lys Leu Arg 70 Gly Leu	lla t Gln Ser Thr Asn 55 Thr Thr Ala	Pro Val Val Ala Arg Leu Ala	Ile Ser 25 Leu Cys Glu Ala 105	Ala 10 Asn Thr Ser Asp Leu 90 Leu	Arg Arg Asn Ala Ile 75 Phe Cys	Val Ala Leu Eeu Leu Leu	Asp Leu 45 Gly Gly Gly Gly	Gly Leu 30 Asp Ile Asn Asn Gln 110	Ala 15 Leu Ser Asn Gly Ala 95 Asn	Ile Ala Asp Tyr Gly 80 Gly Glu
Asn <2110 <211 <212 <213 <400 Met 1 Asn Ala Asp Thr 65 Pro Thr Ile	<pre>/450 /&gt;&gt; SF /&gt;&gt; LE /&gt;&gt; SF /&gt;&gt; OF //&gt;SF Glu Leu Leu Leu Leu Leu Leu Leu Leu Leu Val 50 Leu Leu Val</pre>	EQ III ENGTH (PE: CQUEN Ser Pro Ala 35 Arg Ser Arg Met Leu 115	D NO H: 42 PRT ISM: ICE: Leu Gly 20 Cys His Ala Ala Ala Ang 100 Thr	13 27 Salr 13 Thr 5 Ser Gly Met Asp Ser 85 Pro Gly	none] Leu Lys Lys Leu Arg 70 Gly Leu Glu	lla t Gln Ser Thr Asn 55 Thr Thr Ala Pro	yphi Pro Val Val Ala Arg Leu Ala Arg 120	Ile Ser 25 Leu Leu Cys Glu Ala 105 Met	Ala 10 Asn Thr Ser Asp Leu 90 Leu	Arg Arg Asn Ala Ile 75 Phe Cys Glu	Val Ala Leu 60 Thr Leu Leu Arg	Asp Leu Leu Gly Gly Gly Gly Pro 125	Gly Leu 30 Asp Ile Asn Asn Gln 110 Ile	Ala 15 Leu Ser Asn Gly Ala 95 Asn Gly	Ile Ala Asp Tyr Gly Gly Glu His
<pre>Asn &lt;2110 &lt;211 &lt;212 &lt;212 &lt;213 &lt;400 Met 1 Asn Ala Asp Thr 65 Pro Thr Ile Leu</pre>	<pre>// 450 // 450 // 5 EF // 5 EF // 5 EF // 5 EF // 5 EF // 5 // 10 // 10 //</pre>	EQ II ENGTI PE: CQUEN Ser Pro Ala 35 Arg Arg Arg Met Leu 115 Asp	O NO H: 42 PRT (SM: Leu Gly 20 Cys His Ala Ala Arg 100 Thr Ser	13 27 Salr 13 Thr 5 Ser Gly Met Asp Ser 85 Pro Gly Leu	nonel Leu Lys Lys Leu Arg 70 Gly Leu Glu Arg	lla t Gln Ser Thr Asn 55 Thr Thr Ala Pro Gln 135	ryphi Pro Val Val Ala Arg Leu Ala Arg 120 Gly	Ile Ser 25 Leu Leu Cys Glu Ala 105 Met Gly	Ala 10 Asn Thr Ser Asp Leu 90 Leu Lys Ala	Arg Arg Asn Ala Ile 75 Phe Cys Glu Asn	Val Ala Leu Eeu Leu Leu Leu Arg Ile 140	Asp Leu 45 Gly Gly Gly Gly Pro 125 Asp	Gly Leu 30 Asp Ile Asn Gln 110 Ile Tyr	Ala 15 Leu Ser Asn Gly Ala 95 Asn Gly Leu	Ile Ala Asp Tyr Gly Gly Glu His Glu

127

145 150 155 160	
Asp Ile Glu Val Asp Gly Ser Val Ser Ser Gln Phe Leu Thr Ala Leu 165 170 175	
Leu Met Thr Ala Pro Leu Ala Pro Glu Asp Thr Ile Ile Arg Val Lys 180 185 190	
Gly Glu Leu Val Ser Lys Pro Tyr Ile Asp Ile Thr Leu Asn Leu Met 195 200 205	
Lys Thr Phe Gly Val Glu Ile Ala Asn His His Tyr Gln Gln Phe Val 210 215 220	
Val Lys Gly Gly Gln Gln Tyr His Ser Pro Gly Arg Tyr Leu Val Glu 225 230 235 240	
Gly Asp Ala Ser Ser Ala Ser Tyr Phe Leu Ala Ala Gly Ala Ile Lys 245 250 255	
Gly Gly Thr Val Lys Val Thr Gly Ile Gly Arg Lys Ser Met Gln Gly 260 265 270	
Asp Ile Arg Phe Ala Asp Val Leu Glu Lys Met Gly Ala Thr Ile Thr 275 280 285	
Trp Gly Asp Asp Phe Ile Ala Cys Thr Arg Gly Glu Leu His Ala Ile 290 295 300	
Asp Met Asp Met Asn His Ile Pro Asp Ala Ala Met Thr Ile Ala Thr 305 310 315 320	
Thr Ala Leu Phe Ala Lys Gly Thr Thr Thr Leu Arg Asn Ile Tyr Asn 325 330 335	
Trp Arg Val Lys Glu Thr Asp Arg Leu Phe Ala Met Ala Thr Glu Leu 340 345 350	
Arg Lys Val Gly Ala Glu Val Glu Glu Gly His Asp Tyr Ile Arg Ile 355 360 365	
Thr Pro Pro Ala Lys Leu Gln His Ala Asp Ile Gly Thr Tyr Asn Asp	
His Arg Met Ala Met Cys Phe Ser Leu Val Ala Leu Ser Asp Thr Pro	
Val Thr Ile Leu Asp Pro Lys Cys Thr Ala Lys Thr Phe Pro Asp Tyr	
405 410 415 Phe Glu Glu Leu Ala Arg Met. Ser Thr Pro Ala	
420 425	
<210> SEQ ID NO 14 <211> LENGTH: 9089 <212> TYPE: DNA <213> ORGANISM: artificial sequence <220> FEATURE: <223> OTHER INFORMATION: constructed plasmid	
<400> SEQUENCE: 14	
togogogttt oggtgatgac ggtgaaaaco totgacacat goagotooog gagaoggtoa	60
cagettgtet gtaageggat geegggagea gacaageeeg teagggegeg teagegggtg	120
ttggcgggtg tcggggctgg cttaactatg cggcatcaga gcagattgta ctgagagtgc	180
accataccac agettttcaa ttcaattcat cattttttt ttattettt ttttgattte	240
ggtttctttg aaatttttt gattcggtaa tctccgaaca gaaggaaggaa cgaaggaagg	300
agcacagact tagattggta tatatacgca tatgtagtgt tgaagaaaca tgaaattgcc	360
caguallett aaccoaactg cacagaacaa aaacctgcag gaaacgaaga taaatcatgt	420
cgaaagctac atataaggaa cgtgctgcta ctcatcctag tcctgttgct gccaagctat	480

129

tatatatat gaagatag caasaasat tytigatta attegatgit egisecaas         940           aggattat         gaagatag         gaagatag         600           teggitatat         gaagatag         gaagatag         600           caagaas         tettistet         tetagaga         700           aattagaga         tetagagat         gaagatag         700           aatgaaga         tetagagat         gagaataga         900           gagaatat         taggaacat         900         900           gagaatat         taggaacat         900         900           gagaatat         taggaacat         900         900           gagaatat         taggaacat         900         900           gagaatat         taggaacat         gagaatat         900           gagaatat         taggaatat         1000         1000           gagaatat         taggagatat         900         1000           giggigge         gaaataga         aatggagad         1000         1000           giggigge         gaaataga         taggaagad         1000         1000           giggigge         gaaagaada         taggaagada         1000         1000           giggigge         taggaacat         tatagaaga	<ul> <li>tatatati gaagaaag caacaact tytgtgette attgggett gitagtaacea</li> <li>sig</li> <li>eggettett gagttagt gaagaatg goorgaag ta goorgaag goorgaag gaatatee</li> <li>coagaaa thttisee ttegagge gaatatagaa tagaagag goorgaa gaagaace</li> <li>aattgorga ettegagg gaatatagaa tagaagag goorgaa gaagaace</li> <li>aggettett goorgaagaatg digaagaag goorgaa gaagaage</li> <li>eggeseets aggeettig attiaagaa tagaagag caagagte caasagte</li> <li>gaagaata tagagaatg goorgaaga atgaaggte gaagata acagtag</li> <li>coggtagg ttagaaca goorgaaga atgaaggte gaagata gaagaag</li> <li>coggtagg ttagaaca goorgaaga atgaaggte gaagata gaagaag</li> <li>coggtagg ttagaaca goorgaaga atgaaggte gaagatat gaagaag</li> <li>coggtagg ttagaaca goorgaagaa gaagaag gaagatt gaagaag</li> <li>coggtagg ttagaaca goorgaaga atgaagte gaagatat gaagatag</li> <li>coggtagg ttagaaca goorgaagaa gaagaag goorgaaga attiggaag</li> <li>coggtagg ttagaaca goorgaagaa gaagaag goorgaaga attiggaag</li> <li>coggtagg ttagaaca goorgaagaa gaagaag goorgaaga attiggaag</li> <li>coggtagg ttagaaca goorgaagaa gaagaaga goorgaaga attiggaag</li> <li>diattagge taagaata cooratag aatacca gaagatag acagtaga</li> <li>coorataa ataaccga tagaccaga atgogtag tagatata taaccagaa</li> <li>jaagactaa gooraacaa aaaccga tagacgaa goorgaaga goorgaa atgog aa</li> <li>coorataa ataaccga tagaccaga taggetgg tagatata tagaccagaa</li> <li>jaagactaa tagagaacg gooraacga atagagteg goorgaa atgog aa</li> <li>coorataa ataaccga tagaccaga taggetgg tagatata cooraacgaa</li> <li>jaagactaa toggaacca ataaccaca ataacgaa goorgaag goorgaa goorgaa</li> <li>jaagactaa toggaacaa cagacca ataacgaa goorgaaga goorgaaca goorgaa</li> <li>jaagactaa toggaacaa cagacaga goorgaaga goorgaa goorgaaca</li> <li>jaagacaa tagagaaga gooraacaa tagaccaga goorgaaga goorgaa atag</li> <li>jaagaagaa gooraagaa tagaccaa ataaccaa goorgaaga goorgaaca</li> <li>jaagaagaa gooraagaa tagaccaa atagaaga goorgaa goorgaaca</li> <li>jaagaacaa toggaacaa cagacaga goorgaaga goorgaaca goorgaagaa</li> <li>jaagaacaa toggaacaa caadaacaa goorgaaga goorgaa goorgaaga</li> <li>jaagaacaa toggaacaa c</li></ul>								
aggattartgangtagtgangtagtgangtattartfieldtyptatttortiggggggenagtiagorgstattyfieldattigggttortigggggtigtaggagenagtatggfieldaggatggigggatcadatigtaggagenagtatggifieldaggatggigggatcadagggaggagaagtatgaifieldaggatggigggaggaggagaagtaggifieldfieldaggatggigggaggaggagaagtaggifieldfieldgggatggigggaggaggagaagtaggifieldfieldgggatggiggaggaggigaaggaggigaaggaggifieldgggatggigaaggaggigaaggaggigaaggaggifieldgggatggigaaggaggigaaggaggigaaggaggifieldgggatggigaaggaggigaaggaggigaaggaggifieldgggatggigaaggaggigaaggaggigaaggaggifieldgggaggagagaaggaggigaaggaggifieldfieldgggaggagaggaggagaaggaaggaggifieldfieldgggaggagaggaggagagggaaggaggiggaggagagifieldggaggaggiggaggagaaggaggaggaggifieldfieldggagggiggaggagagggaaggaggiggaggaggifieldggagggiggaggaggiggaaggaggiggaggaggifieldggagggiggaggaggiggaaggaggiggagggifieldggagggiggaaggaggiggaaggaggiggagggifieldggaaggiggaggaggiggaaggaggiggaaggi <td< td=""><td>upgattactgagttagttgagttagttgagttagttgagttagttgagttagttgagttagttgagttagttgagttagttgagttagttgagttagttagttagttagttagtgagttagttagttagttagttagttagttagttagttag</td><td>ttaatatcat</td><td>gcacgaaaag</td><td>caaacaaact</td><td>tgtgtgcttc</td><td>attggatgtt</td><td>cgtaccacca</td><td>540</td><td></td></td<>	upgattactgagttagttgagttagttgagttagttgagttagttgagttagttgagttagttgagttagttgagttagttgagttagttgagttagttagttagttagttagtgagttagttagttagttagttagttagttagttagttag	ttaatatcat	gcacgaaaag	caaacaaact	tgtgtgcttc	attggatgtt	cgtaccacca	540	
typictictt       set0         crangtacan       tittitict       tictacagaa       gammattige       g	typitatiott gartgattt tooratgagag goarastita googotaag goattatoog       660         craagtaraa tuttttatto ttogaagaa gaaattigo tgacattgg aataganca       780         anttgongis otoigoggi guatacaya lagaagang gonganga sagaataca       840         aggatatgi gugocongg attgitagong gitsganga gonganga gaaganca       840         aggatatata tangggatat gitsgang atgataga canagatti gitsgogt       900         gegatatata tangggatat gitsganga atganganga tangangana       900         gegatatad tangggatat gitsganga ganganga tangangana       900         gitsgitto tangggata guganga dagangata capitaga gonganga landanganga       900         gitsgitto tanggatat guangataa guanganga tangangana       1000         gitsgitto tanggatat guangataa guanganga tangangana       1000         gitsgitto tanggatat guangataa guanganga tangangan       1000         gitsgitta gitangana anaotgat tanganga tiggingan tantagana       1000         gitsgitta tanganga tangangan tigginganga tangangana       1000         gitsgitta tangangat acgataga tangangana       1000         gitsgitta tangangat acgataga tangangana       1000         gittatat tangangat gipangata gugangana       1000         gigatatat tanganga tiggataga gugangang gugangang       1000         gigatgitta ataganga tiggataga gugangang       1000         gigatgitta tangangang gugangang gugangang gugangang       1000	aggaattact	ggagttagtt	gaagcattag	gtcccaaaat	ttgtttacta	aaaacacatg	600	
<ul> <li>ccaqtaca tittitato titgaagaa gaaattig tigacattig tatacagtea</li> <li>attigaat ototigaaga guittiga guittigang guiggaaga guittig agaggacat acgatigaa</li> <li>aggaactag aggoctitig atgitagang attigtag cauggocaa guiggacat</li> <li>gagaacta taaggacattig guiggaagag ataggoca caaggitti gittagaga</li> <li>coggitiggi titgaattig cigacattig cauggocaa acagtitti gittagaga</li> <li>coggitiggi titagatga aggagaga guiggagga ataggota cauggitta guitagaga</li> <li>coggitiggi titagatga agggagaga atagagga cauggota acagtitti gittaggat</li> <li>coggitiggi titagatga agggagag ataggota acagtatag accegiggat</li> <li>liatigota aggagacti gacattati titiggaga guitti gaaaggaa tattiggaa</li> <li>gittaggoca gcaaactaa acactgat tataccca logogitiga ataccegac</li> <li>liata atagga titaattaa cogaacag tiggoggaa ataggogaa</li> <li>coggitag guiga guiga guiga guiga guittiga guigagga atattigaa</li> <li>gittaggoca gcaaacag cogaacag taggota atggogaaca</li> <li>liata ataagaa cogaacag taggotaga atggogaaca</li> <li>gittaggoca guigaagag guigagaag agaaggag guigagaa guigaagga</li> <li>coctataa ataaagaa tagaccaa aaactgag tittagaa guigaaggat guigagaca</li> <li>liaagacataa tiggaacca aaaggaga guaaggag guigagaag guigagaaga</li> <li>gittagogoc actadgoga cactacaca aaccegocg gotatagg guigagaag</li> <li>liaagacataa taggagaag guaaggag guigaggag guigaggag guigagaag</li> <li>gittagogo cactagota ticaggaga guigagaga guigaggag guigagaag</li> <li>gittagogo cactago taccego guaacaga guigaggoga tiggigogi guigagaa</li> <li>liaagacataa taggagaag taacgaca guigaggaga tiggigogaa guigagaga</li> <li>gittagogo cactago taccego guaacaga guigaggag guigagaa guigagaga</li> <li>gittagogo cactago taccego guigagaag guigagaag guigagaag guigagaag</li> <li>gittagogo cactago taccego guigagaag guigagaag guigagaag guigagaag</li> <li>gittagota guigagaag guigagaag guigagaag guigagaag guigagaag</li> <li>gittagota titaggaga guigagaag guigagaag guigagaag guigagaag</li> <li>gittagota titaggaga duigagaga guigagaga guigagaaga guigagaga tittigagaa</li> <li>gittagota titaggaga duigagaga guigagaaga guigagag</li></ul>	comptana         Littiacic ticgaagae gaaattige tyacitige taceaque         720           antigaqia cictigggig gistacaque taquaque gyeggicata aquattaque         780           acggingtigg gygocoaggi attigtaque gyttigaaga gygogoaga gaagtacata         840           aggaactag aggocittig attigtaque antigtane anaggicae citabetate         960           gaaataa taaggicate gyttigaaga attigtage aattigta cagtiggit attigtace         960           comptana         920           comptana         920           gagaataa         920           comptana         1200           comptana         1200           comptana         1200           comptana         1200           comptana         1200           comptana <td>tggatatctt</td> <td>gactgatttt</td> <td>tccatggagg</td> <td>gcacagttaa</td> <td>gccgctaaag</td> <td>gcattatccg</td> <td>660</td> <td></td>	tggatatctt	gactgatttt	tccatggagg	gcacagttaa	gccgctaaag	gcattatccg	660	
attgegitgtategegitgtategegatgtategegatgtgegegatgtgegegatgtgegegatgtgegegatgtgegegatgtgegegatgtgegegatgtgegegatgtge	Astgegate detregogget getatecage steggacgate acquirages760acquiriget gegeecangt attgetages gittiganges gegergena gangtacata640aqgaacdata tagggetattig attgetage attgicatig cangquire cetatecatig900gegaacdata tagggittig tittageag attgigetage cangquire gattgigit stategaca1000coggitgiggittig gitgangag atgaaggita gattgigits stategaca1000attgiggetet tacanggatet gefgangag atgaaggita gattgigs stategaca1000gigtgigeta gatgaaggit gatgitage atgaggita gatgitag attgigat1000gigtgigeta gatgaaggit gatgitage atgaggitag atgaggatet1000gigtgigeta ggenaacta anacetget tataagtaa tiggaggata tattgigaa1000gatgogeta gegaaata cacgecacag aactgigaa gigtgiga ataceegac1200gigtgigeta gigaagaa tagacetagi tataagtaa tiggacgiga ataceegac1200gatgogeta gegaaaata cacgecacag aactgigaa cittgigaaca1800cocctataa atcaasaga tagacegag taggitgig tigtgeteca gittgigaa1800gigatgoet attaagaa tiggace cacgittag agetgaggitage gittgigetag1600gigatgoet atcaagaa cataaceaca accoccego gigaaaace gicaacaga1800gigatgoet accgaacet aacggaaga ggaagaag gaaggae gaggatag tiggaggag1800gigatagetig cacegotegi gitaacacea accoccego gigatgoeg cacegaga1800gigatageta tatagetag tiggace cacegattag gigatgee tiggaggae1800gigatageta tatagetag gigaagaag gaaggae gatgetige tigaaggee1800gigatageta tatagetag gigaagaag gaaggae gatgetige tigaaggee1800gigatageta tatagetag gigaagaag gaaggae gatgetige tigaaggee1800gigatageta tatagetag gigaaggae gatgetige tigaaggee1800gigatageta tatagetag gigatgeti tigaagg	ccaagtacaa	tttttactc	ttcgaagaca	gaaaatttgc	tgacattggt	aatacagtca	720	
acggtgtgtgtggggccaggtattgttagcggtttgtagcaggagtacta640aggaactagaggcctttgattgttagcaggaaggagtcoctattagt900gagaatatataagggactgttgagagatgaaggag100coggtgtggtttaagtgacagggaaggatgaggacg100attgtagttaagtgacagggaggatgaggagg110gygatgetaggtaggggtgaacgtatagtaaggacg1200gygatgetaggtaggggtgaacgtatagaaggacg1200gagtgggggaaggacggaacgtatagaaggacg1200gagtgggaggaaggacggaaggacggaaggacg1200gagtgggaggaaggaggagaaggacg1200gagtggagaggaggaaggagaaggacg1200aattaggottatacgtttatacca1200aggtggaaggagaaacggagaaac1200aggtgaaggagaaacggagaaac1200aggtgaaggagaaacgaaggaca1200aggtgaagaagaacagaaggaca1200aggedttataccataagged1200aggedattaggacgaaggacg1200aggedattaggacgaaggacg1200ggagaccaattaggacgaaggacg1200ggagaccattaggacggaaggacg1200ggagaccattaggegggaaggacg1200ggagaccattaggegggaaggacg1200ggagaccattaggegggaaggacg1200ggaaccattaggegggaaggacg1200 <td< td=""><td>Acggtegtegtgggecocagtattgttageggtttgangeaggegectaggggaggaactagaggaactagaggaactagaggaactagbeecoggtegtegtctaaggaactagggtgaggcatggaaggabeeattgtectaaggaactagggtgagggcatggaggabeeattgtegtetataaggaactagggtgagggcatggaggabeeattgtegtetataaggaactagggtaggggcatggaggabeeattgtegtetaaggtagggtgaacgtaggaaggaggabeeggtagggegaaagtagggaaagtaggcatggaggabeeggtagggegaaagtaggagaaagtaggcatggaggabeeggtagggeggaagaatacatgaaggacatggaggabeeggtagggeggaagaatacatgaaggaggaaagtagbeeaattaggettaattagattaagtaattaaggadbeecoggtetagggaaagtaacatgaaggagaaagtaggbeeagateggecataagaattaaggadttaaggadbeeagateggegaagaacaaggaagaagaggaaggagbeeggaagaacaacataagaacataaggagggaagaagabeeggaagaacaacataagaacggtataabeebeeggaagaacaacataagaacggtataagbeebeeggaagaacaacataagaacggtataabeebeeggaagaacaaggaagaagaggaagaagaggaagaagaagbeeggaagaacaacataagaacggaagaacaabeebeeggaagaacaacataggaagaggaagaagaagaaaggaagbe</td><td>aattgcagta</td><td>ctctgcgggt</td><td>gtatacagaa</td><td>tagcagaatg</td><td>ggcagacatt</td><td>acgaatgcac</td><td>780</td><td></td></td<>	Acggtegtegtgggecocagtattgttageggtttgangeaggegectaggggaggaactagaggaactagaggaactagaggaactagbeecoggtegtegtctaaggaactagggtgaggcatggaaggabeeattgtectaaggaactagggtgagggcatggaggabeeattgtegtetataaggaactagggtgagggcatggaggabeeattgtegtetataaggaactagggtaggggcatggaggabeeattgtegtetaaggtagggtgaacgtaggaaggaggabeeggtagggegaaagtagggaaagtaggcatggaggabeeggtagggegaaagtaggagaaagtaggcatggaggabeeggtagggeggaagaatacatgaaggacatggaggabeeggtagggeggaagaatacatgaaggaggaaagtagbeeaattaggettaattagattaagtaattaaggadbeecoggtetagggaaagtaacatgaaggagaaagtaggbeeagateggecataagaattaaggadttaaggadbeeagateggegaagaacaaggaagaagaggaaggagbeeggaagaacaacataagaacataaggagggaagaagabeeggaagaacaacataagaacggtataabeebeeggaagaacaacataagaacggtataagbeebeeggaagaacaacataagaacggtataabeebeeggaagaacaaggaagaagaggaagaagaggaagaagaagbeeggaagaacaacataagaacggaagaacaabeebeeggaagaacaacataggaagaggaagaagaagaaaggaagbe	aattgcagta	ctctgcgggt	gtatacagaa	tagcagaatg	ggcagacatt	acgaatgcac	780	
aggaactag aggcrtttg atgtrageag aattgtrat eaagggrtee etatetat900gagaactae taagggraet stagacga egaaggea eaaggtrat egategget960tattgetea aggageat gygggaagg atgaaggt aegatggt eegateggat1200coggtgggt ttagatge aggggaagg eatggggea eaaggtrag acggtggga1200gigdagetea getaggget gaactatta tigtggaag aggaattit getaaggaa1200gigdagetea getaggget gaactatta tigtggaag aggaagta tattiggaag1200gigdagetea getagget gaactatta tigtggaag at attagaa1200gigdagetea getagget gaactatta tigtagaa toggeaga tattigaaga1200gadegetea gegaaatta eegetaag tattagea egitaattat tigtaaga1200aaattagge toaattaa tatadeat tattaceat tiggeggig aataceace1200gadegetea gegaaatt eegetag tattaee egitaatt tigtaaat1380teectaa ateaaagaa teggeeaga taggetag tigtigteea gitiggaea1500aagaceaa teggaacea tataaggaag eiggaadae gigaagge gigaaggeegg1600gigatgeee eateggeg gigaadaae eigeaagge gigggaaace gigaggeegg1600gigatgeee eategge gigaadaae eigeagge gigogegaa teggigegeegg1600gigatgeee eategeeg gigaadaae gigaaggee teggigegaa teggigegeegg1600gigtageee eategeeg gigaadaae gigaaggee teggigegeegg1600gigtageeg eetategee teggegeg gigeegeege gigeegeegg1600cotteegee gigaaggae tigggeege gigeegeege1600cotteegee gigaaggae tigggeege gigeegeege gigeegeege1600cotteegee gigeaggae tigggeege gigeegeege gigeegeege1600cotteegee gigeegee eetategeege gigeegeege gigeegeege1600cotteegee gigeegee gigeegeege gigeegeegeegeegeegeegeegeegeegeegeegeege	aggaactag aggottttig atgitageag aattgitatig eaaggottee etatetatig900gagaatatae taaggitaet gitgineatig egaaggotge eaaggittig gittageade960ceggitgiggit taaggigae aggiggaeg etatiggite acaggiteg attaggite attageade1020eggitgite taaggite aggiggaeg etatiggite acaggiteg attaggiteg attageade1020gittiggite taaggite gaactatata tigtitiggaeg aggattatt geaaaggiga1140giggitegea gitaaggite gaacgitae gaaageag etiggigaege tattiggaa1200aattaggit teaattata tigtitigaag aggattat geaaageag1200aattaggit teaattata tatategt tataceeta tigeiggiga aataceegee1200aattaggit teaattaa tatategt tataceeta tigeiggigaa ataeeggea1300teeggitaa tittigitaa ategotagi tittigaegi gitgittee gittiggaaca1500gagateeta ateggige giggitagi gitgittee gittiggaaca1500gegatgeeg eaacgaa aggigaeg eaggitege giggitegea1600giggitege gigaangga giggangaag eaggitege giggitegea1740giggitegeg eaacgae tiggitege gigaangge giggitegea1740gigtegegit eacgitege giaancee acaccee geere gettaatege eegetaagi1800centeree attaggitege tateee acaccee geere gettaatege eegetaagi1800centeree attaggitege tiggitegea giggitegea tiggitegea1900aaggateat ateggitege tiggitegea giggitegea tiggitegea1900aaggateat ateggitege gigaategea giggitegea giggitegea1900centeree attaggitege tiggitegea giggitegea2000centeree attaggitege gideacacae gaacgeege giggitegea2100aaggateat ateggitege titteet gitegeagea figgitega aggattag giggitega2100aaattageat ateggitege tateacae ceaggitegea giggitega giggitega210	acggtgtggt	gggcccaggt	attgttagcg	gtttgaagca	ggcggcagaa	gaagtaacaa	840	
gagaatatatataggatatgitgacatig <t< td=""><td>gagaatatataaggatatigitagaatigagaggagagagaggagagataggita<td>aggaacctag</td><td>aggccttttg</td><td>atgttagcag</td><td>aattgtcatg</td><td>caagggctcc</td><td>ctatctactg</td><td>900</td><td></td></td></t<>	gagaatatataaggatatigitagaatigagaggagagagaggagagataggita <td>aggaacctag</td> <td>aggccttttg</td> <td>atgttagcag</td> <td>aattgtcatg</td> <td>caagggctcc</td> <td>ctatctactg</td> <td>900</td> <td></td>	aggaacctag	aggccttttg	atgttagcag	aattgtcatg	caagggctcc	ctatctactg	900	
ttattgötö angaga atg götgganga atgangöt a ögattgöttg attatgaca i 200 cögötgögö titagatga anggögang cattgögten acagtatag acögtögdö i 200 atötögtöt i kanggat genattatta tigitögang aggackatt grannaggan i 140 gögatgöta göngangöt gangöttan ganangeng etgögangön tattgögan i 200 gatgögön gönandeta ananotgi titatadgi at tgöngöta atögöngö i 200 anattagög et chattta titatacög i tataocola igögötgöga atöögöna i 200 anattagög et chattta titatacög i tataocola igögötgöga atöögöna i 200 agatgögina göganada cögönaga attgörag atögötan atögögöna i 200 agatgögina göganada etgösöga attgögöng ögönanados götösögöna i 200 agatgöcina atonagan ingösön adögöngöngö i 200 agagatoka atonagan ingösöngö i 200 agagatoka atongösön ecögattag agettgögö gögötagö gögögöngö i 200 agagatögö enetegöng ginakacac acaccegöng götösöngö i 200 gögötgögö cöntögön ginakocac acaccegöng gögösötagi gögögötgöng i 200 tangörögö göttecön itöngörig gönandgöng gögöngötagi tögögöngön i 200 acaccitögi attaogöxa itöngögöng gögöngöngö i göngögöngö i 200 accottögöt attaogöxa itöngögöng gögöngöngö i göngögöngö i 200 accottögöt attaogöxa itöngögöngö gönöngöngö i 200 accottögi attagöngön itöngöngöng gögöngötagi gönögöngön i 200 accottögi attagöngön itöngöngöngö gögöngötagi gönögöngön i 200 accottögi attagöngön itöngöngöngö gönögöngöngö i 200 accottögöt attagöngön itöngöngöngöngö i 200 accottögöt attagöngön itönögöngöngöngöngöngöngöngöngöngöngöngöng	ttattgötö aagagacatg gytgäagag atgaegyta egattgöttg attatgacce 1020 cogstgögg ttlagatgac aaggagacg cattgögtöca acagtataga acogtgögtg 1140 gygtgötöt tacaggatt gacattatt titgötgjaag aggactatt geaaggaga 1140 gygtgögöd geaaastaa aaactgöt tattagitaa töötgöta attaactoca 1260 aanttagge tocaatta tatatcagt tattacota töögtgöga aatacogoac 1320 agatgögöta geagaaata eogotcag aaatgitaa egatgötga atacogoac 1320 agatgögöta gagaaata eogotcag aaatgitaa egotgata etaactoca 1300 tögötgögö geaaasta ataactgöt attutacos töögtgöga atacogoac 1500 agatgögöta gagaaata eogotcag aaatgitaa egitatat titgitaaat 1360 töögötgögö deaasta ataacgita tuttaaca ataggögaa atogocaa 1500 agagtocat attaagaa tagacogag taggitaga tgitgitoca gittgöga at togogtaga ekseles egiggactoca acgtocaagg gogaaaace gittataga 1560 gegatgögö deaesgaa costaccot aataagitt titggögö aggitgögg 1660 coaacgigö gagaagga gigaagaag egaaggae gögaaagoog jaaagoog 1660 coaacgigö gagaagga gigaagaag egaasgiac gögedgög egotgoca 1740 gigtagoogi cacteogo etacagota eacecogoo gettatgög eogotgaa 1690 eegoetogog cacteogo etacagotag egocactegi gigaaggoog 1860 cottogot attacogoca gitgögaaag gigaaggae gigsdögö tögötgögög 1920 taacgotoag tittocog toacgacgi gigacactec gaaggitag egocgotaa 1960 aegoatogag dittocog toacgacgi gigacactec gaaggitag egocgotaa 2040 tottocat atacgocag tiggögaaag gigaagaag gigaagaag egosgitag egocgotaa 2040 tottocat atacgacat gigogaaagga gigaagaa gigagaaga egosgitag egocgota 2040 aegoatoga etactoci gigotaagga atogigaag agitagi egotgitag egotgita 2020 aagatactat etogaaati gigotacci gigogaaga atgigotagi gigotgitaga 2220 aagatactat etogaati gigotacci egagagaa atgigotagi agigagaaga 2400 attagéaaa agatecttog taacagat tigotgita atgigotagi agigagaaga 2400 attagéaaa agatecticg taacagat ticcatati gigtigaa atgigogaaga 2400 attagéaaa agatecticg taacaagt titacatte tiggigaaga atgigagaaga 2400 attagéaaa agatecticg taacaagt titacagi attigatgi attigatgi 2530 acqtocogi giaattiga attatecta eagatgi attigatgi agigaagaa 2400 cigaocaaga aagatat egocagat titacaga attigagagi tiggigaagaa 2400 attagéaa	gagaatatac	taagggtact	gttgacattg	cgaagagcga	caaagatttt	gttatcggct	960	
coggtggg titigatgar aaggaaga cattgggtea acagtatag acogtggg1000atgtggtet taraggate gacattat titigtggaag aggatatt geaaaggaa1140gggatgeta ggtaaggg gaacgtaa aaaagaga citigggaaga tattggaa1200aaattagag ticaatta tatacagt tataagtaa tgeatgata ciaaaccaa1200aaattagag ticaatta tatacagt tatacocaa tgeggtgga atacogcaa1200agatgogtaa ggaaaata cogoacag aattgaa citiaatti tigtaaaa1300toogetaa tititigtaa atagocag attggaa ggaaaacc gitatata1400toogetaa tataa tagacoga taggetgag tigtggaaga cigtagaaga1500agagtocat attaagaa tiggotoca acocaagg gogaaaacc gitatacag1500agagtacata ateggaacca aaggeaga cogaaggac gggegetag gogaaaacc gitatacag1600gigatggic accegtege gaaacacca accegeeg gitatage gogaagaac1740gitagoggi caccitaga citaaggig gaaaggac gggegetag gogogtegg1800gigatagota taaggeaga tigggaag gggaaggaa gggaaggaa ggaaggaa ggaaggaa ggaaggaa1900gigatagota ataggeega tigggaagge gigocaga cogogtaa1900gigatagota taggegaa tigggaag gggaaggaa ggaaggaa	ccggtgtgggtitagatgac aagggaagg cattgggtca acagtataga accgtggag1000atgtggtcttacaggatct gacattat titgtggaag aggatatt gcaaagggaa1140gggtgtgta gggaagt gaacgttaa gaaagcagg ctgggaaga tattggaa1200aaattagag titcaattaa tatatcagt tattaccat gcggtgtga aataccgcac1320agatgggtaa gggaasat ccgcatcagg aattgtaa cgtaatat tigtagaa1300ccogtataa tittgtaa atagctat tittacca taggggaa ataccgcac1300cogegtaa gggaasat cgcactact aaggtgag gtgtgaa gtgtgtga ataccgcac1500agagtgcaa atagacggaa tagaccgaa atgggtaga gtgtgtga gtgtgtga ataccgcaa1500agagtgcca atacgaaga tagaccgaa taggtgag ggggaagac gtgtggaaggt1500gggtgggg caccgaagag ggggacc ccggattag aggttggg ggggggaagg1600gggatggcc actacgtaa caccacca aaccaggg ggggctagg gggggggg1600cgaacgtgg ggaasaga gggaagaa ggaaggaa gggaagga gggggtagg gggggggg	ttattgctca	aagagacatg	ggtggaagag	atgaaggtta	cgattggttg	attatgacac	1020	
Atgriggtote tacaggate gacattate tigtinggaag aggactatte gacaaggaga1140gggatgota gacaagta ggacagate gacagtate gacaageag etgggaca tattegaga1200gatgoggoca gcaaagtaa aaaaetgtat tatagtaa tigotgtata etaagocaa1200aaattagage tecaattaa tatateagi tattaceet tigoggigga aatacegoca1200agatgogtaa ggagaaata eegocaaga aatggtaa egittata tigtigaaa1300teeetaaa etaagaa tagacegaga taggetgag tigtigteca gittigaaa1500agagtocat attaagaa egiggacea eadeagga taggetgag ggaaaaace gittataegi1600gegatgoge dacaegiga eedacaegi eegocattag gegacaaace gittigteea1600egaacgtegig gagaaagaa egaaggace eegattag agetgegig ggaageegi1600egaacgige dacaegiga eegaacgaa egigeetgag gegeetgegig1800gegeetgeeg eetaegi eegaacgig egaacgig egigeetgag eegetgeegi1800gegeetgeeg eetaegi eegaetgee geactigtig gaaggeegi teggeetgeegi1800eegeetgeeg eetaegi eegeetgee geactigtig gaaggeegi teggeetgeegi1800eegeetgeeg eetaegi eegeetgee geactigtig gaaggeegi teggeetgeegi1800eegeetgeeg eetaegi eegeetgeegi eegeetgeegi eegeetgeegi eigeetgeegi eegeetgeegi eigeetgeegi eigeetgeegi1900eegeetgeegi eegeetgeegi eegeetgeegi eegeetgeegi eigeetgegiegi eigeetgeegiegiegi1900eetteeteet etteeteetiegiegiegiegiegiegiegiegiegiegiegiegiegi	atgrogtete tacaggatet gacattatta tigtiggag aggactatt genanggaga1140giggatgeta ggitaggggt gacgitace ganangengg etgiggangen tattinggan1200gatgeggeen genanetaa aanatigtat tataagtaan tgentigtaa tataeeen1260aaattagage teenattaa tatateegi tattaeeen taiggigtigta aataeegeen1320agatgegtaa ggaanata eegeateegi aattigtaan egitaatatt tigtaanat1380teegetaa tittigtaa ateageeta tittiaeee ataggeegaa ateggeaaa1400teeetaa ateaagaa tageegaga taggetigag tigtigtee aggitegea ate1500agagteedat attaaagaa tigaeegaga taggetigag tigtigtee aggitegea ate1500aggategee ateagaga eentaeee aggingaee eeegattag agetigaeg ggaaageegi1600gegatiggee gaaaagga gigaaggaee eeegattag agetigaegi gegegegegi1600gigatagegi eaegetege giaacaeee aeeeeege gigeegeegegi gigeegegeaa1740gigaaeggae eaegetege giaacaeee aeeeeeegegi gigeegegigeegegigi1860eeeetegegi eaegetege giaacaeee aeeeeeeegegigeegegigeegegigigi1980aeegeedagg titteeea teeaggeigi gigeegeegegigeegegigeegeigi1980aeeeetegi gigeegegig gigeegeegeegegigeegeegegigeegeege2200aeeetegit titteeea teegagatet geegegiga aggeegeege2200agaateatt tittgaagat ateteet teegigigigaga atggegigeigi ageegegie2200gigeegeaaaa agateetegi tigeegaaaate tigeegeaaa2400tetteeaaaa agateetegi tageetegi aggegigi aggeegeaaa2400tetteeaaaa agateetegi tageetegi aggegigi aggegigi aeegagaa2400tetteeaaaa agateetegi tageetete agaaggitigi aeegagaaa2400taaeegaaaa agateetig taaeagat tigeetigi aggegigi ageegaaa2400taaeegaaaa agateetegi taaeagat tigeete	ccggtgtggg	tttagatgac	aagggagacg	cattgggtca	acagtataga	accgtggatg	1080	
gggatgctaaggatggatggaacgttaagaacgttaatatagtaatittggaatittggaagatgcggccagcaacactaaaaactgtattatacgtaatgatgtgtatatatatacgtaatatacgtaaagatgcgtaaggaacaataccgcatcagaaattgtaacgttatatttgtaacatatacgcaatatatcocgttaattttgtaaaccgcatcagaaatggtagtgtggttgtattgtaacatittagatgcgtaggaacaataccgcatcagacaggtagtgtggttgtatittagatgccatattaagaatgtggaacaattttaacattggggaagaacatittgggatggccacacggaacacaaggagaccccatcaggaacagtitttittagacactaatcggaacacaacacggaacggaaggaagatitttittggaacggcccatcacgaaggaaggaagaggaaggaaggatitttittggaacggcccatcacggaagggaaggaaggaaggaaggagtitttittggacggcccatcacgaatacaggaagaggaaggaaggaaggatitttittggaacgacatataggaaggaaggaaggaaggaaggaggatitttitttittggaacgaaggaacgaaggagggaaggaagggaaggaaggatitttittggaacgaataacgaacgaaggaaggaagtitttitttittggaacgaataacgaacgaaggaaggaagtitttitttittggaacgaadtaaggaaggaaggaaggaaggaggaggaaggagtitttittggaacgaadtaaggacaataaggacagaatitttitttittggaacgaadttaggaacataaggacagaa<	gggatgctaaggatgcgacagaaagctaagaaagcagactagagacatatagtaateacticaagatgcggacagcaaactaaaaaactgtattataagtaatataagtaatataagtaatataactaa1300aaattagagattatatatattattaccattatgacgtgaa13001300agatgcgtaaggaagaaatacogcatcagaaatggaga14401300toccttataatcaagaatagacgaatatgacgaa1500agagtccatattaagaatgggatgcatatgacgaaacc1600ggatgcgaccatacggaacatacagtatttggggatggac1600gagacactaatcoggaacccatacggaacggaaggac1740gtgtagcggcacaccacacaccaccaccaccac1900gtgtagcggcacaccacacaccaccaccaccac1900gcogctgcggcataccacacacaccaccaccaccac1900gcogctacggcatacgcattaggacga1900gcogctacggcataccacacaccaccaccaccaccaccaccaccaccaccaccaccac	atgtggtctc	tacaggatct	gacattatta	ttgttggaag	aggactattt	gcaaagggaa	1140	
gatgeggecageaaaactaaaaactegtattataagtaatgegtgtgaaaacceaa1260aaattagagetteaattaattataceattataceetatgeggtgga1320agatgegtaaggagaaaatacegeetaacegeetaa1380teeretaattettaaateageetatettaaeeta1440teeretaaateageaatettaaeetatettaaeeta1500agagteetaateageaatettaaeetagegategeeta1500agageetaaateageaatettaaeeta1500agageactaateggaacetaateageaceta1600cgaacgtggegaaaagaacaaceetaaageactaa1740gtgtageggtcedeetaecedeetaege1800gegetegeecedeetaecedeetaege1800gegetegeecedeetaecedeetaege1800gegetegeecedeetaeteageetaeg1800gegetegeecedeetaeteageetaeg1980aeegaeteeateageetae1980aeegaeteetedeetaegggeaageetee2100aeegaeteetedeetaegggeaageetee2100aeegaeteetedeetaegggeaageaaa2200agaatatttedaaeagatedeetaegaa2400cedeetaatedaeagaatedaeadea2400cedeetaatedaeagaatedaeadea2400cedeetaatedaeagaatedaeadea2400cedeetaatedaeagaatedaeadea2400cedeetaatedaeagaatedaeadea2400cedeetaatedaeagaa	gatgroggeeageaaactaaaaactgtattataqgaatatacogta1260aattagagetetaattaatatateogttattacogta1320agatgogtaaggagaaaatcogeateaggaatteggegaa1380toogogtaaattttgtaaatcageeaa1440toottaaatcagaagatagageeaaa1500agatgogeeaatcagaagatagageeaaa1500agageeeaaatcagaagatgggatggee1600gogatggeeatcaggaaccodeaceea1600cogaactgoggaaacaeacodeaceea1740gtgtagogecatecoggageaacgaeg1800gogoegggaaagaaggaaagaagggaaagaeg1800gogoegcatecoggageaacgaeg1800gogoeggeaacgaeggeaacgaeg1800gogoeggeaacgaeggeaacgaeg1800codecogageaacgaeggeaacgaeg1800codecogageaacgaeggeaacgaeg1800codecocaatcageeggeaacgaeg1800codecocaatcageeggeaacgaeg1800codecocaatcageeggeaacgaeg1990taacgeeaggtttteceg1990taacgeeaggttteceggeaacgaeg2100accettaattaggegaattggetaeggeggatgaegggatogttgeaacteggegegaaagae2200ggaatcatttgactedgeggaagae2200agaatcatttgactedgeggaagae2300cottaactgaagaetagaagaeta2400 <t< td=""><td>gggatgctaa</td><td>ggtagagggt</td><td>gaacgttaca</td><td>gaaaagcagg</td><td>ctgggaagca</td><td>tatttgagaa</td><td>1200</td><td></td></t<>	gggatgctaa	ggtagagggt	gaacgttaca	gaaaagcagg	ctgggaagca	tatttgagaa	1200	
aattagage tteaatta ttaateagt tattaceeta tgeggtgtga aatacegea1320agatgegtaaggagaaata cegecaeg gaattgtaa egttatatt ttgttaaat1380tegegttaatttttgtaa ateageetatttttaeeta ataggeega ateggeeta1500agagteetaataaagaataggegtea1500agagteetaateageaet1500agagteetaateageaet1500agagteetaateageaet1500agagteetaateageaet1500agageaetaateageaetateageaetgtgaageggaacgtggggaaageaegtgaageegaacgtggggaaageaegtgaagegeeceategee1600geegeteegeeateegee1740gtgaagegeeceategee1800geegeteegeeateegee1800ecetteegegtaaceeeaceeegeeaaegeetaataggeegatgggegaagggeegteegeeateegeatggaaggeegeegeteegeeatageeggggaaggeegeegeegeteegeeatageeggggaaggeegeegeteegeeatageeggggaaggeegeegeteegeeatageeggggaaggeegeegeteegeeatageeggtggaaggeegeegeteegeeatageeggtgeaggeetgeegeteegeeatageeggggaaggeegeegeteegeeatageeggtgeaggeetgeegeteegeeatageeggtgeagegeegeegeteegeeatageeggtgeagegeegeegeteegeeatageeggtgeageeggeeggeeggeeggeeggeeggeeggeeggeeg	aasttagage tteaattta ttaateagt tattaceet tgeggtgga aatacegea1320agstgegtaaggagaaata cogeateagg aasttgtaacyttaatatt ttyttaaaattoegegttaatttttgtaaateageetaateeetaataagaatagaeegeatagggtgga tagggtgggegatgeetaateaaagaaagageetaateegaacetaaggaeetaateegaaceggegatgeeteaceecaagegegteetataegeecaagegegeataegeecaagegegeataegeecaagegegeataegeecaagegeecaataegeecaagegeecaataegeecaagegeecaataegeecaagegeecaataegeecaagegeecaata	gatgcggcca	gcaaaactaa	aaaactgtat	tataagtaaa	tgcatgtata	ctaaactcac	1260	
agatgogtaaggagaaaata cogcatcagaattgtaacigtataatt tigtaaa1380toogogtaatittigtaa atcagoogatittiacaca atagoocgaa atcogcaaa1440toccttataatocaaaagaa tagacogag taggigtag tigtigtoca gittiggaaca1500agagtocactattaaagaacgiggacocaaccacaca accacaca1500agagtocactattaagaacgiggacocaaccacaca1620aagacataatoggaacocaaaccacacaaaccacacaaccacaca1680cigaacgiggigaaagaaggigaaagagagigaagagaag1680gigtagogcicacgicgoggitaaccacacicacegicgog cicacag1800gigtagogcicacgicgog giaaccacacicacegicgog gicacag1800gigegedegogcicatcogcaticagoogg1800cicactacgicaticagocagtiggaagaga1990taacgocaggiggacgigggiggacgigga1990aacgactacataaggocagiggacgigga2040ticttoccatataggiggatiggtiggiggacgiggacgiggac2100accettaatticaacetiggigtagaagaagiggagaagaagigtagaaaatgigagaagaatgiggagaagaa2200agaatcatictitgageti titgaggadagiggagaagaa2200gigagaaaagiggigagaagaagiggagaagaa2300cicaactiggigtigaactagiggagaagaa2400ciggeaaaagiacttiggiggigaagaa2400ataageaaagiacttigtitaactit1500gigacataaagatcaticciggiggiga2500ciggaagaagiggaagaagaa <td>agatgogtaaggagaaatacogcatcaggaattgtaacgttatattigttaaattorgogttaatttttgtaaattggcogataggotgaacogagagtcactattaaagaatagacogagtagggttggtgttgttcaggaggaaagaggagtcactattaaagaacacaccacacgcaaaggggaaggcogg1600ggagtgggggaaaggagggaagggaaagcacaccacacgcactag1740gtgtagoggcacaccacacaccaccaacacgcogg1800ggagtgcogcatacgtagggaaggagaag1800ggagtgcogcatacgtagggaaggagaag1800ggagtgcogcatacgtagggaaggagag1800ggegtgogcatacgcoggcatataggggaaggagaggaggggaaagcacacgcoggcatagggg1800cctttcgctattaggcoggcacactaggggaaggagaggagggggcacgcogggcataggcog1800cctttcgctattaggcoggggaaggagg1920taacgcoaggtttaccagtaggotgcog1800acgcatcatattaggcoggggaaggagg1920taacgcoaggtttaccagtaggotgcog2100accattatttocaattogggaaggagg2220agaatattttocaattogggaagaag2400attaggaacaagtaggattocaattog2400cggaagaacaagtaggattocaattog2500ggaagaacaagtaggattaggagga2400catacgtaaattaccatagatcat2400catacgtaaattaccatagatcatt2400cata</td> <td>aaattagagc</td> <td>ttcaatttaa</td> <td>ttatatcagt</td> <td>tattacccta</td> <td>tgcggtgtga</td> <td>aataccgcac</td> <td>1320</td> <td></td>	agatgogtaaggagaaatacogcatcaggaattgtaacgttatattigttaaattorgogttaatttttgtaaattggcogataggotgaacogagagtcactattaaagaatagacogagtagggttggtgttgttcaggaggaaagaggagtcactattaaagaacacaccacacgcaaaggggaaggcogg1600ggagtgggggaaaggagggaagggaaagcacaccacacgcactag1740gtgtagoggcacaccacacaccaccaacacgcogg1800ggagtgcogcatacgtagggaaggagaag1800ggagtgcogcatacgtagggaaggagaag1800ggagtgcogcatacgtagggaaggagag1800ggegtgogcatacgcoggcatataggggaaggagaggaggggaaagcacacgcoggcatagggg1800cctttcgctattaggcoggcacactaggggaaggagaggagggggcacgcogggcataggcog1800cctttcgctattaggcoggggaaggagg1920taacgcoaggtttaccagtaggotgcog1800acgcatcatattaggcoggggaaggagg1920taacgcoaggtttaccagtaggotgcog2100accattatttocaattogggaaggagg2220agaatattttocaattogggaagaag2400attaggaacaagtaggattocaattog2400cggaagaacaagtaggattocaattog2500ggaagaacaagtaggattaggagga2400catacgtaaattaccatagatcat2400catacgtaaattaccatagatcatt2400cata	aaattagagc	ttcaatttaa	ttatatcagt	tattacccta	tgcggtgtga	aataccgcac	1320	
tecgegttaaa ttittigtaa atcagetcat tittiaacca ataggecgaa atcggeaaa1440tececttataa atcaaagaa tagaecgaga tagggttag tgitgtecca gittiggaaca1500agagtecact attaaagaac giggactea acgteaaagg gegaaaaace gietategg1560gegatggee actaegtgaa ceateacet aateaagtit titggggteg gggaagaegg1680cgaacgtgge gagaaagga gggaagaag egaaaggae gggegetagg gegetgag1680gegegtegge ceateegee gitaaceacea caceegee gettaatgee cegetaegg1860geegetege ceateegee titaeggetg geaacgtg ggaaggag deggeggt atggtegg1860cectteget attaegeeag etggegaaag ggggatget tegetagg1920taacgeeagg gitticecag teegeggt ggeececee etggeggeag taggtege1980accetteat tagggegaa tigggtaeg ggeececee etggeggeag taggeegt2040tetttee geeaggegg attateg geeaggegg tegetgege2220agaatette titgaaggat etggtegat gittitete gtittaeg2280tettaaaetg a ggegaatgit etgetgaaga atggegaaggaag2400ataggeaaaa agateetteg tacaaagt tittaeatte titgggtgtaa gggaagaa2400ataggeaaa agateetteg tacaaagt etggegga atggegt atggegaag2500taggetataa eaggagaat tegataag atgggga atgtga taggagaaga2400atageaaaa agateetteg tacaaagt tittaeatte tigggttgaa gggaagaa2400attageaaaa agateetteg tacaaagt tittaeatte tigggtgaaggaagaa2500iggatataa agaggaatti eeataaga atgegga attigga tegtaga2500iggatataa agaggaatti eeataaga teggagga attigga tagegaagaa2400attageeaaa agateetteg tacaagat tittaeatte tiggtegga agaggaagaa2500iggacataa agateetteg tactaaga atatgegg attittee attiggegg2500iggacataa agaggaatti eeataaga atatgegg attittiga agagaaga2600 <td>tegegttaaa ttittgttaa atcagetcat ttittaacca ataggecgaa ateggecaaa1440teeettataa ateaaaagaa tagaeegaga tagggttgag tgttgttee gttggaaca1500agagteetta ataaagaa gtggaetee acgeetaaagg gegaaaaaee gteetateagg1560gegatggee actaegtgaa ecateaceet aateaagtt ttitggggte gggaaageegg1680egaaegtgge gagaaaggaa gggaagaaag egaaagaee gggegetagg eegetagge1800gegegtegge cateegeeg gtaaceace acceegeeg gettaatgeg eegetagg1800gegegtegge cattegea tteaggetg gegaaagae gggegetagg eegetagg1800geegetagee cattagee eegeaaag egggatge tgeaaggea teggtegg1800cectetteget attaegeeag etggegaaag ggggatge tgeaaggeg taggtegg1800aaegeetaa taggegag attagget gegeetagg1800aegaetaat teggeag etggegaaag ggggatge tgeaaggeg taggtegg1920taaegeetag taggegaa ttgggtaeg ggeeetee egaggtega taggtega1980aegaeteat ataggegaa ttgggtaeg ggeeetee egaggtega tggeegtata1980aecetteat tecaaaetg gegteaagga teceggat ggeegtag2100aecetteat tecaaaetg gegteaagg ateceggat ggeegtag gaecaegee2200agaateate tttgaaggat actateet tecaatttg atgetegt aegtgaaaa2340eggeaaaa agateeteg taacaagat tittaeatte tggtgtgaa gegaaaaa2400attaggeaaa aggegattt eeataete aggegaatt eeataete aggegaaaa2400attageeaaa aggeatte eaaagtaf tataetee tetteggeg age2500ittagaetaa aggegatte eataete teggeg agtettet tetteetee2500ittagaetaa aggegatte eaaagat tettaetee teggagaaga2400attageeaaa agateeteg atectaaga atatgegg agtettet atectaetee2500ittagaetaa agegaatte cataatea eaagtag attegga agtetteg agegaagaa2400<tr<< td=""><td>agatgcgtaa</td><td>ggagaaaata</td><td>ccgcatcagg</td><td>aaattgtaaa</td><td>cgttaatatt</td><td>ttgttaaaat</td><td>1380</td><td></td></tr<<></td>	tegegttaaa ttittgttaa atcagetcat ttittaacca ataggecgaa ateggecaaa1440teeettataa ateaaaagaa tagaeegaga tagggttgag tgttgttee gttggaaca1500agagteetta ataaagaa gtggaetee acgeetaaagg gegaaaaaee gteetateagg1560gegatggee actaegtgaa ecateaceet aateaagtt ttitggggte gggaaageegg1680egaaegtgge gagaaaggaa gggaagaaag egaaagaee gggegetagg eegetagge1800gegegtegge cateegeeg gtaaceace acceegeeg gettaatgeg eegetagg1800gegegtegge cattegea tteaggetg gegaaagae gggegetagg eegetagg1800geegetagee cattagee eegeaaag egggatge tgeaaggea teggtegg1800cectetteget attaegeeag etggegaaag ggggatge tgeaaggeg taggtegg1800aaegeetaa taggegag attagget gegeetagg1800aegaetaat teggeag etggegaaag ggggatge tgeaaggeg taggtegg1920taaegeetag taggegaa ttgggtaeg ggeeetee egaggtega taggtega1980aegaeteat ataggegaa ttgggtaeg ggeeetee egaggtega tggeegtata1980aecetteat tecaaaetg gegteaagga teceggat ggeegtag2100aecetteat tecaaaetg gegteaagg ateceggat ggeegtag gaecaegee2200agaateate tttgaaggat actateet tecaatttg atgetegt aegtgaaaa2340eggeaaaa agateeteg taacaagat tittaeatte tggtgtgaa gegaaaaa2400attaggeaaa aggegattt eeataete aggegaatt eeataete aggegaaaa2400attageeaaa aggeatte eaaagtaf tataetee tetteggeg age2500ittagaetaa aggegatte eataete teggeg agtettet tetteetee2500ittagaetaa aggegatte eaaagat tettaetee teggagaaga2400attageeaaa agateeteg atectaaga atatgegg agtettet atectaetee2500ittagaetaa agegaatte cataatea eaagtag attegga agtetteg agegaagaa2400 <tr<< td=""><td>agatgcgtaa</td><td>ggagaaaata</td><td>ccgcatcagg</td><td>aaattgtaaa</td><td>cgttaatatt</td><td>ttgttaaaat</td><td>1380</td><td></td></tr<<>	agatgcgtaa	ggagaaaata	ccgcatcagg	aaattgtaaa	cgttaatatt	ttgttaaaat	1380	
teccttataa atcaaagaa tagaccgaga tagggttgag tgttgtteca gtttggacaa agagtecaet attaagaac gtggacteca acgteaaagg gegaaaaace gtetateagg gegatggece actacgtgaa ceateacet aateaagttt titggggteg aggtgeegta aageateaa teggaacet aaaggagee ecegattag agettgaegg ggaaageegg ife8 eggaegtgge gagaaagga gggagaaag eggaaggae ggegetagg egettaageegg gegetgeg egagaagga gggagaaag egaaggae ggegetagg egettaatgee egettaagee gegetegee ecateegee gtaaceace eceegeeg ettaatgee eegetag gegetegee ecateegee gtaaceace eceegeeg ettaatgee eggeeggeg ife8 ecetteege attaegeeag ettaegeeg egaactgtt ggaaggeega teggtgeegg ife8 ecetteege gtttteeeag teaegaegt gtaaaeaga eggeaggeg teggeeggaa aegaeteaet ataggegaa teggtaeeg geeeeet eggaggeege taggeeggaat ife8 ecetteet eegeag gtttteeeag teaegaegt gtaaaggee ggeegetag eggeegata ife8 ecetteet eegeagegaa teggtaeeg geeeeet eggaggeege taggeegetaa ife8 ecetteet eegeagega teggegaag ggeegeegeat ife8 ecetteet eegeagega teggegaag ggeegegeat ife8 ecetteet eegeagegea teggegaag eggeegetag ife8 ecetteet eegeagegea teggegaag ggeegegea ife8 ecetteet eegeagegea teggegaag geegegeat ife8 ecetteet eegeagegea teggegaag eggeegea ife8 ecetteet eegeagegea teggegaag geegegea ife8 ecetteet eegeagegea teggegaag eggeegea ife8 ecetteet eegeagegea teggegaag ife8 ecetteet eegeagegea ife8 ife8 ecetteet eegeagegegeageage ife8 ife8 ife8 ife8 ife8 ife8 ife8 ife	teccttataa atcaaaagaa tagaccgaga tagggttgag tyttytteca gtttggaaca1500agagtocact attaaagaac gtggacteca acgtoaaagg gegaaaaace gtetateagg1560gegatggee actacgtgaa ceateacet aatcaagtt tttggggte aggtgeegta1620aageactaaa teggaaceet aaagggage eecegattag agettgacgg ggaaageegg1680cgaacgtgge gagaaaggag gggaagaaag eggacgeag ggegetagg geegteagg1800gegegtegge cateegee gtaaccace acceegee gettaatgee ceegtagg1800geegetegge catteegea tteaggetg geaatgeg ggeaggag ggeegetagg geegetagg1800cecttedet ataegeeag etggegaaag ggggatgge tegaaggeg teggtgegg1800cecttedet ataegeeag etggegaaag ggggatgge tegaaggeg taggtgeg1920taacgeeag gttteccag teegegat ggacceet gaggegad geegetgag eegegeat1800aegacteat ataggegga ttgggtaceg geececee etggaged taggegat2040tetttecca ttaggtega ttgggtaceg geececee etggage gagegetagg2220aggateatte tetgaaggat efgtegaagg attgegegteag gagegaagae2240gegegaaaa eagtagat tigaetete gtegaagga atggegatg agegegaa2400tettaacegta actagaatt ggettag gggaagga atggegtag agegaagaa2400cegeagaaa agateetteg taacaagat tttacette teggtgtga agegaagaa2400attageaaaa agateetteg taacaagat tttacette teggtgtga agegaagaa2400taggetataa ageggaatt eetataet eetatteg atgetga etgaagaa2400attageaaaa agateetteg taacaagat tetagga attagga etgaagaa2400taggetataa ageggaatt eetataet eetatte teggagga agtagga aggaagaaga2400taggetataa ageggaatt eetataet eetatteg atetaga eegaagaaga2400acteegeag gagaatteg atetaga attaggeag attaggaagaa attaggaagaaga2400taggetataa agegga	tcgcgttaaa	tttttgttaa	atcagctcat	tttttaacca	ataggccgaa	atcggcaaaa	1440	
agagtccact attaaagaac gtggactca acgtcaaagg gcgaaaaacc gtctatcagg 1560 gcgatggcc actacgtga ccatcaccct aatcaagtt tttggggtcg aggtgccgta 1620 aagcactaa tcggaacct aaagggagc cccgattag agctgacg gggacgcagg 1680 cgaacgtgg gagaaagga gggaagaag cgaaggag gggcgctagg gcgctggca 1740 gtgtgacggt cacgctgcg gtaaccacca cacccgccg gcttaatgcg ccgtacagg 1800 gcgcgtcgc ccattcgca ttcaggctg gcaactgtg ggaaggag tgggtgg 1590 taacgccag gtttcccag tcaggctg gtaaccacc gaccggc gctaagg cgcggtaa 1980 acgactaat ataggcga ttgggaacg gggcactgg ggccatga ggccagtag 1980 acgactact ataggcga ttgggtacg ggcccccct cgaggtcga tgggcatta 29040 tcttcccat attaggtt gcaaggtag aggaagtag aggaggtag ggcgtag ggcgcata 2100 acccttatt tccaaactg gcgcaaggg atccggta ggacgtag ggcgcatag 2220 agaatcatt ttgaagga acttatct tccaatttg atggtggtg aggagatg 2280 taaactgtat actagaatt ggacttcg tgggaaga aggagatg aggagatg actggtga 2280 taaactgtat actagaatt ggacttcg tagggaaga aggagatg aggagatg actgggaa aggatcata ataggaat tggatcteg taggaaga atgggaaga atgggaagaa 2400 attagcaaa agatccttg taacaagt tttacattt tggtgtgaa gggaagaag 2460 tgggcataa aggcgaatt ccatacat cagatttg tattattt tttcctccc 2520 acgtcoggg gaatctggt attatcgt attatag aaattggg agtttttc atgtgaga 2640 tactgactge agcaaat aggcagat tataggaat atggagt tgtgaagga 2700 ggaacgtaa aggaattgg attatacg attatagg atttgatg gtgaaaga 2700 ggaacgtaa aggaattgg attatacg attatagg atttgatg tgtacagag 270	agagtccactattaagaacgtggattccaagtggatccaactacgtga1660gcgatggccactacgtgaaccatcacctatcaagtttttggggtcggagtgccgta1620aagcactaatcggaacctaaagggagccccgattggggaaggggg1680cgaacgtggcgagaaggagaggggaggagaaggggaggggggggggggggggggggggggggggggggggggg	tcccttataa	atcaaaagaa	tagaccgaga	tagggttgag	tgttgttcca	gtttggaaca	1500	
gegatggecc actacgtga ccatcacect aatcaagttt tttggggteg aggtgecgta1620aageactaa teggaacect aaagggage cecegattag agettgaegg ggaaageegg1680cgaacgtgge gagaaagga gggaagaag eggaaggee gggegetagg egegtgea1740gtgtageggt cacgetgeeg etaaccace caceegeeg gettaatgeg eegetag1800geegetegee ceattegeea tteaggetge geaactgttg ggaaggeea teggteegg1860cctetteget attaegeeag etgegaaag ggggatgee tgeaaggeega teggteegg1860cctetteget attaegeeag etgegaaag ggggatgee tgeaaggeeg tageegtaat1980acgacecag gtttteccag teaggaegt ggaaagteag eggeegtaat1980acgactaat tagggegaa ttgggtaeeg ggeeceeet eggagteega teggeegtaat2040tettteceat attaggtte gegeaaggg atcetggta ggtgeteg ggaeeggaat2100accetteatt teteaaaetg gegteaaggg atcetggta ggtgeteag ggaeeggee2220agaateatte tttgaaggat acttateet teceaatttg atgetegt aegtgeage2340cggeagaaa aggaegtteg tageaggaa atgaggaag aggaagaag2400attageaaaa agateetteg taacaagat tttaette tggtgtaa gggaagaa2400attageaaaa agateetteg taacaagat tttaette teggtgtaa gggaagaag2400attageaaa agateetteg taacaagat tttaette teggtgtaa gggaagaa2400attageaaa agateetteg taacaagat tttaette tegtgttgaa gggaagaa2550ggaacattaa tttteaaegt attetaeg aaattegeg agtttttte attetaggea2580gaacattaa tttteaaeg aattegeeg agaatte attgtageega2700ggaacgetaa gagataette agaatta taggeaga attaggeegaata2700ggaacgetaa gagataette agaattag atttgtage tgtaceaga2700ggaacgetaa gagataette cagaatta teteggeega aaaatteet ttgtaaagta2700ggaacgetaa agacatte cagaatta teceggeega aaaatteet tegra	gogatggocc actacgtga ccatcacct atcaagtt tttggggtog aggtgocgt1620aaggactaa toggaacct aagggago cccgattag agttggogg ggaagggg1680cgaacgtgg gagaaggaa gggaagaag cgaaggag gggogtagg gogtggaa1740gtgtagogg cactgtgoe gtaaccaca caccogoog gcttaatgog cogtacag1800gegogtogg coattogoca ttcaggotg gaacatgtg ggaaggag toggtgogg1800cctcttogt attaccaca ttcagotg goacccgt gcaacgg gtaggggg toggtogg1920taacgccag gttttcccag tcacgacgt gtaaacaca ggocagtga gogocatgg cogocg1980acgactact ataggogaa ttgggtacg gocccccct cgaggtoga toggcatta2040tctttcccat attaggtt gcaaggg atcggg at cgggtad ggtgotg tgaaggac2100accttat tccaaactg gogtactgg atcctggt ggaaggad agggggttag tgaaggca2220agaatatte ttgaagat actatect tccaatttg attgetgt acgtggaa2340cggocgaagaa cagtagat tggcttag gggaagaa atgggatg aggaagaa2400attagcaaa agatcett gaacatg acttage actagga attggg aggaagaa2520acgtcaga acagtagat ttgactete cagattt ttacattt tggtgtag gggaagaa2520acgtocgcg gaatctggt attatece tccaatttg tattatt ttecttocc2520acgtcata acggaatt tgacttog tacgagaa atgggatg aggaagaa2640ttaggactat aggaattt catatact cagatttg tattatt ttecttocc2520acgtcogg gaatctgg attataga aattgogg agttttttc atgtaggaa2640tactgactga aggaatat agaacat atgtagt ttatagge2640tactgactga aggaatt tatagge actggg agaatat2640tactgactga aggaattt catatat atgtagg attgtag tgtacgata2640tactgactga aggaattt aggactgat tatagge gaattgg tgtacgata2700ggaacgtaat taggaatt tataggegg aaattggg tataccaata2700ggaacgtaatt caga	agagtccact	attaaagaac	gtggactcca	acgtcaaagg	gcgaaaaacc	gtctatcagg	1560	
aagaactaaa teggaacct aaagggagee ceegattag agettgaegg ggaaageegg1680cgaacgtgge gagaaaggaa gggaagaaag eggaaggage gggegetagg gegetgaea1740gtgtageggt caegetgee gtaaccacca caecegeeg gettaatgee oegetaeagg1800geogetegee caetegee atteageetge geaactgttg ggaagggega teggtgegg1860cettteget attaegeeag etggegaagg ggggatgee tgeaaggeega ttaagttggg1920taaegeeag gttteeceag teagegeag ggggatgee tgeaaggeega taagttggg1980aegaeteaet ataggeegaa ttgggtaeeg ggeeeeceet egggegaag eggegataa2040tetteeea attagatte geeaaggg atceggtgat ggetttateg tttaatgea2100accetteat teteaactg gegteaaggg atceggta ggegegteag2220agaateatte tttgaaggat acttateet teeeattg attgeetgagea2340cggeagaaa caagtagat gggtaagat ttggtgaaggaatag gggaaggaa	aagcactaa teggaacet aaagggagee eccepttag agettgaegg gaaageegg1680cgaacgtgge gagaaaggaa gggaagaaag egaaaggae gggegetagg geegtegaa1740gtgtageggt ecceptegee gtaaceacea eaceegeeg gettaatgeg eegsteagg1800geegteege ceattegee gtaaceacea eaceegeeg gettaatgeg eegsteagg1800cectetteget attaegeeag etgegeagg ggggatge tgeaaggeega teggteggg1920taaegeeagg gtttteeeag eegseagt gtaaaaegae ggeeagtgae egesegtaa1980acgacteaet ataggeegaa teggetaeeg ggeeceeet egagetegae teggeeatta2040tettteeea attagatte geeaageeg gaaagteea gaaagteet tagaegaat2100accetteatt teteaaaetg geeteaagga ateetggta ggtgetaet ggesteaeg2220aggateatte titggagat eegsteat ggagaagaa atggggaagaa atggggaagaa2340cggeagaaa eaagtagat tiggeteeteg tageagaagaa atggggaagaa2400taaaeetgaa aggecatteg taeetteg teetegaagaa2400cggeagaaa eaagtagat tiggeteeteg tageagaa atgagatatg ageegagaa2400cggeagaaa eaagtagat tiggeteteg tagggaagaa atgaggatg aggaagaa2400taaaeetgaa aggeegatteg taataeete eagatteg tateteattettee2520acgteegeg gaatetgeg tatataeet eagatteg atettagat geegagaa2640taetgeaega aggeagatte taagaeag aattgegg agtttttee atetaggee2640taetgaeege eegsaatta aggeatget tataggeegega aaattgeege aaaatteat tegaagat2700ggaacgee aggaaaattee agaatete taggaeag atettggeegega aaaatteet tegaacate2700ggaacgee aggaacatte eegaatte teetgeegga aaaatteege tegaacate2700ggaacgee aaggaaattee eegaatta teetgeegga aaaatteet tegaacate2700ggaacaetaa ageceatee eegaatte teetgeegga aaaatteet tegaacate2800taacaeaaaaaa	gcgatggccc	actacgtgaa	ccatcaccct	aatcaagttt	tttggggtcg	aggtgccgta	1620	
cgaacgtggc gagaaagga gggaagaaag cgaaaggac gggcgctagg gcgctggca1740gtgtagcggt cacgctgcge gtaaccacca cacccocgce gcttaatgcg ccgtacagg1800gcgcgtcgcg ccattcgca ttcaggctg goaactgttg ggaagggcga tcggtgcgg1860cctcttcgct attacgccag ctggcgaaag ggggatgtgc tgcaaggcga ttaagttggg1920taacgccagg gttttcccag tcacgacgt gtaaacaac ggccagtgag cgcgcgtaa1980acgactcact ataggcgaa ttgggtaccg ggccccccc cgaggtcgac tggccattaa2040tctttcccat attaggttg ggctaaggg atctggt ggtatggg tgggggtgg ggcggggaagg2100accettcatt tctcaaactg gcgtcaaggg atcctggta ggtggtcag gagccagcg2220agaatcattc tttgaagga acttacctt tccaatttg attgtcgt acgtgggaa2340cggcagaaa a caagtagatt ggcattgg tgggaagaa atgaggatg agcgagaa2400attagcaaaa agatcctteg taacaagat tttactt tggtgtgaa gggaagaa2400attagcaaa agatctgg taactaac cagatttgt tattattt tttccttccc2520acgtccgcg ggaattt ccatacat cagatttgt attgtagt attgtagg2580gtacattaa ttttcaagt attcataag aaattgcgg agttttttc attgtagtag2640tactgactgc acgcaaatt aggcagtat tatagcg agaatta gtatcatt attgtagta2700ggaacgctaa gagtacttc agaatgt ttataggcag attggc tgtaccgata2700ggaacgcaaa aggcaaatt aggcagtat tataggcag atttgtagg tgtaccgata2700ggaacgcaa aggcaaatt ggatagt tcctggcga aaaattcat ttgtaactt2700ggaacgcaa aggcaaatta ccccaaaatta tcctggcgg aaaaattcat ttgtaactt2700ggaacgcaaagaa aggcaatta cccaaaatta tcctggcgga aaaaattcat ttgtaactt2700gaacgacgaaagaa aggcaaatat tcctggcgga aaaaattcat ttgtaactt2700gaacgacaaagaa agccaatata tagcagtat tcccggcgga aaaaattcat ttgtaacgata2700<	cgaacgtgg gagaaagga gggaagaaag cgaaaggag gggcgctagg gcgctggcaa1740gtgtageggt cacgetgeg gtaaccacca caccegeeg gettaatgeg cegetacagg1800gegegtegeg cattegee atteaggetg gcaactgttg ggaagggega teggtgegg1860cetetteget attacgeeag etggegaag ggggatgtge tgeaaggega teagtgegg1920taacgeeagg gtttteeeag teacgaegt gtaaaacgae ggeeagtgag egeegtaa1980aegaeteaet ataggeegaa ttgggtaceg ggeeeetee etgaggeeget taagtegg2040tettteeeat attagatte geeaagge atceggta ggeegetga gageegee2100accetteat teteaaaetg gegteaagg atcetggta ggtttateg tttattet2160ggttettata geategttt ggaeteet gteecaatta ggeegetgag aggeegee2220agaateatte tttgaaggat acttateet teceaattg atgeggaag atgeggtga atgegtgag2340cggeagaaaa caagtagat tggeetteg taggaagaa atgaggattg agegaagaa2400attageeaaa agateetteg taacaagat tttacett teggtgtaa gggaagaag2400atgagetate ageggaatt ceatagatt tgacetteg tagtetate ttteettee2520acgetegeg gaatetggt attataege atetagge attegge gage ga	aagcactaaa	tcggaaccct	aaagggagcc	cccgatttag	agcttgacgg	ggaaagccgg	1680	
gtgtageggt Cacgetgege gtaaccace caccegeeg gettaatgeg eegetacagg1800gegegtegeg ceattegeea tteaggetge geaactgttg ggaaggega teggtgegg1860cetteteget attaegeeag etggegaaag ggggatgtge tgeaaggega ttaagttggg1920taacgeeagg gttteeeag teaegaegt gtaaaacgae ggeeagtgag eegeegtaa1980acgacteat ataggegaa ttgggtaceg ggeeeeceet egaggtega tggeeagtaa2040tetteeeat attagatte geeaageeat gaaagteea gaaagteet tagaegaat2100accetteatt teteaactg geegteaaggg ateetggtat ggtttateg ttttattet2160ggttettata geategttt ggaetteet gteeeatta ggeeggteag gageeageeg2220agaateatte tttgaaggat acttateett teeaatttg attgetgtat acgttggae2340cegecagaaaa eagtagatt tgaetteet gteeeatta ggeeggteag gageageage2400attageaaaa agateetteg taacaagat ttteet teeaatttg attgetgta acttagae2400attageaaaa agateetteg taacaagat ttteeatte teggegttag gggaaagata2460tgagetatae ageggaatt ceatacae eagtagatt ttteeatte tggtgttgaa gggaaagata2460tgagetatae ageggaatte eattaege atetagata atgtatet atettgeege2500gtacattaa ttteeage attetaag aaattgegg agttttte atgtagata2640tactgaetge acgeaaatt aggeatgtat tataggeatg attegaetga2700ggaacgetaa gagtaette agaategtat teetggegg aaaaattee tttgaaaett2700ggaacgetaa gagtaacte agaatgtta teetggegg attaggee gtacegata2700ggaacgetaa gagtaactte agaategta teetggegga aaaaatteet ttgaaaett2700ggaacgetaa gagtaacte agaatagta teetggeege aaaaatteet ttgaaaett2700ggaacgetaa gagtaacte agaatata teetggeege aaaaatteet ttgaaatt2700gaaacgetaa gagtaactte agaatata teetgeege aaaaaatteet ttgaaaatt2700 </td <td>gtgtgtgggt cacgetgge gtaaccace caccegeeg gettaatgeg cegetaacg1800gegegtegeg cattegea tteaggetge geaactgttg ggaagggega teggtgegg1860cetetteget attaegeeag etggegaaag ggggatgte tgeaaggega ttaagttggg1920taacgeeagg gttteeeag teggegaaag ggggatgte ggeeagtgag egegegtaat1980aegaeteat ataggegga ttgggtaeeg ggeeceeet egaggtegae tggeeata2040tetteeeat attagatte geeaaggeg ateetggat ggaaggteg tgaeggat2100accetteat teteaactg gegteaaggg ateetggat ggtttateg tttattet2160ggttettat geaetgettt ggaetteet gteeeata ggeggteag gageeagee2220agaateatte tttgaaggat acttateet teeaattg geggtteag gageagteg2340ceggeagaaa eagtagat tgaetteet gteeatat gaaggattg agegagaa2340ceggeagaaa agateette gtaecagat ttaetteet teggeggaa atggggaagata2460ttaggetatae ageggaatt ceatateet eagttgga atettagee2520acgteegeg gaatetggt attetateg atetagat atgtatet atettegee2520acgteegeg gaatetggt atetaagaatt ttaetge atetagat atgtatet2580gtacatttaa tttecaagt attetaag aaatteggg agttttte atgtaggatg2640taactgaeag aggataett aggetagt tetaggeag atettgget gtaecagat2700ggaacgetaa gagtaactte agaategt teteggega aaaatteta ttgtaacett2760taaaaaaaaa agceaatate eccaaaata taggege etceatata tagtaacta tetagaaca2800</td> <td>cgaacgtggc</td> <td>gagaaaggaa</td> <td>gggaagaaag</td> <td>cgaaaggagc</td> <td>gggcgctagg</td> <td>gcgctggcaa</td> <td>1740</td> <td></td>	gtgtgtgggt cacgetgge gtaaccace caccegeeg gettaatgeg cegetaacg1800gegegtegeg cattegea tteaggetge geaactgttg ggaagggega teggtgegg1860cetetteget attaegeeag etggegaaag ggggatgte tgeaaggega ttaagttggg1920taacgeeagg gttteeeag teggegaaag ggggatgte ggeeagtgag egegegtaat1980aegaeteat ataggegga ttgggtaeeg ggeeceeet egaggtegae tggeeata2040tetteeeat attagatte geeaaggeg ateetggat ggaaggteg tgaeggat2100accetteat teteaactg gegteaaggg ateetggat ggtttateg tttattet2160ggttettat geaetgettt ggaetteet gteeeata ggeggteag gageeagee2220agaateatte tttgaaggat acttateet teeaattg geggtteag gageagteg2340ceggeagaaa eagtagat tgaetteet gteeatat gaaggattg agegagaa2340ceggeagaaa agateette gtaecagat ttaetteet teggeggaa atggggaagata2460ttaggetatae ageggaatt ceatateet eagttgga atettagee2520acgteegeg gaatetggt attetateg atetagat atgtatet atettegee2520acgteegeg gaatetggt atetaagaatt ttaetge atetagat atgtatet2580gtacatttaa tttecaagt attetaag aaatteggg agttttte atgtaggatg2640taactgaeag aggataett aggetagt tetaggeag atettgget gtaecagat2700ggaacgetaa gagtaactte agaategt teteggega aaaatteta ttgtaacett2760taaaaaaaaa agceaatate eccaaaata taggege etceatata tagtaacta tetagaaca2800	cgaacgtggc	gagaaaggaa	gggaagaaag	cgaaaggagc	gggcgctagg	gcgctggcaa	1740	
gegegtegeg ceattegeea tteaggetge geaactgttg ggaaggega teggtegeg1860cetetteget attaegeeag etggegaaag ggggatgtge tgeaaggega ttaagttggg1920taaegeeagg gttteeeag teaegaegt gtaaaaegae ggeeagtgag eegegetaa1980acgaeteact atagggegaa ttgggtaceg ggeeceeet egaggtegae tggeeattaa2040tettteeeat attagatte geeaageeat gaaagteaa gaaaggtett tagaegaat2100accetteatt teteaaaetg gegteaaggg ateetggtat ggtttateg tttattet2160ggttettaa geategttt ggaetteet gtteeeatta ggeeggteag gageeagege2220agaateatte tttgaaggat acttateett teeaatttg attgetgtt aegttggae2340ceggeagaaaa eagaeatt ggaetteg tgagaagaa atgaggatg agegagaaa2460tgagetatae ageggaatt teeateet eagatttg tatetattt tteeeteetee2520acgteetegg gaateetgg ateetagata atgtggtgaaggaagaa2460tgagetatea ageggaatt ceataeat eagaatttgt ateetaatt tteeetee2520acgteetegg gaateetgg ateetagata atgtgtga ggeaagata2640taatgeetag agegaatt eetaagaat tgagegg agettttte atgtagaga2640taatgaeaaaa agateette agaattgta tataggeag atteggeg aaaaatteat ttgtaacgata2700ggaacgetaa gagtaactte agaatatta teetgegga aaaaatteat ttgtaaaett2700ggaacgetaa gagtaactte agaatatta teetgegga aaaaatteat ttgtaaaett2700ggaacgetaa gagtaactte agaatagta teetggega aaaaatteat ttgtaaaett2700ggaacgetaa agateatte agaatagta teetgeegga aaaaatteat ttgtaaaett2700ggaacgetaa gagtaactte agaatagta teetgeegga aaaaatteat tegtaaaatt2700ggaacgetaa agataatte ceetaatte teetgeegga aaaaatteat tegtaaaatt2700taaaaaaaaa ageeeaataa ageeeaatta teetgeegega aaaaatteat tegtaaaatt2700 <t< td=""><td>gegegtegeg ceattegee itteaggetge geaactgittig ggaaggega teggtegeg1860cetteteget attaegeeag etggegaaag ggggatgige tigeaaggega ttaagtegg1920taacgeeagg gittteeeag teaegaegit gtaaaaegae ggeeagtega etggeeatta1980aegaeteact atagggegaa tigggtaeeg ggeeeeeee egaggtegae tiggeeatta2040tetteeeat attaggeegaa tigggtaeeg ggeeeeeee egaggtegae tiggeeatta2100aeceetteatt etceaaaetig geegteeaggg ateetggtat ggitttateg tittattet2160ggttettata geategttti ggaetteet gtteeeatta ggeggteeag gageeageeg2220agaateatte tittgaaggat actateett teceaattig atggtgataeg gageaggae2340ceageagaaa eaagtagatt tiggetetteg gggaagagaa atgaggattig aeggagaaga2400attaegeaaa agateetteg taacaagatt titteetteg titteettet tettee2520acgeetegg gaatetigt atataetee eaggaagattig tatettatette2520acgeetegg gaatetigt atataetee eaggaagattig titteettet2580gitaeatae ageggaatti ceataetee eagegg agettigt tatetaatti titteettee2520acgeetegg gaatetigt atataetee atetagaat atggtata atgtagget geetegge2580gitaeattaa ageeggaatti ceataege atetagaat atgttatett atetiggee2580gitaeattaa titteeagg ateteag aaatgeegg agettittee atgtagatga2640tateggeege acceaaatta ageeatgatt tataggeegg attaggee titteeagata2700ggaacgeeaa gagtaactie agaatgeegg attaggeege aaaaattee2700ggaacgeeaa ageeatae eeceaaatta teetgeegga aaaaattee titgaaaett2820taaaaaaaa ageeaatae eeceaaatta taaggeege etceattat aaeaaaat2820taaaaaaaa ageeaatae eeceaaatta taaggeeggee etceattata aaeaaaatee2820taaaaaaaaa ageeaatae eeceaaatta taagaage etceeaaaat2820&lt;</td><td>gtgtagcggt</td><td>cacgctgcgc</td><td>gtaaccacca</td><td>caccegeege</td><td>gcttaatgcg</td><td>ccgctacagg</td><td>1800</td><td></td></t<>	gegegtegeg ceattegee itteaggetge geaactgittig ggaaggega teggtegeg1860cetteteget attaegeeag etggegaaag ggggatgige tigeaaggega ttaagtegg1920taacgeeagg gittteeeag teaegaegit gtaaaaegae ggeeagtega etggeeatta1980aegaeteact atagggegaa tigggtaeeg ggeeeeeee egaggtegae tiggeeatta2040tetteeeat attaggeegaa tigggtaeeg ggeeeeeee egaggtegae tiggeeatta2100aeceetteatt etceaaaetig geegteeaggg ateetggtat ggitttateg tittattet2160ggttettata geategttti ggaetteet gtteeeatta ggeggteeag gageeageeg2220agaateatte tittgaaggat actateett teceaattig atggtgataeg gageaggae2340ceageagaaa eaagtagatt tiggetetteg gggaagagaa atgaggattig aeggagaaga2400attaegeaaa agateetteg taacaagatt titteetteg titteettet tettee2520acgeetegg gaatetigt atataetee eaggaagattig tatettatette2520acgeetegg gaatetigt atataetee eaggaagattig titteettet2580gitaeatae ageggaatti ceataetee eagegg agettigt tatetaatti titteettee2520acgeetegg gaatetigt atataetee atetagaat atggtata atgtagget geetegge2580gitaeattaa ageeggaatti ceataege atetagaat atgttatett atetiggee2580gitaeattaa titteeagg ateteag aaatgeegg agettittee atgtagatga2640tateggeege acceaaatta ageeatgatt tataggeegg attaggee titteeagata2700ggaacgeeaa gagtaactie agaatgeegg attaggeege aaaaattee2700ggaacgeeaa ageeatae eeceaaatta teetgeegga aaaaattee titgaaaett2820taaaaaaaa ageeaatae eeceaaatta taaggeege etceattat aaeaaaat2820taaaaaaaa ageeaatae eeceaaatta taaggeeggee etceattata aaeaaaatee2820taaaaaaaaa ageeaatae eeceaaatta taagaage etceeaaaat2820<	gtgtagcggt	cacgctgcgc	gtaaccacca	caccegeege	gcttaatgcg	ccgctacagg	1800	
cctcttcgct attacgccag ctggcgaaag ggggatgtg tgcaaggcga ttaagttgg1920taacgccagg gttttcccag tcacgacgt gtaaaacgac ggccagtgag cgcgcgtaat1980acgactcact atagggcgaa ttgggtaccg ggcccccct cgaggtcgac tggccattaa2040tctttcccat attagatte gccaagccat gaaagttcaa gaaaggtett tagacgaatt2100accettcatt tctcaaactg gcgtcaaggg atcetggtat ggtttateg ttttattet2160ggttettata gcategttt ggacttete gtteccatta ggeggtcag gagccagege2220agaateatte tttgaaggat acttateett tceaatttg attgtetgt acgttggaca2340cggcagaaaa caagtagatt tgacteteg tagggaagaa atgaggattg agcgagaaa2400attagcaaaa agateetteg taacaagatt tttacattee tggtgttgaa gggaaagaa2400tgagetata acggggaatt ccatacaet cagatttgt tatctaatt ttteettee2520acgtecgegg gaatetgtg atattacg atetagatat atgttatet atettggeg2400ttaagcaaaa agateetteg taacaagatt ttacattee tggtgttgaa gggaaagaa2400taagcatata ageggaatt ccatateaet cagatttgt tatetaatt ttteettee2520acgtecgegg gaatetgtg atettaag aaattgegg agtttttte atgtaggtag2640tactgactge acgcaaatt aggcatgat tataggcatg atttgatge tgtaccgata2700ggaaacgetaa gagtaactte agaatgtta teetggegga aaaaatteat ttgtaact2760tacaaaaaaa agccaatta aggcaatta teetggegga aaaaatteat ttgtaaact2700ggaaacgetaa gagtaactte agaatgta teetggegga aaaaatteat ttgtaact2700ggaaacgetaa gagtaactte agaatgta tcteggegga aaaaatteat ttgtaact2700ggaaacgetaa gagtaactte agaatgta teetggegg gtettate2700taaaaaaaaa agccaataa agccaatta teetggegg gtetga aacaaattee2700taaaaaaaaa agccaatta teetggeggg ateetga gteggaaaatta2700taa	cctcttcgct attacgccag ctggcgaaag ggggatgtgc tgcaaggcga ttaagttgg1920taacgccagg gttttcccag tcacgacgtt gtaaaacgac ggccagtgag cgcgcgtaat1980acgactcact atagggcgaa ttgggtaccg ggcccccct cgaggtcgac tggccattaa2040tctttcccat attagattte gccaagcgat gaaagttcaa gaaagtctt tagacgaat2100acccttcatt tctcaaactg gcgtcaaggg atcetggtat ggtttateg tttattte2160ggttcttata gcatcgttt ggactcctc gttcccatta ggcggttcag gagccagcg2220agaatcatte tttgaaggat acttatect tccaatttg attgtcgt acgttggac2280taaactgtat actagaaatt ggacttetg tagggaagaa atgaggatgtg agggagagaa2400attagcaaaa agatcctteg taacaagat tttacatte tggtgttgaa gggaaagat2400attagcaaaa agatcettg taacaagat tttacatte tggtgttgaa gggaaagat2400ggtactatac ageggaatt ccatatcact cagatttgt tatctaattt tttecttec2520acgtccgcgg gaatetggt atattacge atctagata atgtgtagt agtgataga2400tactgactge acgcaaatat aggcatgt tatagcagg agttttttc atgtagtga2580gtacatttaa ttttcaacgt attectaag aaattgcgg agttttttc atgtagtag2700ggaacgctaa gagtaactte agatagtat tataggcag aaaaattcat ttgtaacgat2700ggaacgctaa gagtaactte cccaaatta ttaggcgg acccctattat aactaaaat2820	gegegtegeg	ccattcgcca	ttcaggctgc	gcaactgttg	ggaagggcga	tcggtgcggg	1860	
taacgecagg gtttteeeag teacgacgtt gtaaaacgae ggecagtgag egeggtaat 1980 acgacteact atagggegaa ttgggtaceg ggeececeet egaggtegae tggeeattaa 2040 tettteeeat attagatte geeaageeat gaaagteea gaaaggtett tagaegaat 2100 accetteatt teteaaaetg gegteaaggg ateetggtat ggttttateg ttttatteet 2160 ggttettata geategttt ggaetteet gtteeeatta ggeggteag gageeagege 2220 agaateatte titgaaggat aettateett teeaattig attgeeggt acgtggaea 2280 taaaetgtat aetagaaatt ggaettegt aggtgaaaeta gaagatatgg ateetggaea 2340 eggeagaaaa eaagtagatt tgaetetteg tagggaagaa atgaggatg agggagaagaa 2460 tgagetatae ageggaatte ceatateet eagattigt tateetatte titeetee 2520 acgteegegg gaateetggt atataetge atetagata atgitateet attettee 2520 gtaeattaa titteaaegt atteetaag aaattgeggg agtttitte atgitagatg 2640 taetgaetge acgeaaata aggeatgatt tataggeatg attegatgat 2700 ggaaegeetaa gagtaeette agaategtta teetggegga aaaaatteat tigtaaaett 2760	taacgecagg gtttteccag teacgacgt gtaaaacgae ggecagtgag egeeggtaat 1980 acgaeteaet ataggeegaa ttgggtaeeg ggeeeeeeet egagtegae tggeeattaa 2040 tettteeeat attagatte geeaageeat gaaagteeat gaaagteet tagaegaatt 2100 accetteatt teteaaaetg gegteaaggg ateetggtat ggtttateg tttattet 2160 ggttettata geategttt ggaetteet gtteeeatta ggeeggteag gageeagege 2220 agaateatte tttgaaggat aettateett teeaatttg attgetgtt acgttggaea 2280 taaaetgat aetagaaatt ggaettegt ggtgaaaeta gaagatatgg ateetggaea 2340 eggeagaaaa eaagtagatt tgeeteteg tageggaagaa atgaggattg agegagaaea 2400 attageeaaaa agateetteg taacaagatt tttaeatte tggtgttgaa gggaaagta 2460 tgagetatae ageggaattt ceatateaet eagaattig tatetatett ttteettee 2520 acgteeggg gaatetggt atattaetge atetagata atgttaetet ateetggege 2580 gtaeattaa ttteeaegt atteetaag aaattgeggg agttttte atgtaggatg 2580 gtaeattaa ttteeaegt atteetaag aaattgeggg agttttte atgtagatga 2640 taetgaeegg aggaaaett aggeatgta teetggegg aaaaattee ttgtaaeet 2760 taaaaaaaaa ageeaatat eceaegata teetggegg aaaaattee ttgtaaaett 2760	cctcttcgct	attacgccag	ctggcgaaag	ggggatgtgc	tgcaaggcga	ttaagttggg	1920	
acgactcactatagggggaattgggtaccgggccccccctcgaggtcgactggccattaa2040tctttcccatattagatttcgccaagccatgaaagttcaagaaaggtctttagacgaatt2100acccttcatttctcaaactggcgtcaagggatcctggtatggttttatcg1160ggttcttatagcatcgtttggacttctcgttcccattaggcggtcag2220agaatcattctttgaaggatacttatcctttccaatttgattgtcgtgta2340taaactgtatacagtagattggactttcgtaggggaagaa2400attagcaaaaagatccttcgtaacaagaat2400attagcaaaaagatcgggaattccaattatctgggtgtaaggaa2400tgagctatacagegggaatttccaatatcactcggtgtgtaaggaagaa2400attagcaaaaagatcgggaattccaattatctagtgtgtgaaggaagaa2400acgtccgcgggaatcggaattccaatatacccaggtttgta2520acgtccgcgggaatcggaattccaatacactcagatttgttactaagtttgagctatacagcggaattccatatcactcagatttgttactaattttgagctatacagcggaatttccatatcactcagatttgttatcaatttgtacatttaatttcaacgtattaggcagagttttttc2580gtacatttaatttcaatgcagaataggaattgg2700ggaacgctaagagaacgctaagagaacgtta2700ggaacgctaagagaacgttatccggcggaaatt2760taaaaaaaaaagccaatataccgcaaattattaaggccg2700taaaaaaaaaagccaatataccgcaaaattattaaggccg	acgactcact atagggegaa ttgggtaceg ggececect egaggtegae tggecattaa2040tettteeaa ttagatte gecaageeat gaaagteaa gaaagtett tagaegaatt2100accetteatt teteaaaetg gegteaaggg ateetggtat ggtttateg tttattet2160ggttettata geategttt ggaettete gtteeeatta ggeggteag gageeagee2220agaateatte tttgaaggat actateett teeaatttig attgtetgt acgttggaea2280taaaetgtat actagaaatt ggaettegg ggtggaaaeta gaagatatgg atettgate2340eggeagaaaa eaagtagatt tgaeteteg tagggaagaa atgaggattg agegagaae2400attageaaaa agateeteg taacaagatt tttacatte tggtgtgaa gggaaagata2460tgagetatae ageggaatti ceatateaet eagatttgt tatetattt ttteettee2520acgteegeg gaatetgtg atettage atettage atggtaggaagaa2460tacgacegeg gaatetgtg atattaetge atetageatt atgttatet atettggee2580gtacatttaa ttteeaegt atteetaag aaattgegg agtttttte atgtagatg2640taacagaeta aggeatatt egaatgatt tataggeag atttgatege tgtacegata2700ggaaegetaa gagtaaette egaategtat teteggegg aaaaatteat ttgtaaaett2760taaaaaaaaa ageeaatate eceaaaatta taaggeege eteeattat aaetaaaatt2820	taacgccagg	gttttcccag	tcacgacgtt	gtaaaacgac	ggccagtgag	cgcgcgtaat	1980	
tctttcccat attagattte gecaageeat gaaagttea gaaagtett tagaegaatt2100accetteatt teteaaaetg gegteaaggg ateetggtat ggttttateg ttttattet2160ggttettata geategttt ggaettett gtteeeatta ggeggtteag gageeagee2220agaateatte tttgaaggat acttateett teeaattttg attgtetgtt aegttggaea2280taaaetgtat aetagaaatt ggaetttgat ggtgaaaeta gaagatatgg ateetgatae2340eggeeagaaaa eaagtagatt tgaetetteg tageggaagaa atgaggattg agegagaaae2400attageaaaa agateetteg taacaagatt tttaeeatte tggtgttgaa gggaaagata2460tgagetatae ageggaattt ceatateet eagatttgt tatetaattt tteeettee2520acgteegeg gaatetgtgt atattaetge atetagatat atgttatet atettggege2580gtacatttaa tttteeaegt attetaag aaattgeggg agtttttte atgtagatga2640taetgaetge aegeeaatta aggeatgatt tataggeatg atttgatgge tgtaecgata2700ggaaegetaa gagtaeette agaategtta teetggegga aaaaatteet ttgtaaaett2760taaaaaaaaa ageeaatate ceeaaatta tteetageege eteeattatt aaetaaaatt2820	tcttteccatattagatttegecaagecatgaaagtteaagaaagtetttagaegaatt2100accetteatttetecaaactggegteaagggateetggtatggttttategtttattet2160ggttettatageategtttggaettetetgtteceattaggeggteag2220agaateattetttgaaggatactateetttecaatttgatgtetggaea2280taaactgtatactagaaattggaettetegtagggaagaa2340eggeagaaaacaggtagatttgaettegtagggaagaa2400attageaaaaagateettegtaacaagatttttacattetggggaagaatgagetataeagegggaattceatateacteggettit2520acgteeggggaatetgtgatattaetgeatettaattttteettegtgagetataeageggaattceatateacteggettit2580gtacatttaattteeaegtatettaggegggettittee2580gtacattaattteeaegtatettaggegg2700ggaacgetaagagtaactteagaategtttataggeat2760taaaaaaaaaagecaatatececaaattateetggggege2820teataaaaaaaaagecaatateteetgaegt2820	acgactcact	atagggcgaa	ttgggtaccg	ggccccccct	cgaggtcgac	tggccattaa	2040	
accetteatt teteaaactg gegteaaggg ateetggtat ggttttateg ttttattet2160ggttettata geategttt ggaettete gtteeeatta ggeggteag gageeagege2220agaateatte tttgaaggat acttateett teeaattttg attgtetgtt acgttggaea2280taaactgtat actagaaatt ggaetttgat ggtgaaacta gaagatatgg atettgatae2340eggeeagaaaa eaagtagatt tgaetetteg tagggaagaa atgaggattg agegagaaae2460taageeaaaa agateetteg taacaagatt tttacattee tggtgttgaa gggaaagata2460taggetatae ageggaatte eeataeete eagattegt atettagatae ttteeeteeteeteetee2520acgteegegg gaatetggt atattaeetge atetagatat atgttateet atettggege2580gtacatttaa tttteeaegt atteetaagaata taggegg agtttttte atgtaggtag2640taetgaetge acgeeaatat aggeatgatt tataggeetg attegatgge tgtaecgata2700ggaacgetaa gagtaeette agaategtta teetggega aaaaatteet ttgtaaaett2760taaaaaaaaa ageeeatate eeeaaatta ttaaggeege eteeattatt acetaaeatt2820	accetteatt teteaaaetg gegteaaggg ateetggtat ggtttateg ttttattet2160ggttettata geategttt ggaettete gtteeeatta ggeggteag gageeagege2220agaateatte tttgaaggat acttateett teeaatttg attgtetgtt acgttggaea2280taaaetgtat actagaaatt ggaetttgat ggtgaaaeta gaagatatgg atettgatae2340eggeagaaaa eagtagatt tgaeteteg tagggaagaa atgaggattg agegagaaae2400attageaaaa agateetteg taacaagatt tttacatte tggtgttgaa gggaaagata2460tgagetatae ageggaatte ceatateet eagatttgt tatetaatte ttteettee2520acgteegegg gaatetgtg atattaege atetageg agtttttte atgtaggtag2640taetgaetge acgeaaata aggeatgatt tataggeatg atttgatgge tgtaecegata2700ggaaegetaa gagtaaette agaategtta teetggegga aaaaatteet ttgtaaaett2760taaaaaaaaaa ageeaatate eecaaagta ttaaggege eteecattat aaetaaaatt2820	tctttcccat	attagatttc	gccaagccat	gaaagttcaa	gaaaggtctt	tagacgaatt	2100	
ggttettaa geategttt ggaettete gtteeeatta ggeggteag gageeagege2220agaateatte tttgaaggat aettateett teeaattttg attgtetgtt aegttggaea2280taaaetgtat aetagaaatt ggaetttgat ggtgaaaeta gaagatatgg atettgatae2340eggeagaaaa eaagtagatt tgaetetteg tagggaagaa atgaggattg agegagaae2400attageaaaa agateetteg taacaagatt tttaeette tggtgttgaa gggaaagata2460tgagetatae ageggaattt ceatateet eagatttgt tatetaattt ttteettee2520acgteegeg gaatetgtgt atattaetge atetagata atgttatet atettggege2580gtaeatttaa tttteeaegt attetataag aaattgeggg agtttttte atgtagatga2640tactgaetge aegeaaatat aggeatgatt tataggeatg atttgatgge tgtaeegata2700ggaaegetaa gagtaaette agaategtta teetggegga aaaaatteat ttgtaaaett2760	ggttettata geategttt ggaetteet gtteecatta ggeggteag gagecagege2220agaateatte tttgaaggat aettateett teeaattttg attgtetgtt aegttggaea2280taaaetgtat aetagaaatt ggaetttgat ggtgaaaeta gaagatatgg atettgatae2340eggeagaaaa eaagtagatt tgaetetteg tagggaagaa atgaggattg agegagaae2400attageaaaa agateetteg taacaagatt tttaeattee tggtgttgaa gggaaagata2460tgagetatae ageggaattt ecatateet eagattttgt tatetaatte ttteettee2520acgteegegg gaatetgtgt atattaetge atetagatat atgttatett atettggege2580gtaeatttaa tttteeaegt attetatag aaattgeggg agtttttte atgtagatga2640taetgaetge aegeaaatat aggeatgatt tataggeatg atttgatgge tgtaecgata2700ggaaegetaa gagtaaette agaategtta teetggegga aaaaatteat ttgtaaaett2760taaaaaaaaa ageeaatate cecaaaatta ttaaggeege eteeattat aaetaaaatt2820	acccttcatt	tctcaaactg	gcgtcaaggg	atcctggtat	ggttttatcg	ttttatttct	2160	
agaatcatte titgaaggat acttateett teeaattitg attgtetgtt acgitggaca2280taaactgtat actagaaatt ggaetttgat ggtgaaacta gaagatatgg atettgatae2340cggeagaaaa eaagtagatt tgaetetteg tagggaagaa atgaggattg agegagaae2400attageaaaa agateetteg taacaagatt titaeattee tggtgttgaa gggaaagata2460tgagetatae ageggaatte eeatatee eagattitgt tatetaatte titeetteee2520acgteegeg gaatetgtgt atattaetge atetagatat atgttatett atettggege2580gtaeatttaa titteeaegt attetataag aaattgeggg agtttitte atgtagatga2640taetgaetge acgeaaatat aggeatgatt tataggeatg atttgatege tgtaecgata2700ggaaegetaa gagtaaette agaategtta teetggegga aaaaatteat tigtaaaett2760	agaatcatte titgaaggat aettateett teeaattitg attgtetgtt aegtiggaca2280taaactgtat aetagaaatt ggaetttgat ggtgaaacta gaagatatgg atettgatae2340eggeagaaaa eaagtagatt tgaetetteg tagggaagaa atgaggattg agegagaaae2400attageaaaa agateetteg taacaagatt tittaeattee tggtgttgaa gggaaagata2460tgagetatae ageggaatte eeatateet eagattitgt tatetaatte titeetteee2520acgteegeg gaatetgtgt atattaetge atetagatat atgttatett atettggege2580gtaeatttaa titteeaegt atteetataag aaattgeggg agtttitte atgtagatga2640taetgaetge aegeaaatat aggeatgatt tataggeatg attgatgge tgtaeegata2700ggaaeegetaa gagtaaette agaategtta teetggegga aaaaatteat titgtaaaett2760taaaaaaaaa ageeaatate eecaaaatta ttaaggeege eteecattat aaetaaaatt2820	ggttcttata	gcatcgtttt	ggacttctct	gttcccatta	ggcggttcag	gagccagcgc	2220	
taaactgtat actagaaatt ggactttgat ggtgaaacta gaagatatgg atcttgatac 2340 cggcagaaaa caagtagatt tgactcttog tagggaagaa atgaggattg agoggagaac 2400 attagcaaaa agatocttog taacaagatt tttacattto tggtgttgaa gggaaagata 2460 tgagetatac agoggaattt ocatatcact cagattttgt tatotaattt tttocttoco 2520 acgtoogog gaatotgtgt atattactgo atotagatat atgttatott atottggogo 2580 gtacatttaa ttttcaacgt attotataag aaattgoggg agtttttto atgtagatga 2640 tactgactgo acgcaaatat aggcatgatt tataggcatg atttgatggo tgtaccgata 2700 ggaacgotaa gagtaactto agaatogtta tootggogga aaaaattoat ttgtaaactt 2760	taaactgtat actagaaatt ggactttgat ggtgaaacta gaagatatgg atcttgatac2340cggcagaaaa caagtagatt tgactcttcg tagggaagaa atgaggattg agcgagaaac2400attagcaaaa agatcetteg taacaagatt tttacatte tggtgttgaa gggaaagata2460tgagetatac ageggaattt ecatateaet eagattttgt tatetaattt ttteettee2520acgteegeg gaatetgtgt atattaetge atetagatat atgttatett atettggege2580gtacatttaa tttteeaegt attetataag aaattgeggg agtttttte atgtagatga2640tactgaetge aegeaaatat aggeatgatt tataggeatg atttgatgge tgtaecgata2700ggaaegetaa gagtaaette agaategtta teetggegga aaaaatteat ttgtaaaett2760taaaaaaaaa ageeaatate eccaaaatta ttaagagege eteeattat aaetaaaatt2820	agaatcattc	tttgaaggat	acttatcctt	tccaattttg	attgtctgtt	acgttggaca	2280	
cggcagaaaa caagtagatt tgactetteg tagggaagaa atgaggattg agegagaaac2400attagcaaaa agateetteg taacaagatt tttacatte tggtgttgaa gggaaagata2460tgagetatae ageggaattt ecatateaet eagattttgt tatetaattt ttteetteee2520acgteegeg gaatetgtgt atattaetge atetagatat atgttatett atettggege2580gtacatttaa tttteaaegt attetataag aaattgeggg agtttttte atgtagatga2640taetgaetge aegeaaatat aggeatgatt tataggeatg atttgatgge tgtaeegata2700ggaaegetaa gagtaaette agaategtta teetggegga aaaaatteat ttgtaaaett2760taaaaaaaaa agecaatate eegaaatta ttaaggeege eteetatta aaetaaaatt2820	cggcagaaaa caagtagatt tgactetteg tagggaagaa atgaggattg agegagaaac2400attagcaaaa agateetteg taacaagatt tttacattee tggtgttgaa gggaaagata2460tgagetatae ageggaatte ecatateaet eagatttgt tatetaatte ttteettee2520acgteegegg gaatetgtgt atattaetge atetagatat atgttatett atettggege2580gtacatttaa tttteeaegt attetataag aaattgeggg agtttttte atgtagatga2640taetgaetge acgeaaatat aggeatgatt tataggeatg atttgatgge tgtaeegata2700ggaaegetaa gagtaaette agaategtta teetggegga aaaaatteat ttgtaaaett2760taaaaaaaaa ageeaatate eccaaaatta ttaagagege etceattat aaetaaaatt2820	taaactgtat	actagaaatt	ggactttgat	ggtgaaacta	gaagatatgg	atcttgatac	2340	
attagcaaaa agateetteg taacaagatt tttacattee tggtgttgaa gggaaagata2460tgagetatae ageggaatte ecatateet eagattetgt tatetaatte ttteetteee2520acgteegegg gaatetgtgt atattaetge atetagatat atgttatett atettggege2580gtacatttaa tttteaaegt attetataag aaattgeggg agtttttte atgtagatga2640taetgaetge aegeaaatat aggeatgatt tataggeatg attegatgge tgtaecegata2700ggaaegetaa gagtaaette agaategtta teetggegga aaaaatteat ttgtaaaett2760taaaaaaaaa ageeaatate eegeaaatta ttaaggeege eteeattatt aaetaaaatt2820	attagcaaaa agateetteg taacaagatt tttacattte tggtgttgaa gggaaagata2460tgagetatae ageggaattt eeatteette cagattttgt tatetaattt ttteetteee2520acgteegeg gaatetgtgt atattaetge atetagatat atgttatett atettggege2580gtacatttaa tttteaaegt attetaaag aaattgeggg agtttttte atgtagatga2640taetgaetge aegeaaatat aggeatgatt tataggeatg atttgatgge tgtaeegata2700ggaaegetaa gagtaaette agaategtta teetggegga aaaaatteat ttgtaaaett2760taaaaaaaaa ageeaatate eecaaaatta ttaagagege eteeattat aaetaaaatt2820	cggcagaaaa	caagtagatt	tgactcttcg	tagggaagaa	atgaggattg	agcgagaaac	2400	
tgagetatae ageggaattt eeatateaet eagattttgt tatetaattt ttteetteee 2520 aegteegeg gaatetgtgt atattaetge atetagatat atgttatett atettggege 2580 gtaeatttaa tttteaaegt attetataag aaattgeggg agtttttte atgtagatga 2640 taetgaetge aegeaaatat aggeatgatt tataggeatg atttgatgge tgtaeegata 2700 ggaaegetaa gagtaaette agaategtta teetggegga aaaaatteat ttgtaaaett 2760 taaaaaaaaa ageeaatate eeseaaatta ttaaggegeg eteeattatt aaetaaaatt 2820	tgagctatac agoggaattt ccatatcact cagattttgt tatctaattt tttocttooc2520acgtcogogg gaatctgtgt atattactge atetagatat atgttatett atettggoge2580gtacatttaa ttttcaacgt attetataag aaattgeggg agtttttte atgtagatga2640tactgactge acgeaaatat aggeatgatt tataggeatg attegatgge tgtacegata2700ggaacgetaa gagtaaette agaategtta teetggegga aaaaatteat ttgtaaaett2760taaaaaaaaa ageeaatate eccaaaatta ttaagagege etceattat aaetaaaatt2820	attagcaaaa	agatccttcg	taacaagatt	tttacatttc	tggtgttgaa	gggaaagata	2460	
acgteegeg gaatetgtgt atattaetge atetagatat atgttatett atettggege 2580 gtacatttaa titteaaegt attetataag aaattgeggg agttittte atgtagatga 2640 taetgaetge aegeaaatat aggeatgatt tataggeatg attigatgge tgtaeegata 2700 ggaaegetaa gagtaaette agaategtta teetggegga aaaaatteat tigtaaaett 2760 taaaaaaaaa ageeaatate eegaaatta taaggagge eteeattatt aaetaaaatt 2820	acgteegeg gaatetgtgt atattactge atetagatat atgttatett atettggege2580gtacatttaa ttttcaacgt attetataag aaattgeggg agtttttte atgtagatga2640tactgactge acgeaaatat aggeatgatt tataggeatg attgatgge tgtacegata2700ggaacgetaa gagtaaette agaategtta teetggegga aaaaatteat ttgtaaaett2760taaaaaaaaa ageeaatate eccaaaatta ttaaggege etceattatt aaetaaaatt2820teaeteagea teesacaatgt ateeggtate tastagageg attagatgta geggaaage2820	tgagctatac	agcggaattt	ccatatcact	cagattttgt	tatctaattt	tttccttccc	2520	
gtacatttaa ttttcaacgt attctataag aaattgcggg agttttttc atgtagatga 2640 tactgactgc acgcaaatat aggcatgatt tataggcatg atttgatggc tgtaccgata 2700 ggaacgctaa gagtaacttc agaatcgtta tcctggcgga aaaaattcat ttgtaaactt 2760 taaaaaaaaa agccaatatc cscaaaatta ttaaggaggg stccattatt aactaaaatt 2820	gtacatttaa ttttcaacgt attctataag aaattgcggg agttttttc atgtagatga 2640 tactgactgc acgcaaatat aggcatgatt tataggcatg atttgatggc tgtaccgata 2700 ggaacgctaa gagtaacttc agaatcgtta tcctggcgga aaaaattcat ttgtaaactt 2760 taaaaaaaaa agccaatatc cccaaaatta ttaagagcgc ctccattatt aactaaaatt 2820	acgtccgcgg	gaatctgtgt	atattactgc	atctagatat	atgttatctt	atcttggcgc	2580	
tactgactge acgeaaatat aggeatgatt tataggeatg atttgatgge tgtacegata 2700 ggaaegetaa gagtaaette agaategtta teetggegga aaaaatteat ttgtaaaett 2760 taaaaaaaaa ageeaatate eesaaatta ttaagagege etceattatt aactaaaatt 2820	tactgactge acgeaaatat aggeatgatt tataggeatg atttgatgge tgtacegata 2700 ggaaegetaa gagtaaette agaategtta teetggegga aaaaatteat ttgtaaaett 2760 taaaaaaaaa ageeaatate eecaaaatta ttaagagege eteeattatt aaetaaaatt 2820	gtacatttaa	ttttcaacgt	attctataag	aaattgcggg	agttttttc	atgtagatga	2640	
ggaacgetaa gagtaaette agaategtta teetggegga aaaaatteat ttgtaaaett 2760	ggaacgetaa gagtaaette agaategtta teetggegga aaaaatteat ttgtaaaett 2760 taaaaaaaaa agecaatate eecaaaatta ttaagagege eteeatatt aaetaaaatt 2820	tactgactgc	acgcaaatat	aggcatgatt	tataggcatg	atttgatggc	tgtaccgata	2700	
taaaaaaaa agccaatato oocaaaatta ttaagagogo otocattatt aactaaaatt 2820	taaaaaaaa agccaatate eecaaaatta ttaagagege eteeattatt aaetaaaatt 2820	ggaacgctaa	gagtaacttc	agaatcgtta	tcctggcgga	aaaaattcat	ttgtaaactt	2760	
	tractragra treagatgt atraggtate tactagagat attagatgtg grgaaaaga 2880	taaaaaaaaa	agccaatatc	cccaaaatta	ttaagagcgc	ctccattatt	aactaaaatt	2820	
tcactcagca tccacaatgt atcaggtatc tactacagat attacatgtg gcgaaaaaga 2880	concompon cooncluter accuration cacturague accuration yeyaaaaaaya 2000	tcactcagca	tccacaatgt	atcaggtatc	tactacagat	attacatgtg	gcgaaaaaga	2880	

131

#### -continued

caagaacaat gcaatagcgc atcaagaaaa aacacaaagc tttcaatcaa tgaatcgaaa 2940 atgtcattaa aatagtatat aaattgaaac taagtcataa agctataaaa agaaaattta 3000 tttaaatgca agatttaaag taaattcacg gccctgcagg ccctaacctg ctaggacaca 3060 acgtetttge etggtaaagt ttetagetga egtgatteet teacetgtgg ateeggeaat 3120 tgtaaaggtt gtgaaaccct cagcttcata accgacacct gcaaatgact ttgcattctt 3180 aacaaagata gttgtatcaa tttcacgttc gaatctatta aggttatcga tgttcttaga 3240 ataaatgtag geggaatgtt ttetattetg eteagetate ttggegtatt taatggette 3300 atcaatgtcc ttcactctaa ctataggcaa aattggcatc atcaactccg tcataacgaa 3360 cggatggttt gcgttgactt cacaaataat acactttaca ttacttggtg actctacatc 3420 tatttcatcc aaaaacagtt tagcgtcctt accaacccac ttcttattaa tgaaatattc 3480 ttgagtttca ttgttctttt gaagaacaag gtctatcagc ttggatactt ggtcttcatt 3540 gataatgacg gcgttgtttt tcaacatgtt agagatcaga tcatctgcaa cgttttcaaa 3600 cacqaacact tettttteeq eqatacaaqq aaqattqttq teaaacqaac aacetteaat 3660 aatgettetg ceggeettet egatatetge tgtategtet acaataaeeg gaggattaee 3720 cgcgccagct ccgatggcct ttttaccaga attaagaagg gtttttacca tacccgggcc 3780 accortaccy cacaacaatt ttatggatgg atgtttgata atagggtcta aactttccat 3840 3900 aqttqqqttc tttataqtaq tqacaaqqtt ttcaqqtcca ccacaqctaa ttatqqcttt gtttatcatt tctactgcga aagcgacaca ctttttggcg catgggtgac cattaaatac 3960 aactgcattc cccgcagcta tcatacctat agaattgcag ataacggttt ctgttggatt 4020 cgtgcttgga gttatagcgc cgataactcc gtatggactc atttcaacca ctgttagtcc 4080 attatcgccg gaccatgctg ttgttgtcag atcttcagtg cctggggtat acttggccac 4140 taattcatgt ttcaagattt tatcctcata ccttcccatg tgggtttcct ccaggatcat 4200 tgtggctaag acctetttat tetgtaatge ggettttett attteggtga ttattttete 4260 tetttgttee tttgtgtagt gtagggaaag aatettttgt geatgtaetg eagaagaaat 4320 ggcattetea acatttteaa atacteeaaa acatgaagag ttatetttgt aattetttaa 4380 gttgatgttt tcaccattag tcttcacttt caagtctttg gtggttggga ttaaggtatc 4440 tttatccatg gtgtttgttt atgtgtgttt attcgaaact aagttcttgg tgttttaaaa 4500 ctaaaaaaaa gactaactat aaaagtagaa tttaagaagt ttaagaaata gatttacaga 4560 attacaatca atacctaccg tctttatata cttattagtc aagtaggggga ataatttcag 4620 ggaactggtt tcaacctttt ttttcagctt tttccaaatc agagagagca gaaggtaata 4680 gaaggtgtaa gaaaatgaga tagatacatg cgtgggtcaa ttgccttgtg tcatcattta 4740 ctccaggcag gttgcatcac tccattgagg ttgtgcccgt tttttgcctg tttgtgcccc 4800 tgttctctgt agttgcgcta agagaatgga cctatgaact gatggttggt gaagaaaaca 4860 atattttggt gctgggattc ttttttttc tggatgccag cttaaaaagc gggctccatt 4920 4980 atatttagtg gatgccagga ataaactgtt cacccagaca cctacgatgt tatatattct gtgtaacccg ccccctattt tgggcatgta cgggttacag cagaattaaa aggctaattt 5040 tttgactaaa taaagttagg aaaatcacta ctattaatta tttacgtatt ctttgaaatg 5100 gcagtattga taatgataaa ctcgaactga aaaagcgtgt tttttattca aaatgattct 5160

aactccctta cgtaatcaag gaatcttttt gccttggcct ccgcgtcatt aaacttcttg

5220

133

ttgttgacgc	taacattcaa	cgctagtata	tattcgtttt	tttcaggtaa	gttcttttca	5280
acgggtctta	ctgatgaggc	agtcgcgtct	gaacctgtta	agaggtcaaa	tatgtcttct	5340
tgaccgtacg	tgtcttgcat	gttattagct	ttgggaattt	gcatcaagtc	ataggaaaat	5400
ttaaatcttg	gctctcttgg	gctcaaggtg	acaaggtcct	cgaaaatagg	gcgcgcccca	5460
ccgcggtgga	gctccagctt	ttgttccctt	tagtgagggt	taattgcgcg	cttggcgtaa	5520
tcatggtcat	agctgtttcc	tgtgtgaaat	tgttatccgc	tcacaattcc	acacaacata	5580
ggagccggaa	gcataaagtg	taaagcctgg	ggtgcctaat	gagtgaggta	actcacatta	5640
attgcgttgc	gctcactgcc	cgctttccag	tcgggaaacc	tgtcgtgcca	gctgcattaa	5700
tgaatcggcc	aacgcgcggg	gagaggcggt	ttgcgtattg	ggcgctcttc	cgcttcctcg	5760
ctcactgact	cgctgcgctc	ggtcgttcgg	ctgcggcgag	cggtatcagc	tcactcaaag	5820
gcggtaatac	ggttatccac	agaatcaggg	gataacgcag	gaaagaacat	gtgagcaaaa	5880
ggccagcaaa	aggccaggaa	ccgtaaaaag	gccgcgttgc	tggcgttttt	ccataggctc	5940
cgcccccctg	acgagcatca	caaaaatcga	cgctcaagtc	agaggtggcg	aaacccgaca	6000
ggactataaa	gataccaggc	gtttccccct	ggaagctccc	tcgtgcgctc	tcctgttccg	6060
accctgccgc	ttaccggata	cctgtccgcc	tttctccctt	cgggaagcgt	ggcgctttct	6120
catageteac	gctgtaggta	tctcagttcg	gtgtaggtcg	ttcgctccaa	gctgggctgt	6180
gtgcacgaac	cccccgttca	gcccgaccgc	tgcgccttat	ccggtaacta	tcgtcttgag	6240
tccaacccgg	taagacacga	cttatcgcca	ctggcagcag	ccactggtaa	caggattagc	6300
agagcgaggt	atgtaggcgg	tgctacagag	ttcttgaagt	ggtggcctaa	ctacggctac	6360
actagaagga	cagtatttgg	tatctgcgct	ctgctgaagc	cagttacctt	cggaaaaaga	6420
gttggtagct	cttgatccgg	caaacaaacc	accgctggta	gcggtggttt	ttttgtttgc	6480
aagcagcaga	ttacgcgcag	aaaaaagga	tctcaagaag	atcctttgat	cttttctacg	6540
gggtctgacg	ctcagtggaa	cgaaaactca	cgttaaggga	ttttggtcat	gagattatca	6600
aaaaggatct	tcacctagat	ccttttaaat	taaaaatgaa	gttttaaatc	aatctaaagt	6660
atatatgagt	aaacttggtc	tgacagttac	caatgcttaa	tcagtgaggc	acctatctca	6720
gcgatctgtc	tatttcgttc	atccatagtt	gcctgactcc	ccgtcgtgta	gataactacg	6780
atacgggagg	gcttaccatc	tggccccagt	gctgcaatga	taccgcgaga	cccacgctca	6840
ccggctccag	atttatcagc	aataaaccag	ccagccggaa	gggccgagcg	cagaagtggt	6900
cctgcaactt	tatccgcctc	catccagtct	attaattgtt	gccgggaagc	tagagtaagt	6960
agttcgccag	ttaatagttt	gcgcaacgtt	gttgccattg	ctacaggcat	cgtggtgtca	7020
cgctcgtcgt	ttggtatggc	ttcattcagc	tccggttccc	aacgatcaag	gcgagttaca	7080
tgatccccca	tgttgtgcaa	aaaagcggtt	agctccttcg	gtcctccgat	cgttgtcaga	7140
agtaagttgg	ccgcagtgtt	atcactcatg	gttatggcag	cactgcataa	ttctcttact	7200
gtcatgccat	ccgtaagatg	cttttctgtg	actggtgagt	actcaaccaa	gtcattctga	7260
gaatagtgta	tgcggcgacc	gagttgctct	tgcccggcgt	caatacggga	taataccgcg	7320
ccacatagca	gaactttaaa	agtgctcatc	attggaaaac	gttetteggq	gcgaaaactc	7380
tcaaggatct	taccactatt	 gagatecagt	togatotaac	ccactestas	acccaactca	7440
tetteagest	cttttactt+	caccaccat+	tataataaa	caaaaacacc	aaddcaaaat	7500
accacco	agggaataat	aaaaaaaa	aaatattaat	tactostost	attaatttt	7560
yeegeaaaaa	ayyyaataag	ygcgacacgg	aaatgttgaa	LACECATACT	GELOCETEE	/ 560
caatattatt	gaagcattta	tcagggttat	tgtctcatga	gcggatacat	atttgaatgt	7620

135

atttagaaaa ataaacaaat aggggttccg cgcacatttc cccgaaaagt gccacctgaa

#### -continued

cgaagcatct	gtgcttcatt	ttgtagaaca	aaaatgcaac	gcgagagcgc	taatttttca	7740
aacaaagaat	ctgagctgca	tttttacaga	acagaaatgc	aacgcgaaag	cgctatttta	7800
ccaacgaaga	atctgtgctt	catttttgta	aaacaaaaat	gcaacgcgag	agcgctaatt	7860
tttcaaacaa	agaatctgag	ctgcattttt	acagaacaga	aatgcaacgc	gagagcgcta	7920
ttttaccaac	aaagaatcta	tacttctttt	ttgttctaca	aaaatgcatc	ccgagagcgc	7980
tatttttcta	acaaagcatc	ttagattact	tttttctcc	tttgtgcgct	ctataatgca	8040
gtctcttgat	aactttttgc	actgtaggtc	cgttaaggtt	agaagaaggc	tactttggtg	8100
tctattttct	cttccataaa	aaaagcctga	ctccacttcc	cgcgtttact	gattactagc	8160
gaagctgcgg	gtgcatttt	tcaagataaa	ggcatccccg	attatattct	ataccgatgt	8220
ggattgcgca	tactttgtga	acagaaagtg	atagcgttga	tgattcttca	ttggtcagaa	8280
aattatgaac	ggtttcttct	attttgtctc	tatatactac	gtataggaaa	tgtttacatt	8340
ttcgtattgt	tttcgattca	ctctatgaat	agttcttact	acaattttt	tgtctaaaga	8400
gtaatactag	agataaacat	aaaaatgta	gaggtcgagt	ttagatgcaa	gttcaaggag	8460
cgaaaggtgg	atgggtaggt	tatataggga	tatagcacag	agatatatag	caaagagata	8520
cttttgagca	atgtttgtgg	aagcggtatt	cgcaatattt	tagtagctcg	ttacagtccg	8580
gtgcgttttt	ggttttttga	aagtgcgtct	tcagagcgct	tttggttttc	aaaagcgctc	8640
tgaagttcct	atactttcta	gagaatagga	acttcggaat	aggaacttca	aagcgtttcc	8700
gaaaacgagc	gcttccgaaa	atgcaacgcg	agctgcgcac	atacagctca	ctgttcacgt	8760
cgcacctata	tctgcgtgtt	gcctgtatat	atatatacat	gagaagaacg	gcatagtgcg	8820
tgtttatgct	taaatgcgta	cttatatgcg	tctatttatg	taggatgaaa	ggtagtctag	8880
tacctcctgt	gatattatcc	cattccatgc	ggggtatcgt	atgcttcctt	cagcactacc	8940
ctttagctgt	tctatatgct	gccactcctc	aattggatta	gtctcatcct	tcaatgctat	9000
catttccttt	gatattggat	catactaaga	aaccattatt	atcatgacat	taacctataa	9060
aaataggcgt	atcacgaggc	cctttcgtc				9089
<210> SEQ : <211> LENG <212> TYPE <213> ORGAI <220> FEATU <223> OTHEN	ID NO 15 IH: 5956 : DNA NISM: Artif: JRE: R INFORMATIC	icial Sequer DN: pBTX1 p	nce plasmid			
<400> SEQUI	ENCE: 15					
tcgaggtcga	cggtatcgat	aagcttgata	tcgaattcct	gcagcccggg	ggatctgaaa	60
tgaataacaa	tactgacagt	actaaataat	tgcctacttg	gcttcacata	cgttgcatac	120
gtcgatatag	ataataatga	taatgacagc	aggattatcg	taatacgtaa	tagttgaaaa	180
tctcaaaaat	gtgtgggtca	ttacgtaaat	aatgatagga	atgggattct	tctatttttc	240
ctttttccat	tctagcagcc	gtcgggaaaa	cgtggcatcc	tctctttcgg	gctcaattgg	300
agtcacgctg	ccgtgagcat	cctctcttc	catatctaac	aactgagcac	gtaaccaatg	360
gaaaagcatg	agcttagcgt	tgctccaaaa	aagtattgga	tggttaatac	catttgtctg	420
ttctcttctg	actttgactc	ctcaaaaaaa	aaaaatctac	aatcaacaga	tcgcttcaat	480
tacgccctca	caaaaacttt	tttccttctt	cttcgcccac	gttaaatttt	atccctcatg	540

137

ttgtctaacg	gatttctgca	cttgatttat	tataaaaaga	caaagacata	atacttctct	600	
atcaatttca	gttattgttc	ttccttgcgt	tattcttctg	ttcttcttt	tcttttgtca	660	
tatataacca	taaccaagta	atacatattc	aaatctagag	ggatccctga	ggttaattaa	720	
acgcgtgagt	aagcgaattt	cttatgattt	atgatttta	ttattaaata	agttataaaa	780	
aaaataagtg	tatacaaatt	ttaaagtgac	tcttaggttt	taaaacgaaa	attcttattc	840	
ttgagtaact	ctttcctgta	ggtcaggttg	ctttctcagg	tatagcatga	ggtcgctctt	900	
attgaccaca	cctctaccgg	catgccgagc	aaatgcctgc	aaatcgctcc	ccatttcacc	960	
caattgtaga	tatgctaact	ccagcaatga	gttgatgaat	ctcggtgtgt	attttatgtc	1020	
ctcagaggac	aacacctgtg	gtactagttc	tagagcggcc	gccaccgcgg	tggagctcca	1080	
gcttttgttc	cctttagtga	gggttaattg	cgcgcttggc	gtaatcatgg	tcatagctgt	1140	
ttcctgtgtg	aaattgttat	ccgctcacaa	ttccacacaa	cataggagcc	ggaagcataa	1200	
agtgtaaagc	ctggggtgcc	taatgagtga	ggtaactcac	attaattgcg	ttgcgctcac	1260	
tgcccgcttt	ccagtcggga	aacctgtcgt	gccagctgca	ttaatgaatc	ggccaacgcg	1320	
cggggagagg	cggtttgcgt	attgggcgct	cttccgcttc	ctcgctcact	gactcgctgc	1380	
gctcggtcgt	tcggctgcgg	cgagcggtat	cageteacte	aaaggcggta	atacggttat	1440	
ccacagaatc	aggggataac	gcaggaaaga	acatgtgagc	aaaaggccag	caaaaggcca	1500	
ggaaccgtaa	aaaggccgcg	ttgctggcgt	ttttccatag	gctccgcccc	cctgacgagc	1560	
atcacaaaaa	tcgacgctca	agtcagaggt	ggcgaaaccc	gacaggacta	taaagatacc	1620	
aggcgtttcc	ccctggaagc	tccctcgtgc	gctctcctgt	tccgaccctg	ccgcttaccg	1680	
gatacctgtc	cgcctttctc	ccttcgggaa	gcgtggcgct	ttctcatagc	tcacgctgta	1740	
ggtatctcag	ttcggtgtag	gtcgttcgct	ccaagctggg	ctgtgtgcac	gaaccccccg	1800	
ttcagcccga	ccgctgcgcc	ttatccggta	actatcgtct	tgagtccaac	ccggtaagac	1860	
acgacttatc	gccactggca	gcagccactg	gtaacaggat	tagcagagcg	aggtatgtag	1920	
gcggtgctac	agagttcttg	aagtggtggc	ctaactacgg	ctacactaga	aggacagtat	1980	
ttggtatctg	cgctctgctg	aagccagtta	ccttcggaaa	aagagttggt	agctcttgat	2040	
ccggcaaaca	aaccaccgct	ggtagcggtg	gttttttgt	ttgcaagcag	cagattacgc	2100	
gcagaaaaaa	aggatctcaa	gaagatcctt	tgatcttttc	tacggggtct	gacgctcagt	2160	
ggaacgaaaa	ctcacgttaa	gggattttgg	tcatgagatt	atcaaaaagg	atcttcacct	2220	
agatcctttt	aaattaaaaa	tgaagtttta	aatcaatcta	aagtatatat	gagtaaactt	2280	
ggtctgacag	ttaccaatgc	ttaatcagtg	aggcacctat	ctcagcgatc	tgtctatttc	2340	
gttcatccat	agttgcctga	ctccccgtcg	tgtagataac	tacgatacgg	gagggcttac	2400	
catctggccc	cagtgctgca	atgataccgc	gagacccacg	ctcaccggct	ccagatttat	2460	
cagcaataaa	ccagccagcc	ggaagggccg	agcgcagaag	tggtcctgca	actttatccg	2520	
cctccatcca	gtctattaat	tgttgccggg	aagctagagt	aagtagttcg	ccagttaata	2580	
gtttgcgcaa	cgttgttgcc	attgctacag	gcatcgtggt	gtcacgctcg	tcgtttggta	2640	
tggcttcatt	cageteeggt	tcccaacgat	caaggcgagt	tacatgatcc	cccatgttgt	2700	
gcaaaaaaqc	ggttagetee	tteggteete	cgatcqttqt	cagaaqtaaq	ttggccqcaq	2760	
tattatcact	catggttatg	gcagcactor	ataattetet	tactotcato	ccatccotaa	2820	
gatgetttte	tataactaat	gagtactoo	ccaadtoatt	ctgagaatag	tatatacaga	2880	
gauguluuc	at at t	yaytattaa	anact -	anara	egeacycyge	2000	
yacegagttg	electgedeg	yeyceaatae	yyyataatac	egegecadat	aycagaactt	2940	

139

# 140

taaaagtgct	catcattgga	aaacgttctt	cggggcgaaa	actctcaagg	atcttaccgc	3000
tgttgagatc	cagttcgatg	taacccactc	gtgcacccaa	ctgatcttca	gcatctttta	3060
ctttcaccag	cgtttctggg	tgagcaaaaa	caggaaggca	aaatgccgca	aaaaagggaa	3120
taagggcgac	acggaaatgt	tgaatactca	tactcttcct	ttttcaatat	tattgaagca	3180
tttatcaggg	ttattgtctc	atgagcggat	acatatttga	atgtatttag	aaaaataaac	3240
aaataggggt	tccgcgcaca	tttccccgaa	aagtgccacc	tgggtccttt	tcatcacgtg	3300
ctataaaaat	aattataatt	taaatttttt	aatataaata	tataaattaa	aaatagaaag	3360
taaaaaaga	aattaaagaa	aaaatagttt	ttgttttccg	aagatgtaaa	agactctagg	3420
gggatcgcca	acaaatacta	ccttttatct	tgctcttcct	gctctcaggt	attaatgccg	3480
aattgtttca	tcttgtctgt	gtagaagacc	acacacgaaa	atcctgtgat	tttacatttt	3540
acttatcgtt	aatcgaatgt	atatctattt	aatctgcttt	tcttgtctaa	taaatatata	3600
tgtaaagtac	gctttttgtt	gaaatttttt	aaacctttgt	ttatttttt	ttcttcattc	3660
cgtaactctt	ctaccttctt	tatttacttt	ctaaaatcca	aatacaaaac	ataaaaataa	3720
ataaacacag	agtaaattcc	caaattattc	catcattaaa	agatacgagg	cgcgtgtaag	3780
ttacaggcaa	gcgatccgtc	ctaagaaacc	attattatca	tgacattaac	ctataaaaat	3840
aggcgtatca	cgaggccctt	tcgtctcgcg	cgtttcggtg	atgacggtga	aaacctctga	3900
cacatgcagc	tcccggagac	ggtcacagct	tgtctgtaag	cggatgccgg	gagcagacaa	3960
gcccgtcagg	gcgcgtcagc	gcgtgttggc	gggtgtcggg	gctggcttaa	ctatgcggca	4020
tcagagcaga	ttgtactgag	agtgcaccat	aaattcccgt	tttaagagct	tggtgagcgc	4080
taggagtcac	tgccaggtat	cgtttgaaca	cggcattagt	cagggaagtc	ataacacagt	4140
cctttcccgc	aattttcttt	ttctattact	cttggcctcc	tctagtacac	tctatatttt	4200
tttatgcctc	ggtaatgatt	ttcattttt	ttttcccct	agcggatgac	tcttttttt	4260
tcttagcgat	tggcattatc	acataatgaa	ttatacatta	tataaagtaa	tgtgatttct	4320
tcgaagaata	tactaaaaaa	tgagcaggca	agataaacga	aggcaaagat	gacagagcag	4380
aaagccctag	taaagcgtat	tacaaatgaa	accaagattc	agattgcgat	ctctttaaag	4440
ggtggtcccc	tagcgataga	gcactcgatc	ttcccagaaa	aagaggcaga	agcagtagca	4500
gaacaggcca	cacaatcgca	agtgattaac	gtccacacag	gtatagggtt	tctggaccat	4560
atgatacatg	ctctggccaa	gcattccggc	tggtcgctaa	tcgttgagtg	cattggtgac	4620
ttacacatag	acgaccatca	caccactgaa	gactgcggga	ttgetetegg	tcaagctttt	4680
aaagaggccc	tactggcgcg	tggagtaaaa	aggtttggat	caggatttgc	gcctttggat	4740
gaggcacttt	ccagagcggt	ggtagatett	tcgaacaggc	cgtacgcagt	tgtcgaactt	4800
ggtttgcaaa	gggagaaagt	aggagatctc	tcttgcgaga	tgatcccgca	ttttcttgaa	4860
agctttgcag	aggctagcag	aattaccctc	cacgttgatt	gtctgcgagg	caagaatgat	4920
catcaccgta	gtgagagtgc	gttcaaggct	cttgcggttg	ccataagaga	agccacctcg	4980
cccaatggta	ccaacgatgt	tccctccacc	aaaggtgttc	ttatgtagtg	acaccgatta	5040
tttaaagctg	cagcatacga	tatatataca	tgtgtatata	tgtataccta	tgaatgtcag	5100
taagtatgta	tacgaacagt	atgatactga	agatgacaag	gtaatgcatc	attctatacg	5160
tgtcattctg	aacgaggcgc	gctttccttt	tttctttttg	cttttcttt	tttttctct	5220
tgaactcgac	ggatctatqc	ggtgtgaaat	accgcacaqa	tgcgtaagqa	gaaaataccq	5280

141

### 142

catcaggaaa ttg	taaacgt taatattttg	ttaaaattcg cgt	ttaaattt ttgttaaatc	5340
agctcatttt tta	accaata ggccgaaatc	ggcaaaatcc ctt	tataaatc aaaagaatag	5400
accgagatag ggt	tgagtgt tgttccagtt	tggaacaaga gtc	ccactatt aaagaacgtg	5460
gactccaacg tca	aagggcg aaaaaccgtc	tatcagggcg atg	ggcccact acgtgaacca	5520
tcaccctaat caa	gtttttt ggggtcgagg	tgccgtaaag cac	ctaaatcg gaaccctaaa	5580
gggagccccc gat	ttagagc ttgacgggga	aagccggcga acg	gtggcgag aaaggaaggg	5640
aagaaagcga aag	gageggg egetagggeg	ctggcaagtg tag	geggteae getgegegta	5700
accaccacac ccg	ccgcgct taatgcgccg	ctacagggcg cgt	tcgcgcca ttcgccattc	5760
aggetgegea act	gttggga agggcgatcg	gtgegggeet ett	tcgctatt acgccagctg	5820
gcgaaagggg gat	gtgctgc aaggcgatta	. agttgggtaa cgc	ccagggtt ttcccagtca	5880
cgacgttgta aaa	cgacggc cagtgagcgc	gcgtaatacg act	tcactata gggcgaattg	5940
ggtaccgggc ccc	ccc			5956
210, CEO ID M	0.16			
<210> SEQ 1D N <211> LENGTH: !	570			
<212> IIPE: PR <213> ORGANISM	: Bacillus subtili	s		
<400> SEQUENCE	: 16			
Met Thr Lys Al.	a Thr Lys Glu Gln	Lys Ser Leu Val	l Lys Asn Arg Gly 15	
- Ala Glu Leu Va	l Val Asp Cvs Leu	Val Glu Gln Glv	v Val Thr His Val	
20	1 (al imp of 20a	25	30	
Phe Gly Ile Pro 35	o Gly Ala Lys Ile 40	Asp Ala Val Phe	e Asp Ala Leu Gln 45	
Asp Lys Gly Pro 50	o Glu Ile Ile Val 55	Ala Arg His Glu 60	u Gln Asn Ala Ala	
Phe Met Ala Gli 65	n Ala Val Gly Arg 70	Leu Thr Gly Lys 75	s Pro Gly Val Val 80	
Leu Val Thr Se	r Gly Pro Gly Ala 85	Ser Asn Leu Ala 90	a Thr Gly Leu Leu 95	
Thr Ala Asn Th: 10	r Glu Gly Asp Pro 0	Val Val Ala Leu 105	u Ala Gly Asn Val 110	
Ile Arg Ala As	n Ara Leu Lvs Ara	Thr His Gln Ser	r Leu Asp Asn Ala	
115	120		125	
Ala Leu Phe Gli 130	n Pro Ile Thr Lys 135	Tyr Ser Val Glu 140	u Val Gln Asp Val 0	
Lys Asn Ile Pro 145	o Glu Ala Val Thr 150	Asn Ala Phe Arg 155	g Ile Ala Ser Ala 160	
Gly Gln Ala Gly	y Ala Ala Phe Val 165	Ser Phe Pro Gln 170	n Asp Val Val Asn 175	
Glu Val Thr Asi	n Thr Lys Asn Val 0	Arg Ala Val Ala 185	a Ala Pro Lys Leu 190	
Gly Pro Ala Al	a Asp Asp Ala Ile	Ser Ala Ala Ile	e Ala Lys Ile Gln	
195	200		205	
Thr Ala Lys Lev 210	u Pro Val Val Leu 215	Val Gly Met Lys 220	s Gly Gly Arg Pro 0	
Glu Ala Ile Ly: 225	s Ala Val Arg Lys 230	Leu Leu Lys Lys 235	s Val Gln Leu Pro 240	
Phe Val Glu Th	r Tyr Gln Ala Ala 245	Gly Thr Leu Ser 250	r Arg Asp Leu Glu 255	

-continued

Asp	Gln	Tyr	Phe 260	Gly	Arg	Ile	Gly	Leu 265	Phe	Arg	Asn	Gln	Pro 270	Gly	Asp
Leu	Leu	Leu 275	Glu	Gln	Ala	Asp	Val 280	Val	Leu	Thr	Ile	Gly 285	Tyr	Asp	Pro
Ile	Glu 290	Tyr	Asp	Pro	Lys	Phe 295	Trp	Asn	Ile	Asn	Gly 300	Asp	Arg	Thr	Ile
Ile 305	His	Leu	Asp	Glu	Ile 310	Ile	Ala	Asp	Ile	Asp 315	His	Ala	Tyr	Gln	Pro 320
Asp	Leu	Glu	Leu	Ile 325	Gly	Asp	Ile	Pro	Ser 330	Thr	Ile	Asn	His	Ile 335	Glu
His	Asp	Ala	Val 340	Lys	Val	Glu	Phe	Ala 345	Glu	Arg	Glu	Gln	Lys 350	Ile	Leu
Ser	Asp	Leu 355	Lys	Gln	Tyr	Met	His 360	Glu	Gly	Glu	Gln	Val 365	Pro	Ala	Asp
Trp	Lys 370	Ser	Asp	Arg	Ala	His 375	Pro	Leu	Glu	Ile	Val 380	ГЛа	Glu	Leu	Arg
Asn 385	Ala	Val	Asp	Asp	His 390	Val	Thr	Val	Thr	Сув 395	Asp	Ile	Gly	Ser	His 400
Ala	Ile	Trp	Met	Ser 405	Arg	Tyr	Phe	Arg	Ser 410	Tyr	Glu	Pro	Leu	Thr 415	Leu
Met	Ile	Ser	Asn 420	Gly	Met	Gln	Thr	Leu 425	Gly	Val	Ala	Leu	Pro 430	Trp	Ala
Ile	Gly	Ala 435	Ser	Leu	Val	Lys	Pro 440	Gly	Glu	Lys	Val	Val 445	Ser	Val	Ser
Gly	Asp 450	Gly	Gly	Phe	Leu	Phe 455	Ser	Ala	Met	Glu	Leu 460	Glu	Thr	Ala	Val
Arg 465	Leu	Lys	Ala	Pro	Ile 470	Val	His	Ile	Val	Trp 475	Asn	Asp	Ser	Thr	Tyr 480
Asp	Met	Val	Ala	Phe 485	Gln	Gln	Leu	Lys	Lys 490	Tyr	Asn	Arg	Thr	Ser 495	Ala
Val	Asp	Phe	Gly 500	Asn	Ile	Asp	Ile	Val 505	Lys	Tyr	Ala	Glu	Ser 510	Phe	Gly
Ala	Thr	Gly 515	Leu	Arg	Val	Glu	Ser 520	Pro	Asp	Gln	Leu	Ala 525	Aab	Val	Leu
Arg	Gln 530	Gly	Met	Asn	Ala	Glu 535	Gly	Pro	Val	Ile	Ile 540	Aab	Val	Pro	Val
Asp 545	Tyr	Ser	Asp	Asn	Ile 550	Asn	Leu	Ala	Ser	Asp 555	Lys	Leu	Pro	Lys	Glu 560
Phe	Gly	Glu	Leu	Met 565	Lys	Thr	Lys	Ala	Leu 570						
<21) <21; <21; <21; <22; <22;	<pre>&lt;210&gt; SEQ ID NO 17 &lt;211&gt; LENGTH: 4531 &lt;212&gt; TYPE: DNA &lt;213&gt; ORGANISM: Artificial Sequence &lt;220&gt; FEATURE: &lt;223&gt; OTHER INFORMATION: pUC19 ILV2</pre>														
<40	)> SI	EQUEI	ICE :	17											
gat	cccad	ett ç	gaati	tgaa	ct ta	attai	tcat	t cta	atgad	ctta	atti	ttag	ccg t	catt	agttt
ttt	agaaa	aaa a	aata	tataa	aa gi	tata	cacga	a aaa	aaaat	tata	cato	gacta	acc g	gattt	ggtag
gcti	utcat ttqa	at a	ataaq	yaago tgtti	sa co tt to	yacta yaaat	aata	a ata g tqo	ataa	agt	ctq	tacq	utt (	acto	gaaaat gtaat

145

tcagtctgtt	gtctttcaac	ttctgggtca	aaatttatga	actcgtctag	accgctacca	300	
cctgccacca	ttggcaaaac	aggaactttt	ttatcaactt	ccacttcaag	caaaactggg	360	
cccttggtag	aaacgaattc	tttcaactta	gcgtccaatt	cctcttgctt	cttgactctt	420	
aaacctttta	aacccatagc	ctccgctagt	tttatgaaat	cagggttcaa	ttgatgtgtg	480	
tgggaataac	gatgttcgta	gaacagggat	tgccattgag	taaccatacc	ttgctcttca	540	
ttgttcaaaa	tcaaaatctt	cactggagta	ccagcttgaa	cggcagaact	caattccgtt	600	
agagtcatgt	taaaggatgc	gtcaccatca	atgtcaataa	ccaaagattc	tggctttgca	660	
acttgagcac	cgatggcggc	agggagaccg	taacccatcg	tacctaaacc	acctgatgtg	720	
atgaaagtat	gtggatttct	ccatgtccag	tgttgagcag	cccacatttg	atgttgcccc	780	
acacccgttg	taacaatgac	atgtcttcct	gtgtcgttgg	caaccttgga	tagtttcttt	840	
ataaccgtct	gtggtttaat	tttagatcct	ggagteteet	ccatataagc	gtatgggtat	900	
tccttcttcc	atttatttat	ttgagcaaac	cattcagacc	tctccttaac	tgggaaaatc	960	
tttgacatca	ttttgcccag	attggtcgta	gcatcacctt	ccactgctat	ttgagtttga	1020	
acaaccttgt	ttatgttttt	tggactaacc	tcgaaatgaa	taataccacc	tctaccctcg	1080	
gcagctgcac	gacgagcttc	tggagcgaat	ttagaaatat	taccagtgac	acggtcgtcg	1140	
aatctagcac	caactgcaat	tatcaagtcg	gcattttgca	ctgccaggtt	ggcagtagca	1200	
caaccgtgca	taccaagcat	atccaatgat	tttggatctt	cttggtcgaa	tgaacctaaa	1260	
ccttgtaaag	tagtggtgac	aggtatttga	gcacggtcac	ttaattcttt	tagtaatctt	1320	
ggaccatctg	catggtttaa	aataccagca	ccgacgtata	agacaggttt	ctttgccaag	1380	
ttgatcaaat	ctgctgcttt	attgatactt	tgcatgacaa	attcatcttg	tgcgcgactg	1440	
gttaattggt	ttagtgcgtt	tgatggaaga	gttgttttg	ttggaattgg	atttcttaag	1500	
atagctgctg	taacatcctt	tggtaaatcg	accaagactg	gtcccggtct	accgctcgtg	1560	
gcaatttcaa	aagcctcgtt	aatacgcaat	ggcaattctt	ccacggactt	gaccatgaca	1620	
ttccatttcg	tacaagatct	agaaatacca	acgacgtcag	cctcttggaa	agcatcagta	1680	
ccgatagcac	tagttgggac	ttgccctgta	aagacaacca	ttggaatccc	gtctgcaaag	1740	
gcatctgcca	ttggagtaac	gacattggtg	gcacctggcc	cagaagtaac	caagacaaca	1800	
cctggtttac	cagaagctct	ggcgtagcct	tctgccatgt	gaccggcacc	ttgttcgtgt	1860	
tttggaaggc	atgcaagctt	ggcgtaatca	tggtcatagc	tgtttcctgt	gtgaaattgt	1920	
tatccgctca	caattccaca	caacatacga	gccggaagca	taaagtgtaa	agcctggggt	1980	
gcctaatgag	tgagctaact	cacattaatt	gcgttgcgct	cactgcccgc	tttccagtcg	2040	
ggaaacctgt	cgtgccagct	gcattaatga	atcggccaac	gcgcggggag	aggcggtttg	2100	
cgtattgggc	gctcttccgc	ttcctcgctc	actgactcgc	tgcgctcggt	cgttcggctg	2160	
cggcgagcgg	tatcagctca	ctcaaaggcg	gtaatacggt	tatccacaga	atcaggggat	2220	
aacgcaggaa	agaacatgtg	agcaaaaggc	cagcaaaagg	ccaggaaccg	taaaaaggcc	2280	
gcgttgctgg	cgtttttcca	taggctccgc	ccccctgacg	agcatcacaa	aaatcgacgc	2340	
tcaagtcaga	ggtggcgaaa	cccgacagga	ctataaagat	accaggcgtt	tccccctgga	2400	
agctccctcg	tgcgctctcc	tgttccgacc	ctgccgctta	ccggatacct	gtccgccttt	2460	
ctcccttcgg	gaagcgtggc	gctttctcat	agctcacgct	gtaggtatct	cagttcggtg	2520	
taggtcgttc	gctccaagct	gggctgtgtg	cacgaacccc	ccgttcagcc	cgaccgctgc	2580	
gccttatccg	gtaactatcg	tcttgagtcc	aacccggtaa	gacacgactt	atcgccactg	2640	

147

#### -continued

gcagcagcca ctggtaacag gattagcaga gcgaggtatg taggcggtgc tacagagttc 2700 ttgaagtggt ggcctaacta cggctacact agaaggacag tatttggtat ctgcgctctg 2760 ctgaagccag ttaccttcgg aaaaagagtt ggtagctctt gatccggcaa acaaaccacc 2820 gctggtagcg gtggtttttt tgtttgcaag cagcagatta cgcgcagaaa aaaaggatct 2880 caagaagatc ctttgatctt ttctacgggg tctgacgctc agtggaacga aaactcacgt 2940 3000 taagggattt tggtcatgag attatcaaaa aggatcttca cctagatcct tttaaattaa aaatgaagtt ttaaatcaat ctaaagtata tatgagtaaa cttggtctga cagttaccaa 3060 tgettaatea gtgaggeace tateteageg atetgtetat ttegtteate eatagttgee 3120 tgactccccg tcgtgtagat aactacgata cgggagggct taccatctgg ccccagtgct 3180 gcaatgatac cgcgagaccc acgctcaccg gctccagatt tatcagcaat aaaccagcca 3240 gccggaaggg ccgagcgcag aagtggtcct gcaactttat ccgcctccat ccagtctatt 3300 aattqttqcc qqqaaqctaq aqtaaqtaqt tcqccaqtta ataqtttqcq caacqttqtt 3360 gccattgcta caggcatcgt ggtgtcacgc tcgtcgtttg gtatggcttc attcagctcc 3420 3480 qqttcccaac qatcaaqqcq aqttacatqa tcccccatqt tqtqcaaaaa aqcqqttaqc teetteqqte etceqateqt tqtcaqaaqt aaqttqqceq caqtqttate acteatqqtt 3540 atogcagcac tocataatte tettactote atoccateco taagatgett ttetgtgaet 3600 3660 qqtqaqtact caaccaaqtc attctqaqaa taqtqtatqc qqcqaccqaq ttqctcttqc ccggcgtcaa tacgggataa taccgcgcca catagcagaa ctttaaaagt gctcatcatt 3720 3780 ggaaaacgtt cttcggggcg aaaactctca aggatcttac cgctgttgag atccagttcg atgtaaccca ctcgtgcacc caactgatct tcagcatctt ttactttcac cagcgtttct 3840 gggtgagcaa aaacaggaag gcaaaatgcc gcaaaaaagg gaataagggc gacacggaaa 3900 tgttgaatac tcatactctt cctttttcaa tattattgaa gcatttatca gggttattgt 3960 ctcatgagcg gatacatatt tgaatgtatt tagaaaaata aacaaatagg ggttccgcgc 4020 acattteece gaaaagtgee acetgaegte taagaaacea ttattateat gaeattaace 4080 tataaaaata ggcgtatcac gaggcccttt cgtctcgcgc gtttcggtga tgacggtgaa 4140 aacctctgac acatgcagct cccggagacg gtcacagctt gtctgtaagc ggatgccggg 4200 agcagacaag cccgtcaggg cgcgtcagcg ggtgttggcg ggtgtcgggg ctggcttaac 4260 tatgeggeat cagageagat tgtaetgaga gtgeaceata tgeggtgtga aataeegeae 4320 agatgegtaa ggagaaaata eegcateagg egecattege eatteagget gegeaactgt 4380 tgggaagggc gatcggtgcg ggcctcttcg ctattacgcc agctggcgaa agggggatgt 4440 gctgcaaggc gattaagttg ggtaacgcca gggttttccc agtcacgacg ttgtaaaacg 4500 acggccagtg aattcgagct cggtacccgg g 4531 <210> SEO ID NO 18

<211> LENGTH: 6074 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: pLA59 ILV2-410 <400> SEQUENCE: 18 gatccgcatt gcggattacg tattctaatg ttcagtaccg ttcgtataat gtatgctata cgaagttatg cagattgtac tgagagtgca ccataccacc ttttcaattc atcattttt 148

60

149

ttttattctt	ttttttgatt	tcggtttcct	tgaaattttt	ttgattcggt	aatctccgaa	180	
cagaaggaag	aacgaaggaa	ggagcacaga	cttagattgg	tatatatacg	catatgtagt	240	
gttgaagaaa	catgaaattg	cccagtattc	ttaacccaac	tgcacagaac	aaaaacctgc	300	
aggaaacgaa	gataaatcat	gtcgaaagct	acatataagg	aacgtgctgc	tactcatcct	360	
agtcctgttg	ctgccaagct	atttaatatc	atgcacgaaa	agcaaacaaa	cttgtgtgct	420	
tcattggatg	ttcgtaccac	caaggaatta	ctggagttag	ttgaagcatt	aggtcccaaa	480	
atttgtttac	taaaaacaca	tgtggatatc	ttgactgatt	tttccatgga	gggcacagtt	540	
aagccgctaa	aggcattatc	cgccaagtac	aattttttac	tcttcgaaga	cagaaaattt	600	
gctgacattg	gtaatacagt	caaattgcag	tactctgcgg	gtgtatacag	aatagcagaa	660	
tgggcagaca	ttacgaatgc	acacggtgtg	gtgggcccag	gtattgttag	cggtttgaag	720	
caggcggcag	aagaagtaac	aaaggaacct	agaggccttt	tgatgttagc	agaattgtca	780	
tgcaagggct	ccctatctac	tggagaatat	actaagggta	ctgttgacat	tgcgaagagc	840	
gacaaagatt	ttgttatcgg	ctttattgct	caaagagaca	tgggtggaag	agatgaaggt	900	
tacgattggt	tgattatgac	acccggtgtg	ggtttagatg	acaagggaga	cgcattgggt	960	
caacagtata	gaaccgtgga	tgatgtggtc	tctacaggat	ctgacattat	tattgttgga	1020	
agaggactat	ttgcaaaggg	aagggatgct	aaggtagagg	gtgaacgtta	cagaaaagca	1080	
ggctgggaag	catatttgag	aagatgcggc	cagcaaaact	aaaaactgt	attataagta	1140	
aatgcatgta	tactaaactc	acaaattaga	gcttcaattt	aattatatca	gttattaccc	1200	
tatgcggtgt	gaaataccgc	acagatgcgt	aaggagaaaa	taccgcatca	ggaaattgta	1260	
aacgttaata	ttttgttaaa	attcgcgtta	aatttttgtt	aaatcagctc	atttttaac	1320	
caataggccg	aaatcggcaa	aatcccttat	aaatcaaaag	aatagaccga	gatagggttg	1380	
agtgttgttc	cagtttggaa	caagagtcca	ctattaaaga	acgtggactc	caacgtcaaa	1440	
gggcgaaaaa	ccgtctatca	gggcgatggc	ccactacgtg	aaccatcacc	ctaatcaaga	1500	
taacttcgta	taatgtatgc	tatacgaacg	gtaccagtga	tgatacaacg	agttagccaa	1560	
ggtgaattca	ctggccgtcg	ttttacaacg	tcgtgactgg	gaaaaccctg	gcgttaccca	1620	
acttaatcgc	cttgcagcac	atcccccttt	cgccagctgg	cgtaatagcg	aagaggcccg	1680	
caccgatcgc	ccttcccaac	agttgcgcag	cctgaatggc	gaatggcgcc	tgatgcggta	1740	
ttttctcctt	acgcatctgt	gcggtatttc	acaccgcata	tggtgcactc	tcagtacaat	1800	
ctgctctgat	gccgcatagt	taagccagcc	ccgacacccg	ccaacacccg	ctgacgcgcc	1860	
ctgacgggct	tgtctgctcc	cggcatccgc	ttacagacaa	gctgtgaccg	tctccgggag	1920	
ctgcatgtgt	cagaggtttt	caccgtcatc	accgaaacgc	gcgagacgaa	agggcctcgt	1980	
gatacgccta	tttttatagg	ttaatgtcat	gataataatg	gtttcttaga	cgtcaggtgg	2040	
cacttttcgg	ggaaatgtgc	gcggaacccc	tatttgttta	tttttctaaa	tacattcaaa	2100	
tatgtatccg	ctcatgagac	aataaccctg	ataaatgctt	caataatatt	gaaaaaggaa	2160	
gagtatgagt	attcaacatt	tccgtgtcgc	ccttattccc	tttttgcgg	cattttgcct	2220	
tcctgttttt	gctcacccag	aaacgctggt	gaaagtaaaa	gatgctgaag	atcagttggg	2280	
tgcacqaqtq	ggttacatcq	aactgqatct	caacagcggt	aagatcettq	agagttttcq	2340	
CCCCGAAGAA	cattttccaa	tgatgaggag	ttttaaadtt	ctactatata	acacaatatt	2400	
atcocatatt	dacaccada	aananceert	caatcaccac	atacactett	ctcageetco	2460	
attent	yacyccyyyc	aayaycaadt	aget th	acadadtatt	aaat	2400	
cttggttgag	tactcaccag	tcacagaaaa	gcatcttacg	gatggcatga	cagtaagaga	2520	

151

attatgcagt	gctgccataa	ccatgagtga	taacactgcg	gccaacttac	ttctgacaac	2580
gatcggagga	ccgaaggagc	taaccgcttt	tttgcacaac	atgggggatc	atgtaactcg	2640
ccttgatcgt	tgggaaccgg	agctgaatga	agccatacca	aacgacgagc	gtgacaccac	2700
gatgcctgta	gcaatggcaa	caacgttgcg	caaactatta	actggcgaac	tacttactct	2760
agcttcccgg	caacaattaa	tagactggat	ggaggcggat	aaagttgcag	gaccacttct	2820
gcgctcggcc	cttccggctg	gctggtttat	tgctgataaa	tctggagccg	gtgagcgtgg	2880
gtctcgcggt	atcattgcag	cactgggggcc	agatggtaag	ccctcccgta	tcgtagttat	2940
ctacacgacg	gggagtcagg	caactatgga	tgaacgaaat	agacagatcg	ctgagatagg	3000
tgcctcactg	attaagcatt	ggtaactgtc	agaccaagtt	tactcatata	tactttagat	3060
tgatttaaaa	cttcattttt	aatttaaaag	gatctaggtg	aagatccttt	ttgataatct	3120
catgaccaaa	atcccttaac	gtgagttttc	gttccactga	gcgtcagacc	ccgtagaaaa	3180
gatcaaagga	tcttcttgag	atcctttttt	tctgcgcgta	atctgctgct	tgcaaacaaa	3240
aaaaccaccg	ctaccagcgg	tggtttgttt	gccggatcaa	gagctaccaa	ctctttttcc	3300
gaaggtaact	ggcttcagca	gagcgcagat	accaaatact	gtccttctag	tgtagccgta	3360
gttaggccac	cacttcaaga	actctgtagc	accgcctaca	tacctcgctc	tgctaatcct	3420
gttaccagtg	gctgctgcca	gtggcgataa	gtcgtgtctt	accgggttgg	actcaagacg	3480
atagttaccg	gataaggcgc	agcggtcggg	ctgaacgggg	ggttcgtgca	cacagcccag	3540
cttggagcga	acgacctaca	ccgaactgag	atacctacag	cgtgagctat	gagaaagcgc	3600
cacgetteee	gaagggagaa	aggcggacag	gtatccggta	agcggcaggg	tcggaacagg	3660
agagcgcacg	agggagcttc	caggggggaaa	cgcctggtat	ctttatagtc	ctgtcgggtt	3720
tcgccacctc	tgacttgagc	gtcgattttt	gtgatgctcg	tcagggggggc	ggagcctatg	3780
gaaaaacgcc	agcaacgcgg	cctttttacg	gttcctggcc	ttttgctggc	cttttgctca	3840
catgttcttt	cctgcgttat	cccctgattc	tgtggataac	cgtattaccg	cctttgagtg	3900
agctgatacc	gctcgccgca	gccgaacgac	cgagcgcagc	gagtcagtga	gcgaggaagc	3960
ggaagagcgc	ccaatacgca	aaccgcctct	ccccgcgcgt	tggccgattc	attaatgcag	4020
ctggcacgac	aggtttcccg	actggaaagc	gggcagtgag	cgcaacgcaa	ttaatgtgag	4080
ttagctcact	cattaggcac	cccaggcttt	acactttatg	cttccggctc	gtatgttgtg	4140
tggaattgtg	agcggataac	aatttcacac	aggaaacagc	tatgaccatg	attacgccaa	4200
gcttgcatgc	cttccaaaac	acgaacaagg	tgccggtcac	atggcagaag	gctacgccag	4260
agcttctggt	aaaccaggtg	ttgtcttggt	tacttctggg	ccaggtgcca	ccaatgtcgt	4320
tactccaatg	gcagatgcct	ttgcagacgg	gattccaatg	gttgtcttta	cagggcaagt	4380
ctcaactagt	gctatcggta	ctgatgcttt	ccaagaggct	gacgtcgttg	gtatttctag	4440
atcttgtacg	aaatggaatg	tcatggtcaa	gtccgtggaa	gaattgccat	tgcgtattaa	4500
cgaggctttt	gaaattgcca	cgagcggtag	accgggacca	gtcttggtcg	atttaccaaa	4560
ggatgttaca	gcagctatct	taagaaatcc	aattccaaca	aaaacaactc	ttccatcaaa	4620
cgcactaaac	caattaacca	gtcgcgcaca	agatgaattt	gtcatgcaaa	gtatcaataa	4680
agcagcagat	ttgatcaact	tggcaaagaa	acctgtctta	tacgtcggtg	ctggtattt	4740
aaaccatgca	gatggtccaa	gattactaaa	agaattaagt	gaccgtgctc	aaatacctgt	4800
caccactact	ttacaaggtt	taggttcatt	cgaccaagaa	gatccaaaat	cattggatat	4860

153

### -continued

gcttggtatg tgcagttggt						
tgcagttggt	cacggttgtg	ctactgccaa	cctggcagtg	caaaatgccg	acttgataat	4920
	gctagattcg	acgaccgtgt	cactggtaat	atttctaaat	tcgctccaga	4980
agctcgtcgt	gcagctgccg	agggtagagg	tggtattatt	catttcgagg	ttagtccaaa	5040
aaacataaac	aaggttgttc	aaactcaaat	agcagtggaa	ggtgatgcta	cgaccaatct	5100
gggcaaaatg	atgtcaaaga	ttttcccagt	taaggagagg	tctgaatggt	ttgctcaaat	5160
aaataaatgg	aagaaggaat	acccatacgc	ttatatggag	gagactccag	gatctaaaat	5220
taaaccacag	acggttataa	agaaactatc	caaggttgcc	aacgacacag	gaagacatgt	5280
cattgttaca	acgggtgtgg	ggcaacatca	aatgtgggct	gctcaacact	ggacatggag	5340
aaatccacat	actttcatca	catcaggtgg	tttaggtacg	atgggttacg	gtctccctgc	5400
cgccatcggt	gctcaagttg	caaagccaga	atctttggtt	attgacattg	atggtgacgc	5460
atcctttaac	atgactctaa	cggaattgag	ttctgccgtt	caagctggta	ctccagtgaa	5520
gattttgatt	ttgaacaatg	aagagcaagg	tatggttact	caatggcaat	ccctgttcta	5580
cgaacatcgt	tattcccaca	cacatcaatt	gaaccctgat	ttcataaaac	tagcggaggc	5640
tatgggttta	aaaggtttaa	gagtcaagaa	gcaagaggaa	ttggacgcta	agttgaaaga	5700
attcgtttct	accaagggcc	cagttttgct	tgaagtggaa	gttgataaaa	aagttcctgt	5760
tttgccaatg	gtggcaggtg	gtagcggtct	agacgagttc	ataaattttg	acccagaagt	5820
tgaaagacaa	cagactgaat	tacgtcataa	gcgtacaggc	ggtaagcact	gaatttcaaa	5880
aacatttatt	tcaaaagcat	tttcagtaaa	aaatgcagac	tttattatta	tttaatcgtg	5940
cttcttatat	atgacattct	accaaatcgg	tagtcatgta	tattttttc	gtatatactt	6000
tatatattt	tttctaaaaa	actaatgacg	gctaaaatta	agtcatagat	gaataataag	6060
ttcaattcaa	gtgg					6074
	ID NO 19					
<210> SEQ <211> LENG <212> TYPE <213> ORGA <220> FEAT <223> OTHE	TH: 6827 : DNA NISM: Artif: URE: R INFORMATIC	icial Sequer DN: pLA59 il	nce LvB			
<pre>&lt;210&gt; SEQ &lt;211&gt; LENG &lt;212&gt; TYPE &lt;213&gt; ORGA &lt;220&gt; FEAT &lt;223&gt; OTHE &lt;400&gt; SEQU</pre>	ID NO 19 IH: 6827 : DNA NISM: Artif: URE: R INFORMATIC ENCE: 19	icial Sequer DN: pLA59 il	nce LvB			
<pre>&lt;210&gt; SEQ &lt;211&gt; LENG &lt;212&gt; TYPE &lt;213&gt; ORGA &lt;220&gt; FEAT &lt;223&gt; OTHE &lt;400&gt; SEQU cgcctacttg</pre>	ID NO 19 TH: 6827 : DNA NISM: Artif: URE: R INFORMATIO ENCE: 19 gcttcacata	icial Sequer DN: pLA59 il cgttgcatac	nce lvB gtcgatatag	ataataatga	taatgacagc	60
<pre>&lt;210&gt; SEQ &lt;211&gt; LENG &lt;212&gt; TYPE &lt;213&gt; ORGA &lt;220&gt; FEAT &lt;223&gt; OTHE &lt;400&gt; SEQU cgcctacttg aggattatcg</pre>	ID NO 19 TH: 6827 : DNA NISM: Artif: URE: R INFORMATI( ENCE: 19 gcttcacata taatacgtaa	icial Sequer DN: pLA59 il cgttgcatac tagttgaaaa	nce lvB gtcgatatag tctcaaaaat	ataataatga gtgtgggtca	taatgacagc ttacgtaaat	60 120
<pre>&lt;210&gt; SEQ &lt;211&gt; LENG &lt;212&gt; TYPE &lt;213&gt; ORGA &lt;220&gt; FEAT &lt;223&gt; OTHE &lt;400&gt; SEQU cgcctacttg aggattatcg aatgatagga</pre>	ID NO IS ID NO IS ID NA NISM: Artif: URE: R INFORMATIC ENCE: 19 gcttcacata taatacgtaa atgggattct	icial Sequer DN: pLA59 il cgttgcatac tagttgaaaa tctatttttc	nce lvB gtcgatatag tctcaaaaat ctttttccat	ataataatga gtgtgggtca tctagcagcc	taatgacagc ttacgtaaat gtcgggaaaa	60 120 180
<pre>&lt;210&gt; SEQ &lt;211&gt; LENG &lt;212&gt; TYPE &lt;213&gt; ORGA &lt;220&gt; FEAT &lt;223&gt; OTHE &lt;400&gt; SEQU cgcctacttg aggattatcg aatgatagga cgtggcatcc</pre>	ID NO 19 TH: 6827 : DNA NISM: Artif: URE: R INFORMATIO ENCE: 19 gcttcacata taatacgtaa atgggattct tctctttcgg	icial Sequer DN: pLA59 il cgttgcatac tagttgaaaa tctatttttc gctcaattgg	nce lvB gtcgatatag tctcaaaaat ctttttccat agtcacgctg	ataataatga gtgtgggtca tctagcagcc ccgtgagcat	taatgacagc ttacgtaaat gtcgggaaaa cctctctttc	60 120 180 240
<pre>&lt;210&gt; SEQ &lt;211&gt; LENG &lt;212&gt; TYPE &lt;213&gt; ORGA &lt;220&gt; FEAT &lt;223&gt; OTHE &lt;400&gt; SEQU cgcctacttg aggattatcg aatgatagga cgtggcatcc catatctaac</pre>	ID NO IS ID: 6827 : DNA NISM: Artif: URE: R INFORMATIO ENCE: 19 gcttcacata taatacgtaa atgggattct tctctttcgg aactgagcac	icial Sequer DN: pLA59 il cgttgcatac tagttgaaaa tctatttttc gctcaattgg gtaaccaatg	nce LvB gtcgatatag tctcaaaaat ctttttccat agtcacgctg gaaaagcatg	ataataatga gtgtgggtca tctagcagcc ccgtgagcat agcttagcgt	taatgacagc ttacgtaaat gtcgggaaaa cctctctttc tgctccaaaa	60 120 180 240 300
<pre>&lt;210&gt; SEQ &lt;211&gt; LENG &lt;212&gt; TYPE &lt;213&gt; ORGA &lt;220&gt; FEAT &lt;223&gt; OTHE &lt;400&gt; SEQU cgcctacttg aggattatcg aatgatagga cgtggcatcc catatctaac aagtattgga</pre>	ID NO IS TH: 6827 : DNA NISM: Artif: URE: R INFORMATIC ENCE: 19 gcttcacata taatacgtaa atggggattct tctctttcgg aactgagcac tggttaatac	icial Sequer DN: pLA59 il cgttgcatac tagttgaaaa tctattttc gctcaattgg gtaaccaatg catttgtctg	nce lvB gtcgatatag tctcaaaaat ctttttccat agtcacgctg gaaaagcatg ttctcttctg	ataataatga gtgtgggtca tctagcagcc ccgtgagcat agcttagcgt actttgactc	taatgacagc ttacgtaaat gtcgggaaaa cctctctttc tgctccaaaa ctcaaaaaaa	60 120 180 240 300 360
<pre>&lt;210&gt; SEQ &lt;211&gt; LENG &lt;212&gt; TYPE &lt;213&gt; ORGA &lt;220&gt; FEAT &lt;220&gt; FEAT &lt;223&gt; OTHE &lt;400&gt; SEQU cgcctacttg aggattatcg aatgatagga cgtggcatcc catatctaac aagtattgga aaaaatctac</pre>	ID NO IS ID: 6827 : DNA NISM: Artif: URE: R INFORMATIO ENCE: 19 gcttcacata taatacgtaa atgggattct tctctttcgg aactgagcac tggttaatac aatcaacaga	icial Sequer DN: pLA59 il cgttgcatac tagttgaaaa tctatttttc gctcaattgg gtaaccaatg catttgtctg tcgcttcaat	nce lvB gtcgatatag tctcaaaaat ctttttccat agtcacgctg gaaaagcatg ttctcttctg tacgccctca	ataataatga gtgtgggtca tctagcagcc ccgtgagcat agcttagcgt actttgactc caaaacttt	taatgacagc ttacgtaaat gtcgggaaaa cctctcttc tgctccaaaa ctcaaaaaaa tttccttctt	60 120 180 240 300 360
<pre>&lt;210&gt; SEQ &lt;211&gt; LENG &lt;212&gt; TYPE &lt;213&gt; ORGA &lt;220&gt; FEAT &lt;223&gt; OTHE &lt;400&gt; SEQU cgcctacttg aggattatcg aatgatagga cgtggcatcc catatctaac aagtattgga aaaaatctac cttcgcccac</pre>	ID NO IS TH: 6827 : DNA NISM: Artif: URE: R INFORMATIO ENCE: 19 gcttcacata taatacgtaa atgggattct tctctttcgg aactgagcac tggttaatac aatcaacaga gttaaatttt	icial Sequer DN: pLA59 il cgttgcatac tagttgaaaa tctatttttc gctcaattgg gtaaccaatg catttgtctg tcgcttcaat atccctcatg	nce lvB gtcgatatag tctcaaaaat ctttttccat agtcacgctg gaaaagcatg ttctcttctg tacgccctca ttgtctaacg	ataataatga gtgtgggtca tctagcagcc ccgtgagcat agcttagcgt actttgactc caaaaacttt gatttctgca	taatgacagc ttacgtaaat gtcgggaaaa cctctcttc tgctccaaaa ctcaaaaaaa tttccttctt	60 120 180 240 300 360 420
<pre>&lt;210&gt; SEQ &lt;211&gt; LENG &lt;212&gt; TYPE &lt;213&gt; ORGA &lt;220&gt; FEAT &lt;223&gt; OTHE &lt;400&gt; SEQU cgcctacttg aggattatcg aatgatagga cgtggcatcc catatctaac aagtattgga aaaaatctac cttcgcccac tataaaaaga</pre>	TH: 6827 : DNA NISM: Artif: URE: R INFORMATIO ENCE: 19 gcttcacata taatacgtaa atgggattct tctctttcgg aactgagcacc tggttaatac aatcaacaga gttaaatttt caaagacata	icial Sequer DN: pLA59 il cgttgcatac tagttgaaaa tctatttttc gctcaattgg gtaaccaatg catttgtctg tcgcttcaat atccctcatg atacttctct	nce lvB gtcgatatag tctcaaaaat ctttttccat agtcacgctg gaaaagcatg ttctcttctg tacgccctca ttgtctaacg atcaatttca	ataataatga gtgtgggtca tctagcagcc ccgtgagcat agcttagcgt actttgactc caaaaacttt gatttctgca gttattgttc	taatgacagc ttacgtaaat gtcgggaaaa cctctctttc tgctccaaaa ctcaaaaaaa tttccttctt cttgatttat ttccttgcgt	60 120 180 240 300 360 420 480 540
<pre>&lt;210&gt; SEQ &lt;211&gt; LENG &lt;211&gt; LENG &lt;212&gt; TYPE &lt;213&gt; ORGA &lt;220&gt; FEAT &lt;223&gt; OTHE &lt;400&gt; SEQU cgcctacttg aggattatcg aatgatagga cgtggcatcc catatctaac aagtattgga aaaaatctac cttcgcccac tataaaaaga tattcttctg</pre>	ID NO IS ID: 6827 : DNA NISM: Artif: URE: R INFORMATIO ENCE: 19 gcttcacata taatacgtaa atgggattct tctctttcgg aactgagcac tggttaatac aatcaacaga gttaaatttt caaagacata ttcttctttt	icial Sequer DN: pLA59 il cgttgcatac tagttgaaaa tctatttttc gctcaattgg gtaaccaatg catttgtctg tcgcttcaat atccctcatg atacttctct tcttttgtca	nce lvB gtcgatatag tctcaaaaat ctttttccat agtcacgctg gaaaagcatg ttctcttctg tacgccctca ttgtctaacg atcaatttca tatataacca	ataataatga gtgtgggtca tctagcagcc ccgtgagcat agcttagcgt actttgactc caaaaacttt gatttctgca gttattgttc taaccaagta	taatgacagc ttacgtaaat gtcgggaaaa cctctctttc tgctccaaaa ctcaaaaaaa tttccttctt cttgatttat ttccttgcgt atacatattc	60 120 180 240 300 360 420 480 540
<pre>&lt;210&gt; SEQ &lt;211&gt; LENG &lt;211&gt; LENG &lt;212&gt; TYPE &lt;213&gt; ORGA &lt;220&gt; FEAT &lt;223&gt; OTHE &lt;400&gt; SEQU cgcctacttg aggattatcg aatgatagga cgtggcatcc catatctaac aagtattgga aaaaatctac cttcgcccac tataaaaaga tattcttctg aaggtaccat</pre>	ID NO 19 TH: 6827 : DNA NISM: Artif: URE: R INFORMATIO ENCE: 19 gcttcacata taatacgtaa atgggattct tctctttcgg aactgagcac tggttaatac aatcaacaga gttaaatttt caaagacata ttcttctttt ggcaagttcg	icial Sequer DN: pLA59 il cgttgcatac tagttgaaaa tctatttttc gctcaattgg gtaaccaatg catttgtctg tcgcttcaat atccctcatg atacttctct tcttttgtca ggcacaacat	nce lvB gtcgatatag tctcaaaaat ctttttccat agtcacgctg gaaaagcatg ttctcttctg tacgccctca ttgtctaacg atcaatttca tatataacca cgacgcgtaa	ataataatga gtgtgggtca tctagcagcc ccgtgagcat agcttagcgt actttgactc caaaaacttt gatttctgca gttattgttc taaccaagta gcgctttacc	taatgacagc ttacgtaaat gtcgggaaaa cctctctttc tgctccaaaa cttcaaaaaa tttccttctt cttgatttat ttccttgcgt atacatattc ggcgcagaat	60 120 180 240 300 360 420 480 540 600
<pre>&lt;210&gt; SEQ &lt;211&gt; LENG &lt;211&gt; LENG &lt;212&gt; TYPE &lt;213&gt; ORGA &lt;220&gt; FEAT &lt;223&gt; OTHE &lt;400&gt; SEQU cgcctacttg aggattatcg aggattatcg aatgatagga cgtggcatcc catatctaac aagtattgga aaaaatctac cttcgcccac tataaaaaga tattcttctg aaggtaccat ttatcgttca</pre>	TH: 6827 : DNA NISM: Artif: URE: R INFORMATIO ENCE: 19 gcttcacata taatacgtaa atgggattct tctctttcgg aactgagcac tggttaatac aatcaacaga gttaaatttt caaagacata ttcttctttt ggcaagttcg tttcctggaa	icial Sequer DN: pLA59 il cgttgcatac tagttgaaaa tctatttttc gctcaattgg gtaaccaatg cattgtctg tcgcttcaat atccctcatg atacttctct tcttttgtca ggcacaacat cagcagggca	nce lvB gtcgatatag tctcaaaaat ctttttccat agtcacgctg gaaaagcatg ttctcttctg tacgccctca ttgtctaacg atcaatttca tatataacca cgacgcgtaa ttaagattgt	ataataatga gtgtgggtca tctagcagcc ccgtgagcat agcttagcgt actttgactc caaaaacttt gatttctgca gttattgttc taaccaagta gcgctttacc gacaggcatt	taatgacagc ttacgtaaat gtcgggaaaa cctctotttc tgctccaaaa ctcaaaaaaa tttccttctt cttgatttat ttccttgcgt atacatattc ggcgcagaat ccgggcggtt	60 120 180 240 300 360 420 480 540 600 660
<pre>&lt;210&gt; SEQ &lt;211&gt; LENG &lt;211&gt; LENG &lt;212&gt; TYPE &lt;213&gt; ORGA &lt;220&gt; FEAT &lt;223&gt; OTHE &lt;400&gt; SEQU cgcctacttg aggattatcg aatgatagga cgtggcatcc catatctaac aagtattgga aaaaatctac cttcgcccac tataaaaaga tattcttctg aaggtaccat ttatcgttca ctatcctgcc</pre>	ID NO 19 TH: 6827 : DNA NISM: Artif: URE: R INFORMATIO ENCE: 19 gcttcacata taatacgtaa atgggattct tctctttcgg aactgagcac tggttaatac aatcaacaga gttaaatttt caaagacata ttcttctttt ggcaagttcg tttcctggaa tgtttacgat	icial Sequer DN: pLA59 il cgttgcatac tagttgaaaa tctatttttc gctcaattgg gtaaccaatg cattgtctg tcgcttcaat atccctcatg atacttctct tcttttgtca ggcacaacat cagcagggca gccttaagcc	nce lvB gtcgatatag tctcaaaaat ctttttccat agtcacgctg gaaaagcatg ttctcttctg tacgccctca ttgtctaacg atcaatttca cgacgcgtaa ttaagattgt aaagcacgca	ataataatga gtgtgggtca tctagcagcc ccgtgagcat agcttagcgt actttgactc caaaaacttt gatttctgca gttattgttc taaccaagta gcgctttacc gacaggcatt aatccgccat	taatgacagc ttacgtaaat gtcgggaaaa cctctcttc tgctccaaaa tttccttctt cttgatttat ttccttgcgt atacatattc ggcgcagaat ccgggcggtt attctgccc	<ul> <li>60</li> <li>120</li> <li>180</li> <li>240</li> <li>300</li> <li>360</li> <li>420</li> <li>480</li> <li>540</li> <li>600</li> <li>660</li> <li>720</li> <li>780</li> </ul>

155

cggtctgtat	ggcctgtagc	ggaccgggtg	cgactaacct	ggtgaccgcc	attgccgatg	900
cgcggctgga	ctccatcccg	ctgatttgca	tcactggtca	ggtteeegee	tcgatgatcg	960
gcaccgacgc	cttccaggaa	gtggacacct	acggcatctc	tatccccatc	accaaacaca	1020
actatctggt	cagacatatc	gaagaactcc	cgcaggtcat	gagcgatgcc	ttccgcattg	1080
cgcaatcagg	ccgcccaggc	ccggtgtgga	tagacattcc	taaggatgtg	caaacggcag	1140
tttttgagat	tgaaacacag	cccgctatgg	cagaaaaagc	cgccgccccc	gcctttagcg	1200
aagaaagcat	tcgtgacgca	gcggcgatga	ttaacgctgc	caaacgcccg	gtgctttatc	1260
tgggcggcgg	tgtgatcaat	gcgcccgcac	gggtgcgtga	actggcggag	aaagcgcaac	1320
tgcctaccac	catgacttta	atggcgctgg	gcatgttgcc	aaaagcgcat	ccgttgtcgc	1380
tgggtatgct	ggggatgcac	ggcgtgcgca	gcaccaacta	tattttgcag	gaggcggatt	1440
tgttgatagt	gctcggtgcg	cgttttgatg	accgggcgat	tggcaaaacc	gagcagttct	1500
gtccgaatgc	caaaatcatt	catgtcgata	tcgaccgtgc	agagctgggt	aaaatcaagc	1560
agccgcacgt	ggcgattcag	gcggatgttg	atgacgtgct	ggcgcagttg	atcccgctgg	1620
tggaagcgca	accgcgtgca	gagtggcacc	agttggtagc	ggatttgcag	cgtgagtttc	1680
cgtgtccaat	cccgaaagcg	tgcgatccgt	taagccatta	cggcctgatc	aacgccgttg	1740
ccgcctgtgt	cgatgacaat	gcaattatca	ccaccgacgt	tggtcagcat	cagatgtgga	1800
ccgcgcaagc	ttatccgctc	aatcgcccac	gccagtggct	gacctccggt	gggctgggca	1860
cgatgggttt	tggcctgcct	gcggcgattg	gcgctgcgct	ggcgaacccg	gatcgcaaag	1920
tgttgtgttt	ctccggcgac	ggcagcctga	tgatgaatat	tcaggagatg	gcgaccgcca	1980
gtgaaaatca	gctggatgtc	aaaatcattc	tgatgaacaa	cgaagcgctg	gggctggtgc	2040
atcagcaaca	gagtetgtte	tacgagcaag	gcgtttttgc	cgccacctat	ccgggcaaaa	2100
tcaactttat	gcagattgcc	gccggattcg	gcctcgaaac	ctgtgatttg	aataacgaag	2160
ccgatccgca	ggcttcattg	caggaaatca	tcaatcgccc	tggcccggcg	ctgatccatg	2220
tgcgcattga	tgccgaagaa	aaagtttacc	cgatggtgcc	gccaggtgcg	gcgaatactg	2280
aaatggtggg	ggaataagcg	gccgcgttaa	ttcaaattaa	ttgatatagt	tttttaatga	2340
gtattgaatc	tgtttagaaa	taatggaata	ttattttat	ttatttattt	atattattgg	2400
tcggctcttt	tcttctgaag	gtcaatgaca	aaatgatatg	aaggaaataa	tgatttctaa	2460
aattttacaa	cgtaagatat	ttttacaaaa	gcctagctca	tcttttgtca	tgcactattt	2520
tactcacgct	tgaaattaac	ggccagtcca	ctgcggagtc	atttcaaagt	catcctaatc	2580
gatctatcgt	ttttgatagc	tcattttgga	gttcgcggga	tccgcattgc	ggattacgta	2640
ttctaatgtt	cagtaccgtt	cgtataatgt	atgctatacg	aagttatgca	gattgtactg	2700
agagtgcacc	ataccacctt	ttcaattcat	cattttttt	ttattctttt	ttttgatttc	2760
ggtttccttg	aaatttttt	gattcggtaa	tctccgaaca	gaaggaagaa	cgaaggaagg	2820
agcacagact	tagattggta	tatatacgca	tatgtagtgt	tgaagaaaca	tgaaattgcc	2880
cagtattctt	aacccaactg	cacagaacaa	aaacctgcag	gaaacgaaga	taaatcatgt	2940
cgaaagctac	atataaggaa	cgtgctgcta	ctcatcctag	tcctgttgct	gccaagctat	3000
ttaatatcat	gcacgaaaag	caaacaaact	tgtgtgcttc	attggatgtt	cgtaccacca	3060
aggaattact	ggagttagtt	gaagcattag	gtcccaaaat	ttgtttacta	aaaacacatg	3120
tggatatctt	gactgatttt	tccatggagg	gcacagttaa	gccgctaaag	gcattatccg	3180

157

#### -continued

				0011011	ruou	
ccaagtacaa	tttttactc	ttcgaagaca	gaaaatttgc	tgacattggt	aatacagtca	3240
aattgcagta	ctctgcgggt	gtatacagaa	tagcagaatg	ggcagacatt	acgaatgcac	3300
acggtgtggt	gggcccaggt	attgttagcg	gtttgaagca	ggcggcagaa	gaagtaacaa	3360
aggaacctag	aggccttttg	atgttagcag	aattgtcatg	caagggctcc	ctatctactg	3420
gagaatatac	taagggtact	gttgacattg	cgaagagcga	caaagatttt	gttatcggct	3480
ttattgctca	aagagacatg	ggtggaagag	atgaaggtta	cgattggttg	attatgacac	3540
ccggtgtggg	tttagatgac	aagggagacg	cattgggtca	acagtataga	accgtggatg	3600
atgtggtctc	tacaggatct	gacattatta	ttgttggaag	aggactattt	gcaaagggaa	3660
gggatgctaa	ggtagagggt	gaacgttaca	gaaaagcagg	ctgggaagca	tatttgagaa	3720
gatgcggcca	gcaaaactaa	aaaactgtat	tataagtaaa	tgcatgtata	ctaaactcac	3780
aaattagagc	ttcaatttaa	ttatatcagt	tattacccta	tgcggtgtga	aataccgcac	3840
agatgcgtaa	ggagaaaata	ccgcatcagg	aaattgtaaa	cgttaatatt	ttgttaaaat	3900
tcgcgttaaa	tttttgttaa	atcagctcat	tttttaacca	ataggccgaa	atcggcaaaa	3960
tcccttataa	atcaaaagaa	tagaccgaga	tagggttgag	tgttgttcca	gtttggaaca	4020
agagtccact	attaaagaac	gtggactcca	acgtcaaagg	gcgaaaaacc	gtctatcagg	4080
gcgatggccc	actacgtgaa	ccatcaccct	aatcaagata	acttcgtata	atgtatgcta	4140
tacgaacggt	accagtgatg	atacaacgag	ttagccaagg	tgaattcact	ggccgtcgtt	4200
ttacaacgtc	gtgactggga	aaaccctggc	gttacccaac	ttaatcgcct	tgcagcacat	4260
ccccctttcg	ccagctggcg	taatagcgaa	gaggcccgca	ccgatcgccc	ttcccaacag	4320
ttgcgcagcc	tgaatggcga	atggcgcctg	atgcggtatt	ttctccttac	gcatctgtgc	4380
ggtatttcac	accgcatatg	gtgcactctc	agtacaatct	gctctgatgc	cgcatagtta	4440
agccagcccc	gacacccgcc	aacacccgct	gacgcgccct	gacgggcttg	tctgctcccg	4500
gcatccgctt	acagacaagc	tgtgaccgtc	tccgggagct	gcatgtgtca	gaggttttca	4560
ccgtcatcac	cgaaacgcgc	gagacgaaag	ggcctcgtga	tacgcctatt	tttataggtt	4620
aatgtcatga	taataatggt	ttcttagacg	tcaggtggca	cttttcgggg	aaatgtgcgc	4680
ggaaccccta	tttgtttatt	tttctaaata	cattcaaata	tgtatccgct	catgagacaa	4740
taaccctgat	aaatgcttca	ataatattga	aaaaggaaga	gtatgagtat	tcaacatttc	4800
cgtgtcgccc	ttattccctt	ttttgcggca	ttttgccttc	ctgtttttgc	tcacccagaa	4860
acgctggtga	aagtaaaaga	tgctgaagat	cagttgggtg	cacgagtggg	ttacatcgaa	4920
ctggatctca	acagcggtaa	gatccttgag	agttttcgcc	ccgaagaacg	ttttccaatg	4980
atgagcactt	ttaaagttct	gctatgtggc	gcggtattat	cccgtattga	cgccgggcaa	5040
gagcaactcg	gtcgccgcat	acactattct	cagaatgact	tggttgagta	ctcaccagtc	5100
acagaaaagc	atcttacgga	tggcatgaca	gtaagagaat	tatgcagtgc	tgccataacc	5160
atgagtgata	acactgcggc	caacttactt	ctgacaacga	tcggaggacc	gaaggagcta	5220
accgcttttt	tgcacaacat	gggggatcat	gtaactcgcc	ttgatcgttg	ggaaccggag	5280
ctgaatgaag	ccataccaaa	cgacgagcgt	gacaccacga	tgcctgtagc	aatggcaaca	5340
acqttqcqca	aactattaac	tggcgaacta	cttactctag	cttcccqqca	acaattaata	5400
gactgeater	aggoggataa	agttgcagga	ccacttetee	actegadet	teegaetaae	5460
taatt+++4	ctrateetc	tagaggggg	gaggatgggt	ctococotet	cattocadoa	5520
ataaa	cigaladalC	at an art of	gagegegegge	and an area	callycayda	5520
ccggggccag	atggtaagcc	ctcccgtatc	gtagttatct	acacgacggg	gagtcaggca	5580

159

# 160

actatggatg	aacgaaatag	acagatcgct	gagataggtg	cctcactgat	taagcattgg	5640
taactgtcag	accaagttta	ctcatatata	ctttagattg	atttaaaact	tcatttttaa	5700
tttaaaagga	tctaggtgaa	gatcctttt	gataatctca	tgaccaaaat	cccttaacgt	5760
gagttttcgt	tccactgagc	gtcagacccc	gtagaaaaga	tcaaaggatc	ttcttgagat	5820
cctttttttc	tgcgcgtaat	ctgctgcttg	caaacaaaaa	aaccaccgct	accagcggtg	5880
gtttgtttgc	cggatcaaga	gctaccaact	ctttttccga	aggtaactgg	cttcagcaga	5940
gcgcagatac	caaatactgt	ccttctagtg	tagccgtagt	taggccacca	cttcaagaac	6000
tctgtagcac	cgcctacata	cctcgctctg	ctaatcctgt	taccagtggc	tgctgccagt	6060
ggcgataagt	cgtgtcttac	cgggttggac	tcaagacgat	agttaccgga	taaggcgcag	6120
cggtcgggct	gaacggggggg	ttcgtgcaca	cagcccagct	tggagcgaac	gacctacacc	6180
gaactgagat	acctacagcg	tgagctatga	gaaagcgcca	cgcttcccga	agggagaaag	6240
gcggacaggt	atccggtaag	cggcagggtc	ggaacaggag	agcgcacgag	ggagetteca	6300
gggggaaacg	cctggtatct	ttatagtcct	gtcgggtttc	gccacctctg	acttgagcgt	6360
cgatttttgt	gatgctcgtc	agggggggcgg	agcctatgga	aaaacgccag	caacgcggcc	6420
tttttacggt	tcctggcctt	ttgctggcct	tttgctcaca	tgttctttcc	tgcgttatcc	6480
cctgattctg	tggataaccg	tattaccgcc	tttgagtgag	ctgataccgc	tcgccgcagc	6540
cgaacgaccg	agcgcagcga	gtcagtgagc	gaggaagcgg	aagagcgccc	aatacgcaaa	6600
ccgcctctcc	ccgcgcgttg	gccgattcat	taatgcagct	ggcacgacag	gtttcccgac	6660
tggaaagcgg	gcagtgagcg	caacgcaatt	aatgtgagtt	agctcactca	ttaggcaccc	6720
caggetttae	actttatgct	tccggctcgt	atgttgtgtg	gaattgtgag	cggataacaa	6780
tttcacacag	gaaacagcta	tgaccatgat	tacgccaagc	ttgcatg		6827
<210> SEQ : <211> LENG' <212> TYPE <213> ORGAN <400> SEOU	ID NO 20 FH: 1713 : DNA NISM: Lacto ENCE: 20	coccus lact:	is			
ctataaatca	gtcacacage	cttcactqqc	taaacacata	agtttagcaa	atttagcaag	60
aaccccacqc	gttgctttag	gaggaggttt	ttgataattg	gcacgtcgtc	tagcaatctc	120
atcatcagca	accttcaaac	taatgetgtt	attgaccgca	tcaatctcga	taatatcatc	180
atcttcgacc	aagccaatca	atccqccctc	aaccocttco	qqqacaatqt	ggccaaccac	240
aaagccatgt	gtgccaccag	agaaacgtcc	gtcagtaatc	aaagcacaag	attttccaag	300
acctoctcca	atcaaggetg	aggtcggttt	taacatttct	ggcattcctg	aacctccaac	360
tggaccgata	ttacqqatqa	cagcgacatc	tectacatae	aaacgacctg	actcaattcc	420
gtcaataaaa	tattattcac	catcaaagag	acqqqccqtt	cctttaaaaa	attegeette	480
ttttaaaaaa	otttttaaao	aaaaaaaaaa	ttagagggegee	ttagaataaa	gaatttggaa	E40
	attttgcca	cagaaccacc	Lucaycaaya	Ligicalaaa	gaatttytaa	540
augacetgtt	yctttgattg	yattttcaag	ıggtegeata	acgtettgge	LEECAAAGEC	600
caaatccagt	gctgttttaa	cattttcagc	tagagtttta	ccagtaaccg	tcaaacagtc	660
gccgtgaagc	ttgccctctt	tgagcaaata	tttcaaaaca	gcaggcacgc	caccaatttt	720
gtgcaaatct	tccaccatgt	acttgccact	tggcttgaag	tcgccaagca	caggcgtgac	780
atcggaaata	cgttgaaaat	catcttgcgt	aatttcgaca	ccgatggcat	tagccatggc	840

-continued

161

aatgatatga agcacagcat tagttgaacc accgagaacc atgacgatcg ttatggcatt	900
ttcaaaagct tccttggtca tgatatcgct tggcttgatg tctttttcta gcaaattttt	960
gattgecaga ccaattteat cacattegte ttettttet tgaetgaeeg etggatttga	1020
agatgaataa ggtaaactca ttcctaaagt ttcaatagca gcagccaggg tattggcagt	1080
gtacatteet eegcaageae ettgteetgg aatggeattg eaaataaege eatgataate	1140
ttcatcagaa atatttccag tgattttttg ccccagagct tcaaaggccg aaacaatatt	1200
taacttttcg cctttgtatt cgccatgttc aatcgttcca ccataaacca taattgacgg	1260
acgattgagc cgagccatac cgataattga gcctggcata tttttgtcac aaccgggaac	1320
agcgacaatt gcatcgtaat attcggcacc agcgttggtt tcaatgctgt cagcaataac	1380
ttcacgactg accaaggaat atctcatgcc aagttttccg ttggcaatcc catcagaaac	1440
cccaatcgtg tgaaattgaa gtccaatcag tccgtcagtt tgattgacag aatttttgat	1500
tttactacca agtgtcccca agtgcatgtt acagggattt ccgtcccaat ccatgctgac	1560
gatteegaee tgagettttt tgaagteete atetttaaaa eeaatgeeat agtaeatgge	1620
ttgggtcgct ggctgtgttg ggtcttgtgt caatgttttt gaatacttat tgagctctat	1680
tgattcaact tttccgttat atttgaattc cat	1713
<210> SEQ ID NO 21 <211> LENGTH: 570 <212> TYPE: PRT <213> ORGANISM: Lactococcus lactis	
<400> SEQUENCE: 21	
Met Glu Phe Lys Tyr Asn Gly Lys Val Glu Ser Ile Glu Leu Asn Lys 1 5 10 15	
Tyr Ser Lys Thr Leu Thr Gln Asp Pro Thr Gln Pro Ala Thr Gln Ala 20 25 30	
Met Tyr Tyr Gly Ile Gly Phe Lys Asp Glu Asp Phe Lys Lys Ala Gln 35 40 45	
Val Gly Ile Val Ser Met Asp Trp Asp Gly Asn Pro Cys Asn Met His 50 55 60	
Leu Gly Thr Leu Gly Ser Lys Ile Lys Asn Ser Val Asn Gln Thr Asp 65 70 75 80	
Gly Leu Ile Gly Leu Gln Phe His Thr Ile Gly Val Ser Asp Gly Ile	
85 90 95	
Ala Ash Giy Lys Leu Giy Met Arg Tyr Ser Leu Val Ser Arg Glu Val 100 105 110	
Ile Ala Asp Ser Ile Glu Thr Asn Ala Gly Ala Glu Tyr Tyr Asp Ala 115 120 125	
Ile Val Ala Val Pro Gly Cys Asp Lys Asn Met Pro Gly Ser Ile Ile 130 135 140	
Gly Met Ala Arg Leu Asn Arg Pro Ser Ile Met Val Tyr Gly Gly Thr	
145 150 155 160	
Ile Glu His Gly Glu Tyr Lys Gly Glu Lys Leu Asn Ile Val Ser Ala 165 170 175	
Phe Glu Ala Leu Gly Gln Lys Ile Thr Gly Asn Ile Ser Asp Glu Asp 180 185 190	
Tyr His Gly Val Ile Cys Asn Ala Ile Pro Gly Gln Gly Ala Cys Gly 195 200 205	
Gly Met Tyr Thr Ala Asn Thr Leu Ala Ala Ala Ile Glu Thr Leu Gly 210 215 220	

Met Ser Leu Pro Tyr Ser Ser Ser Asn Pro Ala Val Ser Gln Glu Lys Glu Asp Glu Cys Asp Glu Ile Gly Leu Ala Ile Lys Asn Leu Leu Glu Lys Asp Ile Lys Pro Ser Asp Ile Met Thr Lys Glu Ala Phe Glu Asn Ala Ile Thr Ile Val Met Val Leu Gly Gly Ser Thr Asn Ala Val Leu His Ile Ile Ala Met Ala Asn Ala Ile Gly Val Glu Ile Thr Gln Asp Asp Phe Gln Arg Ile Ser Asp Val Thr Pro Val Leu Gly Asp Phe Lys Pro Ser Gly Lys Tyr Met Val Glu Asp Leu His Lys Ile Gly Gly Val Pro Ala Val Leu Lys Tyr Leu Leu Lys Glu Gly Lys Leu His Gly Asp Cys Leu Thr Val Thr Gly Lys Thr Leu Ala Glu Asn Val Lys Thr Ala Leu Asp Leu Asp Phe Glu Ser Gln Asp Ile Met Arg Pro Leu Glu Asn Pro Ile Lys Ala Thr Gly His Leu Gln Ile Leu Tyr Gly Asn Leu Ala Glu Gly Gly Ser Val Ala Lys Ile Ser Gly Lys Glu Gly Glu Phe Phe Lys Gly Thr Ala Arg Val Phe Asp Gly Glu Gln His Phe Ile Asp Gly Ile Glu Ser Gly Arg Leu His Ala Gly Asp Val Ala Val Ile Arg Asn Ile Gly Pro Val Gly Gly Pro Gly Met Pro Glu Met Leu Lys Pro Thr Ser Ala Leu Ile Gly Ala Gly Leu Gly Lys Ser Cys Ala Leu Ile Thr Asp Gly Arg Phe Ser Gly Gly Thr His Gly Phe Val Val Gly His Ile Val Pro Glu Ala Val Glu Gly Gly Leu Ile Gly Leu Val Glu Asp Asp Asp Ile Ile Glu Ile Asp Ala Val Asn Asn Ser Ile Ser Leu Lys Val Ala Asp Asp Glu Ile Ala Arg Arg Arg Ala Asn Tyr Gln Lys Pro Ala Pro Lys Ala Thr Arg Gly Val Leu Ala Lys Phe Ala Lys Leu Thr Arg Pro Ala Ser Glu Gly Cys Val Thr Asp Leu <210> SEQ ID NO 22 <211> LENGTH: 12298 <212> TYPE: DNA

<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: pLH804 L2V4

<400> SEQUENCE: 22

tcccattacc gacatttggg cgctatacgt gcatatgttc atgtatgtat ctgtatttaa 60

165

#### continued

				-contir	Iuea		
aacacttttg	tattatttt	cctcatatat	gtgtataggt	ttatacggat	gatttaatta	120	
ttacttcacc	accctttatt	tcaggctgat	atcttagcct	tgttactaga	ttaatcatgt	180	
aattagttat	gtcacgctta	cattcacgcc	ctccccccac	atccgctcta	accgaaaagg	240	
aaggagttag	acaacctgaa	gtctaggtcc	ctatttattt	ttttatagtt	atgttagtat	300	
taagaacgtt	atttatattt	caaatttttc	tttttttct	gtacagacgc	gtgtacgcat	360	
gtaacattat	actgaaaacc	ttgcttgaga	aggttttggg	acgctcgaag	gctttaattt	420	
gcgggcggcc	gcacctggta	aaacctctag	tggagtagta	gatgtaatca	atgaagcgga	480	
agccaaaaga	ccagagtaga	ggcctataga	agaaactgcg	ataccttttg	tgatggctaa	540	
acaaacagac	atctttttat	atgtttttac	ttctgtatat	cgtgaagtag	taagtgataa	600	
gcgaatttgg	ctaagaacgt	tgtaagtgaa	caagggacct	cttttgcctt	tcaaaaaagg	660	
attaaatgga	gttaatcatt	gagatttagt	tttcgttaga	ttctgtatcc	ctaaataact	720	
cccttacccg	acgggaaggc	acaaaagact	tgaataatag	caaacggcca	gtagccaaga	780	
ccaaataata	ctagagttaa	ctgatggtct	taaacaggca	ttacgtggtg	aactccaaga	840	
ccaatataca	aaatatcgat	aagttattct	tgcccaccaa	tttaaggagc	ctacatcagg	900	
acagtagtac	cattcctcag	agaagaggta	tacataacaa	gaaaatcgcg	tgaacacctt	960	
atataactta	gcccgttatt	gagctaaaaa	accttgcaaa	atttcctatg	aataagaata	1020	
cttcagacgt	gataaaaatt	tactttctaa	ctcttctcac	gctgccccta	tctgttcttc	1080	
cgctctaccg	tgagaaataa	agcatcgagt	acggcagttc	gctgtcactg	aactaaaaca	1140	
ataaggctag	ttcgaatgat	gaacttgctt	gctgtcaaac	ttctgagttg	ccgctgatgt	1200	
gacactgtga	caataaattc	aaaccggtta	tagcggtctc	ctccggtacc	ggttctgcca	1260	
cctccaatag	agctcagtag	gagtcagaac	ctctgcggtg	gctgtcagtg	actcatccgc	1320	
gtttcgtaag	ttgtgcgcgt	gcacatttcg	cccgttcccg	ctcatcttgc	agcaggcgga	1380	
aattttcatc	acgctgtagg	acgcaaaaaa	aaaataatta	atcgtacaag	aatcttggaa	1440	
aaaaattga	aaaattttgt	ataaaaggga	tgacctaact	tgactcaatg	gcttttacac	1500	
ccagtatttt	ccctttcctt	gtttgttaca	attatagaag	caagacaaaa	acatatagac	1560	
aacctattcc	taggagttat	attttttac	cctaccagca	atataagtaa	aaaactgttt	1620	
aaacagtatg	gaagaatgta	agatggctaa	gatttactac	caagaagact	gtaacttgtc	1680	
cttgttggat	ggtaagacta	tcgccgttat	cggttacggt	tctcaaggtc	acgctcatgc	1740	
cctgaatgct	aaggaatccg	gttgtaacgt	tatcattggt	ttatacgaag	gtgcggagga	1800	
gtggaaaaga	gctgaagaac	aaggtttcga	agtctacacc	gctgctgaag	ctgctaagaa	1860	
ggctgacatc	attatgatct	tgatcccaga	tgaaaagcag	gctaccatgt	acaaaaacga	1920	
catcgaacca	aacttggaag	ccggtaacat	gttgatgttc	gctcacggtt	tcaacatcca	1980	
tttcggttgt	attgttccac	caaaggacgt	tgatgtcact	atgatcgctc	caaagggtcc	2040	
aggtcacacc	gttagatccg	aatacgaaga	aggtaaaggt	gtcccatgct	tggttgctgt	2100	
cgaacaagac	gctactggca	aggetttgga	tatggctttg	gcctacgctt	tagccatcgg	2160	
tggtgctaga	gccggtgtct	tggaaactac	cttcagaacc	gaaactgaaa	ccgacttgtt	2220	
cggtgaacaa	gctgttttat	gtggtggtgt	ctgcgctttg	atgcaggccg	gttttgaaac	2280	
cttggttgaa	gccggttacq	acccaagaaa	cgcttacttc	gaatgtatcc	acgaaatgaa	2340	
qttgatcgtt.	qacttgatct	accaatctoo	tttctccaat.	atqcqttact	ctatctccaa	2400	
cactoctora	tacaataact	acattaccor	tocaaagate	attactoaad	ataccaadaa	2460	
Joeguu							

167

#### -continued

ggctatgaag	aagattttgt	ctgacattca	agatggtacc	tttgccaagg	acttcttggt	2520
tgacatgtct	gatgctggtt	cccaggtcca	cttcaaggct	atgagaaagt	tggcctccga	2580
acacccagct	gaagttgtcg	gtgaagaaat	tagatccttg	tactcctggt	ccgacgaaga	2640
caagttgatt	aacaactgag	gccctgcagg	ccagaggaaa	ataatatcaa	gtgctggaaa	2700
ctttttctct	tggaatttt	gcaacatcaa	gtcatagtca	attgaattga	cccaatttca	2760
catttaagat	tttttttt	tcatccgaca	tacatctgta	cactaggaag	ccctgttttt	2820
ctgaagcagc	ttcaaatata	tatattttt	acatatttat	tatgattcaa	tgaacaatct	2880
aattaaatcg	aaaacaagaa	ccgaaacgcg	aataaataat	ttatttagat	ggtgacaagt	2940
gtataagtcc	tcatcgggac	agctacgatt	tctctttcgg	ttttggctga	gctactggtt	3000
gctgtgacgc	agcggcatta	gcgcggcgtt	atgagctacc	ctcgtggcct	gaaagatggc	3060
gggaataaag	cggaactaaa	aattactgac	tgagccatat	tgaggtcaat	ttgtcaactc	3120
gtcaagtcac	gtttggtgga	cggccccttt	ccaacgaatc	gtatatacta	acatgcgcgc	3180
gcttcctata	tacacatata	catatatata	tatatatata	tgtgtgcgtg	tatgtgtaca	3240
cctgtattta	atttccttac	tcgcgggttt	ttctttttc	tcaattcttg	gcttcctctt	3300
tctcgagcgg	accggatcct	cgcgaactcc	aaaatgagct	atcaaaaacg	atagatcgat	3360
taggatgact	ttgaaatgac	tccgcagtgg	actggccgtt	aatttcaagc	gtgagtaaaa	3420
tagtgcatga	caaaagatga	gctaggcttt	tgtaaaaata	tcttacgttg	taaaatttta	3480
gaaatcatta	tttccttcat	atcattttgt	cattgacctt	cagaagaaaa	gagccgacca	3540
ataatataaa	taaataaata	aaaataatat	tccattattt	ctaaacagat	tcaatactca	3600
ttaaaaaact	atatcaatta	atttgaatta	acttaattaa	ttatttttg	ccagtttctt	3660
caggcttcca	aaagtctgtt	acggeteece	tagaagcaga	cgaaacgatg	tgagcatatt	3720
taccaaggat	accgcgtgaa	tagagcggtg	gcaattcaat	ggtctcttga	cgatgtttta	3780
actcttcatc	ggagatatca	aagtgtaatt	ccttagtgtc	ttggtcaata	gtgactatgt	3840
ctcctgtttg	caggtaggcg	attggaccgc	catcttgtgc	ttcaggagcg	atatgaccca	3900
cgacaagacc	ataagtacca	cctgagaagc	ggccatctgt	cagaagggca	actttttcac	3960
cttgcccttt	accaacaatc	attgatgaaa	gggaaagcat	ttcaggcata	ccaggaccgc	4020
cctttggtcc	tacaaaacgt	acgacaacaa	catcaccatc	aacaatatca	tcattcaaga	4080
cagcttcaat	ggcttcttct	tcagaattaa	agaccttagc	aggaccgaca	tgacgacgca	4140
cttttacacc	agaaactttg	gcaacggcac	cgtctggagc	caagttacca	tggagaataa	4200
tgaccggacc	atcttcacgt	ttaggatttt	caagcggcat	aataaccttt	tgaccaggtg	4260
ttaaatcatc	aaaagccttc	aaattttcag	cgactgtttt	gccagtacaa	gtgatacggt	4320
caccatgaag	gaagccattt	ttaaggagat	atttcataac	tgctggtacc	cctccgacct	4380
tgtaaaggtc	ttggaataca	tattgaccag	aaggtttcaa	atcagccaaa	tgaggaactt	4440
tttcttggaa	agtattgaaa	tcatcaagtg	tcaattccac	attagcagca	tgggcaatag	4500
ctaagaggtg	aagggttgag	ttggttgaac	ctcccagagc	catagttaca	gtaatagcat	4560
cttcaaaagc	ttcacgcgtt	aaaatgtcag	aaggttttaa	gcccatttcg	agcattttga	4620
caacagcgcg	accagettet	tcaatatctg	ctttctttc	tgcggattca	gccgggtgag	4680
aagatgaacc	cggaaggcta	agtcccaaaa	cttcaatagc	tgtcgccatt	gtgttagcag	4740
tatacatacc	accgcagcct	ccaggaccgg	gacaagcatt	acattccaaa	gctttaactt	4800

169

#### continued

				-0011011	lued	
cttctttggt	catatcgccg	tggttccaat	ggccgacacc	ttcaaagaca	gagactaaat	4860
cgatatcttt	gccgtctaaa	ttaccaggtg	caattgttcc	gccgtaagca	aaaatggctg	4920
ggatatccat	gttagccata	gcgataacag	aaccgggcat	gtttttatca	caaccgccaa	4980
tggctacaaa	agcatccgca	ttatgacctc	ccatggctgc	ttcaatagaa	tctgcaataa	5040
tatcacgaga	tgtcaaggag	aaacgcattc	cttgggttcc	catggcgatt	ccatcagaaa	5100
ccgtgattgt	tccgaactga	actggccaag	caccagcttc	cttaacaccg	actttggcta	5160
gtttaccaaa	gtcatgtaag	tggatattac	aaggtgtgtt	ttcagcccaa	gttgaaatga	5220
caccgacgat	aggtttttca	aagtcttcat	cttgcatacc	agttgcacgc	aacatagcac	5280
gattaggtga	tttaaccatt	gaatcgtaaa	cagaactacg	atttcttaag	tctttaagag	5340
ttttttgtc	agtcatactc	acgtgaaact	tagattagat	tgctatgctt	tctttccaat	5400
gagcaagaag	taaaaaaagt	tgtaatagaa	caggaaaaat	gaagctgaaa	cttgagaaat	5460
tgaagaccgt	ttgttaactc	aaatatcaat	gggaggtcgt	cgaaagagaa	caaaatcgaa	5520
aaaaaagttt	tcaagagaaa	gaaacgtgat	aaaaattttt	attgccttct	ccgacgaaga	5580
aaaagggacg	aggcggtctc	ttttccttt	tccaaacctt	tagtacgggt	aattaacggc	5640
accctagagg	aaggaggagg	gggaatttag	tatgctgtgc	ttgggtgttt	tgaagtggta	5700
cggcggtgcg	cggagtccga	gaaaatctgg	aagagtaaaa	aaggagtaga	gacattttga	5760
agctatgccg	gcagatctat	ttaaatggcg	cgccgacgtc	aggtggcact	tttcggggaa	5820
atgtgcgcgg	aacccctatt	tgtttattt	tctaaataca	ttcaaatatg	tatccgctca	5880
tgagacaata	accctgataa	atgcttcaat	aatattgaaa	aaggaagagt	atgagtattc	5940
aacatttccg	tgtcgccctt	attccctttt	ttgcggcatt	ttgccttcct	gtttttgctc	6000
acccagaaac	gctggtgaaa	gtaaaagatg	ctgaagatca	gttgggtgca	cgagtgggtt	6060
acatcgaact	ggatctcaac	agcggtaaga	tccttgagag	ttttcgcccc	gaagaacgtt	6120
ttccaatgat	gagcactttt	aaagttctgc	tatgtggcgc	ggtattatcc	cgtattgacg	6180
ccgggcaaga	gcaactcggt	cgccgcatac	actattctca	gaatgacttg	gttgagtact	6240
caccagtcac	agaaaagcat	cttacggatg	gcatgacagt	aagagaatta	tgcagtgctg	6300
ccataaccat	gagtgataac	actgcggcca	acttacttct	gacaacgatc	ggaggaccga	6360
aggagctaac	cgcttttttg	cacaacatgg	gggatcatgt	aactcgcctt	gatcgttggg	6420
aaccggagct	gaatgaagcc	ataccaaacg	acgagcgtga	caccacgatg	cctgtagcaa	6480
tggcaacaac	gttgcgcaaa	ctattaactg	gcgaactact	tactctagct	tcccggcaac	6540
aattaataga	ctggatggag	gcggataaag	ttgcaggacc	acttctgcgc	teggeeette	6600
cggctggctg	gtttattgct	gataaatctg	gagccggtga	gcgtgggtct	cgcggtatca	6660
ttgcagcact	ggggccagat	ggtaagccct	cccgtatcgt	agttatctac	acgacgggga	6720
gtcaggcaac	tatggatgaa	cgaaatagac	agatcgctga	gataggtgcc	tcactgatta	6780
agcattggta	actgtcagac	caagtttact	catatatact	ttagattgat	ttaaaacttc	6840
atttttaatt	taaaaggatc	taggtgaaga	tcctttttga	taatctcatg	accaaaatcc	6900
cttaacgtga	gttttcgttc	cactgagcgt	cagaccccgt	agaaaagatc	aaaggatctt	6960
cttgagatcc	ttttttctg	cgcgtaatct	gctgcttgca	aacaaaaaaa	ccaccgctac	7020
cagcggtggt	ttgtttgccg	gatcaagagc	taccaactct	ttttccgaag	gtaactggct	7080
tcagcagagc	gcagatacca	aatactgttc	ttctagtgta	gccgtagtta	ggccaccact	7140
tcaagaactc	tgtagcaccg	cctacatacc	tcgctctgct	aatcctgtta	ccagtggctg	7200
	-			-		

171

ctgccagtgg	cgataagtcg	tgtcttaccg	ggttggactc	aagacgatag	ttaccggata	7260
aggcgcagcg	gtcgggctga	acgggggggtt	cgtgcacaca	gcccagcttg	gagcgaacga	7320
cctacaccga	actgagatac	ctacagcgtg	agctatgaga	aagcgccacg	cttcccgaag	7380
ggagaaaggc	ggacaggtat	ccggtaagcg	gcagggtcgg	aacaggagag	cgcacgaggg	7440
agcttccagg	gggaaacgcc	tggtatcttt	atagtcctgt	cgggtttcgc	cacctctgac	7500
ttgagcgtcg	atttttgtga	tgctcgtcag	ggggggcggag	cctatggaaa	aacgccagca	7560
acgcggcctt	tttacggttc	ctggcctttt	gctggccttt	tgctcacatg	ttettteetg	7620
cgttatcccc	tgattctgtg	gataaccgta	ttaccgcctt	tgagtgagct	gataccgctc	7680
gccgcagccg	aacgaccgag	cgcagcgagt	cagtgagcga	ggaagcggaa	gagcgcccaa	7740
tacgcaaacc	gcctctcccc	gcgcgttggc	cgattcatta	atgcagctgg	cacgacaggt	7800
ttcccgactg	gaaagcgggc	agtgagcgca	acgcaattaa	tgtgagttag	ctcactcatt	7860
aggcacccca	ggctttacac	tttatgcttc	cggctcgtat	gttgtgtgga	attgtgagcg	7920
gataacaatt	tcacacagga	aacagctatg	accatgatta	cgccaagctt	tttctttcca	7980
atttttttt	tttcgtcatt	ataaaaatca	ttacgaccga	gattcccggg	taataactga	8040
tataattaaa	ttgaagctct	aatttgtgag	tttagtatac	atgcatttac	ttataataca	8100
gttttttagt	tttgctggcc	gcatcttctc	aaatatgctt	cccagcctgc	ttttctgtaa	8160
cgttcaccct	ctaccttagc	atcccttccc	tttgcaaata	gtcctcttcc	aacaataata	8220
atgtcagatc	ctgtagagac	cacatcatcc	acggttctat	actgttgacc	caatgcgtct	8280
cccttgtcat	ctaaacccac	accgggtgtc	ataatcaacc	aatcgtaacc	ttcatctctt	8340
ccacccatgt	ctctttgagc	aataaagccg	ataacaaaat	ctttgtcgct	cttcgcaatg	8400
tcaacagtac	ccttagtata	ttctccagta	gatagggagc	ccttgcatga	caattctgct	8460
aacatcaaaa	ggcctctagg	ttcctttgtt	acttcttctg	ccgcctgctt	caaaccgcta	8520
acaatacctg	ggcccaccac	accgtgtgca	ttcgtaatgt	ctgcccattc	tgctattctg	8580
tatacacccg	cagagtactg	caatttgact	gtattaccaa	tgtcagcaaa	ttttctgtct	8640
tcgaagagta	aaaaattgta	cttggcggat	aatgccttta	gcggcttaac	tgtgccctcc	8700
atggaaaaat	cagtcaagat	atccacatgt	gtttttagta	aacaaatttt	gggacctaat	8760
gcttcaacta	actccagtaa	ttccttggtg	gtacgaacat	ccaatgaagc	acacaagttt	8820
gtttgctttt	cgtgcatgat	attaaatagc	ttggcagcaa	caggactagg	atgagtagca	8880
gcacgttcct	tatatgtagc	tttcgacatg	atttatcttc	gtttcctgca	ggtttttgtt	8940
ctgtgcagtt	gggttaagaa	tactgggcaa	tttcatgttt	cttcaacact	acatatgcgt	9000
atatatacca	atctaagtct	gtgctccttc	cttcgttctt	ccttctgttc	ggagattacc	9060
gaatcaaaaa	aatttcaagg	aaaccgaaat	caaaaaaaag	aataaaaaaa	aaatgatgaa	9120
ttgaaaagct	tgcatgcctg	caggtcgact	ctagtatact	ccgtctactg	tacgatacac	9180
ttccgctcag	gtccttgtcc	tttaacgagg	ccttaccact	cttttgttac	tctattgatc	9240
cageteagea	aaggcagtgt	gatctaagat	tctatcttcg	cgatgtagta	aaactagcta	9300
gaccgagaaa	gagactagaa	atgcaaaagg	cacttctaca	atggctgcca	tcattattat	9360
ccgatgtgac	gctgcatttt	tttttttt	tttttttt	tttttttt	tttttttt	9420
tttttttgt	acaaatatca	taaaaaaaga	gaatctttt	aagcaaqqat	tttcttaact	9480
tcttcqqcqa	cagcatcacc	gacttcqqtq	gtactqttqq	aaccacctaa	atcaccaqtt	9540
					J	
173

### -continued

	9600
aagttcaatg acaatttcaa catcattgca gcagacaaga tagtggcgat agggttgacc	9660
ttattetttg geaaatetgg ageggaacea tggeatggtt egtaeaaaee aaatgeggtg	9720
ttettgtetg geaaagagge caaggaegea gatggeaaca aaceeaagga geetgggata	9780
acggaggett categgagat gatateacea aacatgttge tggtgattat aataeeattt	9840
aggtgggttg ggttcttaac taggatcatg gcggcagaat caatcaattg atgttgaact	9900
ttcaatgtag ggaattcgtt cttgatggtt tcctccacag tttttctcca taatcttgaa	9960
gaggccaaaa cattagcttt atccaaggac caaataggca atggtggctc atgttgtagg 1	0020
gccatgaaag cggccattct tgtgattctt tgcacttctg gaacggtgta ttgttcacta 1	0080
teecaagega caccateace ategtettee tttetettae caaagtaaat aceteecaet 1	0140
aattetetaa caacaacgaa gteagtaeet ttageaaatt gtggettgat tggagataag 1	0200
tctaaaagag agtcggatgc aaagttacat ggtcttaagt tggcgtacaa ttgaagttct 1	0260
ttacggattt ttagtaaacc ttgttcaggt ctaacactac cggtacccca tttaggacca 1	0320
cccacagcac ctaacaaaac ggcatcagcc ttcttggagg cttccagcgc ctcatctgga 1	0380
agtggaacac ctgtagcatc gatagcagca ccaccaatta aatgattttc gaaatcgaac 1	0440
ttgacattgg aacgaacatc agaaatagct ttaagaacct taatggcttc ggctgtgatt 1	0500
tettgaceaa egtggteace tggeaaaaeg aegatettet taggggeaga eattaeaatg 1	0560
gtatateett gaaatatata taaaaaaaaa aaaaaaaaaa	0620
tcaatgatat tcgaatacgc tttgaggaga tacagcctaa tatccgacaa actgttttac 1	0680
agatttacga tcgtacttgt tacccatcat tgaattttga acatccgaac ctgggagttt 1	0740
tccctgaaac agatagtata tttgaacctg tataataata tatagtctag cgctttacgg 1	0800
aagacaatgt atgtatttcg gttcctggag aaactattgc atctattgca taggtaatct 1	0860
tgcacgtcgc atccccggtt cattttctgc gtttccatct tgcacttcaa tagcatatct 1	0920
ttgttaacga agcatctgtg cttcattttg tagaacaaaa atgcaacgcg agagcgctaa 1	0980
tttttcaaac aaagaatctg agctgcattt ttacagaaca gaaatgcaac gcgaaagcgc 1:	1040
tattttacca acgaagaatc tgtgcttcat ttttgtaaaa caaaaatgca acgcgagagc 1	1100
gctaattttt caaacaaaga atctgagctg catttttaca gaacagaaat gcaacgcgag 1	1160
agegetattt taccaacaaa gaatetatae ttetttttg ttetacaaaa atgeateeeg 1	1220
agagegetat ttttetaaca aageatetta gattaetttt ttteteettt gtgegeteta 1	1280
taatgcagtc tcttgataac tttttgcact gtaggtccgt taaggttaga agaaggctac 1	1340
tttggtgtct attttctctt ccataaaaaa agcctgactc cacttcccgc gtttactgat 1	1400
tactagogaa gotgogggtg catttttca agataaaggo atoooogatt atattotata 1	1460
ccgatgtgga ttgcgcatac tttgtgaaca gaaagtgata gcgttgatga ttcttcattg 1	1520
gtcagaaaat tatgaacggt ttcttctatt ttgtctctat atactacgta taggaaatgt 1	1580
ttacattttc gtattgtttt cgattcactc tatgaatagt tcttactaca atttttttgt 1	1640
ctaaagagta atactagaga taaacataaa aaatgtagag gtcgagttta gatgcaagtt $1^\circ$	1700
caaggagcga aaggtggatg ggtaggttat atagggatat agcacagaga tatatagcaa 1	1760
agagatactt ttgagcaatg tttgtggaag cggtattcgc aatattttag tagctcgtta $1^\circ$	1820
cagtccggtg cgtttttggt tttttgaaag tgcgtcttca gagcgctttt ggttttcaaa 1	1880
agegetetga agtteetata etttetagag aataggaaet teggaatagg aaetteaaag 1	1940

175

#### -continued

cgtttccgaa aacgagcgct tccgaaaatg caacgcgagc tgcgcacata cagctcactg 12000 ttcacgtcgc acctatatct gcgtgttgcc tgtatatata tatacatgag aagaacggca 12060 tagtgcgtgt ttatgcttaa atgcgtactt atatgcgtct atttatgtag gatgaaaggt 12120 agtctagtac ctcctgtgat attatcccat tccatgcggg gtatcgtatg cttccttcag 12180 cactaccett tagetgttet atatgetgee acteeteaat tggattagte teateettea 12240 atgetateat tteetttgat attggateat atgeatagta eegagaaaet agaggate 12298 <210> SEQ ID NO 23 <211> LENGTH: 16387 <212> TYPE: DNA <213> ORGANISM: Artificial sequence <220> FEATURE: <223> OTHER INFORMATION: pLH475-JEA1 plasmid <400> SEQUENCE: 23 teccattace gacatttggg egetataegt geatatgtte atgtatgtat etgtatttaa 60 aacacttttg tattattttt cctcatatat gtgtataggt ttatacggat gatttaatta 120 ttacttcacc accctttatt tcaggctgat atcttagcct tgttactagt tagaaaaaga 180 catttttgct gtcagtcact gtcaagagat tcttttgctg gcatttcttc tagaagcaaa 240 aagagggatg cgtcttttcc gctgaaccgt tccagcaaaa aagactacca acgcaatatg 300 gattgtcaga atcatataaa agagaagcaa ataactcctt gtcttgtatc aattgcatta 360 taatatette ttgttagtge aatateatat agaagteate gaaatagata ttaagaaaaa 420 caaactgtac aatcaatcaa tcaatcatcg ctgaggatgt tgacaaaaagc aacaaaagaa 480 caaaaatccc ttgtgaaaaa cagaggggcg gagcttgttg ttgattgctt agtggagcaa 540 ggtgtcacac atgtatttgg cattccaggt gcaaaaattg atgcggtatt tgacgcttta 600 caagataaag gacctgaaat tatcgttgcc cggcacgaac aaaacgcagc attcatggcc 660 caagcagtcg gccgtttaac tggaaaaccg ggagtcgtgt tagtcacatc aggaccgggt 720 gcctctaact tggcaacagg cctgctgaca gcgaacactg aaggagaccc tgtcgttgcg 780 cttgctggaa acgtgatccg tgcagatcgt ttaaaacgga cacatcaatc tttggataat 840 gcggcgctat tccagccgat tacaaaatac agtgtagaag ttcaagatgt aaaaaatata 900 ccggaagctg ttacaaatgc atttaggata gcgtcagcag ggcaggctgg ggccgctttt 960 gtgagettte egeaagatgt tgtgaatgaa gteacaaata egaaaaaegt gegtgetgtt 1020 gcagcgccaa aactcggtcc tgcagcagat gatgcaatca gtgcggccat agcaaaaatc 1080 caaacagcaa aactteetgt egttttggte ggeatgaaag geggaagaee ggaagcaatt 1140 aaagcggttc gcaagctttt gaaaaaggtt cagcttccat ttgttgaaac atatcaagct 1200 gccggtaccc tttctagaga tttagaggat caatattttg gccgtatcgg tttgttccgc 1260 aaccageetg gegatttaet getagageag geagatgttg ttetgaegat eggetatgae 1320 ccgattgaat atgatccgaa attctggaat atcaatggag accggacaat tatccattta 1380 gacgagatta togotgacat tgatcatgot taccagootg atottgaatt gatcggtgac 1440 atteegteea egateaatea tategaacae gatgetgtga aagtggaatt tgeagagegt 1500 gagcagaaaa tcctttctga tttaaaacaa tatatgcatg aaggtgagca ggtgcctgca 1560 gattggaaat cagacagagc gcaccctctt gaaatcgtta aagagttgcg taatgcagtc 1620 gatgatcatg ttacagtaac ttgcgatatc ggttcgcacg ccatttggat gtcacgttat 1680

177

### -continued

ttccgcagct	acgagccgtt	aacattaatg	atcagtaacg	gtatgcaaac	actcggcgtt	1740
gcgcttcctt	gggcaatcgg	cgcttcattg	gtgaaaccgg	gagaaaaagt	ggtttctgtc	1800
tctggtgacg	gcggtttctt	attctcagca	atggaattag	agacagcagt	tcgactaaaa	1860
gcaccaattg	tacacattgt	atggaacgac	agcacatatg	acatggttgc	attccagcaa	1920
ttgaaaaaat	ataaccgtac	atctgcggtc	gatttcggaa	atatcgatat	cgtgaaatat	1980
gcggaaagct	tcggagcaac	tggcttgcgc	gtagaatcac	cagaccagct	ggcagatgtt	2040
ctgcgtcaag	gcatgaacgc	tgaaggtcct	gtcatcatcg	atgtcccggt	tgactacagt	2100
gataacatta	atttagcaag	tgacaagctt	ccgaaagaat	tcggggaact	catgaaaacg	2160
aaagctctct	agttaattaa	tcatgtaatt	agttatgtca	cgcttacatt	cacgccctcc	2220
ccccacatcc	gctctaaccg	aaaaggaagg	agttagacaa	cctgaagtct	aggtccctat	2280
ttatttttt	atagttatgt	tagtattaag	aacgttattt	atatttcaaa	ttttcttt	2340
ttttctgtac	agacgcgtgt	acgcatgtaa	cattatactg	aaaaccttgc	ttgagaaggt	2400
tttgggacgc	tcgaaggctt	taatttgcgg	gcggccgctc	tagaactagt	accacaggtg	2460
ttgtcctctg	aggacataaa	atacacaccg	agattcatca	actcattgct	ggagttagca	2520
tatctacaat	tgggtgaaat	ggggagcgat	ttgcaggcat	ttgctcggca	tgccggtaga	2580
ggtgtggtca	ataagagcga	cctcatgcta	tacctgagaa	agcaacctga	cctacaggaa	2640
agagttactc	aagaataaga	attttcgttt	taaaacctaa	gagtcacttt	aaaatttgta	2700
tacacttatt	ttttttataa	cttatttaat	aataaaaatc	ataaatcata	agaaattcgc	2760
ttactcttaa	ttaatcaagc	atctaaaaca	caaccgttgg	aagcgttgga	aaccaactta	2820
gcatacttgg	atagagtacc	tcttgtgtaa	cgaggtggag	gtgcaaccca	actttgttta	2880
cgttgagcca	tttccttatc	agagactaat	aggtcaatct	tgttattatc	agcatcaatg	2940
ataatctcat	cgccgtctct	gaccaacccg	ataggaccac	cttcagcggc	ttcgggaaca	3000
atgtggccga	ttaagaaccc	gtgagaacca	ccagagaatc	taccatcagt	caacaatgca	3060
acatctttac	ccaaaccgta	acccatcaga	gcagaggaag	gctttagcat	ttcaggcata	3120
cctggtgcac	ctcttggacc	ttcatatctg	ataacaacaa	cggttttttc	accettettg	3180
atttcacctc	tttccaaggc	ttcaataaag	gcaccttcct	cttcgaacac	acgtgctcta	3240
cccttgaagt	aagtaccttc	cttaccggta	attttaccca	cagctccacc	tggtgccaat	3300
gaaccgtaca	gaatttgcaa	gtgaccgttg	gccttgattg	ggtgggagag	tggcttaata	3360
atctcttgtc	cttcaggtag	gcttggtgct	ttctttgcac	gttctgccaa	agtgtcaccg	3420
gtaacagtca	ttgtgttacc	gtgcaacatg	ttgttttcat	atagatactt	aatcacagat	3480
tgggtaccac	caacgttaat	caaatcggcc	atgacgtatt	taccagaagg	tttgaagtca	3540
ccgatcaatg	gtgtagtatc	actgattctt	tggaaatcat	ctggtgacaa	cttgacaccc	3600
gcagagtgag	caacagccac	caaatgcaaa	acagcattag	tggacccacc	ggttgcaacg	3660
acataagtaa	tggcgttttc	aaaagcctct	tttgtgagga	tatcacgagg	taaaataccc	3720
aattccattg	tcttcttgat	gtattcacca	atgttgtcac	actcagctaa	cttctccttg	3780
gaaacggctg	ggaaggaaga	ggagtttgga	atggtcaaac	ctagcacttc	agcggcagaa	3840
gccattgtgt	tggcagtata	cataccacca	caagaaccag	gacctgggca	tgcatgttcc	3900
acaacatctt	ctctttcttc	ttcagtgaat	tgcttggaaa	tatattcacc	gtaggattgg	3960
aacgcagaga	cgatatcgat	gtttttagaq	atcctgttaa	aacctctagt	ggagtagtag	4020
atqtaatcaa	tqaaqcqqaa	qccaaaadac	caqaqtaqaq	gcctatagaa	qaaactqcqa	4080
- )	J	5		J	5	

179

#### -continued

taccttttgt	gatggctaaa	caaacagaca	tctttttata	tgtttttact	tctgtatatc	4140
gtgaagtagt	aagtgataag	cgaatttggc	taagaacgtt	gtaagtgaac	aagggacctc	4200
ttttgccttt	caaaaaagga	ttaaatggag	ttaatcattg	agatttagtt	ttcgttagat	4260
tctgtatccc	taaataactc	ccttacccga	cgggaaggca	caaaagactt	gaataatagc	4320
aaacggccag	tagccaagac	caaataatac	tagagttaac	tgatggtctt	aaacaggcat	4380
tacgtggtga	actccaagac	caatatacaa	aatatcgata	agttattctt	gcccaccaat	4440
ttaaggagcc	tacatcagga	cagtagtacc	attcctcaga	gaagaggtat	acataacaag	4500
aaaatcgcgt	gaacacctta	tataacttag	cccgttattg	agctaaaaaa	ccttgcaaaa	4560
tttcctatga	ataagaatac	ttcagacgtg	ataaaaattt	actttctaac	tcttctcacg	4620
ctgcccctat	ctgttcttcc	gctctaccgt	gagaaataaa	gcatcgagta	cggcagttcg	4680
ctgtcactga	actaaaacaa	taaggctagt	tcgaatgatg	aacttgcttg	ctgtcaaact	4740
tctgagttgc	cgctgatgtg	acactgtgac	aataaattca	aaccggttat	agcggtctcc	4800
tccggtaccg	gttctgccac	ctccaataga	gctcagtagg	agtcagaacc	tctgcggtgg	4860
ctgtcagtga	ctcatccgcg	tttcgtaagt	tgtgcgcgtg	cacatttcgc	ccgttcccgc	4920
tcatcttgca	gcaggcggaa	attttcatca	cgctgtagga	cgcaaaaaaa	aaataattaa	4980
tcgtacaaga	atcttggaaa	aaaaattgaa	aaattttgta	taaaagggat	gacctaactt	5040
gactcaatgg	cttttacacc	cagtattttc	cctttccttg	tttgttacaa	ttatagaagc	5100
aagacaaaaa	catatagaca	acctattcct	aggagttata	ttttttacc	ctaccagcaa	5160
tataagtaaa	aaactagtat	gaaagttttc	tacgataaag	actgcgacct	gtcgatcatc	5220
caaggtaaga	aagttgccat	catcggcttc	ggttcccagg	gccacgctca	agcactcaac	5280
ctgaaggatt	ccggcgtaga	cgtgactgtt	ggcctgccta	aaggctttgc	tgatgtagcc	5340
aaggctgaag	cccacggctt	taaagtgacc	gacgttgctg	cagccgttgc	cggtgccgac	5400
ttggtcatga	tcctgattcc	ggacgagttc	cagtcccagc	tgtacaagaa	cgaaatcgag	5460
ccgaacatca	agaagggcgc	cactctggcc	ttctcccacg	gcttcgcgat	ccactacaac	5520
caggttgtgc	ctcgtgccga	cctcgacgtg	atcatgatcg	cgccgaaggc	tccaggccac	5580
accgtacgtt	ccgagttcgt	caagggcgga	ggtattcctg	acctgatcgc	gatctaccag	5640
gacgtttccg	gcaacgccaa	gaacgtcgcc	ctgtcctacg	ccgcaggcgt	gggcggcggc	5700
cgtaccggca	tcatcgaaac	caccttcaag	gacgagactg	aaaccgacct	gttcggtgag	5760
caggetgtte	tgtgtggcgg	taccgtcgag	ctggtcaaag	ccggtttcga	aaccctggtt	5820
gaagctggct	acgctccaga	aatggcctac	ttcgagtgcc	tgcacgaact	gaagctgatc	5880
gttgacctca	tgtacgaagg	cggtatcgcc	aacatgaact	actcgatctc	caacaacgct	5940
gaatacggcg	agtacgtgac	tggtccagaa	gtcatcaacg	ccgaatcccg	tcaggccatg	6000
cgcaatgctc	tgaagcgcat	ccaggacggc	gaatacgcga	agatgttcat	cagcgaaggc	6060
gctaccggct	acccatcgat	gaccgccaag	cgtcgtaaca	acgctgctca	cggtatcgaa	6120
atcatcggcg	agcaactgcg	ctcgatgatg	ccttggatcg	gtgccaacaa	aatcgtcgac	6180
aaagccaaga	actaaggccc	tgcaggccta	tcaagtgctg	gaaacttttt	ctcttggaat	6240
ttttgcaaca	tcaagtcata	gtcaattgaa	ttgacccaat	ttcacattta	agatttttt	6300
tttttcatcc	gacatacatc	tgtacactag	gaagccctgt	ttttctgaag	cagcttcaaa	6360
tatatatatt	ttttacatat	ttattatgat	tcaatgaaca	atctaattaa	atcgaaaaca	6420

181

### -continued

agaaccgaaa	cgcgaataaa	taatttattt	agatggtgac	aagtgtataa	gtcctcatcg	6480
ggacagctac	gatttctctt	tcggttttgg	ctgagctact	ggttgctgtg	acgcagcggc	6540
attagcgcgg	cgttatgagc	taccctcgtg	gcctgaaaga	tggcgggaat	aaagcggaac	6600
taaaaattac	tgactgagcc	atattgaggt	caatttgtca	actcgtcaag	tcacgtttgg	6660
tggacggccc	ctttccaacg	aatcgtatat	actaacatgc	gcgcgcttcc	tatatacaca	6720
tatacatata	tatatatata	tatatgtgtg	cgtgtatgtg	tacacctgta	tttaatttcc	6780
ttactcgcgg	gtttttcttt	tttctcaatt	cttggcttcc	tctttctcga	gtatataatt	6840
tttcaggtaa	aatttagtac	gatagtaaaa	tacttctcga	actcgtcaca	tatacgtgta	6900
cataatgtct	gaaccagctc	aaaagaaaca	aaaggttgct	aacaactctc	tagagcggcc	6960
gcccgcaaat	taaagccttc	gagcgtccca	aaaccttctc	aagcaaggtt	ttcagtataa	7020
tgttacatgc	gtacacgcgt	ctgtacagaa	aaaaaagaaa	aatttgaaat	ataaataacg	7080
ttcttaatac	taacataact	ataaaaaaat	aaatagggac	ctagacttca	ggttgtctaa	7140
ctccttcctt	ttcggttaga	gcggatgtgg	ggggagggggg	tgaatgtaag	cgtgacataa	7200
ctaattacat	gattaattaa	ttattggttt	tctggtctca	actttctgac	ttccttacca	7260
accttccaga	tttccatgtt	tctgatggtg	tctaattcct	tttctagctt	ttctctgtag	7320
tcaggttgag	agttgaattc	caaagatctc	ttggtttcgg	taccgttctt	ggtagattcg	7380
tacaagtctt	ggaaaacagg	cttcaaagca	ttcttgaaga	ttgggtacca	gtccaaagca	7440
cctcttctgg	cggtggtgga	acaagcatcg	tacatgtaat	ccataccgta	cttaccgatc	7500
aatgggtata	gagattgggt	agcttcttcg	acggtttcgt	tgaaagcttc	agatggggag	7560
tgaccgtttt	ctctcaagac	gtcgtattga	gccaagaaca	taccgtggat	accacccatt	7620
aaacaacctc	tttcaccgta	caagtcagag	ttgacttctc	tttcgaaagt	ggtttggtaa	7680
acgtaaccgg	aaccaatggc	aacggccaaa	gcttgggcct	tttcgtgagc	cttaccggtg	7740
acatcgttcc	agacggcgta	agaagagtta	ataccacgac	cttccttgaa	caaagatctg	7800
acagttctac	cggaaccctt	tggagcaacc	aagataacat	ctaagtcctt	tggtggttca	7860
acgtgagtca	agtccttgaa	gactggggag	aaaccgtggg	agaagtacaa	agtcttaccc	7920
ttggtcaaca	atggcttgat	agcaggccag	gtttctgatt	gagcggcatc	ggacaacaag	7980
ttcataacgt	aactacctct	cttgatagca	tcttcaacag	tgaacaagtt	cttgcctgga	8040
acccaaccgt	cttcgatggc	agccttccaa	gaagcaccat	ctttacggac	accaatgata	8100
acgttcaaac	cgttgtctct	caagttcaaa	ccttgaccgt	aaccttggga	accgtaaccg	8160
atcaaagcaa	aagtgtcgtt	cttgaagtag	tccaacaact	tttctcttgg	ccagtcagct	8220
ctttcgtaga	cggtttcaac	agtaccaccg	aagttgattt	gcttcaacat	cctcagctct	8280
agatttgaat	atgtattact	tggttatggt	tatatatgac	aaaagaaaaa	gaagaacaga	8340
agaataacgc	aaggaagaac	aataactgaa	attgatagag	aagtattatg	tctttgtctt	8400
tttataataa	atcaagtgca	gaaatccgtt	agacaacatg	agggataaaa	tttaacgtgg	8460
gcgaagaaga	aggaaaaaag	tttttgtgag	ggcgtaattg	aagcgatctg	ttgattgtag	8520
attttttt	tttgaggagt	caaagtcaga	agagaacaga	caaatggtat	taaccatcca	8580
atacttttt	ggagcaacgc	taagctcatg	cttttccatt	ggttacgtgc	tcagttgtta	8640
gatatggaaa	gagaggatgc	tcacggcagc	gtgactccaa	ttgagcccga	aagagaggat	8700
gccacgtttt	cccgacggct	gctagaatgq	aaaaaggaaa	aatagaagaa	tcccattcct	8760
atcattattt	acqtaatqac	ccacacattt	ttgagatttt	caactattac	qtattacqat	8820
			5 5			

183

184

aatcctgctg	tcattatcat	tattatctat	atcgacgtat	gcaacgtatg	tgaagccaag	8880
taggcaatta	tttagtactg	tcagtattgt	tattcatttc	agatctatcc	gcggtggagc	8940
tcgaattcac	tggccgtcgt	tttacaacgt	cgtgactggg	aaaaccctgg	cgttacccaa	9000
cttaatcgcc	ttgcagcaca	tcccctttc	gccagctggc	gtaatagcga	agaggcccgc	9060
accgatcgcc	cttcccaaca	gttgcgcagc	ctgaatggcg	aatggcgcct	gatgcggtat	9120
tttctcctta	cgcatctgtg	cggtatttca	caccgcatac	gtcaaagcaa	ccatagtacg	9180
cgccctgtag	cggcgcatta	agcgcggcgg	gtgtggtggt	tacgcgcagc	gtgaccgcta	9240
cacttgccag	cgccttagcg	cccgctcctt	tcgctttctt	cccttccttt	ctcgccacgt	9300
tcgccggctt	tccccgtcaa	gctctaaatc	ggggggttccc	tttagggttc	cgatttagtg	9360
ctttacggca	cctcgacccc	aaaaaacttg	atttgggtga	tggttcacgt	agtgggccat	9420
cgccctgata	gacggttttt	cgccctttga	cgttggagtc	cacgttcttt	aatagtggac	9480
tcttgttcca	aactggaaca	acactcaact	ctatctcggg	ctattcttt	gatttataag	9540
ggattttgcc	gatttcggtc	tattggttaa	aaaatgagct	gatttaacaa	aaatttaacg	9600
cgaattttaa	caaaatatta	acgtttacaa	ttttatggtg	cactctcagt	acaatctgct	9660
ctgatgccgc	atagttaagc	cagccccgac	acccgccaac	acccgctgac	gcgccctgac	9720
gggcttgtct	gctcccggca	tccgcttaca	gacaagctgt	gaccgtctcc	gggagctgca	9780
tgtgtcagag	gttttcaccg	tcatcaccga	aacgcgcgag	acgaaagggc	ctcgtgatac	9840
gcctatttt	ataggttaat	gtcatgataa	taatggtttc	ttagacgtca	ggtggcactt	9900
ttcggggaaa	tgtgcgcgga	acccctattt	gtttatttt	ctaaatacat	tcaaatatgt	9960
atccgctcat	gagacaataa	ccctgataaa	tgcttcaata	atattgaaaa	aggaagagta	10020
tgagtattca	acatttccgt	gtcgccctta	ttcccttttt	tgcggcattt	tgccttcctg	10080
tttttgctca	cccagaaacg	ctggtgaaag	taaaagatgc	tgaagatcag	ttgggtgcac	10140
gagtgggtta	catcgaactg	gatctcaaca	gcggtaagat	ccttgagagt	tttcgccccg	10200
aagaacgttt	tccaatgatg	agcactttta	aagttctgct	atgtggcgcg	gtattatccc	10260
gtattgacgc	cgggcaagag	caactcggtc	gccgcataca	ctattctcag	aatgacttgg	10320
ttgagtactc	accagtcaca	gaaaagcatc	ttacggatgg	catgacagta	agagaattat	10380
gcagtgctgc	cataaccatg	agtgataaca	ctgcggccaa	cttacttctg	acaacgatcg	10440
gaggaccgaa	ggagctaacc	gcttttttgc	acaacatggg	ggatcatgta	actcgccttg	10500
atcgttggga	accggagctg	aatgaagcca	taccaaacga	cgagcgtgac	accacgatgc	10560
ctgtagcaat	ggcaacaacg	ttgcgcaaac	tattaactgg	cgaactactt	actctagctt	10620
cccggcaaca	attaatagac	tggatggagg	cggataaagt	tgcaggacca	cttctgcgct	10680
cggcccttcc	ggctggctgg	tttattgctg	ataaatctgg	agccggtgag	cgtgggtctc	10740
gcggtatcat	tgcagcactg	gggccagatg	gtaagccctc	ccgtatcgta	gttatctaca	10800
cgacggggag	tcaggcaact	atggatgaac	gaaatagaca	gatcgctgag	ataggtgcct	10860
cactgattaa	gcattggtaa	ctgtcagacc	aagtttactc	atatatactt	tagattgatt	10920
taaaacttca	tttttaattt	aaaaggatct	aggtgaagat	cctttttgat	aatctcatga	10980
ccaaaatccc	ttaacgtgag	ttttcgttcc	actgagcgtc	agaccccgta	gaaaagatca	11040
aaggatcttc	ttgagatcct	ttttttctgc	gcgtaatctg	ctgcttgcaa	acaaaaaaac	11100
caccqctacc	aqcqqtqqtt	tqtttqccaa	atcaaqaqct	accaactctt	tttccqaaqq	11160

185

# 186

taactggctt	cagcagagcg	cagataccaa	atactgttct	tctagtgtag	ccgtagttag	11220
gccaccactt	caagaactct	gtagcaccgc	ctacatacct	cgctctgcta	atcctgttac	11280
cagtggctgc	tgccagtggc	gataagtcgt	gtcttaccgg	gttggactca	agacgatagt	11340
taccggataa	ggcgcagcgg	tcgggctgaa	cgggggggttc	gtgcacacag	cccagcttgg	11400
agcgaacgac	ctacaccgaa	ctgagatacc	tacagcgtga	gctatgagaa	agcgccacgc	11460
ttcccgaagg	gagaaaggcg	gacaggtatc	cggtaagcgg	cagggtcgga	acaggagagc	11520
gcacgaggga	gcttccaggg	ggaaacgcct	ggtatcttta	tagtcctgtc	gggtttcgcc	11580
acctctgact	tgagcgtcga	tttttgtgat	gctcgtcagg	ggggcggagc	ctatggaaaa	11640
acgccagcaa	cgcggccttt	ttacggttcc	tggccttttg	ctggcctttt	gctcacatgt	11700
tctttcctgc	gttatcccct	gattctgtgg	ataaccgtat	taccgccttt	gagtgagctg	11760
ataccgctcg	ccgcagccga	acgaccgagc	gcagcgagtc	agtgagcgag	gaagcggaag	11820
agcgcccaat	acgcaaaccg	cctctccccg	cgcgttggcc	gattcattaa	tgcagctggc	11880
acgacaggtt	tcccgactgg	aaagcgggca	gtgagcgcaa	cgcaattaat	gtgagttagc	11940
tcactcatta	ggcaccccag	gctttacact	ttatgcttcc	ggctcgtatg	ttgtgtggaa	12000
ttgtgagcgg	ataacaattt	cacacaggaa	acagctatga	ccatgattac	gccaagcttt	12060
ttctttccaa	tttttttt	ttcgtcatta	taaaaatcat	tacgaccgag	attcccgggt	12120
aataactgat	ataattaaat	tgaagctcta	atttgtgagt	ttagtataca	tgcatttact	12180
tataatacag	tttttagtt	ttgctggccg	catcttctca	aatatgcttc	ccagcctgct	12240
tttctgtaac	gttcaccctc	taccttagca	tcccttccct	ttgcaaatag	tcctcttcca	12300
acaataataa	tgtcagatcc	tgtagagacc	acatcatcca	cggttctata	ctgttgaccc	12360
aatgcgtctc	ccttgtcatc	taaacccaca	ccgggtgtca	taatcaacca	atcgtaacct	12420
tcatctcttc	cacccatgtc	tctttgagca	ataaagccga	taacaaaatc	tttgtcgctc	12480
ttcgcaatgt	caacagtacc	cttagtatat	tctccagtag	atagggagcc	cttgcatgac	12540
aattctgcta	acatcaaaag	gcctctaggt	tcctttgtta	cttcttctgc	cgcctgcttc	12600
aaaccgctaa	caatacctgg	geccaccaca	ccgtgtgcat	tcgtaatgtc	tgcccattct	12660
gctattctgt	atacacccgc	agagtactgc	aatttgactg	tattaccaat	gtcagcaaat	12720
tttctgtctt	cgaagagtaa	aaaattgtac	ttggcggata	atgcctttag	cggcttaact	12780
gtgeceteca	tggaaaaatc	agtcaagata	tccacatgtg	tttttagtaa	acaaattttg	12840
ggacctaatg	cttcaactaa	ctccagtaat	tccttggtgg	tacgaacatc	caatgaagca	12900
cacaagtttg	tttgcttttc	gtgcatgata	ttaaatagct	tggcagcaac	aggactagga	12960
tgagtagcag	cacgttcctt	atatgtagct	ttcgacatga	tttatcttcg	tttcctgcag	13020
gtttttgttc	tgtgcagttg	ggttaagaat	actgggcaat	ttcatgtttc	ttcaacacta	13080
catatgcgta	tatataccaa	tctaagtctg	tgeteettee	ttcgttcttc	cttctgttcg	13140
gagattaccg	aatcaaaaaa	atttcaagga	aaccgaaatc	aaaaaaaga	ataaaaaaaa	13200
aatgatgaat	tgaaaagctt	gcatgcctgc	aggtcgactc	tagtatactc	cgtctactgt	13260
acgatacact	tccgctcagg	tccttgtcct	ttaacgaggc	cttaccactc	ttttgttact	13320
ctattgatcc	agctcagcaa	aggcagtgtg	atctaagatt	ctatcttcgc	gatgtagtaa	13380
aactagctag	accgagaaag	agactagaaa	tgcaaaaggc	acttctacaa	tggctgccat	13440
cattattatc	cgatgtgacg	ctgcatttt	tttttttt	tttttttt	tttttttt	13500
tttttttt	ttttttgta	caaatatcat	aaaaaagag	aatcttttta	agcaaggatt	13560

187

ttcttaactt	cttcggcgac	agcatcaccg	acttcggtgg	tactgttgga	accacctaaa	13620
tcaccagttc	tgatacctgc	atccaaaacc	tttttaactg	catcttcaat	ggctttacct	13680
tcttcaggca	agttcaatga	caatttcaac	atcattgcag	cagacaagat	agtggcgata	13740
gggttgacct	tattctttgg	caaatctgga	gcggaaccat	ggcatggttc	gtacaaacca	13800
aatgcggtgt	tcttgtctgg	caaagaggcc	aaggacgcag	atggcaacaa	acccaaggag	13860
cctgggataa	cggaggcttc	atcggagatg	atatcaccaa	acatgttgct	ggtgattata	13920
ataccattta	ggtgggttgg	gttcttaact	aggatcatgg	cggcagaatc	aatcaattga	13980
tgttgaactt	tcaatgtagg	gaattcgttc	ttgatggttt	cctccacagt	ttttctccat	14040
aatcttgaag	aggccaaaac	attagcttta	tccaaggacc	aaataggcaa	tggtggctca	14100
tgttgtaggg	ccatgaaagc	ggccattctt	gtgattcttt	gcacttctgg	aacggtgtat	14160
tgttcactat	cccaagcgac	accatcacca	tcgtcttcct	ttctcttacc	aaagtaaata	14220
cctcccacta	attctctaac	aacaacgaag	tcagtacctt	tagcaaattg	tggcttgatt	14280
ggagataagt	ctaaaagaga	gtcggatgca	aagttacatg	gtcttaagtt	ggcgtacaat	14340
tgaagttctt	tacggatttt	tagtaaacct	tgttcaggtc	taacactacc	ggtaccccat	14400
ttaggaccac	ccacagcacc	taacaaaacg	gcatcagcct	tcttggaggc	ttccagcgcc	14460
tcatctggaa	gtggaacacc	tgtagcatcg	atagcagcac	caccaattaa	atgattttcg	14520
aaatcgaact	tgacattgga	acgaacatca	gaaatagctt	taagaacctt	aatggcttcg	14580
gctgtgattt	cttgaccaac	gtggtcacct	ggcaaaacga	cgatcttctt	aggggcagac	14640
attacaatgg	tatatccttg	aaatatatat	aaaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	14700
tgcagcttct	caatgatatt	cgaatacgct	ttgaggagat	acagcctaat	atccgacaaa	14760
ctgttttaca	gatttacgat	cgtacttgtt	acccatcatt	gaattttgaa	catccgaacc	14820
tgggagtttt	ccctgaaaca	gatagtatat	ttgaacctgt	ataataatat	atagtctagc	14880
gctttacgga	agacaatgta	tgtatttcgg	ttcctggaga	aactattgca	tctattgcat	14940
aggtaatctt	gcacgtcgca	tccccggttc	attttctgcg	tttccatctt	gcacttcaat	15000
agcatatctt	tgttaacgaa	gcatctgtgc	ttcattttgt	agaacaaaaa	tgcaacgcga	15060
gagcgctaat	ttttcaaaca	aagaatctga	gctgcatttt	tacagaacag	aaatgcaacg	15120
cgaaagcgct	attttaccaa	cgaagaatct	gtgcttcatt	tttgtaaaac	aaaaatgcaa	15180
cgcgagagcg	ctaatttttc	aaacaaagaa	tctgagctgc	atttttacag	aacagaaatg	15240
caacgcgaga	gcgctatttt	accaacaaag	aatctatact	tctttttgt	tctacaaaaa	15300
tgcatcccga	gagcgctatt	tttctaacaa	agcatcttag	attactttt	ttctcctttg	15360
tgcgctctat	aatgcagtct	cttgataact	ttttgcactg	taggtccgtt	aaggttagaa	15420
gaaggctact	ttggtgtcta	ttttctcttc	cataaaaaaa	gcctgactcc	acttcccgcg	15480
tttactgatt	actagcgaag	ctgcgggtgc	atttttcaa	gataaaggca	tccccgatta	15540
tattctatac	cgatgtggat	tgcgcatact	ttgtgaacag	aaagtgatag	cgttgatgat	15600
tcttcattgg	tcagaaaatt	atgaacggtt	tcttctattt	tgtctctata	tactacgtat	15660
aggaaatgtt	tacattttcg	tattgttttc	gattcactct	atgaatagtt	cttactacaa	15720
ttttttgtc	taaagagtaa	tactagagat	aaacataaaa	aatgtagagg	tcgagtttag	15780
atgcaagttc	aaggaqcqaa	aggtggatgg	gtaggttata	tagggatata	gcacagagat	15840
atatagcaaa	gagatacttt	tgagcaatgt	ttgtggaage	ggtattcgca	atattttagt	15900
-		-		-	-	

189

# 190

agctcgttac agtccggtgc gtttttggtt ttttgaaagt gcgtcttcag agcgcttttg	15960
gttttcaaaa gcgctctgaa gttcctatac tttctagaga ataggaactt cggaatagga	16020
acttcaaagc gtttccgaaa acgagcgctt ccgaaaatgc aacgcgagct gcgcacatac	16080
ageteactgt teacgtegea ectatatetg egtgttgeet gtatatatat atacatgaga	16140
agaacggcat agtgcgtgtt tatgcttaaa tgcgtactta tatgcgtcta tttatgtagg	16200
atgaaaggta gtctagtacc tcctgtgata ttatcccatt ccatgcgggg tatcgtatgc	16260
ttccttcagc actacccttt agctgttcta tatgctgcca ctcctcaatt ggattagtct	16320
catcottcaa tgotatcatt tootttgata ttggatcata tgoatagtac oggaaacta	16380
gaggate	16387
<210> SEQ ID NO 24 <211> LENGTH: 448 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Saccharomyces cerevisiae CUP1 promoter	
<400> SEQUENCE: 24	
cccattaccg acatttgggc gctatacgtg catatgttca tgtatgtatc tgtatttaaa	60
acacttttgt attatttttc ctcatatatg tgtataggtt tatacggatg atttaattat	120
tacttcacca ccctttattt caggetgata tettageett gttactagtt agaaaaagae	180
atttttgctg tcagtcactg tcaagagatt cttttgctgg catttcttct agaagcaaaa	240
agagegatge gtetttteeg etgaacegtt eeageaaaaa agaetaeeaa egeaatatgg	300
attgtcagaa tcatataaaa gagaagcaaa taactccttg tcttgtatca attgcattat	360
aatatettet tgttagtgca atateatata gaagteateg aaatagatat taagaaaaae	420
aaactgtaca atcaatcaat caatcatc	448
<210> SEQ ID NO 25 <211> LENGTH: 1716 <212> TYPE: DNA <213> ORGANISM: Bacillus subtilis	
<400> SEQUENCE: 25	
atgttgacaa aagcaacaaa agaacaaaaa tcccttgtga aaaacagagg ggcggagctt	60
gttgttgatt gcttagtgga gcaaggtgtc acacatgtat ttggcattcc aggtgcaaaa	120
attgatgcgg tatttgacgc tttacaagat aaaggacctg aaattatcgt tgcccggcac	180
gaacaaaacg cagcattcat ggcccaagca gtcggccgtt taactggaaa accgggagtc	240
gtgttagtca catcaggacc gggtgcctct aacttggcaa caggcctgct gacagcgaac	300
actgaaggag accetgtegt tgegettget ggaaaegtga teegtgeaga tegtttaaaa	360
cggacacatc aatctttgga taatgcggcg ctattccagc cgattacaaa atacagtgta	420
gaagttcaag atgtaaaaaa tataccggaa gctgttacaa atgcatttag gatagcgtca	480
gcagggcagg ctggggccgc ttttgtgagc tttccgcaag atgttgtgaa tgaagtcaca	540
aatacgaaaa acgtgcgtgc tgttgcagcg ccaaaactcg gtcctgcagc agatgatgca	600
atcagtgcgg ccatagcaaa aatccaaaca gcaaaacttc ctgtcgtttt ggtcggcatg	660
aaaggcggaa gaccggaagc aattaaagcg gttcgcaagc ttttgaaaaa ggttcagctt	720
ccatttgttg aaacatatca agctgccggt accctttcta qaqatttaqa qqatcaatat	780
tttggccqta tcggtttgtt ccgcaaccag cctggcgatt tactgctaga gcaggaga	840
3. 3	

#### -continued

gttgttctga cgatcggcta tgacccgatt gaatatgatc cgaaattctg gaatatcaat ggagaccgga caattatcca tttagacgag attatcgctg acattgatca tgcttaccag cctgatcttg aattgatcgg tgacattccg tccacgatca atcatatcga acacgatgct gtgaaagtgg aatttgcaga gcgtgagcag aaaatccttt ctgatttaaa acaatatatg catgaaggtg agcaggtgcc tgcagattgg aaatcagaca gagcgcaccc tcttgaaatc gttaaagagt tgegtaatge agtegatgat catgttaeag taaettgega tateggtteg cacgccattt ggatgtcacg ttatttccgc agctacgagc cgttaacatt aatgatcagt aacggtatge aaacaetegg egttgegett eettggggeaa teggegette attggtgaaa ccgggagaaa aagtggtttc tgtctctggt gacggcggtt tcttattctc agcaatggaa ttaqaqacaq caqttcqact aaaaqcacca attqtacaca ttqtatqqaa cqacaqcaca tatgacatgg ttgcattcca gcaattgaaa aaatataacc gtacatctgc ggtcgatttc ggaaatatcg atatcgtgaa atatgcggaa agcttcggag caactggctt gcgcgtagaa tcaccagacc agctggcaga tgttctgcgt caaggcatga acgctgaagg tcctgtcatc atcgatgtcc cggttgacta cagtgataac attaatttag caagtgacaa gcttccgaaa gaattcqqqq aactcatqaa aacqaaaqct ctctaq <210> SEO ID NO 26 <211> LENGTH: 571 <212> TYPE: PRT <213> ORGANISM: Bacillus subtilis <400> SEQUENCE: 26 Met Leu Thr Lys Ala Thr Lys Glu Gln Lys Ser Leu Val Lys Asn Arg Gly Ala Glu Leu Val Val Asp Cys Leu Val Glu Gln Gly Val Thr His Val Phe Gly Ile Pro Gly Ala Lys Ile Asp Ala Val Phe Asp Ala Leu Gln Asp Lys Gly Pro Glu Ile Ile Val Ala Arg His Glu Gln Asn Ala Ala Phe Met Ala Gln Ala Val Gly Arg Leu Thr Gly Lys Pro Gly Val Val Leu Val Thr Ser Gly Pro Gly Ala Ser Asn Leu Ala Thr Gly Leu Leu Thr Ala Asn Thr Glu Gly Asp Pro Val Val Ala Leu Ala Gly Asn Val Ile Arg Ala Asp Arg Leu Lys Arg Thr His Gln Ser Leu Asp Asn Ala Ala Leu Phe Gln Pro Ile Thr Lys Tyr Ser Val Glu Val Gln Asp Val Lys Asn Ile Pro Glu Ala Val Thr Asn Ala Phe Arg Ile Ala Ser Ala Gly Gln Ala Gly Ala Ala Phe Val Ser Phe Pro Gln Asp Val Val 

Gln Thr Ala Lys Leu Pro Val Val Leu Val Gly Met Lys Gly Gly Arg

Asn Glu Val Thr Asn Thr Lys Asn Val Arg Ala Val Ala Ala Pro Lys

Leu Gly Pro Ala Ala Asp Asp Ala Ile Ser Ala Ala Ile Ala Lys Ile

193

-cont	inued
CONC.	LIIUCU

	210					215					220				
Pro 225	Glu	Ala	Ile	Lys	Ala 230	Val	Arg	Lys	Leu	Leu 235	Lys	Lys	Val	Gln	Leu 240
Pro	Phe	Val	Glu	Thr 245	Tyr	Gln	Ala	Ala	Gly 250	Thr	Leu	Ser	Arg	Asp 255	Leu
Glu	Asp	Gln	Tyr 260	Phe	Gly	Arg	Ile	Gly 265	Leu	Phe	Arg	Asn	Gln 270	Pro	Gly
Asp	Leu	Leu 275	Leu	Glu	Gln	Ala	Asp 280	Val	Val	Leu	Thr	Ile 285	Gly	Tyr	Asp
Pro	Ile 290	Glu	Tyr	Asp	Pro	Lys 295	Phe	Trp	Asn	Ile	Asn 300	Gly	Asp	Arg	Thr
Ile 305	Ile	His	Leu	Aab	Glu 310	Ile	Ile	Ala	Asp	Ile 315	Asb	His	Ala	Tyr	Gln 320
Pro	Asp	Leu	Glu	Leu 325	Ile	Gly	Asp	Ile	Pro 330	Ser	Thr	Ile	Asn	His 335	Ile
Glu	His	Aab	Ala 340	Val	Lys	Val	Glu	Phe 345	Ala	Glu	Arg	Glu	Gln 350	Lys	Ile
Leu	Ser	Asp 355	Leu	Lys	Gln	Tyr	Met 360	His	Glu	Gly	Glu	Gln 365	Val	Pro	Ala
Asp	Trp 370	Lys	Ser	Asb	Arg	Ala 375	His	Pro	Leu	Glu	Ile 380	Val	Lys	Glu	Leu
Arg 385	Asn	Ala	Val	Asp	Aap 390	His	Val	Thr	Val	Thr 395	Сүз	Asp	Ile	Gly	Ser 400
His	Ala	Ile	Trp	Met 405	Ser	Arg	Tyr	Phe	Arg 410	Ser	Tyr	Glu	Pro	Leu 415	Thr
Leu	Met	Ile	Ser 420	Asn	Gly	Met	Gln	Thr 425	Leu	Gly	Val	Ala	Leu 430	Pro	Trp
Ala	Ile	Gly 435	Ala	Ser	Leu	Val	Lys 440	Pro	Gly	Glu	Lys	Val 445	Val	Ser	Val
Ser	Gly 450	Asp	Gly	Gly	Phe	Leu 455	Phe	Ser	Ala	Met	Glu 460	Leu	Glu	Thr	Ala
Val 465	Arg	Leu	Lys	Ala	Pro 470	Ile	Val	His	Ile	Val 475	Trp	Asn	Asp	Ser	Thr 480
Tyr	Asp	Met	Val	Ala 485	Phe	Gln	Gln	Leu	Lys 490	Lys	Tyr	Asn	Arg	Thr 495	Ser
Ala	Val	Asp	Phe 500	Gly	Asn	Ile	Asp	Ile 505	Val	Lys	Tyr	Ala	Glu 510	Ser	Phe
Gly	Ala	Thr 515	Gly	Leu	Arg	Val	Glu 520	Ser	Pro	Asp	Gln	Leu 525	Ala	Asp	Val
Leu	Arg 530	Gln	Gly	Met	Asn	Ala 535	Glu	Gly	Pro	Val	Ile 540	Ile	Asp	Val	Pro
Val 545	Asp	Tyr	Ser	Asp	Asn 550	Ile	Asn	Leu	Ala	Ser 555	Asp	ГЛа	Leu	Pro	Lys 560
Glu	Phe	Gly	Glu	Leu 565	Met	Lys	Thr	Lys	Ala 570	Leu					
<210 <211 <212 <212 <212	0> SI L> LI 2> T 3> OI 0> FI	EQ II ENGTH YPE : RGANI EATUH	D NO H: 2! DNA ISM: RE:	27 50 Art:	ific:	ial S	Seque	ence							
<223	3> 0	THER	INF	ORMA'	FION	: Sad	ccha	romy	ces (	cere	visia	ae C'	YC1 t	erm:	inator 2
<400	i> SI	ZÕAFI	NCE:	27											

ccgcaaatta aagcettega gegteecaaa aeetteteaa geaaggtttt eagtataatg 60

1	0	4
	ノフ	J

ttacatqcqt acacqcqtct qtacaqaaaa aaaaqaaaaa tttqaaatat aaataacqtt	120
cttaatacta acataactat aaaaaaataa atagggacct agacttcagg thetetaact	180
cetteetttt eggttagage ggatatagag ggatatagage taasataast	240
aattacatga	250
<210> SEQ ID NO 28 <211> LENGTH: 1181 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Saccharomyces cerevisiae ILV5 promoter	
<400> SEQUENCE: 28	
taaaacctct agtggagtag tagatgtaat caatgaagcg gaagccaaaa gaccagagta	60
gaggcctata gaagaaactg cgataccttt tgtgatggct aaacaaacag acatcttttt	120
atatgttttt acttctgtat atcgtgaagt agtaagtgat aagcgaattt ggctaagaac	180
gttgtaagtg aacaagggac ctcttttgcc tttcaaaaaa ggattaaatg gagttaatca	240
ttgagattta gttttcgtta gattctgtat ccctaaataa ctcccttacc cgacgggaag	300
gcacaaaaga cttgaataat agcaaacggc cagtagccaa gaccaaataa tactagagtt	360
aactgatggt cttaaacagg cattacgtgg tgaactccaa gaccaatata caaaatatcg	420
ataagttatt cttgcccacc aatttaagga gcctacatca ggacagtagt accattcctc	480
agagaagagg tatacataac aagaaaatcg cgtgaacacc ttatataact tagcccgtta	540
ttgagctaaa aaaccttgca aaatttccta tgaataagaa tacttcagac gtgataaaaa	600
tttactttct aactcttctc acgctgcccc tatctgttct tccgctctac cgtgagaaat	660
aaagcatcga gtacggcagt tcgctgtcac tgaactaaaa caataaggct agttcgaatg	720
atgaacttgc ttgctgtcaa acttctgagt tgccgctgat gtgacactgt gacaataaat	780
tcaaaccggt tatagcggtc teeteeggta eeggttetge caeeteeaat agageteagt	840
aggagtcaga acctctgcgg tggctgtcag tgactcatcc gcgtttcgta agttgtgcgc	900
gtgcacattt cgcccgttcc cgctcatctt gcagcaggcg gaaattttca tcacgctgta	960
ggacgcaaaa aaaaaataat taatcgtaca agaatcttgg aaaaaaaatt gaaaaatttt	1020
gtataaaagg gatgacctaa cttgactcaa tggcttttac acccagtatt ttccctttcc	1080
ttgtttgtta caattataga agcaagacaa aaacatatag acaacctatt cctaggagtt	1140
atatttttt accctaccag caatataagt aaaaaactag t	1181
<210> SEQ ID NO 29 <211> LENGTH: 1017 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Pf5.IlvC-JEA1 coding region	
<400> SEQUENCE: 29	
atgaaagttt tctacgataa agactgcgac ctgtcgatca tccaaggtaa gaaagttgcc	60
atcatcggct tcggttccca gggccacgct caagcactca acctgaagga ttccggcgta	120
gacgtgactg ttggcctgcc taaaggcttt gctgatgtag ccaaggctga agcccacggc	180
tttaaagtga ccgacgttgc tgcagccgtt gccggtgccg acttggtcat gatcctgatt	240
ccggacgagt tccagtccca gctgtacaag aacgaaatcg agccgaacat caagaagggc	300

197

gccactctgg ccttctccca cggcttcgcg atccactaca accaggttgt gcctcgtgcc	360
gacetegaeg tgateatgat egegeegaag geteeaggee acaeegtaeg tteegagtte	420
gtcaagggcg gaggtattcc tgacctgatc gcgatctacc aggacgtttc cggcaacgcc	480
aagaacgteg eeetgteeta egeegeagge gtgggeggeg geegtaeegg eateategaa	540
accaccttca aggacgagac tgaaaccgac ctgttcggtg agcaggctgt tctgtgtggc	600
ggtaccgtcg agctggtcaa agccggtttc gaaaccctgg ttgaagctgg ctacgctcca	660
gaaatggeet aettegagtg eetgeaegaa etgaagetga tegttgaeet eatgtaegaa	720
ggcggtatcg ccaacatgaa ctactcgatc tccaacaacg ctgaatacgg cgagtacgtg	780
actggtccag aagtcatcaa cgccgaatcc cgtcaggcca tgcgcaatgc tctgaagcgc	840
atccaggacg gcgaatacgc gaagatgttc atcagcgaag gcgctaccgg ctacccatcg	900
atgaccgcca agcgtcgtaa caacgctgct cacggtatcg aaatcatcgg cgagcaactg	960
cgctcgatga tgccttggat cggtgccaac aaaatcgtcg acaaagccaa gaactaa	1017
<210> SEQ ID NO 30 <211> LENGTH: 338 <212> TYPE: PRT <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Pf5.IlvC-JEA1 coding region	
<400> SEQUENCE: 30	
Met Lys Val Phe Tyr Asp Lys Asp Cys Asp Leu Ser Ile Ile Gln Gly 1 5 10 15	
Lys Lys Val Ala Ile Ile Gly Phe Gly Ser Gln Gly His Ala Gln Ala 20 25 30	
Leu Asn Leu Lys Asp Ser Gly Val Asp Val Thr Val Gly Leu Pro Lys 35 40 45	
Gly Phe Ala Asp Val Ala Lys Ala Glu Ala His Gly Phe Lys Val Thr 50 55 60	
Asp Val Ala Ala Ala Val Ala Gly Ala Asp Leu Val Met Ile Leu Ile 65 70 75 80	
Pro Asp Glu Phe Gln Ser Gln Leu Tyr Lys Asn Glu Ile Glu Pro Asn 85 90 95	
Ile Lys Lys Gly Ala Thr Leu Ala Phe Ser His Gly Phe Ala Ile His 100 105 110	
Tyr Asn Gln Val Val Pro Arg Ala Asp Leu Asp Val Ile Met Ile Ala 115 120 125	
Pro Lys Ala Pro Gly His Thr Val Arg Ser Glu Phe Val Lys Gly Gly 130 135 140	
Gly Ile Pro Asp Leu Ile Ala Ile Tyr Gln Asp Val Ser Gly Asn Ala 145 150 155 160	
Lys Asn Val Ala Leu Ser Tyr Ala Ala Gly Val Gly Gly Gly Arg Thr 165 170 175	
Gly Ile Ile Glu Thr Thr Phe Lys Asp Glu Thr Glu Thr Asp Leu Phe 180 185 190	
Gly Glu Gln Ala Val Leu Cys Gly Gly Thr Val Glu Leu Val Lys Ala 195 200 205	
Gly Phe Glu Thr Leu Val Glu Ala Gly Tyr Ala Pro Glu Met Ala Tyr 210 215 220	
Phe Glu Cys Leu His Glu Leu Lys Leu Ile Val Asp Leu Met Tyr Glu 225 230 235 240	

# 

-continued
------------

Gly Gly Ile Ala Asn Met Asn Tyr Ser Ile Ser Asn Asn Ala Glu Tyr 245 250 255	
Gly Glu Tyr Val Thr Gly Pro Glu Val Ile Asn Ala Glu Ser Arg Gln	
260 265 270	
275 280 285	
Met Phe Ile Ser Glu Gly Ala Thr Gly Tyr Pro Ser Met Thr Ala Lys 290 295 300	
Arg Arg Asn Asn Ala Ala His Gly Ile Glu Ile Ile Gly Glu Gln Leu305310315320	
Arg Ser Met Met Pro Trp Ile Gly Ala Asn Lys Ile Val Asp Lys Ala 325 330 335	
Lys Asn	
<210> SEQ ID NO 31 <211> LENGTH: 759 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Saccharomyces cerevisiae ILV5 terminator	
<400> SEQUENCE: 31	
ggccctgcag gcctatcaag tgctggaaac tttttctctt ggaatttttg caacatcaag	60
tcatagtcaa ttgaattgac ccaatttcac atttaagatt ttttttttt catccgacat	120
acatetgtae actaggaage eetgtttte tgaageaget teaaatatat atattttta	180
catatttatt atgattcaat gaacaatcta attaaatcga aaacaagaac cgaaacgcga	240
ataaataatt tatttagatg gtgacaagtg tataagtoot catogggaca gotaogattt	300
ctctttcggt tttggctgag ctactggttg ctgtgacgca gcggcattag cgcggcgtta	360
tgagctaccc tcgtggcctg aaagatggcg ggaataaagc ggaactaaaa attactgact	420
gagccatatt gaggtcaatt tgtcaactcg tcaagtcacg tttggtggac ggcccctttc	480
caacgaatcg tatatactaa catgcgcgcg cttcctatat acacatatac atatatat	540
atatatatat gtgtgcgtgt atgtgtacac ctgtatttaa tttccttact cgcgggtttt	600
tetttttet caattettgg etteetett etegagtata taatttttea ggtaaaattt	660
agtacgatag taaaatactt ctcgaactcg tcacatatac gtgtacataa tgtctgaacc	720
ageteaaaag aaacaaaagg ttgetaacaa etetetaga	759
<210> SEQ ID NO 32 <211> LENGTH: 643 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Saccharomyces cerevisiae FBA1 promoter	
<400> SEQUENCE: 32	
gaaatgaata acaatactga cagtactaaa taattgccta cttggcttca catacgttgc	60
atacgtcgat atagataata atgataatga cagcaggatt atcgtaatac gtaatagttg	120
aaaatctcaa aaatgtgtgg gtcattacgt aaataatgat aggaatggga ttcttctatt	180
ttteettttt eeattetage ageegteggg aaaaegtgge ateetetett tegggeteaa	240
ttggagtcac gctgccgtga gcatcctctc tttccatatc taacaactga gcacgtaacc	300
aatggaaaag catgagctta gcgttgctcc aaaaaagtat tggatggtta ataccatttg	360
tetqttetet tetqaetttq acteetcaaa aaaaaaaaat etacaatcaa caqategett	420

#### -continued

caattacgcc ctcacaaaaa cttttttcct tctt	cttogo coacgttaaa ttttato	cct 480
catgttgtct aacggatttc tgcacttgat ttat	tataaa aagacaaaga cataata	ctt 540
ctctatcaat ttcagttatt gttcttcctt gcgt	tattet tetgttette ttttet	ttt 600
gtcatatata accataacca agtaatacat attc	aaatct aga	643
<210> SEQ ID NO 33 <211> LENGTH: 1188 <212> TYPE: DNA <213> ORGANISM: Saccharomyces cerevis	iae	
<400> SEQUENCE: 33		
atgttgagaa ctcaagccgc cagattgatc tgca	actece gtgteateae tgetaag	aga 60
acctttgett tggecacceg tgetgetget taca	gcagac cagctgcccg tttcgtt	aag 120
ccaatgatca ctacccgtgg tttgaagcaa atca	acttcg gtggtactgt tgaaacc	gtc 180
tacgaaagag ctgactggcc aagagaaaag ttgt	tggact acttcaagaa cgacact	ttt 240
getttgateg gttaeggtte ceaaggttae ggte	aaggtt tgaacttgag agacaac	ggt 300
ttgaacgtta tcattggtgt ccgtaaagat ggtg	cttctt ggaaggctgc catcgaa	gac 360
ggttgggttc caggcaagaa cttgttcact gttg	aagatg ctatcaagag aggtagt	tac 420
gttatgaact tgttgtccga tgccgctcaa tcag	aaacct ggcctgctat caagcca	ttg 480
ttgaccaagg gtaagacttt gtacttctcc cacg	gtttet eeccagtett caaggae	ttg 540
actcacgttg aaccaccaaa ggacttagat gtta	tcttgg ttgctccaaa gggttcc	ggt 600
agaactgtca gatctttgtt caaggaaggt cgtg	gtatta actcttctta cgccgtc	tgg 660
aacgatgtca coggtaaggo toacgaaaag good	aagctt tggccgttgc cattggt	tcc 720
ggttacgttt accaaaccac tttcgaaaga gaag	tcaact ctgacttgta cggtgaa	aga 780
ggttgtttaa tgggtggtat ccacggtatg ttct	tggctc aatacgacgt cttgaga	gaa 840
aacggtcact ccccatctga agctttcaac gaaa	ccgtcg aagaagctac ccaatct	cta 900
tacccattga toggtaagta oggtatggat taca	tgtacg atgcttgttc caccacc	gcc 960
agaagaggtg ctttggactg gtacccaatc ttca	agaatg ctttgaagcc tgttttc	caa 1020
gacttgtacg aatctaccaa gaacggtacc gaaa	ccaaga gatctttgga attcaac	tct 1080
caacctgact acagagaaaa gctagaaaag gaat	tagaca ccatcagaaa catggaa	atc 1140
tggaaggttg gtaaggaagt cagaaagttg agac	cagaaa accaataa	1188
<210> SEQ ID NO 34 <211> LENGTH: 395 <212> TYPE: PRT <213> ORGANISM: Saccharomyces cerevis	iae	
<400> SEQUENCE: 34		
Met Leu Arg Thr Gln Ala Ala Arg Leu I 1 5 1	le Cys Asn Ser Arg Val Il 0 15	e
Thr Ala Lys Arg Thr Phe Ala Leu Ala Th 20 25	hr Arg Ala Ala Ala Tyr Se 30	r
Arg Pro Ala Ala Arg Phe Val Lys Pro M 35 40	et Ile Thr Thr Arg Gly Le 45	u
Lys Gln Ile Asn Phe Gly Gly Thr Val G 50 55	lu Thr Val Tyr Glu Arg Al 60	a
Asp Trp Pro Arg Glu Lys Leu Leu Asp T 65 70	yr Phe Lys Asn Asp Thr Ph 75 80	e

### -continued

Ala Leu Ile Gly Tyr Gly Ser Gln Gly Tyr Gly Gln Gly Leu Asn Leu 85 90 95 Arg Asp Asn Gly Leu Asn Val Ile Ile Gly Val Arg Lys Asp Gly Ala Ser Trp Lys Ala Ala Ile Glu Asp Gly Trp Val Pro Gly Lys Asn Leu Phe Thr Val Glu Asp Ala Ile Lys Arg Gly Ser Tyr Val Met Asn Leu Leu Ser Asp Ala Ala Gln Ser Glu Thr Trp Pro Ala Ile Lys Pro Leu Leu Thr Lys Gly Lys Thr Leu Tyr Phe Ser His Gly Phe Ser Pro Val Phe Lys Asp Leu Thr His Val Glu Pro Pro Lys Asp Leu Asp Val Ile Leu Val Ala Pro Lys Gly Ser Gly Arg Thr Val Arg Ser Leu Phe Lys Glu Gly Arg Gly Ile Asn Ser Ser Tyr Ala Val Trp Asn Asp Val Thr Gly Lys Ala His Glu Lys Ala Gln Ala Leu Ala Val Ala Ile Gly Ser Gly Tyr Val Tyr Gln Thr Thr Phe Glu Arg Glu Val Asn Ser Asp Leu Tyr Gly Glu Arg Gly Cys Leu Met Gly Gly Ile His Gly Met Phe Leu Ala Gln Tyr Asp Val Leu Arg Glu Asn Gly His Ser Pro Ser Glu Ala Phe Asn Glu Thr Val Glu Glu Ala Thr Gln Ser Leu Tyr Pro Leu Ile Gly Lys Tyr Gly Met Asp Tyr Met Tyr Asp Ala Cys Ser Thr Thr Ala Arg Arg Gly Ala Leu Asp Trp Tyr Pro Ile Phe Lys Asn Ala Leu Lys Pro Val Phe Gln Asp Leu Tyr Glu Ser Thr Lys Asn Gly Thr Glu Thr Lys Arg Ser Leu Glu Phe Asn Ser Gln Pro Asp Tyr Arg Glu Lys Leu Glu Lys Glu Leu Asp Thr Ile Arg Asn Met Glu Ile Trp Lys Val Gly Lys Glu Val Arg Lys Leu Arg Pro Glu Asn Gln <210> SEQ ID NO 35 <211> LENGTH: 244 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Saccharomyces cerevisiae CYC1 terminator <400> SEQUENCE: 35 attaaagcct tcgagcgtcc caaaaccttc tcaagcaagg ttttcagtat aatgttacat gcgtacacgc gtctgtacag aaaaaaaaga aaaatttgaa atataaataa cgttcttaat actaacataa ctataaaaaa ataaataggg acctagactt caggttgtct aactccttcc ttttcggtta gagcggatgt gggggggggg cgtgaatgta agcgtgacat aactaattac

-continued

atga

<210> SEQ ID NO 36 <211> LENGTH: 570 <212> TYPE: PRT <213> ORGANISM: Bacillus subtilis <400> SEQUENCE: 36 Met Thr Lys Ala Thr Lys Glu Gln Lys Ser Leu Val Lys Asn Arg Gly Ala Glu Leu Val Val Asp Cys Leu Val Glu Gln Gly Val Thr His Val Phe Gly Ile Pro Gly Ala Lys Ile Asp Ala Val Phe Asp Ala Leu Gln 35 40 45 Asp Lys Gly Pro Glu Ile Ile Val Ala Arg His Glu Gln Asn Ala Ala Phe Met Ala Gln Ala Val Gly Arg Leu Thr Gly Lys Pro Gly Val Val Leu Val Thr Ser Gly Pro Gly Ala Ser Asn Leu Ala Thr Gly Leu Leu Thr Ala Asn Thr Glu Gly Asp Pro Val Val Ala Leu Ala Gly Asn Val Ile Arg Ala Asp Arg Leu Lys Arg Thr His Gln Ser Leu Asp Asn Ala Ala Leu Phe Gln Pro Ile Thr Lys Tyr Ser Val Glu Val Gln Asp Val Lys Asn Ile Pro Glu Ala Val Thr Asn Ala Phe Arg Ile Ala Ser Ala Gly Gln Ala Gly Ala Ala Phe Val Ser Phe Pro Gln Asp Val Val Asn Glu Val Thr Asn Thr Lys Asn Val Arg Ala Val Ala Ala Pro Lys Leu Gly Pro Ala Ala Asp Asp Ala Ile Ser Ala Ala Ile Ala Lys Ile Gln Thr Ala Lys Leu Pro Val Val Leu Val Gly Met Lys Gly Gly Arg Pro Glu Ala Ile Lys Ala Val Arg Lys Leu Leu Lys Lys Val Gln Leu Pro Phe Val Glu Thr Tyr Gln Ala Ala Gly Thr Leu Ser Arg Asp Leu Glu Asp Gln Tyr Phe Gly Arg Ile Gly Leu Phe Arg Asn Gln Pro Gly Asp Leu Leu Glu Gln Ala Asp Val Val Leu Thr Ile Gly Tyr Asp Pro Ile Glu Tyr Asp Pro Lys Phe Trp Asn Ile Asn Gly Asp Arg Thr Ile Ile His Leu Asp Glu Ile Ile Ala Asp Ile Asp His Ala Tyr Gln Pro Asp Leu Glu Leu Ile Gly Asp Ile Pro Ser Thr Ile As<br/>n His Ile Glu His Asp Ala Val Lys Val Glu Phe Ala Glu Arg Glu Gln Lys Ile Leu Ser Asp Leu Lys Gln Tyr Met His Glu Gly Glu Gln Val Pro Ala Asp

208

Trp	Lys 370	Ser	Aab	Arg	Ala	His 375	Pro	Leu	Glu	Ile	Val 380	Lys	Glu	Leu	Arg	
Asn 385	Ala	Val	Aab	Asp	His 390	Val	Thr	Val	Thr	Cys 395	Asp	Ile	Gly	Ser	His 400	
Ala	Ile	Trp	Met	Ser 405	Arg	Tyr	Phe	Arg	Ser 410	Tyr	Glu	Pro	Leu	Thr 415	Leu	
Met	Ile	Ser	Asn 420	Gly	Met	Gln	Thr	Leu 425	Gly	Val	Ala	Leu	Pro 430	Trp	Ala	
Ile	Gly	Ala 435	Ser	Leu	Val	Lys	Pro 440	Gly	Glu	Lys	Val	Val 445	Ser	Val	Ser	
Gly	Asp 450	Gly	Gly	Phe	Leu	Phe 455	Ser	Ala	Met	Glu	Leu 460	Glu	Thr	Ala	Val	
Arg 465	Leu	Lys	Ala	Pro	Ile 470	Val	His	Ile	Val	Trp 475	Asn	Asp	Ser	Thr	Tyr 480	
Asp	Met	Val	Ala	Phe 485	Gln	Gln	Leu	Lys	Lys 490	Tyr	Asn	Arg	Thr	Ser 495	Ala	
Val	Asp	Phe	Gly 500	Asn	Ile	Asp	Ile	Val 505	Lys	Tyr	Ala	Glu	Ser 510	Phe	Gly	
Ala	Thr	Gly 515	Leu	Arg	Val	Glu	Ser 520	Pro	Asp	Gln	Leu	Ala 525	Asp	Val	Leu	
Arg	Gln 530	Gly	Met	Asn	Ala	Glu 535	Gly	Pro	Val	Ile	Ile 540	Asp	Val	Pro	Val	
Asp 545	Tyr	Ser	Asp	Asn	Ile 550	Asn	Leu	Ala	Ser	Asp 555	ГЛа	Leu	Pro	Lys	Glu 560	
Phe	Gly	Glu	Leu	Met 565	Lys	Thr	Lys	Ala	Leu 570							
<210 <210 <210	)> SH L> LH 2> TY	EQ II ENGTH (PE :	) NO 1: 34 PRT	37 13	orogi	ino		1420								
<210 <211 <212 <212 <213	0> SH L> LH 2> TY 3> OH 0> SH	EQ II ENGTH (PE : RGAN] EQUEN	D NO H: 34 PRT ISM: NCE:	37 13 Anae 37	erost	ipe	s ca	ccae								
<210 <211 <212 <212 <400 Met 1	D> SH L> LH 2> TY 3> OH D> SH Glu	EQ II ENGTH (PE : RGANI EQUEN Glu	) NO H: 34 PRT ISM: ICE: Cys	37 13 Anae 37 Lys 5	erost Met	ipe: Ala	з сас Lys	ccae Ile	Tyr 10	Tyr	Gln	Glu	Asp	Суз 15	Asn	
<210 <211 <211 <211 <400 Met 1 Leu	D> SH L> LH 2> TY 3> OF Glu Ser	EQ II ENGTH (PE: RGANI EQUEN Glu Leu	) NO H: 34 PRT ISM: ICE: Cys Leu 20	37 43 Anac 37 Lys 5 Asp	erost Met Gly	Lys	s cao Lys Thr	Ile Ile 25	Tyr 10 Ala	Tyr Val	Gln Ile	Glu Gly	Asp Tyr 30	Cys 15 Gly	Asn Ser	
<210 <211 <211 <211 <400 Met 1 Leu Gln	D> SH L> LH 2> TY 3> OH D> SH Glu Ser Gly	EQ II ENGTH (PE: (GAN) EQUEN Glu Leu His 35	) NO H: 34 PRT (SM: ICE: Cys Leu 20 Ala	37 Anae 37 Lys 5 Asp His	Met Gly Ala	Leu	Lys Thr Asn 40	Ile Ile 25 Ala	Tyr 10 Ala Lys	Tyr Val Glu	Gln Ile Ser	Glu Gly Gly 45	Asp Tyr 30 Cys	Cys 15 Gly Asn	Asn Ser Val	
<210 <211 <212 <212 <400 Met 1 Leu Gln Ile	D> SH L> LH 2> TY 3> OF Glu Ser Gly Ile 50	EQ II ENGTH (PE: RGANI EQUEN Glu Leu His 35 Gly	) NO H: 34 PRT ISM: ICE: Cys Leu 20 Ala Leu	37 Anad 37 Lys 5 Asp His Tyr	Met Gly Ala Glu	Leu Gly 55	Lys Thr Asn 40 Ala	Ile Ile 25 Ala Lys	Tyr 10 Ala Lys Glu	Tyr Val Glu Trp	Gln Ile Ser Lys 60	Glu Gly Gly 45 Arg	Asp Tyr 30 Cys Ala	Cys 15 Gly Asn Glu	Asn Ser Val Glu	
<211 <211 <212 <212 <400 Met 1 Leu Gln 65	<pre>&gt;&gt; SF &gt;&gt; LF &gt;&gt; LF &gt;&gt; TY &gt;&gt; OF Glu Ser Gly Ile 50 Gly</pre>	EQ II ENGTH (PE: CRANI EQUEN Glu Leu His 35 Gly Phe	) NO H: 34 PRT (SM: ISM: Cys Leu 20 Ala Leu Glu	37 13 Anaa 37 Lys 5 Asp His Tyr Val	Met Gly Ala Glu Tyr 70	Lipe: Ala Lys Leu Gly 55 Thr	Lys Thr Asn 40 Ala Ala	Ile Ile 25 Ala Lys Ala	Tyr 10 Ala Lys Glu Glu	Tyr Val Glu Trp Ala 75	Gln Ile Ser Lys 60 Ala	Glu Gly Gly 45 Arg Lys	Asp Tyr 30 Cys Ala Lys	Cys 15 Gly Asn Glu Ala	Asn Ser Val Glu Asp 80	
<211 <211 <212 <211 <400 Met 1 Leu Gln 65 Ile	<pre>&gt;&gt; SF &gt;&gt; LL &gt;&gt; TY &gt;&gt; OF Glu Ser Gly Ile 50 Gly Ile</pre>	EQ III ENGTH (PE: CGAN) EQUEN Glu Leu His 35 Gly Phe Met	) NO H: 33 PRT ISM: CYS Leu 20 Ala Leu Glu Ile	37 Anae 37 Lys 5 Asp His Tyr Val Leu 85	Met Gly Ala Glu Tyr 70 Ile	Ala Lys Leu Gly 55 Thr Asn	Lys Thr Asn Ala Ala Asp	Ile Ile 25 Ala Lys Ala Glu	Tyr 10 Ala Lys Glu Glu Lys 90	Tyr Val Glu Trp Ala 75 Gln	Gln Ile Ser Lys 60 Ala Ala	Glu Gly Gly 45 Arg Lys Thr	Asp Tyr 30 Cys Ala Lys Met	Cys 15 Gly Asn Glu Ala Tyr 95	Asn Ser Val Glu Asp 80 Lys	
<211 <211 <212 <212 <400 Met 1 Leu Gln 1 le Gln 65 Ile Asn	<pre>&gt;&gt; SF &gt;&gt; L &gt;&gt; T &gt;&gt; T &gt;&gt; OF Glu Ser Glu Ser Gly Ile So Gly Ile Asp</pre>	EQ III ENGTH (PE: :CGANI) EQUED Glu Leu His 35 Gly Phe Met Ile	) NO H: 34 PRT ISM: ICE: Cys Leu 20 Ala Leu Glu Ile Glu 100	37 43 Anae 37 Lys 5 Asp His Tyr Val Leu 85 Pro	Met Gly Ala Glu Tyr 70 Ile Asn	Lipe: Ala Lys Leu Gly 55 Thr Asn Leu	Lys Thr Asn Ala Ala Asp Glu	Ile Ile 25 Ala Lys Ala Glu Ala 105	Tyr 10 Ala Lys Glu Lys 90 Gly	Tyr Val Glu Trp Ala 75 Gln Asn	Gln Ile Ser Lys 60 Ala Ala Met	Glu Gly Gly 45 Arg Lys Thr Leu	Asp Tyr 30 Cys Ala Lys Met 110	Cys 15 Gly Asn Glu Ala Tyr 95 Phe	Asn Ser Val Glu Asp 80 Lys Ala	
<211 <211 <212 <212 <400 Met 1 Leu Gln 65 Ile Asn His	<pre>&gt;&gt; SI &gt;&gt; I &gt;&gt; T &gt;&gt; T &gt;&gt; T &gt;&gt; O Glu Ser Gly Ile Gly Ile Asp Gly</pre>	SQ III ENGTH (PE::CGAN) SQUEN Glu Leu His 35 Gly Phe Ile Phe 115	) NO H: 34 PRT ISM: ICE: Cys Leu 20 Ala Leu Glu 11e Glu 100 Asn	37 Anae 37 Lys 5 Asp His Tyr Val Leu 85 Pro Ile	Met Gly Ala Glu Tyr 70 Ile Asn His	Leu Ala Lys Leu Gly 55 Thr Asn Leu Phe	Lys Thr Asn 40 Ala Ala Glu Gly 120	Ile Ile 25 Ala Lys Ala Glu Ala 105 Cys	Tyr 10 Ala Lys Glu Glu Lys 90 Gly Ile	Tyr Val Glu Trp Ala 75 Gln Asn Val	Gln Ile Ser Lys 60 Ala Ala Met Pro	Glu Gly 45 Arg Lys Thr Leu Pro 125	Asp Tyr 30 Cys Ala Lys Met 110 Lys	Cys 15 Gly Asn Glu Ala Tyr 95 Phe Asp	Asn Ser Val Glu Asp 80 Lys Ala Val	
<211 <211 <211 <400 Met 1 Leu Gln Ile Gln 65 Ile Asn His Asp	<pre>&gt;&gt; SI &gt;&gt; SI &gt;&gt; TY &gt;&gt; TY &gt;&gt; OF Glu Ser Gly Ser Gly Ile So Gly Gly Cly Val 130</pre>	EQ II ENGTH CYPE:: CGANJ EQUEN Glu Leu His 35 Gly Phe Ile Phe 115 Thr	) NO H: 34 PRT (SM: (SM: Cys Leu 20 Ala Leu Glu 100 Asn Met	37 Anac 37 Lys 5 Asp His Tyr Val Leu 85 Pro Ile	Met Gly Ala Glu Tyr 70 Ile Asn His Ala	Ala Lys Leu Gly 55 Thr Asn Leu Phe Pro 135	Lys Thr Asn 40 Ala Ala Glu Glu 120 Lys	Ile Ile 25 Ala Lys Ala Glu Ala 105 Cys Gly	Tyr 10 Ala Lys Glu Glu Lys 90 Gly Ile Pro	Tyr Val Glu Trp Ala 75 Gln Asn Val Gly	Gln Ile Ser Lys 60 Ala Ala Met Pro His 140	Glu Gly 45 Arg Lys Thr Leu Pro 125 Thr	Asp Tyr 30 Cys Ala Lys Met 110 Lys Val	Cys 15 Gly Asn Glu Ala Tyr 95 Phe Asp Arg	Asn Ser Val Glu Asp 80 Lys Ala Val Ser	
<211 <211 <212 <211 <400 Met 1 Leu Gln 65 Ile Asn His Asp Glu 145	<pre>&gt;&gt; SI &gt;&gt; C &gt;&gt; T &gt;&gt; T &gt;&gt; T &gt;&gt; T &gt;&gt; T &gt;&gt; T =&gt; T =&gt; T</pre>	I le	) NO H: 34 PRT ISM: ICE: Cys Leu 20 Ala Leu Glu 100 Asn Met Glu	37 Anae 37 Lys 5 Asp His Tyr Val Leu 85 Pro Ile Gly	Met Gly Ala Glu Tyr 70 Ile Asn His Ala Lys 150	Leu Gly 55 Thr Asn Leu Phe Pro 135 Gly	Lys Thr Asn Ala Ala Glu Gly 120 Lys Val	Ile Ile 25 Ala Lys Ala Glu Ala 105 Cys Gly Pro	Tyr 10 Ala Lys Glu Glu Glu Lys 90 Gly Ile Pro Cys	Tyr Val Glu Trp Ala 75 Gln Asn Val Gly Leu 155	Gln Ile Ser Lys 60 Ala Ala Ala Pro His 140 Val	Glu Gly Gly 45 Lys Thr Leu Pro 125 Thr Ala	Asp Tyr 30 Cys Ala Lys Met 110 Lys Val Val	Cys 15 Gly Asn Glu Ala Tyr 95 Phe Asp Arg Glu	Asn Ser Val Glu Asp 80 Lys Ala Val Ser Gln 160	

209

т																
-	le	Gly	Gly	Ala 180	Arg	Ala	Gly	Val	Leu 185	Glu	Thr	Thr	Phe	Arg 190	Thr	Glu
Τ	hr	Glu	Thr 195	Asp	Leu	Phe	Gly	Glu 200	Gln	Ala	Val	Leu	Сув 205	Gly	Gly	Val
С	ÇAa	Ala 210	Leu	Met	Gln	Ala	Gly 215	Phe	Glu	Thr	Leu	Val 220	Glu	Ala	Gly	Tyr
A 2	Asp 25	Pro	Arg	Asn	Ala	Tyr 230	Phe	Glu	Суз	Ile	His 235	Glu	Met	Lys	Leu	Ile 240
V	Val	Asp	Leu	Ile	Tyr 245	Gln	Ser	Gly	Phe	Ser 250	Gly	Met	Arg	Tyr	Ser 255	Ile
S	Ser	Asn	Thr	Ala 260	Glu	Tyr	Gly	Asp	Tyr 265	Ile	Thr	Gly	Pro	Lys 270	Ile	Ile
Т	hr	Glu	Asp 275	Thr	Lys	Lys	Ala	Met 280	Lys	Lys	Ile	Leu	Ser 285	Aab	Ile	Gln
A	/ab	Gly 290	Thr	Phe	Ala	rÀa	Asp 295	Phe	Leu	Val	Asp	Met 300	Ser	Aab	Ala	Gly
S 3	Ser 805	Gln	Val	His	Phe	Lys 310	Ala	Met	Arg	Гла	Leu 315	Ala	Ser	Glu	His	Pro 320
A	Ala	Glu	Val	Val	Gly 325	Glu	Glu	Ile	Arg	Ser 330	Leu	Tyr	Ser	Trp	Ser 335	Asp
Ģ	Ju	Asp	Lys	Leu 340	Ile	Asn	Asn									
<210> SEQ ID NO 38 <211> LENGTH: 343 <212> TYPE: PRT <213> ORGANISM: Anaerostipes caccae																
<	:400	)> SE	EQUEN	ICE :	38											
M 1	let	Glu	Glu	Сув	Lys 5	Met	Ala	Гла	Ile	Tyr 10	Tyr	Gln	Glu	Asp	Сув 15	Asn
L	Jeu		_				Tite	Thr								
G		Ser	Leu	Leu 20	Asp	GIY	цув	1111	Ile 25	Ala	Val	Ile	Gly	Tyr 30	Gly	Ser
	Jln	Ser Gly	Leu His 35	Leu 20 Ala	Asp His	Gly Ala	Leu	Asn 40	Ile 25 Ala	Ala Lys	Val Glu	Ile Ser	Gly Gly 45	Tyr 30 Cys	Gly Asn	Ser Val
I	3ln :le	Ser Gly Ile 50	Leu His 35 Gly	Leu 20 Ala Leu	Asp His Tyr	Gly Ala Glu	Leu Gly 55	Asn 40 Ala	Ile 25 Ala Lys	Ala Lys Asp	Val Glu Trp	Ile Ser Lys 60	Gly Gly 45 Arg	Tyr 30 Cys Ala	Gly Asn Glu	Ser Val Glu
I G 6	3ln 1e 3ln 55	Ser Gly Ile 50 Gly	Leu His 35 Gly Phe	Leu 20 Ala Leu Glu	Asp His Tyr Val	Gly Ala Glu Tyr 70	Leu Gly 55 Thr	Asn 40 Ala Ala	Ile 25 Ala Lys Ala	Ala Lys Asp Glu	Val Glu Trp Ala 75	Ile Ser Lys 60 Ala	Gly Gly 45 Arg Lys	Tyr 30 Cys Ala Lys	Gly Asn Glu Ala	Ser Val Glu Asp 80
1 6 6	3ln 1e 3ln 55	Ser Gly Ile 50 Gly Ile	Leu His 35 Gly Phe Met	Leu 20 Ala Leu Glu Ile	Asp His Tyr Val Leu 85	Gly Ala Glu Tyr 70 Ile	Leu Gly 55 Thr Asn	Asn 40 Ala Ala Asp	Ile 25 Ala Lys Ala Glu	Ala Lys Asp Glu Lys 90	Val Glu Trp Ala 75 Gln	Ile Ser Lys 60 Ala Ala	Gly Gly 45 Arg Lys Thr	Tyr 30 Cys Ala Lys Met	Gly Asn Glu Ala Tyr 95	Ser Val Glu Asp 80 Lys
I G G G G G G G G G G G G G G G G G G G	Sln Sln S Sle	Ser Gly Ile 50 Gly Ile Asp	Leu His 35 Gly Phe Met Ile	Leu 20 Ala Leu Glu Ile Glu 100	Asp His Tyr Val Leu 85 Pro	Gly Ala Glu Tyr 70 Ile Asn	Leu Gly 55 Thr Asn Leu	Asn 40 Ala Ala Asp Glu	Ile 25 Ala Lys Ala Glu Ala 105	Ala Lys Asp Glu Lys 90 Gly	Val Glu Trp Ala 75 Gln Asn	Ile Ser Lys 60 Ala Ala Met	Gly Gly 45 Arg Lys Thr Leu	Tyr 30 Cys Ala Lys Met 110	Gly Asn Glu Ala Tyr 95 Phe	Ser Val Glu Asp 80 Lys Ala
I G G G G G G G G G G G G G G G G G G G	Sln Ile Sln 55 Ile Asn His	Ser Gly Ile 50 Gly Ile Asp Gly	Leu His 35 Gly Phe Met Ile Phe 115	Leu 20 Ala Leu Glu 11e Glu 100 Asn	Asp His Tyr Val Leu 85 Pro Ile	Gly Ala Glu Tyr 70 Ile Asn His	Leu Gly 55 Thr Asn Leu Phe	Asn 40 Ala Ala Asp Glu Glu 120	Ile 25 Ala Lys Ala Glu Ala 105 Cys	Ala Lys Asp Glu Lys 90 Gly Ile	Val Glu Trp Ala 75 Gln Asn Val	Ile Ser Lys 60 Ala Ala Met Pro	Gly Gly 45 Arg Lys Thr Leu Pro 125	Tyr 30 Cys Ala Lys Met 110 Lys	Gly Asn Glu Ala Tyr 95 Phe Asp	Ser Val Glu Asp 80 Lys Ala Val
I G G G G G G G G G G G G G G G G G G G	Sln Sln Sln Sl Sle Asn His	Ser Gly Ile 50 Gly Ile Asp Gly Val 130	Leu His 35 Gly Phe Met Ile Phe 115 Thr	Leu 20 Ala Leu Glu 10e Asn Met	Asp His Tyr Val Leu 85 Pro Ile Ile	Gly Ala Glu Tyr 70 Ile Asn His Ala	Leu Gly 55 Thr Asn Leu Phe Pro 135	Asn 40 Ala Ala Asp Glu Gly 120 Lys	Ile 25 Ala Lys Ala Glu Ala 105 Cys Gly	Ala Lys Asp Glu Lys 90 Gly Ile Pro	Val Glu Trp Ala 75 Gln Asn Val Gly	Ile Ser Lys 60 Ala Ala Met Pro His 140	Gly Gly 45 Arg Lys Thr Leu Pro 125 Thr	Tyr 30 Cys Ala Lys Met 110 Lys Val	Gly Asn Glu Ala Tyr 95 Phe Asp Arg	Ser Val Glu Asp 80 Lys Ala Val Ser
I A H A C I	Sln Sle Sle Asn His Asp Slu -45	Ser Gly Ile 50 Gly Ile Asp Gly Val 130 Tyr	Leu His 35 Gly Phe Met Ile Phe 115 Thr Glu	Leu 20 Ala Leu Glu 100 Asn Met Glu	Asp His Tyr Val Leu 85 Pro Ile Ile Gly	Gly Ala Glu Tyr 70 Ile Asn His Ala Lys 150	Lys Leu Gly 55 Thr Asn Leu Phe Pro 135 Gly	Asn 40 Ala Ala Asp Glu Gly 120 Lys Val	Ile 25 Ala Lys Ala Glu Ala 105 Cys Gly Pro	Ala Lys Glu Lys Gly Gly Ile Pro Cys	Val Glu Trp Ala 75 Gln Asn Val Gly Leu 155	Ile Ser Lys 60 Ala Ala Met Pro His 140 Val	Gly Gly 45 Arg Lys Thr Leu Pro 125 Thr Ala	Tyr 30 Cys Ala Lys Met 110 Lys Val Val	Gly Asn Glu Ala Tyr 95 Phe Asp Arg Glu	Ser Val Glu Asp 80 Lys Ala Val Ser Gln 160
I G G I A H A G I A	Sln Sln S Sln S Slu Asp Slu Asp	Ser Gly Ile 50 Gly Ile Asp Gly Val 130 Tyr Ala	Leu His 35 Gly Phe Met Ile Phe 115 Thr Glu Thr	Leu 20 Ala Leu Glu 100 Asn Glu Glu Glu	Asp His Tyr Val Leu 85 Pro Ile Gly Lys 165	Gly Ala Glu Tyr 70 Ile Asn His Ala Lys 150 Ala	Leu Gly 55 Thr Asn Leu Phe Pro 135 Gly Leu	Asn 40 Ala Ala Asp Glu Glu 120 Lys Val Asp	Ile 25 Ala Lys Ala Glu Ala 105 Cys Gly Pro Met	Ala Lys Glu Lys 90 Gly Ile Pro Cys Ala 170	Val Glu Trp Ala 75 Gln Asn Val Gly Leu 155 Leu	Ile Ser Lys 60 Ala Ala Met Pro His 140 Val Ala	Gly Gly 45 Arg Lys Thr Leu Pro 125 Thr Ala Tyr	Tyr 30 Cys Ala Lys Met 110 Lys Val Val	Gly Asn Glu Ala Tyr 95 Phe Asp Arg Glu Leu 175	Ser Val Glu Asp 80 Lys Ala Val Ser Gln 160 Ala
I G G G G G G G G G G G G G G G G G G G	Sln Sle Sln S S Le Asp S Lu S S Lu S S Lu S S S S S S S S S S	Ser Gly Ile 50 Gly Ile Asp Gly Val 130 Tyr Ala Gly	Leu His 35 Gly Phe Met Ile Phe 115 Thr Glu Thr Gly	Leu 20 Ala Glu 11e Glu 100 Asn Met Glu Glu Glu 20 Ala 180	Asp His Tyr Val Leu 85 Pro Ile Gly Lys 165 Arg	Gly Ala Glu Tyr 70 Ile Asn His Ala Lys 150 Ala Ala	Leu Gly 55 Thr Asn Leu Phe Pro 135 Gly Leu Gly	Asn 40 Ala Ala Asp Glu 120 Lys Val Asp Val	Ile 25 Ala Lys Ala Glu Ala 105 Cys Gly Pro Met Leu 185	Ala Lys Glu Lys Gly Ile Pro Cys Ala 170 Glu	Val Glu Trp Ala 75 Gln Asn Val Gly Leu 155 Leu Thr	Ile Ser Lys 60 Ala Ala Met Pro His 140 Val Ala Thr	Gly Gly 45 Arg Lys Thr Leu Pro 125 Thr Ala Tyr Phe	Tyr 30 Cys Ala Lys Met 110 Lys Val Val Ala Arg 190	Gly Asn Glu Ala Tyr 95 Phe Asp Glu Leu 175 Thr	Ser Val Glu Asp 80 Lys Ala Val Ser Gln 160 Ala Glu

211

	195				200					205				
Cys Ala 210	Leu Me	t Gln	Ala	Gly 215	Phe	Glu	Thr	Leu	Val 220	Glu	Ala	Gly	Tyr	
Asp Pro 225	Arg As	n Ala	Tyr 230	Phe	Glu	Суз	Ile	His 235	Glu	Met	Lys	Leu	Ile 240	
Val Asr	) Leu Il	e Tyr 245	Gln	Ser	Gly	Phe	Ser 250	Gly	Met	Arg	Tyr	Ser 255	Ile	
Ser Asr	1 Thr Al 26	a Glu 0	Tyr	Gly	Asp	Tyr 265	Ile	Thr	Gly	Pro	Lys 270	Ile	Ile	
Thr Glu	Asp Th 275	r Lys	Lys	Ala	Met 280	Lys	Lys	Ile	Leu	Ser 285	Asp	Ile	Gln	
Asp Gly 290	Thr Ph	e Ala	Lys	Asp 295	Phe	Leu	Val	Asp	Met 300	Ser	Asp	Ala	Gly	
Ser Glr 305	ı Val Hi	s Phe	Lys 310	Ala	Met	Arg	Lys	Leu 315	Ala	Ser	Glu	His	Pro 320	
Ala Glu	Val Va	1 Gly 325	Glu	Glu	Ile	Arg	Ser 330	Leu	Tyr	Ser	Trp	Ser 335	Asp	
Glu Asp	Lys Le 34	u Ile 0	Asn	Asn										
<210> S <211> I <212> T <213> C <220> F <223> C	EQ ID N ENGTH: YPE: DN RGANISM EATURE: THER IN	0 39 15539 A : Art: FORMA	ific: TION	ial : pLl	H468									
<400> 5	EQUENCE	: 39												
tcgcgcg	ittt cgg	tgatga	ac g	gtga	aaacc	tct	gaca	acat	gcaç	geteo	ccg g	gagad	eggtca	60
cagette	ıtct gta	agegga	at g	ccgg	gagca	a gao	caago	cccg	tcaç	gggcg	gag t	ccago	gggtg	120
ttggcgg	gtg tcg	gggct	gg ci	ttaa	ctatg	g cgé	gcato	caga	gcaç	gatto	gta 🤇	ctgag	gagtgc	180
accataa	att ccc	gtttta	aa g	agcti	tggtg	g ago	geta	agga	gtca	actgo	cca 🤉	ggtat	cgttt	240
gaacaco	Igca tta	gtcag	gg a	agtca	ataac	aca	agtco	ttt	cccç	gcaat	tt t	cttt	ttcta	300
ttactct	tgg cct	cctcta	ag ta	acaci	tctat	att:	tttt	tat	gcct	cggt	caa t	gatt	ttcat	360
tttttt	ttt cca	cctago	cg ga	atga	ctctt	ttt	ttt	ctt	agco	gatto	ggc a	attat	cacat	420
aatgaat	tat aca	ttata	ta a	agta	atgtg	g att	tctt	cga	agaa	atata	act a	aaaaa	aatgag	480
caggcaa	igat aaa	cgaago	gc ai	aagat	tgaca	ı gaç	gcaga	aaag	ccct	agta	aaa q	gegta	attaca	540
aatgaaa	icca aga	ttcaga	at to	gcgai	tctct	tta	aago	ggtg	gtco	ccta	age é	gatag	Jagcac	600
tcgatct	tcc cag	aaaaa	ga ge	gcaga	aagca	ı gta	igcaç	jaac	aggo	caca	aca a	atcgo	caagtg	660
teccost	aat caa	caggta	at ag	yggti gagti	utetg	g gao	cata	tga •taa	taca	arget	.ct (	ygeea	agcat	720
actores	act ccc	aatt	gu ti	gagu	gualt atcae	, ggt	.yact	aaad	acat	ayac	.ya (	acco	acacc	840
accyaac	act geg	yyarto	yu u	aasti	teas	, yet		aady	ayyo		•99 9	19000	agget c	64U 000
gyaguaa	aaa yyt	acacc	cu ag	tada				Jary	ayyo	acut		ayag	Joygug	900
gragate	tat att	acaygo	oc gi	acgo	agut	. yıc	-yaao	July	911	-ycaa	iay g	Jyaya	aayta	360
ggagato	tag or	ycgag:	at gi	atco	ogeat		.cttç	Jaaa	gett	.cgca	aga g	ygeta	igcaga	1020
attacco	ate tt	cugati	uy to	ouge	Jaggo	aaq	Jaato	Jate	atCa	accgt	ag t	-yaga	ayugog	1140
LLCAAGO	and ttg	oggtte	90 Ci	ataa	Jagaa	, geo	acct	.cgc	ccaa	auggt	ac (	aado	Jacgtt	1000
CCCTCCa	ieca aag	yrgtt(	et ta	augta	aytga	cad	cgat:	Lat	τταα	aget	.gc a	₄ycat	acgat	T200

		-continued	
atatatacat gtgtatatat	gtatacctat gaatgtcagt	aagtatgtat acgaacagta	1260
tgatactgaa gatgacaagg	taatgcatca ttctatacgt	gtcattctga acgaggcgcg	1320
ctttcctttt ttctttttgc	tttttctttt tttttctctt	gaactcgacg gatctatgcg	1380
gtgtgaaata ccgcacagat	gcgtaaggag aaaataccgo	c atcaggaaat tgtaagcgtt	1440
aatattttgt taaaattcgc	gttaaatttt tgttaaatca	a gctcattttt taaccaatag	1500
gccgaaatcg gcaaaatccc	ttataaatca aaagaataga	a ccgagatagg gttgagtgtt	1560
gttccagttt ggaacaagag	tccactatta aagaacgtgg	g actccaacgt caaagggcga	1620
aaaaccgtct atcagggcga	tggcccacta cgtgaaccat	c caccctaatc aagttttttg	1680
gggtcgaggt gccgtaaagc	actaaatcgg aaccctaaaq	g ggageeeeeg atttagaget	1740
tgacgggggaa agccggcgaa	cgtggcgaga aaggaaggga	a agaaagcgaa aggagcgggc	1800
gctagggcgc tggcaagtgt	agcggtcacg ctgcgcgtaa	a ccaccacacc cgccgcgctt	1860
aatgcgccgc tacagggcgc	gtccattcgc cattcaggct	t gcgcaactgt tgggaagggc	1920
gcggtgcggg cctcttcgct	attacgccag ctggcgaaag	g ggggatgtgc tgcaaggcga	1980
ttaagttggg taacgccagg	gttttcccag tcacgacgtt	: gtaaaacgac ggccagtgag	2040
cgcgcgtaat acgactcact	atagggcgaa ttgggtacco	g ggeeeeeet egaggtegae	2100
ggcgcgccac tggtagagag	cgactttgta tgccccaatt	gegaaaceeg egatateett	2160
ctcgattctt tagtacccga	ccaggacaag gaaaaggagg	g tcgaaacgtt tttgaagaaa	2220
caagaggaac tacacggaag	ctctaaagat ggcaaccago	c cagaaactaa gaaaatgaag	2280
ttgatggatc caactggcac	cgctggcttg aacaacaata	a ccagccttcc aacttctgta	2340
aataacggcg gtacgccagt	gccaccagta ccgttacctt	t toggtatadd tootttoodd	2400
atgtttccaa tgcccttcat	gcctccaacg gctactatca	a caaatcctca tcaagctgac	2460
gcaagcccta agaaatgaat	aacaatactg acagtactaa	a ataattgeet aettggette	2520
acatacgttg catacgtcga	tatagataat aatgataato	g acagcaggat tatcgtaata	2580
cgtaatagct gaaaatctca	aaaatgtgtg ggtcattaco	g taaataatga taggaatggg	2640
attettetat tttteetttt	tccattctag cagccgtcgg	g gaaaacgtgg catcctctct	2700
ttcgggctca attggagtca	cgctgccgtg agcatcctct	: ctttccatat ctaacaactg	2760
agcacgtaac caatggaaaa	gcatgagett agegttgete	c caaaaaagta ttggatggtt	2820
aataccattt gtctgttctc	ttctgacttt gactcctcaa	a aaaaaaaaat ctacaatcaa	2880
cagatcgctt caattacgcc	ctcacaaaaa cttttttcct	t tettettege ceaegttaaa	2940
ttttatccct catgttgtct	aacggatttc tgcacttgat	t ttattataaa aagacaaaga	3000
cataatactt ctctatcaat	ttcagttatt gttcttcctt	gegttattet tetgttette	3060
tttttctttt gtcatatata	accataacca agtaatacat	attcaaacta gtatgactga	3120
caaaaaaact cttaaagact	taagaaatcg tagttctgtt	tacgattcaa tggttaaatc	3180
acctaatcgt gctatgttgc	gtgcaactgg tatgcaagat	gaagactttg aaaaacctat	3240
cgtcggtgtc atttcaactt	gggctgaaaa cacaccttgt	aatatccact tacatgactt	3300
tggtaaacta gccaaagtcg	gtgttaagga agctggtgct	t tggccagttc agttcggaac	3360
aatcacggtt tctgatggaa	tcgccatggg aacccaagga	a atgcgtttct ccttgacatc	3420
tcgtgatatt attgcagatt	ctattgaagc agccatggga	a ggtcataatg cggatgcttt	3480
tgtagccatt ggcggttgtg	ataaaaacat gcccggttct	: gttatcgcta tggctaacat	3540
ggatatccca gccattttg	cttacggcgg aacaattgca	a cctggtaatt tagacggcaa	3600

215

agatatcgat	ttagtctctg	tctttgaagg	tgtcggccat	tggaaccacg	gcgatatgac	3660
caaagaagaa	gttaaagctt	tggaatgtaa	tgcttgtccc	ggtcctggag	gctgcggtgg	3720
tatgtatact	gctaacacaa	tggcgacagc	tattgaagtt	ttgggactta	gccttccggg	3780
ttcatcttct	cacccggctg	aatccgcaga	aaagaaagca	gatattgaag	aagctggtcg	3840
cgctgttgtc	aaaatgctcg	aaatgggctt	aaaaccttct	gacattttaa	cgcgtgaagc	3900
ttttgaagat	gctattactg	taactatggc	tctgggaggt	tcaaccaact	caacccttca	3960
cctcttagct	attgcccatg	ctgctaatgt	ggaattgaca	cttgatgatt	tcaatacttt	4020
ccaagaaaaa	gttcctcatt	tggctgattt	gaaaccttct	ggtcaatatg	tattccaaga	4080
cctttacaag	gtcggagggg	taccagcagt	tatgaaatat	ctccttaaaa	atggcttcct	4140
tcatggtgac	cgtatcactt	gtactggcaa	aacagtcgct	gaaaatttga	aggettttga	4200
tgatttaaca	cctggtcaaa	aggttattat	gccgcttgaa	aatcctaaac	gtgaagatgg	4260
tccgctcatt	attctccatg	gtaacttggc	tccagacggt	gccgttgcca	aagtttctgg	4320
tgtaaaagtg	cgtcgtcatg	tcggtcctgc	taaggtettt	aattetgaag	aagaagccat	4380
tgaagctgtc	ttgaatgatg	atattgttga	tggtgatgtt	gttgtcgtac	gttttgtagg	4440
accaaagggc	ggtcctggta	tgcctgaaat	gettteeett	tcatcaatga	ttgttggtaa	4500
agggcaaggt	gaaaaagttg	cccttctgac	agatggccgc	ttctcaggtg	gtacttatgg	4560
tcttgtcgtg	ggtcatatcg	ctcctgaagc	acaagatggc	ggtccaatcg	cctacctgca	4620
aacaggagac	atagtcacta	ttgaccaaga	cactaaggaa	ttacactttg	atatctccga	4680
tgaagagtta	aaacatcgtc	aagagaccat	tgaattgcca	ccgctctatt	cacgcggtat	4740
ccttggtaaa	tatgctcaca	tcgtttcgtc	tgcttctagg	ggagccgtaa	cagacttttg	4800
gaagcctgaa	gaaactggca	aaaaatgttg	tcctggttgc	tgtggttaag	cggccgcgtt	4860
aattcaaatt	aattgatata	gtttttaat	gagtattgaa	tctgtttaga	aataatggaa	4920
tattatttt	atttatttat	ttatattatt	ggtcggctct	tttcttctga	aggtcaatga	4980
caaaatgata	tgaaggaaat	aatgatttct	aaaattttac	aacgtaagat	atttttacaa	5040
aagcctagct	catcttttgt	catgcactat	tttactcacg	cttgaaatta	acggccagtc	5100
cactgcggag	tcatttcaaa	gtcatcctaa	tcgatctatc	gtttttgata	gctcattttg	5160
gagttcgcga	ttgtcttctg	ttattcacaa	ctgttttaat	ttttatttca	ttctggaact	5220
cttcgagttc	tttgtaaagt	ctttcatagt	agcttacttt	atcctccaac	atatttaact	5280
tcatgtcaat	ttcggctctt	aaattttcca	catcatcaag	ttcaacatca	tcttttaact	5340
tgaatttatt	ctctagctct	tccaaccaag	cctcattgct	ccttgattta	ctggtgaaaa	5400
gtgatacact	ttgcgcgcaa	tccaggtcaa	aactttcctg	caaagaattc	accaatttct	5460
cgacatcata	gtacaatttg	ttttgttctc	ccatcacaat	ttaatatacc	tgatggattc	5520
ttatgaagcg	ctgggtaatg	gacgtgtcac	tctacttcgc	cttttccct	actcctttta	5580
gtacggaaga	caatgctaat	aaataagagg	gtaataataa	tattattaat	cggcaaaaaa	5640
gattaaacgc	caagcgttta	attatcagaa	agcaaacgtc	gtaccaatcc	ttgaatgctt	5700
cccaattgta	tattaagagt	catcacagca	acatattctt	gttattaaat	taattattat	5760
tgatttttga	tattgtataa	aaaaaccaaa	tatgtataaa	aaaagtgaat	aaaaaatacc	5820
aagtatggag	aaatatatta	gaagtctata	cgttaaacca	cccgggcccc	ccctcgaggt	5880
cgacggtatc	gataagcttg	atatcgaatt	cctgcagccc	gggggatcca	ctagttctag	5940

217

agcggccgct	ctagaactag	taccacaggt	gttgtcctct	gaggacataa	aatacacacc	6000
gagattcatc	aactcattgc	tggagttagc	atatctacaa	ttgggtgaaa	tggggagcga	6060
tttgcaggca	tttgctcggc	atgccggtag	aggtgtggtc	aataagagcg	acctcatgct	6120
atacctgaga	aagcaacctg	acctacagga	aagagttact	caagaataag	aattttcgtt	6180
ttaaaaccta	agagtcactt	taaaatttgt	atacacttat	ttttttata	acttatttaa	6240
taataaaaat	cataaatcat	aagaaattcg	cttactctta	attaatcaaa	aagttaaaat	6300
tgtacgaata	gattcaccac	ttcttaacaa	atcaaaccct	tcattgattt	tctcgaatgg	6360
caatacatgt	gtaattaaag	gatcaagagc	aaacttcttc	gccataaagt	cggcaacaag	6420
ttttggaaca	ctatccttgc	tcttaaaacc	gccaaatata	gctcccttcc	atgtacgacc	6480
gcttagcaac	agcataggat	tcatcgacaa	attttgtgaa	tcaggaggaa	cacctacgat	6540
cacactgact	ccatatgcct	cttgacagca	ggacaacgca	gttaccatag	tatcaagacg	6600
gcctataact	tcaaaagaga	aatcaactcc	accgtttgac	atttcagtaa	ggacttcttg	6660
tattggtttc	ttataatctt	gagggttaac	acattcagta	gccccgacct	ccttagcttt	6720
tgcaaatttg	tccttattga	tgtctacacc	tataatcctc	gctgcgcctg	cagctttaca	6780
ccccataata	acgcttagtc	ctactcctcc	taaaccgaat	actgcacaag	tcgaaccctg	6840
tgtaaccttt	gcaactttaa	ctgcggaacc	gtaaccggtg	gaaaatccgc	accctatcaa	6900
gcaaactttt	tccagtggtg	aagctgcatc	gattttagcg	acagatatct	cgtccaccac	6960
tgtgtattgg	gaaaatgtag	aagtaccaag	gaaatggtgt	ataggtttcc	ctctgcatgt	7020
aaatctgctt	gtaccatcct	gcatagtacc	tctaggcata	gacaaatcat	ttttaaggca	7080
gaaattaccc	tcaggatgtt	tgcagactct	acacttacca	cattgaggag	tgaacagtgg	7140
gatcacttta	tcaccaggac	gaacagtggt	aacaccttca	cctatggatt	caacgattcc	7200
ggcagcctcg	tgtcccgcga	ttactggcaa	aggagtaact	agagtgccac	tcaccacatg	7260
gtcgtcggat	ctacagattc	cggtggcaac	catcttgatt	ctaacctcgt	gtgcttttgg	7320
tggcgctact	tctacttctt	ctatgctaaa	cggctttttc	tcttcccaca	aaactgccgc	7380
tttacactta	ataactttac	cggctgttga	catcctcagc	tagctattgt	aatatgtgtg	7440
tttgtttgga	ttattaagaa	gaataattac	aaaaaaatt	acaaaggaag	gtaattacaa	7500
cagaattaag	aaaggacaag	aaggaggaag	agaatcagtt	cattatttct	tctttgttat	7560
ataacaaacc	caagtagcga	tttggccata	cattaaaagt	tgagaaccac	cctccctggc	7620
aacagccaca	actcgttacc	attgttcatc	acgatcatga	aactcgctgt	cagctgaaat	7680
ttcacctcag	tggatctctc	ttttattct	tcatcgttcc	actaaccttt	ttccatcagc	7740
tggcagggaa	cggaaagtgg	aatcccattt	agcgagcttc	ctctttctt	caagaaaaga	7800
cgaagcttgt	gtgtgggtgc	gcgcgctagt	atctttccac	attaagaaat	ataccataaa	7860
ggttacttag	acatcactat	ggctatatat	atatatatat	atatatgtaa	cttagcacca	7920
tcgcgcgtgc	atcactgcat	gtgttaaccg	aaaagtttgg	cgaacacttc	accgacacgg	7980
tcatttagat	ctgtcgtctg	cattgcacgt	cccttagcct	taaatcctag	gcgggagcat	8040
tctcgtgtaa	ttgtgcagcc	tgcgtagcaa	ctcaacatag	cgtagtctac	ccagtttttc	8100
aagggtttat	cgttagaaga	ttctcccttt	tcttcctgct	cacaaatctt	aaagtcatac	8160
attgcacgac	taaatgcaag	catgcggatc	ccccgggctg	caggaattcg	atatcaagct	8220
tatcgatacc	gtcgactggc	cattaatctt	tcccatatta	gatttcgcca	agccatgaaa	8280
gttcaagaaa	ggtetttaga	cgaattaccc	ttcatttctc	aaactggcgt	caagggatcc	8340

219

tggtatggtt	ttatcgtttt	atttctggtt	cttatagcat	cgttttggac	ttctctgttc	8400
ccattaggcg	gttcaggagc	cagcgcagaa	tcattctttg	aaggatactt	atcctttcca	8460
attttgattg	tctgttacgt	tggacataaa	ctgtatacta	gaaattggac	tttgatggtg	8520
aaactagaag	atatggatct	tgataccggc	agaaaacaag	tagatttgac	tcttcgtagg	8580
gaagaaatga	ggattgagcg	agaaacatta	gcaaaaagat	ccttcgtaac	aagattttta	8640
catttctggt	gttgaaggga	aagatatgag	ctatacagcg	gaatttccat	atcactcaga	8700
ttttgttatc	taattttttc	cttcccacgt	ccgcgggaat	ctgtgtatat	tactgcatct	8760
agatatatgt	tatcttatct	tggcgcgtac	atttaatttt	caacgtattc	tataagaaat	8820
tgcgggagtt	tttttcatgt	agatgatact	gactgcacgc	aaatataggc	atgatttata	8880
ggcatgattt	gatggctgta	ccgataggaa	cgctaagagt	aacttcagaa	tcgttatcct	8940
ggcggaaaaa	attcatttgt	aaactttaaa	aaaaaagcc	aatatcccca	aaattattaa	9000
gagcgcctcc	attattaact	aaaatttcac	tcagcatcca	caatgtatca	ggtatctact	9060
acagatatta	catgtggcga	aaaagacaag	aacaatgcaa	tagcgcatca	agaaaaaaca	9120
caaagctttc	aatcaatgaa	tcgaaaatgt	cattaaaata	gtatataaat	tgaaactaag	9180
tcataaagct	ataaaaagaa	aatttattta	aatgcaagat	ttaaagtaaa	ttcacggccc	9240
tgcaggcctc	agctcttgtt	ttgttctgca	aataacttac	ccatctttt	caaaacttta	9300
ggtgcaccct	cctttgctag	aataagttct	atccaataca	tcctatttgg	atctgcttga	9360
gcttctttca	tcacggatac	gaattcattt	tctgttctca	caattttgga	cacaactctg	9420
tcttccgttg	ccccgaaact	ttctggcagt	tttgagtaat	tccacatagg	aatgtcatta	9480
taactctggt	tcggaccatg	aatttccctc	tcaaccgtgt	aaccatcgtt	attaatgata	9540
aagcagattg	ggtttatctt	ctctctaatg	gctagtccta	attcttggac	agtcagttgc	9600
aatgatccat	ctccgataaa	caataaatgt	ctagattctt	tatctgcaat	ttggctgcct	9660
agagetgegg	ggaaagtgta	tcctatagat	ccccacaagg	gttgaccaat	aaaatgtgat	9720
ttcgatttca	gaaatataga	tgaggcaccg	aagaaagaag	tgccttgttc	agccacgatc	9780
gtctcattac	tttgggtcaa	attttcgaca	gcttgccaca	gtctatcttg	tgacaacagc	9840
gcgttagaag	gtacaaaatc	ttcttgcttt	ttatctatgt	acttgccttt	atattcaatt	9900
tcggacaagt	caagaagaga	tgatatcagg	gattcgaagt	cgaaattttg	gattettteg	9960
ttgaaaattt	taccttcatc	gatattcaag	gaaatcattt	tattttcatt	aagatggtga	10020
gtaaatgcac	ccgtactaga	atcggtaagc	tttacaccca	acataagaat	aaaatcagca	10080
gattccacaa	attccttcaa	gtttggctct	gacagagtac	cgttgtaaat	ccccaaaaat	10140
gagggcaatg	cttcatcaac	agatgattta	ccaaagttca	aagtagtaat	aggtaactta	10200
gtctttgaaa	taaactgagt	aacagtcttc	tctaggccga	acgatataat	ttcatggcct	10260
gtgattacaa	ttggtttctt	ggcattcttc	agactttcct	gtattttgtt	cagaatctct	10320
tgatcagatg	tattcgacgt	ggaattttcc	ttcttaagag	gcaaggatgg	tttttcagcc	10380
ttagcggcag	ctacatctac	aggtaaattg	atgtaaaccg	gctttctttc	ctttagtaag	10440
gcagacaaca	ctctatcaat	ttcaacagtt	gcattctcgg	ctgtcaataa	agtcctggca	10500
gcagtaaccg	gttcgtgcat	cttcataaag	tgcttgaaat	caccatcagc	caacgtatgg	10560
tgaacaaact	taccttcgtt	ctgcactttc	gaggtaggag	atcccacgat	ctcaacaaca	10620
ggcaggttct	cagcatagga	gcccgctaag	ccattaactg	cggataattc	gccaacacca	10680

221

#### -continued

aatgtagtca agaatgccgc agcctttttc gttcttgcgt acccgtcggc catataggag 10740 gcatttaact cattagcatt tcccacccat ttcatatctt tgtgtgaaat aatttgatct 10800 agaaattgca aattgtagtc acctggtact ccgaatattt cttctatacc taattcgtgt 10860 aatctgtcca acagatagtc acctactgta tacattttgt ttactagttt atgtgtgttt 10920 attcgaaact aagttcttgg tgttttaaaa ctaaaaaaaa gactaactat aaaagtagaa 10980 tttaagaagt ttaagaaata gatttacaga attacaatca atacctaccg tctttatata 11040 cttattagtc aagtaggggga ataatttcag ggaactggtt tcaacctttt ttttcagctt 11100 tttccaaatc agagagagca gaaggtaata gaaggtgtaa gaaaatgaga tagatacatg 11160 cgtgggtcaa ttgccttgtg tcatcattta ctccaggcag gttgcatcac tccattgagg 11220 ttgtgcccgt tttttgcctg tttgtgcccc tgttctctgt agttgcgcta agagaatgga 11280 cctatgaact gatggttggt gaagaaaaca atattttggt gctgggattc ttttttttc 11340 tggatgccag cttaaaaaqc gggctccatt atatttagtg gatgccagga ataaactgtt 11400 cacccagaca cctacgatgt tatatattct gtgtaacccg ccccctattt tgggcatgta 11460 cgggttacag cagaattaaa aggctaattt tttgactaaa taaagttagg aaaatcacta 11520 ctattaatta tttacgtatt ctttgaaatg gcagtattga taatgataaa ctcgaactga 11580 aaaagcgtgt tttttattca aaatgattct aactccctta cgtaatcaag gaatcttttt 11640 gccttggcct ccgcgtcatt aaacttcttg ttgttgacgc taacattcaa cgctagtata 11700 tattcgtttt tttcaggtaa gttcttttca acgggtctta ctgatgaggc agtcgcgtct 11760 gaacctgtta agaggtcaaa tatgtcttct tgaccgtacg tgtcttgcat gttattagct 11820 ttgggaattt gcatcaagtc ataggaaaat ttaaatcttg gctctcttgg gctcaaggtg 11880 acaaggteet egaaaatagg gegegeeeea eegeggtgga geteeagett ttgtteeett 11940 tagtgagggt taattgcgcg cttggcgtaa tcatggtcat agctgtttcc tgtgtgaaat 12000 tgttatccgc tcacaattcc acacaacata cgagccggaa gcataaagtg taaagcctgg 12060 ggtgcctaat gagtgagcta actcacatta attgcgttgc gctcactgcc cgctttccag 12120 tcgggaaacc tgtcgtgcca gctgcattaa tgaatcggcc aacgcgcggg gagaggcggt 12180 ttgcgtattg ggcgctcttc cgcttcctcg ctcactgact cgctgcgctc ggtcgttcgg 12240 ctgcggcgag cggtatcagc tcactcaaag gcggtaatac ggttatccac agaatcaggg 12300 gataacgcag gaaagaacat gtgagcaaaa ggccagcaaa aggccaggaa ccgtaaaaag 12360 gccgcgttgc tggcgttttt ccataggctc cgccccctg acgagcatca caaaaatcga 12420 cgctcaagtc agaggtggcg aaacccgaca ggactataaa gataccaggc gtttccccct 12480 ggaageteee tegtgegete teetgtteeg accetgeege ttaceggata cetgteegee 12540 tttctccctt cgggaagcgt ggcgctttct catagctcac gctgtaggta tctcagttcg 12600 gtgtaggtcg ttcgctccaa gctgggctgt gtgcacgaac cccccgttca gcccgaccgc 12660 tgegeettat eeggtaacta tegtettgag teeaaceegg taagacaega ettategeea 12720 ctggcagcag ccactggtaa caggattagc agagcgaggt atgtaggcgg tgctacagag 12780 ttettgaagt ggtggeetaa etaeggetae aetagaagaa eagtatttgg tatetgeget 12840 ctgctgaagc cagttacctt cggaaaaaga gttggtagct cttgatccgg caaacaaacc 12900 accgctggta gcggtggttt ttttgtttgc aagcagcaga ttacgcgcag aaaaaaagga 12960 teteaagaag ateetttgat ettttetaeg gggtetgaeg eteagtggaa egaaaaetea 13020 cgttaaggga ttttggtcat gagattatca aaaaggatct tcacctagat ccttttaaat 13080

223

#### -continued

taaaaatgaa gttttaaatc aatctaaagt atatatgagt aaacttggtc tgacagttac 13140 caatgettaa teagtgagge acctatetea gegatetgte tatttegtte atceatagtt 13200 gcctgactcc ccgtcgtgta gataactacg atacgggagg gcttaccatc tggccccagt 13260 gctgcaatga taccgcgaga cccacgctca ccggctccag atttatcagc aataaaccag 13320 ccagccggaa gggccgagcg cagaagtggt cctgcaactt tatccgcctc catccagtct 13380 attaattgtt gccgggaagc tagagtaagt agttcgccag ttaatagttt gcgcaacgtt 13440 gttgccattg ctacaggcat cgtggtgtca cgctcgtcgt ttggtatggc ttcattcagc 13500 tccggttccc aacgatcaag gcgagttaca tgatccccca tgttgtgcaa aaaagcggtt 13560 ageteetteg gteeteegat egttgteaga agtaagttgg eegeagtgtt ateacteatg 13620 gttatggcag cactgcataa ttctcttact gtcatgccat ccgtaagatg cttttctgtg 13680 actggtgagt actcaaccaa gtcattctga gaatagtgta tgcggcgacc gagttgctct 13740 tgcccggcgt caatacggga taataccgcg ccacatagca gaactttaaa agtgctcatc 13800 attggaaaac gttcttcggg gcgaaaactc tcaaggatct taccgctgtt gagatccagt 13860 tcgatgtaac ccactcgtgc acccaactga tcttcagcat cttttacttt caccagcgtt 13920 tctgggtgag caaaaacagg aaggcaaaat gccgcaaaaa agggaataag ggcgacacgg 13980 aaatgttgaa tactcatact cttccttttt caatattatt gaagcattta tcagggttat 14040 tqtctcatqa qcqqatacat atttqaatqt atttaqaaaa ataaacaaat aqqqqttccq 14100 cgcacatttc cccgaaaagt gccacctgaa cgaagcatct gtgcttcatt ttgtagaaca 14160 aaaatgcaac gcgagagcgc taatttttca aacaaagaat ctgagctgca tttttacaga 14220 acagaaatgc aacgcgaaag cgctatttta ccaacgaaga atctgtgctt catttttgta 14280 aaacaaaaat gcaacgcgag agcgctaatt tttcaaacaa agaatctgag ctgcattttt 14340 acagaacaga aatgcaacgc gagagcgcta ttttaccaac aaagaatcta tacttctttt 14400 ttgttctaca aaaatgcatc ccgagagcgc tatttttcta acaaagcatc ttagattact 14460 ttttttctcc tttgtgcgct ctataatgca gtctcttgat aactttttgc actgtaggtc 14520 cgttaaggtt agaagaaggc tactttggtg tctattttct cttccataaa aaaagcctga 14580 ctccacttcc cgcgtttact gattactagc gaagctgcgg gtgcattttt tcaagataaa 14640 ggcatccccg attatattct ataccgatgt ggattgcgca tactttgtga acagaaagtg 14700 atagcgttga tgattettea ttggteagaa aattatgaae ggtttettet attttgtete 14760 tatatactac gtataggaaa tgtttacatt ttcgtattgt tttcgattca ctctatgaat 14820 agttcttact acaatttttt tgtctaaaga gtaatactag agataaacat aaaaaatgta 14880 gaggtcgagt ttagatgcaa gttcaaggag cgaaaggtgg atgggtaggt tatataggga 14940 tatagcacag agatatatag caaagagata cttttgagca atgtttgtgg aagcggtatt 15000 cgcaatattt tagtagctcg ttacagtccg gtgcgttttt ggttttttga aagtgcgtct 15060 tcagagcgct tttggttttc aaaagcgctc tgaagttcct atactttcta gagaatagga 15120 actteggaat aggaaettea aagegttee gaaaaegage getteegaaa atgeaaegeg 15180 agetgegeae atacagetea etgtteaegt egeaeetata tetgegtgtt geetgtatat 15240 atatatacat gagaagaacg gcatagtgcg tgtttatgct taaatgcgta cttatatgcg 15300 tetatttatg taggatgaaa ggtagtetag taeeteetgt gatattatee catteeatge 15360 qqqqtatcqt atqcttcctt caqcactacc ctttaqctqt tctatatqct qccactcctc 15420

226

#### -continued

aattggatta gtctcatcct tcaatgctat catttccttt gatattggat catactaaga 15480 aaccattatt atcatgacat taacctataa aaataggegt atcaegagge eetttegte 15539 <210> SEQ ID NO 40 <211> LENGTH: 1128 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Codon-optimized polynucleotide sequence-Horseliver ADH <400> SEQUENCE: 40 atgtcaacag ccggtaaagt tattaagtgt aaagcggcag ttttgtggga agagaaaaag 60 ccgtttagca tagaagaagt agaagtagcg ccaccaaaag cacacgaggt tagaatcaag 120 atggttgcca ccggaatctg tagatccgac gaccatgtgg tgagtggcac tctagttact 180 cctttgccag taatcgcggg acacgaggct gccggaatcg ttgaatccat aggtgaaggt 240 gttaccactg ttcgtcctgg tgataaagtg atcccactgt tcactcctca atgtggtaag 300 tgtagagtet geaaacatee tgagggtaat ttetgeetta aaaatgattt gtetatgeet 360 aqaqqtacta tqcaqqatqq tacaaqcaqa tttacatqca qaqqqaaacc tatacaccat 420 tteettggta ettetacatt tteecaatac acagtggtgg acgagatate tgtegetaaa 480 atcgatgcag cttcaccact ggaaaaagtt tgcttgatag ggtgcggatt ttccaccggt 540 tacggttccg cagttaaagt tgcaaaggtt acacagggtt cgacttgtgc agtattcggt 600 ttaggaggag taggactaag cgttattatg gggtgtaaag ctgcaggcgc agcgaggatt 660 ataggtgtag acatcaataa ggacaaattt gcaaaagcta aggaggtcgg ggctactgaa 720 tgtgttaacc ctcaagatta taagaaacca atacaagaag tccttactga aatgtcaaac 780 ggtggagttg atttctcttt tgaagttata ggccgtcttg atactatggt aactgcgttg 840 teetgetgte aagaggeata tggagteagt gtgategtag gtgtteetee tgatteacaa 900 aatttgtcga tgaatcctat gctgttgcta agcggtcgta catggaaggg agctatattt 960 ggcggtttta agagcaagga tagtgttcca aaacttgttg ccgactttat ggcgaagaag 1020 tttgctcttg atcctttaat tacacatgta ttgccattcg agaaaatcaa tgaagggttt 1080 gatttgttaa gaagtggtga atctattcgt acaattttaa ctttttga 1128 <210> SEQ ID NO 41 <211> LENGTH: 1023 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Saccharomyces cerevisiae TDH3 promoter <400> SEQUENCE: 41 caccgcggtg gggcgcgccc tattttcgag gaccttgtca ccttgagccc aagagagcca 60 agatttaaat tttcctatga cttgatgcaa attcccaaag ctaataacat gcaagacacg 120 tacggtcaag aagacatatt tgacctctta acaggttcag acgcgactgc ctcatcagta 180 agacccgttg aaaagaactt acctgaaaaa aacgaatata tactagcgtt gaatgttagc 240 gtcaacaaca agaagtttaa tgacgcggag gccaaggcaa aaagattcct tgattacgta 300 agggagttag aatcattttg aataaaaaac acgctttttc agttcgagtt tatcattatc 360 aatactgcca tttcaaagaa tacgtaaata attaatagta gtgattttcc taactttatt 420 tagtcaaaaa attagcettt taattetget gtaaceegta catgeecaaa atagggggeg 480

- 1	1	-
- Z	z	1

#### -continued

ggttacacag aatatataac atcgtaggtg tctgggtgaa cagtttattc ctggcatcca	540
ctaaatataa tggagcccgc tttttaagct ggcatccaga aaaaaaaaga atcccagcac	600
caaaatattg ttttcttcac caaccatcag ttcataggtc cattctctta gcgcaactac	660
agagaacagg ggcacaaaca ggcaaaaaac gggcacaacc tcaatggagt gatgcaacct	720
gcctggagta aatgatgaca caaggcaatt gacccacgca tgtatctatc tcattttctt	780
acaccttcta ttaccttctg ctctctcga tttggaaaaa gctgaaaaaa aaggttgaaa	840
ccagttccct gaaattattc ccctacttga ctaataagta tataaagacg gtaggtattg	900
attgtaatto tgtaaatota tttottaaao ttottaaatt otaottttat agttagtott	960
ttttttagtt ttaaaacacc aagaacttag tttcgaataa acacacataa actagtaaac	1020
aaa	1023
<210> SEQ ID NO 42 <211> LENGTH: 21 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Primer OT1068	
<400> SEQUENCE: 42	
caaaagctga gctccaccgc g	21
<210> SEQ ID NO 43 <211> LENGTH: 44 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Primer OT1067	
<400> SEQUENCE: 43	
gtttactagt ttatgtgtgt ttattcgaaa ctaagttett ggtg	44
<210> SEQ ID NO 44 <211> LENGTH: 9333 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: pLH467	
<400> SEQUENCE: 44	
tegegegttt eggtgatgae ggtgaaaace tetgacaeat geageteeeg gagaeggtea	60
cagettgtet gtaageggat geegggagea gacaageeeg teagggegeg teagegggtg	120
ttggcgggtg tcggggctgg cttaactatg cggcatcaga gcagattgta ctgagagtgc	180
accataccac agetttteaa tteaatteat eattttttt ttattetttt ttttgattte	240
ggtttetttg aaatttttt gatteggtaa teteegaaca gaaggaagaa egaaggaagg	300
agcacagact tagattggta tatatacgca tatgtagtgt tgaagaaaca tgaaattgcc	360
cagtattett aacceaactg cacagaacaa aaacetgeag gaaaegaaga taaateatgt	420
cgaaagctac atataaggaa cgtgctgcta ctcatcctag teetgttget gecaagetat	480
ttaatatcat gcacgaaaag caaacaaact tgtgtgcttc attggatgtt cgtaccacca	540
aggaattact ggagttagtt gaagcattag gtcccaaaat ttgtttacta aaaacacatg	600
tggatatett gaetgatttt teeatggagg geaeagttaa geegetaaag geattateeg	660
ccaagtacaa ttttttactc ttcgaagaca gaaaatttgc tgacattggt aatacagtca	720
aattgcagta ctctgcgggt gtatacagaa tagcagaatg ggcagacatt acgaatgcac	780

229

acggtgtggt	gggcccaggt	attgttagcg	gtttgaagca	ggcggcagaa	gaagtaacaa	840
aggaacctag	aggccttttg	atgttagcag	aattgtcatg	caagggctcc	ctatctactg	900
gagaatatac	taagggtact	gttgacattg	cgaagagcga	caaagatttt	gttatcggct	960
ttattgctca	aagagacatg	ggtggaagag	atgaaggtta	cgattggttg	attatgacac	1020
ccggtgtggg	tttagatgac	aagggagacg	cattgggtca	acagtataga	accgtggatg	1080
atgtggtctc	tacaggatct	gacattatta	ttgttggaag	aggactattt	gcaaagggaa	1140
gggatgctaa	ggtagagggt	gaacgttaca	gaaaagcagg	ctgggaagca	tatttgagaa	1200
gatgcggcca	gcaaaactaa	aaaactgtat	tataagtaaa	tgcatgtata	ctaaactcac	1260
aaattagagc	ttcaatttaa	ttatatcagt	tattacccta	tgcggtgtga	aataccgcac	1320
agatgcgtaa	ggagaaaata	ccgcatcagg	aaattgtaaa	cgttaatatt	ttgttaaaat	1380
tcgcgttaaa	tttttgttaa	atcagctcat	tttttaacca	ataggccgaa	atcggcaaaa	1440
tcccttataa	atcaaaagaa	tagaccgaga	tagggttgag	tgttgttcca	gtttggaaca	1500
agagtccact	attaaagaac	gtggactcca	acgtcaaagg	gcgaaaaacc	gtctatcagg	1560
gcgatggccc	actacgtgaa	ccatcaccct	aatcaagttt	tttggggtcg	aggtgccgta	1620
aagcactaaa	tcggaaccct	aaagggagcc	cccgatttag	agcttgacgg	ggaaagccgg	1680
cgaacgtggc	gagaaaggaa	gggaagaaag	cgaaaggagc	gggcgctagg	gcgctggcaa	1740
gtgtagcggt	cacgctgcgc	gtaaccacca	cacccgccgc	gcttaatgcg	ccgctacagg	1800
gcgcgtcgcg	ccattcgcca	ttcaggctgc	gcaactgttg	ggaagggcga	tcggtgcggg	1860
cctcttcgct	attacgccag	ctggcgaaag	ggggatgtgc	tgcaaggcga	ttaagttggg	1920
taacgccagg	gttttcccag	tcacgacgtt	gtaaaacgac	ggccagtgag	cgcgcgtaat	1980
acgactcact	atagggcgaa	ttgggtaccg	ggccccccct	cgaggtcgac	tggccattaa	2040
tctttcccat	attagatttc	gccaagccat	gaaagttcaa	gaaaggtctt	tagacgaatt	2100
accetteatt	tctcaaactg	gcgtcaaggg	atcctggtat	ggttttatcg	ttttatttct	2160
ggttcttata	gcatcgtttt	ggacttctct	gttcccatta	ggcggttcag	gagccagcgc	2220
agaatcattc	tttgaaggat	acttatcctt	tccaattttg	attgtctgtt	acgttggaca	2280
taaactgtat	actagaaatt	ggactttgat	ggtgaaacta	gaagatatgg	atcttgatac	2340
cggcagaaaa	caagtagatt	tgactcttcg	tagggaagaa	atgaggattg	agcgagaaac	2400
attagcaaaa	agatccttcg	taacaagatt	tttacatttc	tggtgttgaa	gggaaagata	2460
tgagctatac	agcggaattt	ccatatcact	cagattttgt	tatctaattt	tttccttccc	2520
acgtccgcgg	gaatctgtgt	atattactgc	atctagatat	atgttatctt	atcttggcgc	2580
gtacatttaa	ttttcaacgt	attctataag	aaattgcggg	agttttttc	atgtagatga	2640
tactgactgc	acgcaaatat	aggcatgatt	tataggcatg	atttgatggc	tgtaccgata	2700
ggaacgctaa	gagtaacttc	agaatcgtta	tcctggcgga	aaaaattcat	ttgtaaactt	2760
taaaaaaaaa	agccaatatc	cccaaaatta	ttaagagcgc	ctccattatt	aactaaaatt	2820
tcactcagca	tccacaatgt	atcaggtatc	tactacagat	attacatgtg	gcgaaaaaga	2880
caagaacaat	gcaatagcgc	atcaagaaaa	aacacaaagc	tttcaatcaa	tgaatcgaaa	2940
atgtcattaa	aatagtatat	aaattgaaac	taagtcataa	agctataaaa	agaaaattta	3000
tttaaatgca	agatttaaag	taaattcacg	gccctgcagg	cctcagctct	tgttttgttc	3060
tgcaaataac	ttacccatct	ttttcaaaac	tttaggtgca	ccctcctttg	ctagaataag	3120

231

#### -continued

ttctatccaa	tacatcctat	ttggatctgc	ttgagcttct	ttcatcacgg	atacgaattc	3180
attttctgtt	ctcacaattt	tggacacaac	tctgtcttcc	gttgccccga	aactttctgg	3240
cagttttgag	taattccaca	taggaatgtc	attataactc	tggttcggac	catgaatttc	3300
cctctcaacc	gtgtaaccat	cgttattaat	gataaagcag	attgggttta	tettetetet	3360
aatggctagt	cctaattctt	ggacagtcag	ttgcaatgat	ccatctccga	taaacaataa	3420
atgtctagat	tctttatctg	caatttggct	gcctagagct	gcggggaaag	tgtatcctat	3480
agatccccac	aagggttgac	caataaaatg	tgatttcgat	ttcagaaata	tagatgaggc	3540
accgaagaaa	gaagtgcctt	gttcagccac	gatcgtctca	ttactttggg	tcaaattttc	3600
gacagettge	cacagtctat	cttgtgacaa	cagcgcgtta	gaaggtacaa	aatcttcttg	3660
ctttttatct	atgtacttgc	ctttatattc	aatttcggac	aagtcaagaa	gagatgatat	3720
cagggattcg	aagtcgaaat	tttggattct	ttcgttgaaa	attttacctt	catcgatatt	3780
caaggaaatc	attttattt	cattaagatg	gtgagtaaat	gcacccgtac	tagaatcggt	3840
aagctttaca	cccaacataa	gaataaaatc	agcagattcc	acaaattcct	tcaagtttgg	3900
ctctgacaga	gtaccgttgt	aaatccccaa	aaatgagggc	aatgcttcat	caacagatga	3960
tttaccaaag	ttcaaagtag	taataggtaa	cttagtcttt	gaaataaact	gagtaacagt	4020
cttctctagg	ccgaacgata	taatttcatg	gcctgtgatt	acaattggtt	tcttggcatt	4080
cttcagactt	tcctgtattt	tgttcagaat	ctcttgatca	gatgtattcg	acgtggaatt	4140
ttccttctta	agaggcaagg	atggtttttc	agccttagcg	gcagctacat	ctacaggtaa	4200
attgatgtaa	accggctttc	tttcctttag	taaggcagac	aacactctat	caatttcaac	4260
agttgcattc	tcggctgtca	ataaagtcct	ggcagcagta	accggttcgt	gcatcttcat	4320
aaagtgcttg	aaatcaccat	cagccaacgt	atggtgaaca	aacttacctt	cgttctgcac	4380
tttcgaggta	ggagatccca	cgatctcaac	aacaggcagg	ttctcagcat	aggagcccgc	4440
taagccatta	actgcggata	attcgccaac	accaaatgta	gtcaagaatg	ccgcagcctt	4500
tttcgttctt	gcgtacccgt	cggccatata	ggaggcattt	aactcattag	catttcccac	4560
ccatttcata	tctttgtgtg	aaataatttg	atctagaaat	tgcaaattgt	agtcacctgg	4620
tactccgaat	atttcttcta	tacctaattc	gtgtaatctg	tccaacagat	agtcacctac	4680
tgtatacatt	ttgtttacta	gtttatgtgt	gtttattcga	aactaagttc	ttggtgtttt	4740
aaaactaaaa	aaaagactaa	ctataaaagt	agaatttaag	aagtttaaga	aatagattta	4800
cagaattaca	atcaatacct	accgtcttta	tatacttatt	agtcaagtag	gggaataatt	4860
tcagggaact	ggtttcaacc	tttttttca	gctttttcca	aatcagagag	agcagaaggt	4920
aatagaaggt	gtaagaaaat	gagatagata	catgcgtggg	tcaattgcct	tgtgtcatca	4980
tttactccag	gcaggttgca	tcactccatt	gaggttgtgc	ccgtttttg	cctgtttgtg	5040
cccctgttct	ctgtagttgc	gctaagagaa	tggacctatg	aactgatggt	tggtgaagaa	5100
aacaatattt	tggtgctggg	attcttttt	tttctggatg	ccagcttaaa	aagcgggctc	5160
cattatattt	agtggatgcc	aggaataaac	tgttcaccca	gacacctacg	atgttatata	5220
ttctgtgtaa	cccgccccct	attttgggca	tgtacgggtt	acagcagaat	taaaaggcta	5280
attttttgac	taaataaagt	taggaaaatc	actactatta	attatttacg	tattctttga	5340
aatggcagta	ttgataatga	taaactcgaa	ctgaaaaagc	gtgttttta	ttcaaaatga	5400
ttctaactcc	cttacgtaat	caaggaatct	ttttgccttq	geeteegeqt	cattaaactt	5460
cttgttqttq	acgctaacat	tcaacqctaq	tatatattcq	tttttttcaq	gtaagttett	5520
	-	5 5	5	5		

233

ttcaacgggt	cttactgatg	aggcagtcgc	gtctgaacct	gttaagaggt	caaatatgtc	5580
ttcttgaccg	tacgtgtctt	gcatgttatt	agctttggga	atttgcatca	agtcatagga	5640
aaatttaaat	cttggctctc	ttgggctcaa	ggtgacaagg	tcctcgaaaa	tagggcgcgc	5700
cccaccgcgg	tggagctcca	gcttttgttc	cctttagtga	gggttaattg	cgcgcttggc	5760
gtaatcatgg	tcatagctgt	ttcctgtgtg	aaattgttat	ccgctcacaa	ttccacacaa	5820
cataggagcc	ggaagcataa	agtgtaaagc	ctggggtgcc	taatgagtga	ggtaactcac	5880
attaattgcg	ttgcgctcac	tgcccgcttt	ccagtcggga	aacctgtcgt	gccagctgca	5940
ttaatgaatc	ggccaacgcg	cggggagagg	cggtttgcgt	attgggcgct	cttccgcttc	6000
ctcgctcact	gactcgctgc	gctcggtcgt	tcggctgcgg	cgagcggtat	cageteacte	6060
aaaggcggta	atacggttat	ccacagaatc	aggggataac	gcaggaaaga	acatgtgagc	6120
aaaaggccag	caaaaggcca	ggaaccgtaa	aaaggccgcg	ttgctggcgt	ttttccatag	6180
gctccgcccc	cctgacgagc	atcacaaaaa	tcgacgctca	agtcagaggt	ggcgaaaccc	6240
gacaggacta	taaagatacc	aggcgtttcc	ccctggaagc	tccctcgtgc	gctctcctgt	6300
tccgaccctg	ccgcttaccg	gatacctgtc	cgcctttctc	ccttcgggaa	gcgtggcgct	6360
ttctcatagc	tcacgctgta	ggtatctcag	ttcggtgtag	gtcgttcgct	ccaagctggg	6420
ctgtgtgcac	gaaccccccg	ttcagcccga	ccgctgcgcc	ttatccggta	actatcgtct	6480
tgagtccaac	ccggtaagac	acgacttatc	gccactggca	gcagccactg	gtaacaggat	6540
tagcagagcg	aggtatgtag	gcggtgctac	agagttcttg	aagtggtggc	ctaactacgg	6600
ctacactaga	aggacagtat	ttggtatctg	cgctctgctg	aagccagtta	ccttcggaaa	6660
aagagttggt	agctcttgat	ccggcaaaca	aaccaccgct	ggtagcggtg	gttttttgt	6720
ttgcaagcag	cagattacgc	gcagaaaaaa	aggatctcaa	gaagatcctt	tgatcttttc	6780
tacggggtct	gacgctcagt	ggaacgaaaa	ctcacgttaa	gggattttgg	tcatgagatt	6840
atcaaaaagg	atcttcacct	agatcctttt	aaattaaaaa	tgaagtttta	aatcaatcta	6900
aagtatatat	gagtaaactt	ggtctgacag	ttaccaatgc	ttaatcagtg	aggcacctat	6960
ctcagcgatc	tgtctatttc	gttcatccat	agttgcctga	ctccccgtcg	tgtagataac	7020
tacgatacgg	gagggcttac	catctggccc	cagtgctgca	atgataccgc	gagacccacg	7080
ctcaccggct	ccagatttat	cagcaataaa	ccagccagcc	ggaagggccg	agcgcagaag	7140
tggtcctgca	actttatccg	cctccatcca	gtctattaat	tgttgccggg	aagctagagt	7200
aagtagttcg	ccagttaata	gtttgcgcaa	cgttgttgcc	attgctacag	gcatcgtggt	7260
gtcacgctcg	tcgtttggta	tggcttcatt	cageteeggt	tcccaacgat	caaggcgagt	7320
tacatgatcc	cccatgttgt	gcaaaaaagc	ggttagctcc	tteggteete	cgatcgttgt	7380
cagaagtaag	ttggccgcag	tgttatcact	catggttatg	gcagcactgc	ataattctct	7440
tactgtcatg	ccatccgtaa	gatgcttttc	tgtgactggt	gagtactcaa	ccaagtcatt	7500
ctgagaatag	tgtatgcggc	gaccgagttg	ctcttgcccg	gcgtcaatac	gggataatac	7560
cgcgccacat	agcagaactt	taaaagtgct	catcattgga	aaacgttctt	cggggcgaaa	7620
actctcaagg	atcttaccgc	tgttgagatc	cagttcgatg	taacccactc	gtgcacccaa	7680
ctgatcttca	gcatctttta	ctttcaccag	cgtttctggg	tgagcaaaaa	caggaaggca	7740
aaatgccgca	aaaaagggaa	taagggcgac	acggaaatgt	tgaatactca	tactcttcct	7800
ttttcaatat	tattgaagca	tttatcaggg	ttattgtctc	atgagcggat	acatatttga	7860

-continued

235

atgtatttag aaaaataaac aaataggggt teegegeaca ttteeeegaa aagtgeeace

7920

600

236

tgaacgaagc atctgtgctt cattttgtag aacaaaaatg caacgcgaga gcgctaattt 7980 8040 tttaccaacg aagaatctgt gcttcatttt tgtaaaacaa aaatgcaacg cgagagcgct 8100 aatttttcaa acaaagaatc tgagctgcat ttttacagaa cagaaatgca acgcgagagc 8160 gctattttac caacaaagaa tctatacttc ttttttgttc tacaaaaatg catcccgaga 8220 gcgctatttt tctaacaaag catcttagat tacttttttt ctcctttgtg cgctctataa 8280 tgcagtetet tgataaettt ttgeaetgta ggteegttaa ggttagaaga aggetaettt 8340 ggtgtctatt ttctcttcca taaaaaaagc ctgactccac ttcccgcgtt tactgattac 8400 tagcgaaget gegggtgeat ttttteaaga taaaggeate eeegattata ttetataeeg 8460 atgtggattg cgcatacttt gtgaacagaa agtgatagcg ttgatgattc ttcattggtc 8520 agaaaattat gaacggtttc ttctattttg tctctatata ctacgtatag gaaatgttta 8580 cattttcgta ttgttttcga ttcactctat gaatagttct tactacaatt tttttgtcta 8640 aagagtaata ctagagataa acataaaaaa tgtagaggtc gagtttagat gcaagttcaa 8700 ggagcgaaag gtggatgggt aggttatata gggatatagc acagagatat atagcaaaga 8760 gatacttttg agcaatgttt gtggaagcgg tattcgcaat attttagtag ctcgttacag 8820 tccggtgcgt ttttggtttt ttgaaagtgc gtcttcagag cgcttttggt tttcaaaagc 8880 gctctgaagt tcctatactt tctagagaat aggaacttcg gaataggaac ttcaaagcgt 8940 ttccgaaaac gagcgcttcc gaaaatgcaa cgcgagctgc gcacatacag ctcactgttc 9000 acgtcgcacc tatatctgcg tgttgcctgt atatatat acatgagaag aacggcatag 9060 tgcgtgttta tgcttaaatg cgtacttata tgcgtctatt tatgtaggat gaaaggtagt 9120 ctagtacete etgtgatatt ateceattee atgeggggta tegtatgett eetteageae 9180 taccetttag etgttetata tgetgeeaet eeteaattgg attagtetea teetteaatg 9240 ctatcatttc ctttgatatt ggatcatact aagaaaccat tattatcatg acattaacct 9300 ataaaaatag gcgtatcacg aggccctttc gtc 9333 <210> SEQ ID NO 45 <211> LENGTH: 8994 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Synthetic construct pRS425::GPM-sadB <400> SEQUENCE: 45 ctagttetag ageggeegee accgeggtgg ageteeaget tttgtteeet ttagtgaggg 60 ttaattgege gettggegta atcatggtea tagetgttte etgtgtgaaa ttgttateeg 120 ctcacaattc cacaacaat aggagccqga agcataaagt gtaaagcctg gggtgcctaa 180 tgagtgaggt aactcacatt aattgcgttg cgctcactgc ccgctttcca gtcgggaaac 240 ctgtcgtgcc agctgcatta atgaatcggc caacgcgcgg ggagaggcgg tttgcgtatt 300 gggcgctctt ccgcttcctc gctcactgac tcgctgcgct cggtcgttcg gctgcggcga 360 gcggtatcag ctcactcaaa ggcggtaata cggttatcca cagaatcagg ggataacgca 420 ggaaagaaca tgtgagcaaa aggccagcaa aaggccagga accgtaaaaa ggccgcgttg 480 ctggcgtttt tccataggct ccgcccccct gacgagcatc acaaaaatcg acgctcaagt 540

cagaggtggc gaaacccgac aggactataa agataccagg cgtttccccc tggaagctcc

237

ctcgtgcgct	ctcctgttcc	gaccctgccg	cttaccggat	acctgtccgc	ctttctccct	660
tcgggaagcg	tggcgctttc	tcatagctca	cgctgtaggt	atctcagttc	ggtgtaggtc	720
gttcgctcca	agctgggctg	tgtgcacgaa	ccccccgttc	agcccgaccg	ctgcgcctta	780
tccggtaact	atcgtcttga	gtccaacccg	gtaagacacg	acttatcgcc	actggcagca	840
gccactggta	acaggattag	cagagcgagg	tatgtaggcg	gtgctacaga	gttcttgaag	900
tggtggccta	actacggcta	cactagaagg	acagtatttg	gtatctgcgc	tctgctgaag	960
ccagttacct	tcggaaaaag	agttggtagc	tcttgatccg	gcaaacaaac	caccgctggt	1020
agcggtggtt	ttttgtttg	caagcagcag	attacgcgca	gaaaaaaagg	atctcaagaa	1080
gateetttga	tcttttctac	ggggtctgac	gctcagtgga	acgaaaactc	acgttaaggg	1140
attttggtca	tgagattatc	aaaaaggatc	ttcacctaga	tccttttaaa	ttaaaaatga	1200
agttttaaat	caatctaaag	tatatatgag	taaacttggt	ctgacagtta	ccaatgctta	1260
atcagtgagg	cacctatctc	agcgatctgt	ctatttcgtt	catccatagt	tgcctgactc	1320
cccgtcgtgt	agataactac	gatacgggag	ggcttaccat	ctggccccag	tgctgcaatg	1380
ataccgcgag	acccacgctc	accggctcca	gatttatcag	caataaacca	gccagccgga	1440
agggccgagc	gcagaagtgg	tcctgcaact	ttatccgcct	ccatccagtc	tattaattgt	1500
tgccgggaag	ctagagtaag	tagttcgcca	gttaatagtt	tgcgcaacgt	tgttgccatt	1560
gctacaggca	tcgtggtgtc	acgctcgtcg	tttggtatgg	cttcattcag	ctccggttcc	1620
caacgatcaa	ggcgagttac	atgatccccc	atgttgtgca	aaaaagcggt	tagctccttc	1680
ggtcctccga	tcgttgtcag	aagtaagttg	gccgcagtgt	tatcactcat	ggttatggca	1740
gcactgcata	attctcttac	tgtcatgcca	tccgtaagat	gcttttctgt	gactggtgag	1800
tactcaacca	agtcattctg	agaatagtgt	atgcggcgac	cgagttgctc	ttgcccggcg	1860
tcaatacggg	ataataccgc	gccacatagc	agaactttaa	aagtgctcat	cattggaaaa	1920
cgttcttcgg	ggcgaaaact	ctcaaggatc	ttaccgctgt	tgagatccag	ttcgatgtaa	1980
cccactcgtg	cacccaactg	atcttcagca	tcttttactt	tcaccagcgt	ttctgggtga	2040
gcaaaaacag	gaaggcaaaa	tgccgcaaaa	aagggaataa	gggcgacacg	gaaatgttga	2100
atactcatac	tcttcctttt	tcaatattat	tgaagcattt	atcagggtta	ttgtctcatg	2160
agcggataca	tatttgaatg	tatttagaaa	aataaacaaa	taggggttcc	gcgcacattt	2220
ccccgaaaag	tgccacctga	acgaagcatc	tgtgcttcat	tttgtagaac	aaaaatgcaa	2280
cgcgagagcg	ctaatttttc	aaacaaagaa	tctgagctgc	atttttacag	aacagaaatg	2340
caacgcgaaa	gcgctatttt	accaacgaag	aatctgtgct	tcatttttgt	aaaacaaaaa	2400
tgcaacgcga	gagcgctaat	ttttcaaaca	aagaatctga	gctgcatttt	tacagaacag	2460
aaatgcaacg	cgagagcgct	attttaccaa	caaagaatct	atacttcttt	tttgttctac	2520
aaaaatgcat	cccgagagcg	ctatttttct	aacaaagcat	cttagattac	ttttttctc	2580
ctttgtgcgc	tctataatgc	agtctcttga	taactttttg	cactgtaggt	ccgttaaggt	2640
tagaagaagg	ctactttggt	gtctattttc	tcttccataa	aaaaagcctg	actccacttc	2700
ccgcgtttac	tgattactag	cgaagctgcg	ggtgcatttt	ttcaagataa	aggcatecce	2760
gattatattc	tataccgatg	tggattgcgc	atactttgtg	aacagaaagt	gatagcgttg	2820
atgattcttc	attggtcaga	aaattatgaa	cggtttcttc	tattttgtct	ctatatacta	2880
cgtataggaa	atgtttacat	tttcgtattg	ttttcgattc	actctatgaa	tagttettae	2940

239

### -continued

tacaatttt	ttgtctaaag	agtaatacta	gagataaaca	taaaaaatgt	agaggtcgag	3000
tttagatgca	agttcaagga	gcgaaaggtg	gatgggtagg	ttatataggg	atatagcaca	3060
gagatatata	gcaaagagat	acttttgagc	aatgtttgtg	gaagcggtat	tcgcaatatt	3120
ttagtagctc	gttacagtcc	ggtgcgtttt	tggtttttg	aaagtgcgtc	ttcagagcgc	3180
ttttggtttt	caaaagcgct	ctgaagttcc	tatactttct	agagaatagg	aacttcggaa	3240
taggaacttc	aaagcgtttc	cgaaaacgag	cgcttccgaa	aatgcaacgc	gagetgegea	3300
catacagete	actgttcacg	tcgcacctat	atctgcgtgt	tgcctgtata	tatatataca	3360
tgagaagaac	ggcatagtgc	gtgtttatgc	ttaaatgcgt	acttatatgc	gtctatttat	3420
gtaggatgaa	aggtagtcta	gtacctcctg	tgatattatc	ccattccatg	cggggtatcg	3480
tatgcttcct	tcagcactac	cctttagctg	ttctatatgc	tgccactcct	caattggatt	3540
agtctcatcc	ttcaatgcta	tcatttcctt	tgatattgga	tcatactaag	aaaccattat	3600
tatcatgaca	ttaacctata	aaaataggcg	tatcacgagg	ccctttcgtc	tcgcgcgttt	3660
cggtgatgac	ggtgaaaacc	tctgacacat	gcagctcccg	gagacggtca	cagettgtet	3720
gtaagcggat	gccgggagca	gacaagcccg	tcagggcgcg	tcagcgggtg	ttggcgggtg	3780
tcggggctgg	cttaactatg	cggcatcaga	gcagattgta	ctgagagtgc	accatatcga	3840
ctacgtcgta	aggccgtttc	tgacagagta	aaattcttga	gggaactttc	accattatgg	3900
gaaatgcttc	aagaaggtat	tgacttaaac	tccatcaaat	ggtcaggtca	ttgagtgttt	3960
tttatttgtt	gtatttttt	tttttagag	aaaatcctcc	aatatcaaat	taggaatcgt	4020
agtttcatga	ttttctgtta	cacctaactt	tttgtgtggt	gccctcctcc	ttgtcaatat	4080
taatgttaaa	gtgcaattct	ttttccttat	cacgttgagc	cattagtatc	aatttgctta	4140
cctgtattcc	tttactatcc	tcctttttct	ccttcttgat	aaatgtatgt	agattgcgta	4200
tatagtttcg	tctaccctat	gaacatattc	cattttgtaa	tttcgtgtcg	tttctattat	4260
gaatttcatt	tataaagttt	atgtacaaat	atcataaaaa	aagagaatct	ttttaagcaa	4320
ggattttctt	aacttcttcg	gcgacagcat	caccgacttc	ggtggtactg	ttggaaccac	4380
ctaaatcacc	agttctgata	cctgcatcca	aaaccttttt	aactgcatct	tcaatggcct	4440
taccttcttc	aggcaagttc	aatgacaatt	tcaacatcat	tgcagcagac	aagatagtgg	4500
cgatagggtc	aaccttattc	tttggcaaat	ctggagcaga	accgtggcat	ggttcgtaca	4560
aaccaaatgc	ggtgttcttg	tctggcaaag	aggccaagga	cgcagatggc	aacaaaccca	4620
aggaacctgg	gataacggag	gcttcatcgg	agatgatatc	accaaacatg	ttgctggtga	4680
ttataatacc	atttaggtgg	gttgggttct	taactaggat	catggcggca	gaatcaatca	4740
attgatgttg	aaccttcaat	gtagggaatt	cgttcttgat	ggtttcctcc	acagtttttc	4800
tccataatct	tgaagaggcc	aaaagattag	ctttatccaa	ggaccaaata	ggcaatggtg	4860
gctcatgttg	tagggccatg	aaagcggcca	ttcttgtgat	tctttgcact	tctggaacgg	4920
tgtattgttc	actatcccaa	gcgacaccat	caccatcgtc	ttcctttctc	ttaccaaagt	4980
aaatacctcc	cactaattct	ctgacaacaa	cgaagtcagt	acctttagca	aattgtggct	5040
tgattggaga	taagtctaaa	agagagtcgg	atgcaaagtt	acatggtctt	aagttggcgt	5100
acaattgaag	ttctttacgg	atttttagta	aaccttgttc	aggtctaaca	ctaccggtac	5160
cccatttagg	accagccaca	gcacctaaca	aaacggcatc	aaccttcttg	gaggetteca	5220
gcgcctcatc	tggaagtggg	acacctgtag	catcgatagc	agcaccacca	attaaatgat	5280
tttcgaaatc	gaacttqaca	ttggaacqaa	catcaqaaat	agetttaaqa	accttaatqq	5340
5	5 5		5	5		

241

cttcggctgt	gatttcttga	ccaacgtggt	cacctggcaa	aacgacgatc	ttettagggg	5400
cagacatagg	ggcagacatt	agaatggtat	atccttgaaa	tatatatata	tattgctgaa	5460
atgtaaaagg	taagaaaagt	tagaaagtaa	gacgattgct	aaccacctat	tggaaaaaac	5520
aataggtcct	taaataatat	tgtcaacttc	aagtattgtg	atgcaagcat	ttagtcatga	5580
acgcttctct	attctatatg	aaaagccggt	tccggcctct	cacctttcct	ttttctccca	5640
atttttcagt	tgaaaaaggt	atatgcgtca	ggcgacctct	gaaattaaca	aaaaatttcc	5700
agtcatcgaa	tttgattctg	tgcgatagcg	cccctgtgtg	ttctcgttat	gttgaggaaa	5760
aaaataatgg	ttgctaagag	attcgaactc	ttgcatctta	cgatacctga	gtattcccac	5820
agttaactgc	ggtcaagata	tttcttgaat	caggcgcctt	agaccgctcg	gccaaacaac	5880
caattacttg	ttgagaaata	gagtataatt	atcctataaa	tataacgttt	ttgaacacac	5940
atgaacaagg	aagtacagga	caattgattt	tgaagagaat	gtggattttg	atgtaattgt	6000
tgggattcca	tttttaataa	ggcaataata	ttaggtatgt	ggatatacta	gaagttetee	6060
tcgaccgtcg	atatgcggtg	tgaaataccg	cacagatgcg	taaggagaaa	ataccgcatc	6120
aggaaattgt	aaacgttaat	attttgttaa	aattcgcgtt	aaatttttgt	taaatcagct	6180
catttttaa	ccaataggcc	gaaatcggca	aaatccctta	taaatcaaaa	gaatagaccg	6240
agatagggtt	gagtgttgtt	ccagtttgga	acaagagtcc	actattaaag	aacgtggact	6300
ccaacgtcaa	agggcgaaaa	accgtctatc	agggcgatgg	cccactacgt	gaaccatcac	6360
cctaatcaag	tttttggggg	tcgaggtgcc	gtaaagcact	aaatcggaac	cctaaaggga	6420
gcccccgatt	tagagettga	cggggaaagc	cggcgaacgt	ggcgagaaag	gaagggaaga	6480
aagcgaaagg	agcgggcgct	agggcgctgg	caagtgtagc	ggtcacgctg	cgcgtaacca	6540
ccacacccgc	cgcgcttaat	gcgccgctac	agggcgcgtc	gcgccattcg	ccattcaggc	6600
tgcgcaactg	ttgggaaggg	cgatcggtgc	gggcetette	gctattacgc	cagctggcga	6660
aaggggggatg	tgctgcaagg	cgattaagtt	gggtaacgcc	agggttttcc	cagtcacgac	6720
gttgtaaaac	gacggccagt	gagcgcgcgt	aatacgactc	actatagggc	gaattgggta	6780
ccgggccccc	cctcgaggtc	gacggtatcg	ataagcttga	tatcgaattc	ctgcagcccg	6840
ggggatccgc	atgcttgcat	ttagtcgtgc	aatgtatgac	tttaagattt	gtgagcagga	6900
agaaaaggga	gaatcttcta	acgataaacc	cttgaaaaac	tgggtagact	acgctatgtt	6960
gagttgctac	gcaggctgca	caattacacg	agaatgctcc	cgcctaggat	ttaaggctaa	7020
gggacgtgca	atgcagacga	cagatctaaa	tgaccgtgtc	ggtgaagtgt	tcgccaaact	7080
tttcggttaa	cacatgcagt	gatgcacgcg	cgatggtgct	aagttacata	tatatatata	7140
tatatatata	tagccatagt	gatgtctaag	taacctttat	ggtatatttc	ttaatgtgga	7200
aagatactag	cgcgcgcacc	cacacacaag	cttcgtcttt	tcttgaagaa	aagaggaagc	7260
tcgctaaatg	ggattccact	ttccgttccc	tgccagctga	tggaaaaagg	ttagtggaac	7320
gatgaagaat	aaaaagagag	atccactgag	gtgaaatttc	agctgacagc	gagtttcatg	7380
atcgtgatga	acaatggtaa	cgagttgtgg	ctgttgccag	ggagggtggt	tctcaacttt	7440
taatgtatgg	ccaaatcgct	acttgggttt	gttatataac	aaagaagaaa	taatgaactg	7500
attctcttcc	tccttcttgt	cctttcttaa	ttctgttgta	attaccttcc	tttgtaattt	7560
tttttgtaat	tattcttctt	aataatccaa	acaaacacac	atattacaat	agctagctga	7620
ggatgaaggc	attagtttat	catggggatc	acaaaatttc	gttagaagac	aaaccaaaac	7680

243

#### -continued

ccactctgca gaaaccaaca gacgttgtgg ttagggtgtt gaaaacaaca atttgcggta	7740						
ctgacttggg aatatacaaa ggtaagaatc ctgaagtggc agatggcaga atcctgggtc	7800						
atgagggggt tggcgtcatt gaagaagtgg gcgaatccgt gacacaattc aaaaaggggg	7860						
ataaagtttt aatctcctgc gttactagct gtggatcgtg tgattattgc aagaagcaac	7920						
tgtattcaca ctgtagagac ggtggctgga ttttaggtta catgatcgac ggtgtccaag	7980						
ccgaatacgt cagaatacca catgctgaca attcattgta taagatcccg caaactatcg	8040						
atgatgaaat tgcagtacta ctgtccgata ttttacctac tggacatgaa attggtgttc	8100						
aatatggtaa cgttcaacca ggcgatgctg tagcaattgt aggagcaggt cctgttggaa	8160						
tgtcagtttt gttaactgct caattttact cgcctagtac cattattgtt atcgacatgg	8220						
acgaaaaccg tttacaatta gcgaaggagc ttggggccac acacactatt aactccggta	8280						
ctgaaaatgt tgtcgaagct gtgcatcgta tagcagccga aggagtggat gtagcaatag	8340						
aagctgttgg tatacccgca acctgggaca tctgtcagga aattgtaaaa cccggcgctc	8400						
atattgccaa cgtgggagtt catggtgtta aggtggactt tgaaattcaa aagttgtgga	8460						
ttaagaatct aaccatcacc actggtttgg ttaacactaa tactacccca atgttgatga	8520						
aggtagcete taetgataaa ttgeetttaa agaaaatgat taeteacagg tttgagttag	8580						
ctgaaatcga acacgcatat caggttttct tgaatggcgc taaagaaaaa gctatgaaga	8640						
ttattctatc taatgcaggt gccgcctaat taattaagag taagcgaatt tcttatgatt	8700						
tatgattttt attattaaat aagttataaa aaaaataagt gtatacaaat tttaaagtga	8760						
ctcttaggtt ttaaaacgaa aattcttatt cttgagtaac tctttcctgt aggtcaggtt	8820						
gettteteag gtatageatg aggtegetet tattgaceae acetetaeeg geatgeegag	8880						
caaatgeetg caaategete eccattteae ceaattgtag atatgetaae tecageaatg	8940						
agttgatgaa teteggtgtg tattttatgt eeteagagga caacaeetgt ggta	8994						
<210> SEQ ID NO 46 <211> LENGTH: 753 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Saccharomyces cerevisiae GPM1 promoter							
<400> SEQUENCE: 46							
gcatgcttgc atttagtcgt gcaatgtatg actttaagat ttgtgagcag gaagaaaagg	60						
gagaatette taacgataaa eeettgaaaa aetgggtaga etaegetatg ttgagttget	120						
acgcaggctg cacaattaca cgagaatgct cccgcctagg atttaaggct aagggacgtg	180						
caatgcagac gacagatcta aatgaccgtg tcggtgaagt gttcgccaaa cttttcggtt	240						
aacacatgca gtgatgcacg cgcgatggtg ctaagttaca tatatatat tatagccata	300						
gtgatgtcta agtaaccttt atggtatatt tcttaatgtg gaaagatact agcgcgcgca	360						
cccacacaca agcttcgtct tttcttgaag aaaagaggaa gctcgctaaa tgggattcca	420						
ctttccgttc cctgccagct gatggaaaaa ggttagtgga acgatgaaga ataaaaagag	480						
agatecaetg aggtgaaatt teagetgaea gegagtttea tgategtgat gaacaatggt	540						

600

660

720

753

244

gtgatgtcta agta cccacacaca agct ctttccgttc cctg agatccactg aggtg igctga gcgagt igategiga : ga -gg aacgagttgt ggctgttgcc agggagggtg gttctcaact tttaatgtat ggccaaatcg ctacttgggt ttgttatata acaaagaaga aataatgaac tgattctctt cctccttctt gtcctttctt aattctgttg taattacctt cctttgtaat tttttttgta attattcttc ttaataatcc aaacaaacac acatattaca ata
<210	)> SH	EQ II	) NO	47											
<212	L> L1 2> T3	CNG11 (PE :	DNA	14 /											
<213	3 > OF	RGANI	[SM:	Achi	romok	bacte	er xy	losc	xida	ins					
<400	)> SI	EQUEI	ICE :	47											
atga	aaago	ctc t	ggtt	tato	a co	gtga	accac	aag	jatct	cgc	ttga	agad	caa q	geeca	agccc
acco	ettea	aaa a	ageed	cacgo	ja to	gtagt	agta	ı cgo	gttt	tga	agad	caco	gat d	ctgcc	Igcacg
gato	ctcgg	gca t	ctad	caaaq	là ca	agaa	atcca	a gaç	gtco	leed	acgo	gcgo	cat o	cctgg	gccat
gaag	ggggt	ag g	gegto	categ	ja go	gaagt	gggg	gao	gagto	gtca	cgca	gtto	caa q	gaaag	Jgcgac
aago	gteet	:ga t	ttco	ctgcc	gt ca	actto	cttgo	, ddo	tcgt	gcg	acta	actgo	caa q	gaago	agett
tact	cccca	att ç	geege	gaco	là cò	gggtg	ggato	cto	ggtt	aca	tgat	cgat	gg (	gtgo	aggee
gaat	cacgt	ccc ç	gcato	cccgo	a to	geega	acaac	ago	ctct	aca	agat	cccc	cca ç	gacaa	attgac
gaco	gaaat	cg d	ccgto	ctgo	t ga	agcga	acato	cto	JCCCS	ıccg	gcca	acgaa	aat d	ggag	gtccag
tato	gggaa	atg t	ccaç	gaago	là cò	gatgo	ggtg	g gct	atto	Jtcg	gcgo	adad	ccc d	gtcg	gcatg
tccç	gtact	gt t	gaco	gcco	a gt	tcta	actco	: ccc	tcga	icca	tcat	cgto	gat d	cgaca	atggac
gaga	aatco	gee t	ccaç	getee	jc ca	aagga	ageto	a dda	gcaa	ıcgc	acad	cato	caa d	ctccc	Igcacg
gaga	aacgt	tg t	cgaa	ageee	gt go	cataç	ggatt	gee	gcaç	Jagg	gagt	cgat	gt t	gega	atcgag
gcgé	gtggg	gca t	acco	ggcga	ac tt	ggga	acato	tgc	cago	Jaga	tcgt	caaç	gee d	ggag	JCGCAC
atco	gccaa	acg t	cggo	gtgo	a to	ggcgt	caag	g gtt	gact	tcg	agat	tcaç	gaa q	gctct	ggatc
aaga	aacct	:ga d	gato	cacca	ic go	ggact	ggtg	j aac	acga	aca	cgad	gcco	cat q	gctga	itgaag
gtco	geeto	cga d	ccgad	caago	t to	cgtt	gaag	g aag	jatga	itta	ccca	atcgo	ett d	cgago	tggcc
gaga	gagategage aegeetatea ggtatteete aatggegeea aggagaagge gatgaagate														
atco	atcetetega aegeaggege tgeetga														
~ ~ ~															
<210	)> SH L> LH	SQ 11 ENGTH	) NO 1: 34	48 18											
<212	2> TY	PE:	PRT	7 aba	amak			.1							
<213		CGAN	ISM:	ACIII	. Olior	Jacte	эг ху	TORC	XIU	uis					
<400	)> SI	SÕOEL	ICE:	48			<b>G</b> ]				<b>T</b> ] -	<b>a</b>		<b>6</b> 1	
Met 1	ГЛЗ	Ala	Leu	Val 5	Tyr	Hls	GIY	Asp	H15 10	ГЛа	lle	Ser	Leu	GIU 15	Asp
Lys	Pro	Lys	Pro 20	Thr	Leu	Gln	Lys	Pro 25	Thr	Asp	Val	Val	Val 30	Arg	Val
Leu	Lys	Thr 35	Thr	Ile	Сүз	Gly	Thr 40	Asp	Leu	Gly	Ile	Tyr 45	Lys	Gly	Lys
Asn	Pro 50	Glu	Val	Ala	Aap	Gly 55	Arg	Ile	Leu	Gly	His 60	Glu	Gly	Val	Gly
Val 65	Ile	Glu	Glu	Val	Gly 70	Glu	Ser	Val	Thr	Gln 75	Phe	Lys	Lys	Gly	Asp 80
Lys	Val	Leu	Ile	Ser 85	Суз	Val	Thr	Ser	Cys 90	Gly	Ser	Cys	Asp	Tyr 95	Суз
Lys	Lys	Gln	Leu 100	Tyr	Ser	His	Cys	Arg 105	Asp	Gly	Gly	Trp	Ile 110	Leu	Gly
Tyr	Met	Ile 115	Asp	Gly	Val	Gln	Ala 120	Glu	Tyr	Val	Arg	Ile 125	Pro	His	Ala
Asp	Asn 130	Ser	Leu	Tyr	Гла	Ile 135	Pro	Gln	Thr	Ile	Asp 140	Asp	Glu	Ile	Ala

248

Val Leu Leu Ser Asp Ile Leu Pro Thr Gly His Glu Ile Gly Val 145 150 155	Gln 160									
Tyr Gly Asn Val Gln Pro Gly Asp Ala Val Ala Ile Val Gly Ala 165 170 175	Gly									
Pro Val Gly Met Ser Val Leu Leu Thr Ala Gln Phe Tyr Ser Pro 180 185 190	Ser									
Thr Ile Ile Val Ile Asp Met Asp Glu Asn Arg Leu Gln Leu Ala 195 200 205	Lys									
Glu Leu Gly Ala Thr His Thr Ile Asn Ser Gly Thr Glu Asn Val 210 215 220	Val									
Glu Ala Val His Arg Ile Ala Ala Glu Gly Val Asp Val Ala Ile 225 230 235	Glu 240									
Ala Val Gly Ile Pro Ala Thr Trp Asp Ile Cys Gln Glu Ile Val 245 250 255	Lys									
Pro Gly Ala His Ile Ala Asn Val Gly Val His Gly Val Lys Val 260 265 270	Asp									
Phe Glu Ile Gln Lys Leu Trp Ile Lys Asn Leu Thr Ile Thr Thr 275 280 285	Gly									
Leu Val Asn Thr Asn Thr Thr Pro Met Leu Met Lys Val Ala Ser 290 295 300	Thr									
Asp Lys Leu Pro Leu Lys Lys Met Ile Thr His Arg Phe Glu Leu 305 310 315	Ala 320									
Glu Ile Glu His Ala Tyr Gln Val Phe Leu Asn Gly Ala Lys Glu 325 330 335	Гла									
Ala Met Lys Ile Ile Leu Ser Asn Ala Gly Ala Ala 340 345										
<210> SEQ ID NO 49 <211> LENGTH: 316 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Saccharomyces cerevisiae ADH1 terminator										
<400> SEQUENCE: 49	aaata 60									
aqtqtataca aattttaaaq tqactcttaq qttttaaaac qaaaattctt attc!	tqaqt 120									
aactetttee tgtaggteag gttgetttet eaggtatage atgaggtege tetta	attgac 180									
cacaceteta coggeatgee gageaaatge etgeaaateg eteceeattt cace	caattg 240									
tagatatgct aactccagca atgagttgat gaatctcggt gtgtatttta tgtcc	ctcaga 300									
ggacaacacc tgtggt	316									
<210> SEQ ID NO 50 <211> LENGTH: 39 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Primer OT1074 <400> SEQUENCE: 50										
cacacatatt acaatagcta gctgaggatg aaagctctg	39									
<210> SEQ ID NO 51 <211> LENGTH: 39 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence										

<pre>&lt;220&gt; FEATURE: &lt;223&gt; OTHER INFORMATION: Primer OT1075</pre>	
<400> SEQUENCE: 51	
cagagettte ateetcaget agetattgta atatgtgtg	39
<210> SEQ ID NO 52 <211> LENGTH: 9075 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: pLH435	
<400> SEQUENCE: 52	
ctagttctag ageggeegee acegeggtgg ageteeaget tttgtteeet ttagtgagg	gg 60
ttaattgege gettggegta atcatggtea tagetgttte etgtgtgaaa ttgttate	cg 120
ctcacaattc cacacaacat aggagccgga agcataaagt gtaaagcctg gggtgccta	aa 180
tgagtgaggt aactcacatt aattgcgttg cgctcactgc ccgctttcca gtcgggaaa	ac 240
ctgtcgtgcc agctgcatta atgaatcggc caacgcgcgg ggagaggcgg tttgcgtat	tt 300
gggcgctctt ccgcttcctc gctcactgac tcgctgcgct cggtcgttcg gctgcggcg	ga 360
gcggtatcag ctcactcaaa ggcggtaata cggttatcca cagaatcagg ggataacgo	ca 420
ggaaagaaca tgtgagcaaa aggccagcaa aaggccagga accgtaaaaa ggccgcgti	tg 480
ctggcgtttt tccataggct ccgccccct gacgagcatc acaaaaatcg acgctcaa	gt 540
cagaggtggc gaaacccgac aggactataa agataccagg cgtttccccc tggaagct	cc 600
ctcgtgcgct ctcctgttcc gaccctgccg cttaccggat acctgtccgc ctttctccc	ct 660
togggaagog tggogottto toatagotoa ogotgtaggt atotoagtto ggtgtagg	tc 720
gttcgctcca agctgggctg tgtgcacgaa ccccccgttc agcccgaccg ctgcgcct	ta 780
teeggtaact ategtettga gteeaaceeg gtaagacaeg acttategee aetggeage	ca 840
gccactggta acaggattag cagagcgagg tatgtaggcg gtgctacaga gttcttgaa	ag 900
tggtggccta actacggcta cactagaagg acagtatttg gtatctgcgc tctgctgaa	ag 960
ccagttacct toggaaaaag agttggtago tottgatoog goaaacaaac cacogotg	gt 1020
agoggtggtt tttttgtttg caagcagcag attacgogca gaaaaaaagg atotcaaga	aa 1080
gateetttga tettttetae ggggtetgae geteagtgga aegaaaaete aegttaag	gg 1140
attttggtca tgagattatc aaaaaggatc ttcacctaga tccttttaaa ttaaaaat	ga 1200
agttttaaat caatctaaag tatatatgag taaacttggt ctgacagtta ccaatgcti	ta 1260
atcagtgagg cacctatete agegatetgt etatttegtt eatecatagt tgeetgaet	tc 1320
cccgtcgtgt agataactac gatacgggag ggcttaccat ctggccccag tgctgcaat	tg 1380
ataccgcgag acccacgctc accggctcca gatttatcag caataaacca gccagccg	ga 1440
agggccgagc gcagaagtgg tcctgcaact ttatccgcct ccatccagtc tattaatt	gt 1500
tgccgggaag ctagagtaag tagttcgcca gttaatagtt tgcgcaacgt tgttgcca	tt 1560
gctacaggca tegtggtgte acgetegteg tttggtatgg etteatteag etceggtte	cc 1620
caacgatcaa ggcgagttac atgatccccc atgttgtgca aaaaagcggt tagctcctt	tc 1680
agtecteda testtateas aastaastta secseastat tateasteat estateas	ca 1740
gradigeta atteinting tetesteese teestasse activities atteint	, ag 1800
geaciyeara arreteriae igreargeea teegtaagat gettteetgt gaetggtga	ay 1800
tactcaacca agtcattctg agaatagtgt atgcggcgac cgagttgctc ttgcccgg	cg 1860

251

### -continued

tcaatacggg ataataccgc gccacatagc agaactttaa aagtgctcat cattggaaaa 1920 cgttcttcgg ggcgaaaact ctcaaggatc ttaccgctgt tgagatccag ttcgatgtaa 1980 cccactcgtg cacccaactg atcttcagca tcttttactt tcaccagcgt ttctgggtga 2040 gcaaaaacag gaaggcaaaa tgccgcaaaa aagggaataa gggcgacacg gaaatgttga 2100 atactcatac tetteettt teaatattat tgaageattt ateagggtta ttgteteatg 2160 ageggataea tatttgaatg tatttagaaa aataaacaaa taggggttee gegeacattt 2220 ccccgaaaag tgccacctga acgaagcatc tgtgcttcat tttgtagaac aaaaatgcaa 2280 cgcgagagcg ctaatttttc aaacaaagaa tctgagctgc atttttacag aacagaaatg 2340 caacgcgaaa gcgctatttt accaacgaag aatctgtgct tcatttttgt aaaacaaaaa 2400 tgcaacgcga gagcgctaat ttttcaaaca aagaatctga gctgcatttt tacagaacag 2460 aaatgcaacg cgagagcgct attttaccaa caaagaatct atacttcttt tttgttctac 2520 aaaaatgcat cccgagagcg ctatttttct aacaaagcat cttagattac tttttttctc 2580 ctttgtgcgc tctataatgc agtctcttga taactttttg cactgtaggt ccgttaaggt 2640 tagaagaagg ctactttggt gtctattttc tcttccataa aaaaagcctg actccacttc 2700 ccgcgtttac tgattactag cgaagctgcg ggtgcatttt ttcaagataa aggcatcccc 2760 gattatattc tataccgatg tggattgcgc atactttgtg aacagaaagt gatagcgttg 2820 atgattette attggteaga aaattatgaa eggtttette tattttgtet etatataeta 2880 cgtataggaa atgtttacat tttcgtattg ttttcgattc actctatgaa tagttcttac 2940 tacaattttt ttgtctaaag agtaatacta gagataaaca taaaaaatgt agaggtcgag 3000 tttagatgca agttcaagga gcgaaaggtg gatgggtagg ttatataggg atatagcaca 3060 gagatatata gcaaagagat acttttgagc aatgtttgtg gaagcggtat tcgcaatatt 3120 ttagtagete gttacagtee ggtgegtttt tggttttttg aaagtgegte tteagagege 3180 ttttggtttt caaaagcgct ctgaagttcc tatactttct agagaatagg aacttcggaa 3240 taggaacttc aaagegtttc egaaaaegag egetteegaa aatgeaaege gagetgegea 3300 catacagete actgtteacg tegeacetat atetgegtgt tgeetgtata tatatataea 3360 tgagaagaac ggcatagtgc gtgtttatgc ttaaatgcgt acttatatgc gtctatttat 3420 gtaggatgaa aggtagteta gtaceteetg tgatattate ceatteeatg eggggtateg 3480 tatgetteet teageactae cetttagetg ttetatatge tgecaeteet caattggatt 3540 agteteatee tteaatgeta teattteett tgatattgga teataetaag aaaceattat 3600 tatcatgaca ttaacctata aaaataggcg tatcacgagg ccctttcgtc tcgcgcgttt 3660 cggtgatgac ggtgaaaaacc tctgacacat gcagctcccg gagacggtca cagcttgtct 3720 gtaageggat geegggagea gacaageeeg teagggegeg teagegggtg ttggegggtg 3780 tcqqqqctqq cttaactatq cqqcatcaqa qcaqattqta ctqaqaqtqc accatatcqa 3840 ctacgtcgta aggccgtttc tgacagagta aaattcttga gggaactttc accattatgg 3900 gaaatgette aagaaggtat tgaettaaac teeatcaaat ggteaggtea ttgagtgttt 3960 tttatttgtt gtatttttt ttttttagag aaaateetee aatateaaat taggaategt 4020 agtttcatga ttttctgtta cacctaactt tttgtgtggt gccctcctcc ttgtcaatat 4080 taatgttaaa gtgcaattet tttteettat caegttgage cattagtate aatttgetta 4140 cctgtattcc tttactatcc tcctttttct ccttcttgat aaatgtatgt agattgcgta 4200

tatagtttcg tctaccctat gaacatattc cattttgtaa tttcgtgtcg tttctattat

4260

253

gaatttcatt	tataaagttt	atgtacaaat	atcataaaaa	aagagaatct	ttttaagcaa	4320
ggattttctt	aacttcttcg	gcgacagcat	caccgacttc	ggtggtactg	ttggaaccac	4380
ctaaatcacc	agttctgata	cctgcatcca	aaaccttttt	aactgcatct	tcaatggcct	4440
taccttcttc	aggcaagttc	aatgacaatt	tcaacatcat	tgcagcagac	aagatagtgg	4500
cgatagggtc	aaccttattc	tttggcaaat	ctggagcaga	accgtggcat	ggttcgtaca	4560
aaccaaatgc	ggtgttcttg	tctggcaaag	aggccaagga	cgcagatggc	aacaaaccca	4620
aggaacctgg	gataacggag	gcttcatcgg	agatgatatc	accaaacatg	ttgctggtga	4680
ttataatacc	atttaggtgg	gttgggttct	taactaggat	catggcggca	gaatcaatca	4740
attgatgttg	aaccttcaat	gtagggaatt	cgttcttgat	ggtttcctcc	acagtttttc	4800
tccataatct	tgaagaggcc	aaaagattag	ctttatccaa	ggaccaaata	ggcaatggtg	4860
gctcatgttg	tagggccatg	aaagcggcca	ttcttgtgat	tctttgcact	tctggaacgg	4920
tgtattgttc	actatcccaa	gcgacaccat	caccatcgtc	ttcctttctc	ttaccaaagt	4980
aaatacctcc	cactaattct	ctgacaacaa	cgaagtcagt	acctttagca	aattgtggct	5040
tgattggaga	taagtctaaa	agagagtcgg	atgcaaagtt	acatggtctt	aagttggcgt	5100
acaattgaag	ttctttacgg	atttttagta	aaccttgttc	aggtctaaca	ctaccggtac	5160
cccatttagg	accagccaca	gcacctaaca	aaacggcatc	aaccttcttg	gaggetteca	5220
gcgcctcatc	tggaagtggg	acacctgtag	catcgatagc	agcaccacca	attaaatgat	5280
tttcgaaatc	gaacttgaca	ttggaacgaa	catcagaaat	agctttaaga	accttaatgg	5340
cttcggctgt	gatttcttga	ccaacgtggt	cacctggcaa	aacgacgatc	ttcttagggg	5400
cagacatagg	ggcagacatt	agaatggtat	atccttgaaa	tatatatata	tattgctgaa	5460
atgtaaaagg	taagaaaagt	tagaaagtaa	gacgattgct	aaccacctat	tggaaaaaac	5520
aataggtcct	taaataatat	tgtcaacttc	aagtattgtg	atgcaagcat	ttagtcatga	5580
acgettetet	attctatatg	aaaagccggt	teeggeetet	cacctttcct	ttttctccca	5640
atttttcagt	tgaaaaaggt	atatgcgtca	ggcgacctct	gaaattaaca	aaaaatttcc	5700
agtcatcgaa	tttgattctg	tgcgatagcg	cccctgtgtg	ttctcgttat	gttgaggaaa	5760
aaaataatgg	ttgctaagag	attcgaactc	ttgcatctta	cgatacctga	gtattcccac	5820
agttaactgc	ggtcaagata	tttcttgaat	caggcgcctt	agaccgctcg	gccaaacaac	5880
caattacttg	ttgagaaata	gagtataatt	atcctataaa	tataacgttt	ttgaacacac	5940
atgaacaagg	aagtacagga	caattgattt	tgaagagaat	gtggattttg	atgtaattgt	6000
tgggattcca	ttttaataa	ggcaataata	ttaggtatgt	ggatatacta	gaagttctcc	6060
tcgaccgtcg	atatgcggtg	tgaaataccg	cacagatgcg	taaggagaaa	ataccgcatc	6120
aggaaattgt	aaacgttaat	attttgttaa	aattcgcgtt	aaatttttgt	taaatcagct	6180
catttttaa	ccaataggcc	gaaatcggca	aaatccctta	taaatcaaaa	gaatagaccg	6240
agatagggtt	gagtgttgtt	ccagtttgga	acaagagtcc	actattaaag	aacgtggact	6300
ccaacgtcaa	agggcgaaaa	accgtctatc	agggcgatgg	cccactacgt	gaaccatcac	6360
cctaatcaag	tttttgggg	tcgaggtgcc	gtaaagcact	aaatcggaac	cctaaaggga	6420
gcccccgatt	tagagcttga	cggggaaagc	cggcgaacgt	ggcgagaaag	gaagggaaga	6480
aagcgaaagg	agcgggcgct	agggcgctgg	caagtgtagc	ggtcacgctg	cgcgtaacca	6540
ccacacccgc	cgcgcttaat	gcgccgctac	agggcgcgtc	gcgccattcg	ccattcaggc	6600

255

### -continued

tgcgcaactg	ttgggaaggg	cgatcggtgc	gggcctcttc	gctattacgc	cagctggcga	6660
aaggggggatg	tgctgcaagg	cgattaagtt	gggtaacgcc	agggttttcc	cagtcacgac	6720
gttgtaaaac	gacggccagt	gagcgcgcgt	aatacgactc	actatagggc	gaattgggta	6780
ccgggccccc	cctcgaggtc	gacggtatcg	ataagcttga	tatcgaattc	ctgcagcccg	6840
ggggatccgc	atgcttgcat	ttagtcgtgc	aatgtatgac	tttaagattt	gtgagcagga	6900
agaaaaggga	gaatcttcta	acgataaacc	cttgaaaaac	tgggtagact	acgctatgtt	6960
gagttgctac	gcaggctgca	caattacacg	agaatgctcc	cgcctaggat	ttaaggctaa	7020
gggacgtgca	atgcagacga	cagatctaaa	tgaccgtgtc	ggtgaagtgt	tcgccaaact	7080
tttcggttaa	cacatgcagt	gatgcacgcg	cgatggtgct	aagttacata	tatatata	7140
tatatatata	tagccatagt	gatgtctaag	taacctttat	ggtatatttc	ttaatgtgga	7200
aagatactag	cgcgcgcacc	cacacacaag	cttcgtcttt	tcttgaagaa	aagaggaagc	7260
tcgctaaatg	ggattccact	ttccgttccc	tgccagctga	tggaaaaagg	ttagtggaac	7320
gatgaagaat	aaaaagagag	atccactgag	gtgaaatttc	agctgacagc	gagtttcatg	7380
atcgtgatga	acaatggtaa	cgagttgtgg	ctgttgccag	ggagggtggt	tctcaacttt	7440
taatgtatgg	ccaaatcgct	acttgggttt	gttatataac	aaagaagaaa	taatgaactg	7500
attctcttcc	tccttcttgt	cctttcttaa	ttctgttgta	attaccttcc	tttgtaattt	7560
tttttgtaat	tattcttctt	aataatccaa	acaaacacac	atattacaat	agctagctga	7620
ggatgtcaac	agccggtaaa	gttattaagt	gtaaagcggc	agttttgtgg	gaagagaaaa	7680
agccgtttag	catagaagaa	gtagaagtag	cgccaccaaa	agcacacgag	gttagaatca	7740
agatggttgc	caccggaatc	tgtagatccg	acgaccatgt	ggtgagtggc	actctagtta	7800
ctcctttgcc	agtaatcgcg	ggacacgagg	ctgccggaat	cgttgaatcc	ataggtgaag	7860
gtgttaccac	tgttcgtcct	ggtgataaag	tgatcccact	gttcactcct	caatgtggta	7920
agtgtagagt	ctgcaaacat	cctgagggta	atttctgcct	taaaaatgat	ttgtctatgc	7980
ctagaggtac	tatgcaggat	ggtacaagca	gatttacatg	cagagggaaa	cctatacacc	8040
atttccttgg	tacttctaca	ttttcccaat	acacagtggt	ggacgagata	tctgtcgcta	8100
aaatcgatgc	agcttcacca	ctggaaaaag	tttgcttgat	agggtgcgga	ttttccaccg	8160
gttacggttc	cgcagttaaa	gttgcaaagg	ttacacaggg	ttcgacttgt	gcagtattcg	8220
gtttaggagg	agtaggacta	agcgttatta	tggggtgtaa	agctgcaggc	gcagcgagga	8280
ttataggtgt	agacatcaat	aaggacaaat	ttgcaaaagc	taaggaggtc	ggggctactg	8340
aatgtgttaa	ccctcaagat	tataagaaac	caatacaaga	agtccttact	gaaatgtcaa	8400
acggtggagt	tgatttctct	tttgaagtta	taggccgtct	tgatactatg	gtaactgcgt	8460
tgtcctgctg	tcaagaggca	tatggagtca	gtgtgatcgt	aggtgttcct	cctgattcac	8520
aaaatttgtc	gatgaatcct	atgctgttgc	taagcggtcg	tacatggaag	ggagctatat	8580
ttggcggttt	taagagcaag	gatagtgttc	caaaacttgt	tgccgacttt	atggcgaaga	8640
agtttgctct	tgatccttta	attacacatg	tattgccatt	cgagaaaatc	aatgaagggt	8700
ttgatttgtt	aagaagtggt	gaatctattc	gtacaatttt	aactttttga	ttaattaaga	8760
gtaagcgaat	ttcttatgat	ttatgatttt	tattattaaa	taagttataa	aaaaaataag	8820
tgtatacaaa	ttttaaaqtq	actcttaqqt	tttaaaacqa	aaattcttat	tettgagtaa	8880
ctctttccta	taqqtcaqqt	tqctttctca	qqtataqcat	qaqqtcqctc	ttattgacca	8940
cacctctacc	adcatacca	gcaaatgoot	gcaaatcoct	ccccatttca	cccaattota	9000
Laccocact	Jacardeeda	Junaucycet	Junanegel	uccuca	uccyca	

### -continued

gatatgctaa	ctccagcaat	gagttgatga	atctcggtgt	gtattttatg	tcctcagagg	9060					
acaacacctg	tggta					9075					
<210> SEQ I <211> LENGT <212> TYPE: <213> ORGAN <220> FEATU <223> OTHER	210> SEQ ID NO 53 211> LENGTH: 9491 212> TYPE: DNA 2213> ORGANISM: Artificial Sequence 220> FEATURE: 2223> OTHER INFORMATION: Synthetic construct pRS423 FBA ilvD(Strep)										
<400> SEQUE	ENCE: 53										
tcgcgcgttt	cggtgatgac	ggtgaaaacc	tctgacacat	gcagctcccg	gagacggtca	60					
cagcttgtct	gtaagcggat	gccgggagca	gacaagcccg	tcagggcgcg	tcagcgggtg	120					
ttggcgggtg	tcggggctgg	cttaactatg	cggcatcaga	gcagattgta	ctgagagtgc	180					
accataaatt	cccgttttaa	gagcttggtg	agcgctagga	gtcactgcca	ggtatcgttt	240					
gaacacggca	ttagtcaggg	aagtcataac	acagtccttt	cccgcaattt	tctttttcta	300					
ttactcttgg	cctcctctag	tacactctat	attttttat	gcctcggtaa	tgattttcat	360					
tttttttt	cccctagcgg	atgactcttt	ttttttctta	gcgattggca	ttatcacata	420					
atgaattata	cattatataa	agtaatgtga	tttcttcgaa	gaatatacta	aaaaatgagc	480					
aggcaagata	aacgaaggca	aagatgacag	agcagaaagc	cctagtaaag	cgtattacaa	540					
atgaaaccaa	gattcagatt	gcgatctctt	taaagggtgg	tcccctagcg	atagagcact	600					
cgatcttccc	agaaaaagag	gcagaagcag	tagcagaaca	ggccacacaa	tcgcaagtga	660					
ttaacgtcca	cacaggtata	gggtttctgg	accatatgat	acatgctctg	gccaagcatt	720					
ccggctggtc	gctaatcgtt	gagtgcattg	gtgacttaca	catagacgac	catcacacca	780					
ctgaagactg	cgggattgct	ctcggtcaag	cttttaaaga	ggccctactg	gcgcgtggag	840					
taaaaaggtt	tggatcagga	tttgcgcctt	tggatgaggc	actttccaga	gcggtggtag	900					
atctttcgaa	caggccgtac	gcagttgtcg	aacttggttt	gcaaagggag	aaagtaggag	960					
atctctcttg	cgagatgatc	ccgcattttc	ttgaaagctt	tgcagaggct	agcagaatta	1020					
ccctccacgt	tgattgtctg	cgaggcaaga	atgatcatca	ccgtagtgag	agtgcgttca	1080					
aggetettge	ggttgccata	agagaagcca	cctcgcccaa	tggtaccaac	gatgttccct	1140					
ccaccaaagg	tgttcttatg	tagtgacacc	gattatttaa	agctgcagca	tacgatatat	1200					
atacatgtgt	atatatgtat	acctatgaat	gtcagtaagt	atgtatacga	acagtatgat	1260					
actgaagatg	acaaggtaat	gcatcattct	atacgtgtca	ttctgaacga	ggcgcgcttt	1320					
ccttttttct	ttttgctttt	tcttttttt	tctcttgaac	tcgacggatc	tatgcggtgt	1380					
gaaataccgc	acagatgcgt	aaggagaaaa	taccgcatca	ggaaattgta	aacgttaata	1440					
ttttgttaaa	attcgcgtta	aatttttgtt	aaatcagctc	atttttaac	caataggccg	1500					
aaatcggcaa	aatcccttat	aaatcaaaag	aatagaccga	gatagggttg	agtgttgttc	1560					
cagtttggaa	caagagtcca	ctattaaaga	acgtggactc	caacgtcaaa	gggcgaaaaa	1620					
ccgtctatca	gggcgatggc	ccactacgtg	aaccatcacc	ctaatcaagt	ttttggggt	1680					
cgaggtgccg	taaagcacta	aatcggaacc	ctaaagggag	cccccgattt	agagettgae	1740					
ggggaaagcc	ggcgaacgtg	gcgagaaagg	aagggaagaa	agcgaaagga	gcgggcgcta	1800					
gggcgctggc	aagtgtagcg	gtcacgctgc	gcgtaaccac	cacacccgcc	gcgcttaatg	1860					
cgccgctaca	gggcgcgtcg	cgccattcgc	cattcaggct	gcgcaactgt	tgggaagggc	1920					

259

### -continued

gateggtgeg ggeetetteg etattaegee agetggegaa agggggatgt getgeaagge	1980
gattaagttg ggtaacgcca gggttttccc agtcacgacg ttgtaaaacg acggccagtg	2040
agcgcgcgta atacgactca ctatagggcg aattgggtac cgggcccccc ctcgaggtcg	2100
acggcgcgcc actggtagag agcgactttg tatgccccaa ttgcgaaacc cgcgatatcc	2160
ttctcgattc tttagtaccc gaccaggaca aggaaaagga ggtcgaaacg tttttgaaga	2220
aacaagagga actacacgga agctctaaag atggcaacca gccagaaact aagaaaatga	2280
agttgatgga tccaactggc accgctggct tgaacaacaa taccagcctt ccaacttctg	2340
taaataacgg cggtacgcca gtgccaccag taccgttacc tttcggtata cctcctttcc	2400
ccatgtttcc aatgcccttc atgcctccaa cggctactat cacaaatcct catcaagctg	2460
acgcaagccc taagaaatga ataacaatac tgacagtact aaataattgc ctacttggct	2520
tcacatacgt tgcatacgtc gatatagata ataatgataa tgacagcagg attatcgtaa	2580
tacgtaatag ttgaaaatct caaaaatgtg tgggtcatta cgtaaataat gataggaatg	2640
ggattettet attttteett ttteeattet ageageegte gggaaaaegt ggeateetet	2700
ctttcgggct caattggagt cacgctgccg tgagcatcct ctctttccat atctaacaac	2760
tgagcacgta accaatggaa aagcatgagc ttagcgttgc tccaaaaaag tattggatgg	2820
ttaataccat ttgtctgttc tcttctgact ttgactcctc aaaaaaaaaa	2880
caacagateg etteaattae geeeteacaa aaaetttttt eettettett egeeeaegtt	2940
aaattttatc cctcatgttg tctaacggat ttctgcactt gatttattat aaaaagacaa	3000
agacataata cttetetate aattteagtt attgttette ettgegttat tettetgtte	3060
ttctttttct tttgtcatat ataaccataa ccaagtaata catattcaaa ctagtatgac	3120
tgacaaaaaa actottaaag acttaagaaa togtagttot gtttaogatt caatggttaa	3180
atcacctaat cgtgctatgt tgcgtgcaac tggtatgcaa gatgaagact ttgaaaaacc	3240
tatogtoggt gtoatttoaa ottgggotga aaacacaoot tgtaatatoo aottacatga	3300
ctttggtaaa ctagccaaag tcggtgttaa ggaagctggt gcttggccag ttcagttcgg	3360
aacaatcacg gtttctgatg gaatcgccat gggaacccaa ggaatgcgtt tctccttgac	3420
atetegtgat attattgeag attetattga ageageeatg ggaggteata atgeggatge	3480
ttttgtagcc attggcggtt gtgataaaaa catgcccggt tctgttatcg ctatggctaa	3540
catggatatc ccagccattt ttgcttacgg cggaacaatt gcacctggta atttagacgg	3600
caaagatatc gatttagtct ctgtctttga aggtgtcggc cattggaacc acggcgatat	3660
gaccaaagaa gaagttaaag ctttggaatg taatgcttgt cccggtcctg gaggctgcgg	3720
tggtatgtat actgctaaca caatggcgac agctattgaa gttttgggac ttagccttcc	3780
gggttcatct tctcacccgg ctgaatccgc agaaaagaaa	3840
tegegetgtt gteaaaatge tegaaatggg ettaaaaeet tetgaeattt taaegegtga	3900
agettttgaa gatgetatta etgtaaetat ggetetggga ggtteaaeea aeteaaeeet	3960
tcacctctta gctattgccc atgctgctaa tgtggaattg acacttgatg atttcaatac	4020
tttccaagaa aaagttcctc atttggctga tttgaaacct tctggtcaat atgtattcca	4080
agacetttae aaggteggag gggtaeeage agttatgaaa tateteetta aaaatggett	4140
ccttcatggt gaccgtatca cttgtactgg caaaacagtc gctgaaaatt tgaaggcttt	4200
tgatgattta acacctggtc aaaaggttat tatgccgctt gaaaatccta aacgtgaaga	4260
tggtccgctc attattctcc atggtaactt ggctccagac ggtgccgttg ccaaagtttc	4320

261

### -continued

tggtgtaaaa	gtgcgtcgtc	atgtcggtcc	tgctaaggtc	tttaattctg	aagaagaagc	4380
cattgaagct	gtcttgaatg	atgatattgt	tgatggtgat	gttgttgtcg	tacgttttgt	4440
aggaccaaag	ggcggtcctg	gtatgcctga	aatgctttcc	ctttcatcaa	tgattgttgg	4500
taaagggcaa	ggtgaaaaag	ttgcccttct	gacagatggc	cgcttctcag	gtggtactta	4560
tggtcttgtc	gtgggtcata	tcgctcctga	agcacaagat	ggcggtccaa	tcgcctacct	4620
gcaaacagga	gacatagtca	ctattgacca	agacactaag	gaattacact	ttgatatctc	4680
cgatgaagag	ttaaaacatc	gtcaagagac	cattgaattg	ccaccgctct	attcacgcgg	4740
tatccttggt	aaatatgctc	acatcgtttc	gtctgcttct	aggggagccg	taacagactt	4800
ttggaagcct	gaagaaactg	gcaaaaaatg	ttgtcctggt	tgctgtggtt	aagcggccgc	4860
gttaattcaa	attaattgat	atagttttt	aatgagtatt	gaatctgttt	agaaataatg	4920
gaatattatt	tttatttatt	tatttatatt	attggtcggc	tctttcttc	tgaaggtcaa	4980
tgacaaaatg	atatgaagga	aataatgatt	tctaaaattt	tacaacgtaa	gatattttta	5040
caaaagccta	gctcatcttt	tgtcatgcac	tattttactc	acgcttgaaa	ttaacggcca	5100
gtccactgcg	gagtcatttc	aaagtcatcc	taatcgatct	atcgtttttg	atagctcatt	5160
ttggagttcg	cgattgtctt	ctgttattca	caactgtttt	aatttttatt	tcattctgga	5220
actcttcgag	ttctttgtaa	agtctttcat	agtagcttac	tttatcctcc	aacatattta	5280
acttcatgtc	aatttcggct	cttaaatttt	ccacatcatc	aagttcaaca	tcatctttta	5340
acttgaattt	attctctagc	tcttccaacc	aagceteatt	gctccttgat	ttactggtga	5400
aaagtgatac	actttgcgcg	caatccaggt	caaaactttc	ctgcaaagaa	ttcaccaatt	5460
tctcgacatc	atagtacaat	ttgttttgtt	ctcccatcac	aatttaatat	acctgatgga	5520
ttcttatgaa	gcgctgggta	atggacgtgt	cactctactt	cgcctttttc	cctactcctt	5580
ttagtacgga	agacaatgct	aataaataag	agggtaataa	taatattatt	aatcggcaaa	5640
aaagattaaa	cgccaagcgt	ttaattatca	gaaagcaaac	gtcgtaccaa	tccttgaatg	5700
cttcccaatt	gtatattaag	agtcatcaca	gcaacatatt	cttgttatta	aattaattat	5760
tattgatttt	tgatattgta	taaaaaaacc	aaatatgtat	aaaaaagtg	aataaaaaat	5820
accaagtatg	gagaaatata	ttagaagtct	atacgttaaa	ccaccgcggt	ggagctccag	5880
cttttgttcc	ctttagtgag	ggttaattgc	gcgcttggcg	taatcatggt	catagctgtt	5940
tcctgtgtga	aattgttatc	cgctcacaat	tccacacaac	ataggagccg	gaagcataaa	6000
gtgtaaagcc	tggggtgcct	aatgagtgag	gtaactcaca	ttaattgcgt	tgcgctcact	6060
gcccgctttc	cagtcgggaa	acctgtcgtg	ccagctgcat	taatgaatcg	gccaacgcgc	6120
ggggagaggc	ggtttgcgta	ttgggcgctc	ttccgcttcc	tcgctcactg	actcgctgcg	6180
ctcggtcgtt	cggctgcggc	gagcggtatc	agctcactca	aaggcggtaa	tacggttatc	6240
cacagaatca	ggggataacg	caggaaagaa	catgtgagca	aaaggccagc	aaaaggccag	6300
gaaccgtaaa	aaggccgcgt	tgctggcgtt	tttccatagg	ctccgccccc	ctgacgagca	6360
tcacaaaaat	cgacgctcaa	gtcagaggtg	gcgaaacccg	acaggactat	aaagatacca	6420
ggcgtttccc	cctggaagct	ccctcgtgcg	ctctcctgtt	ccgaccctgc	cgcttaccgg	6480
atacctgtcc	gcctttctcc	cttcgggaag	cgtggcgctt	tctcatagct	cacgctgtag	6540
gtatctcagt	tcggtgtagg	tcgttcgctc	caagetggge	tgtgtgcacg	aaccccccgt	6600
tcagcccgac	cgctgcgcct	tatccggtaa	ctatcgtctt	gagtccaacc	cggtaagaca	6660

263

### -continued

cgacttatcg	ccactggcag	cagccactgg	taacaggatt	agcagagcga	ggtatgtagg	6720
cggtgctaca 🤅	gagttettga	agtggtggcc	taactacggc	tacactagaa	ggacagtatt	6780
tggtatctgc g	gctctgctga	agccagttac	cttcggaaaa	agagttggta	gctcttgatc	6840
cggcaaacaa	accaccgctg	gtagcggtgg	ttttttgtt	tgcaagcagc	agattacgcg	6900
cagaaaaaaa	ggatctcaag	aagatccttt	gatcttttct	acggggtctg	acgctcagtg	6960
gaacgaaaac	tcacgttaag	ggattttggt	catgagatta	tcaaaaagga	tcttcaccta	7020
gatcctttta .	aattaaaaat	gaagttttaa	atcaatctaa	agtatatatg	agtaaacttg	7080
gtctgacagt	taccaatgct	taatcagtga	ggcacctatc	tcagcgatct	gtctatttcg	7140
ttcatccata 🤉	gttgcctgac	tccccgtcgt	gtagataact	acgatacggg	agggettace	7200
atctggcccc .	agtgctgcaa	tgataccgcg	agacccacgc	tcaccggctc	cagatttatc	7260
agcaataaac	cagccagccg	gaagggccga	gcgcagaagt	ggtcctgcaa	ctttatccgc	7320
ctccatccag ·	tctattaatt	gttgccggga	agctagagta	agtagttcgc	cagttaatag	7380
tttgcgcaac 🤅	gttgttgcca	ttgctacagg	catcgtggtg	tcacgctcgt	cgtttggtat	7440
ggetteatte	agctccggtt	cccaacgatc	aaggcgagtt	acatgatccc	ccatgttgtg	7500
caaaaaagcg g	gttagctcct	tcggtcctcc	gatcgttgtc	agaagtaagt	tggccgcagt	7560
gttatcactc .	atggttatgg	cagcactgca	taattctctt	actgtcatgc	catccgtaag	7620
atgcttttct g	gtgactggtg	agtactcaac	caagtcattc	tgagaatagt	gtatgcggcg	7680
accgagttgc	tettgeeegg	cgtcaatacg	ggataatacc	gcgccacata	gcagaacttt	7740
aaaagtgctc	atcattggaa	aacgttcttc	ggggcgaaaa	ctctcaagga	tcttaccgct	7800
gttgagatcc .	agttcgatgt	aacccactcg	tgcacccaac	tgatcttcag	catcttttac	7860
tttcaccagc 🤅	gtttctgggt	gagcaaaaac	aggaaggcaa	aatgccgcaa	aaaagggaat	7920
aagggcgaca	cggaaatgtt	gaatactcat	actcttcctt	tttcaatatt	attgaagcat	7980
ttatcagggt '	tattgtctca	tgagcggata	catatttgaa	tgtatttaga	aaaataaaca	8040
aataggggtt	ccgcgcacat	ttccccgaaa	agtgccacct	gaacgaagca	tctgtgcttc	8100
attttgtaga	acaaaaatgc	aacgcgagag	cgctaatttt	tcaaacaaag	aatctgagct	8160
gcatttttac	agaacagaaa	tgcaacgcga	aagcgctatt	ttaccaacga	agaatctgtg	8220
cttcattttt g	gtaaaacaaa	aatgcaacgc	gagagcgcta	atttttcaaa	caaagaatct	8280
gagetgeatt	tttacagaac	agaaatgcaa	cgcgagagcg	ctattttacc	aacaaagaat	8340
ctatacttct ·	tttttgttct	acaaaaatgc	atcccgagag	cgctatttt	ctaacaaagc	8400
atcttagatt .	actttttttc	teetttgtge	gctctataat	gcagtetett	gataactttt	8460
tgcactgtag g	gtccgttaag	gttagaagaa	ggctactttg	gtgtctattt	tctcttccat	8520
aaaaaagcc	tgactccact	tcccgcgttt	actgattact	agcgaagctg	cgggtgcatt	8580
ttttcaagat .	aaaggcatcc	ccgattatat	tctataccga	tgtggattgc	gcatactttg	8640
tgaacagaaa 🤉	gtgatagcgt	tgatgattct	tcattggtca	gaaaattatg	aacggtttct	8700
tctattttgt	ctctatatac	tacgtatagg	aaatgtttac	attttcgtat	tgttttcgat	8760
tcactctatg	aatagttett	actacaattt	ttttgtctaa	agagtaatac	tagagataaa	8820
cataaaaaat g	gtagaggtcg	agtttagatg	caagttcaag	gagcgaaagg	tggatgggta	8880
ggttatatag g	ggatatagca	cagagatata	tagcaaagag	atacttttga	gcaatgtttg	8940
tggaagcggt	attcgcaata	ttttagtagc	tcgttacagt	ccggtgcgtt	tttggttttt	9000
tgaaagtgcg	tcttcagagc	gcttttggtt	ttcaaaagcg	ctctgaagtt	cctatacttt	9060

-	1	5
- L	o	Э.

### -continued

266

ctagagaata ggaacttcgg aataggaac	t tcaaagcgtt tccgaaaacg agcgcttc	cg 9120
aaaatgcaac gcgagctgcg cacatacag	c tcactgttca cgtcgcacct atatctgc	gt 9180
gttgcctgta tatatatat catgagaag	a acggcatagt gcgtgtttat gcttaaat	gc 9240
gtacttatat gcgtctattt atgtaggat	g aaaggtagtc tagtacctcc tgtgatat	ta 9300
teccatteca tgeggggtat egtatgett	c cttcagcact accctttagc tgttctat	at 9360
gctgccactc ctcaattgga ttagtctca	t cetteaatge tateatttee tttgatat	tg 9420
gatcatctaa gaaaccatta ttatcatga	c attaacctat aaaaataggc gtatcacg	ag 9480
gccctttcgt c		9491
<pre>&lt;210&gt; SEQ ID NO 54 &lt;211&gt; LENGTH: 1000 &lt;212&gt; TYPE: DNA &lt;213&gt; ORGANISM: Artificial Sequ &lt;220&gt; FEATURE: &lt;223&gt; OTHER INFORMATION: FBA1 t &lt;400&gt; SEQUENCE: 54</pre>	ence erminator nt 4861 to 5860	
gttaattcaa attaattgat atagttttt	t aatgagtatt gaatctgttt agaaataa	tg 60
gaatattatt tttatttatt tatttatat	t attggtoggo tottttotto tgaaggto	aa 120
tgacaaaatg atatgaagga aataatgat	t tctaaaattt tacaacgtaa gatatttt	ta 180
caaaagccta gctcatcttt tgtcatgca	c tattttactc acgcttgaaa ttaacggc	ca 240
gtccactgcg gagtcatttc aaagtcatc	c taatcgatct atcgtttttg atagctca	tt 300
ttggagttcg cgattgtctt ctgttattc	a caactgtttt aatttttatt tcattctg	ga 360
actcttcgag ttctttgtaa agtctttca	t agtagettae tttateetee aacatatt	ta 420
acttcatgtc aatttcggct cttaaattt	t ccacatcatc aagttcaaca tcatcttt	ta 480
acttgaattt attctctagc tcttccaac	c aageeteatt geteettgat ttaetggt	ga 540
aaagtgatac actttgcgcg caatccagg	t caaaactttc ctgcaaagaa ttcaccaa	tt 600
tctcgacatc atagtacaat ttgttttgt	t ctcccatcac aatttaatat acctgatg	ga 660
ttcttatgaa gcgctgggta atggacgtg	t cactctactt cgcctttttc cctactcc	tt 720
ttagtacgga agacaatgct aataaataa	g agggtaataa taatattatt aatcggca	aa 780
aaagattaaa cgccaagcgt ttaattatc	a gaaagcaaac gtcgtaccaa tccttgaa	tg 840
cttcccaatt gtatattaag agtcatcac	a gcaacatatt cttgttatta aattaatt	at 900
tattgatttt tgatattgta taaaaaaac	c aaatatgtat aaaaaaagtg aataaaaa	at 960
accaagtatg gagaaatata ttagaagtc	t atacgttaaa	1000

<210> SEQ ID NO 55 <211> LENGTH: 1713 <212> TYPE: DNA <213> ORGANISM: Streptococcus mutans

<400> SEQUENCE: 55

atgactgaca	aaaaaactct	taaagactta	agaaatcgta	gttctgttta	cgattcaatg	60
gttaaatcac	ctaatcgtgc	tatgttgcgt	gcaactggta	tgcaagatga	agactttgaa	120
aaacctatcg	tcggtgtcat	ttcaacttgg	gctgaaaaca	caccttgtaa	tatccactta	180
catgactttg	gtaaactagc	caaagtcggt	gttaaggaag	ctggtgcttg	gccagttcag	240
ttcggaacaa	tcacggtttc	tgatggaatc	gccatgggaa	cccaaggaat	gcgtttctcc	300

267

### -continued

268

\_

-continued	
ttgacatete gtgatattat tgeagattet attgaageag ceatgggagg teataatgeg	360
gatgcttttg tagccattgg cggttgtgat aaaaacatgc ccggttctgt tatcgctatg	420
gctaacatgg atatcccagc catttttgct tacggcggaa caattgcacc tggtaattta	480
gacggcaaag atatcgattt agtctctgtc tttgaaggtg tcggccattg gaaccacggc	540
gatatgacca aagaagaagt taaagctttg gaatgtaatg cttgtcccgg tcctggaggc	600
tgcggtggta tgtatactgc taacacaatg gcgacagcta ttgaagtttt gggacttagc	660
cttccgggtt catcttctca cccggctgaa tccgcagaaa agaaagcaga tattgaagaa	720
gctggtcgcg ctgttgtcaa aatgctcgaa atgggcttaa aaccttctga cattttaacg	780
cgtgaagctt ttgaagatgc tattactgta actatggctc tgggaggttc aaccaactca	840
accetteace tettagetat tgeccatget getaatgtgg aattgacaet tgatgattte	900
aatactttcc aagaaaaagt tcctcatttg gctgatttga aaccttctgg tcaatatgta	960
ttccaagacc tttacaaggt cggaggggta ccagcagtta tgaaatatct ccttaaaaat	1020
ggetteette atggtgaeeg tateaettgt aetggeaaaa eagtegetga aaatttgaag	1080
gcttttgatg atttaacacc tggtcaaaag gttattatgc cgcttgaaaa tcctaaacgt	1140
gaagatggtc cgctcattat tctccatggt aacttggctc cagacggtgc cgttgccaaa	1200
gtttctggtg taaaagtgcg tcgtcatgtc ggtcctgcta aggtctttaa ttctgaagaa	1260
gaagccattg aagctgtctt gaatgatgat attgttgatg gtgatgttgt tgtcgtacgt	1320
tttgtaggac caaagggcgg tcctggtatg cctgaaatgc tttccctttc atcaatgatt	1380
gttggtaaag ggcaaggtga aaaagttgcc cttctgacag atggccgctt ctcaggtggt	1440
acttatggtc ttgtcgtggg tcatatcgct cctgaagcac aagatggcgg tccaatcgcc	1500
tacctgcaaa caggagacat agtcactatt gaccaagaca ctaaggaatt acactttgat	1560
atctccgatg aagagttaaa acatcgtcaa gagaccattg aattgccacc gctctattca	1620
cgcggtatcc ttggtaaata tgctcacatc gtttcgtctg cttctagggg agccgtaaca	1680
gacttttgga agcctgaaga aactggcaaa aaa	1713
<210> SEQ ID NO 56 <211> LENGTH: 571 <212> TYPE: PRT <213> ORGANISM: Streptococcus mutans	
<400> SEQUENCE: 56	
Met Thr Asp Lys Lys Thr Leu Lys Asp Leu Arg Asn Arg Ser Ser Val151015	
Tyr Asp Ser Met Val Lys Ser Pro Asn Arg Ala Met Leu Arg Ala Thr 20 25 30	
Gly Met Gln Asp Glu Asp Phe Glu Lys Pro Ile Val Gly Val Ile Ser 35 40 45	
Thr Trp Ala Glu Asn Thr Pro Cys Asn Ile His Leu His Asp Phe Gly 50 55 60	
Lys Leu Ala Lys Val Gly Val Lys Glu Ala Gly Ala Trp Pro Val Gln 65 70 75 80	
Phe Gly Thr Ile Thr Val Ser Asp Gly Ile Ala Met Gly Thr Gln Gly 85 90 95	
Met Arg Phe Ser Leu Thr Ser Arg Asp Ile Ile Ala Asp Ser Ile Glu 100 105 110	
Ala Ala Met Gly Gly His Asn Ala Asp Ala Phe Val Ala Ile Gly Gly 115 120 125	

Сүз	Asp 130	Lys	Asn	Met	Pro	Gly 135	Ser	Val	Ile	Ala	Met 140	Ala	Asn	Met	Asp
Ile 145	Pro	Ala	Ile	Phe	Ala 150	Tyr	Gly	Gly	Thr	Ile 155	Ala	Pro	Gly	Asn	Leu 160
Asp	Gly	Lys	Aab	Ile 165	Asp	Leu	Val	Ser	Val 170	Phe	Glu	Gly	Val	Gly 175	His
Trp	Asn	His	Gly 180	Asp	Met	Thr	Lys	Glu 185	Glu	Val	Lya	Ala	Leu 190	Glu	Сүз
Asn	Ala	Cys 195	Pro	Gly	Pro	Gly	Gly 200	Сүз	Gly	Gly	Met	Tyr 205	Thr	Ala	Asn
Thr	Met 210	Ala	Thr	Ala	Ile	Glu 215	Val	Leu	Gly	Leu	Ser 220	Leu	Pro	Gly	Ser
Ser 225	Ser	His	Pro	Ala	Glu 230	Ser	Ala	Glu	Lys	Lys 235	Ala	Asp	Ile	Glu	Glu 240
Ala	Gly	Arg	Ala	Val 245	Val	Lys	Met	Leu	Glu 250	Met	Gly	Leu	Lys	Pro 255	Ser
Asp	Ile	Leu	Thr 260	Arg	Glu	Ala	Phe	Glu 265	Asp	Ala	Ile	Thr	Val 270	Thr	Met
Ala	Leu	Gly 275	Gly	Ser	Thr	Asn	Ser 280	Thr	Leu	His	Leu	Leu 285	Ala	Ile	Ala
His	Ala 290	Ala	Asn	Val	Glu	Leu 295	Thr	Leu	Asp	Asp	Phe 300	Asn	Thr	Phe	Gln
Glu 305	Lys	Val	Pro	His	Leu 310	Ala	Asp	Leu	Lys	Pro 315	Ser	Gly	Gln	Tyr	Val 320
Phe	Gln	Asp	Leu	Tyr 325	Lys	Val	Gly	Gly	Val 330	Pro	Ala	Val	Met	Lys 335	Tyr
Leu	Leu	Lys	Asn 340	Gly	Phe	Leu	His	Gly 345	Asp	Arg	Ile	Thr	Cys 350	Thr	Gly
Lys	Thr	Val 355	Ala	Glu	Asn	Leu	Lys 360	Ala	Phe	Asp	Asp	Leu 365	Thr	Pro	Gly
Gln	Lys 370	Val	Ile	Met	Pro	Leu 375	Glu	Asn	Pro	Lys	Arg 380	Glu	Asp	Gly	Pro
Leu 385	Ile	Ile	Leu	His	Gly 390	Asn	Leu	Ala	Pro	Asp 395	Gly	Ala	Val	Ala	Lys 400
Val	Ser	Gly	Val	Lys 405	Val	Arg	Arg	His	Val 410	Gly	Pro	Ala	Lys	Val 415	Phe
Asn	Ser	Glu	Glu 420	Glu	Ala	Ile	Glu	Ala 425	Val	Leu	Asn	Asb	Asp 430	Ile	Val
Asp	Gly	Asp 435	Val	Val	Val	Val	Arg 440	Phe	Val	Gly	Pro	Lys 445	Gly	Gly	Pro
Gly	Met 450	Pro	Glu	Met	Leu	Ser 455	Leu	Ser	Ser	Met	Ile 460	Val	Gly	Lys	Gly
Gln 465	Gly	Glu	Lys	Val	Ala 470	Leu	Leu	Thr	Asp	Gly 475	Arg	Phe	Ser	Gly	Gly 480
Thr	Tyr	Gly	Leu	Val 485	Val	Gly	His	Ile	Ala 490	Pro	Glu	Ala	Gln	Asp 495	Gly
Gly	Pro	Ile	Ala 500	Tyr	Leu	Gln	Thr	Gly 505	Asp	Ile	Val	Thr	Ile 510	Asp	Gln
Asp	Thr	Lys 515	Glu	Leu	His	Phe	Asp 520	Ile	Ser	Aap	Glu	Glu 525	Leu	Lys	His
Arg	Gln 530	Glu	Thr	Ile	Glu	Leu 535	Pro	Pro	Leu	Tyr	Ser 540	Arg	Gly	Ile	Leu

Gly Lys Tyr Ala His Ile Val Ser Ser Ala Ser Arg Gly Ala Val Thr 545 550 555 560	
Asp Phe Trp Lys Pro Glu Glu Thr Gly Lys Lys	
565 570	
<210> SEQ ID NO 57 <211> LENGTH: 4280 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: pUC19-URA3r	
<400> SEQUENCE: 57	
ggggateete tagagtegae etgeaggeat geaagettgg egtaateatg gteatagetg	60
tttcctgtgt gaaattgtta tccgctcaca attccacaca acatacgagc cggaagcata	120
aagtgtaaag cctggggtgc ctaatgagtg agctaactca cattaattgc gttgcgctca	180
ctgcccgctt tccagtcggg aaacctgtcg tgccagctgc attaatgaat cggccaacgc	240
gcggggagag gcggtttgcg tattgggcgc tettecgett eetegeteae tgaetegetg	300
cgctcggtcg ttcggctgcg gcgagcggta tcagctcact caaaggcggt aatacggtta	360
tccacagaat caggggataa cgcaggaaag aacatgtgag caaaaggcca gcaaaaggcc	420
aggaaccgta aaaaggccgc gttgctggcg tttttccata ggctccgccc ccctgacgag	480
catcacaaaa atcgacgctc aagtcagagg tggcgaaacc cgacaggact ataaagatac	540
caggogttto coortggaag otoootogtg ogototootg ttoogacoot googottaco	600
ggatacetgt cegeetttet ecetteggga agegtggege ttteteatag eteaegetgt	660
aggtatetea gtteggtgta ggtegttege teeaagetgg getgtgtgea egaaceeeee	720
gttcagcccg accgctgcgc cttatccggt aactatcgtc ttgagtccaa cccggtaaga	780
cacgacttat cgccactggc agcagccact ggtaacagga ttagcagagc gaggtatgta	840
ggcggtgcta cagagttett gaagtggtgg eetaaetaeg getaeaetag aaggaeagta	900
tttggtatet gegetetget gaageeagtt acetteggaa aaagagttgg tagetettga	960
teeggeaaac aaaceaeege tggtageggt ggtttttttg tttgeaagea geagattaeg	1020
cgcagaaaaa aaggatctca agaagatcct ttgatctttt ctacggggtc tgacgctcag	1080
tggaacgaaa actcacgtta agggattttg gtcatgagat tatcaaaaag gatcttcacc	1140
tagateettt taaattaaaa atgaagtttt aaateaatet aaagtatata tgagtaaaet	1200
tggtctgaca gttaccaatg cttaatcagt gaggcaccta tctcagcgat ctgtctattt	1260
cgttcatcca tagttgcctg actccccgtc gtgtagataa ctacgatacg ggagggctta	1320
ccatctggcc ccagtgctgc aatgataccg cgagacccac gctcaccggc tccagattta	1380
tcagcaataa accagccagc cggaagggcc gagcgcagaa gtggtcctgc aactttatcc	1440
gcetecatee agtetattaa ttgttgeegg gaagetagag taagtagtte geeagttaat	1500
agtttgegea aegttgttge cattgetaea ggeategtgg tgteaegete gtegtttggt	1560
atggetteat teageteegg tteecaaega teaaggegag ttaeatgate ecceatgttg	1620
tgcaaaaaag cggttagctc cttcggtcct ccgatcgttg tcagaagtaa gttggccgca	1680
gtgttatcac tcatggttat ggcagcactg cataattete ttactgtcat gecateegta	1740
agatgetttt etgtgaetgg tgagtaetea aceaagteat tetgagaata gtgtatgegg	1800
cgaccgagtt gctcttgccc ggcgtcaata cgggataata ccgcgccaca tagcagaact	1860
ttaaaagtgc tcatcattgg aaaacgttct tcggggcgaa aactctcaag gatcttaccg	1920

273

### -continued

ctgttgagat	ccagttcgat	gtaacccact	cgtgcaccca	actgatcttc	agcatcttt	1980
actttcacca	gcgtttctgg	gtgagcaaaa	acaggaaggc	aaaatgccgc	aaaaaggga	2040
ataagggcga	cacggaaatg	ttgaatactc	atactcttcc	tttttcaata	ttattgaagc	2100
atttatcagg	gttattgtct	catgagcgga	tacatatttg	aatgtattta	gaaaaataaa	2160
caaatagggg	ttccgcgcac	atttccccga	aaagtgccac	ctgacgtcta	agaaaccatt	2220
attatcatga	cattaaccta	taaaaatagg	cgtatcacga	ggccctttcg	tctcgcgcgt	2280
ttcggtgatg	acggtgaaaa	cctctgacac	atgcagctcc	cggagacggt	cacagcttgt	2340
ctgtaagcgg	atgccgggag	cagacaagcc	cgtcagggcg	cgtcagcggg	tgttggcggg	2400
tgtcggggct	ggcttaacta	tgcggcatca	gagcagattg	tactgagagt	gcaccatatg	2460
cggtgtgaaa	taccgcacag	atgcgtaagg	agaaaatacc	gcatcaggcg	ccattcgcca	2520
ttcaggctgc	gcaactgttg	ggaagggcga	tcggtgcggg	cctcttcgct	attacgccag	2580
ctggcgaaag	ggggatgtgc	tgcaaggcga	ttaagttggg	taacgccagg	gttttcccag	2640
tcacgacgtt	gtaaaacgac	ggccagtgaa	ttcgagctcg	gtacccccgg	ctctgagaca	2700
gtagtaggtt	agtcatcgct	ctaccgacgc	gcaggaaaag	aaagaagcat	tgcggattac	2760
gtattctaat	gttcagcccg	cggaacgcca	gcaaatcacc	acccatgcgc	atgatactga	2820
gtcttgtaca	cgctgggctt	ccagtgtact	gagagtgcac	cataccacag	cttttcaatt	2880
caattcatca	tttttttt	attcttttt	ttgatttcgg	tttctttgaa	attttttga	2940
ttcggtaatc	tccgaacaga	aggaagaacg	aaggaaggag	cacagactta	gattggtata	3000
tatacgcata	tgtagtgttg	aagaaacatg	aaattgccca	gtattcttaa	cccaactgca	3060
cagaacaaaa	acctgcagga	aacgaagata	aatcatgtcg	aaagctacat	ataaggaacg	3120
tgctgctact	catcctagtc	ctgttgctgc	caagctattt	aatatcatgc	acgaaaagca	3180
aacaaacttg	tgtgcttcat	tggatgttcg	taccaccaag	gaattactgg	agttagttga	3240
agcattaggt	cccaaaattt	gtttactaaa	aacacatgtg	gatatcttga	ctgattttc	3300
catggagggc	acagttaagc	cgctaaaggc	attatccgcc	aagtacaatt	ttttactctt	3360
cgaagacaga	aaatttgctg	acattggtaa	tacagtcaaa	ttgcagtact	ctgcgggtgt	3420
atacagaata	gcagaatggg	cagacattac	gaatgcacac	ggtgtggtgg	gcccaggtat	3480
tgttagcggt	ttgaagcagg	cggcagaaga	agtaacaaag	gaacctagag	gccttttgat	3540
gttagcagaa	ttgtcatgca	agggctccct	atctactgga	gaatatacta	agggtactgt	3600
tgacattgcg	aagagcgaca	aagattttgt	tatcggcttt	attgctcaaa	gagacatggg	3660
tggaagagat	gaaggttacg	attggttgat	tatgacaccc	ggtgtgggtt	tagatgacaa	3720
gggagacgca	ttgggtcaac	agtatagaac	cgtggatgat	gtggtctcta	caggatctga	3780
cattattatt	gttggaagag	gactatttgc	aaagggaagg	gatgctaagg	tagagggtga	3840
acgttacaga	aaagcaggct	gggaagcata	tttgagaaga	tgcggccagc	aaaactaaaa	3900
aactgtatta	taagtaaatg	catgtatact	aaactcacaa	attagagett	caatttaatt	3960
atatcagtta	ttaccctatg	cggtgtgaaa	taccgcacag	atgcgtaagg	agaaaatacc	4020
gcatcaggaa	attgtaaacg	ttaatatttt	gttaaaattc	gcgttaaatt	tttgttaaat	4080
cagctcattt	tttaaccaat	aggccgaaat	cggcaaaatc	ttcagcccgc	ggaacgccag	4140
caaatcacca	cccatgcgca	tgatactgag	tcttgtacac	gctgggcttc	cagtgatgat	4200
acaacgagtt	agccaaggtg	agcacggatg	tctaaattag	aattacgttt	taatatcttt	4260

275

-continued

ttttccatat ctagggctag	4280
cocould couggeoug	1200
<210> SEQ ID NO 58 <211> LENGTH: 30 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Primer 114117-11A	
<400> SEQUENCE: 58	
gcatgcttgc atttagtcgt gcaatgtatg	30
<210> SEQ ID NO 59 <211> LENGTH: 54 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Primer 114117-11B <400> SEQUENCE: 59	
gaacattaga atacgtaatc cgcaatgcac tagtaccaca ggtgttgtcc tctg	54
<210> SEQ ID NO 60 <211> LENGTH: 54 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Primer 114117-11C	
<400> SEQUENCE: 60	
cagaggacaa cacctgtggt actagtgcat tgcggattac gtattctaat gttc	54
<210> SEQ ID NO 61 <211> LENGTH: 28 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Primer 114117-11D	
<400> SEQUENCE: 61	
caccttggct aactcgttgt atcatcac	28
<210> SEQ ID NO 62 <211> LENGTH: 100 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Primer 114117-13A	
<400> SEQUENCE: 62	
ttttaagccg aatgagtgac agaaaaagcc cacaacttat caagtgatat tgaacaaagg	60
gcgaaacttc gcatgcttgc atttagtcgt gcaatgtatg	100
<210> SEQ ID NO 63 <211> LENGTH: 98 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Primer 114117-13B	
<400> SEQUENCE: 63	
cccaattggt aaatattcaa caagagacgc gcagtacgta acatgcgaat tgcgtaattc	60
acggcgataa caccttggct aactcgttgt atcatcac	98

<210> SEQ ID NO 64	
<211> LENGTH: 29	
<212> TYPE: DNA	
<213> ORGANISM: Artificial Sequence	
<220> FEATURE:	
<223> OTHER INFORMATION: Primer 112590-34G	
<400> SEQUENCE: 64	
caaaageeea tgteeeacae caaaggatg	29
<210> SEQ ID NO 65	
<211> LENGTH: 26	
<212> TYPE: DNA	
<213> ORGANISM: Artificial Sequence	
<220> FEATURE:	
<2233 OTHER INFORMATION: FILMET 112590-34H	
<400> SEOUENCE: 65	
caccatcgcg cgtgcatcac tgcatg	26
<210> SEQ ID NO 66	
<211> LENGTH: 28	
<212> TYPE: DNA	
<2213> ORGANISM: ATTITICIAL Sequence	
<223> OTHER INFORMATION: Primer 112590-34F	
<400> SEQUENCE: 66	
tcggtttttg caatatgacc tgtgggcc	28
<210> SEQ ID NO 67	
<zii> LENGIH: ZZ</zii>	
<212> IIPE: DNA	
<220> FEATURE:	
<223> OTHER INFORMATION: Primer 112590-49E	
<400> SEQUENCE: 67	
gagaagatgc ggccagcaaa ac	22
-2105 SEO ID NO 68	
<211> LENGTH · 99	
<212> TYPE: DNA	
<213> ORGANISM: artificial sequence	
<220> FEATURE:	
<223> OTHER INFORMATION: Primer 114117-27A	
<400> SEQUENCE: 68	
	<u> </u>
teettetea attattattt tetaeteata aceteaegea aaataaeaea gteaaateaa	60
tcaaagtatg actgacaaaa aaactottaa agacttaag	99
countyputy accymount and coordan agaecounty	
<210> SEQ ID NO 69	
<211> LENGTH: 77	
<212> TYPE: DNA	
<213> ORGANISM: artificial sequence	
<220> FEATURE:	
<223> OTHER INFORMATION: Primer 114117-27B	
<#AAA PEÄNENCE: 0A	
assastere staatsata aansataatt atttattta aatttaaaat ataaattat	60
yaavallaya alacylaalo ogoaalyoll ottoottto ogtttaacgt alagadttot	00
aatatattto tooatao	77
<210> SEO ID NO 70	
~	

<211> LENGTH: 45 <212> TYPE: DNA -continued

-continued	
<213> ORGANISM: artificial sequence <220> FEATURE:	
<223> OTHER INFORMATION: Primer 114117-27C	
<400> SEQUENCE: 70	
aaacggaaaa gaaagaagca ttgcggatta cgtattctaa tgttc	45
<210> SEQ ID NO 71 <211> LENGTH: 88 <212> TYPE: DNA <213> ORGANISM: artificial sequence <220> FEATURE: <223> OTHER INFORMATION: Primer 114117-27D	
<400> SEQUENCE: 71	
tatttttcgt tacataaaaa tgcttataaa actttaacta ataattagag attaaatcgc	60
caccttggct aactcgttgt atcatcac	88
<210> SEQ ID NO 72 <211> LENGTH: 27 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Primer 114117-36D	
<400> SEQUENCE: 72	
gacttttgga agcetgaaga aactgge	27
<210> SEQ ID NO 73 <211> LENGTH: 20 <212> TYPE: DNA <213> ORGANISM: artificial sequence <220> FEATURE: <223> OTHER INFORMATION: Primer 135 <400> SEQUENCE: 73	
cttqqcaqca acaqqactaq	20
<210> SEQ ID NO 74 <211> LENGTH: 26 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: 112590-30F <400> SEQUENCE: 74	
ccaggccaat tcaacagact gtcggc	26
<210> SEQ ID NO 75 <211> LENGTH: 2347 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: URA3r2 template DNA	
<400> SEQUENCE: 75	
gcattgcgga ttacgtattc taatgttcag gtgctggaag aagagctgct taaccgccgc	60
gcccagggtg aagatccacg ctactttacc ctgcgtcgtc tggatttcgg cggctgtcgt	120
ctttcgctgg caacgccggt tgatgaagcc tgggacggtc cgctctcctt aaacggtaaa	180
cgtategeea eetettatee teacetgete aagegttate tegaceagaa aggeatetet	240
tttaaateet gettaetgaa eggttetgtt gaagtegeee egegtgeegg aetggeggat	300
gcgatttgcg atctggtttc caccggtgcc acgctggaag ctaacggcct gcgcgaagtc	360

281

#### -continued

gaagttatct atcgctcgaa agcctgcctg attcaacgcg atggcgaaat ggaagaatcc 420 aaacagcaac tgatcgacaa actgctgacc cgtattcagg gtgtgatcca ggcgcgcgaa 480 tcaaaataca tcatgatgca cgcaccgacc gaacgtctgg atgaagtcat ggtacctact 540 gagagtgcac cataccacag cttttcaatt caattcatca ttttttttt attcttttt 600 ttgatttcgg tttctttgaa attttttga ttcggtaatc tccgaacaga aggaagaacg 660 aaggaaggag cacagactta gattggtata tatacgcata tgtagtgttg aagaaacatg 720 aaattgccca gtattettaa eecaactgca cagaacaaaa aeetgcagga aacgaagata 780 aatcatgtcg aaagctacat ataaggaacg tgctgctact catcctagtc ctgttgctgc 840 caagctattt aatatcatgc acgaaaagca aacaaacttg tgtgcttcat tggatgttcg 900 taccaccaag gaattactgg agttagttga agcattaggt cccaaaattt gtttactaaa 960 aacacatgtg gatatettga etgattttte eatggaggge acagttaage egetaaagge 1020 1080 attatccqcc aaqtacaatt ttttactctt cqaaqacaqa aaatttqctq acattqqtaa tacagtcaaa ttgcagtact ctgcgggtgt atacagaata gcagaatggg cagacattac 1140 1200 qaatqcacac qqtqtqqtqq qcccaqqtat tqttaqcqqt ttqaaqcaqq cqqcaqaaqa agtaacaaag gaacctagag gccttttgat gttagcagaa ttgtcatgca agggctccct 1260 atctactgga gaatatacta agggtactgt tgacattgcg aagagcgaca aagattttgt 1320 1380 tatcggcttt attgctcaaa gagacatggg tggaagagat gaaggttacg attggttgat tatgacaccc ggtgtgggtt tagatgacaa gggagacgca ttgggtcaac agtatagaac 1440 cgtggatgat gtggtctcta caggatctga cattattatt gttggaagag gactatttgc 1500 aaagggaagg gatgctaagg tagagggtga acgttacaga aaagcaggct gggaagcata 1560 tttgagaaga tgcggccagc aaaactaaaa aactgtatta taagtaaatg catgtatact 1620 aaactcacaa attagagctt caatttaatt atatcagtta ttaccctatg cggtgtgaaa 1680 taccgcacag atgcgtaagg agaaaatacc gcatcaggaa attgtaaacg ttaatatttt 1740 gttaaaattc gcgttaaatt tttgttaaat cagctcattt tttaaccaat aggccgaaat 1800 cggcaaaatc tctagagtgc tggaagaaga gctgcttaac cgccgcgccc agggtgaaga 1860 tccacgctac tttaccctgc gtcgtctgga tttcggcggc tgtcgtcttt cgctggcaac 1920 1980 geoggttgat gaageetggg acggteeget eteettaaae ggtaaaegta tegeeaeete ttateeteac etgeteaage gttatetega eeagaaagge atetettta aateetgett 2040 actgaacggt tctgttgaag tcgccccgcg tgccggactg gcggatgcga tttgcgatct 2100 ggtttccacc ggtgccacgc tggaagctaa cggcctgcgc gaagtcgaag ttatctatcg 2160 ctcgaaagcc tgcctgattc aacgcgatgg cgaaatggaa gaatccaaac agcaactgat 2220 cgacaaactg ctgacccgta ttcagggtgt gatccaggcg cgcgaatcaa aatacatcat 2280 gatgcacgca ccgaccgaac gtctggatga agtcatccag tgatgataca acgagttagc 2340 caaggtg 2347

<210> SEQ ID NO 76 <211> LENGTH: 80 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Primer 114117-45A

<400> SEQUENCE: 76

284

### 283

### continued

 cttcgaagaa tatactaaaa aatgagcagg caagataaac gaaggcaaag gcattgcgga	60
ttacgtattc taatgttcag	80
<210> SEQ ID NO 77 <211> LENGTH: 81 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Primer 114117-45B	
<400> SEQUENCE: 77	
tatacacatg tatatatatc gtatgctgca gctttaaata atcggtgtca caccttggct	60
aactcgttgt atcatcactg g	81
<210> SEQ ID NO 78 <211> LENGTH: 90 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Primer 384	
<400> SEQUENCE: 78	
atggttcatt taggtccaaa aaaaccacaa gccagaaagg gttccatggc cgatgtgcca	60
gcattgcgga ttacgtattc taatgttcag	90
<210> SEQ ID NO 79 <211> LENGTH: 91 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Primer 385	
<400> SEQUENCE: 79	
ttaagcaccg atgataccaa cggacttacc ttcagcaatt cttttttggg ccaaagcagc	60
caccttgget aactegttgt atcateaetg g	91
<210> SEQ ID NO 80 <211> LENGTH: 24 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Primer N869	
<400> SEQUENCE: 80	
ctaggatgag tagcagcacg ttcc	24
<210> SEQ ID NO 81 <211> LENGTH: 26 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Primer N871	
<400> SEQUENCE: 81	
ccaatteegt gatgtetett tgttge	26
<210> SEQ ID NO 82 <211> LENGTH: 20 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Primer N946	

<400> SEQUENCE: 82

<b>_</b>	o	5
- 2	o	Э.

-continued

-continued	
gtgaacgagt tcacaaccgc	20
<pre>&lt;210&gt; SEQ ID NO 83 &lt;211&gt; LENGTH: 22 &lt;212&gt; TYPE: DNA &lt;213&gt; ORGANISM: Artificial Sequence &lt;220&gt; FEATURE: &lt;223&gt; OTHER INFORMATION: Primer N947 &lt;400&gt; SECUENCE: 83</pre>	
	22
gttegtteea gaattateae ge	22
<210> SEQ ID NO 84 <211> LENGTH: 26 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Primer PDC5:KanMXF	
<400> SEQUENCE: 84	
gacttgaata atgcagcggc gcttgc	26
<210> SEQ ID NO 85 <211> LENGTH: 30 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Primer PDC5::KanMXR	
<400> SEQUENCE: 85	
ccaccctctt caattagcta agatcatagc	30
<210> SEQ ID NO 86 <211> LENGTH: 49 <212> TYPE: DNA <213> ORGANISM: Artificial sequence <220> FEATURE: <223> OTHER INFORMATION: Primer oBP457	
<400> SEQUENCE: 86	
ccagaaaccc tatacctgtg tggacgtaag gccatgaagc tttttcttt	49
<210> SEQ ID NO 87 <211> LENGTH: 49 <212> TYPE: DNA <213> ORGANISM: Artificial sequence <220> FEATURE: <223> OTHER INFORMATION: Primer oBP458	
<400> SEQUENCE: 87	
attggaaaga aaaagettea tggeettaeg teeacaeagg tatagggtt	49
<210> SEQ ID NO 88 <211> LENGTH: 22 <212> TYPE: DNA <213> ORGANISM: Artificial sequence <220> FEATURE: <223> OTHER INFORMATION: Primer oBP459	
<400> SEQUENCE: 88	
cataagaaca cctttggtgg ag	22
<210> SEQ ID NO 89 <211> LENGTH: 24 <212> TYPE: DNA	

-continued	
<pre>&lt;213&gt; ORGANISM: Artificial sequence &lt;220&gt; FEATURE: &lt;223&gt; OTHER INFORMATION: Primer oBP452</pre>	
<400> SEQUENCE: 89	
ttctcgacgt gggcctttt cttg 24	
<pre>&lt;210&gt; SEQ ID NO 90 &lt;211&gt; LENGTH: 49 &lt;212&gt; TYPE: DNA &lt;213&gt; ORGANISM: Artificial sequence &lt;220&gt; FEATURE: &lt;220&gt; OTHER INFORMATION: Primer ORP455</pre>	
<400> SEQUENCE: 90	
tatggaccet gaaaccacag ceacattgta accaccacga cggttgttg 49	
<pre>&lt;210&gt; SEQ ID NO 91 &lt;211&gt; LENGTH: 49 &lt;212&gt; TYPE: DNA</pre>	
<213> ORGANISM: Artificial sequence <220> FEATURE:	
<223> OTHER INFORMATION: Primer oBP456	
<400> SEQUENCE: 91	
tttagcaaca accgtcgtgg tggttacaat gtggctgtgg tttcagggt 49	
<210> SEQ ID NO 92 <211> LENGTH: 512 <212> TYPE: PRT <213> ORGANISM: Aspergillus niger	
<400> SEQUENCE: 92	
Met Ala Leu Leu Ala Val Ala Gly Val Tyr Ala Phe Ala Ala Leu Leu 1 5 10 15	
Val Ala Ile Val Leu Asn Val Thr Arg Gln Leu Leu Phe Arg Asn Glu 20 25 30	
Lys Glu Pro Pro Val Val Phe His Trp Ile Pro Phe Leu Gly Ser Thr 35 40 45	
Ile Ser Tyr Gly Met Asp Pro Tyr Thr Phe Phe Phe Ser Cys Arg Lys 50 55 60	
Lys Tyr Gly Asp Ile Phe Thr Phe Val Leu Leu Gly Gln Lys Thr Thr65707580	
Val Tyr Leu Gly Val Gln Gly Asn Asp Phe Ile Leu Asn Gly Lys Leu 85 90 95	
Lys Asp Val Ser Ala Glu Glu Val Tyr Ser Pro Leu Thr Thr Pro Val 100 105 110	
Phe Gly Ser Asp Val Val Tyr Asp Cys Pro Asn Ser Lys Leu Met Glu 115 120 125	
Gln Lys Lys Phe Ile Lys Phe Gly Leu Thr Gln Ala Ala Leu Glu Ser 130 135 140	
His Val Gln Leu Ile Glu Lys Glu Thr Leu Asp Tyr Leu Arg Asp Ser 145 150 155 160	
Pro Arg Phe Asn Gly Ala Ser Gly Val Ile Asp Ile Pro Ala Ala Met 165 170 175	
Ala Glu Ile Thr Ile Tyr Thr Ala Ala Arg Ala Leu Gln Gly Glu Glu 180 185 190	
Val Arg Lys Leu Thr Ala Glu Phe Ala Glu Leu Tyr His Asp Leu 195 200 205	

22

290

Asp	Lys 210	Gly	Phe	Ser	Pro	Ile 215	Asn	Phe	Met	Leu	Pro 220	Trp	Ala	Pro	Leu	
Pro 225	His	Asn	Arg	Lys	Arg 230	Asp	Ala	Ala	His	Ala 235	Arg	Met	Arg	Glu	Ile 240	
Tyr	Thr	Asp	Ile	Ile 245	Asn	Glu	Arg	Arg	Lys 250	Asn	Pro	Asp	Glu	Glu 255	Lys	
Ser	Asp	Met	Ile 260	Trp	Asn	Leu	Met	His 265	Суз	Thr	Tyr	Гла	Ser 270	Gly	Gln	
Pro	Val	Pro 275	Asp	Lys	Glu	Ile	Ala 280	His	Met	Met	Ile	Thr 285	Leu	Leu	Met	
Ala	Gly 290	Gln	His	Ser	Ser	Ser 295	Ser	Ile	Ser	Ser	Trp 300	Ile	Met	Leu	Arg	
Leu 305	Ala	Ser	Glu	Pro	Gln 310	Val	Leu	Glu	Glu	Leu 315	Tyr	Gln	Glu	Gln	Leu 320	
Ala	Ser	Leu	Ser	Asn 325	Arg	Asn	Gly	Val	Phe 330	Glu	Pro	Leu	Gln	Tyr 335	Gln	
Asp	Leu	Asp	Lys 340	Leu	Pro	Phe	Leu	Gln 345	Ser	Val	Ile	ГЛа	Glu 350	Thr	Leu	
Arg	Ile	His 355	Ser	Ser	Ile	His	Ser 360	Ile	Met	Arg	Lya	Val 365	Lys	Asn	Pro	
Leu	. Pro 370	Val	Pro	Gly	Thr	Ser 375	Tyr	Ile	Ile	Pro	Glu 380	Aab	His	Val	Leu	
Leu 385	. Ala	Ser	Pro	Gly	Val 390	Thr	Ala	Leu	Ser	Asp 395	Glu	Tyr	Phe	Pro	Asn 400	
Ala	Thr	Arg	Trp	Asp 405	Pro	His	Arg	Trp	Glu 410	Asn	Gln	Pro	Asp	Lys 415	Glu	
Glu	. Asp	Gly	Glu 420	Met	Val	Asp	Tyr	Gly 425	Tyr	Gly	Ser	Val	Ser 430	Lys	Gly	
Thr	Ala	Ser 435	Pro	Tyr	Leu	Pro	Phe 440	Gly	Ala	Gly	Arg	His 445	Arg	Суз	Ile	
Gly	Glu 450	Lys	Phe	Ala	Tyr	Val 455	Asn	Leu	Gly	Val	Ile 460	Ile	Ala	Thr	Ile	
Val 465	Arg	His	Leu	Гла	Leu 470	Phe	Asn	Val	Asp	Gly 475	Arg	ГЛЗ	Gly	Val	Pro 480	
Glγ	Thr	Asp	Tyr	Ser 485	Thr	Leu	Phe	Ser	Gly 490	Pro	Met	ГЛЗ	Pro	Ala 495	Ile	
Val	Gly	Trp	Glu 500	Arg	Arg	Phe	Pro	Asp 505	Asn	Ile	Гла	Gly	Ser 510	Met	Asn	
<210> SEQ ID NO 93 <211> LENGTH: 22 <212> TYPE: DNA <213> ORGANISM: Artificial sequence <220> FEATURE: <223> OTHER INFORMATION: Primer BP460																
<40	0> SI	EQUEI	NCE :	93												
agg	aggattatca ttcataagtt tc															
<21 <21 <21 <22 <22 <22	<210> SEQ ID NO 94 <211> LENGTH: 20 <212> TYPE: DNA <213> ORGANISM: Artificial sequence <220> FEATURE: <223> OTHER INFORMATION: Primer LA135															
		~ ~ ~ ~														

-continued

cttggcagca acaggactag		20
<210> SEQ ID NO 95 <211> LENGTH: 23 <212> TYPE: DNA <213> ORGANISM: Artificial sequence <220> FEATURE: <223> OTHER INFORMATION: Primer BP461		
<400> SEQUENCE: 95		
ttettggage tgggacatgt ttg		23
<210> SEQ ID NO 96 <211> LENGTH: 22 <212> TYPE: DNA <213> ORGANISM: Artificial sequence <220> FEATURE: <223> OTHER INFORMATION: Primer LA92		
<400> SEQUENCE: 96		
gagaagatgc ggccagcaaa ac		22
<210> SEQ ID NO 97 <211> LENGTH: 4242 <212> TYPE: DNA <213> ORGANISM: Artificial <220> FEATURE: <223> OTHER INFORMATION: pLA59		
<400> SEQUENCE: 97		
aaacgccagc aacgcggcct ttttacggtt ccto	gcettt tgetggeett ttgeteacat	60
gttettteet gegttateee etgattetgt ggat	aaccgt attaccgcct ttgagtgagc	2 120
tgataccgct cgccgcagcc gaacgaccga gcgo	agcgag tcagtgagcg aggaagcgga	a 180
agagegeeca ataegeaaae egeeteteee ege	cgttgg ccgattcatt aatgcagctg	g 240
gcacgacagg tttcccgact ggaaagcggg cagt	gagcgc aacgcaatta atgtgagtta	a 300
gctcactcat taggcacccc aggctttaca cttt	atgett ceggetegta tgttgtgtgg	360
aattgtgagc ggataacaat ttcacacagg aaac	agctat gaccatgatt acgccaagct	420
tgcatgcctg caggtcgact ctagaggatc cgca	atgcgg atccgcattg cggattacgt	480
attetaatgt teagtacegt tegtataatg tate	ctatac gaagttatgc agattgtact	540
gagagtgcac cataccacct tttcaattca tcat	tttttt tttattcttt tttttgattt	600
cggtttcctt gaaatttttt tgattcggta atct	ccgaac agaaggaaga acgaaggaag	J 660
gagcacagac ttagattggt atatatacgc atat	gtagtg ttgaagaaac atgaaattgc	2 720
ccagtattet taacceaact geacagaaca aaaa	octgca ggaaacgaag ataaatcatg	780
tcgaaagcta catataagga acgtgctgct acto	atccta gtcctgttgc tgccaagcta	a 840
tttaatatca tgcacgaaaa gcaaacaaac ttg	gtgett cattggatgt tegtaceaee	900
aaggaattac tggagttagt tgaagcatta ggto	ccaaaa tttgtttact aaaaacacat	960
gtggatatet tgaetgattt tteeatggag ggea	cagtta agccgctaaa ggcattatcc	2 1020
gccaagtaca attttttact cttcgaagac aqaa	aatttg ctgacattgg taatacaqtc	e 1080
aaattqcaqt actctqcqqq tqtatacaqa ataq	caqaat qqqcaqacat tacqaatqca	a 1140
cacqqtqtqq tqqqcccaqq tattattaqc qqt	tqaaqc aqqcqqcaqa aqaaqtaaca	a 1200
aaddaaccta gaddetttt datdttade daat	tatcat gcaaggatta cotatotaat	1260
	agagg agaaggett tettet	1220
yyagaatata ctaagggtac tgttgacatt gcga	ayageg acaaagattt tgttategge	; 13∠U

293

#### -continued

tttattgctc aaagagacat gggtggaaga gatgaaggtt acgattggtt gattatgaca 1380 cccggtgtgg gtttagatga caagggagac gcattgggtc aacagtatag aaccgtggat 1440 gatgtggtct ctacaggatc tgacattatt attgttggaa gaggactatt tgcaaaggga 1500 agggatgcta aggtagaggg tgaacgttac agaaaagcag gctgggaagc atatttgaga 1560 agatgcggcc agcaaaacta aaaaactgta ttataagtaa atgcatgtat actaaactca 1620 1680 caaattagag cttcaattta attatatcag ttattaccct atgcggtgtg aaataccgca 1740 cagatgcgta aggagaaaat accgcatcag gaaattgtaa acgttaatat tttgttaaaa ttcgcgttaa atttttgtta aatcagctca ttttttaacc aataggccga aatcggcaaa 1800 atcccttata aatcaaaaga atagaccgag atagggttga gtgttgttcc agtttggaac 1860 aagagteeac tattaaagaa egtggaetee aaegteaaag ggegaaaaae egtetateag 1920 ggcgatggcc cactacgtga accatcaccc taatcaagat aacttcgtat aatgtatgct 1980 atacgaacgg taccagtgat gatacaacga gttagccaag gtgaattcac tggccgtcgt 2040 tttacaacgt cgtgactggg aaaaccetgg cgttaceeaa ettaategee ttgeageaca 2100 2160 teccecttte gecagetgge gtaatagega agaggeeege acegategee etteccaaca gttgcgcagc ctgaatggcg aatggcgcct gatgcggtat tttctcctta cgcatctgtg 2220 cggtatttca caccgcatat ggtgcactct cagtacaatc tgctctgatg ccgcatagtt 2280 2340 aagccagccc cgacacccgc caacacccgc tgacgcgccc tgacgggctt gtctgctccc ggcatccgct tacagacaag ctgtgaccgt ctccgggagc tgcatgtgtc agaggttttc 2400 accgtcatca ccgaaacgcg cgagacgaaa gggcctcgtg atacgcctat ttttataggt 2460 taatgtcatg ataataatgg tttcttagac gtcaggtggc acttttcggg gaaatgtgcg 2520 cqqaacccct atttqtttat ttttctaaat acattcaaat atqtatccqc tcatqaqaca 2580 ataaccctga taaatgcttc aataatattg aaaaaggaag agtatgagta ttcaacattt 2640 ccgtgtcgcc cttattccct tttttgcggc attttgcctt cctgtttttg ctcacccaga 2700 aacgctggtg aaagtaaaag atgctgaaga tcagttgggt gcacgagtgg gttacatcga 2760 actggatete aacageggta agateettga gagttttege eeegaagaae gtttteeaat 2820 gatgagcact tttaaagttc tgctatgtgg cgcggtatta tcccgtattg acgccgggca 2880 agagcaactc ggtcgccgca tacactattc tcagaatgac ttggttgagt actcaccagt 2940 cacagaaaaag catcttacgg atggcatgac agtaagagaa ttatgcagtg ctgccataac 3000 catgagtgat aacactgcgg ccaacttact tctgacaacg atcggaggac cgaaggagct 3060 aaccgctttt ttgcacaaca tggggggatca tgtaactcgc cttgatcgtt gggaaccgga 3120 3180 gctgaatgaa gccataccaa acgacgagcg tgacaccacg atgcctgtag caatggcaac aacgttgcgc aaactattaa ctggcgaact acttactcta gcttcccggc aacaattaat 3240 agactggatg gaggeggata aagttgeagg aceaettetg egeteggeee tteeggetgg 3300 ctggtttatt gctgataaat ctggagccgg tgagcgtggg tctcgcggta tcattgcagc 3360 actgggggcca gatggtaagc cctcccgtat cgtagttatc tacacgacgg ggagtcaggc 3420 aactatggat gaacgaaata gacagatcgc tgagataggt gcctcactga ttaagcattg 3480 gtaactgtca gaccaagttt actcatatat actttagatt gatttaaaac ttcatttta 3540 atttaaaagg atctaggtga agatcctttt tgataatctc atgaccaaaa tcccttaacg 3600

tgagttttcg ttccactgag cgtcagaccc cgtagaaaag atcaaaggat cttcttgaga

3660

295

### continued

-concinded	
teetttttt etgegegtaa tetgetgett geaaacaaaa aaaceaeege taeeageggt	3720
ggtttgtttg ccggatcaag agctaccaac tctttttccg aaggtaactg gcttcagcag	3780
agegeagata ceaaataetg teettetagt gtageegtag ttaggeeace aetteaagaa	3840
ctctgtagca ccgcctacat acctcgctct gctaatcctg ttaccagtgg ctgctgccag	3900
tggcgataag tcgtgtctta ccgggttgga ctcaagacga tagttaccgg ataaggcgca	3960
geggteggge tgaaeggggg gttegtgeae acageeeage ttggagegaa egaeetaeae	4020
cgaactgaga tacctacagc gtgagctatg agaaagcgcc acgcttcccg aagggagaaa	4080
ggcggacagg tatccggtaa gcggcagggt cggaacagga gagcgcacga gggagcttcc	4140
agggggaaac geetggtate tttatagtee tgtegggttt egeeaeetet gaettgageg	4200
tcgatttttg tgatgctcgt caggggggcg gagcctatgg aa	4242
<210> SEQ ID NO 98 <211> LENGTH: 80 <212> TYPE: DNA <213> ORGANISM: Artificial <220> FEATURE: <223> OTHER INFORMATION: LA678	
<400> SEQUENCE: 98	
caacgttaac accgttttcg gtttgccagg tgacttcaac ttgtccttgt gcattgcgga	60
ttacgtattc taatgttcag	80
<210> SEQ ID NO 99 <211> LENGTH: 81 <212> TYPE: DNA <213> ORGANISM: Artificial <220> FEATURE: <223> OTHER INFORMATION: LA679 <400> SEQUENCE: 99	
gtggagcatc gaagactggc aacatgattt caatcattct gatcttagag caccttggct	60
aactcgttgt atcatcactg g	81
<210> SEQ ID NO 100 <211> LENGTH: 4586 <212> TYPE: DNA <213> ORGANISM: Artificial sequence <220> FEATURE: <223> OTHER INFORMATION: pLA54 template DNA	
<400> SEQUENCE: 100	
gggtaccgag ctcgaattca ctggccgtcg ttttacaacg tcgtgactgg gaaaaccctg	60
gcgttaccca acttaatcgc cttgcagcac atcccccttt cgccagctgg cgtaatagcg	120
aagaggcccg caccgatcgc ccttcccaac agttgcgcag cctgaatggc gaatggcgcc	180
tgatgeggta tttteteett aegeatetgt geggtattte aeaeegeata tggtgeaete	240
tcagtacaat ctgctctgat gccgcatagt taagccagcc ccgacacccg ccaacacccg	300
ctgacgcgcc ctgacgggct tgtctgctcc cggcatccgc ttacagacaa gctgtgaccg	360
tctccgggag ctgcatgtgt cagaggtttt caccgtcatc accgaaacgc gcgagacgaa	420
agggcotogt gatacgoota tttttatagg ttaatgtoat gataataatg gtttottaga	480
cgtcaggtgg cacttttcgg ggaaatgtgc gcggaacccc tatttgttta tttttctaaa	540
tacattcaaa tatgtatccg ctcatgagac aataaccctg ataaatgctt caataatatt	600
gaaaaaggaa gagtatgagt attcaacatt tccgtgtcgc ccttattccc ttttttgcgg	660

297

### -continued

cattttgcct	tcctgttttt	gctcacccag	aaacgctggt	gaaagtaaaa	gatgctgaag	720
atcagttggg	tgcacgagtg	ggttacatcg	aactggatct	caacagcggt	aagatccttg	780
agagttttcg	ccccgaagaa	cgttttccaa	tgatgagcac	ttttaaagtt	ctgctatgtg	840
gcgcggtatt	atcccgtatt	gacgccgggc	aagagcaact	cggtcgccgc	atacactatt	900
ctcagaatga	cttggttgag	tactcaccag	tcacagaaaa	gcatcttacg	gatggcatga	960
cagtaagaga	attatgcagt	gctgccataa	ccatgagtga	taacactgcg	gccaacttac	1020
ttctgacaac	gatcggagga	ccgaaggagc	taaccgcttt	tttgcacaac	atgggggatc	1080
atgtaactcg	ccttgatcgt	tgggaaccgg	agctgaatga	agccatacca	aacgacgagc	1140
gtgacaccac	gatgcctgta	gcaatggcaa	caacgttgcg	caaactatta	actggcgaac	1200
tacttactct	agcttcccgg	caacaattaa	tagactggat	ggaggcggat	aaagttgcag	1260
gaccacttct	gcgctcggcc	cttccggctg	gctggtttat	tgctgataaa	tctggagccg	1320
gtgagcgtgg	gtctcgcggt	atcattgcag	cactgggggcc	agatggtaag	ccctcccgta	1380
tcgtagttat	ctacacgacg	gggagtcagg	caactatgga	tgaacgaaat	agacagatcg	1440
ctgagatagg	tgcctcactg	attaagcatt	ggtaactgtc	agaccaagtt	tactcatata	1500
tactttagat	tgatttaaaa	cttcattttt	aatttaaaag	gatctaggtg	aagatccttt	1560
ttgataatct	catgaccaaa	atcccttaac	gtgagttttc	gttccactga	gcgtcagacc	1620
ccgtagaaaa	gatcaaagga	tcttcttgag	atcctttttt	tctgcgcgta	atctgctgct	1680
tgcaaacaaa	aaaaccaccg	ctaccagcgg	tggtttgttt	gccggatcaa	gagctaccaa	1740
ctctttttcc	gaaggtaact	ggcttcagca	gagcgcagat	accaaatact	gtccttctag	1800
tgtagccgta	gttaggccac	cacttcaaga	actctgtagc	accgcctaca	tacctcgctc	1860
tgctaatcct	gttaccagtg	gctgctgcca	gtggcgataa	gtcgtgtctt	accgggttgg	1920
actcaagacg	atagttaccg	gataaggcgc	agcggtcggg	ctgaacgggg	ggttcgtgca	1980
cacagcccag	cttggagcga	acgacctaca	ccgaactgag	atacctacag	cgtgagctat	2040
gagaaagcgc	cacgcttccc	gaagggagaa	aggcggacag	gtatccggta	agcggcaggg	2100
tcggaacagg	agagcgcacg	agggagcttc	cagggggaaa	cgcctggtat	ctttatagtc	2160
ctgtcgggtt	tcgccacctc	tgacttgagc	gtcgatttt	gtgatgctcg	tcagggggggc	2220
ggagcctatg	gaaaaacgcc	agcaacgcgg	cctttttacg	gttcctggcc	ttttgctggc	2280
cttttgctca	catgttcttt	cctgcgttat	cccctgattc	tgtggataac	cgtattaccg	2340
cctttgagtg	agctgatacc	gctcgccgca	gccgaacgac	cgagcgcagc	gagtcagtga	2400
gcgaggaagc	ggaagagcgc	ccaatacgca	aaccgcctct	ccccgcgcgt	tggccgattc	2460
attaatgcag	ctggcacgac	aggtttcccg	actggaaagc	gggcagtgag	cgcaacgcaa	2520
ttaatgtgag	ttagctcact	cattaggcac	cccaggcttt	acactttatg	cttccggctc	2580
gtatgttgtg	tggaattgtg	agcggataac	aatttcacac	aggaaacagc	tatgaccatg	2640
attacgccaa	gcttgcatgc	ctgcaggtcg	actctagagg	atccccgcat	tgcggattac	2700
gtattctaat	gttcagataa	cttcgtatag	catacattat	acgaagttat	ctagggattc	2760
ataaccattt	tctcaatcga	attacacaga	acacaccgta	caaacctctc	tatcataact	2820
acttaatagt	cacacacgta	ctcgtctaaa	tacacatcat	cgtcctacaa	gttcatcaaa	2880
gtgttggaca	gacaactata	ccagcatgga	tctcttgtat	cggttctttt	ctcccgctct	2940
ctcgcaataa	caatgaacac	tgggtcaatc	atagcctaca	caggtgaaca	gagtagcgtt	3000

299

tatacagggt ttatacggtg attectacgg caaaaatttt teatttetaa aaaaaaaaag

### -continued

aaaaattttt ctttccaacg ctagaaggaa aagaaaaatc taattaaatt gatttggtga 3120 ttttctgaga gttccctttt tcatatatcg aattttgaat ataaaaggag atcgaaaaaa 3180 tttttctatt caatctgttt tctggtttta tttgatagtt tttttgtgta ttattattat 3240 ggattagtac tggtttatat gggtttttct gtataacttc tttttatttt agtttgttta 3300 atettatttt gagttaeatt atagtteeet aaetgeaaga gaagtaaeat taaaaetega 3360 gatgggtaag gaaaagactc acgtttcgag gccgcgatta aattccaaca tggatgctga 3420 tttatatggg tataaatggg ctcgcgataa tgtcgggcaa tcaggtgcga caatctatcg 3480 attgtatggg aagcccgatg cgccagagtt gtttctgaaa catggcaaag gtagcgttgc 3540 caatgatgtt acagatgaga tggtcagact aaactggctg acggaattta tgcctcttcc 3600 gaccatcaag cattttatcc gtactcctga tgatgcatgg ttactcacca ctgcgatccc 3660 cqqcaaaaca qcattccaqq tattaqaaqa atatcctqat tcaqqtqaaa atattqttqa 3720 tgcgctggca gtgttcctgc gccggttgca ttcgattcct gtttgtaatt gtccttttaa 3780 cagcgatcgc gtatttcgtc tcgctcaggc gcaatcacga atgaataacg gtttggttga 3840 tgcgagtgat tttgatgacg agcgtaatgg ctggcctgtt gaacaagtct ggaaagaaat 3900 gcataagctt ttgccattct caccggattc agtcgtcact catggtgatt tctcacttga 3960 taaccttatt tttqacqaqq qqaaattaat aqqttqtatt qatqttqqac qaqtcqqaat 4020 cgcagaccga taccaggatc ttgccatcct atggaactgc ctcggtgagt tttctccttc 4080 attacagaaa cggctttttc aaaaatatgg tattgataat cctgatatga ataaattgca 4140 gtttcatttg atgctcgatg agtttttcta agtttaactt gatactacta gattttttct 4200 cttcatttat aaaatttttg gttataattg aagctttaga agtatgaaaa aatccttttt 4260 tttcattctt tgcaaccaaa ataagaagct tcttttattc attgaaatga tgaatataaa 4320 cctaacaaaa gaaaaagact cgaatatcaa acattaaaaa aaaataaaaag aggttatctg 4380 ttttcccatt tagttggagt ttgcattttc taatagatag aactctcaat taatgtggat 4440ttagtttctc tgttcgtttt tttttgtttt gttctcactg tatttacatt tctatttagt 4500 atttagttat tcatataatc tataacttcg tatagcatac attatacgaa gttatccagt 4560 gatgatacaa cgagttagcc aaggtg 4586 <210> SEQ ID NO 101 <211> LENGTH: 80 <212> TYPE: DNA <213> ORGANISM: Artificial <220> FEATURE: <223> OTHER INFORMATION: BK505 primer <400> SEQUENCE: 101 ttccqqtttc tttqaaattt ttttqattcq qtaatctccq aqcaqaaqqa qcattqcqqa 60 ttacqtattc taatqttcaq 80 <210> SEQ ID NO 102 <211> LENGTH: 81 <212> TYPE: DNA <213> ORGANISM: Artificial <220> FEATURE: <223> OTHER INFORMATION: BK506 primer <400> SEQUENCE: 102

#### gggtaataac tgatataatt aaattgaagc tctaatttgt gagtttagta caccttggct 60

300

301

-continued

aactcgttgt atcatcactg g 81	
<210> SEQ ID NO 103 <211> LENGTH: 38 <212> TYPE: DNA <213> ORGANISM: Artificial sequence <220> FEATURE:	
<223> OTHER INFORMATION: BK468 Primer	
<400> SEQUENCE: 103	
gcetegagtt ttaatgttae ttetettgea gttaggga 38	
<pre>&lt;210&gt; SEQ ID NO 104 &lt;211&gt; LENGTH: 31 &lt;212&gt; TYPE: DNA &lt;213&gt; ORGANISM: Artificial &lt;220&gt; FEATURE: &lt;223&gt; OTHER INFORMATION: LA492 primer</pre>	
<400> SEQUENCE: 104	
gctaaattcg agtgaaacac aggaagacca g 31	
<210> SEQ ID NO 105 <211> LENGTH: 23 <212> TYPE: DNA <213> ORGANISM: Artificial <220> FEATURE: <223> OTHER INFORMATION: AK109-1	
<400> SEQUENCE: 105	
agtcacatca agatcgttta tgg 23	
<210> SEQ ID NO 106 <211> LENGTH: 23 <212> TYPE: DNA <213> ORGANISM: Artificial <220> FEATURE: <223> OTHER INFORMATION: AK109-2	
<400> SEQUENCE: 106	
gcacggaata tgggactact tcg 23	
<210> SEQ ID NO 107 <211> LENGTH: 23 <212> TYPE: DNA <213> ORGANISM: Artificial <220> FEATURE: <223> OTHER INFORMATION: AK109-3	
<400> SEQUENCE: 107	
actccacttc aagtaagagt ttg 23	
<210> SEQ ID NO 108 <211> LENGTH: 548 <212> TYPE: PRT <213> ORGANISM: Lactococcus lactis	
<400> SEQUENCE: 108	
Met Tyr Thr Val Gly Asp Tyr Leu Leu Asp Arg Leu His Glu Leu Gly 1 5 10 15	
Ile Glu Glu Ile Phe Gly Val Pro Gly Asp Tyr Asn Leu Gln Phe Leu 20 25 30	
Asp Gln Ile Ile Ser His Lys Asp Met Lys Trp Val Gly Asn Ala Asn 35 40 45	

Glu	Leu	Asn	Ala	Ser	Tyr	Met	Ala	Asp	Gly	Tyr	Ala	Arg	Thr	Lys	Lys
Ala	5U Ala	Ala	Phe	Leu	Thr	55 Thr	Phe	Glv	Val	Glv	60 Glu	Leu	Ser	Ala	Val
65				204	70			017		75	014	Doa	201		80
Asn	Gly	Leu	Ala	Gly 85	Ser	Tyr	Ala	Glu	Asn 90	Leu	Pro	Val	Val	Glu 95	Ile
Val	Gly	Ser	Pro 100	Thr	Ser	Lys	Val	Gln 105	Asn	Glu	Gly	LYa	Phe 110	Val	His
His	Thr	Leu 115	Ala	Asp	Gly	Asp	Phe 120	Lys	His	Phe	Met	Lys 125	Met	His	Glu
Pro	Val 130	Thr	Ala	Ala	Arg	Thr 135	Leu	Leu	Thr	Ala	Glu 140	Asn	Ala	Thr	Val
Glu 145	Ile	Asp	Arg	Val	Leu 150	Ser	Ala	Leu	Leu	Lys 155	Glu	Arg	Lys	Pro	Val 160
Tyr	Ile	Asn	Leu	Pro 165	Val	Asp	Val	Ala	Ala 170	Ala	Lys	Ala	Glu	Lys 175	Pro
Ser	Leu	Pro	Leu 180	ГЛа	Lys	Glu	Asn	Ser 185	Thr	Ser	Asn	Thr	Ser 190	Asp	Gln
Glu	Ile	Leu 195	Asn	ГÀа	Ile	Gln	Glu 200	Ser	Leu	Lys	Asn	Ala 205	Lys	Lys	Pro
Ile	Val 210	Ile	Thr	Gly	His	Glu 215	Ile	Ile	Ser	Phe	Gly 220	Leu	Glu	Lys	Thr
Val 225	Thr	Gln	Phe	Ile	Ser 230	Lys	Thr	Lys	Leu	Pro 235	Ile	Thr	Thr	Leu	Asn 240
Phe	Gly	Lys	Ser	Ser 245	Val	Asp	Glu	Ala	Leu 250	Pro	Ser	Phe	Leu	Gly 255	Ile
Tyr	Asn	Gly	Thr 260	Leu	Ser	Glu	Pro	Asn 265	Leu	Lys	Glu	Phe	Val 270	Glu	Ser
Ala	Asp	Phe 275	Ile	Leu	Met	Leu	Gly 280	Val	Lys	Leu	Thr	Asp 285	Ser	Ser	Thr
Gly	Ala 290	Phe	Thr	His	His	Leu 295	Asn	Glu	Asn	Lys	Met 300	Ile	Ser	Leu	Asn
Ile 305	Aab	Glu	Gly	Lys	Ile 310	Phe	Asn	Glu	Arg	Ile 315	Gln	Asn	Phe	Aab	Phe 320
Glu	Ser	Leu	Ile	Ser 325	Ser	Leu	Leu	Asp	Leu 330	Ser	Glu	Ile	Glu	Tyr 335	Lys
Gly	Lys	Tyr	Ile 340	Asp	LÀa	LYa	Gln	Glu 345	Asp	Phe	Val	Pro	Ser 350	Asn	Ala
Leu	Leu	Ser 355	Gln	Asp	Arg	Leu	Trp 360	Gln	Ala	Val	Glu	Asn 365	Leu	Thr	Gln
Ser	Asn 370	Glu	Thr	Ile	Val	Ala 375	Glu	Gln	Gly	Thr	Ser 380	Phe	Phe	Gly	Ala
Ser 385	Ser	Ile	Phe	Leu	Lys 390	Ser	Lys	Ser	His	Phe 395	Ile	Gly	Gln	Pro	Leu 400
Trp	Gly	Ser	Ile	Gly 405	Tyr	Thr	Phe	Pro	Ala 410	Ala	Leu	Gly	Ser	Gln 415	Ile
Ala	Asp	Lys	Glu 420	Ser	Arg	His	Leu	Leu 425	Phe	Ile	Gly	Asp	Gly 430	Ser	Leu
Gln	Leu	Thr 435	Val	Gln	Glu	Leu	Gly 440	Leu	Ala	Ile	Arg	Glu 445	Lys	Ile	Asn
Pro	Ile 450	Cys	Phe	Ile	Ile	Asn 455	Asn	Asp	Gly	Tyr	Thr 460	Val	Glu	Arg	Glu

### -continued

Ile His Gly Pro Asn Gln Ser Tyr Asn Asp Ile Pro Met Trp Asn Tyr465470475480
Ser Lys Leu Pro Glu Ser Phe Gly Ala Thr Glu Asp Arg Val Val Ser 485 490 495
Lys Ile Val Arg Thr Glu Asn Glu Phe Val Ser Val Met Lys Glu Ala 500 505 510
Gln Ala Asp Pro Asn Arg Met Tyr Trp Ile Glu Leu Ile Leu Ala Lys 515 520 525
Glu Gly Ala Pro Lys Val Leu Lys Lys Met Gly Lys Leu Phe Ala Glu 530 535 540
Gln Asn Lys Ser 545
<210> SEQ ID NO 109 <211> LENGTH: 49 <212> TYPE: DNA <213> ORGANISM: Artificial sequence <220> FEATURE: <223> OTHER INFORMATION: Primer oBP453
<400> SEQUENCE: 109
tgcagettta aataateggt gtcactaett tgeettegtt tatettgee 49
<pre>&lt;210&gt; SEQ ID NO 110 &lt;211&gt; LENGTH: 49 &lt;212&gt; TYPE: DNA &lt;213&gt; ORGANISM: Artificial sequence &lt;220&gt; FEATURE: &lt;223&gt; OTHER INFORMATION: Primer oBP454 &lt;400&gt; SEQUENCE: 110</pre>
gagcaggcaa gataaacgaa ggcaaagtag tgacaccgat tatttaaag 49
<210> SEQ ID NO 111 <211> LENGTH: 23 <212> TYPE: DNA <213> ORGANISM: Artificial sequence <220> FEATURE: <223> OTHER INFORMATION: Primer LA337
<400> SEQUENCE: 111
ctcatttgaa tcagcttatg gtg 23
<210> SEQ ID NO 112 <211> LENGTH: 24 <212> TYPE: DNA <213> ORGANISM: Artificial sequence <220> FEATURE: <223> OTHER INFORMATION: LA692
<400> SEQUENCE: 112
ggaagtcatt gacaccatct tggc 24
<210> SEQ ID NO 113 <211> LENGTH: 24 <212> TYPE: DNA <213> ORGANISM: Artificial sequence <220> FEATURE: <223> OTHER INFORMATION: Primer LA693
<400> SEQUENCE: 113
agaagctggg acagcagcgt tagc 24

aont	4	nuad	

-continued
<pre><li>&lt;210&gt; SEQ ID NO 114 &lt;211&gt; LENGTH: 41 &lt;212&gt; TYPE: DNA &lt;213&gt; ORGANISM: Artificial sequence &lt;220&gt; FEATURE:</li></pre>
<223> OTHER INFORMATION: Primer LA684
<400> SEQUENCE: 114
aatcactgca tgccttccaa aacacgaaca aggtgccggt c 41
<210> SEQ ID NO 115 <211> LENGTH: 45 <212> TYPE: DNA <213> ORGANISM: Artificial sequence <220> FEATURE: <223> OTHER INFORMATION: Primer LA685
<400> SEQUENCE: 115
ttaagtagga tcccacttga attgaactta ttattcatct atgac 45
<210> SEQ ID NO 116 <211> LENGTH: 38 <212> TYPE: DNA <213> ORGANISM: Artificial sequence <220> FEATURE: <223> OTHER INFORMATION: Primer LA682
<400> SEQUENCE: 116
gtetttaeag ggeaagtete aaetagtget ateggtae 38
<210> SEQ ID NO 117 <211> LENGTH: 38 <212> TYPE: DNA <213> ORGANISM: Artificial sequence <220> FEATURE: <223> OTHER INFORMATION: Primer LA683
<400> SEQUENCE: 117
gtaccgatag cactagttga gacttgccct gtaaagac 38
<210> SEQ ID NO 118 <211> LENGTH: 546 <212> TYPE: PRT <213> ORGANISM: Macrococcus caseolyticus
<400> SEQUENCE: 118
Met Lys Gln Arg Ile Gly Gln Tyr Leu Ile Asp Ala Leu His Val Asn 1 5 10 15
Gly Val Asp Lys Ile Phe Gly Val Pro Gly Asp Phe Thr Leu Ala Phe 20 25 30
Leu Asp Asp Ile Ile Arg His Asp Asn Val Glu Trp Val Gly Asn Thr 35 40 45
Asn Glu Leu Asn Ala Ala Tyr Ala Ala Asp Gly Tyr Ala Arg Val Asn 50 55 60
Gly Leu Ala Ala Val Ser Thr Thr Phe Gly Val Gly Glu Leu Ser Ala 65 70 75 80
Val Asn Gly Ile Ala Gly Ser Tyr Ala Glu Arg Val Pro Val Ile Lys 85 90 95
Ile Ser Gly Gly Pro Ser Ser Val Ala Gln Gln Glu Gly Arg Tyr Val
His His Ser Leu Gly Glu Gly Ile Phe Asp Ser Tyr Ser Lys Met Tyr
120 120

-continued

309

310

Ala	His 130	Ile	Thr	Ala	Thr	Thr 135	Thr	Ile	Leu	Ser	Val 140	Asp	Asn	Ala	Val
Asp 145	Glu	Ile	Asp	Arg	Val 150	Ile	His	Суз	Ala	Leu 155	Lys	Glu	Lys	Arg	Pro 160
Val	His	Ile	His	Leu 165	Pro	Ile	Asp	Val	Ala 170	Leu	Thr	Glu	Ile	Glu 175	Ile
Pro	His	Ala	Pro 180	Lys	Val	Tyr	Thr	His 185	Glu	Ser	Gln	Asn	Val 190	Asp	Ala
Tyr	Ile	Gln 195	Ala	Val	Glu	Гла	Lys 200	Leu	Met	Ser	Ala	Lys 205	Gln	Pro	Val
Ile	Ile 210	Ala	Gly	His	Glu	Ile 215	Asn	Ser	Phe	Lys	Leu 220	His	Glu	Gln	Leu
Glu 225	Gln	Phe	Val	Asn	Gln 230	Thr	Asn	Ile	Pro	Val 235	Ala	Gln	Leu	Ser	Leu 240
Gly	Lys	Ser	Ala	Phe 245	Asn	Glu	Glu	Asn	Glu 250	His	Tyr	Leu	Gly	Ile 255	Tyr
Asp	Gly	Lys	Ile 260	Ala	Lys	Glu	Asn	Val 265	Arg	Glu	Tyr	Val	Asp 270	Asn	Ala
Asp	Val	Ile 275	Leu	Asn	Ile	Gly	Ala 280	Lys	Leu	Thr	Asp	Ser 285	Ala	Thr	Ala
Gly	Phe 290	Ser	Tyr	Lys	Phe	Asp 295	Thr	Asn	Asn	Ile	Ile 300	Tyr	Ile	Asn	His
Asn 305	Aab	Phe	Lys	Ala	Glu 310	Asp	Val	Ile	Ser	Asp 315	Asn	Val	Ser	Leu	Ile 320
Asp	Leu	Val	Asn	Gly 325	Leu	Asn	Ser	Ile	Asp 330	Tyr	Arg	Asn	Glu	Thr 335	His
Tyr	Pro	Ser	Tyr 340	Gln	Arg	Ser	Asp	Met 345	Lys	Tyr	Glu	Leu	Asn 350	Asp	Ala
Pro	Leu	Thr 355	Gln	Ser	Asn	Tyr	Phe 360	Lys	Met	Met	Asn	Ala 365	Phe	Leu	Glu
ГЛа	Asp 370	Asp	Ile	Leu	Leu	Ala 375	Glu	Gln	Gly	Thr	Ser 380	Phe	Phe	Gly	Ala
Tyr 385	Asp	Leu	Ser	Leu	Tyr 390	ГЛа	Gly	Asn	Gln	Phe 395	Ile	Gly	Gln	Pro	Leu 400
Trp	Gly	Ser	Ile	Gly 405	Tyr	Thr	Phe	Pro	Ser 410	Leu	Leu	Gly	Ser	Gln 415	Leu
Ala	Asp	Met	His 420	Arg	Arg	Asn	Ile	Leu 425	Leu	Ile	Gly	Asp	Gly 430	Ser	Leu
Gln	Leu	Thr 435	Val	Gln	Ala	Leu	Ser 440	Thr	Met	Ile	Arg	Lys 445	Asp	Ile	Lys
Pro	Ile 450	Ile	Phe	Val	Ile	Asn 455	Asn	Asp	Gly	Tyr	Thr 460	Val	Glu	Arg	Leu
Ile 465	His	Gly	Met	Glu	Glu 470	Pro	Tyr	Asn	Asp	Ile 475	Gln	Met	Trp	Asn	Tyr 480
ГЛа	Gln	Leu	Pro	Glu 485	Val	Phe	Gly	Gly	Lys 490	Asp	Thr	Val	Гла	Val 495	His
Asp	Ala	Lys	Thr 500	Ser	Asn	Glu	Leu	Lys 505	Thr	Val	Met	Asp	Ser 510	Val	Lys
Ala	Asp	Lys 515	Asp	His	Met	His	Phe 520	Ile	Glu	Val	His	Met 525	Ala	Val	Glu
Asp	Ala 530	Pro	Lys	Lys	Leu	Ile 535	Asp	Ile	Ala	Lys	Ala 540	Phe	Ser	Asp	Ala

Asn Lys

-continued

312

545		
<210> SEQ ID NO 119 <211> LENGTH: 28 <212> TYPE: DNA <213> ORGANISM: Artificial sequence <220> FEATURE: <223> OTHER INFORMATION: LA686		
<400> SEQUENCE: 119		
cttccaaaac acgaacaagg tgccggtc	28	
<210> SEQ ID NO 120 <211> LENGTH: 97 <212> TYPE: DNA <213> ORGANISM: Artificial sequence <220> FEATURE: <223> OTHER INFORMATION: LA687		
<400> SEQUENCE: 120		
gttaaaacac accatttgaa tacatatgct acgtatccac tttagattta tcaaatacta	60	
ccaactcacc ttggctaact cgttgtatca tcactgg	97	
<210> SEQ ID NO 121 <211> LENGTH: 7555 <212> TYPE: DNA <213> ORGANISM: Artificial <220> FEATURE: <223> OTHER INFORMATION: pRS423::PGAL1-cre		
<400> SEQUENCE: 121		
tcgcgcgttt cggtgatgac ggtgaaaacc tctgacacat gcagctcccg gagacggtca	60	
cagettgtet gtaageggat geegggagea gacaageeeg teagggegeg teagegggtg	120	
ttggcgggtg tcggggctgg cttaactatg cggcatcaga gcagattgta ctgagagtgc	180	
accataaatt cccgttttaa gagcttggtg agcgctagga gtcactgcca ggtatcgttt	240	
gaacacggca ttagtcaggg aagtcataac acagtccttt cccgcaattt tcttttcta	300	
ttactcttgg cctcctctag tacactctat atttttttat gcctcggtaa tgattttcat	360	
ttttttttt cccctagogg atgactettt ttttttetta gogattggea ttateacata	420	
atgaattata cattatataa agtaatgtga tttcttcgaa gaatatacta aaaaatgagc	480	
aggcaagata aacgaaggca aagatgacag agcagaaagc cctagtaaag cgtattacaa	540	
atgaaaccaa gattcagatt gcgatctctt taaagggtgg tcccctagcg atagagcact	600	
cgatcttccc agaaaaagag gcagaagcag tagcagaaca ggccacacaa tcgcaagtga	660	
ttaacgtcca cacaggtata gggtttctgg accatatgat acatgctctg gccaagcatt	720	
ccggctggtc gctaatcgtt gagtgcattg gtgacttaca catagacgac catcacacca	780	
ctgaagactg cgggattgct ctcggtcaag cttttaaaga ggccctactg gcgcgtggag	840	
taaaaaggtt tggatcagga tttgcgcctt tggatgaggc actttccaga gcggtggtag	900	
atctttcgaa caggccgtac gcagttgtcg aacttggttt gcaaagggag aaagtaggag	960	
atctctcttg cgagatgatc ccgcattttc ttgaaagctt tgcagaggct agcagaatta	1020	
ccctccacgt tgattgtctg cgaggcaaga atgatcatca ccgtagtgag agtgcgttca	1080	
aggetettge ggttgecata agagaageea eetegeeeaa tggtaeeaae gatgtteeet	1140	
ccaccaaagg tgttcttatg tagtgacacc gattatttaa agctgcagca tacgatatat	1200	

atacatgtgt atatatgtat acctatgaat gtcagtaagt atgtatacga acagtatgat 1260

313

314

actgaagatg	acaaggtaat	gcatcattct	atacgtgtca	ttctgaacga	ggcgcgcttt	1320
ccttttttct	ttttgctttt	tctttttt	tctcttgaac	tcgacggatc	tatgcggtgt	1380
gaaataccgc	acagatgcgt	aaggagaaaa	taccgcatca	ggaaattgta	aacgttaata	1440
ttttgttaaa	attcgcgtta	aatttttgtt	aaatcagctc	atttttaac	caataggccg	1500
aaatcggcaa	aatcccttat	aaatcaaaag	aatagaccga	gatagggttg	agtgttgttc	1560
cagtttggaa	caagagtcca	ctattaaaga	acgtggactc	caacgtcaaa	gggcgaaaaa	1620
ccgtctatca	gggcgatggc	ccactacgtg	aaccatcacc	ctaatcaagt	tttttggggt	1680
cgaggtgccg	taaagcacta	aatcggaacc	ctaaagggag	cccccgattt	agagettgae	1740
ggggaaagcc	ggcgaacgtg	gcgagaaagg	aagggaagaa	agcgaaagga	gcgggcgcta	1800
gggcgctggc	aagtgtagcg	gtcacgctgc	gcgtaaccac	cacacccgcc	gcgcttaatg	1860
cgccgctaca	gggcgcgtcg	cgccattcgc	cattcaggct	gcgcaactgt	tgggaagggc	1920
gatcggtgcg	ggcctcttcg	ctattacgcc	agctggcgaa	aggggggatgt	gctgcaaggc	1980
gattaagttg	ggtaacgcca	gggttttccc	agtcacgacg	ttgtaaaacg	acggccagtg	2040
agcgcgcgta	atacgactca	ctatagggcg	aattgggtac	cgggcccccc	ctcgaggtcg	2100
acggtatcga	taagcttgat	tagaagccgc	cgagcgggcg	acagecetee	gacggaagac	2160
tctcctccgt	gcgtcctcgt	cttcaccggt	cgcgttcctg	aaacgcagat	gtgcctcgcg	2220
ccgcactgct	ccgaacaata	aagattctac	aatactagct	tttatggtta	tgaagaggaa	2280
aaattggcag	taacctggcc	ccacaaacct	tcaaattaac	gaatcaaatt	aacaaccata	2340
ggatgataat	gcgattagtt	ttttagcctt	atttctgggg	taattaatca	gcgaagcgat	2400
gatttttgat	ctattaacag	atatataaat	ggaaaagctg	cataaccact	ttaactaata	2460
ctttcaacat	tttcagtttg	tattacttct	tattcaaatg	tcataaaagt	atcaacaaaa	2520
aattgttaat	atacctctat	actttaacgt	caaggagaaa	aatgtccaat	ttactgcccg	2580
tacaccaaaa	tttgcctgca	ttaccggtcg	atgcaacgag	tgatgaggtt	cgcaagaacc	2640
tgatggacat	gttcagggat	cgccaggcgt	tttctgagca	tacctggaaa	atgettetgt	2700
ccgtttgccg	gtcgtgggcg	gcatggtgca	agttgaataa	ccggaaatgg	tttcccgcag	2760
aacctgaaga	tgttcgcgat	tatcttctat	atcttcaggc	gcgcggtctg	gcagtaaaaa	2820
ctatccagca	acatttgggc	cagctaaaca	tgcttcatcg	tcggtccggg	ctgccacgac	2880
caagtgacag	caatgctgtt	tcactggtta	tgcggcggat	ccgaaaagaa	aacgttgatg	2940
ccggtgaacg	tgcaaaacag	gctctagcgt	tcgaacgcac	tgatttcgac	caggttcgtt	3000
cactcatgga	aaatagcgat	cgctgccagg	atatacgtaa	tctggcattt	ctggggattg	3060
cttataacac	cctgttacgt	atagccgaaa	ttgccaggat	cagggttaaa	gatatctcac	3120
gtactgacgg	tgggagaatg	ttaatccata	ttggcagaac	gaaaacgctg	gttagcaccg	3180
caggtgtaga	gaaggcactt	agcctggggg	taactaaact	ggtcgagcga	tggatttccg	3240
tctctggtgt	agctgatgat	ccgaataact	acctgttttg	ccgggtcaga	aaaaatggtg	3300
ttgccgcgcc	atctgccacc	agccagctat	caactcgcgc	cctggaaggg	atttttgaag	3360
caactcatcg	attgatttac	ggcgctaagg	atgactctgg	tcagagatac	ctggcctggt	3420
ctggacacag	tgcccgtgtc	ggagccgcgc	gagatatggc	ccgcgctgga	gtttcaatac	3480
cggagatcat	gcaagctggt	ggctggacca	atgtaaatat	tgtcatgaac	tatatccgta	3540
acctggatag	tgaaacaggg	gcaatggtgc	gcctgctgga	agatggcgat	taggagtaag	3600

315

cgaatttctt	atgatttatg	atttttatta	ttaaataagt	tataaaaaaa	ataagtgtat	3660
acaaatttta	aagtgactct	taggttttaa	aacgaaaatt	cttattcttg	agtaactctt	3720
tcctgtaggt	caggttgctt	tctcaggtat	agcatgaggt	cgctcttatt	gaccacacct	3780
ctaccggcat	gccgagcaaa	tgcctgcaaa	tcgctcccca	tttcacccaa	ttgtagatat	3840
gctaactcca	gcaatgagtt	gatgaatctc	ggtgtgtatt	ttatgtcctc	agaggacaac	3900
acctgtggtg	ttctagagcg	gccgccaccg	cggtggagct	ccagcttttg	ttccctttag	3960
tgagggttaa	ttgcgcgctt	ggcgtaatca	tggtcatagc	tgtttcctgt	gtgaaattgt	4020
tatccgctca	caattccaca	caacatagga	gccggaagca	taaagtgtaa	agcctggggt	4080
gcctaatgag	tgaggtaact	cacattaatt	gcgttgcgct	cactgcccgc	tttccagtcg	4140
ggaaacctgt	cgtgccagct	gcattaatga	atcggccaac	gcgcggggag	aggcggtttg	4200
cgtattgggc	gctcttccgc	ttcctcgctc	actgactcgc	tgcgctcggt	cgttcggctg	4260
cggcgagcgg	tatcagctca	ctcaaaggcg	gtaatacggt	tatccacaga	atcaggggat	4320
aacgcaggaa	agaacatgtg	agcaaaaggc	cagcaaaagg	ccaggaaccg	taaaaaggcc	4380
gcgttgctgg	cgtttttcca	taggctccgc	ccccctgacg	agcatcacaa	aaatcgacgc	4440
tcaagtcaga	ggtggcgaaa	cccgacagga	ctataaagat	accaggcgtt	tccccctgga	4500
agctccctcg	tgcgctctcc	tgttccgacc	ctgccgctta	ccggatacct	gtccgccttt	4560
ctcccttcgg	gaagcgtggc	gctttctcat	agctcacgct	gtaggtatct	cagttcggtg	4620
taggtcgttc	gctccaagct	gggctgtgtg	cacgaacccc	ccgttcagcc	cgaccgctgc	4680
gccttatccg	gtaactatcg	tcttgagtcc	aacccggtaa	gacacgactt	atcgccactg	4740
gcagcagcca	ctggtaacag	gattagcaga	gcgaggtatg	taggcggtgc	tacagagttc	4800
ttgaagtggt	ggcctaacta	cggctacact	agaaggacag	tatttggtat	ctgcgctctg	4860
ctgaagccag	ttaccttcgg	aaaaagagtt	ggtagctctt	gatccggcaa	acaaaccacc	4920
gctggtagcg	gtggttttt	tgtttgcaag	cagcagatta	cgcgcagaaa	aaaaggatct	4980
caagaagatc	ctttgatctt	ttctacgggg	tctgacgctc	agtggaacga	aaactcacgt	5040
taagggattt	tggtcatgag	attatcaaaa	aggatcttca	cctagatcct	tttaaattaa	5100
aaatgaagtt	ttaaatcaat	ctaaagtata	tatgagtaaa	cttggtctga	cagttaccaa	5160
tgcttaatca	gtgaggcacc	tatctcagcg	atctgtctat	ttcgttcatc	catagttgcc	5220
tgactccccg	tcgtgtagat	aactacgata	cgggagggct	taccatctgg	ccccagtgct	5280
gcaatgatac	cgcgagaccc	acgctcaccg	gctccagatt	tatcagcaat	aaaccagcca	5340
gccggaaggg	ccgagcgcag	aagtggtcct	gcaactttat	ccgcctccat	ccagtctatt	5400
aattgttgcc	gggaagctag	agtaagtagt	tcgccagtta	atagtttgcg	caacgttgtt	5460
gccattgcta	caggcatcgt	ggtgtcacgc	tcgtcgtttg	gtatggcttc	attcagctcc	5520
ggttcccaac	gatcaaggcg	agttacatga	tcccccatgt	tgtgcaaaaa	agcggttagc	5580
tccttcggtc	ctccgatcgt	tgtcagaagt	aagttggccg	cagtgttatc	actcatggtt	5640
atggcagcac	tgcataattc	tcttactgtc	atgccatccg	taagatgctt	ttctgtgact	5700
ggtgagtact	caaccaagtc	attctgagaa	tagtgtatgc	ggcgaccgag	ttgctcttgc	5760
ccggcgtcaa	tacgggataa	taccgcgcca	catagcagaa	ctttaaaagt	gctcatcatt	5820
ggaaaacgtt	cttcggggcg	aaaactctca	aggatettae	cgctgttgag	atccagttcg	5880
atgtaaccca	ctcgtgcacc	caactgatct	tcagcatctt	ttactttcac	cagegtttet	5940
gggtgaqcaa	aaacagqaaq	gcaaaatqcc	gcaaaaaaqq	gaataaqqqc	gacacggaaa	6000
	55 -5		- 55			
317

tgttgaatac tcatactctt cctttttcaa tattattgaa gcatttatca gggttattgt

### -continued

ctcatgagcg gatacatatt tgaatgtatt tagaaaaata aacaaatagg ggttccgcgc acatttcccc gaaaagtgcc acctgaacga agcatctgtg cttcattttg tagaacaaaa atgcaacgcg agagcgctaa tttttcaaac aaagaatctg agctgcattt ttacagaaca gaaatgcaac gcgaaagcgc tattttacca acgaagaatc tgtgcttcat ttttgtaaaa caaaaatgca acgcgagagc gctaattttt caaacaaaga atctgagctg catttttaca gaacagaaat gcaacgcgag agcgctattt taccaacaaa gaatctatac ttctttttg ttctacaaaa atgcatcccg agagegetat ttttctaaca aagcatctta gattactttt tttctccttt gtgcgctcta taatgcagtc tcttgataac tttttgcact gtaggtccgt taaggttaga agaaggctac tttggtgtct attttctctt ccataaaaaa agcctgactc cactteeege gtttaetgat taetagegaa getgegggtg catttttea agataaagge atccccgatt atattctata ccgatgtgga ttgcgcatac tttgtgaaca gaaagtgata gcgttgatga ttcttcattg gtcagaaaat tatgaacggt ttcttctatt ttgtctctat atactacgta taggaaatgt ttacattttc gtattgtttt cgattcactc tatgaatagt tottactaca attttttgt otaaagagta atactagaga taaacataaa aaatgtagag gtcgagttta gatgcaagtt caaggagcga aaggtggatg ggtaggttat atagggatat agcacagaga tatatagcaa agagatactt ttgagcaatg tttgtggaag cggtattcgc aatattttag tagetegtta cagteeggtg egtttttggt tttttgaaag tgegtettea gagcgctttt ggttttcaaa agcgctctga agttcctata ctttctagag aataggaact toggaatagg aacttoaaag ogtttoogaa aacgagogot toogaaaatg caacgogago tgcgcacata cageteactg tteacgtege acetatatet gegtgttgee tgtatatata tatacatgag aagaacggca tagtgcgtgt ttatgcttaa atgcgtactt atatgcgtct atttatgtag gatgaaaggt agtctagtac ctcctgtgat attatcccat tccatgcggg gtatcgtatg cttccttcag cactaccctt tagctgttct atatgctgcc actcctcaat tggattagtc tcatccttca atgctatcat ttcctttgat attggatcat ctaagaaacc attattatca tgacattaac ctataaaaat aggcgtatca cgaggccctt tcgtc <210> SEQ ID NO 122 <211> LENGTH: 30 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Primer T001 <400> SEQUENCE: 122 tcaaggtacc atggcaagtt cgggcacaac <210> SEQ ID NO 123 <211> LENGTH: 32 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Primer T002 <400> SEQUENCE: 123 taacgeggee gettatteee ceaceattte ag

<210> SEQ ID NO 124

6060

6120

6180

6240

6300

6360 6420

6480

6540

6600

6660

6720

6780

6840

6900

6960 7020

7080

7140

7200

7260

7320

7380

7440

7500

7555

30

ntinued

320

-continued	
<2115 LENGTH: 36	
<212> TYPE: DNA	
<213> ORGANISM: Artificial Sequence	
<220> FEATURE:	
<223> OTHER INFORMATION: Primer T003	
<400> SEQUENCE: 124	
atcattgcat gcgcctactt ggcttcacat acgttg	36
<210> SEO TO NO 125	
<211> LENGTH: 28	
<212> TYPE: DNA	
<213> ORGANISM: Artificial Sequence	
<220> FEATURE: <223> OTHER INFORMATION: Primer T004	
<400> SEQUENCE: 125	
catggtacct tgaatatgta ttacttgg	28
<210> SEQ ID NO 126	
<211> LENGTH: 36	
<212> TYPE: DNA	
<213> ORGANISM: Artificial Sequence	
<223> OTHER INFORMATION: Primer T005	
<400> SEQUENCE: 126	
	26
alaageggee gegllaalle aaallaallg alalag	30
<2105 SEC TD NO 127	
<211> LENGTH: 36	
<212> TYPE: DNA	
<213> ORGANISM: Artificial Sequence	
<220> FEATURE:	
<223> OTHER INFORMATION: Primer T006	
<400> SEQUENCE: 127	
ttagattgga tcccgcgaac tccaaaatga gctatc	36
<210> SEQ ID NO 128	
<211> LENGTH: 94	
<212> TYPE: DNA	
<213> ORGANISM: Artificial Sequence	
<223> OTHER INFORMATION: Primer T007	
<400> SEQUENCE: 128	
atgagagett tggcatattt caagaagggt gatatteaet teaetaatga tateeetagg	60
ccaggcctac ttggcttcac atacgttgca tacg	94
110 JPO TO NO 100	
<2107 SEQ 10 129 <2115 LENGTH: 91	
<212> TYPE: DNA	
<213> ORGANISM: Artificial Sequence	
<220> FEATURE:	
<223> OTHER INFORMATION: Primer T008	
<400> SEQUENCE: 129	
ttacttcatt tcaccgtgat tgttaggcgt caatagaatc ttaacgttgg attccttgtg	60
cacettgget aactegttgt ateateaetg g	91
<210> SEQ ID NO 130	
<211> LENGTH: 2064	

<212> TYPE: DNA <213> ORGANISM: Saccharomyces cerevisiae

<400> SEQUENCE: 130 atgatcagac aatctacgct aaaaaacttc gctattaagc gttgctttca acatatagca 60 taccgcaaca cacctgccat gagatcagta gctctcgcgc agcgctttta tagttcgtct 120 tecegttatt acagtgegte tecattacea geetetaaaa ggeeagagee tgeteeaagt 180 ttcaatgttg atccattaga acagcccgct gaaccttcaa aattggctaa gaaactacgc 240 getgageetg acatggatae etetttegte ggtttaaetg gtggteaaat atttaaegaa 300 atgatgtcca gacaaaacgt tgatactgta tttggttatc caggtggtgc tatcctacct 360 gtttacgatg ccattcataa cagtgataaa ttcaacttcg ttcttccaaa acacgaacaa 420 ggtgccggtc acatggcaga aggctacgcc agagcttctg gtaaaccagg tgttgtcttg 480 gttacttctg ggccaggtgc caccaatgtc gttactccaa tggcagatgc ctttgcagac 540 gggattccaa tggttgtctt tacagggcaa gtcccaacta gtgctatcgg tactgatgct 600 ttccaagagg ctgacgtcgt tggtatttct agatcttgta cgaaatggaa tgtcatggtc 660 aagteegtgg aagaattgee attgegtatt aaegaggett ttgaaattge caegageggt 720 agaccgggac cagtettggt egatttacca aaggatgtta cageagetat ettaagaaat 780 ccaattecaa caaaaacaac tettecatea aacgeaetaa accaattaac cagtegegea 840 caagatgaat ttgtcatgca aagtatcaat aaagcagcag atttgatcaa cttggcaaag 900 960 aaacctgtct tatacgtcgg tgctggtatt ttaaaccatg cagatggtcc aagattacta aaagaattaa gtgaccgtgc tcaaatacct gtcaccacta ctttacaagg tttaggttca 1020 1080 ttcgaccaag aagatccaaa atcattggat atgcttggta tgcacggttg tgctactgcc aacctggcag tgcaaaatgc cgacttgata attgcagttg gtgctagatt cgacgaccgt 1140 1200 gtcactggta atatttctaa attcgctcca gaagctcgtc gtgcagctgc cgagggtaga ggtggtatta ttcatttcga ggttagtcca aaaaacataa acaaggttgt tcaaactcaa 1260 atagcagtgg aaggtgatgc tacgaccaat ctgggcaaaa tgatgtcaaa gattttccca 1320 gttaaggaga ggtctgaatg gtttgctcaa ataaataaat ggaagaagga atacccatac 1380 gcttatatgg aggagactcc aggatctaaa attaaaccac agacggttat aaagaaacta 1440 tccaaggttg ccaacgacac aggaagacat gtcattgtta caacgggtgt ggggcaacat 1500 caaatgtggg ctgctcaaca ctggacatgg agaaatccac atactttcat cacatcaggt 1560 ggtttaggta cgatgggtta cggtctccct gccgccatcg gtgctcaagt tgcaaagcca 1620 gaatctttgg ttattgacat tgatggtgac gcatccttta acatgactct aacggaattg 1680 agttctgccg ttcaagctgg tactccagtg aagattttga ttttgaacaa tgaagagcaa 1740 ggtatggtta ctcaatggca atccctgttc tacgaacatc gttattccca cacacatcaa 1800 ttgaaccctg atttcataaa actagcggag gctatgggtt taaaaggttt aagagtcaag 1860 aagcaagagg aattggacgc taagttgaaa gaattcgttt ctaccaaggg cccagttttg 1920 cttgaagtgg aagttgataa aaaagttcct gttttgccaa tggtggcagg tggtagcggt 1980 ctagacgagt tcataaattt tgacccagaa gttgaaagac aacagactga attacgtcat 2040 aagcgtacag gcggtaagca ctga 2064

<210> SEQ ID NO 131 <211> LENGTH: 687 <212> TYPE: PRT <213> ORGANISM: Saccharomyces cerevisiae

<400	)> SH	EQUEI	ICE :	131											
Met 1	Ile	Arg	Gln	Ser 5	Thr	Leu	Lys	Asn	Phe 10	Ala	Ile	Lys	Arg	Cys 15	Phe
Gln	His	Ile	Ala 20	Tyr	Arg	Asn	Thr	Pro 25	Ala	Met	Arg	Ser	Val 30	Ala	Leu
Ala	Gln	Arg 35	Phe	Tyr	Ser	Ser	Ser 40	Ser	Arg	Tyr	Tyr	Ser 45	Ala	Ser	Pro
Leu	Pro 50	Ala	Ser	Lys	Arg	Pro 55	Glu	Pro	Ala	Pro	Ser 60	Phe	Asn	Val	Asp
Pro 65	Leu	Glu	Gln	Pro	Ala 70	Glu	Pro	Ser	Lys	Leu 75	Ala	LÀa	Lys	Leu	Arg 80
Ala	Glu	Pro	Asp	Met 85	Asp	Thr	Ser	Phe	Val 90	Gly	Leu	Thr	Gly	Gly 95	Gln
Ile	Phe	Asn	Glu 100	Met	Met	Ser	Arg	Gln 105	Asn	Val	Asp	Thr	Val 110	Phe	Gly
Tyr	Pro	Gly 115	Gly	Ala	Ile	Leu	Pro 120	Val	Tyr	Asp	Ala	Ile 125	His	Asn	Ser
Asp	Lys 130	Phe	Asn	Phe	Val	Leu 135	Pro	Lys	His	Glu	Gln 140	Gly	Ala	Gly	His
Met 145	Ala	Glu	Gly	Tyr	Ala 150	Arg	Ala	Ser	Gly	Lys 155	Pro	Gly	Val	Val	Leu 160
Val	Thr	Ser	Gly	Pro 165	Gly	Ala	Thr	Asn	Val 170	Val	Thr	Pro	Met	Ala 175	Aap
Ala	Phe	Ala	Asp 180	Gly	Ile	Pro	Met	Val 185	Val	Phe	Thr	Gly	Gln 190	Val	Pro
Thr	Ser	Ala 195	Ile	Gly	Thr	Asp	Ala 200	Phe	Gln	Glu	Ala	Asp 205	Val	Val	Gly
Ile	Ser 210	Arg	Ser	Суз	Thr	Lys 215	Trp	Asn	Val	Met	Val 220	ГЛЗ	Ser	Val	Glu
Glu 225	Leu	Pro	Leu	Arg	Ile 230	Asn	Glu	Ala	Phe	Glu 235	Ile	Ala	Thr	Ser	Gly 240
Arg	Pro	Gly	Pro	Val 245	Leu	Val	Asp	Leu	Pro 250	Lys	Asp	Val	Thr	Ala 255	Ala
Ile	Leu	Arg	Asn 260	Pro	Ile	Pro	Thr	Lys 265	Thr	Thr	Leu	Pro	Ser 270	Asn	Ala
Leu	Asn	Gln 275	Leu	Thr	Ser	Arg	Ala 280	Gln	Aab	Glu	Phe	Val 285	Met	Gln	Ser
Ile	Asn 290	Lys	Ala	Ala	Asb	Leu 295	Ile	Asn	Leu	Ala	Lys 300	Lys	Pro	Val	Leu
Tyr 305	Val	Gly	Ala	Gly	Ile 310	Leu	Asn	His	Ala	Asp 315	Gly	Pro	Arg	Leu	Leu 320
Lys	Glu	Leu	Ser	Asp 325	Arg	Ala	Gln	Ile	Pro 330	Val	Thr	Thr	Thr	Leu 335	Gln
Gly	Leu	Gly	Ser 340	Phe	Asp	Gln	Glu	Asp 345	Pro	Lys	Ser	Leu	Asp 350	Met	Leu
Gly	Met	His 355	Gly	Суа	Ala	Thr	Ala 360	Asn	Leu	Ala	Val	Gln 365	Asn	Ala	Asp
Leu	Ile 370	Ile	Ala	Val	Gly	Ala 375	Arg	Phe	Asp	Asp	Arg 380	Val	Thr	Gly	Asn
Ile 385	Ser	Lys	Phe	Ala	Pro 390	Glu	Ala	Arg	Arg	Ala 395	Ala	Ala	Glu	Gly	Arg 400
Gly	Gly	Ile	Ile	His 405	Phe	Glu	Val	Ser	Pro 410	Lys	Asn	Ile	Asn	Lys 415	Val

Val	Gln	Thr	Gln 420	Ile	Ala	Val	Glu	Gly 425	Asp	Ala	Thr	Thr	Asn 430	Leu	Gly	
Lys	Met	Met 435	Ser	Lys	Ile	Phe	Pro 440	Val	Lys	Glu	Arg	Ser 445	Glu	Trp	Phe	
Ala	Gln 450	Ile	Asn	Lys	Trp	Lys 455	Lys	Glu	Tyr	Pro	Tyr 460	Ala	Tyr	Met	Glu	
Glu 465	Thr	Pro	Gly	Ser	Lys 470	Ile	Lys	Pro	Gln	Thr 475	Val	Ile	LÀa	Гла	Leu 480	
Ser	Lys	Val	Ala	Asn 485	Asp	Thr	Gly	Arg	His 490	Val	Ile	Val	Thr	Thr 495	Gly	
Val	Gly	Gln	His 500	Gln	Met	Trp	Ala	Ala 505	Gln	His	Trp	Thr	Trp 510	Arg	Asn	
Pro	His	Thr 515	Phe	Ile	Thr	Ser	Gly 520	Gly	Leu	Gly	Thr	Met 525	Gly	Tyr	Gly	
Leu	Pro 530	Ala	Ala	Ile	Gly	Ala 535	Gln	Val	Ala	Lys	Pro 540	Glu	Ser	Leu	Val	
Ile 545	Asp	Ile	Asp	Gly	Asp 550	Ala	Ser	Phe	Asn	Met 555	Thr	Leu	Thr	Glu	Leu 560	
Ser	Ser	Ala	Val	Gln 565	Ala	Gly	Thr	Pro	Val 570	Lys	Ile	Leu	Ile	Leu 575	Asn	
Asn	Glu	Glu	Gln 580	Gly	Met	Val	Thr	Gln 585	Trp	Gln	Ser	Leu	Phe 590	Tyr	Glu	
His	Arg	Tyr 595	Ser	His	Thr	His	Gln 600	Leu	Asn	Pro	Asp	Phe 605	Ile	Lys	Leu	
Ala	Glu 610	Ala	Met	Gly	Leu	Lys 615	Gly	Leu	Arg	Val	Lys 620	LYa	Gln	Glu	Glu	
Leu 625	Asp	Ala	Lys	Leu	Lys 630	Glu	Phe	Val	Ser	Thr 635	Lys	Gly	Pro	Val	Leu 640	
Leu	Glu	Val	Glu	Val 645	Asp	Гла	Lys	Val	Pro 650	Val	Leu	Pro	Met	Val 655	Ala	
Gly	Gly	Ser	Gly 660	Leu	Asp	Glu	Phe	Ile 665	Asn	Phe	Asp	Pro	Glu 670	Val	Glu	
Arg	Gln	Gln 675	Thr	Glu	Leu	Arg	His 680	ГÀа	Arg	Thr	Gly	Gly 685	LÀa	His		
<21) <21; <21; <21; <40;	0 > SI 1 > LI 2 > T 3 > OF 0 > SI	EQ II ENGTH YPE : RGANI EOUEI	D NO H: 10 DNA ISM: NCE:	132 589 Escl 132	nerio	chia	col:	Ĺ								
atg	gcaa	gtt d	cggg	cacaa	ac at	ccga	cgcgt	: aaq	geget	tta	ccg	gegea	aga a	attta	atcgtt	60
cat	ttcct	agg a	aacaq	gcago	gg ca	atta	agatt	t gte	gacaq	ggca	ttc	cggg	cgg 1	ttcta	atcctg	120
cct	gttta	acg a	atgc	cttaa	ag co	caaa	gcaco	g caa	aatco	cgcc	ata	tct	ggc (	ccgto	catgaa	180
cag	ggcgo	cgg g	gctti	tatco	gc to	cagg	gaato	g gcé	gegea	accg	acg	gtaaa	acc g	ggcgé	gtctgt	240
atg	geetç	gta 🤉	gegga	accg	gg tạ	gcga	ctaad	c ctę	ggtga	accg	ccat	tgc	cga 1	tgcgo	eggetg	300
gac	tccat	cee d	cgct	gatti	cg ca	atca	ctggt	c caę	ggtt	cccg	cct	gate	gat (	cggca	accgac	360
gcc.	ttcca	agg a	aagto	ggaca	ac ci	cacg	gcato	e te	atco	ccca	tca	ccaa	aca	caact	atctg	420
gtc	agaca	ata t	tcgaa	agaa	ct co	ccgc	aggto	c ato	gage	gatg	ccti	cccg	cat 1	tgcgo	caatca	480
ggc	cgcco	cag g	geec	ggtgi	cg ga	ataga	acatt	c cct	aag	gatg	tgea	aaaco	ggc a	agtti	ttgag	540
att	gaaa	cac a	ageco	cgcta	at go	gcaga	aaaaa	a gco	gccé	geee	ccg	cctt	tag (	cgaaq	gaaagc	600

327

attcg	tgacg	cagc	ggcga	ıt ga	attaa	acgct	gco	caaa	cgcc	cggt	gctt	ta	tctg	ggcggc	660
ggtgt	gatca	atgc	gcccg	jc ad	cgggt	tgcgt	: gaa	actg	gcgg	agaa	aagco	gca	actgo	cctacc	720
accat	gactt	taat	ggcgc	t g	ggcat	tgttg	g cca	aaaaq	gcgc	atco	gtt	gtc	gctg	ggtatg	780
ctggg	gatgo	acgg	cgtgc	g ca	agca	ccaad	c tat	catt	tgc	agga	aggco	gga	tttgt	ttgata	840
gtgct	cggtg	cgcg	ttttg	ja to	gacco	gggcg	g att	ggca	aaaa	ccga	agcaç	gtt	ctgto	ccgaat	900
gccaa	aatca	ttca	tgtcg	ja ta	atcga	accgt	c gea	agago	ctgg	gtaa	aaato	caa	gcago	ccgcac	960
gtggc	gatto	aggc	ggatg	ıt tç	gatga	acgto	g ctç	ggcgo	cagt	tgat	cccc	gct	ggtgg	gaagcg	1020
caacc	gcgtg	caga	gtggc	a co	cagti	tggta	a geç	ggati	tgc	agco	gtgag	gtt	tccgt	tgtcca	1080
atccc	gaaag	cgtg	cgato	c gi	taaq	gccat	t tao	cggco	ctga	tcaa	acgco	gt	tgccç	gcctgt	1140
gtcga	tgaca	atgc	aatta	it ca	acca	ccgao	gtt	ggt	cage	atca	agato	gtg	gacco	gcgcaa	1200
gctta	tccgc	tcaa	tcgcc	c a	cgcca	agtgo	g ctç	gacct	cccg	gtg	ggcto	<u>a</u> aa	cacga	atgggt	1260
tttgg	cctgo	ctgc	ggcga	it tạ	ggcgo	ctgco	g ctç	ggcga	aacc	cgga	atcgo	caa	agtgt	tgtgt	1320
ttctc	cggcg	acgg	cagco	t ga	atgat	tgaat	t att	ccagę	gaga	tggo	cgaco	cgc	cagto	gaaaat	1380
caget	ggatg	tcaa	aatca	it to	ctgat	tgaad	c aad	cgaaq	gcgc	tggg	ggcto	ggt	gcato	cagcaa	1440
cagag	tctgt	tcta	cgagc	a aq	ggcgt	tttt	gco	cgcca	acct	atco	gggg	caa	aatca	aacttt	1500
atgca	gattg	ccgc	cggat	t co	ggcct	tcgaa	a aco	ctgt	gatt	tgaa	ataad	cga	agcco	gatccg	1560
cagge	ttcat	tgca	ggaaa	it ca	atcaa	atcgo	c cct	ggco	ccgg	cgct	gato	cca	tgtgo	cgcatt	1620
gatge	cgaag	aaaa	agttt	a co	ccgat	tggt	g ccé	gccaq	ggtg	cggo	cgaat	ac	tgaaa	atggtg	1680
gggga	ataa														1689
<210> <211> <212> <213> <400>	SEQ LENG TYPE ORGA SEOU	ID NO TH: 5 : PRT NISM: ENCE:	133 62 Esch 133	erio	chia	col:	L								
Met A	la Se	r Ser	Gly	Thr	Thr	Ser	Thr	Arg	Lys	Arg	Phe	Thr	Gly	Ala	
1	_	_	5			_	_	10	_	_		_	15		
Glu Pl	he Il	e Val 20	His	Phe	Leu	Glu	Gln 25	Gln	Gly	Ile	Lys	Ile 30	Val	Thr	
Gly I	le Pr 35	o Gly	Gly	Ser	Ile	Leu 40	Pro	Val	Tyr	Asp	Ala 45	Leu	Ser	Gln	
Ser Ti 5	hr Gl 0	n Ile	Arg	His	Ile 55	Leu	Ala	Arg	His	Glu 60	Gln	Gly	Ala	Gly	
Phe I 65	le Al	a Gln	Gly	Met 70	Ala	Arg	Thr	Asp	Gly 75	ГЛа	Pro	Ala	Val	СУЗ 80	
Met A	la Cy	s Ser	Gly 85	Pro	Gly	Ala	Thr	Asn 90	Leu	Val	Thr	Ala	Ile 95	Ala	
Asp A	la Ar	g Leu 100	Asp	Ser	Ile	Pro	Leu 105	Ile	Сув	Ile	Thr	Gly 110	Gln	Val	
Pro A	la Se 11	r Met 5	Ile	Gly	Thr	Asp 120	Ala	Phe	Gln	Glu	Val 125	Asp	Thr	Tyr	
Gly I	le Se 30	r Ile	Pro	Ile	Thr 135	Lys	His	Asn	Tyr	Leu 140	Val	Arg	His	Ile	
Glu G	lu Le	u Pro	Gln	Val	Met	Ser	Asp	Ala	Phe	Arg	Ile	Ala	Gln	Ser	
145				120					т22					τ00	
Gly A:	rg Pr	o Gly	Pro 165	Val	Trp	Ile	Asp	Ile 170	Pro	гла	Asp	Val	Gln 175	Thr	

Ala Val Phe Glu Ile Glu Thr Gln Pro Ala Met Ala Glu Lys Ala Ala Ala Pro Ala Phe Ser Glu Glu Ser Ile Arg Asp Ala Ala Ala Met Ile Asn Ala Ala Lys Arg Pro Val Leu Tyr Leu Gly Gly Gly Val Ile Asn Ala Pro Ala Arg Val Arg Glu Leu Ala Glu Lys Ala Gln Leu Pro Thr Thr Met Thr Leu Met Ala Leu Gly Met Leu Pro Lys Ala His Pro Leu Ser Leu Gly Met Leu Gly Met His Gly Val Arg Ser Thr Asn Tyr Ile Leu Gln Glu Ala Asp Leu Leu Ile Val Leu Gly Ala Arg Phe Asp Asp Arg Ala Ile Gly Lys Thr Glu Gln Phe Cys Pro Asn Ala Lys Ile Ile His Val Asp Ile Asp Arg Ala Glu Leu Gly Lys Ile Lys Gln Pro His Val Ala Ile Gln Ala Asp Val Asp Asp Val Leu Ala Gln Leu Ile Pro Leu Val Glu Ala Gln Pro Arg Ala Glu Trp His Gln Leu Val Ala Asp Leu Gln Arg Glu Phe Pro Cys Pro Ile Pro Lys Ala Cys Asp Pro Leu Ser His Tyr Gly Leu Ile Asn Ala Val Ala Ala Cys Val Asp Asp Asn Ala Ile Ile Thr Thr Asp Val Gly Gln His Gln Met Trp Thr Ala Gln Ala Tyr Pro Leu Asn Arg Pro Arg Gln Trp Leu Thr Ser Gly Gly Leu Gly Thr Met Gly Phe Gly Leu Pro Ala Ala Ile Gly Ala Ala Leu Ala Asn Pro Asp Arg Lys Val Leu Cys Phe Ser Gly Asp Gly Ser Leu Met Met Asn Ile Gln Glu Met Ala Thr Ala Ser Glu Asn Gln Leu Asp Val Lys Ile Ile Leu Met Asn Asn Glu Ala Leu Gly Leu Val His Gln Gln Gln Ser Leu Phe Tyr Glu Gln Gly Val Phe Ala Ala Thr Tyr Pro Gly Lys Ile Asn Phe Met Gln Ile Ala Ala Gly Phe Gly Leu Glu Thr Cys Asp Leu Asn Asn Glu Ala Asp Pro Gln Ala Ser Leu Gln Glu Ile Ile Asn Arg Pro Gly Pro Ala Leu Ile His Val Arg Ile Asp Ala Glu Glu Lys Val Tyr Pro Met Val Pro Pro Gly Ala Ala Asn Thr Glu Met Val 

Gly Glu

<210> SEQ ID NO 134 <211> LENGTH: 338 <212> TYPE: PRT <213> ORGANISM: Pseudomonas fluorescens

<400> SEQUENCE:	134				
Met Lys Val Phe 1	Tyr Asp Ly: 5	a Aap Cya A 1	Asp Leu Ser LO	Ile Ile	Gln Gly 15
Lys Lys Val Ala 20	Ile Ile Gl	y Tyr Gly S 25	Ser Gln Gly	His Ala 30	Gln Ala
Cys Asn Leu Lys 35	Asp Ser Gly	y Val Asp V 40	Val Thr Val	Gly Leu 45	Arg Lys
Gly Ser Ala Thr 50	Val Ala Ly: 55	s Ala Glu A	Ala His Gly 60	Leu Lys	Val Thr
Asp Val Ala Ala 65	Ala Val Al. 70	a Gly Ala A	Asp Leu Val 75	Met Ile	Leu Thr 80
Pro Asp Glu Phe	Gln Ser Gli 85	n Leu Tyr L 9	Jys Asn Glu 90	Ile Glu	Pro Asn 95
Ile Lys Lys Gly 100	Ala Thr Le	u Ala Phe S 105	Ser His Gly	Phe Ala 110	Ile His
Tyr Asn Gln Val 115	Val Pro Arg	g Ala Asp I 120	leu Asp Val	Ile Met 125	Ile Ala
Pro Lys Ala Pro 130	Gly His Th: 13	r Val Arg S 5	Ser Glu Phe 140	Val Lys	Gly Gly
Gly Ile Pro Asp 145	Leu Ile Al. 150	a Ile Tyr G	3ln Asp Ala 155	Ser Gly	Asn Ala 160
Lys Asn Val Ala	Leu Ser Ty: 165	r Ala Ala G 1	Gly Val Gly 170	Gly Gly	Arg Thr 175
Gly Ile Ile Glu 180	Thr Thr Ph	e Lys Asp G 185	3lu Thr Glu	Thr Asp 190	Leu Phe
Gly Glu Gln Ala 195	Val Leu Cya	s Gly Gly T 200	Thr Val Glu	Leu Val 205	Lys Ala
Gly Phe Glu Thr 210	Leu Val Glu 21	u Ala Gly T 5	fyr Ala Pro 220	Glu Met	Ala Tyr
Phe Glu Cys Leu 225	His Glu Le 230	u Lys Leu I	lle Val Asp 235	Leu Met	Tyr Glu 240
Gly Gly Ile Ala	Asn Met Ası 245	n Tyr Ser I 2	lle Ser Asn 250	Asn Ala	Glu Tyr 255
Gly Glu Tyr Val 260	Thr Gly Pro	o Glu Val I 265	Ile Asn Ala	Glu Ser 270	Arg Gln
Ala Met Arg Asn 275	Ala Leu Ly:	s Arg Ile G 280	Gln Asp Gly	Glu Tyr 285	Ala Lys
Met Phe Ile Ser 290	Glu Gly Al 29	a Thr Gly I 5	Tyr Pro Ser 300	Met Thr	Ala Lys
Arg Arg Asn Asn 305	Ala Ala Hi 310	s Gly Ile G	Glu Ile Ile 315	Gly Glu	Gln Leu 320
Arg Ser Met Met	Pro Trp Il 325	e Gly Ala A 3	Asn Lys Ile 330	Val Asp	Lys Ala 335
Lys Asn					
<210> SEQ ID NO <211> LENGTH: 5 <212> TYPE: PRT <213> ORGANISM:	135 71 Streptococo	cus mutans			
<400> SEQUENCE:	135				
Met Thr Asp Lys 1	Lys Thr Lev 5	u Lys Asp I 1	leu Arg Asn 10	Arg Ser	Ser Val 15

Tyr Asp Ser Met Val Lys Ser Pro Asn Arg Ala Met Leu Arg Ala Thr Gly Met Gln Asp Glu Asp Phe Glu Lys Pro Ile Val Gly Val Ile Ser Thr Trp Ala Glu Asn Thr Pro Cys Asn Ile His Leu His Asp Phe Gly Lys Leu Ala Lys Val Gly Val Lys Glu Ala Gly Ala Trp Pro Val Gln Phe Gly Thr Ile Thr Val Ser Asp Gly Ile Ala Met Gly Thr Gln Gly Met Arg Phe Ser Leu Thr Ser Arg Asp Ile Ile Ala Asp Ser Ile Glu 100 105 110 Ala Ala Met Gly Gly His Asn Ala Asp Ala Phe Val Ala Ile Gly Gly Cys Asp Lys Asn Met Pro Gly Ser Val Ile Ala Met Ala Asn Met Asp Ile Pro Ala Ile Phe Ala Tyr Gly Gly Thr Ile Ala Pro Gly Asn Leu Asp Gly Lys Asp Ile Asp Leu Val Ser Val Phe Glu Gly Val Gly His Trp Asn His Gly Asp Met Thr Lys Glu Glu Val Lys Ala Leu Glu Cys Asn Ala Cys Pro Gly Pro Gly Gly Cys Gly Gly Met Tyr Thr Ala Asn Thr Met Ala Thr Ala Ile Glu Val Leu Gly Leu Ser Leu Pro Gly Ser Ser Ser His Pro Ala Glu Ser Ala Glu Lys Lys Ala Asp Ile Glu Glu Ala Gly Arg Ala Val Val Lys Met Leu Glu Met Gly Leu Lys Pro Ser Asp Ile Leu Thr Arg Glu Ala Phe Glu Asp Ala Ile Thr Val Thr Met Ala Leu Gly Gly Ser Thr Asn Ser Thr Leu His Leu Leu Ala Ile Ala His Ala Ala Asn Val Glu Leu Thr Leu Asp Asp Phe Asn Thr Phe Gln Glu Lys Val Pro His Leu Ala Asp Leu Lys Pro Ser Gly Gln Tyr Val Phe Gln Asp Leu Tyr Lys Val Gly Gly Val Pro Ala Val Met Lys Tyr Leu Leu Lys Asn Gly Phe Leu His Gly Asp Arg Ile Thr Cys Thr Gly Lys Thr Val Ala Glu Asn Leu Lys Ala Phe Asp Asp Leu Thr Pro Gly Gln Lys Val Ile Met Pro Leu Glu Asn Pro Lys Arg Glu Asp Gly Pro Leu Ile Ile Leu His Gly Asn Leu Ala Pro Asp Gly Ala Val Ala Lys Val Ser Gly Val Lys Val Arg Arg His Val Gly Pro Ala Lys Val Phe Asn Ser Glu Glu Glu Ala Ile Glu Ala Val Leu Asn Asp Asp Ile Val 

	Gly	Asp 435	Val	Val	Val	Val	Arg 440	Phe	Val	Gly	Pro	Lys 445	Gly	Gly	Pro
Gly	Met 450	Pro	Glu	Met	Leu	Ser 455	Leu	Ser	Ser	Met	Ile 460	Val	Gly	Lys	Gly
Gln 465	Gly	Glu	Lys	Val	Ala 470	Leu	Leu	Thr	Asp	Gly 475	Arg	Phe	Ser	Gly	Gly 480
Thr	Tyr	Gly	Leu	Val 485	Val	Gly	His	Ile	Ala 490	Pro	Glu	Ala	Gln	Asp 495	Gly
Gly	Pro	Ile	Ala 500	Tyr	Leu	Gln	Thr	Gly 505	Asp	Ile	Val	Thr	Ile 510	Asp	Gln
Asp	Thr	Lys 515	Glu	Leu	His	Phe	Asp 520	Ile	Ser	Asp	Glu	Glu 525	Leu	Lys	His
Arg	Gln 530	Glu	Thr	Ile	Glu	Leu 535	Pro	Pro	Leu	Tyr	Ser 540	Arg	Gly	Ile	Leu
Gly 545	Lys	Tyr	Ala	His	Ile 550	Val	Ser	Ser	Ala	Ser 555	Arg	Gly	Ala	Val	Thr 560
Asp	Phe	Trp	Lys	Pro 565	Glu	Glu	Thr	Gly	Lys 570	Lys					
<21	0> SI	EQ II	1. E.	136 18											
<21 <21 <21	2 > T 3 > OF	YPE : RGANI	PRT SM:	±⊙ List	ceria	a gra	ayi								
<40	0> SI	EQUEI	ICE :	136											
Met 1	Tyr	Thr	Val	Gly 5	Gln	Tyr	Leu	Val	Asp 10	Arg	Leu	Glu	Glu	Ile 15	Gly
Ile	Asp	Lys	Val 20	Phe	Gly	Val	Pro	Gly 25	Asp	Tyr	Asn	Leu	Thr 30	Phe	Leu
Asp	Tyr	Ile 35	Gln	Asn	His	Glu	Gly 40	Leu	Ser	Trp	Gln	Gly 45	Asn	Thr	Asn
Glu	Leu 50	Asn	Ala	Ala	Tyr	Ala 55	Ala	Asp	Gly	Tyr	Ala 60	Arg	Glu	Arg	Gly
Val 65	Ser	-				Thr	-	Gly	Val	Gly	Glu	Lou	<i>a</i>	Ala	Ile
		Ala	Leu	Val	Thr 70	1111	Phe			75		цец	ser		80
Asn	Gly	Ala Thr	Leu Ala	Val Gly 85	Thr 70 Ser	Phe	Phe Ala	Glu	Gln 90	75 Val	Pro	Val	Ile	His 95	80 Ile
Asn Val	Gly Gly	Ala Thr Ser	Leu Ala Pro 100	Val Gly 85 Thr	Thr 70 Ser Met	Phe Asn	Ala Val	Glu Gln 105	Gln 90 Ser	75 Val Asn	Pro Lys	Val Lys	Ile Leu 110	His 95 Val	80 Ile His
Asn Val His	Gly Gly Ser	Ala Thr Ser Leu 115	Leu Ala Pro 100 Gly	Val Gly 85 Thr Met	Thr 70 Ser Met Gly	Phe Asn Asn	Phe Ala Val Phe 120	Glu Gln 105 His	Gln 90 Ser Asn	75 Val Asn Phe	Pro Lys Ser	Val Lys Glu 125	Ile Leu 110 Met	His 95 Val Ala	80 Ile His Lys
Asn Val His Glu	Gly Gly Ser Val 130	Ala Thr Ser Leu 115 Thr	Leu Ala Pro 100 Gly Ala	Val Gly 85 Thr Met Ala	Thr 70 Ser Met Gly Thr	Phe Asn Asn Thr 135	Phe Ala Val Phe 120 Met	Glu Gln 105 His Leu	Gln 90 Ser Asn Thr	75 Val Asn Phe Glu	Pro Lys Ser Glu 140	Val Lys Glu 125 Asn	Ile Leu 110 Met Ala	His 95 Val Ala Ala	80 Ile His Lys Ser
Asn Val His Glu Glu 145	Gly Gly Ser Val 130 Ile	Ala Thr Ser Leu 115 Thr Asp	Leu Ala Pro 100 Gly Ala Arg	Val Gly 85 Thr Met Ala Val	Thr 70 Ser Met Gly Thr Leu 150	Phe Asn Asn Thr 135 Glu	Phe Ala Val Phe 120 Met Thr	Glu Gln 105 His Leu Ala	Gln 90 Ser Asn Thr Leu	75 Val Asn Phe Glu Leu 155	Pro Lys Ser Glu 140 Glu	Val Lys Glu 125 Asn Lys	Ile Leu 110 Met Ala Arg	His 95 Val Ala Ala Pro	80 Ile His Lys Ser Val 160
Asn Val His Glu 145 Tyr	Gly Gly Ser Val 130 Ile Ile	Ala Thr Ser Leu 115 Thr Asp Asn	Leu Ala Pro 100 Gly Ala Arg Leu	Val Gly 85 Thr Met Ala Val Pro 165	Thr 70 Ser Met Gly Thr Leu 150 Ile	Phe Asn Asn Thr 135 Glu Asp	Ala Val Phe 120 Met Thr Ile	Glu Gln 105 His Leu Ala Ala	Gln 90 Ser Asn Thr Leu His 170	75 Val Asn Phe Glu Leu 155 Lys	Pro Lys Ser Glu 140 Glu Ala	Val Lys Glu 125 Asn Lys Ile	Ile Leu 110 Met Ala Arg Val	His 95 Val Ala Ala Pro Lys 175	80 Ile His Lys Ser Val 160 Pro
Asn Val His Glu 145 Tyr Ala	Gly Gly Ser Val 130 Ile Ile	Ala Thr Ser Leu 115 Thr Asp Asn Ala	Leu Ala Pro Gly Ala Arg Leu Leu	Val Gly 85 Thr Met Ala Val Pro 165 Gln	Thr 70 Ser Met Gly Thr Leu 150 Ile Thr	Phe Asn Asn Thr 135 Glu Asp Glu	Phe Ala Val Phe 120 Met Thr Ile Lys	Glu Gln 105 His Leu Ala Ala Ser 185	Gln 90 Ser Asn Thr Leu His 170 Ser	75 Val Asn Phe Glu Leu Lys Gly	Pro Lys Ser Glu 140 Glu Ala Glu	Val Lys Glu 125 Asn Lys Ile Arg	Ile Leu 110 Met Ala Arg Val Glu 190	His 95 Val Ala Ala Pro Lys 175 Ala	80 Ile His Lys Ser Val 160 Pro Gln
Asn Val His Glu 145 Tyr Ala Leu	Gly Gly Ser Val 130 Ile Lys Ala	Ala Thr Ser Leu 115 Thr Asp Asn Ala Glu 195	Leu Ala Pro 100 Gly Ala Arg Leu Leu 180	Val Gly 85 Thr Met Ala Val Pro 165 Gln Ile	Thr 70 Ser Met Gly Thr Leu 150 Ile Thr Leu	Phe Asn Asn Thr 135 Glu Asp Glu Ser	Ala Val Phe 120 Met Thr Ile Lys His 200	Glu Gln 105 His Leu Ala Ala Ser 185 Leu	Gln 90 Ser Asn Thr Leu His 170 Ser Glu	75 Val Asn Phe Glu Leu Lys Gly Lys	Pro Lys Ser Glu Glu Ala Glu Ala	Val Lys Glu 125 Asn Lys Ile Arg Ala 205	Ile Leu 110 Met Ala Arg Val Glu 190 Gln	His 95 Val Ala Ala Pro Lys 175 Ala Pro	80 Ile His Lys Ser Val 160 Pro Gln Ile
Asn Val His Glu 145 Tyr Ala Leu Val	Gly Gly Ser Val 130 Ile Lys Ala Ile 210	Ala Thr Ser Leu 115 Thr Asp Asn Ala Glu 195 Ala	Leu Ala Pro Gly Ala Arg Leu Leu 180 Ile Gly	Val Gly 85 Thr Met Ala Val Pro 165 Gln Ile His	Thr Ser Met Gly Thr Leu 150 Ile Thr Leu Glu	Phe Asn Asn Thr 135 Glu Asp Glu Ser Ile 215	Ala Val Phe 120 Met Thr Ile Lys Hiss 200 Ala	Glu Gln 105 His Leu Ala Ser 185 Leu Arg	Gln 90 Ser Asn Thr Leu His 170 Ser Glu Phe	75 Val Asn Phe Glu Leu Sly Gly Lys Gln	Pro Lys Ser Glu 140 Glu Ala Glu Ala Ile 220	Val Lys Glu 125 Asn Lys Ile Arg Ala 205 Arg	Ile Leu 110 Met Ala Arg Val Glu 190 Gln Glu	His 95 Val Ala Ala Pro Lys Ala Pro Ala	80 Ile His Lys Ser Val 160 Pro Gln Ile Phe

Gly	Lys	Gly	Ser	Phe 245	Asn	Glu	Glu	Asn	Glu 250	His	Phe	Ile	Gly	Thr 255	Tyr
Tyr	Pro	Ala	Phe 260	Ser	Asp	Lys	Asn	Val 265	Leu	Asp	Tyr	Val	Asp 270	Asn	Ser
Asp	Phe	Val 275	Leu	His	Phe	Gly	Gly 280	Lys	Ile	Ile	Asp	Asn 285	Ser	Thr	Ser
Ser	Phe 290	Ser	Gln	Gly	Phe	Lys 295	Thr	Glu	Asn	Thr	Leu 300	Thr	Ala	Ala	Asn
Asp 305	Ile	Ile	Met	Leu	Pro 310	Asp	Gly	Ser	Thr	Tyr 315	Ser	Gly	Ile	Ser	Leu 320
Asn	Gly	Leu	Leu	Ala 325	Glu	Leu	Glu	ГÀа	Leu 330	Asn	Phe	Thr	Phe	Ala 335	Asp
Thr	Ala	Ala	Lys 340	Gln	Ala	Glu	Leu	Ala 345	Val	Phe	Glu	Pro	Gln 350	Ala	Glu
Thr	Pro	Leu 355	Lys	Gln	Aab	Arg	Phe 360	His	Gln	Ala	Val	Met 365	Asn	Phe	Leu
Gln	Ala 370	Aab	Aab	Val	Leu	Val 375	Thr	Glu	Gln	Gly	Thr 380	Ser	Ser	Phe	Gly
Leu 385	Met	Leu	Ala	Pro	Leu 390	Lys	Lys	Gly	Met	Asn 395	Leu	Ile	Ser	Gln	Thr 400
Leu	Trp	Gly	Ser	Ile 405	Gly	Tyr	Thr	Leu	Pro 410	Ala	Met	Ile	Gly	Ser 415	Gln
Ile	Ala	Ala	Pro 420	Glu	Arg	Arg	His	Ile 425	Leu	Ser	Ile	Gly	Asp 430	Gly	Ser
Phe	Gln	Leu 435	Thr	Ala	Gln	Glu	Met 440	Ser	Thr	Ile	Phe	Arg 445	Glu	Lys	Leu
Thr	Pro 450	Val	Ile	Phe	Ile	Ile 455	Asn	Asn	Asp	Gly	Tyr 460	Thr	Val	Glu	Arg
Ala 465	Ile	His	Gly	Glu	Asp 470	Glu	Ser	Tyr	Asn	Asp 475	Ile	Pro	Thr	Trp	Asn 480
Leu	Gln	Leu	Val	Ala 485	Glu	Thr	Phe	Gly	Gly 490	Asp	Ala	Glu	Thr	Val 495	Asp
Thr	His	Asn	Val 500	Phe	Thr	Glu	Thr	Asp 505	Phe	Ala	Asn	Thr	Leu 510	Ala	Ala
Ile	Asp	Ala 515	Thr	Pro	Gln	Lys	Ala 520	His	Val	Val	Glu	Val 525	His	Met	Glu
Gln	Met 530	Asp	Met	Pro	Glu	Ser 535	Leu	Arg	Gln	Ile	Gly 540	Leu	Ala	Leu	Ser
Lys 545	Gln	Asn	Ser												
<210 <211	)> SH L> LH	EQ II ENGTI	) NO 1: 10	137 541											
<212 <213	2 > TY 3 > OF	(PE : RGANI	DNA ISM:	Macı	rocod	ccus	case	eolyt	icu	3					
<400	)> SH	equei	ICE :	137											
atga	aaaca	aac g	gtato	cgggo	ca at	actt	gato	c gat	geed	ctac	acgt	taat	tgg t	gtc	gataag
atct	ttg	gag t	ccca	aggto	ga ti	tcad	ettta	a gco	ettt	tgg	acga	atato	cat a	aagao	catgac
aaco	gtgga	aat g	gggtg	gggaa	aa ta	actaa	atgag	y ttç	jaaco	geeg	ctta	acgco	cgc t	gate	ggttac
gcta	agagt	ta a	atgga	attaç	ge eé	gctgt	atct	acc	cactt	ttg	gggt	tgg	cga q	gttat	tctgct
gtga	aatgo	gta t	tget	ggaa	ag ti	cacgo	cagag	g ogt	gtto	cctg	taat	caaa	aat d	ctcaç	ggcggt

	55		340
		-continued	
ccttcatcag ttgctcaaca aga	agggtaga tatgtccacc	attcattggg tgaaggaatc	360
tttgattcat attcaaagat gta	acgctcac ataaccgcaa	caactacaat cttatccgtt	420
gacaacgcag tcgacgaaat tga	atagagtt attcattgtg	ctttgaagga aaagaggcca	480
gtgcatattc atttgcctat tga	acgtagcc ttaactgaga	ttgaaatccc tcatgcacca	540
aaagtttaca cacacgaatc cca	agaacgtc gatgcttaca	ttcaagctgt tgagaaaaag	600
ttaatgtctg caaaacaacc agt	caatcata gcaggtcatg	aaatcaattc attcaagttg	660
cacgaacaac tggaacagtt tgt	caatcag acaaacatcc	ctgttgcaca actttccttg	720
ggtaagtetg ettteaatga aga	agaatgaa cattaccttg	gtatctacga tggcaaaatc	780
gcaaaggaaa atgtgagaga gta	acgtcgac aatgctgatg	tcatattgaa cataggtgcc	840
aaactgactg attctgctac ago	etggattt teetacaagt	tcgatacaaa caacataatc	900
tacattaacc ataatgactt caa	aagctgaa gatgtgattt	ctgataatgt ttcactgatt	960
gatettgtga atggeetgaa tte	ctattgac tatagaaatg	aaacacacta cccatcttat	1020
caaagatctg atatgaaata cga	aattgaat gacgcaccac	ttacacaatc taactatttc	1080
aaaatgatga acgcttttct aga	aaaaagat gacatcctac	tagctgaaca aggtacatcc	1140
tttttcggcg catatgactt atc	ccctatac aagggaaatc	agtttatcgg tcagccttta	1200
tgggggtcaa tagggtatac ttt	tccatct ttactaggaa	gtcaactagc agacatgcat	1260
aggagaaaca ttttgcttat ago	gcgatggt agtttacaac	ttactgttca agccctaagt	1320
acaatgatta gaaaggatat caa	aaccaatc attttcgtta	tcaataacga cggttacacc	1380
gtcgaaagac ttatccacgg cat	tggaagag ccatacaatg	atatccaaat gtggaactac	1440
aagcaattgc cagaagtatt tgg	gtggaaaa gatactgtaa	aagttcatga tgctaaaacc	1500
tccaacgaac tgaaaactgt aat	eggattet gttaaageag	acaaagatca catgcatttc	1560
attgaagtgc atatggcagt aga	aggacgcc ccaaagaagt	tgattgatat agctaaagcc	1620
tttagtgatg ctaacaagta a			1641
<pre>&lt;210&gt; SEQ ID NO 138 &lt;211&gt; LENGTH: 347 &lt;212&gt; TYPE: PRT &lt;213&gt; ORGANISM: Beijerick &lt;400&gt; SEQUENCE: 138</pre>	cia indica		
Met Lyg Ala Leu Val Tyr Z	ra Gly Pro Gly Gln	Lyc Leu Val Clu Clu	
1 5	10	15	
Arg Gln Lys Pro Glu Leu I 20	Lys Glu Pro Gly Asp 25	Ala Ile Val Lys Val 30	
Thr Lys Thr Thr Ile Cys G 35	Gly Thr Asp Leu His 40	Ile Leu Lys Gly Asp 45	
Val Ala Thr Cys Lys Pro C 50 5	Gly Arg Val Leu Gly 55	His Glu Gly Val Gly 60	
Val Ile Glu Ser Val Gly S 65 70	Ser Gly Val Thr Ala 75	Phe Gln Pro Gly Asp 80	
Arg Val Leu Ile Ser Cys I 85	lle Ser Ser Cys Gly 90	Lys Cys Ser Phe Cys 95	
Arg Arg Gly Met Phe Ser H 100	His Cys Thr Thr Gly 105	Gly Trp Ile Leu Gly 110	
Asn Glu Ile Asp Gly Thr C 115	Gln Ala Glu Tyr Val 120	Arg Val Pro His Ala 125	
Asp Thr Ser Leu Tyr Arg I	Ile Pro Ala Gly Ala	Asp Glu Glu Ala Leu	

	130					135					140				
Val 145	Met	Leu	Ser	Asp	Ile 150	Leu	Pro	Thr	Gly	Phe 155	Glu	Cys	Gly	Val	Leu 160
Asn	Gly	Lys	Val	Ala 165	Pro	Gly	Ser	Ser	Val 170	Ala	Ile	Val	Gly	Ala 175	Gly
Pro	Val	Gly	Leu 180	Ala	Ala	Leu	Leu	Thr 185	Ala	Gln	Phe	Tyr	Ser 190	Pro	Ala
Glu	Ile	Ile 195	Met	Ile	Asp	Leu	Asp 200	Asp	Asn	Arg	Leu	Gly 205	Leu	Ala	Lys
Gln	Phe 210	Gly	Ala	Thr	Arg	Thr 215	Val	Asn	Ser	Thr	Gly 220	Gly	Asn	Ala	Ala
Ala 225	Glu	Val	Lys	Ala	Leu 230	Thr	Glu	Gly	Leu	Gly 235	Val	Asp	Thr	Ala	Ile 240
Glu	Ala	Val	Gly	Ile 245	Pro	Ala	Thr	Phe	Glu 250	Leu	Сүз	Gln	Asn	Ile 255	Val
Ala	Pro	Gly	Gly 260	Thr	Ile	Ala	Asn	Val 265	Gly	Val	His	Gly	Ser 270	Lys	Val
Asp	Leu	His 275	Leu	Glu	Ser	Leu	Trp 280	Ser	His	Asn	Val	Thr 285	Ile	Thr	Thr
Arg	Leu 290	Val	Asp	Thr	Ala	Thr 295	Thr	Pro	Met	Leu	Leu 300	Lys	Thr	Val	Gln
Ser 305	His	Lys	Leu	Asp	Pro 310	Ser	Arg	Leu	Ile	Thr 315	His	Arg	Phe	Ser	Leu 320
Asp	Gln	Ile	Leu	Asp 325	Ala	Tyr	Glu	Thr	Phe 330	Gly	Gln	Ala	Ala	Ser 335	Thr
Gln	Ala	Leu	Lys 340	Val	Ile	Ile	Ser	Met 345	Glu	Ala					
<210 <211 <212 <213	)> SH L> LH 2> TY 3> OF	EQ II ENGTH (PE : RGANI	) NO 1: 34 PRT [SM:	139 18 Achi	romol	pacte	er xy	vloso	oxida	ans					
<400	)> SE	EQUEI	ICE :	139											
Met 1	Lys	Ala	Leu	Val 5	Tyr	His	Gly	Asp	His 10	Lys	Ile	Ser	Leu	Glu 15	Asp
Lys	Pro	Lys	Pro 20	Thr	Leu	Gln	Lys	Pro 25	Thr	Asp	Val	Val	Val 30	Arg	Val
Leu	Lys	Thr 35	Thr	Ile	Сүз	Gly	Thr 40	Asp	Leu	Gly	Ile	Tyr 45	Lys	Gly	Гла
Asn	Pro 50	Glu	Val	Ala	Asp	Gly 55	Arg	Ile	Leu	Gly	His 60	Glu	Gly	Val	Gly
Val 65	Ile	Glu	Glu	Val	Gly 70	Glu	Ser	Val	Thr	Gln 75	Phe	Lya	Lys	Gly	Asp 80
Lys	Val	Leu	Ile	Ser 85	Сүз	Val	Thr	Ser	Cys 90	Gly	Ser	Суз	Asp	Tyr 95	СЛа
Lys	Lys	Gln	Leu 100	Tyr	Ser	His	Cys	Arg 105	Asp	Gly	Gly	Trp	Ile 110	Leu	Gly
Tyr	Met	Ile 115	Asp	Gly	Val	Gln	Ala 120	Glu	Tyr	Val	Arg	Ile 125	Pro	His	Ala
Asp	Asn 130	Ser	Leu	Tyr	Гла	Ile 135	Pro	Gln	Thr	Ile	Asp 140	Asp	Glu	Ile	Ala
Val 145	Leu	Leu	Ser	Asp	Ile 150	Leu	Pro	Thr	Gly	His 155	Glu	Ile	Gly	Val	Gln 160

Tyr	Gly	Asn	Val	Gln 165	Pro	Gly	Asp	Ala	Val 170	Ala	Ile	Val	Gly	Ala 175	Gly
Pro	Val	Gly	Met 180	Ser	Val	Leu	Leu	Thr 185	Ala	Gln	Phe	Tyr	Ser 190	Pro	Ser
Thr	Ile	Ile 195	Val	Ile	Asp	Met	Asp 200	Glu	Asn	Arg	Leu	Gln 205	Leu	Ala	Lys
Glu	Leu 210	Gly	Ala	Thr	His	Thr 215	Ile	Asn	Ser	Gly	Thr 220	Glu	Asn	Val	Val
Glu 225	Ala	Val	His	Arg	Ile 230	Ala	Ala	Glu	Gly	Val 235	Asp	Val	Ala	Ile	Glu 240
Ala	Val	Gly	Ile	Pro 245	Ala	Thr	Trp	Asp	Ile 250	Суз	Gln	Glu	Ile	Val 255	Lys
Pro	Gly	Ala	His 260	Ile	Ala	Asn	Val	Gly 265	Val	His	Gly	Val	Lys 270	Val	Asp
Phe	Glu	Ile 275	Gln	Гла	Leu	Trp	Ile 280	Гла	Asn	Leu	Thr	Ile 285	Thr	Thr	Gly
Leu	Val 290	Asn	Thr	Asn	Thr	Thr 295	Pro	Met	Leu	Met	Lуа 300	Val	Ala	Ser	Thr
Asp 305	Lys	Leu	Pro	Leu	Lys 310	ГЛа	Met	Ile	Thr	His 315	Arg	Phe	Glu	Leu	Ala 320
Glu	Ile	Glu	His	Ala 325	Tyr	Gln	Val	Phe	Leu 330	Asn	Gly	Ala	Lys	Glu 335	Гла
Ala	Met	Lys	Ile 340	Ile	Leu	Ser	Asn	Ala 345	Gly	Ala	Ala				
<21	0> SH	EQ II	o no	140											
<40	0> SI	EQUEI	ICE :	140											
000															
<21) <213 <213	0> SH 1> LH 2> TY	EQ II ENGTH ZPE :	D NO H: 54 PRT	141 18			1								
< 40	3> 0F	COUEN	ICE ·	141	2000	cus	Taci	.15							
Met	Tyr	Thr	Val	Gly	Asp	Tyr	Leu	Leu	Asp	Arg	Leu	His	Glu	Leu 15	Gly
- Ile	Glu	Glu	Ile 20	Phe	Gly	Val	Pro	Gly 25	Asp	Tyr	Asn	Leu	Gln 30	Phe	Leu
Asp	Gln	Ile 35	Ile	Ser	His	Lys	Asp 40	Met	Гла	Trp	Val	Gly 45	Asn	Ala	Asn
Glu	Leu	Asn	Ala	Ser	Tvr	Met	Ala	Asp	Glv	Tvr	Ala	۸ra	Thr	Lvs	Lvs
	50				-1-	55		T	1	-1-	60	Arg		Lys	-
Ala 65	Ala	Ala	Phe	Leu	Thr 70	55 Thr	Phe	Gly	Val	Gly 75	60 Glu	Leu	Ser	Ala	Val 80
Ala 65 Asn	Ala Gly	Ala Leu	Phe Ala	Leu Gly 85	Thr 70 Ser	55 Thr Tyr	Phe Ala	Gly Glu	Val Asn 90	Gly 75 Leu	60 Glu Pro	Leu Val	Ser Val	Ala Glu 95	Val 80 Ile
Ala 65 Asn Val	Ala Gly Gly	Ala Leu Ser	Phe Ala Pro 100	Leu Gly 85 Thr	Thr 70 Ser Ser	55 Thr Tyr Lys	Phe Ala Val	Gly Glu Gln 105	Val Asn 90 Asn	Gly 75 Leu Glu	60 Glu Pro Gly	Leu Val Lys	Ser Val Phe 110	Ala Glu 95 Val	Val 80 Ile His
Ala 65 Asn Val His	Ala Gly Gly Thr	Ala Leu Ser Leu 115	Phe Ala Pro 100 Ala	Leu Gly 85 Thr Asp	Thr 70 Ser Ser Gly	55 Thr Tyr Lys Asp	Phe Ala Val Phe 120	Gly Glu Gln 105 Lys	Val Asn 90 Asn His	Gly 75 Leu Glu Phe	60 Glu Pro Gly Met	Leu Val Lys Lys 125	Ser Val Phe 110 Met	Ala Glu 95 Val His	Val 80 Ile His Glu
Ala 65 Asn Val His Pro	Ala Gly Gly Thr Val	Ala Leu Ser Leu 115 Thr	Phe Ala Pro 100 Ala Ala	Leu Gly 85 Thr Asp Ala	Thr 70 Ser Ser Gly Arg	55 Thr Tyr Lys Asp Thr 135	Phe Ala Val Phe 120 Leu	Gly Glu Gln 105 Lys Leu	Val Asn 90 Asn His Thr	Gly 75 Leu Glu Phe Ala	Glu Glu Pro Gly Met Glu 140	Leu Val Lys Lys 125 Asn	Ser Val Phe 110 Met Ala	Ala Glu 95 Val His Thr	Val 80 Ile His Glu Val

345

145					150					155					160
Tyr	Ile	Asn	Leu	Pro 165	Val	Asp	Val	Ala	Ala 170	Ala	Lys	Ala	Glu	Lys 175	Pro
Ser	Leu	Pro	Leu 180	Lys	Lys	Glu	Asn	Ser 185	Thr	Ser	Asn	Thr	Ser 190	Asb	Gln
Glu	Ile	Leu 195	Asn	Гла	Ile	Gln	Glu 200	Ser	Leu	Lys	Asn	Ala 205	Lys	Lys	Pro
Ile	Val 210	Ile	Thr	Gly	His	Glu 215	Ile	Ile	Ser	Phe	Gly 220	Leu	Glu	Lys	Thr
Val 225	Thr	Gln	Phe	Ile	Ser 230	ГЛа	Thr	Lys	Leu	Pro 235	Ile	Thr	Thr	Leu	Asn 240
Phe	Gly	Lys	Ser	Ser 245	Val	Asp	Glu	Ala	Leu 250	Pro	Ser	Phe	Leu	Gly 255	Ile
Tyr	Asn	Gly	Thr 260	Leu	Ser	Glu	Pro	Asn 265	Leu	Lys	Glu	Phe	Val 270	Glu	Ser
Ala	Asp	Phe 275	Ile	Leu	Met	Leu	Gly 280	Val	Lys	Leu	Thr	Asp 285	Ser	Ser	Thr
Gly	Ala 290	Phe	Thr	His	His	Leu 295	Asn	Glu	Asn	Lys	Met 300	Ile	Ser	Leu	Asn
Ile 305	Asp	Glu	Gly	Lys	Ile 310	Phe	Asn	Glu	Arg	Ile 315	Gln	Asn	Phe	Asp	Phe 320
Glu	Ser	Leu	Ile	Ser 325	Ser	Leu	Leu	Asp	Leu 330	Ser	Glu	Ile	Glu	Tyr 335	Lys
Gly	Lys	Tyr	Ile 340	Asp	Lys	Гла	Gln	Glu 345	Asp	Phe	Val	Pro	Ser 350	Asn	Ala
Leu	Leu	Ser 355	Gln	Aap	Arg	Leu	Trp 360	Gln	Ala	Val	Glu	Asn 365	Leu	Thr	Gln
Ser	Asn 370	Glu	Thr	Ile	Val	Ala 375	Glu	Gln	Gly	Thr	Ser 380	Phe	Phe	Gly	Ala
Ser 385	Ser	Ile	Phe	Leu	Lys 390	Ser	Lys	Ser	His	Phe 395	Ile	Gly	Gln	Pro	Leu 400
Trp	Gly	Ser	Ile	Gly 405	Tyr	Thr	Phe	Pro	Ala 410	Ala	Leu	Gly	Ser	Gln 415	Ile
Ala	Asp	Lys	Glu 420	Ser	Arg	His	Leu	Leu 425	Phe	Ile	Gly	Asp	Gly 430	Ser	Leu
Gln	Leu	Thr	420 Val	Gln	Glu	Leu	Gly	425 Leu	Ala	Ile	Arg	Glu	Lys	Ile	Asn
Pro	Ile	435 Сув	Phe	Ile	Ile	Asn	440 Asn	Asp	Gly	Tyr	Thr	445 Val	Glu	Arg	Glu
Ile	450 His	Gly	Pro	Asn	Gln	455 Ser	Tyr	Asn	Asp	Ile	460 Pro	Met	Trp	Asn	Tyr
465 Ser	Lys	Leu	Pro	Glu	470 Ser	Phe	Gly	Ala	Thr	475 Glu	Asp	Arg	Val	Val	480 Ser
Lys	Ile	Val	Arg	485 Thr	Glu	Asn	Glu	Phe	490 Val	Ser	Val	Met	Lys	495 Glu	Ala
Gln	Ala	Asn	500 Pro	Asn	Ara	Met	Tvr	505 Trp	Ile	Glu	Leu	Ile	510 Leu	Ala	Lvs
<u></u>		515	D	T	y	Lee	520		Mat	<i>a</i> 1-	u	525			-1~
GIU	GТХ 530	AIA	Pro	гЛа	va⊥	ьец 535	гЛа	гда	Met	сту	цуз 540	ьeu	rne	AIA	GLU
Gln 545	Asn	Lys	Ser												

<211> LENGTH: 375 <212> TYPE: PRT															
<213	8 > OF	RGANI	SM:	Equi	is ca	abal]	lus								
<400	)> SI	EQUEN	ICE :	142											
Met 1	Ser	Thr	Ala	Gly 5	Lys	Val	Ile	Lys	Cys 10	Lys	Ala	Ala	Val	Leu 15	Trp
Glu	Glu	Lys	Lys 20	Pro	Phe	Ser	Ile	Glu 25	Glu	Val	Glu	Val	Ala 30	Pro	Pro
Lys	Ala	His 35	Glu	Val	Arg	Ile	Lys 40	Met	Val	Ala	Thr	Gly 45	Ile	Cys	Arg
Ser	Asp 50	Asp	His	Val	Val	Ser 55	Gly	Thr	Leu	Val	Thr 60	Pro	Leu	Pro	Val
Ile 65	Ala	Gly	His	Glu	Ala 70	Ala	Gly	Ile	Val	Glu 75	Ser	Ile	Gly	Glu	Gly 80
Val	Thr	Thr	Val	Arg 85	Pro	Gly	Asp	Lys	Val 90	Ile	Pro	Leu	Phe	Thr 95	Pro
Gln	Суз	Gly	Lys 100	Суз	Arg	Val	Cys	Lys 105	His	Pro	Glu	Gly	Asn 110	Phe	Суз
Leu	Lys	Asn 115	Asp	Leu	Ser	Met	Pro 120	Arg	Gly	Thr	Met	Gln 125	Asp	Gly	Thr
Ser	Arg 130	Phe	Thr	CAa	Arg	Gly 135	Lys	Pro	Ile	His	His 140	Phe	Leu	Gly	Thr
Ser 145	Thr	Phe	Ser	Gln	Tyr 150	Thr	Val	Val	Asp	Glu 155	Ile	Ser	Val	Ala	Lys 160
Ile	Asp	Ala	Ala	Ser 165	Pro	Leu	Glu	Lys	Val 170	Суз	Leu	Ile	Gly	Cys 175	Gly
Phe	Ser	Thr	Gly 180	Tyr	Gly	Ser	Ala	Val 185	Lys	Val	Ala	Lys	Val 190	Thr	Gln
Gly	Ser	Thr 195	Сув	Ala	Val	Phe	Gly 200	Leu	Gly	Gly	Val	Gly 205	Leu	Ser	Val
Ile	Met 210	Gly	Суз	Lys	Ala	Ala 215	Gly	Ala	Ala	Arg	Ile 220	Ile	Gly	Val	Asp
Ile 225	Asn	Lys	Asp	ГÀа	Phe 230	Ala	Lys	Ala	Lys	Glu 235	Val	Gly	Ala	Thr	Glu 240
Суз	Val	Asn	Pro	Gln 245	Asp	Tyr	Lys	Гла	Pro 250	Ile	Gln	Glu	Val	Leu 255	Thr
Glu	Met	Ser	Asn 260	Gly	Gly	Val	Asp	Phe 265	Ser	Phe	Glu	Val	Ile 270	Gly	Arg
Leu	Asp	Thr 275	Met	Val	Thr	Ala	Leu 280	Ser	Суз	Суз	Gln	Glu 285	Ala	Tyr	Gly
Val	Ser 290	Val	Ile	Val	Gly	Val 295	Pro	Pro	Aab	Ser	Gln 300	Asn	Leu	Ser	Met
Asn 305	Pro	Met	Leu	Leu	Leu 310	Ser	Gly	Arg	Thr	Trp 315	Lys	Gly	Ala	Ile	Phe 320
Gly	Gly	Phe	Lys	Ser 325	Lys	Asp	Ser	Val	Pro 330	Lys	Leu	Val	Ala	Asp 335	Phe
Met	Ala	Lys	Lys 340	Phe	Ala	Leu	Asp	Pro 345	Leu	Ile	Thr	His	Val 350	Leu	Pro
Phe	Glu	Lys 355	Ile	Asn	Glu	Gly	Phe 360	Asp	Leu	Leu	Arg	Ser 365	Gly	Glu	Ser
Ile	Arg 370	Thr	Ile	Leu	Thr	Phe 375									

349

<210> SEQ ID NO 143 <211> LENGTH: 1206 <212> TYPE: DNA

<220> FEATURE:

<213> ORGANISM: Artificial Sequence

-continued

<223> OTHER INFORMATION: Aureobasidin A resistance (AUR1-C) <400> SEQUENCE: 143 atggcaaacc ctttttcgag atggtttcta tcagagagac ctccaaactg ccatgtagcc 60 gatttagaaa caagtttaga tccccatcaa acgttgttga aggtgcaaaa atacaaaccc 120 getttaageg actgggtgea ttacatette ttgggateea teatgetgtt tgtgtteatt 180 actaatcccg caccttggat cttcaagatc ctttttatt gtttcttggg cactttattc 240 atcatterag ctargtrara gtttttette aatgeettge ceatertaar atgggtggeg 300 ctgtatttca cttcatcgta ctttccagat gaccgcaggc ctcctattac tgtcaaagtg 360 ttaccagegg tggaaacaat tttataegge gacaatttaa gtgatattet tgeaacateg 420 acgaatteet ttttggacat tttageatgg ttaeegtaeg gaetatttea ttatggggee 480 ccatttgtcg ttgctgccat cttattcgta tttggtccac caactgtttt gcaaggttat 540 gcttttgcat ttggttatat gaacctgttt ggtgttatca tgcaaaatgt ctttccagcc 600 gctcccccat ggtataaaat tctctatgga ttgcaatcag ccaactatga tatgcatggc 660 tcgcctggtg gattagctag aattgataag ctactcggta ttaatatgta tactacatgt 720 ttttcaaatt cctccgtcat tttcggtgct tttccttcac tgcattccgg gtgtgctact 780 atggaageee tgtttttetg ttattgtttt ceaaaattga ageeettgtt tattgettat 840 gtttgctggt tatggtggtc aactatgtat ctgacacacc attattttgt agaccttatg 900 gcaggttctg tgctgtcata cgttattttc cagtacacaa agtacacaca tttaccaatt 960 gtagatacat ctcttttttg cagatggtca tacacttcaa ttgagaaata cgatatatca 1020 aagagtgatc cattggctgc agattcaaac gatatcgaaa gtgtcccttt gtccaacttg 1080 gaacttgact ttgatcttaa tatgactgat gaacccagtg taagcccttc gttatttgat 1140 ggatctactt ctgtttctcg ttcgtccgcc acgtctataa cgtcactagg tgtaaagagg 1200 gcttaa 1206 <210> SEQ ID NO 144 <211> LENGTH: 401 <212> TYPE: PRT <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Aureobasidin A resistance (AUR1-C) <400> SEQUENCE: 144 Met Ala Asn Pro Phe Ser Arg Trp Phe Leu Ser Glu Arg Pro Pro Asn 10 15 Cys His Val Ala Asp Leu Glu Thr Ser Leu Asp Pro His Gln Thr Leu 20 25 30 Leu Lys Val Gln Lys Tyr Lys Pro Ala Leu Ser Asp Trp Val His Tyr 40 45 Ile Phe Leu Gly Ser Ile Met Leu Phe Val Phe Ile Thr Asn Pro Ala 50 55 60 Pro Trp Ile Phe Lys Ile Leu Phe Tyr Cys Phe Leu Gly Thr Leu Phe 75 65 70 80 Ile Ile Pro Ala Thr Ser Gln Phe Phe Phe Asn Ala Leu Pro Ile Leu 85 90 95

Thr Trp Val Ala Leu Tyr Phe Thr Ser Ser Tyr Phe Pro Asp Asp Arg 100 105 110
Arg Pro Pro Ile Thr Val Lys Val Leu Pro Ala Val Glu Thr Ile Leu 115 120 125
Tyr Gly Asp Asn Leu Ser Asp Ile Leu Ala Thr Ser Thr Asn Ser Phe 130 135 140
Leu Asp Ile Leu Ala Trp Leu Pro Tyr Gly Leu Phe His Tyr Gly Ala 145 150 155 160
Pro Phe Val Val Ala Ala Ile Leu Phe Val Phe Gly Pro Pro Thr Val 165 170 175
Leu Gln Gly Tyr Ala Phe Ala Phe Gly Tyr Met Asn Leu Phe Gly Val 180 185 190
Ile Met Gln Asn Val Phe Pro Ala Ala Pro Pro Trp Tyr Lys Ile Leu 195 200 205
Tyr Gly Leu Gln Ser Ala Asn Tyr Asp Met His Gly Ser Pro Gly Gly 210 215 220
Leu Ala Arg Ile Asp Lys Leu Leu Gly Ile Asn Met Tyr Thr Thr Cys 225 230 235 240
Phe Ser Asn Ser Ser Val Ile Phe Gly Ala Phe Pro Ser Leu His Ser 245 250 255
Gly Cys Ala Thr Met Glu Ala Leu Phe Phe Cys Tyr Cys Phe Pro Lys 260 265 270
Leu Lys Pro Leu Phe Ile Ala Tyr Val Cys Trp Leu Trp Trp Ser Thr 275 280 285
Met Tyr Leu Thr His His Tyr Phe Val Asp Leu Met Ala Gly Ser Val 290 295 300
Leu Ser Tyr Val Ile Phe Gln Tyr Thr Lys Tyr Thr His Leu Pro Ile 305 310 315 320
Val Asp Thr Ser Leu Phe Cys Arg Trp Ser Tyr Thr Ser Ile Glu Lys 325 330 335
Tyr Asp Ile Ser Lys Ser Asp Pro Leu Ala Ala Asp Ser Asn Asp Ile 340 345 350
Glu Ser Val Pro Leu Ser Asn Leu Glu Leu Asp Phe Asp Leu Asn Met 355 360 365
Thr Asp Glu Pro Ser Val Ser Pro Ser Leu Phe Asp Gly Ser Thr Ser 370 375 380
Val Ser Arg Ser Ser Ala Thr Ser Ile Thr Ser Leu Gly Val Lys Arg 385 390 395 400
Ala
<210> SEQ ID NO 145 <211> LENGTH: 552 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: bialiphos resistance protein
<400> SEQUENCE: 145
atgageeeag aacgaegeee ggeegacate egeegtgeea eegaggegga catgeeggeg 60
gtetgeacea tegteaacea etacategag acaageaegg teaactteeg tacegageeg 120
caggaacege aggagtggae ggaegaeete gteegtetge gggagegeta teeetggete 180
accocctace actegacede egadtegace statacetet secondeca chaecedace 300

2	5	2
Э	Э	Э

## continued

ggactgggct ccacgctcta cacccacctg ctgaagtccc tggaggcaca gggcttcaag 360
agcgtggtcg ctgtcatcgg gctgcccaac gacccgagcg tgcgcatgca cgaggcgctc 420
ggatatgeee eeegeggeat getgegggeg geeggettea ageaegggaa etggeatgae 480
gtgggtttet ggeagetgga etteageetg eeggtaeege eeegteeggt eetgeeegte 540
accgagattt ga 552
<210> SEQ ID NO 146 <211> LENGTH: 183 <212> TYPE: PRT <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: bialiphos resistance protein
<400> SEQUENCE: 146
Met Ser Pro Glu Arg Arg Pro Ala Asp Ile Arg Arg Ala Thr Glu Ala 1 5 10 15
Asp Met Pro Ala Val Cys Thr Ile Val Asn His Tyr Ile Glu Thr Ser 20 25 30
Thr Val Asn Phe Arg Thr Glu Pro Gln Glu Pro Gln Glu Trp Thr Asp 35 40 45
Asp Leu Val Arg Leu Arg Glu Arg Tyr Pro Trp Leu Val Ala Glu Val 50 55 60
Asp Gly Glu Val Ala Gly Ile Ala Tyr Ala Gly Pro Trp Lys Ala Arg 65 70 75 80
Asn Ala Tyr Asp Trp Thr Ala Glu Ser Thr Val Tyr Val Ser Pro Arg 85 90 95
His Gln Arg Thr Gly Leu Gly Ser Thr Leu Tyr Thr His Leu Lys 100 105 110
Ser Leu Glu Ala Gln Gly Phe Lys Ser Val Val Ala Val Ile Gly Leu 115 120 125
Pro Asn Asp Pro Ser Val Arg Met His Glu Ala Leu Gly Tyr Ala Pro 130 135 140
Arg Gly Met Leu Arg Ala Ala Gly Phe Lys His Gly Asn Trp His Asp 145 150 155 160
Val Gly Phe Trp Gln Leu Asp Phe Ser Leu Pro Val Pro Pro Arg Pro 165 170 175
Val Leu Pro Val Thr Glu Ile 180
<210> SEQ ID NO 147 <211> LENGTH: 1953 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: cerulenin resistance YML007W Chr 13
<400> SEQUENCE: 147
atgagtgtgt ctaccgccaa gaggtcgctg gatgtcgttt ctccgggttc attagcggag 60
tttgagggtt caaaatctcg tcacgatgaa atagaaaatg aacatagacg tactggtaca 120
cgtgatggcg aggatagcga gcaaccgaag aagaagggta gcaaaactag caaaaagcaa 180
gatttggatc ctgaaactaa gcagaagagg actgcccaaa atcgggccgc tcaaagagct 240
tttagggaac gtaaggagag gaagatgaag gaattggaga agaaggtaca aagtttagag 300
agtattcagc agcaaaatga agtggaagct acttttttga gggaccagtt aatcactctg 360
gtgaatgagt taaaaaaata tagaccagag acaagaaatg actcaaaagt gctggaatat 420

355

ttagcaaggc gagatccta	a tttgcatttt	tcaaaaaata	acgttaacca	cagcaatagc	480
gagccaattg acacaccca	a tgatgacata	caagaaaatg	ttaaacaaaa	gatgaatttc	540
acgtttcaat atccgctto	a taacgacaac	gacaacgaca	acagtaaaaa	tgtggggaaa	600
caattacctt caccaaato	a tccaagtcat	tcggctccta	tgcctataaa	tcagacacaa	660
aagaaattaa gtgacgcta	c agatteetee	agcgctactt	tggattccct	ttcaaatagt	720
aacgatgttc ttaataaca	c accaaactcc	tccacttcga	tggattggtt	agataatgta	780
atatatacta acaggttto	t gtcaggtgat	gatggcagca	atagtaaaac	taagaattta	840
gacagtaata tgttttcta	a tgactttaat	tttgaaaacc	aatttgatga	acaagtttcg	900
gagttttgtt cgaaaatga	a ccaggtatgt	ggaacaaggc	aatgtcccat	tcccaagaaa	960
cccatctcgg ctcttgata	a agaagttttc	gcgtcatctt	ctatactaag	ttcaaattct	1020
cctgctttaa caaatactt	g ggaatcacat	tctaatatta	cagataatac	tcctgctaat	1080
gtcattgcta ctgatgcta	c taaatatgaa	aattccttct	ccggttttgg	ccgacttggt	1140
ttcgatatga gtgccaato	a ttacgtcgtg	aatgataata	gcactggtag	cactgatagc	1200
actggtagca ctggcaata	a gaacaaaaag	aacaataata	atagcgatga	tgtactccca	1260
ttcatatccg agtcaccgt	t tgatatgaac	caagttacta	attttttag	tccgggatct	1320
accggcatcg gcaataato	c tgcctctaac	accaatccca	gcctactgca	aagcagcaaa	1380
gaggatatac cttttatca	a cgcaaatctg	gctttcccag	acgacaattc	aactaatatt	1440
caattacaac ctttctctc	a atctcaatct	caaaataagt	ttgactacga	catgtttttt	1500
agagattcat cgaaggaag	g taacaattta	tttggagagt	ttttagagga	tgacgatgat	1560
gacaaaaaag ccgctaata	t gtcagacgat	gagtcaagtt	taatcaagaa	ccagttaatt	1620
aacgaagaac cagagette	c gaaacaatat	ctacaatcgg	taccaggaaa	tgaaagcgaa	1680
atctcacaaa aaaatggca	g tagtttacag	aatgctgaca	aaatcaataa	tggcaatgat	1740
aacgataatg ataatgatg	t cgttccatct	aaggaaggct	ctttactaag	gtgttcggaa	1800
atttgggata gaataacaa	c acatccgaaa	tactcagata	ttgatgtcga	tggtttatgt	1860
tccgagctaa tggcaaago	c aaaatgttca	gaaagagggg	ttgtcatcaa	tgcagaagac	1920
gttcaattag ctttgaata	a gcatatgaac	taa			1953
<pre>&lt;210&gt; SEQ ID NO 148 &lt;211&gt; LENGTH: 650 &lt;212&gt; TYPE: PRT &lt;213&gt; ORGANISM: Arti &lt;220&gt; FEATURE: &lt;223&gt; OTHER INFORMAT</pre>	ficial Seque ION: cerulen	nce in resistanc	ce YML007W	Chr 13	
<400> SEQUENCE: 148					
Met Ser Val Ser Thr 1 5	Ala Lys Arg	Ser Leu Asp 10	Val Val Sei	Pro Gly 15	
Ser Leu Ala Glu Phe 20	Glu Gly Ser	Lys Ser Arg 25	His Asp Glu 30	ı Ile Glu	
Asn Glu His Arg Arg 35	Thr Gly Thr 40	Arg Asp Gly	Glu Asp Sei 45	r Glu Gln	
Pro Lys Lys Lys Gly 50	Ser Lys Thr 55	Ser Lys Lys	Gln Asp Leu 60	ı Asp Pro	
Glu Thr Lys Gln Lys 65	Arg Thr Ala 70	Gln Asn Arg 75	Ala Ala Glr	n Arg Ala 80	
Phe Arg Glu Arg Lys	Glu Arg Lys	Met Lys Glu	Leu Glu Ly:	s Lys Val	

-continued
------------

				85					90					95	
Gln	Ser	Leu	Glu 100	Ser	Ile	Gln	Gln	Gln 105	Asn	Glu	Val	Glu	Ala 110	Thr	Phe
Leu	Arg	Asp 115	Gln	Leu	Ile	Thr	Leu 120	Val	Asn	Glu	Leu	Lys 125	Lys	Tyr	Arg
Pro	Glu 130	Thr	Arg	Asn	Asp	Ser 135	Lys	Val	Leu	Glu	Tyr 140	Leu	Ala	Arg	Arg
Asp 145	Pro	Asn	Leu	His	Phe 150	Ser	Lys	Asn	Asn	Val 155	Asn	His	Ser	Asn	Ser 160
Glu	Pro	Ile	Asp	Thr 165	Pro	Asn	Asp	Asp	Ile 170	Gln	Glu	Asn	Val	Lys 175	Gln
Lys	Met	Asn	Phe 180	Thr	Phe	Gln	Tyr	Pro 185	Leu	Asp	Asn	Asp	Asn 190	Asp	Asn
Asp	Asn	Ser 195	Lys	Asn	Val	Gly	Lys 200	Gln	Leu	Pro	Ser	Pro 205	Asn	Asp	Pro
Ser	His 210	Ser	Ala	Pro	Met	Pro 215	Ile	Asn	Gln	Thr	Gln 220	Lys	Lys	Leu	Ser
Asp 225	Ala	Thr	Asp	Ser	Ser 230	Ser	Ala	Thr	Leu	Asp 235	Ser	Leu	Ser	Asn	Ser 240
Asn	Asp	Val	Leu	Asn 245	Asn	Thr	Pro	Asn	Ser 250	Ser	Thr	Ser	Met	Asp 255	Trp
Leu	Asp	Asn	Val 260	Ile	Tyr	Thr	Asn	Arg 265	Phe	Val	Ser	Gly	Asp 270	Asp	Gly
Ser	Asn	Ser 275	Lys	Thr	ГЛа	Asn	Leu 280	Asp	Ser	Asn	Met	Phe 285	Ser	Asn	Asp
Phe	Asn 290	Phe	Glu	Asn	Gln	Phe 295	Asp	Glu	Gln	Val	Ser 300	Glu	Phe	Суз	Ser
Lys 305	Met	Asn	Gln	Val	Суз 310	Gly	Thr	Arg	Gln	Cys 315	Pro	Ile	Pro	Гла	Lys 320
Pro	Ile	Ser	Ala	Leu 325	Asp	Lys	Glu	Val	Phe 330	Ala	Ser	Ser	Ser	Ile 335	Leu
Ser	Ser	Asn	Ser 340	Pro	Ala	Leu	Thr	Asn 345	Thr	Trp	Glu	Ser	His 350	Ser	Asn
Ile	Thr	Asp 355	Asn	Thr	Pro	Ala	Asn 360	Val	Ile	Ala	Thr	Asp 365	Ala	Thr	Lys
Tyr	Glu 370	Asn	Ser	Phe	Ser	Gly 375	Phe	Gly	Arg	Leu	Gly 380	Phe	Asp	Met	Ser
Ala 385	Asn	His	Tyr	Val	Val 390	Asn	Asp	Asn	Ser	Thr 395	Gly	Ser	Thr	Asp	Ser 400
Thr	Gly	Ser	Thr	Gly 405	Asn	Lys	Asn	Lys	Lys 410	Asn	Asn	Asn	Asn	Ser 415	Aap
Asp	Val	Leu	Pro 420	Phe	Ile	Ser	Glu	Ser 425	Pro	Phe	Asp	Met	Asn 430	Gln	Val
Thr	Asn	Phe 435	Phe	Ser	Pro	Gly	Ser 440	Thr	Gly	Ile	Gly	Asn 445	Asn	Ala	Ala
Ser	Asn 450	Thr	Asn	Pro	Ser	Leu 455	Leu	Gln	Ser	Ser	Lys 460	Glu	Asp	Ile	Pro
Phe 465	Ile	Asn	Ala	Asn	Leu 470	Ala	Phe	Pro	Asp	Asp 475	Asn	Ser	Thr	Asn	Ile 480
Gln	Leu	Gln	Pro	Phe 485	Ser	Glu	Ser	Gln	Ser 490	Gln	Asn	ГЛа	Phe	Asp 495	Tyr
Asp	Met	Phe	Phe 500	Arg	Asp	Ser	Ser	Lys 505	Glu	Gly	Asn	Asn	Leu 510	Phe	Gly

-continued

Glu Phe Leu Glu Asp Asp Asp Asp Asp Lys Lys Ala Ala Asn Met Ser 515 520 525	
Asp Asp Glu Ser Ser Leu Ile Lys Asn Gln Leu Ile Asn Glu Glu Pro 530 535 540	
Glu Leu Pro Lys Gln Tyr Leu Gln Ser Val Pro Gly Asn Glu Ser Glu 545 550 555 560	
Ile Ser Gln Lys Asn Gly Ser Ser Leu Gln Asn Ala Asp Lys Ile Asn 565 570 575	
Asn Gly Asn Asp Asn Asp Asn Asp Asn Asp Val Val Pro Ser Lys Glu 580 585 590	
Gly Ser Leu Leu Arg Cys Ser Glu Ile Trp Asp Arg Ile Thr Thr His 595 600 605	
Pro Lys Tyr Ser Asp Ile Asp Val Asp Gly Leu Cys Ser Glu Leu Met 610 615 620	
Ala Lys Ala Lys Cys Ser Glu Arg Gly Val Val Ile Asn Ala Glu Asp 625 630 635 640	
Val Gln Leu Ala Leu Asn Lys His Met Asn 645 650	
<210> SEQ ID NO 149 <211> LENGTH: 810 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Geneticin resistance (kanMX) <400> SEQUENCE: 149	
(HOUN DEQUERCE. 14)	
atgggtaagg aaaagactca cgtttcgagg ccgcgattaa attccaacat ggatgctgat (	50
ttatatgggt ataaatgggc tcgcgataat gtcgggcaat caggtgcgac aatctatcga 12	20
rigranggga ageeegange geeagaging ninengaaae anggeaaagg nagegingee in	30
aatgatgtta cagatgagat ggtcagacta aactggctga cggaatttat gcctcttccg 24	40
accatcaagc attitateeg taeteetgat gatgeatggt taeteaceae tgegateeee 30	00
ggcaaaacag catteeaggt attagaagaa tateetgatt caggtgaaaa tattgttgat 36	60
gcgctggcag tgttcctgcg ccggttgcat tcgattcctg tttgtaattg tccttttaac 42	20
agegategeg tatttegtet egeteaggeg eaateaegaa tgaataaegg tttggttgat 48	80
gcgagtgatt ttgatgacga gcgtaatggc tggcctgttg aacaagtctg gaaagaaatg 54	40
cataagettt tgecattete aceggattea gtegteacte atggtgattt eteacttgat 60	00
aaccttattt ttgacgaggg gaaattaata ggttgtattg atgttggacg agtcggaatc 66	60
gcagaccgat accaggatet tgecateeta tggaactgee teggtgagtt tteteettea 72	20
ttacagaaac ggctttttca aaaatatggt attgataatc ctgatatgaa taaattgcag 75	80
tttcatttga tgctcgatga gtttttctaa 8:	10
<210> SEQ ID NO 150 <211> LENGTH: 269 <212> TYPE: PRT <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Geneticin resistance (kanMX) <400> SEQUENCE: 150	
Met Gly Lys Glu Lys Thr His Val Ser Arg Pro Arg Leu Asn Ser Asn 1 5 10 15	

## -continued

Met Asp Ala Asp Leu Tyr Gly Tyr Lys Trp Ala Arg Asp Asn Val Gly 20 25 30
Gln Ser Gly Ala Thr Ile Tyr Arg Leu Tyr Gly Lys Pro Asp Ala Pro
Glu Leu Phe Leu Lys His Gly Lys Gly Ser Val Ala Asn Asp Val Thr
50 55 60
Asp Glu Met Val Arg Leu Asn Trp Leu Thr Glu Phe Met Pro Leu Pro 65 70 75 80
Thr Ile Lys His Phe Ile Arg Thr Pro Asp Asp Ala Trp Leu Leu Thr 85 90 95
Thr Ala Ile Pro Gly Lys Thr Ala Phe Gln Val Leu Glu Glu Tyr Pro 100 105 110
Asp Ser Gly Glu Asn Ile Val Asp Ala Leu Ala Val Phe Leu Arg Arg
Leu His Ser Ile Pro Val Cys Asn Cys Pro Phe Asn Ser Asp Arg Val
130 135 140 Phe Arg Leu Ala Gln Ala Gln Ser Arg Met Asn Asn Glv Leu Val Asp
145 150 155 160
Ala Ser Asp Phe Asp Asp Glu Arg Asn Gly Trp Pro Val Glu Gln Val 165 170 175
Trp Lys Glu Met His Lys Leu Leu Pro Phe Ser Pro Asp Ser Val Val 180 185 190
Thr His Gly Asp Phe Ser Leu Asp Asn Leu Ile Phe Asp Glu Gly Lys 195 200 205
Leu Ile Gly Cys Ile Asp Val Gly Arg Val Gly Ile Ala Asp Arg Tyr
210 215 220 Gln Asp Leu Ala Ile Leu Trp Asp Cys Leu Gly Glu Phe Ser Pro Ser
225     230     235     240
Leu Gln Lys Arg Leu Phe Gln Lys Tyr Gly Ile Asp Asn Pro Asp Met 245 250 255
Asn Lys Leu Gln Phe His Leu Met Leu Asp Glu Phe Phe 260 265
<210> SEQ ID NO 151
<211> LENGTH: 1026 <212> TYPE: DNA
<213> ORGANISM: Artificial Sequence <220> FEATURE: <222> OFMER INFORMATION, Hygromygin R registance (HygR)
<223> OTHER INFORMATION: HygromyClin & resiscance (Hygr)
atgaaaaagc ctgaactcac cgcgacgtct gtcgagaagt ttctgatcga aaagttcgac 60
agegteteeg acetgatgea geteteggag ggegaagaat etegtgettt eagettegat 120
gtaggagggc gtggatatgt cctgcgggta aatagctgcg ccgatggttt ctacaaagat 180
cgttatgttt atcggcactt tgcatcggcc gcgctcccga ttccggaagt gcttgacatt 240
ggggagttca gcgagagcct gacctattgc atctcccgcc gtgcacaggg tgtcacgttg 300
caagacctgc ctgaaaccga actgcccgct gttctccagc cggtcgcgga ggccatggat 360
gcgatcgctg cggccgatct tagccagacg agcgggttcg gcccattcgg accgcaagga 420
atcggtcaat acactacatg gcgtgatttc atatgcgcga ttgctgatcc ccatgtgtat 480
cactggcaaa ctgtgatgga cgacaccgtc agtgcgtccg tcgcgcaggc tctcgatgag 540
ctgatgcttt gggccgagga ctgccccgaa gtccggcacc tcgtgcatgc ggatttcggc 600
tccaacaatg tcctgacgga caatggccgc ataacagcgg tcattgactg gagcgaggcg 660

363

364

atgttcgggg att	cccaata cgaggtcgcc	aacatettet tetggaggee gtggtt	gget 720
tgtatggagc ago	agacgcg ctacttcgag	cggaggcatc cggagcttgc aggato	gccg 780
cgcctccggg cgt	atatgct ccgcattggt	cttgaccaac tctatcagag cttggt	tgac 840
ggcaatttcg atg	atgcagc ttgggcgcag	ggtcgatgcg acgcaatcgt ccgato	ccgga 900
gccgggactg tcg	ggcgtac acaaatcgcc	cgcagaagcg cggccgtctg gaccga	atggc 960
tgtgtagaag tac	tcgccga tagtggaaac	cgacgcccca gcactcgtcc gagggo	caaag 1020
gaatag			1026
<pre>&lt;210&gt; SEQ ID N &lt;211&gt; LENGTH: &lt;212&gt; TYPE: PR &lt;213&gt; ORGANISM &lt;220&gt; FEATURE: &lt;223&gt; OTHER IN</pre>	O 152 341 T : Artificial Seque FORMATION: Hygromy	nce cin B resistance (HygR)	
<400> SEQUENCE	: 152		
Met Lys Lys Pr 1	o Glu Leu Thr Ala 5	Thr Ser Val Glu Lys Phe Leu I 10 15	lle
Glu Lys Phe As 20	p Ser Val Ser Asp	Leu Met Gln Leu Ser Glu Gly ( 25	Ju
Glu Ser Arg Al 35	a Phe Ser Phe Asp 40	Val Gly Gly Arg Gly Tyr Val I 45	Jeu
Arg Val Asn Se 50	r Cys Ala Asp Gly 55	Phe Tyr Lys Asp Arg Tyr Val 3 60	Tyr
Arg His Phe Al 65	a Ser Ala Ala Leu 1 70	Pro Ile Pro Glu Val Leu Asp I 75 8	11e 30
Gly Glu Phe Se	r Glu Ser Leu Thr 85	Tyr Cys Ile Ser Arg Arg Ala ( 90 95	Jln
Gly Val Thr Le	u Gln Asp Leu Pro 9 0	Glu Thr Glu Leu Pro Ala Val I 105 110	Jeu
Gln Pro Val Al	a Glu Ala Met Asp . 120	Ala Ile Ala Ala Ala Asp Leu S	Ser
Gln Thr Ser Gl	y Phe Gly Pro Phe	Gly Pro Gln Gly Ile Gly Gln :	Yr
Thr Thr Trp Ar	g Asp Phe Ile Cys	140 Ala Ile Ala Asp Pro His Val 1	lyr
145 His Trp Gln Th	150 r Val Met Asp Asp	155 Thr Val Ser Ala Ser Val Ala (	50 51n
Ala Leu Asp Gl	165 u Leu Met Leu Trp .	170 175 Ala Glu Asp Cys Pro Glu Val <i>2</i>	<i>j</i> ra
18 His Leu Val Hi	0 s Ala Asp Phe Glv	185 190 Ser Asn Asn Val Leu Thr Asn 2	\sn
195	200	205	
Gly Arg Ile Th 210	r Ala Val Ile Asp 215	Trp Ser Glu Ala Met Phe Gly 2 220	yab
Ser Gln Tyr Gl 225	u Val Ala Asn Ile 230	Phe Phe Trp Arg Pro Trp Leu 2 235 2	Ala 240
Cys Met Glu Gl	n Gln Thr Arg Tyr 245	Phe Glu Arg Arg His Pro Glu I 250 255	eu
Ala Gly Ser Pr 26	o Arg Leu Arg Ala 0	Tyr Met Leu Arg Ile Gly Leu 2 265 270	yab
Gln Leu Tyr Gl 275	n Ser Leu Val Asp 280	Gly Asn Phe Asp Asp Ala Ala 7 285	Trp

## -continued

Ala Gln Gly Arg Cys Asp Ala Ile Val Arg Ser Gly Ala Gly Thr Val 290 295 300
Gly Arg Thr Gln Ile Ala Arg Arg Ser Ala Ala Val Trp Thr Asp Gly 305 310 315 320
Cys Val Glu Val Leu Ala Asp Ser Gly Asn Arg Arg Pro Ser Thr Arg 325 330 335
Pro Arg Ala Lys Glu 340
<210> SEQ ID NO 153 <211> LENGTH: 573 <212> TYPE: DNA <213> ORGANISM: Streptomyces noursei
<400> SEQUENCE: 153
atgaccactc ttgacgacac ggcttaccgg taccgcacca gtgtcccggg ggacgccgag 60
gccatcgagg cactggatgg gtccttcacc accgacaccg tcttccgcgt caccgccacc 120
ggggacgget teaceetgeg ggaggtgeeg gtggaceege eeetgaceaa ggtgtteeee 180
gacgacgaat cggacgacga atcggacgac ggggaggacg gcgacccgga ctcccggacg 240
ttcgtcgcgt acggggacga cggcgacctg gcgggcttcg tggtcatctc gtactcggcg 300
tggaaccgcc ggctgaccgt cgaggacatc gaggtcgccc cggagcaccg ggggcacggg 360
gtcgggcgcg cgttgatggg gctcgcgacg gagttcgccg gcgagcgggg cgccggggcac 420
ctctggctgg aggtcaccaa cgtcaacgca ccggcgatcc acgcgtaccg gcggatgggg 480
ttcaccctct gcggcctgga caccgccctg tacgacggca ccgcctcgga cggcgagcgg 540
caggegetet acatgageat geeetgeeee tag 573
<210> SEQ ID NO 154 <211> LENGTH: 190 <212> TYPE: PRT <213> ORGANISM: Streptomyces noursei
<400> SEQUENCE: 154
Met Thr Thr Leu Asp Asp Thr Ala Tyr Arg Tyr Arg Thr Ser Val Pro 1 5 10 15
Gly Asp Ala Glu Ala Ile Glu Ala Leu Asp Gly Ser Phe Thr Thr Asp 20 25 30
Thr Val Phe Arg Val Thr Ala Thr Gly Asp Gly Phe Thr Leu Arg Glu 35 40 45
Val Pro Val Asp Pro Pro Leu Thr Lys Val Phe Pro Asp Asp Glu Ser 50 55 60
Asp Asp Glu Ser Asp Asp Gly Glu Asp Gly Asp Pro Asp Ser Arg Thr 65 70 75 80
Phe Val Ala Tyr Gly Asp Asp Gly Asp Leu Ala Gly Phe Val Val Ile 85 90 95
Ser Tyr Ser Ala Trp Asn Arg Arg Leu Thr Val Glu Asp Ile Glu Val 100 105 110
Ala Pro Glu His Arg Gly His Gly Val Gly Arg Ala Leu Met Gly Leu 115 120 125
Ala Thr Glu Phe Ala Gly Glu Arg Gly Ala Gly His Leu Trp Leu Glu 130 135 140
Val Thr Asn Val Asn Ala Pro Ala Ile His Ala Tyr Arg Arg Met Gly 145 150 155 160

	-continued	
165	170 175	
Asp Gly Glu Arg Gln Ala Leu Tyr Met 180 181	t Ser Met Pro Cys Pro 5 190	
<210> SEQ ID NO 155 <211> LENGTH: 375 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Phleomycin	e n zeocin binding protein	
<400> SEQUENCE: 155		
atggccaagt tgaccagtgc cgttccggtg c	tcaccgcgc gcgacgtcgc cggagcggtc	60
gagttetgga eegaeegget egggttetee eg	gggacttcg tggaggacga cttcgccggt	120
gtggtccggg acgacgtgac cctgttcatc ag	gegeggtee aggaeeaggt ggtgeeggae	180
aacaccctgg cctgggtgtg ggtgcgcggc c	tggacgagc tgtacgccga gtggtcggag	240
gtcgtgtcca cgaacttccg ggacgcctcc g	ggccggcca tgaccgagat cggcgagcag	300
ccgtggggggc gggagttcgc cctgcgcgac co	cggccggca actgcgtgca cttcgtggcc	360
gaggagcagg actga		375
<210> SEQ ID NO 156 <211> LENGTH: 124 <212> TYPE: PRT <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Phleomycin	e n zeocin binding protein	
<400> SEQUENCE: 156		
Met Ala Lys Leu Thr Ser Ala Val Pro 1 5	o Val Leu Thr Ala Arg Asp Val 10 15	
Ala Gly Ala Val Glu Phe Trp Thr Asp 20 25	p Arg Leu Gly Phe Ser Arg Asp 30	
Phe Val Glu Asp Asp Phe Ala Gly Va 35 40	l Val Arg Asp Asp Val Thr Leu 45	
Phe Ile Ser Ala Val Gln Asp Gln Va 50 55	l Val Pro Asp Asn Thr Leu Ala 60	
Trp Val Trp Val Arg Gly Leu Asp Gl 65 70	u Leu Tyr Ala Glu Trp Ser Glu 75 80	
Val Val Ser Thr Asn Phe Arg Asp Al. 85	a Ser Gly Pro Ala Met Thr Glu 90 95	
Ile Gly Glu Gln Pro Trp Gly Arg Glu 100 10!	u Phe Ala Leu Arg Asp Pro Ala 5 110	
Gly Asn Cys Val His Phe Val Ala Glu 115 120	u Glu Gln Asp	
<210> SEQ ID NO 157 <211> LENGTH: 1539 <212> TYPE: DNA <213> ORGANISM: Aspergillus niger		
<400> SEQUENCE: 157		
atggcattac ttgctgtcgc aggcgtctac g	ctttcgcag cgttgctcgt tgcgatagtc	60
ttaaatgtca cgcgccaatt gctctttcgc a	acgagaaag aaccacccgt cgtcttccat	120
tggatcccct tcttgggaag cacaatcagc ta	atggaatgg acccctatac attcttcttc	180
teetgeagaa aaaagtaegg ggacatette a	ccttcgtgc ttctgggcca gaagaccact	240

369

gtatacttgg gcgttcaagg caacgatttc atcctcaatg gcaaactcaa ggacgtgag	c 300
goggaagagg tetacageee eetcaceace eeggtgtteg ggteegatgt tgtgtaega	c 360
tgccctaatt ccaagctgat ggagcaaaaa aagttcatca agtttggcct cacgcaagc	g 420
gcgctcgagt cacacgtcca gctgatcgaa aaggaaactc tcgactatct ccgggactc	t 480
ccacgettea aeggegegag tggagteatt gatatteetg etgeeatgge tgagattae	a 540
atctatactg ctgcgcgcgc gttgcagggc gaggaggtcc gcaagaagct cacggcaga	g 600
ttcgctgaac tgtaccacga tctagacaag ggattcagcc ccattaactt catgctccc	t 660
tgggctccat tgccgcacaa ccggaagcgt gatgctgctc atgctcggat gagagaaat	c 720
tacacggaca ttatcaacga acgacgcaag aacccagacg aggagaagtc agacatgat	c 780
tggaatetga tgcattgcae etacaagagt ggecageegg teeeggacaa agagattge	t 840
cacatgatga tcactctgtt gatggcaggg caacactcgt cttcttcgat tagttcttg	g 900
atcatgctgc gattggcctc ggagcctcag gtgcttgaag agctctacca agaacagct	g 960
gccagcetta gcaacagaaa tggagtette gageegetge agtateagga eettgacaa	g 1020
ctgccattcc tccagagtgt catcaaggag actctacgga tccactcgtc catccactc	g 1080
atcatgegea aggtgaaaaa eeegetaeea gtaeetggea eeteetaeat tatteeega	a 1140
gaccatgttc tactcgcctc accaggcgta accgcgctta gtgacgaata ctttcctaa	c 1200
gcaaccaggt gggatccgca tcgttgggag aatcagcctg acaaagagga ggatggaga	g 1260
atggtggact acggatatgg cagcgtgtcg aagggcactg ctagtcccta tctaccttt	t 1320
ggcgctggcc gtcaccgctg cattggagag aagttcgcct acgtcaactt gggcgtcat	t 1380
atcgcgacca tagtgcgcca cttgaagcta ttcaatgtgg atggcaggaa aggagtgcc	c 1440
ggaaccgatt actcgaccct cttctccggt cccatgaagc ctgctatagt gggttggga	g 1500
cgacgcttcc cggacaacat caaagggtcc atgaactaa	1539
<210> SEQ ID NO 158 <211> LENGTH: 4519 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: pLA54	
<400> SEQUENCE: 158	
cacettgget aactegttgt ateateactg gataactteg tataatgtat getataega	a 60
gttatcgaac agagaaacta aatccacatt aattgagagt tctatctatt agaaaatgc	a 120
aactccaact aaatgggaaa acagataacc tcttttattt ttttttaatg tttgatatt	c 180
gagtettttt ettttgttag gtttatatte ateattteaa tgaataaaag aagettett	a 240
ttttggttgc aaagaatgaa aaaaaaggat tttttcatac ttctaaagct tcaattata	a 300
ccaaaaattt tataaatgaa gagaaaaaat ctagtagtat caagttaaac ttagaaaaa	c 360
tcatcgagca tcaaatgaaa ctgcaattta ttcatatcag gattatcaat accatattt	t 420
tgaaaaagcc gtttctgtaa tgaaggagaa aactcaccga ggcagttcca taggatggc	a 480
agateetggt ateggtetge gatteegaet egteeaacat caatacaaee tattaattt	c 540
ccctcgtcaa aaataaggtt atcaagtgag aaatcaccat qaqtqacqac tqaatccqq	t 600
gagaatggca aaagcttatg cattlette cagacttatt caacaggcca gooattag	c 660
tegteateaa aateactege ateaaceaaa centattea ttegtgatte	a 720
agacdaaata condatoot ottaaaadda caattacaaa caogaatoo atoosadoo	g 780
uyuuyuuuuu uyuyatuyet yttaaaayya CaattaCaaa Cayyaatuya algCadCCg	2 /00

371

840	ttcttctaat	aatcaggata	ttttcacctg	atcaacaata	ctgccagcgc	cgcaggaaca
900	atcaggagta	accatgcatc	gtggtgagta	ggggatcgca	ctgttttgcc	acctggaatg
960	tagtctgacc	tcagccagtt	ataaattccg	cggaagaggc	gcttgatggt	cggataaaat
1020	caactctggc	gtttcagaaa	cctttgccat	ggcaacgcta	taacatcatt	atctcatctg
1080	attatcgcga	attgcccgac	gtcgcacctg	tcgatagatt	tcccatacaa	gcatcgggct
1140	cctcgaaacg	ttaatcgcgg	atgttggaat	atcagcatcc	acccatataa	gcccatttat
1200	tagggaacta	ctcttgcagt	aatgttactt	ctcgagtttt	ccttacccat	tgagtettt
1260	cagaaaaacc	agaagttata	taaaataaaa	ttaaacaaac	caaaataaga	taatgtaact
1320	ataaaaccag	aaactatcaa	atacacaaaa	cataataata	agtactaatc	catataaacc
1380	tcgatatatg	tattcaaaat	tctcctttta	attttttcga	gaatagaaaa	aaaacagatt
1440	tttccttcta	agatttttct	caatttaatt	atcaccaaat	ctctcagaaa	aaaaagggaa
1500	gccgtaggaa	aaaaattttt	tttagaaatg	tcttttttt	gaaaaatttt	gcgttggaaa
1560	tgattgaccc	gtgtaggcta	ctgttcacct	aaacgctact	aaccctgtat	tcaccgtata
1620	atccatgctg	gatacaagag	gaaaagaacc	gagagcggga	gttattgcga	agtgttcatt
1680	atttagacga	gatgatgtgt	cttgtaggac	ctttgatgaa	ctgtccaaca	gtatagttgt
1740	ttctgtgtaa	gtacggtgtg	agagaggttt	tagttatgat	gactattaag	gtacgtgtgt
1800	gctatacgaa	tataatgtat	gataacttcg	tgaatcccta	aaaatggtta	ttcgattgag
1860	tcgacctgca	tcctctagag	aatgcgggga	cgtaatccgc	cattagaata	gttatctgaa
1920	tgttatccgc	tgtgtgaaat	agctgtttcc	tcatggtcat	cttggcgtaa	ggcatgcaag
1980	ggtgcctaat	taaagcctgg	gcataaagtg	cgagccggaa	acacaacata	tcacaattcc
2040	tcgggaaacc	cgctttccag	gctcactgcc	attgcgttgc	actcacatta	gagtgagcta
2100	ttgcgtattg	gagaggcggt	aacgcgcggg	tgaatcggcc	gctgcattaa	tgtcgtgcca
2160	ctgcggcgag	ggtcgttcgg	cgctgcgctc	ctcactgact	cgcttcctcg	ggcgctcttc
2220	gataacgcag	agaatcaggg	ggttatccac	gcggtaatac	tcactcaaag	cggtatcagc
2280	gccgcgttgc	ccgtaaaaag	aggccaggaa	ggccagcaaa	gtgagcaaaa	gaaagaacat
2340	cgctcaagtc	caaaaatcga	acgagcatca	cgcccccctg	ccataggctc	tggcgttttt
2400	ggaagctccc	gtttccccct	gataccaggc	ggactataaa	aaacccgaca	agaggtggcg
2460	tttctccctt	cctgtccgcc	ttaccggata	accctgccgc	tcctgttccg	tcgtgcgctc
2520	gtgtaggtcg	tctcagttcg	gctgtaggta	catagctcac	ggcgctttct	cgggaagcgt
2580	tgcgccttat	gcccgaccgc	cccccgttca	gtgcacgaac	gctgggctgt	ttcgctccaa
2640	ctggcagcag	cttatcgcca	taagacacga	tccaacccgg	tcgtcttgag	ccggtaacta
2700	ttcttgaagt	tgctacagag	atgtaggcgg	agagcgaggt	caggattagc	ccactggtaa
2760	ctgctgaagc	tatctgcgct	cagtatttgg	actagaagga	ctacggctac	ggtggcctaa
2820	accgctggta	caaacaaacc	cttgatccgg	gttggtagct	cggaaaaaga	cagttacctt
2880	tctcaagaag	aaaaaagga	ttacgcgcag	aagcagcaga	ttttgtttgc	gcggtggttt
2940	cgttaaggga	cgaaaactca	ctcagtggaa	gggtctgacg	cttttctacg	atcctttgat
3000	taaaaatgaa	ccttttaaat	tcacctagat	aaaaggatct	gagattatca	ttttggtcat
3060	caatgcttaa	tgacagttac	aaacttggtc	atatatgagt	aatctaaagt	gttttaaatc
3120	gcctgactcc	atccatagtt	tatttcgttc	gcgatctgtc	acctatctca	tcagtgaggc

COD	÷	п.	n	11	Δ	<u>a</u>
COIL	-	1	. 1 1	u	-	u

-continued	
ccgtcgtgta gataactacg atacgggagg gcttaccatc tggccccagt gctgcaatga	3180
taccgcgaga cccacgctca ccggctccag atttatcagc aataaaccag ccagccggaa	3240
gggccgagcg cagaagtggt cctgcaactt tatccgcctc catccagtct attaattgtt	3300
gccgggaagc tagagtaagt agttcgccag ttaatagttt gcgcaacgtt gttgccattg	3360
ctacaggcat cgtggtgtca cgctcgtcgt ttggtatggc ttcattcagc tccggttccc	3420
aacgatcaag gcgagttaca tgatccccca tgttgtgcaa aaaagcggtt agctccttcg	3480
gtcctccgat cgttgtcaga agtaagttgg ccgcagtgtt atcactcatg gttatggcag	3540
cactgcataa ttctcttact gtcatgccat ccgtaagatg cttttctgtg actggtgagt	3600
actcaaccaa gtcattctga gaatagtgta tgcggcgacc gagttgctct tgcccggcgt	3660
caatacggga taataccgcg ccacatagca gaactttaaa agtgctcatc attggaaaac	3720
gttetteggg gegaaaaete teaaggatet taeegetgtt gagateeagt tegatgtaae	3780
ccactcgtgc acccaactga tettcageat ettttaettt eaccagegtt tetgggtgag	3840
caaaaacagg aaggcaaaat gccgcaaaaa agggaataag ggcgacacgg aaatgttgaa	3900
tactcatact cttccttttt caatattatt gaagcattta tcagggttat tgtctcatga	3960
geggatacat atttgaatgt atttagaaaa ataaacaaat aggggtteeg egeacattte	4020
cccgaaaagt gccacctgac gtctaagaaa ccattattat catgacatta acctataaaa	4080
ataggegtat caegaggeee tttegteteg egegtttegg tgatgaeggt gaaaaeetet	4140
gacacatgca gctcccggag acggtcacag cttgtctgta agcggatgcc gggagcagac	4200
aageeegtea gggegegtea gegggtgttg gegggtgteg gggetggett aaetatgegg	4260
catcagagca gattgtactg agagtgcacc atatgcggtg tgaaataccg cacagatgcg	4320
taaggagaaa ataccgcatc aggcgccatt cgccattcag gctgcgcaac tgttgggaag	4380
ggcgatcggt gcgggcctct tcgctattac gccagctggc gaaaggggga tgtgctgcaa	4440
ggcgattaag ttgggtaacg ccagggtttt cccagtcacg acgttgtaaa acgacggcca	4500
gtgaattega geteggtae	4519
<210> SEQ ID NO 159 <211> LENGTH: 25 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: 749 <400> SEQUENCE: 159	
caagtetttt gtgeetteee gtegg	25
<210> SEQ ID NO 160 <211> LENGTH: 25 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: 413	
<400> SEQUENCE: 160	
ggacataaaa tacacaccga gattc	25
<210> SEQ ID NO 161 <211> LENGTH: 38 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: LA468	

<400> SEQUI	<400> SEQUENCE: 161								
gcctcgagtt	ttaatgttac	ttctcttgca	gttaggga			38			
<210> SEQ : <211> LENG <212> TYPE <213> ORGAN <220> FEATU <223> OTHEN	<pre>&lt;210&gt; SEQ ID NO 162 &lt;211&gt; LENGTH: 10934 &lt;212&gt; TYPE: DNA &lt;213&gt; ORGANISM: Artificial Sequence &lt;220&gt; FEATURE: &lt;223&gt; OTHER INFORMATION: pRS423::TEF(M4)-xpk1+EN01-eutD</pre>								
<400> SEQUI	ENCE: 162								
ggtggagctc	cagcttttgt	tccctttagt	gagggttaat	tgcgcgcttg	gcgtaatcat	60			
ggtcatagct	gtttcctgtg	tgaaattgtt	atccgctcac	aattccacac	aacataggag	120			
ccggaagcat	aaagtgtaaa	gcctggggtg	cctaatgagt	gaggtaactc	acattaattg	180			
cgttgcgctc	actgcccgct	ttccagtcgg	gaaacctgtc	gtgccagctg	cattaatgaa	240			
tcggccaacg	cgcggggaga	ggcggtttgc	gtattgggcg	ctcttccgct	tcctcgctca	300			
ctgactcgct	gcgctcggtc	gttcggctgc	ggcgagcggt	atcagctcac	tcaaaggcgg	360			
taatacggtt	atccacagaa	tcaggggata	acgcaggaaa	gaacatgtga	gcaaaaggcc	420			
agcaaaaggc	caggaaccgt	aaaaaggccg	cgttgctggc	gtttttccat	aggeteegee	480			
cccctgacga	gcatcacaaa	aatcgacgct	caagtcagag	gtggcgaaac	ccgacaggac	540			
tataaagata	ccaggcgttt	ccccctggaa	gctccctcgt	gcgctctcct	gttccgaccc	600			
tgccgcttac	cggatacctg	tccgcctttc	tcccttcggg	aagcgtggcg	ctttctcata	660			
gctcacgctg	taggtatctc	agttcggtgt	aggtcgttcg	ctccaagctg	ggctgtgtgc	720			
acgaaccccc	cgttcagccc	gaccgctgcg	ccttatccgg	taactatcgt	cttgagtcca	780			
acccggtaag	acacgactta	tcgccactgg	cagcagccac	tggtaacagg	attagcagag	840			
cgaggtatgt	aggcggtgct	acagagttct	tgaagtggtg	gcctaactac	ggctacacta	900			
gaaggacagt	atttggtatc	tgcgctctgc	tgaagccagt	taccttcgga	aaaagagttg	960			
gtagctcttg	atccggcaaa	caaaccaccg	ctggtagcgg	tggtttttt	gtttgcaagc	1020			
agcagattac	gcgcagaaaa	aaaggatctc	aagaagatcc	tttgatcttt	tctacggggt	1080			
ctgacgctca	gtggaacgaa	aactcacgtt	aagggatttt	ggtcatgaga	ttatcaaaaa	1140			
ggatcttcac	ctagatcctt	ttaaattaaa	aatgaagttt	taaatcaatc	taaagtatat	1200			
atgagtaaac	ttggtctgac	agttaccaat	gcttaatcag	tgaggcacct	atctcagcga	1260			
tctgtctatt	tcgttcatcc	atagttgcct	gactccccgt	cgtgtagata	actacgatac	1320			
gggagggctt	accatctggc	cccagtgctg	caatgatacc	gcgagaccca	cgctcaccgg	1380			
ctccagattt	atcagcaata	aaccagccag	ccggaagggc	cgagcgcaga	agtggtcctg	1440			
caactttatc	cgcctccatc	cagtctatta	attgttgccg	ggaagctaga	gtaagtagtt	1500			
cgccagttaa	tagtttgcgc	aacgttgttg	ccattgctac	aggcatcgtg	gtgtcacgct	1560			
cgtcgtttgg	tatggcttca	ttcagctccg	gttcccaacg	atcaaggcga	gttacatgat	1620			
cccccatgtt	gtgcaaaaaa	gcggttagct	ccttcggtcc	tccgatcgtt	gtcagaagta	1680			
agttggccgc	agtgttatca	ctcatggtta	tggcagcact	gcataattct	cttactgtca	1740			
tgccatccgt	aagatgcttt	tctgtgactg	gtgagtactc	aaccaagtca	ttctgagaat	1800			
agtgtatgcg	gcgaccgagt	tgetettgee	cggcgtcaat	acgggataat	accgcgccac	1860			
atagcagaac	tttaaaagtg	ctcatcattg	gaaaacgttc	ttcggggcga	aaactctcaa	1920			

377

378

ggatcttacc	gctgttgaga	tccagttcga	tgtaacccac	tcgtgcaccc	aactgatctt	1980
cagcatcttt	tactttcacc	agcgtttctg	ggtgagcaaa	aacaggaagg	caaaatgccg	2040
caaaaaaggg	aataagggcg	acacggaaat	gttgaatact	catactcttc	ctttttcaat	2100
attattgaag	catttatcag	ggttattgtc	tcatgagcgg	atacatattt	gaatgtattt	2160
agaaaaataa	acaaataggg	gttccgcgca	catttccccg	aaaagtgcca	cctgaacgaa	2220
gcatctgtgc	ttcattttgt	agaacaaaaa	tgcaacgcga	gagcgctaat	ttttcaaaca	2280
aagaatctga	gctgcatttt	tacagaacag	aaatgcaacg	cgaaagcgct	attttaccaa	2340
cgaagaatct	gtgcttcatt	tttgtaaaac	aaaaatgcaa	cgcgagagcg	ctaatttttc	2400
aaacaaagaa	tctgagctgc	atttttacag	aacagaaatg	caacgcgaga	gcgctatttt	2460
accaacaaag	aatctatact	tctttttgt	tctacaaaaa	tgcatcccga	gagcgctatt	2520
tttctaacaa	agcatcttag	attactttt	ttctcctttg	tgcgctctat	aatgcagtct	2580
cttgataact	ttttgcactg	taggtccgtt	aaggttagaa	gaaggctact	ttggtgtcta	2640
ttttctcttc	cataaaaaaa	gcctgactcc	acttcccgcg	tttactgatt	actagcgaag	2700
ctgcgggtgc	attttttcaa	gataaaggca	tccccgatta	tattctatac	cgatgtggat	2760
tgcgcatact	ttgtgaacag	aaagtgatag	cgttgatgat	tcttcattgg	tcagaaaatt	2820
atgaacggtt	tcttctattt	tgtctctata	tactacgtat	aggaaatgtt	tacattttcg	2880
tattgttttc	gattcactct	atgaatagtt	cttactacaa	ttttttgtc	taaagagtaa	2940
tactagagat	aaacataaaa	aatgtagagg	tcgagtttag	atgcaagttc	aaggagcgaa	3000
aggtggatgg	gtaggttata	tagggatata	gcacagagat	atatagcaaa	gagatacttt	3060
tgagcaatgt	ttgtggaagc	ggtattcgca	atattttagt	agctcgttac	agtccggtgc	3120
gtttttggtt	ttttgaaagt	gcgtcttcag	agcgcttttg	gttttcaaaa	gcgctctgaa	3180
gttcctatac	tttctagaga	ataggaactt	cggaatagga	acttcaaagc	gtttccgaaa	3240
acgagcgctt	ccgaaaatgc	aacgcgagct	gcgcacatac	agctcactgt	tcacgtcgca	3300
cctatatctg	cgtgttgcct	gtatatatat	atacatgaga	agaacggcat	agtgcgtgtt	3360
tatgcttaaa	tgcgtactta	tatgcgtcta	tttatgtagg	atgaaaggta	gtctagtacc	3420
tcctgtgata	ttatcccatt	ccatgcgggg	tatcgtatgc	ttccttcagc	actacccttt	3480
agctgttcta	tatgctgcca	ctcctcaatt	ggattagtct	catccttcaa	tgctatcatt	3540
tcctttgata	ttggatcatc	taagaaacca	ttattatcat	gacattaacc	tataaaaata	3600
ggcgtatcac	gaggcccttt	cgtctcgcgc	gtttcggtga	tgacggtgaa	aacctctgac	3660
acatgcagct	cccggagacg	gtcacagctt	gtctgtaagc	ggatgccggg	agcagacaag	3720
cccgtcaggg	cgcgtcagcg	ggtgttggcg	ggtgtcgggg	ctggcttaac	tatgcggcat	3780
cagagcagat	tgtactgaga	gtgcaccata	aattcccgtt	ttaagagctt	ggtgagcgct	3840
aggagtcact	gccaggtatc	gtttgaacac	ggcattagtc	agggaagtca	taacacagtc	3900
ctttcccgca	attttcttt	tctattactc	ttggcctcct	ctagtacact	ctatatttt	3960
ttatgcctcg	gtaatgattt	tcatttttt	ttttccccta	gcggatgact	cttttttt	4020
cttagcgatt	ggcattatca	cataatgaat	tatacattat	ataaagtaat	gtgatttctt	4080
cgaagaatat	actaaaaaat	gagcaggcaa	gataaacgaa	ggcaaagatg	acagagcaga	4140
aagccctagt	aaagcgtatt	acaaatgaaa	ccaagattca	gattgcgatc	tctttaaagg	4200
gtggtcccct	agcgatagag	cactcgatct	tcccagaaaa	agaggcagaa	gcagtagcag	4260
aacaggccac	acaatogcaa	gtgattaacq	tccacacagq	tatagggttt	ctggaccata	4320
	-	_ 0	55			

379

## -continued

tgatacatgc	tctggccaag	cattccggct	ggtcgctaat	cgttgagtgc	attggtgact	4380
tacacataga	cgaccatcac	accactgaag	actgcgggat	tgctctcggt	caagctttta	4440
aagaggccct	actggcgcgt	ggagtaaaaa	ggtttggatc	aggatttgcg	cctttggatg	4500
aggcactttc	cagagcggtg	gtagatcttt	cgaacaggcc	gtacgcagtt	gtcgaacttg	4560
gtttgcaaag	ggagaaagta	ggagatctct	cttgcgagat	gatcccgcat	tttcttgaaa	4620
gctttgcaga	ggctagcaga	attaccctcc	acgttgattg	tctgcgaggc	aagaatgatc	4680
atcaccgtag	tgagagtgcg	ttcaaggctc	ttgcggttgc	cataagagaa	gccacctcgc	4740
ccaatggtac	caacgatgtt	ccctccacca	aaggtgttct	tatgtagtga	caccgattat	4800
ttaaagctgc	agcatacgat	atatatacat	gtgtatatat	gtatacctat	gaatgtcagt	4860
aagtatgtat	acgaacagta	tgatactgaa	gatgacaagg	taatgcatca	ttctatacgt	4920
gtcattctga	acgaggcgcg	ctttcctttt	ttctttttgc	ttttcttt	tttttctctt	4980
gaactcgacg	gatctatgcg	gtgtgaaata	ccgcacagat	gcgtaaggag	aaaataccgc	5040
atcaggaaat	tgtaaacgtt	aatattttgt	taaaattcgc	gttaaatttt	tgttaaatca	5100
gctcatttt	taaccaatag	gccgaaatcg	gcaaaatccc	ttataaatca	aaagaataga	5160
ccgagatagg	gttgagtgtt	gttccagttt	ggaacaagag	tccactatta	aagaacgtgg	5220
actccaacgt	caaagggcga	aaaaccgtct	atcagggcga	tggcccacta	cgtgaaccat	5280
caccctaatc	aagttttttg	gggtcgaggt	gccgtaaagc	actaaatcgg	aaccctaaag	5340
ggagcccccg	atttagagct	tgacgggggaa	agccggcgaa	cgtggcgaga	aaggaaggga	5400
agaaagcgaa	aggagcgggc	gctagggcgc	tggcaagtgt	agcggtcacg	ctgcgcgtaa	5460
ccaccacacc	cgccgcgctt	aatgcgccgc	tacagggcgc	gtcgcgccat	tcgccattca	5520
ggctgcgcaa	ctgttgggaa	gggcgatcgg	tgegggeete	ttcgctatta	cgccagctgg	5580
cgaaaggggg	atgtgctgca	aggcgattaa	gttgggtaac	gccagggttt	tcccagtcac	5640
gacgttgtaa	aacgacggcc	agtgagcgcg	cgtaatacga	ctcactatag	ggcgaattgg	5700
gtaccgggcc	ccccctcgag	gtcgacggta	tcgataagct	tgatatcgaa	tteetgegee	5760
cgggccacta	gtcagatgcc	gcgggcactt	gagcacctca	tgcacagcaa	taacacaaca	5820
caatggttag	tagcaacctg	aattcggtca	ttgatgcatg	catgtgccgt	gaagcggggac	5880
aaccagaaaa	gtcgtctata	aatgeeggea	cgtgcgatca	tcgtggcggg	gttttaagag	5940
tgcatatcac	aaattgtcgc	attaccgcgg	aaccgccaga	tattcattac	ttgacgcaaa	6000
agcgtttgaa	ataatgacga	aaaagaagga	agaaaaaaaa	agaaaaatac	cgcttctagg	6060
cgggttatct	actgatccga	gcttccacta	ggatagcacc	caaacacctg	catatttgga	6120
cgacctttac	ttacaccacc	aaaaaccact	ttegeetete	ccgcccctga	taacgtccac	6180
taattgagcg	attacctgag	cggtcctctt	ttgtttgcag	catgagactt	gcatactgca	6240
aatcgtaagt	agcaacgtct	caaggtcaaa	actgtatgga	aaccttgtca	cctcacttaa	6300
ttctagctag	cctaccctgc	aagtcaagag	gtctccgtga	ttcctagcca	cctcaaggta	6360
tgcctctccc	cggaaactgt	ggccttttct	ggcacacatg	atctccacga	tttcaacata	6420
taaatagctt	ttgataatgg	caatattaat	caaatttatt	ttacttcttt	cttgtaacat	6480
ctctcttgta	atcccttatt	ccttctagct	atttttcata	aaaaaccaag	caactgctta	6540
tcaacacaca	aacactaaat	caaagctgag	gatggattta	tttgagtcat	tagcacaaaa	6600
aattactggt	aaagatcaaa	caattgtttt	ccctgaagga	actgaacccc	gaattgtcgg	6660

381

### -continued

	6720
caaagttcag gctgtggcta acgatttgaa tgcggattta acaggcgttc aagtccttga	6780
teetgegaca taeceegetg aagataagea ageaatgett gatgeeeteg ttgaaeggeg	6840
gaaaggtaag aatacgccag aacaagcggc taaaatgctg gaagatgaaa actactttgg	6900
cacgatgete gtttatatgg gcaaagegga tgggatggtt teaggtgeaa teeateeaae	6960
tggtgatacg gtacggccag cgttacaaat tattaagacc aagcccggtt cacaccgaat	7020
ctcgggtgca tttatcatgc aaaagggtga ggaacgctac gtctttgctg actgtgccat	7080
caatattgat cccgatgccg atacgttagc ggaaattgcc actcagagtg cggctactgc	7140
taaggtette gatattgace egaaagttge gatgeteage tteteaaeta agggttegge	7200
taagggtgaa atggtcacta aagtgcaaga agcaacggcc aaggcgcaag ctgctgaacc	7260
ggaattggct atcgatggtg aacttcaatt tgacgcggcc ttcgttgaaa aagttggttt	7320
gcaaaagget eetggtteea aagtagetgg teatgeeaat gtetttgtat tteeagaget	7380
tcagtctggt aatattggct ataagattgc gcaacgattt ggtcattttg aagcggtggg	7440
teetgtettg caaggeetga acaageeggt eteegaettg teaegtggat geagtgaaga	7500
agacgtttat aaggttgcga ttattacagc agcccaagga ttagcttaat taattaagag	7560
taagcgaatt tottatgatt tatgattttt attattaaat aagttataaa aaaaataagt	7620
gtatacaaat tttaaagtga ctcttaggtt ttaaaacgaa aattcttatt cttgagtaac	7680
tettteetgt aggteaggtt gettteteag gtatageatg aggtegetet tattgaceae	7740
acctctaccg gcatgccgag caaatgcctg caaatcgctc cccatttcac ccaattgtag	7800
atatgctaac tccagcaatg agttgatgaa tctcggtgtg tattttatgt cctcagagga	7860
caacacctgt ggtactagtt ctagagcggc cgcccgcaaa ttaaagcctt cgagcgtccc	7920
aaaaccttct caagcaaggt tttcagtata atgttacatg cgtacacgcg tttgtacaga	7980
aaaaaaagaa aaatttgaaa tataaataac gttcttaata ctaacataac tattaaaaaa	8040
aataaatagg gacctagact tcaggttgtc taactccttc cttttcggtt agagcggatg	8100
tgggaggagg gcgtgaatgt aagcgtgaca taactaatta catgattaat taattattt	8160
aaacccttcc attgccaatc attaacttct ggcaagtcag ttccggcatc ccggatatag	8220
gcattgtgtt tagcaagcat attatccatg gattgaacga aggeegeace agtgttttee	8280
attgetggtt gegeegeaat tgeegaetta getaagtega ageggteeat etggtteatg	8340
accogtaogt ogaatggtgt ggtaatatoa ooattttoao ggtaacogtg gaogtataag	8400
ttatggttgt gacgatcaaa gaagatgtca cgaactaagt cttcgtaacc gtggaaagca	8460
aagaccactg gtttgtcctt agtaaagtaa tggtcaaact cagcatctga caagccccgc	8520
ggateetttt eaggaetaeg taaetteaag atgtegaeea egtteaegaa aegaatette	8580
atetetggga aactgtegtg tagtaattgg atggeageea aegttteaag egttggttee	8640
gtcccagcag ctgcaaagac aatgtctggt tcgctacctt ggtccgtact tgcccaatca	8700
atgataccaa gaccattgtc aactaattgc ttagcttctt caatgctgaa ccattgttga	8760
cgtgggtgtt ttgacgtaac cacgtagttg atcttttctt ggctccggaa aatgacgtca	8820
ccgacagcta ataacgtgtt ggcatcggct ggtaaatatt cacgaatgta ttctggtttc	8880
ttttcggcca aatgagttaa tgcacctgga tcttggtggg tataaccatt atggtcttgt	8940
tggaatacag ttgaagccgc gataatgtta agtgatgggt actttttacg ccaatcaagt	9000
tcattggctt tacgtaacca cttgaagtgt tgcgtcaaca ttgagtccac aacgcgtagg	9060

383

#### -continued

aaggcttcat	aactggcaaa	taacccatga	cgtccagtta	agacgtaacc	ttctaaccaa	9120		
ccttcagctt	ggtgttcaga	taactgagca	tctaagaccc	ggccagctgg	tgcttcatat	9180		
tggtcactat	ctggatgaat	gtcttccatc	cattgacgat	tagtggtttc	gaagacacca	9240		
tataaacggt	tagacatggt	ttcatcaggt	ccgaacaacc	ggaagttatc	aggatttttc	9300		
ttgatgacat	cccgcaaata	gtctgaccaa	acgatcatat	cttgcttaac	attcgcgcct	9360		
tctttggacg	tatcgaccgc	ataatcacgg	aagtttggta	agttcaaggc	tttcggatcg	9420		
accccaccat	tggtgattgg	gttagcagcc	atccgactgt	ccccagtagg	aataatttct	9480		
ttaatatcat	ccttcaaaga	gccatcttca	ttgaagagtt	cttttggttg	atatgattcg	9540		
agccaatcaa	ctaaagcatc	cgcatgttcc	atgtcatttt	gatcaacagg	aatcggaatt	9600		
tgatgagcac	ggaatgaacc	ttcgatctta	tcaccgtccc	atgacttcgg	accagtccag	9660		
cccttaggtg	cgcggaagac	gatcattggc	catactggca	atgttgcatc	gttattttcg	9720		
cgagcatgct	tctggattgc	cttgatcttt	tcaacggctt	catccatggc	cttagctaag	9780		
gctgggtgaa	ccttttcagg	atcgtcacct	tcaacgaaga	ttggttccca	attcatgctt	9840		
tcgaagtatt	ccttaatctt	agcatcagaa	gtccgaccaa	aaatcgttgg	attagaaatc	9900		
ttaaaaccat	ttaagttcaa	gattggtaaa	acagccccgt	cgttgattgg	gttaatgaac	9960		
ttcgttgatt	gccatgaagt	tgctaatgga	cccgtttcgg	attccccatc	accaacaaca	10020		
accgcggcga	tttcgtcagg	attgtcaaga	attgccccaa	ccccgtgtga	aattgagtaa	10080		
ccaagttcgc	caccttcgtg	gattgaaccg	ggtgtttcag	gtgccgcatg	ggaagcaacc	10140		
ccacctggga	atgagaattg	cttgaagagc	ttttgcatcc	cttcaacatc	ctgcgtaatt	10200		
tctggataaa	tatcggtgta	agtaccgtca	aggtaagagt	ttgaaaccat	cacttgacca	10260		
ccatgacctg	gaccttcaac	gtagaacatc	ttcaaaccgt	acttgttgat	gacccggtta	10320		
agatgagcat	agataaagtt	ttgaccggca	atcgtccccc	agtgaccaat	tggatgaacc	10380		
ttaacgtcac	tggccttcaa	tggccgttgt	aatagtggat	tatcttttaa	ataaagttga	10440		
ccaactgata	agtagttggc	agcacgccag	tacttatcaa	ctttttgcaa	atatgctggt	10500		
gatgagtaat	ctgttgtcat	cctcagctgg	aacttagatt	agattgctat	gctttctctc	10560		
taacgagcaa	gaagtaaaaa	aagttgtaat	agaacaagaa	aaatgaaact	gaagettgag	10620		
aaattgaaga	ccgtttatta	gcttaaatat	caatgggagg	tcatcgaaag	agaaaaaaat	10680		
caagaaagaa	actctcaaga	aaaagaaacg	tgataaaaat	ttttattgcc	tetetegaeg	10740		
aagagaaaga	aacgaggcgg	tcccttttt	cttttccaaa	cctttagtac	gggtaattag	10800		
cgacacccta	gaggaagaaa	gaggggaaat	ttagtatgct	gtgcttgggt	gtcttgaagt	10860		
ggtacggcga	tgcgcggagt	ccgagaaaat	ctggaagagt	aaaaaggggg	tagaagcgtt	10920		
ttgaagctat	cege					10934		
<210> SEQ I <211> LENGT <212> TYPE: <213> ORGAN <220> FEATU <223> OTHER <400> SEQUE	<210> SEQ ID NO 163 <211> LENGTH: 80 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: N1341							
gttgcaagaa	atgcattatg	caattttttg	attatgacaa	tctctcgaaa	atagcttcaa	60		
aacgcttcta	cccccttttt	J	-	-		80		

#### 385

-continued

<210> SEQ ID NO 164 <211> LENGTH: 80 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: N1338 <400> SEQUENCE: 164 catacattat acgaacggta ctgaacatta gaatacgtaa tccgcaatgc ccgcaaatta 60 80 aageettega gegteecaaa <210> SEQ ID NO 165 <211> LENGTH: 30 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: N1033c <400> SEQUENCE: 165 gcattgcgga ttacgtattc taatgttcag 30 <210> SEQ ID NO 166 <211> LENGTH: 81 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: N1342 <400> SEQUENCE: 166 acatatgtga aaaaaaatag ttgatatttt aaaccaaatc agaaatttat caccttggct 60 aactcgttgt atcatcactg g 81 <210> SEQ ID NO 167 <211> LENGTH: 30 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: N1364 <400> SEQUENCE: 167 atgacaacag attactcatc accagcatat 30 <210> SEQ ID NO 168 <211> LENGTH: 604 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: L8 <400> SEQUENCE: 168 gcctacttgg cttcacatac gttgcatacg acgatataga aaataatgat aatgacagca 60 qqattatcqt ataacqtaat aqtcqaaaaa tctcaaaaat ctqtqqqtca ttacqtaaat 120 aatgatagga atgtgattet tetattttte ettttteeat tetggeagee gtegggaaaa 180 cgtggettee tetettegg getetattgg agtaacgetg cegtgagett cetettte 240 catatctaac aactgagcac gtaaccaatg gtaaagcatg agcttagcgt tgctccaaag 300 aagtattgga aggttaatac catgtgtctg ttctcttctg actttgactc ctcaaataaa 360 aaaaaattet acaateaaca gategettea attaegetet cacaaaaaet ttttteette 420 ttcttcgccc acgttaaatt ttaaccctca tgctgtctaa cggatttctg cacttaattt 480 attataaaac gacaaagaca taatacttct ctatcaattt cagttattgt tcttcattgc 540

388

\_

- 2	o	7
Э	o	1

#### -continued

-continued	
attactcttc tgttcttctt tttcatttgt catatacaac cataaccaaa taatacatat	600
tcaa	604
<210> SEQ ID NO 169 <211> LENGTH: 80 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: N1366	
<400> SEQUENCE: 169	
gttgcaagaa atgcattatg caattttttg attatgacaa tctctcgaaa gcctacttgg	60
cttcacatac gttgcatacg	80
<pre>&lt;210&gt; SEQ ID NO 170 &lt;211&gt; LENGTH: 65 &lt;212&gt; TYPE: DNA &lt;213&gt; ORGANISM: Artificial Sequence &lt;220&gt; FEATURE: &lt;223&gt; OTHER INFORMATION: N1368</pre>	
<400> SEQUENCE: 170	<u>(</u> )
atatgeiggt gatgagtaat eigtigteat titgaatatg tattattigg ttatggtigt	60
atatg	60
<210> SEQ ID NO 171 <211> LENGTH: 80 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: N1371	
<400> SEQUENCE: 171	
aaaaactaat acgtaaacct gcattaaggt aagattatat cagaaaatgt gttgcaagaa	60
atgcattatg caatttttg	80
<210> SEQ ID NO 172 <211> LENGTH: 85 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: N1372	
<400> SEQUENCE: 172	
tagaagctaa tetttaaeet ggaagacagg acagaaaagt aattacaaga acatatgtga	60
aaaaaaatag ttgatatttt aaacc	85
<210> SEQ ID NO 173 <211> LENGTH: 25 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: BK93 <400> SEQUENCE: 173	
aaaaattgat teteategta aatge	25
<210> SEQ ID NO 174 <211> LENGTH: 30 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE:	

<220> FEATURE:
-continued	
<223> OTHER INFORMATION: N1114	
<400> SEQUENCE: 174	
atatgetggt gatgagtaat etgttgteat	30
<210> SEQ ID NO 175 <211> LENGTH: 6728 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: pJT254	
<400> SEQUENCE: 175	
tegegegttt eggtgatgae ggtgaaaaee tetgaeaeat geageteeeg gagaeggtea	60
cagettgtet gtaageggat geegggagea gacaageeeg teagegegeg teagegegtg 1	120
ttggcgggtg tcggggctgg cttaactatg cggcatcaga gcagattgta ctgagagtgc 💦 1	180
accataaatt cccgttttaa gagcttggtg agcgctagga gtcactgcca ggtatcgttt 2	240
gaacacggca ttagtcaggg aagtcataac acagtcettt eeegcaattt tettttteta 3	300
ttactcttgg cctcctctag tacactctat attttttat gcctcggtaa tgattttcat	360
ttttttttt cccctagcgg atgactettt ttttttetta gegattggea ttateacata 4	120
atgaattata cattatata agtaatgtga tttcttcgaa gaatatacta aaaaatgagc 🛛 4	180
aggcaagata aacgaaggca aagatgacag agcagaaagc cctagtaaag cgtattacaa 5	540
atgaaaccaa gattcagatt gcgatctctt taaagggtgg tcccctagcg atagagcact 6	500
cgatcttccc agaaaaagag gcagaagcag tagcagaaca ggccacacaa tcgcaagtga 🦷 6	560
ttaacgtcca cacaggtata gggtttctgg accatatgat acatgctctg gccaagcatt 7	720
ccggctggtc gctaatcgtt gagtgcattg gtgacttaca catagacgac catcacacca 7	780
ctgaagactg cgggattgct ctcggtcaag cttttaaaga ggccctactg gcgcgtggag	340
taaaaaggtt tggatcagga tttgcgcctt tggatgaggc actttccaga gcggtggtag	900
atctttcgaa caggccgtac gcagttgtcg aacttggttt gcaaagggag aaagtaggag 9	960
atctctcttg cgagatgatc ccgcattttc ttgaaagctt tgcagaggct agcagaatta 10	020
ccctccacgt tgattgtctg cgaggcaaga atgatcatca ccgtagtgag agtgcgttca 10	080
aggetettge ggttgecata agagaageea eetegeeeaa tggtaeeaae gatgtteeet 11	140
ccaccaaagg tgttcttatg tagtgacacc gattatttaa agctgcagca tacgatatat 12	200
atacatgtgt atatatgtat acctatgaat gtcagtaagt atgtatacga acagtatgat 12	260
actgaagatg acaaggtaat gcatcattct atacgtgtca ttctgaacga ggcgcgcttt 13	320
cetttttet ttttgetttt tettttttt tetettgaae tegaeggate tatgeggtgt 13	380
gaaataccgc acagatgcgt aaggagaaaa taccgcatca ggaaattgta aacgttaata 14	440
ttttgttaaa attcgcgtta aatttttgtt aaatcagctc attttttaac caataggccg 15	500
aaatcggcaa aatcccttat aaatcaaaag aatagaccga gatagggttg agtgttgttc 15	560
cagtttggaa caagagtcca ctattaaaga acgtggactc caacgtcaaa gggcgaaaaa 16	520
ccgtctatca gggcgatggc ccactacgtg aaccatcacc ctaatcaagt tttttggggt 16	580
cgaggtgccg taaagcacta aatcggaacc ctaaagggag cccccgattt agagcttgac 17	740
ggggaaagcc ggcgaacgtg gcgagaaagg aagggaagaa agcgaaagga gcggggcgcta 18	300
gggegetgge aagtgtageg gteaegetge gegtaaceae eacaeeegee gegettaatg 18	360
cgccgctaca gggcgcgtcg cgccattcgc cattcaggct gcgcaactgt tgggaagggc 19	920

391

### -continued

gatcggtgcg	ggcctcttcg	ctattacgcc	agctggcgaa	aggggggatgt	gctgcaaggc	1980
gattaagttg	ggtaacgcca	gggttttccc	agtcacgacg	ttgtaaaacg	acggccagtg	2040
agcgcgcgta	atacgactca	ctatagggcg	aattgggtac	cgggcccccc	ctcgaggtcg	2100
acggtatcga	taagcttgat	tagaagccgc	cgagcgggcg	acagccctcc	gacggaagac	2160
tctcctccgt	gcgtcctcgt	cttcaccggt	cgcgttcctg	aaacgcagat	gtgcctcgcg	2220
ccgcactgct	ccgaacaata	aagattctac	aatactagct	tttatggtta	tgaagaggaa	2280
aaattggcag	taacctggcc	ccacaaacct	tcaaattaac	gaatcaaatt	aacaaccata	2340
ggatgataat	gcgattagtt	ttttagcctt	atttctgggg	taattaatca	gcgaagcgat	2400
gatttttgat	ctattaacag	atatataaat	ggaaaagctg	cataaccact	ttaactaata	2460
ctttcaacat	tttcagtttg	tattacttct	tattcaaatg	tcataaaagt	atcaacaaaa	2520
aattgttaat	atacctctat	actttaacgt	caaggagaaa	aatgtccaat	ttactgcccg	2580
tacaccaaaa	tttgcctgca	ttaccggtcg	atgcaacgag	tgatgaggtt	cgcaagaacc	2640
tgatggacat	gttcagggat	cgccaggcgt	tttctgagca	tacctggaaa	atgcttctgt	2700
ccgtttgccg	gtcgtgggcg	gcatggtgca	agttgaataa	ccggaaatgg	tttcccgcag	2760
aacctgaaga	tgttcgcgat	tatcttctat	atcttcaggc	gcgcggtctg	gcagtaaaaa	2820
ctatccagca	acatttgggc	cagctaaaca	tgcttcatcg	tcggtccggg	ctgccacgac	2880
caagtgacag	caatgctgtt	tcactggtta	tgcggcggat	ccgaaaagaa	aacgttgatg	2940
ccggtgaacg	tgcaaaacag	gctctagcgt	tcgaacgcac	tgatttcgac	caggttcgtt	3000
cactcatgga	aaatagcgat	cgctgccagg	atatacgtaa	tctggcattt	ctggggattg	3060
cttataacac	cctgttacgt	atagccgaaa	ttgccaggat	cagggttaaa	gatatctcac	3120
gtactgacgg	tgggagaatg	ttaatccata	ttggcagaac	gaaaacgctg	gttagcaccg	3180
caggtgtaga	gaaggcactt	agcctggggg	taactaaact	ggtcgagcga	tggatttccg	3240
tctctggtgt	agctgatgat	ccgaataact	acctgttttg	ccgggtcaga	aaaaatggtg	3300
ttgeegegee	atctgccacc	agccagctat	caactcgcgc	cctggaaggg	atttttgaag	3360
caactcatcg	attgatttac	ggcgctaagg	atgactctgg	tcagagatac	ctggcctggt	3420
ctggacacag	tgcccgtgtc	ggagccgcgc	gagatatggc	ccgcgctgga	gtttcaatac	3480
cggagatcat	gcaagctggt	ggctggacca	atgtaaatat	tgtcatgaac	tatatccgta	3540
acctggatag	tgaaacaggg	gcaatggtgc	gcctgctgga	agatggcgat	taggagtaag	3600
cgaatttctt	atgatttatg	atttttatta	ttaaataagt	tataaaaaaa	ataagtgtat	3660
acaaatttta	aagtgactct	taggttttaa	aacgaaaatt	cttattcttg	agtaactctt	3720
tcctgtaggt	caggttgctt	tctcaggtat	agcatgaggt	cgctcttatt	gaccacacct	3780
ctaccggcat	gccgagcaaa	tgcctgcaaa	tcgctcccca	tttcacccaa	ttgtagatat	3840
gctaactcca	gcaatgagtt	gatgaatctc	ggtgtgtatt	ttatgtcctc	agaggacaac	3900
acctgtggtg	ttctagagcg	gccgccaccg	cggtggagct	ccagettttg	ttccctttag	3960
tgagggttaa	ttgcgcgctt	ggcgtaatca	tggtcatagc	tgtttcctgt	gtgaaattgt	4020
tatccgctca	caattccaca	caacatagga	gccggaagca	taaagtgtaa	agcctggggt	4080
gcctaatgag	tgaggtaact	cacattaatt	gcgttgcgct	cactgcccgc	tttccagtcg	4140
ggaaacctgt	cgtgccagct	gcattaatga	atcggccaac	gcgcggggag	aggcggtttg	4200
cgtattgggc	gctcttccgc	ttcctcgctc	actgactcgc	tgcgctcggt	cgttcggctg	4260

393

### continued

				-0011011	Iueu	
cggcgagcgg	tatcagctca	ctcaaaggcg	gtaatacggt	tatccacaga	atcaggggat	4320
aacgcaggaa	agaacatgtg	agcaaaaggc	cagcaaaagg	ccaggaaccg	taaaaaggcc	4380
gcgttgctgg	cgtttttcca	taggctccgc	ccccctgacg	agcatcacaa	aaatcgacgc	4440
tcaagtcaga	ggtggcgaaa	cccgacagga	ctataaagat	accaggcgtt	tccccctgga	4500
agctccctcg	tgcgctctcc	tgttccgacc	ctgccgctta	ccggatacct	gtccgccttt	4560
ctcccttcgg	gaagcgtggc	gctttctcat	agctcacgct	gtaggtatct	cagttcggtg	4620
taggtcgttc	gctccaagct	gggctgtgtg	cacgaacccc	ccgttcagcc	cgaccgctgc	4680
gccttatccg	gtaactatcg	tcttgagtcc	aacccggtaa	gacacgactt	atcgccactg	4740
gcagcagcca	ctggtaacag	gattagcaga	gcgaggtatg	taggcggtgc	tacagagttc	4800
ttgaagtggt	ggcctaacta	cggctacact	agaaggacag	tatttggtat	ctgcgctctg	4860
ctgaagccag	ttaccttcgg	aaaaagagtt	ggtagctctt	gatccggcaa	acaaaccacc	4920
gctggtagcg	gtggttttt	tgtttgcaag	cagcagatta	cgcgcagaaa	aaaaggatct	4980
caagaagatc	ctttgatctt	ttctacgggg	tctgacgctc	agtggaacga	aaactcacgt	5040
taagggattt	tggtcatgag	attatcaaaa	aggatcttca	cctagatcct	tttaaattaa	5100
aaatgaagtt	ttaaatcaat	ctaaagtata	tatgagtaaa	cttggtctga	cagttaccaa	5160
tgcttaatca	gtgaggcacc	tatctcagcg	atctgtctat	ttcgttcatc	catagttgcc	5220
tgactccccg	tcgtgtagat	aactacgata	cgggagggct	taccatctgg	ccccagtgct	5280
gcaatgatac	cgcgagaccc	acgctcaccg	gctccagatt	tatcagcaat	aaaccagcca	5340
gccggaaggg	ccgagcgcag	aagtggtcct	gcaactttat	ccgcctccat	ccagtctatt	5400
aattgttgcc	gggaagctag	agtaagtagt	tcgccagtta	atagtttgcg	caacgttgtt	5460
gccattgcta	caggcatcgt	ggtgtcacgc	tcgtcgtttg	gtatggcttc	attcagctcc	5520
ggttcccaac	gatcaaggcg	agttacatga	tcccccatgt	tgtgcaaaaa	agcggttagc	5580
tccttcggtc	ctccgatcgt	tgtcagaagt	aagttggccg	cagtgttatc	actcatggtt	5640
atggcagcac	tgcataattc	tcttactgtc	atgccatccg	taagatgctt	ttctgtgact	5700
ggtgagtact	caaccaagtc	attctgagaa	tagtgtatgc	ggcgaccgag	ttgctcttgc	5760
ccggcgtcaa	tacgggataa	taccgcgcca	catagcagaa	ctttaaaagt	gctcatcatt	5820
ggaaaacgtt	cttcgggggcg	aaaactctca	aggatcttac	cgctgttgag	atccagttcg	5880
atgtaaccca	ctcgtgcacc	caactgatct	tcagcatctt	ttactttcac	cagcgtttct	5940
gggtgagcaa	aaacaggaag	gcaaaatgcc	gcaaaaaagg	gaataagggc	gacacggaaa	6000
tgttgaatac	tcatactctt	cctttttcaa	tattattgaa	gcatttatca	gggttattgt	6060
ctcatgagcg	gatacatatt	tgaatgtatt	tagaaaaata	aacaaatagg	ggtteegege	6120
acatttcccc	gaaaagtgcc	acctgggtcc	ttttcatcac	gtgctataaa	aataattata	6180
atttaaattt	tttaatataa	atatataaat	taaaaataga	aagtaaaaaa	agaaattaaa	6240
gaaaaaatag	tttttgtttt	ccgaagatgt	aaaagactct	aggggggatcg	ccaacaaata	6300
ctacctttta	tcttgctctt	cctgctctca	ggtattaatg	ccgaattgtt	tcatcttgtc	6360
tgtgtagaag	accacacacg	aaaatcctgt	gattttacat	tttacttatc	gttaatcgaa	6420
tgtatatcta	tttaatctgc	ttttcttgtc	taataaatat	atatgtaaag	tacgcttttt	6480
gttgaaattt	tttaaacctt	tgtttatttt	tttttcttca	ttccgtaact	cttctacctt	6540
ctttatttac	tttctaaaat	ccaaatacaa	aacataaaaa	taaataaaca	cagagtaaat	6600
tcccaaatta	ttccatcatt	aaaagatacg	aggcgcgtgt	aagttacagg	caagcgatcc	6660

US	9,771	,602	<b>B</b> 2

- 2	0	5
ം	"	2

gtoctaagaa accattatta tcatgacatt aacctataaa aataggogta tcaogaggoo	6720
ctttcgtc	6728
<210> SEQ ID NO 176 <211> LENGTH: 22 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: N160SeqF5	
<400> SEQUENCE: 176	
cctgaagtct aggtccctat tt	22
<210> SEQ ID NO 177 <211> LENGTH: 1650 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: AMN1	
<400> SEQUENCE: 177	
atgaagctgg agcgcgtgag ttctaacggg agctttaagc gtggccgtga catccaaagt	60
ttggagagtc cgtgtacccg cccattaaag aaaatgtcgc catcaccttc atttacgagc	120
ctgaagatgg aaaaaccgtt taaggacatt gttcgaaaat acggggggtca cctgcaccag	180
teetegtata acceaggtte tteaaaagtt gaaetegtge gteeggaeet gagettgaaa	240
acggaccaat catttttgca gagcagcgtg cagacaaccc cgaacaaaaa gagttgtaac	300
gagtatetgt ecacaecega agecaeteee ettaagaaca eggeeacega gaatgegtgg	360
gctacgtcaa gggtggtgag cgcatcaagc ctgtcaatcg tcacgccgac cgaaatcaaa	420
aatatactgg ttgacgagtt tagtgaacta aaacttggtc agcccttaac agcccagcac	480
attractor anarchice totacate approximate and a contractor approximate	540
adtatagan attecettet natutataaa nacgangaac negraagaa ageatorte	660
accarding action to according to acquire to a acquire according to acc	720
ctatttaget geatgatgat caacegeetg togttgaatg teacgegtee gttettattt	780
aaqteteteqe attteaaate aqtqeacaae tteaaaqaat ttetqeqeae aaqteaqqaa	840
accacgcaag tgatgaggee atcgcacttt atcetgeata aattgcaeca qqtaacqcaq	900
ccggatattg agagactgtc tagaatggaa tgccagaacc tcaagtggtt ggaattttat	960
gtatgtcccc gtattacacc tccactgtct tggttcgaca atttgcataa gttagaaaaa	1020
- ttaatcatcc ccggaaacaa gaatatcgac gataatttcc tcttacggct gtctcagagt	1080
attoctaacc tgaaacacct cgtgcttogt gcttgcgaca atgtttooga tagtggtgta	1140
gtttgtatcg ccctgaactg ccctaagctg aagacgttca acatcggacg tcatcgccgc	1200
ggcaatetga ttacateagt tagettggtt geeetgggta agtataegea agttgagaee	1260
gttggttttg caggctgcga tgtggacgac gcaggcatat gggagttcgc gcgtttaaac	1320
gggaaaaacg tcgagcgcct gtcactcaac agttgccggc ttttaaccqa ctataqcttq	1380
ccaatcotgt ttgccottaa tagtttoccg aacottgogg tgttggaaat toqaaacoto	1440
gataaaatta cagatgtoog coattttgtg aaatataato tgtqgaagaa atcactggat	1500
actectatec trattgarge atgegaacge ataacaaage trattgatea ggaagaaga	1560
greener ogwoogwygo yogogwwogo wouwouwugo ogwoogwoow yyaayadaac	

cgggtcaaac gcataaatag cctggtcgct ttaaaggata tgaccgcgtg ggtgaacgct	1620
gacgatgaaa ttgaaaacaa cgtcgattga	1650
<210> SEQ ID NO 178 <211> LENGTH: 549 <212> TYPE: PRT <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: AMN1	
<400> SEQUENCE: 178	
Met Lys Leu Glu Arg Val Ser Ser Asn Gly Ser Phe Lys Arg Gly Arg 1 5 10 15	
Asp Ile Gln Ser Leu Glu Ser Pro Cys Thr Arg Pro Leu Lys Lys Met 20 25 30	
Ser Pro Ser Pro Ser Phe Thr Ser Leu Lys Met Glu Lys Pro Phe Lys 35 40 45	
Asp Ile Val Arg Lys Tyr Gly Gly His Leu His Gln Ser Ser Tyr Asn 50 55 60	
Pro Gly Ser Ser Lys Val Glu Leu Val Arg Pro Asp Leu Ser Leu Lys65707580	
Thr Asp Gln Ser Phe Leu Gln Ser Ser Val Gln Thr Thr Pro Asn Lys 85 90 95	
Lys Ser Cys Asn Glu Tyr Leu Ser Thr Pro Glu Ala Thr Pro Leu Lys 100 105 110	
Asn Thr Ala Thr Glu Asn Ala Trp Ala Thr Ser Arg Val Val Ser Ala 115 120 125	
Ser Ser Leu Ser Ile Val Thr Pro Thr Glu Ile Lys Asn Ile Leu Val 130 135 140	
Asp Glu Phe Ser Glu Leu Lys Leu Gly Gln Pro Leu Thr Ala Gln His 145 150 155 160	
Gln Arg Ser His Ala Val Phe Glu Ile Pro Glu Ile Val Glu Asn Ile 165 170 175	
Ile Lys Met Ile Val Ser Leu Glu Ser Ala Asn Ile Pro Lys Glu Arg 180 185 190	
Pro Cys Leu Arg Arg Asn Pro Gln Ser Tyr Glu His Ser Leu Leu Met 195 200 205	
Tyr Lys Asp Glu Glu Arg Ala Lys Lys Ala Trp Ser Ala Ala Gln Gln 210 215 220	
Leu Arg Asp Pro Pro Leu Val Gly His Lys Glu Lys Lys Gln Gly Ala 225 230 235 240	
Leu Phe Ser Cys Met Met Val Asn Arg Leu Trp Leu Asn Val Thr Arg 245 250 255	
Pro Phe Leu Phe Lys Ser Leu His Phe Lys Ser Val His Asn Phe Lys 260 265 270	
Glu Phe Leu Arg Thr Ser Gln Glu Thr Thr Gln Val Met Arg Pro Ser 275 280 285	
His Phe Ile Leu His Lys Leu His Gln Val Thr Gln Pro Asp Ile Glu 290 295 300	
Arg Leu Ser Arg Met Glu Cys Gln Asn Leu Lys Trp Leu Glu Phe Tyr 305 310 315 320	
Val Cys Pro Arg Ile Thr Pro Pro Leu Ser Trp Phe Asp Asn Leu His 325 330 335	
Lys Leu Glu Lys Leu Ile Ile Pro Gly Asn Lys Asn Ile Asp Asn Asn 340 345 350	

```
-continued
```

Phe	Leu	Leu 355	Arg	Leu	Ser	Gln	Ser 360	Ile	Pro	Asn	Leu	Lys 365	His	Leu	Val	
Leu	Arg 370	Ala	Cys	Asp	Asn	Val 375	Ser	Asp	Ser	Gly	Val 380	Val	Сув	Ile	Ala	
Leu 385	Asn	Сув	Pro	Lys	Leu 390	Lys	Thr	Phe	Asn	Ile 395	Gly	Arg	His	Arg	Arg 400	
Gly	Asn	Leu	Ile	Thr 405	Ser	Val	Ser	Leu	Val 410	Ala	Leu	Gly	Lys	Tyr 415	Thr	
Gln	Val	Glu	Thr 420	Val	Gly	Phe	Ala	Gly 425	Суз	Asp	Val	Asp	Asp 430	Ala	Gly	
Ile	Trp	Glu 435	Phe	Ala	Arg	Leu	Asn 440	Gly	ГЛа	Asn	Val	Glu 445	Arg	Leu	Ser	
Leu	Asn 450	Ser	Суз	Arg	Leu	Leu 455	Thr	Asp	Tyr	Ser	Leu 460	Pro	Ile	Leu	Phe	
Ala 465	Leu	Asn	Ser	Phe	Pro 470	Asn	Leu	Ala	Val	Leu 475	Glu	Ile	Arg	Asn	Leu 480	
Asp	Lys	Ile	Thr	Asp 485	Val	Arg	His	Phe	Val 490	Lys	Tyr	Asn	Leu	Trp 495	Lys	
Lys	Ser	Leu	Asp 500	Ala	Pro	Ile	Leu	Ile 505	Glu	Ala	Cys	Glu	Arg 510	Ile	Thr	
Lys	Leu	Ile 515	Asp	Gln	Glu	Glu	Asn 520	Arg	Val	Lys	Arg	Ile 525	Asn	Ser	Leu	
Val	Ala 530	Leu	Lys	Asp	Met	Thr 535	Ala	Trp	Val	Asn	Ala 540	Asb	Asp	Glu	Ile	
Glu 545	Asn	Asn	Val	Asp												
<210 <211 <212 <213 <220 <223	)> SH L> LH 2> TY 3> OF 3> OF 3> OT	EQ II ENGTH PE: RGAN EATUH THER	D NO H: 66 DNA ISM: RE: INFC	179 538 Art: DRMA	lfic: FION	ial s : pL#	Seque 467	ence								
<400	)> SE	EQUEI	NCE :	179												
aaao	geea	igc a	aacgo	cggc	et ti	ttad	ggtt	c cct	ggco	cttt	tget	ggco	ett 1	tget	cacat	60
9000		et e	geget	ale		gati	aaa	, gga	acaad	aaaa	taci	acege		Lyay	JUGAGO	120
agag	Jegeo	ca a	ataco	gcaa	ac co	geeto	etcco	c cgo	rgegt	tgg	ccga	attca	att a	aggad	cagetg	240
gcad	gaca	igg t	tttc	ccga	t g	gaaaq	lcddd	g caç	gtgag	jege	aaco	gcaat	ta a	atgto	gagtta	300
gcto	cacto	cat t	taggo	cacco	cc aç	ggctt	taca	a ctt	tate	gctt	ccgé	geteç	gta 1	gtt	gtgtgg	360
aatt	gtga	age g	ggata	aacaa	at ti	caca	acago	g aaa	acago	ctat	gaco	catga	att a	acgco	caagct	420
tgca	atgco	ctg d	caggt	cga	ct ct	agag	gato	c cgo	catto	gegg	atta	acgta	att (	ctaat	gttca	480
gtad	cgtt	cg t	tataa	atgta	at go	ctata	acgaa	a gti	atgo	caga	ttgt	cacto	gag a	agtgo	caccat	540
acca	acago	ett t	ttcaa	attca	aa ti	cato	attt	ttt	tttt	att	ctt	ttt	tg a	attto	ggttt	600
cttt	gaaa	att t	tttt	gati	c go	gtaat	ctco	c gaa	acaga	aagg	aaga	aacga	aag g	gaago	gagcac	660
agad	ettag	gat t	tggta	atata	at ad	cgcat	atgt	: agt	gtt	gaag	aaa	catga	aaa 1	tgc	ccagta	720
ttct	taad	ccc a	aacto	gcaca	ag aa	acaaa	aaco	tgo	cagga	aaac	gaaq	gataa	aat (	catg	cgaaa	780
gcta	acata	ata a	aggaa	acgt	ge tç	gctad	ctcat	c cct	agto	cctg	ttgo	ctgco	caa q	gctat	ttaat	840
atca	atgca	acg a	aaaaq	gcaaa	ac aa	aactt	gtgt	: gct	tcat	tgg	atgi	tcgt	cac (	cacca	aaggaa	900

-continued

401

ttactggagt	tagttgaagc	attaggtccc	aaaatttgtt	tactaaaaac	acatgtggat	960
atcttgactg	atttttccat	ggagggcaca	gttaagccgc	taaaggcatt	atccgccaag	1020
tacaattttt	tactcttcga	agacagaaaa	tttgctgaca	ttggtaatac	agtcaaattg	1080
cagtactctg	cgggtgtata	cagaatagca	gaatgggcag	acattacgaa	tgcacacggt	1140
gtggtgggcc	caggtattgt	tagcggtttg	aagcaggcgg	cagaagaagt	aacaaaggaa	1200
cctagaggcc	ttttgatgtt	agcagaattg	tcatgcaagg	gctccctatc	tactggagaa	1260
tatactaagg	gtactgttga	cattgcgaag	agcgacaaag	attttgttat	cggctttatt	1320
gctcaaagag	acatgggtgg	aagagatgaa	ggttacgatt	ggttgattat	gacacccggt	1380
gtgggtttag	atgacaaggg	agacgcattg	ggtcaacagt	atagaaccgt	ggatgatgtg	1440
gtctctacag	gatctgacat	tattattgtt	ggaagaggac	tatttgcaaa	gggaagggat	1500
gctaaggtag	agggtgaacg	ttacagaaaa	gcaggctggg	aagcatattt	gagaagatgc	1560
ggccagcaaa	actaaaaaac	tgtattataa	gtaaatgcat	gtatactaaa	ctcacaaatt	1620
agagcttcaa	tttaattata	tcagttatta	ccctatgcgg	tgtgaaatac	cgcacagatg	1680
cgtaaggaga	aaataccgca	tcaggaaatt	gtaaacgtta	atattttgtt	aaaattcgcg	1740
ttaaattttt	gttaaatcag	ctcattttt	aaccaatagg	ccgaaatcgg	caaaatccct	1800
tataaatcaa	aagaatagac	cgagataggg	ttgagtgttg	ttccagtttg	gaacaagagt	1860
ccactattaa	agaacgtgga	ctccaacgtc	aaagggcgaa	aaaccgtcta	tcagggcgat	1920
ggcccactac	gtgaaccatc	accctaatca	agataacttc	gtataatgta	tgctatacga	1980
acggtaccag	tgatgataca	acgagttagc	caaggtgaat	tcgacttagg	atgtctcatc	2040
aatcatctta	ttcctgctgg	tgtttttgt	atcgccttgc	cttggagtgt	ttatgcttgt	2100
cctttgttca	gtaaccattc	ttcaagtttg	tttcaagtag	taggatacct	tcagatatac	2160
gaaagaaagg	gagtatagtt	gtggatatat	atatatatag	caacccttct	ttataagggt	2220
cctatagact	atactcttca	cactttaaag	tacggaatta	aggcccaagg	gaactaacaa	2280
aaacgttcaa	aaagttttaa	aactatatgt	gttaactgta	caaaaataac	ttatttatca	2340
tatcattttt	ttctctgttt	atttcttcta	gaacttatac	ctgtctttc	cttttattct	2400
ttgaatttgk	tttaatatcc	ctttttgktt	taatatccat	ccattccttt	cacttagaac	2460
taataattcc	cttcgtttga	taatttatca	ttttcctttt	ctgttagtaa	agtacccatt	2520
aaatgaagct	ggagcgcgtg	agttctaacg	ggagctttaa	gcgtggccgt	gacatccaaa	2580
gtttggagag	tccgtgtacc	cgcccattaa	agaaaatgtc	gccatcacct	tcatttacga	2640
gcctgaagat	ggaaaaaccg	tttaaggaca	ttgttcgaaa	atacggggggt	cacctgcacc	2700
agtcctcgta	taacccaggt	tcttcaaaag	ttgaactcgt	gcgtccggac	ctgagcttga	2760
aaacggacca	atcatttttg	cagagcagcg	tgcagacaac	cccgaacaaa	aagagttgta	2820
acgagtatct	gtccacaccc	gaagccactc	cccttaagaa	cacggccacc	gagaatgcgt	2880
gggctacgtc	aagggtggtg	agcgcatcaa	gcctgtcaat	cgtcacgccg	accgaaatca	2940
aaaatatact	ggttgacgag	tttagtgaac	taaaacttgg	tcagccctta	acagcccagc	3000
accaacggag	ccatgcagtt	ttcgagatac	ctgagatcgt	agagaacata	atcaagatga	3060
tcgtttccct	cgagagcgcc	aatattccga	aagaacgtcc	gtgcctgcgt	cgcaacccgc	3120
agagttatga	gcattccctt	ctgatgtata	aagacgagga	acgcgcgaag	aaagcatggt	3180
ccgcggctca	acaactgcgc	gatccgccgc	tggtgggtca	taaggaaaaa	aaacagggcg	3240

-continued	
ctctgtttag ctgcatgatg gtcaaccgcc tgtggttgaa tgtcacgcgt ccgt	tcttat 3300
ttaagtetet geattteaaa teagtgeaca aetteaaaga atttetgege acaa	gtcagg 3360
aaaccacgca agtgatgagg ccatcgcact ttatcctgca taaattgcac cagg	taacgc 3420
agccggatat tgagagactg tctagaatgg aatgccagaa cctcaagtgg ttgg	aatttt 3480
atgtatgtcc ccgtattaca cctccactgt cttggttcga caatttgcat aagt	tagaaa 3540
aattaatcat ccccggaaac aagaatatcg acgataattt cctcttacgg ctgt	ctcaga 3600
gtatteetaa eetgaaacae etegtgette gtgettgega caatgtttee gata	gtggtg 3660
tagtttgtat cgccctgaac tgccctaagc tgaagacgtt caacatcgga cgtc	atcgcc 3720
gcggcaatct gattacatca gttagcttgg ttgccctggg taagtatacg caag	ttgaga 3780
ccgttggttt tgcaggctgc gatgtggacg acgcaggcat atgggagttc gcgc	gtttaa 3840
acgggaaaaa cgtcgagcgc ctgtcactca acagttgccg gcttttaacc gact	atagct 3900
tgccaatcct gtttgccctt aatagtttcc cgaaccttgc ggtgttggaa atte	gaaacc 3960
tcgataaaat tacagatgtc cgccattttg tgaaatataa tctgtggaag aaat	cactgg 4020
atgeteetat eetgattgag gegtgegaae geataacaaa getgattgat eagg	aagaga 4080
accgggtcaa acgcataaat agcctggtcg ctttaaagga tatgaccgcg tggg	tgaacg 4140
ctgacgatga aattgaaaac aacgtcgatt gagacgatga aattgaaaac aacg	tcgatt 4200
gaggtaccat ggtttttgtg actttaccta taaatagtac acaacagacc acca	gtaatt 4260
ctacacactt cttaactgat aatattatta taattgtaac tttttagcag cact	aaattt 4320
aatgaataca tagattttta actagcattt tactattctg tactttttac ttga	aattee 4380
agaagggccg aagaaaccag aatteettea cagaaaacga atteactgge egte	gtttta 4440
caacgtogtg actgggaaaa cootggogtt acccaactta atogoottgo agoa	catece 4500
cetttegeca getggegtaa tagegaagag geeegeaceg ategeeette eeaa	cagttg 4560
cgcageetga atggegaatg gegeetgatg eggtatttte teettaegea tetg	tgcggt 4620
atttcacacc gcatatggtg cactctcagt acaatctgct ctgatgccgc atag	ttaagc 4680
cageceegae accegeeaae accegetgae gegeeetgae gggettgtet gete	ccggca 4740
teegettaca gacaagetgt gacegtetee gggagetgea tgtgteagag gttt	tcaccg 4800
tcatcaccga aacgcgcgag acgaaagggc ctcgtgatac gcctatttt atag	gttaat 4860
gtcatgataa taatggtttc ttagacgtca ggtggcactt ttcggggaaa tgtg	cgcgga 4920
accoctattt gtttattttt ctaaatacat toaaatatgt atoogotoat gaga	caataa 4980
ccctgataaa tgcttcaata atattgaaaa aggaagagta tgagtattca acat	ttccgt 5040
gtcgccctta ttcccttttt tgcggcattt tgccttcctg tttttgctca ccca	gaaacg 5100
ctggtgaaag taaaagatgc tgaagatcag ttgggtgcac gagtgggtta cato	gaactg 5160
gateteaaca geggtaagat eettgagagt tttegeeeeg aagaaegttt teea	atgatg 5220
agcactttta aagttetget atgtggegeg gtattateee gtattgaege eggg	caagag 5280
caactoggto googcataca ctattotoag aatgaottgg ttgagtaoto acca	gtcaca 5340
gaaaagcatc ttacggatgg catgacagta agagaattat gcagtgctgc cata	accatg 5400
agtgataaca ctgcggccaa cttacttctg acaacgatcg gaggaccgaa ggag	ctaacc 5460
gcttttttgc acaacatggg ggatcatgta actcgccttg atcgttggga accg	gagetg 5520
aatgaagcca taccaaacga cgagcgtgac accacgatgc ctgtagcaat ggca	acaacg 5580
ttgcgcaaac tattaactgg cgaactactt actctagctt cccggcaaca atta	atagac 5640

405

### -

406

-continued tggatggagg cggataaagt tgcaggacca cttctgcgct cggcccttcc ggctggctgg 5700 tttattgctg ataaatctgg agccggtgag cgtgggtctc gcggtatcat tgcagcactg 5760 gggccagatg gtaagccctc ccgtatcgta gttatctaca cgacggggag tcaggcaact 5820 atggatgaac gaaatagaca gatcgctgag ataggtgcct cactgattaa gcattggtaa 5880 ctgtcagacc aagtttactc atatatactt tagattgatt taaaacttca tttttaattt 5940 aaaaggatet aggtgaagat eetttttgat aateteatga eeaaaateee ttaaegtgag 6000 ttttcgttcc actgagcgtc agaccccgta gaaaagatca aaggatcttc ttgagatcct 6060 ttttttctgc gcgtaatctg ctgcttgcaa acaaaaaaac caccgctacc agcggtggtt 6120 tgtttgccgg atcaagagct accaactctt tttccgaagg taactggctt cagcagagcg 6180 cagataccaa atactgtcct tctagtgtag ccgtagttag gccaccactt caagaactct 6240 gtagcaccgc ctacatacct cgctctgcta atcctgttac cagtggctgc tgccagtggc 6300 gataagtogt gtottacogg gttggactoa agaogatagt tacoggataa ggogoagogg 6360 tcgggctgaa cggggggttc gtgcacacag cccagcttgg agcgaacgac ctacaccgaa 6420 ctqaqatacc tacaqcqtqa qctatqaqaa aqcqccacqc ttcccqaaqq qaqaaaqqcq 6480 gacaggtatc cggtaagcgg cagggtcgga acaggagagc gcacgaggga gcttccaggg 6540 ggaaacgeet ggtatettta tagteetgte gggtttegee acetetgaet tgagegtega 6600 6638 tttttgtgat gctcgtcagg ggggcggagc ctatggaa <210> SEQ ID NO 180 <211> LENGTH: 100 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: LA712 <400> SEQUENCE: 180 cttaattgaa agaaagaatt tccttcaact tcggtttcct ggttccgcta tttctcgctt 60 gtttcttcta gcattgcgga ttacgtattc taatgttcag 100 <210> SEQ ID NO 181 <211> LENGTH: 30 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: LA746 <400> SEQUENCE: 181 30 gttttctgtg aaggaattct ggtttcttcg <210> SEQ ID NO 182 <211> LENGTH: 343 <212> TYPE: PRT <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Anaerostipes caccae KARI variant K9JB4P <400> SEQUENCE: 182 Met Glu Glu Cys Lys Met Ala Lys Ile Tyr Tyr Gln Glu Asp Cys Asn 1 Leu Ser Leu Leu Asp Gly Lys Thr Ile Ala Val Ile Gly Tyr Gly Ser 20 25 30 Gln Gly His Ala His Ala Leu Asn Ala Lys Glu Ser Gly Cys Asn Val 35 40 45

### -continued

Ile	Ile 50	Gly	Leu	Tyr	Glu	Gly 55	Ala	Glu	Glu	Trp	Lys 60	Arg	Ala	Glu	Glu
Gln 65	Gly	Phe	Glu	Val	Tyr 70	Thr	Ala	Ala	Glu	Ala 75	Ala	Lys	Lys	Ala	Asp 80
Ile	Ile	Met	Ile	Leu 85	Ile	Pro	Asp	Glu	Lys 90	Gln	Ala	Thr	Met	Tyr 95	Lys
Asn	Asp	Ile	Glu 100	Pro	Asn	Leu	Glu	Ala 105	Gly	Asn	Met	Leu	Met 110	Phe	Ala
His	Gly	Phe 115	Asn	Ile	His	Phe	Gly 120	Суз	Ile	Val	Pro	Pro 125	Гλа	Aab	Val
Asp	Val 130	Thr	Met	Ile	Ala	Pro 135	Lys	Gly	Pro	Gly	His 140	Thr	Val	Arg	Ser
Glu 145	Tyr	Glu	Glu	Gly	Lys 150	Gly	Val	Pro	Суз	Leu 155	Val	Ala	Val	Glu	Gln 160
Asp	Ala	Thr	Gly	Lys 165	Ala	Leu	Asp	Met	Ala 170	Leu	Ala	Tyr	Ala	Leu 175	Ala
Ile	Gly	Gly	Ala 180	Arg	Ala	Gly	Val	Leu 185	Glu	Thr	Thr	Phe	Arg 190	Thr	Glu
Thr	Glu	Thr 195	Asp	Leu	Phe	Gly	Glu 200	Gln	Ala	Val	Leu	Сув 205	Gly	Gly	Val
Cys	Ala 210	Leu	Met	Gln	Ala	Gly 215	Phe	Glu	Thr	Leu	Val 220	Glu	Ala	Gly	Tyr
Asp 225	Pro	Arg	Asn	Ala	Tyr 230	Phe	Glu	Суз	Ile	His 235	Glu	Met	Lys	Leu	Ile 240
Val	Asp	Leu	Ile	Tyr 245	Gln	Ser	Gly	Phe	Ser 250	Gly	Met	Arg	Tyr	Ser 255	Ile
Ser	Asn	Thr	Ala 260	Glu	Tyr	Gly	Asp	Tyr 265	Ile	Thr	Gly	Pro	Lys 270	Ile	Ile
Thr	Glu	Asp 275	Thr	Lys	Lys	Ala	Met 280	Lys	Lys	Ile	Leu	Ser 285	Asp	Ile	Gln
Asp	Gly 290	Thr	Phe	Ala	Lys	Asp 295	Phe	Leu	Val	Asp	Met 300	Ser	Asp	Ala	Gly
Ser 305	Gln	Val	His	Phe	Lys 310	Ala	Met	Arg	Гла	Leu 315	Ala	Ser	Glu	His	Pro 320
Ala	Glu	Val	Val	Gly 325	Glu	Glu	Ile	Arg	Ser 330	Leu	Tyr	Ser	Trp	Ser 335	Asp
Glu	Asp	ГЛа	Leu 340	Ile	Asn	Asn									
<210 <211	)> SE L> LE	EQ II ENGTH	) NO 1: 5'	183 71											
<212	2> TY	PE:	PRT	Art.	fia	1.01 (									
<213 <220 <223	3> 0F )> FE 3> 07	CATUR	RE: INFO	DRMA	rion	: Sti	repto		cus r	nutar	ns DI	HAD 1	varia	ant I	J2V4
<400	)> SB	EQUEI	ICE :	183											
Met 1	Thr	Asp	Lys	Lys 5	Thr	Leu	Lys	Asp	Leu 10	Arg	Asn	Arg	Ser	Ser 15	Val
Tyr	Asp	Ser	Met 20	Val	Гла	Ser	Pro	Asn 25	Arg	Ala	Met	Leu	Arg 30	Ala	Thr
Gly	Met	Gln 35	Asp	Glu	Aap	Phe	Glu 40	Lys	Pro	Ile	Val	Gly 45	Val	Ile	Ser
Thr	Trp 50	Ala	Glu	Asn	Thr	Pro 55	Суз	Asn	Ile	His	Leu 60	His	Asp	Phe	Gly

Lys Leu Ala Lys Val Gly Val Lys Glu Ala Gly Ala Trp Pro Val Gln Phe Gly Thr Ile Thr Val Ser Asp Gly Ile Ala Met Gly Thr Gln Gly Met Arg Phe Ser Leu Thr Ser Arg Asp Ile Ile Ala Asp Ser Ile Glu Ala Ala Met Gly Gly His Asn Ala Asp Ala Phe Val Ala Ile Gly Gly Cys Asp Lys Asn Met Pro Gly Ser Val Ile Ala Met Ala Asn Met Asp Ile Pro Ala Ile Phe Ala Tyr Gly Gly Thr Ile Ala Pro Gly Asn Leu Asp Gly Lys Asp Ile Asp Leu Val Ser Val Phe Glu Gly Val Gly His Trp Asn His Gly Asp Met Thr Lys Glu Glu Val Lys Ala Leu Glu Cys Asn Ala Cys Pro Gly Pro Gly Gly Cys Gly Gly Met Tyr Thr Ala Asn Thr Met Ala Thr Ala Ile Glu Val Leu Gly Leu Ser Leu Pro Gly Ser Ser Ser His Pro Ala Glu Ser Ala Glu Lys Lys Ala Asp Ile Glu Glu Ala Gly Arg Ala Val Val Lys Met Leu Glu Met Gly Leu Lys Pro Ser Asp Ile Leu Thr Arg Glu Ala Phe Glu Asp Ala Ile Thr Val Thr Met Ala Leu Gly Gly Ser Thr Asn Ser Thr Leu His Leu Leu Ala Ile Ala His Ala Ala Asn Val Glu Leu Thr Leu Asp Asp Phe Asn Thr Phe Gln Glu Lys Val Pro His Leu Ala Asp Leu Lys Pro Ser Gly Gln Tyr Val Phe Gln Asp Leu Tyr Lys Val Gly Gly Val Pro Ala Val Met Lys Tyr Leu Leu Lys Asn Gly Phe Leu His Gly Asp Arg Ile Thr Cys Thr Gly Lys Thr Val Ala Glu Asn Leu Lys Ala Phe Asp Asp Leu Thr Pro Gly Gln Lys Val Ile Met Pro Leu Glu Asn Pro Lys Arg Glu Asp Gly Pro Val Ile Ile Leu His Gly Asn Leu Ala Pro Asp Gly Ala Val Ala Lys Val Ser Gly Val Lys Val Arg Arg His Val Gly Pro Ala Lys Val Phe Asn Ser Glu Glu Glu Ala Ile Glu Ala Val Leu Asn Asp Asp Ile Val Asp Gly Asp Val Val Val Val Arg Phe Val Gly Pro Lys Gly Gly Pro Gly Met Pro Glu Met Leu Ser Leu Ser Ser Met Ile Val Gly Lys Gly 

### -continued

Thr Tyr Gly Leu Val Val Gly His Ile Ala Pro Glu Ala Gln Asp Gly 485 490 495 Gly Pro Ile Ala Tyr Leu Gln Thr Gly Asp Ile Val Thr Ile Asp Gln 500 505 510 Asp Thr Lys Glu Leu His Phe Asp Ile Ser Asp Glu Glu Leu Lys His 515 520 525 Arg Gln Glu Thr Ile Glu Leu Pro Pro Leu Tyr Ser Arg Gly Ile Leu 530 535 540 Gly Lys Tyr Ala His Ile Val Ser Ser Ala Ser Arg Gly Ala Val Thr 545 550 555 560 Asp Phe Trp Lys Pro Glu Glu Thr Gly Lys Lys 565 <210> SEQ ID NO 184 <211> LENGTH: 7523 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: pLA34 <400> SEQUENCE: 184 ccagettttg tteeetttag tgagggttaa ttgegegett ggegtaatea tggteatage 60 tqtttcctqt qtqaaattqt tatccqctca caattccaca caacataqqa qccqqaaqca 120 taaagtgtaa agcctggggt gcctaatgag tgaggtaact cacattaatt gcgttgcgct 180 cactgecege tttccagteg ggaaacetgt egtgecaget geattaatga ateggecaae 240 300 gcgcggggag aggcggtttg cgtattgggc gctcttccgc ttcctcgctc actgactcgc tgcgctcggt cgttcggctg cggcgagcgg tatcagctca ctcaaaggcg gtaatacggt 360 tatccacaga atcaggggat aacgcaggaa agaacatgtg agcaaaaggc cagcaaaagg 420 ccaggaaccg taaaaaggcc gcgttgctgg cgtttttcca taggctccgc ccccctgacg 480 agcatcacaa aaatcgacgc tcaagtcaga ggtggcgaaa cccgacagga ctataaagat 540 accaggegtt tececetgga ageteeeteg tgegetetee tgtteegaee etgeegetta 600 ccggatacct gtccgccttt ctcccttcgg gaagcgtggc gctttctcat agctcacgct 660 gtaggtatet cagtteggtg taggtegtte getecaaget gggetgtgtg caegaaceee 720 ccgttcagcc cgaccgctgc gccttatccg gtaactatcg tcttgagtcc aacccggtaa 780 gacacgactt atcgccactg gcagcagcca ctggtaacag gattagcaga gcgaggtatg 840 taggcggtgc tacagagttc ttgaagtggt ggcctaacta cggctacact agaaggacag 900 tatttggtat ctgcgctctg ctgaagccag ttaccttcgg aaaaagagtt ggtagctctt 960 gateeggeaa acaaaceace getggtageg gtggtttttt tgtttgeaag eageagatta 1020 cgcgcagaaa aaaaggatct caagaagatc ctttgatctt ttctacgggg tctgacgctc 1080 aqtqqaacqa aaactcacqt taaqqqattt tqqtcatqaq attatcaaaa aqqatcttca 1140 cctagatcct tttaaattaa aaatgaagtt ttaaatcaat ctaaagtata tatgagtaaa 1200 cttggtctga cagttaccaa tgcttaatca gtgaggcacc tatctcagcg atctgtctat 1260 ttcgttcatc catagttgcc tgactccccg tcgtgtagat aactacgata cgggagggct 1320 taccatctgg ccccagtgct gcaatgatac cgcgagaccc acgctcaccg gctccagatt 1380 tatcaqcaat aaaccaqcca qccqqaaqqq ccqaqcqcaq aaqtqqtcct qcaactttat 1440 ccgcctccat ccagtctatt aattgttgcc gggaagctag agtaagtagt tcgccagtta 1500

1560

ataqtttqcq caacqttqtt qccattqcta caqqcatcqt qqtqtcacqc tcqtcqtttq

413

gtatggcttc	attcagctcc	ggttcccaac	gatcaaggcg	agttacatga	tcccccatgt	1620
tgtgcaaaaa	agcggttagc	tccttcggtc	ctccgatcgt	tgtcagaagt	aagttggccg	1680
cagtgttatc	actcatggtt	atggcagcac	tgcataattc	tcttactgtc	atgccatccg	1740
taagatgctt	ttctgtgact	ggtgagtact	caaccaagtc	attctgagaa	tagtgtatgc	1800
ggcgaccgag	ttgctcttgc	ccggcgtcaa	tacgggataa	taccgcgcca	catagcagaa	1860
ctttaaaagt	gctcatcatt	ggaaaacgtt	cttcgggggcg	aaaactctca	aggatcttac	1920
cgctgttgag	atccagttcg	atgtaaccca	ctcgtgcacc	caactgatct	tcagcatctt	1980
ttactttcac	cagcgtttct	gggtgagcaa	aaacaggaag	gcaaaatgcc	gcaaaaaagg	2040
gaataagggc	gacacggaaa	tgttgaatac	tcatactctt	cctttttcaa	tattattgaa	2100
gcatttatca	gggttattgt	ctcatgagcg	gatacatatt	tgaatgtatt	tagaaaaata	2160
aacaaatagg	ggttccgcgc	acatttcccc	gaaaagtgcc	acctgaacga	agcatctgtg	2220
cttcattttg	tagaacaaaa	atgcaacgcg	agagcgctaa	tttttcaaac	aaagaatctg	2280
agctgcattt	ttacagaaca	gaaatgcaac	gcgaaagcgc	tattttacca	acgaagaatc	2340
tgtgcttcat	ttttgtaaaa	caaaaatgca	acgcgagagc	gctaattttt	caaacaaaga	2400
atctgagctg	catttttaca	gaacagaaat	gcaacgcgag	agcgctattt	taccaacaaa	2460
gaatctatac	ttctttttg	ttctacaaaa	atgcatcccg	agagcgctat	ttttctaaca	2520
aagcatctta	gattactttt	tttctccttt	gtgcgctcta	taatgcagtc	tcttgataac	2580
tttttgcact	gtaggtccgt	taaggttaga	agaaggctac	tttggtgtct	attttctctt	2640
ccataaaaaa	agcetgaete	cacttcccgc	gtttactgat	tactagcgaa	gctgcgggtg	2700
cattttttca	agataaaggc	atccccgatt	atattctata	ccgatgtgga	ttgcgcatac	2760
tttgtgaaca	gaaagtgata	gcgttgatga	ttcttcattg	gtcagaaaat	tatgaacggt	2820
ttcttctatt	ttgtctctat	atactacgta	taggaaatgt	ttacattttc	gtattgtttt	2880
cgattcactc	tatgaatagt	tcttactaca	attttttgt	ctaaagagta	atactagaga	2940
taaacataaa	aaatgtagag	gtcgagttta	gatgcaagtt	caaggagcga	aaggtggatg	3000
ggtaggttat	atagggatat	agcacagaga	tatatagcaa	agagatactt	ttgagcaatg	3060
tttgtggaag	cggtattcgc	aatattttag	tagetegtta	cagtccggtg	cgtttttggt	3120
tttttgaaag	tgcgtcttca	gagcgctttt	ggttttcaaa	agcgctctga	agttcctata	3180
ctttctagag	aataggaact	tcggaatagg	aacttcaaag	cgtttccgaa	aacgagcgct	3240
tccgaaaatg	caacgcgagc	tgcgcacata	cageteactg	ttcacgtcgc	acctatatct	3300
gcgtgttgcc	tgtatatata	tatacatgag	aagaacggca	tagtgcgtgt	ttatgcttaa	3360
atgcgtactt	atatgcgtct	atttatgtag	gatgaaaggt	agtctagtac	ctcctgtgat	3420
attatcccat	tccatgcggg	gtatcgtatg	cttccttcag	cactaccctt	tagetgttet	3480
atatgctgcc	actcctcaat	tggattagtc	tcatccttca	atgctatcat	ttcctttgat	3540
attggatcat	ctaagaaacc	attattatca	tgacattaac	ctataaaaat	aggcgtatca	3600
cgaggccctt	tcgtctcgcg	cgtttcggtg	atgacggtga	aaacctctga	cacatgcagc	3660
tcccggagac	ggtcacagct	tgtctgtaag	cggatgccgg	gagcagacaa	gcccgtcagg	3720
gcgcgtcagc	gggtgttggc	gggtgtcggg	gctggcttaa	ctatgcggca	tcagagcaga	3780
ttgtactgag	agtgcaccat	aaattcccgt	tttaagagct	tggtgagcgc	taggagtcac	3840
tgccaggtat	cgtttgaaca	cggcattaqt	cagggaagtc	ataacacaqt	cettteeege	3900

415

### -continued

416

ggtaatgatt ttcattttt tttttcccct ageggatgac tcttttttt tcttagegat tggcattatc acataatgaa ttatacatta tataaagtaa tgtgatttct tcgaagaata tactaaaaaa tgagcaggca agataaacga aggcaaagat gacagagcag aaagccctag taaagcgtat tacaaatgaa accaagattc agattgcgat ctctttaaag ggtggtcccc tagcgataga gcactcgatc ttcccagaaa aagaggcaga agcagtagca gaacaggcca cacaatcgca agtgattaac gtccacacag gtatagggtt tctggaccat atgatacatg ctctggccaa gcattccggc tggtcgctaa tcgttgagtg cattggtgac ttacacatag acgaccatca caccactgaa gactgcggga ttgctctcgg tcaagctttt aaagaggccc tactggcgcg tggagtaaaa aggtttggat caggatttgc gcctttggat gaggcacttt ccagageggt ggtagatett tegaacagge egtaegeagt tgtegaaett ggtttgeaaa gggagaaagt aggagatete tettgegaga tgateeegea ttttettgaa agetttgeag aggetageag aattaccete caegttgatt gtetgegagg caagaatgat cateacegta gtgagagtgc gttcaaggct cttgcggttg ccataagaga agccacctcg cccaatggta ccaacgatgt tccctccacc aaaggtgttc ttatgtagtg acaccgatta tttaaagctg caqcatacqa tatatataca tqtqtatata tqtataccta tqaatqtcaq taaqtatqta tacqaacaqt atqatactqa aqatqacaaq qtaatqcatc attctatacq tqtcattctq aacgaggege gettteettt tttettttg ettttettt tttttetet tgaactegae qqatctatqc qqtqtqaaat accqcacaqa tqcqtaaqqa qaaaataccq catcaqqaaa ttgtaaacgt taatattttg ttaaaattcg cgttaaattt ttgttaaatc agctcatttt ttaaccaata ggccgaaatc ggcaaaatcc cttataaatc aaaagaatag accgagatag ggttgagtgt tgttccagtt tggaacaaga gtccactatt aaagaacgtg gactccaacg tcaaagggcg aaaaaccgtc tatcagggcg atggcccact acgtgaacca tcaccctaat caagtttttt ggggtcgagg tgccgtaaag cactaaatcg gaaccctaaa gggagccccc gatttagagc ttgacgggga aagccggcga acgtggcgag aaaggaaggg aagaaagcga aaggagcggg cgctagggcg ctggcaagtg tagcggtcac gctgcgcgta accaccacac ccgccgcgct taatgcgccg ctacagggcg cgtcgcgcca ttcgccattc aggctgcgca actgttggga agggcgatcg gtgcgggcct cttcgctatt acgccagctg gcgaaagggg gatgtgctgc aaggcgatta agttgggtaa cgccagggtt ttccccagtca cgacgttgta aaacgacggc cagtgagcgc gcgtaatacg actcactata gggcgaattg ggtaccgggc cccccctcga ggtattagaa gccgccgagc gggcgacagc cctccgacgg aagactctcc tecgtgegte etegtettea eeggtegegt teetgaaaeg eagatgtgee tegegeegea ctqctccqaa caataaaqat tctacaatac taqcttttat qqttatqaaq aqqaaaaatt ggcagtaacc tggccccaca aaccttcaaa ttaacgaatc aaattaacaa ccataggatg 5940 ataatgegat tagtttttta geettattte tggggtaatt aateagegaa gegatgattt 6000 ttgatctatt aacagatata taaatggaaa agctgcataa ccactttaac taatactttc 6060 aacattttca gtttgtatta cttcttattc aaatgtcata aaagtatcaa caaaaaattg 6120 ttaatatacc tctatacttt aacgtcaagg agaaaaatgt ccaatttact gcccgtacac 6180 caaaatttgc ctgcattacc ggtcgatgca acgagtgatg aggttcgcaa gaacctgatg 6240 gacatgttca gggatcgcca ggcgttttct gagcatacct ggaaaatgct tctgtccgtt 6300

417

tgccggt	cgt	gggcggcatg	gtgcaagttg	aataaccgga	aatggtttcc	cgcagaacct	6360
gaagatg	ltc	gcgattatct	tctatatctt	caggcgcgcg	gtctggcagt	aaaaactatc	6420
cagcaac	att	tgggccagct	aaacatgctt	catcgtcggt	ccgggctgcc	acgaccaagt	6480
gacagca	atg	ctgtttcact	ggttatgcgg	cggatccgaa	aagaaaacgt	tgatgccggt	6540
gaacgtg	Icaa	aacaggctct	agcgttcgaa	cgcactgatt	tcgaccaggt	tcgttcactc	6600
atggaaa	ata	gcgatcgctg	ccaggatata	cgtaatctgg	catttctggg	gattgcttat	6660
aacaccc	tgt	tacgtatagc	cgaaattgcc	aggatcaggg	ttaaagatat	ctcacgtact	6720
gacggtg	Igga	gaatgttaat	ccatattggc	agaacgaaaa	cgctggttag	caccgcaggt	6780
gtagaga	agg	cacttagcct	gggggtaact	aaactggtcg	agcgatggat	ttccgtctct	6840
ggtgtag	lctg	atgatccgaa	taactacctg	ttttgccggg	tcagaaaaaa	tggtgttgcc	6900
gcgccat	ctg	ccaccagcca	gctatcaact	cgcgccctgg	aagggatttt	tgaagcaact	6960
catcgat	tga	tttacggcgc	taaggatgac	tctggtcaga	gatacctggc	ctggtctgga	7020
cacagtg	laca	gtgtcggagc	cgcgcgagat	atggcccgcg	ctggagtttc	aataccggag	7080
atcatgo	aag	ctggtggctg	gaccaatgta	aatattgtca	tgaactatat	ccgtaacctg	7140
gatagtg	Jaaa	caggggcaat	ggtgcgcctg	ctggaagatg	gcgattagga	gtaagcgaat	7200
ttcttat	gat	ttatgatttt	tattattaaa	taagttataa	aaaaaataag	tgtatacaaa	7260
ttttaaa	gtg	actcttaggt	tttaaaacga	aaattcttat	tcttgagtaa	ctctttcctg	7320
taggtca	ggt	tgctttctca	ggtatagcat	gaggtcgctc	ttattgacca	cacctctacc	7380
ggcatgo	cga	gcaaatgcct	gcaaatcgct	ccccatttca	cccaattgta	gatatgctaa	7440
ctccage	aat	gagttgatga	atctcggtgt	gtattttatg	tcctcagagg	acaacacctg	7500
tggtccg	Icca	ccgcggtgga	gct				7523
<210> SEQ ID NO 185 <211> LENGTH: 96 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: LA722							
taccent	+ =+	ttacctaaac	atctataacc	ttaaaaata	222222taca	casacattaa	60
atcatca	act	taactaacta	attatatat	cactoo	aaaaaacaca	caaacyctya	96
accacca		eggeeddeee	geegeaceae	cacegg			50
<210> S <211> L <212> T <213> C <220> F <223> C	EQ ENG YPE RGAI EATU THE	ID NO 186 TH: 80 : DNA VISM: Artif: JRE: R INFORMATIC	icial Sequer DN: LA733	nce			
<400> S	EQUI	ENCE: 186					
cataatc	aat	ctcaaagaga	acaacacaat	acaataacaa	gaagaacaaa	gcattgcgga	60
ttacgta	ittc	taatgttcag					80
<210> S <211> L <212> T <213> O <220> E	EQ : ENG YPE RGAI	ID NO 187 FH: 30 : DNA NISM: Artif: IRF:	icial Sequer	nce			

<sup>&</sup>lt;223> OTHER INFORMATION: LA453

	-1	~
		· •
_		

,....

420

-continued	
<400> SEQUENCE: 187	
caccgaagaa gaatgcaaaa atttcagctc	30
<210> SEQ ID NO 188 <211> LENGTH: 25 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: LA694	
<400> SEQUENCE: 188	
gctgaagttg ttagaactgt tgttg	25
<210> SEQ ID NO 189 <211> LENGTH: 21 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: LA695	
CAUDA SEQUENCE: 189	21
<pre>&lt;210&gt; SEQ ID NO 190 &lt;211&gt; LENGTH: 22 &lt;212&gt; TYPE: DNA &lt;213&gt; ORGANISM: Artificial Sequence &lt;220&gt; FEATURE: &lt;223&gt; OTHER INFORMATION: oBP594</pre>	21
<400> SEQUENCE: 190	
agetgteteg tgttgtgggt tt	22
<pre>&lt;210&gt; SEQ ID NO 191 &lt;211&gt; LENGTH: 49 &lt;212&gt; TYPE: DNA &lt;213&gt; ORGANISM: Artificial Sequence &lt;220&gt; FEATURE: &lt;223&gt; OTHER INFORMATION: oBP595 &lt;400&gt; SEQUENCE: 191</pre>	
cttaataata gaacaatato atootttacg gocatottat agtgtogtt	49
<210> SEQ ID NO 192 <211> LENGTH: 49 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: oBP596	
<400> SEQUENCE: 192	
gcgccaacga cactataaga tgcccgtaaa ggatgatatt gttctatta	49
<pre>&lt;210&gt; SEQ ID NO 193 &lt;211&gt; LENGTH: 49 &lt;212&gt; TYPE: DNA &lt;213&gt; ORGANISM: Artificial Sequence &lt;220&gt; FEATURE: &lt;223&gt; OTHER INFORMATION: oBP597 </pre>	
<400> SEQUENCE: 193	
tatggaccct gaaaccacag ccacattgca acgacgacaa tgccaaacc	49
-210- CEO ID NO 104	

<210> SEQ ID NO 194 <211> LENGTH: 49

continued

-continued	
<pre>&lt;212&gt; TYPE: DNA &lt;213&gt; ORGANISM: Artificial Sequence &lt;220&gt; FEATURE:</pre>	
<223> OTHER INFORMATION: OBP598	
<400> SEQUENCE: 194	
teettggttt ggeattgteg tegttgeaat gtggetgtgg ttteagggt	49
<210> SEQ ID NO 195 <211> LENGTH: 49 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: oBP599	
<400> SEQUENCE: 195	
atcetetege ggagteeetg tteagtaaag geeatgaage tttttettt	49
<210> SEQ ID NO 196 <211> LENGTH: 49 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: oBP600	
<400> SEQUENCE: 196	
attggaaaga aaaagcttca tggcctttac tgaacaggga ctccgcgag	49
<pre>&lt;210&gt; SEQ ID NO 197 &lt;211&gt; LENGTH: 22 &lt;212&gt; TYPE: DNA &lt;213&gt; ORGANISM: Artificial Sequence &lt;220&gt; FEATURE: &lt;223&gt; OTHER INFORMATION: oBP601 &lt;400&gt; SEOUENCE: 197</pre>	
tcataccaca atcttagacc at	22
<pre>&lt;210&gt; SEQ ID NO 198 &lt;211&gt; LENGTH: 21 &lt;212&gt; TYPE: DNA &lt;213&gt; ORGANISM: Artificial Sequence &lt;220&gt; FEATURE: &lt;223&gt; OTHER INFORMATION: oBP602 &lt;400&gt; SEOUENCE: 198</pre>	
	21
<pre>&lt;210&gt; SEQ ID NO 199 &lt;211&gt; LENGTH: 22 &lt;212&gt; TYPE: DNA &lt;213&gt; ORGANISM: Artificial Sequence &lt;220&gt; FEATURE: &lt;223&gt; OTHER INFORMATION: oBP603 &lt;400&gt; SEQUENCE: 199</pre>	
tgtteccaca atctattacc ta	22
<210> SEQ ID NO 200 <211> LENGTH: 31 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: LA811	
<400> SEQUENCE: 200	

-continued

aacgaagcat ctgtgcttca ttttgtagaa c	31
<210> SEQ ID NO 201 <211> LENGTH: 59 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: LA817	
<400> SEQUENCE: 201	
cgatccactt gtatatttgg atgaattttt gaggaattct gaaccagtcc taaaacgag	59
<pre>&lt;210&gt; SEQ ID NO 202 &lt;211&gt; LENGTH: 31 212&gt; TYPE: DNA &lt;213&gt; ORGANISM: Artificial Sequence &lt;220&gt; FEATURE: &lt;223&gt; OTHER INFORMATION: LA812 &lt;400&gt; SEQUENCE: 202</pre>	
aacaaagata tootattoaa gtocaagato g	31
<pre>&lt;210&gt; SEQ ID NO 203 &lt;211&gt; LENGTH: 33 &lt;212&gt; TYPE: DNA &lt;213&gt; ORGANISM: Artificial Sequence &lt;220&gt; FEATURE: &lt;223&gt; OTHER INFORMATION: LA818</pre>	
<400> SEQUENCE: 203	
ctcaaaaatt catccaaata tacaagtgga tcg	33
<210> SEQ ID NO 204 <211> LENGTH: 90 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: LA512	
<400> SEQUENCE: 204	
gtattttggt agattcaatt ctctttccct ttccttttcc ttcgctcccc ttccttatca	60
gcattgcgga ttacgtattc taatgttcag	90
<210> SEQ ID NO 205 <211> LENGTH: 90 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: LA513	
<400> SEQUENCE: 205	
ttggttgggg gaaaaagagg caacaggaaa gatcagaggg ggaggggggg ggagagtgtc	60
accttggcta actcgttgta tcatcactgg	90
<210> SEQ ID NO 206 <211> LENGTH: 29 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: LA516	
<400> SEQUENCE: 206	
ctcgaaacaa taagacgacg atggctctg	29

continued

	-concinu	lea	
<pre>&lt;210&gt; SEQ ID NO 207 &lt;211&gt; LENGTH: 30 &lt;212&gt; TYPE: DNA &lt;213&gt; ORGANISM: Artificial Sequence &lt;220&gt; FEATURE: &lt;223&gt; OTHER INFORMATION: LA514</pre>			
<400> SEQUENCE: 207			
cactatctgg tgcaaacttg gcaccggaag		30	
<pre>&lt;210&gt; SEQ ID NO 208 &lt;211&gt; LENGTH: 29 &lt;212&gt; TYPE: DNA &lt;213&gt; ORGANISM: Artificial Sequence &lt;220&gt; FEATURE: &lt;223&gt; OTHER INFORMATION: LA515</pre>			
<400> SEQUENCE: 208			
tgtttgtagc cactcgtgaa cttctctgc		29	
<pre>&lt;210&gt; SEQ ID NO 209 &lt;211&gt; LENGTH: 6903 &lt;212&gt; TYPE: DNA &lt;213&gt; ORGANISM: Artificial Sequence &lt;220&gt; FEATURE: &lt;223&gt; OTHER INFORMATION: pLA71</pre>			
<400> SEQUENCE: 209			
aaacgccagc aacgcggcct ttttacggtt cctgg	cettt tgetggeett t	tgctcacat 60	
gttettteet gegttateee etgattetgt ggata	accgt attaccgcct t	tgagtgagc 120	
tgataccgct cgccgcagcc gaacgaccga gcgca	gcgag tcagtgagcg a	lggaagcgga 180	
agagegeeca ataegeaaae egeeteteee egege	gttgg ccgattcatt a	atgcagctg 240	
gcacgacagg tttcccgact ggaaagcggg cagto	agcgc aacgcaatta a	tgtgagtta 300	
gctcactcat taggcacccc aggctttaca cttta	tgett eeggetegta t	gttgtgtgg 360	
aattgtgagc ggataacaat ttcacacagg aaaca	gctat gaccatgatt a	cgccaagct 420	
tgcatgcgat ctgaaatgaa taacaatact gacag	tagat ctgaaatgaa t	aacaatact 480	
gacagtacta aataattgcc tacttggctt cacat	acgtt gcatacgtcg a	tatagataa 540	
taatgataat gacagcagga ttatcgtaat acgta	atagt tgaaaatctc a	aaaatgtgt 600	
gggtcattac gtaaataatg ataggaatgg gatto	ttcta tttttccttt t	tccattcta 660	
gcagccgtcg ggaaaacgtg gcatcctctc tttcc	ggctc aattggagtc a	cgctgccgt 720	
gagcateete tettteeata tetaacaaet gagea	cgtaa ccaatggaaa a	gcatgagct 780	
tagcgttgct ccaaaaaagt attggatggt taata	ccatt tgtctgttct c	ttctgactt 840	
tgactcctca aaaaaaaaaa atctacaatc aacag	atcgc ttcaattacg c	cctcacaaa 900	
aacttttttc cttcttcttc gcccacgtta aattt	tatee eteatgttgt e	taacggatt 960	
tctgcacttg atttattata aaaagacaaa gacat	aatac ttctctatca a	tttcagtta 1020	
ttgttettee ttgegttatt ettetgttet tettt	ttett ttgteatata t	aaccataac 1080	
caagtaatac atattcaaat ctagagctga ggatg	ttgac aaaagcaaca a	aagaacaaa 1140	
aatcoottgt gaaaaacaga ggggoggago ttgtt	gttga ttgcttagtg g	agcaaggtg 1200	
tcacacatgt atttggcatt ccaggtgcaa aaatt	gatgc ggtatttgac g	ctttacaag 1260	
ataaaggacc tgaaattatc gttgcccggc acgaa	caaaa cgcagcattc a	tggcccaag 1320	
cagteggeeg tttaaetgga aaaeegggag tegte	ttagt cacatcagga c	cgggtgcct 1380	

				-contir	nued	
ctaacttggc	aacaggcctg	ctgacagcga	acactgaagg	agaccctgtc	gttgcgcttg	1440
ctggaaacgt	gatccgtgca	gatcgtttaa	aacggacaca	tcaatctttg	gataatgcgg	1500
cgctattcca	gccgattaca	aaatacagtg	tagaagttca	agatgtaaaa	aatataccgg	1560
aagctgttac	aaatgcattt	aggatagcgt	cagcagggca	ggctggggcc	gcttttgtga	1620
gctttccgca	agatgttgtg	aatgaagtca	caaatacgaa	aaacgtgcgt	gctgttgcag	1680
cgccaaaact	cggtcctgca	gcagatgatg	caatcagtgc	ggccatagca	aaaatccaaa	1740
cagcaaaact	tcctgtcgtt	ttggtcggca	tgaaaggcgg	aagaccggaa	gcaattaaag	1800
cggttcgcaa	gcttttgaaa	aaggttcagc	ttccatttgt	tgaaacatat	caagetgeeg	1860
gtaccctttc	tagagattta	gaggatcaat	attttggccg	tatcggtttg	ttccgcaacc	1920
agcctggcga	tttactgcta	gagcaggcag	atgttgttct	gacgatcggc	tatgacccga	1980
ttgaatatga	tccgaaattc	tggaatatca	atggagaccg	gacaattatc	catttagacg	2040
agattatcgc	tgacattgat	catgcttacc	agcctgatct	tgaattgatc	ggtgacattc	2100
cgtccacgat	caatcatatc	gaacacgatg	ctgtgaaagt	ggaatttgca	gagcgtgagc	2160
agaaaatcct	ttctgattta	aaacaatata	tgcatgaagg	tgagcaggtg	cctgcagatt	2220
ggaaatcaga	cagagcgcac	cctcttgaaa	tcgttaaaga	gttgcgtaat	gcagtcgatg	2280
atcatgttac	agtaacttgc	gatatcggtt	cgcacgccat	ttggatgtca	cgttatttcc	2340
gcagctacga	gccgttaaca	ttaatgatca	gtaacggtat	gcaaacactc	ggcgttgcgc	2400
ttccttgggc	aatcggcgct	tcattggtga	aaccgggaga	aaaagtggtt	tctgtctctg	2460
gtgacggcgg	tttcttattc	tcagcaatgg	aattagagac	agcagttcga	ctaaaagcac	2520
caattgtaca	cattgtatgg	aacgacagca	catatgacat	ggttgcattc	cagcaattga	2580
aaaaatataa	ccgtacatct	gcggtcgatt	tcggaaatat	cgatatcgtg	aaatatgcgg	2640
aaagcttcgg	agcaactggc	ttgcgcgtag	aatcaccaga	ccagctggca	gatgttctgc	2700
gtcaaggcat	gaacgctgaa	ggtcctgtca	tcatcgatgt	cccggttgac	tacagtgata	2760
acattaattt	agcaagtgac	aagcttccga	aagaattcgg	ggaactcatg	aaaacgaaag	2820
ctctctagtt	aattaatcat	gtaattagtt	atgtcacgct	tacattcacg	ccctcccccc	2880
acatccgctc	taaccgaaaa	ggaaggagtt	agacaacctg	aagtctaggt	ccctatttat	2940
ttttttatag	ttatgttagt	attaagaacg	ttatttatat	ttcaaatttt	tctttttt	3000
ctgtacagac	gcgtgtacgc	atgtaacatt	atactgaaaa	ccttgcttga	gaaggttttg	3060
ggacgctcga	aggctttaat	ttaggttttg	ggacgctcga	aggctttaat	ttggatccgc	3120
attgcggatt	acgtattcta	atgttcagta	ccgttcgtat	aatgtatgct	atacgaagtt	3180
atgcagattg	tactgagagt	gcaccatacc	acagetttte	aattcaattc	atcattttt	3240
ttttattctt	ttttttgatt	tcggtttctt	tgaaattttt	ttgattcggt	aatctccgaa	3300
cagaaggaag	aacgaaggaa	ggagcacaga	cttagattgg	tatatatacg	catatgtagt	3360
gttgaagaaa	catgaaattg	cccagtattc	ttaacccaac	tgcacagaac	aaaaacctgc	3420
aggaaacgaa	gataaatcat	gtcgaaagct	acatataagg	aacgtgctgc	tactcatcct	3480
agtcctgttg	ctgccaagct	atttaatatc	atgcacgaaa	agcaaacaaa	cttgtgtgct	3540
tcattggatg	ttcgtaccac	caaggaatta	ctggagttag	ttgaagcatt	aggtcccaaa	3600
atttgtttac	taaaaacaca	tgtggatatc	ttgactgatt	tttccatgga	gggcacagtt	3660
aagccgctaa	aggcattatc	cgccaagtac	aatttttac	tcttcgaaga	cagaaaattt	3720
gctgacattg	gtaatacagt	caaattgcag	tactctgcgg	gtgtatacag	aatagcagaa	3780

429

### -continued

tgggcagaca	ttacgaatgc	acacggtgtg	gtgggcccag	gtattgttag	cggtttgaag	3840
caggcggcag	aagaagtaac	aaaggaacct	agaggccttt	tgatgttagc	agaattgtca	3900
tgcaagggct	ccctatctac	tggagaatat	actaagggta	ctgttgacat	tgcgaagagc	3960
gacaaagatt	ttgttatcgg	ctttattgct	caaagagaca	tgggtggaag	agatgaaggt	4020
tacgattggt	tgattatgac	acccggtgtg	ggtttagatg	acaagggaga	cgcattgggt	4080
caacagtata	gaaccgtgga	tgatgtggtc	tctacaggat	ctgacattat	tattgttgga	4140
agaggactat	ttgcaaaggg	aagggatgct	aaggtagagg	gtgaacgtta	cagaaaagca	4200
ggctgggaag	catatttgag	aagatgcggc	cagcaaaact	aaaaaactgt	attataagta	4260
aatgcatgta	tactaaactc	acaaattaga	gcttcaattt	aattatatca	gttattaccc	4320
tatgcggtgt	gaaataccgc	acagatgcgt	aaggagaaaa	taccgcatca	ggaaattgta	4380
aacgttaata	ttttgttaaa	attcgcgtta	aatttttgtt	aaatcagctc	attttttaac	4440
caataggccg	aaatcggcaa	aatcccttat	aaatcaaaag	aatagaccga	gatagggttg	4500
agtgttgttc	cagtttggaa	caagagtcca	ctattaaaga	acgtggactc	caacgtcaaa	4560
gggcgaaaaa	ccgtctatca	gggcgatggc	ccactacgtg	aaccatcacc	ctaatcaaga	4620
taacttcgta	taatgtatgc	tatacgaacg	gtaccagtga	tgatacaacg	agttagccaa	4680
ggtgaattca	ctggccgtcg	ttttacaacg	tcgtgactgg	gaaaaccctg	gcgttaccca	4740
acttaatcgc	cttgcagcac	atcccccttt	cgccagctgg	cgtaatagcg	aagaggcccg	4800
caccgatcgc	ccttcccaac	agttgcgcag	cctgaatggc	gaatggcgcc	tgatgcggta	4860
ttttctcctt	acgcatctgt	gcggtatttc	acaccgcata	tggtgcactc	tcagtacaat	4920
ctgctctgat	gccgcatagt	taagccagcc	ccgacacccg	ccaacacccg	ctgacgcgcc	4980
ctgacgggct	tgtctgctcc	cggcatccgc	ttacagacaa	gctgtgaccg	tctccgggag	5040
ctgcatgtgt	cagaggtttt	caccgtcatc	accgaaacgc	gcgagacgaa	agggcctcgt	5100
gatacgccta	tttttatagg	ttaatgtcat	gataataatg	gtttcttaga	cgtcaggtgg	5160
cacttttcgg	ggaaatgtgc	gcggaacccc	tatttgttta	tttttctaaa	tacattcaaa	5220
tatgtatccg	ctcatgagac	aataaccctg	ataaatgctt	caataatatt	gaaaaaggaa	5280
gagtatgagt	attcaacatt	tccgtgtcgc	ccttattccc	tttttgcgg	cattttgcct	5340
tcctgttttt	gctcacccag	aaacgctggt	gaaagtaaaa	gatgctgaag	atcagttggg	5400
tgcacgagtg	ggttacatcg	aactggatct	caacagcggt	aagatccttg	agagttttcg	5460
ccccgaagaa	cgttttccaa	tgatgagcac	ttttaaagtt	ctgctatgtg	gcgcggtatt	5520
atcccgtatt	gacgccgggc	aagagcaact	cggtcgccgc	atacactatt	ctcagaatga	5580
cttggttgag	tactcaccag	tcacagaaaa	gcatcttacg	gatggcatga	cagtaagaga	5640
attatgcagt	gctgccataa	ccatgagtga	taacactgcg	gccaacttac	ttctgacaac	5700
gatcggagga	ccgaaggagc	taaccgcttt	tttgcacaac	atgggggatc	atgtaactcg	5760
ccttgatcgt	tgggaaccgg	agctgaatga	agccatacca	aacgacgagc	gtgacaccac	5820
gatgcctgta	gcaatggcaa	caacgttgcg	caaactatta	actggcgaac	tacttactct	5880
agetteeegg	caacaattaa	tagactggat	ggaggcggat	aaagttgcag	gaccacttct	5940
gcgctcggcc	cttccggctg	gctggtttat	tgctgataaa	tctggagccg	gtgagcgtgg	6000
gtctcgcggt	atcattgcag	cactggggcc	agatggtaag	ccctcccgta	tcgtagttat	6060
ctacacgacg	gggagtcagg	caactatgga	tgaacgaaat	agacagatcg	ctgagatagg	6120

431

# -continued

	6180
tgatttaaaa cttcattttt aatttaaaag gatctaggtg aagatccttt ttgataatct	6240
catgaccaaa atcccttaac gtgagttttc gttccactga gcgtcagacc ccgtagaaaa	6300
gatcaaagga tettettgag ateettttt tetgegegta atetgetget tgeaaacaaa	6360
aaaaccaccg ctaccagcgg tggtttgttt gccggatcaa gagctaccaa ctctttttcc	6420
gaaggtaact ggcttcagca gagcgcagat accaaatact gtccttctag tgtagccgta	6480
gttaggccac cacttcaaga actetgtage acegeetaca taeetegete tgetaateet	6540
gttaccagtg gctgctgcca gtggcgataa gtcgtgtctt accgggttgg actcaagacg	6600
atagttaccg gataaggcgc agcggtcggg ctgaacgggg ggttcgtgca cacagcccag	6660
cttggagcga acgacctaca ccgaactgag atacctacag cgtgagctat gagaaagcgc	6720
cacgetteee gaagggagaa aggeggacag gtateeggta ageggeaggg teggaacagg	6780
agagcgcacg agggagcttc caggggggaaa cgcctggtat ctttatagtc ctgtcgggtt	6840
tcgccacctc tgacttgagc gtcgattttt gtgatgctcg tcagggggggg ggagcctatg	6900
gaa	6903
<210> SEQ ID NO 210 <211> LENGTH: 6924 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: pLA78 <400> SEQUENCE: 210	
qatccqcatt qcqqattacq tattctaatq ttcaqtaccq ttcqtataat qtatgctata	60
cgaagttatg cagattgtac tgagagtgca ccataccacc ttttcaattc atcattttt	120
ttttattett ttttttgatt teggttteet tgaaattttt ttgatteggt aateteegaa	180
cagaaggaag aacgaaggaa ggagcacaga cttagattgg tatatatacg catatgtagt	240
gttgaagaaa catgaaattg cccagtattc ttaacccaac tgcacagaac aaaaacctgc	300
aggaaacgaa gataaatcat gtcgaaagct acatataagg aacgtgctgc tactcatcct	360
agteetgttg etgecaaget atttaatate atgeaegaaa ageaaaeaaa ettgtgtget	420
tcattggatg ttcgtaccac caaggaatta ctggagttag ttgaagcatt aggtcccaaa	480
attigittac taaaaacaca igiggataic tigacigati titiccaigga gggcacagit	540
aagccgctaa aggcattatc cgccaagtac aattttttac tcttcgaaga cagaaaattt	600
aagccgctaa aggcattatc cgccaagtac aattttttac tcttcgaaga cagaaaattt gctgacattg gtaatacagt caaattgcag tactctgcgg gtgtatacag aatagcagaa	600 660
aagccgctaa aggcattatc cgccaagtac aattttttac tcttcgaaga cagaaaattt gctgacattg gtaatacagt caaattgcag tactctgcgg gtgtatacag aatagcagaa tgggcagaca ttacgaatgc acacggtgtg gtgggcccag gtattgttag cggtttgaag	600 660 720
aagccgctaa aggcattatc cgccaagtac aatttttac tottogaaga cagaaaattt gotgacattg gtaatacagt caaattgcag tactotgogg gtgtatacag aatagcagaa tgggcagaca ttacgaatgc acacggtgtg gtggggcccag gtattgttag oggtttgaag caggoggcag aagaagtaac aaaggaacot agaggcottt tgatgttagc agaattgtca	600 660 720 780
aagccgctaa aggcattatc cgccaagtac aatttttac tcttcgaaga cagaaaattt gctgacattg gtaatacagt caaattgcag tactctgcgg gtgtatacag aatagcagaa tgggcagaca ttacgaatgc acacggtgtg gtgggcccag gtattgttag cggtttgaag caggcggcag aagaagtaac aaaggaacct agaggccttt tgatgttagc agaattgtca tgcaagggct ccctatctac tggagaatat actaagggta ctgttgacat tgcgaagagc	600 660 720 780 840
aagccgctaa aggcattatc cgccaagtac aatttttac tottogaaga cagaaaattt gotgacattg gtaatacagt caaattgcag tactotgogg gtgtatacag aatagcagaa tgggcagaca ttacgaatgc acacggtgtg gtgggcccag gtattgttag oggtttgaag caggoggcag aagaagtaac aaaggaacot agaggoottt tgatgttagc agaattgtca tgcaagggot occtatotac tggagaatat actaagggta otgttgacat tgogaagagc gacaaagatt ttgttatogg otttattgot caaagagaca tgggtggaag agatgaaggt	600 660 720 780 840 900
aagccgctaa aggcattatc cgccaagtac aatttttac tottogaaga cagaaaattt gotgacattg gtaatacagt caaattgcag tactotgogg gtgtatacag aatagcagaa tgggcagaca ttacgaatgc acacggtgtg gtgggcccag gtattgttag oggtttgaag caggoggcag aagaagtaac aaaggaacot agaggcottt tgatgttagc agaattgtca tgcaagggot cootatotac tggagaatat actaagggta otgttgacat tgogaagago gacaaagatt ttgttatogg otttattgot caaagagaca tgggtggaag agatgaaggt tacgattggt tgattatgac accoggtgtg ggtttagatg acaagggaga cgoattgggt	600 660 720 780 840 900 960
aagccgctaa aggcattatc cgccaagtac aatttttac tcttcgaaga cagaaaattt gctgacattg gtaatacagt caaattgcag tactctgcgg gtgtatacag aatagcagaa tgggcagaca ttacgaatgc acacggtgtg gtgggcccag gtattgttag cggtttgaag caggcggcag aagaagtaac aaaggaacct agaggccttt tgatgttagc agaattgtca tgcaagggct ccctatctac tggagaatat actaagggta ctgttgacat tgcgaagagc gacaaagatt ttgttatcgg ctttattgct caaagagaca tgggtggaag agatgaaggt tacgattggt tgattatgac acccggtgtg ggtttagatg acaagggaga cgcattgggt caacagtata gaaccgtgga tgatgtggtc tctacaggat ctgacattat tattgttgga	600 660 720 780 840 900 960 1020
aagccgctaa aggcattatc cgccaagtac aatttttac tottogaaga cagaaaattt gotgacattg gtaatacagt caaattgcag tactotgogg gtgtatacag aatagcagaa tgggcagaca ttacgaatgo acacggtgtg gtgggcocag gtattgttag oggtttgaag caggoggcag aagaagtaac aaaggaacot agaggoottt tgatgttago agaattgtoa tgcaagggot ocotatotac tggagaatat actaagggta otgttgacat tgogaagago gacaaagatt ttgttatogg otttattgot caaagagaca tgggtggaag agatgaaggt tacgattggt tgattatgac accoggtgtg ggtttagatg acaagggaga ogcattgggt caacagtata gaaccgtgga tgatgtggto totacaggat otgacattat tattgttgga agaggactat ttgcaaaggg aagggatgot aaggtagagg gtgaacgtta cagaaaagca	600 660 720 780 840 900 960 1020
aagccgctaa aggcattatc cgccaagtac aatttttac tottogaaga cagaaaattt gotgacattg gtaatacagt caaattgcag tactotgogg gtgtatacag aatagcagaa tgggcagaca ttacgaatgc acacggtgtg gtgggcccag gtattgttag oggttgaag caggoggcag aagaagtaac aaaggaacot agaggoottt tgatgttago agaattgtca tgcaagggot ocotatotac tggagaatat actaagggta otgttgacat tgogaagago gacaaagatt ttgttatogg otttattgot caaagagaca tggggtggaag agatgaaggt tacgattggt tgattatgac accoggtgtg ggtttagatg acaagggaga ogoattgggt caacagtata gaacogtgga tgatgtggto totacaggat otgacattat tattgttgga agaggactat ttgcaaaggg aagggatgot aaggtagagg gtgaacgtta cagaaaagca ggctgggaag catattgag agatgoggo cagcaaaact aaaaaactgt attataagta	600 660 720 780 840 900 960 1020 1080 1140
aagccgctaa aggcattatc cgccaagtac aatttttac tottogaaga cagaaaattt gotgacattg gtaatacagt caaattgcag tactotgogg gtgtatacag aatagcagaa tgggcagaca ttacgaatgc acacggtgtg gtgggcccag gtattgttag oggtttgaag caggoggcag aagaagtaac aaaggaacot agaggocttt tgatgttago agaattgtoa tgcaagggot ocotatotac tggagaatat actaagggta otgttgacat tgogaagago gacaaagatt ttgttatogg otttattgot caaagagaca tgggtggaag agatgaaggt tacgattggt tgattatgac accoggtgtg ggtttagatg acaagggaga ogcattgggt caacagtata gaacogtgga tgatgtggto totacaggat otgacattat tattgttgga agaggactat ttgcaaaggg aagggatgot aaggtagagg gtgaacgtta cagaaaagca ggotgggaag catatttgag aagatgoggo cagcaaaact aaaaaactgt attataagta aatgcatgta tactaaacto acaaattaga gottcaattt aattataca gttattacoo	600 660 720 780 840 900 960 1020 1080 1140 1200

433

aacgttaata	ttttgttaaa	attcgcgtta	aatttttgtt	aaatcagctc	atttttaac	1320
caataggccg	aaatcggcaa	aatcccttat	aaatcaaaag	aatagaccga	gatagggttg	1380
agtgttgttc	cagtttggaa	caagagtcca	ctattaaaga	acgtggactc	caacgtcaaa	1440
gggcgaaaaa	ccgtctatca	gggcgatggc	ccactacgtg	aaccatcacc	ctaatcaaga	1500
taacttcgta	taatgtatgc	tatacgaacg	gtaccagtga	tgatacaacg	agttagccaa	1560
ggtgaattca	ctggccgtcg	ttttacaacg	tcgtgactgg	gaaaaccctg	gcgttaccca	1620
acttaatcgc	cttgcagcac	atcccccttt	cgccagctgg	cgtaatagcg	aagaggcccg	1680
caccgatcgc	ccttcccaac	agttgcgcag	cctgaatggc	gaatggcgcc	tgatgcggta	1740
ttttctcctt	acgcatctgt	gcggtatttc	acaccgcata	tggtgcactc	tcagtacaat	1800
ctgctctgat	gccgcatagt	taagccagcc	ccgacacccg	ccaacacccg	ctgacgcgcc	1860
ctgacgggct	tgtctgctcc	cggcatccgc	ttacagacaa	gctgtgaccg	tctccgggag	1920
ctgcatgtgt	cagaggtttt	caccgtcatc	accgaaacgc	gcgagacgaa	agggcctcgt	1980
gatacgccta	tttttatagg	ttaatgtcat	gataataatg	gtttcttaga	cgtcaggtgg	2040
cacttttcgg	ggaaatgtgc	gcggaacccc	tatttgttta	tttttctaaa	tacattcaaa	2100
tatgtatccg	ctcatgagac	aataaccctg	ataaatgctt	caataatatt	gaaaaaggaa	2160
gagtatgagt	attcaacatt	tccgtgtcgc	ccttattccc	tttttgcgg	cattttgcct	2220
tcctgttttt	gctcacccag	aaacgctggt	gaaagtaaaa	gatgctgaag	atcagttggg	2280
tgcacgagtg	ggttacatcg	aactggatct	caacagcggt	aagatccttg	agagttttcg	2340
ccccgaagaa	cgttttccaa	tgatgagcac	ttttaaagtt	ctgctatgtg	gcgcggtatt	2400
atcccgtatt	gacgccgggc	aagagcaact	cggtcgccgc	atacactatt	ctcagaatga	2460
cttggttgag	tactcaccag	tcacagaaaa	gcatcttacg	gatggcatga	cagtaagaga	2520
attatgcagt	gctgccataa	ccatgagtga	taacactgcg	gccaacttac	ttctgacaac	2580
gatcggagga	ccgaaggagc	taaccgcttt	tttgcacaac	atgggggatc	atgtaactcg	2640
ccttgatcgt	tgggaaccgg	agctgaatga	agccatacca	aacgacgagc	gtgacaccac	2700
gatgcctgta	gcaatggcaa	caacgttgcg	caaactatta	actggcgaac	tacttactct	2760
agcttcccgg	caacaattaa	tagactggat	ggaggcggat	aaagttgcag	gaccacttct	2820
gcgctcggcc	ctteeggetg	gctggtttat	tgctgataaa	tctggagccg	gtgagcgtgg	2880
gtctcgcggt	atcattgcag	cactgggggcc	agatggtaag	ccctcccgta	tcgtagttat	2940
ctacacgacg	gggagtcagg	caactatgga	tgaacgaaat	agacagatcg	ctgagatagg	3000
tgcctcactg	attaagcatt	ggtaactgtc	agaccaagtt	tactcatata	tactttagat	3060
tgatttaaaa	cttcattttt	aatttaaaag	gatctaggtg	aagatccttt	ttgataatct	3120
catgaccaaa	atcccttaac	gtgagttttc	gttccactga	gcgtcagacc	ccgtagaaaa	3180
gatcaaagga	tcttcttgag	atcctttttt	tctgcgcgta	atctgctgct	tgcaaacaaa	3240
aaaaccaccg	ctaccagcgg	tggtttgttt	gccggatcaa	gagctaccaa	ctctttttcc	3300
gaaggtaact	ggcttcagca	gagcgcagat	accaaatact	gtccttctag	tgtagccgta	3360
gttaggccac	cacttcaaga	actctgtagc	accgcctaca	tacctcgctc	tgctaatcct	3420
gttaccagtg	gctgctgcca	gtggcgataa	gtcgtgtctt	accgggttgg	actcaagacg	3480
atagttaccg	gataaggcgc	agcggtcggg	ctgaacgggg	ggttcgtgca	cacageceag	3540
cttggagcga	acgacctaca	ccgaactgag	atacctacag	cgtgagctat	gagaaagcgc	3600

435

## -continued

cacgcttccc	gaagggagaa	aggcggacag	gtatccggta	agcggcaggg	tcggaacagg	3660
agagcgcacg	agggagcttc	cagggggaaa	cgcctggtat	ctttatagtc	ctgtcgggtt	3720
tcgccacctc	tgacttgagc	gtcgattttt	gtgatgctcg	tcagggggggc	ggagcctatg	3780
gaaaaacgcc	agcaacgcgg	cctttttacg	gttcctggcc	ttttgctggc	cttttgctca	3840
catgttcttt	cctgcgttat	cccctgattc	tgtggataac	cgtattaccg	cctttgagtg	3900
agctgatacc	gctcgccgca	gccgaacgac	cgagcgcagc	gagtcagtga	gcgaggaagc	3960
ggaagagcgc	ccaatacgca	aaccgcctct	ccccgcgcgt	tggccgattc	attaatgcag	4020
ctggcacgac	aggtttcccg	actggaaagc	gggcagtgag	cgcaacgcaa	ttaatgtgag	4080
ttagctcact	cattaggcac	cccaggcttt	acactttatg	cttccggctc	gtatgttgtg	4140
tggaattgtg	agcggataac	aatttcacac	aggaaacagc	tatgaccatg	attacgccaa	4200
gcttccaatt	accgtcgctc	gtgatttgtt	tgcaaaaaga	acaaaactga	aaaaacccag	4260
acacgctcga	cttcctgtct	tcctattgat	tgcagcttcc	aatttcgtca	cacaacaagg	4320
tcctgtcgac	gcctacttgg	cttcacatac	gttgcatacg	tcgatataga	taataatgat	4380
aatgacagca	ggattatcgt	aatacgtaat	agttgaaaat	ctcaaaaatg	tgtgggtcat	4440
tacgtaaata	atgataggaa	tgggattctt	ctattttcc	tttttccatt	ctagcagccg	4500
tcgggaaaac	gtggcatcct	ctctttcggg	ctcaattgga	gtcacgctgc	cgtgagcatc	4560
ctctctttcc	atatctaaca	actgagcacg	taaccaatgg	aaaagcatga	gcttagcgtt	4620
gctccaaaaa	agtattggat	ggttaatacc	atttgtctgt	tctcttctga	ctttgactcc	4680
tcaaaaaaaa	aaaatctaca	atcaacagat	cgcttcaatt	acgccctcac	aaaaactttt	4740
ttccttcttc	ttcgcccacg	ttaaatttta	tccctcatgt	tgtctaacgg	atttctgcac	4800
ttgatttatt	ataaaaagac	aaagacataa	tacttctcta	tcaatttcag	ttattgttct	4860
tccttgcgtt	attettetgt	tettetttt	cttttgtcat	atataaccat	aaccaagtaa	4920
tacatattca	agtttaaaca	tgtataccgt	aggacagtac	ttggtagata	gactagaaga	4980
gattggtatc	gataaggttt	tcggtgtgcc	aggggattac	aatttgactt	ttctagatta	5040
cattcaaaat	cacgaaggac	tttcctggca	agggaatact	aatgaactaa	acgcagcata	5100
tgcagcagat	ggctacgccc	gtgaaagagg	cgtatcagct	cttgttacta	cattcggagt	5160
gggtgaactg	tcagccatta	acggaacagc	tggtagttt	gcagaacaag	tccctgtcat	5220
ccacatcgtg	ggttctccaa	ctatgaatgt	gcaatccaac	aaaaagctgg	ttcatcattc	5280
cttaggaatg	ggtaactttc	ataactttag	tgaaatggct	aaggaagtca	ctgccgctac	5340
aaccatgctt	actgaagaga	atgcagette	agagatcgac	agagtattag	aaacagcctt	5400
gttggaaaag	aggccagtat	acatcaatct	tccaattgat	atagctcata	aagcaatagt	5460
taaacctgca	aaagcactac	aaacagagaa	atcatctggt	gagagagagg	cacaacttgc	5520
agaaatcata	ctatcacact	tagaaaaggc	cgctcaacct	atcgtaatcg	ccggtcatga	5580
gatcgcccgt	ttccagataa	gagaaagatt	tgaaaactgg	ataaaccaaa	caaagttgcc	5640
agtaaccaat	ttggcatatg	gcaaaggctc	tttcaatgaa	gagaacgaac	atttcattgg	5700
tacctattac	ccagcttttt	ctgacaaaaa	cgttctggat	tacgttgaca	atagtgactt	5760
cgttttacat	tttggtggga	aaatcattga	caattctacc	tcctcatttt	ctcaaggctt	5820
taagactgaa	aacactttaa	ccgctgcaaa	tgacatcatt	atgetgeeaq	atgggtctac	5880
ttactctqqq	atttctctta	acggtctttt	ggcagaqctq	gaaaaactaa	actttacttt	5940
tqctqatact	qctqctaaac	aaqctqaatt	aqctqttttc	qaaccacaaa	ccqaaacacc	6000
5 5		5 5	5 5			

437

## 438

### -continued

actaaagcaa	gacagatttc	accaagctgt	tatgaacttt	ttgcaagctg	atgatgtgtt	6060
ggtcactgag	caggggacat	catctttcgg	tttgatgttg	gcacctctga	aaaagggtat	6120
gaatttgatc	agtcaaacat	tatgggggtc	cataggatac	acattacctg	ctatgattgg	6180
ttcacaaatt	gctgccccag	aaaggagaca	cattctatcc	atcggtgatg	gatcttttca	6240
actgacagca	caggaaatgt	ccaccatctt	cagagagaaa	ttgacaccag	tgatattcat	6300
tatcaataac	gatggctata	cagtcgaaag	agccatccat	ggagaggatg	agagttacaa	6360
tgatatacca	acttggaact	tgcaattagt	tgctgaaaca	tttggtggtg	atgccgaaac	6420
tgtcgacact	cacaacgttt	tcacagaaac	agacttcgct	aatactttag	ctgctatcga	6480
tgctactcct	caaaaagcac	atgtcgttga	agttcatatg	gaacaaatgg	atatgccaga	6540
atcattgaga	cagattggct	tagccttatc	taagcaaaac	tcttaagttt	aaactaagcg	6600
aatttcttat	gatttatgat	ttttattatt	aaataagtta	taaaaaaaat	aagtgtatac	6660
aaattttaaa	gtgactctta	ggttttaaaa	cgaaaattct	tattcttgag	taactctttc	6720
ctgtaggtca	ggttgctttc	tcaggtatag	catgaggtcg	ctcttattga	ccacacctct	6780
accggcatgc	cgagcaaatg	cctgcaaatc	gctccccatt	tcacccaatt	gtagatatgc	6840
taactccagc	aatgagttga	tgaatctcgg	tgtgtatttt	atgtcctcag	aggacaacac	6900
ctgttgtaat	cgttcttcca	cacg				6924
<211> LENG' <212> TYPE <213> ORGAI <220> FEAT <223> OTHEI	TH: 6761 : DNA NISM: Artif: JRE: R INFORMATIC	icial Sequer DN: pLA65	nce			
<400> SEQUI	ENCE: 211					
gatccgcatt	gcggattacg	tattctaatg	ttcagtaccg	ttcgtataat	gtatgctata	60
cgaagttatg	cagattgtac	tgagagtgca	ccataccacc	ttttcaattc	atcattttt	120
ttttattctt	ttttttgatt	tcggtttcct	tgaaattttt	ttgattcggt	aatctccgaa	180
cagaaggaag	aacgaaggaa	ggagcacaga	cttagattgg	tatatatacg	catatgtagt	240
gttgaagaaa	catgaaattg	cccagtattc	ttaacccaac	tgcacagaac	aaaaacctgc	300
aggaaacgaa	gataaatcat	gtcgaaagct	acatataagg	aacgtgctgc	tactcatcct	360
agtcctgttg	ctgccaagct	atttaatatc	atgcacgaaa	agcaaacaaa	cttgtgtgct	420
tcattggatg	ttcgtaccac	caaggaatta	ctggagttag	ttgaagcatt	aggtcccaaa	480
atttgtttac	taaaaacaca	tgtggatatc	ttgactgatt	tttccatgga	gggcacagtt	540
aagccgctaa	aggcattatc	cgccaagtac	aattttttac	tcttcgaaga	cagaaaattt	600
gctgacattg	gtaatacagt	caaattgcag	tactctgcgg	gtgtatacag	aatagcagaa	660
tgggcagaca	ttacgaatgc	acacggtgtg	gtgggcccag	gtattgttag	cggtttgaag	720
caggcggcag	aagaagtaac	aaaggaacct	agaggccttt	tgatgttagc	agaattgtca	780
tgcaagggct	ccctatctac	tggagaatat	actaagggta	ctgttgacat	tgcgaagagc	840
gacaaagatt	ttgttatcgg	ctttattgct	caaagagaca	tgggtggaag	agatgaaggt	900

caacagtata gaaccgtgga tgatgtggtc tctacaggat ctgacattat tattgttgga 1020 agaggactat ttgcaaaggg aagggatgct aaggtagagg gtgaacgtta cagaaaagca 1080

960

tacgattggt tgattatgac acccggtgtg ggtttagatg acaagggaga cgcattgggt

439

**440** 

ggctgggaag	catatttgag	aagatgcggc	cagcaaaact	aaaaactgt	attataagta	1140
aatgcatgta	tactaaactc	acaaattaga	gcttcaattt	aattatatca	gttattaccc	1200
tatgcggtgt	gaaataccgc	acagatgcgt	aaggagaaaa	taccgcatca	ggaaattgta	1260
aacgttaata	ttttgttaaa	attcgcgtta	aatttttgtt	aaatcagctc	atttttaac	1320
caataggccg	aaatcggcaa	aatcccttat	aaatcaaaag	aatagaccga	gatagggttg	1380
agtgttgttc	cagtttggaa	caagagtcca	ctattaaaga	acgtggactc	caacgtcaaa	1440
gggcgaaaaa	ccgtctatca	gggcgatggc	ccactacgtg	aaccatcacc	ctaatcaaga	1500
taacttcgta	taatgtatgc	tatacgaacg	gtaccagtga	tgatacaacg	agttagccaa	1560
ggtgaattca	ctggccgtcg	ttttacaacg	tcgtgactgg	gaaaaccctg	gcgttaccca	1620
acttaatcgc	cttgcagcac	atcccccttt	cgccagctgg	cgtaatagcg	aagaggcccg	1680
caccgatcgc	ccttcccaac	agttgcgcag	cctgaatggc	gaatggcgcc	tgatgcggta	1740
ttttctcctt	acgcatctgt	gcggtatttc	acaccgcata	tggtgcactc	tcagtacaat	1800
ctgctctgat	gccgcatagt	taagccagcc	ccgacacccg	ccaacacccg	ctgacgcgcc	1860
ctgacgggct	tgtctgctcc	cggcatccgc	ttacagacaa	gctgtgaccg	tctccgggag	1920
ctgcatgtgt	cagaggtttt	caccgtcatc	accgaaacgc	gcgagacgaa	agggcctcgt	1980
gatacgccta	tttttatagg	ttaatgtcat	gataataatg	gtttcttaga	cgtcaggtgg	2040
cacttttcgg	ggaaatgtgc	gcggaacccc	tatttgttta	tttttctaaa	tacattcaaa	2100
tatgtatccg	ctcatgagac	aataaccctg	ataaatgctt	caataatatt	gaaaaaggaa	2160
gagtatgagt	attcaacatt	tccgtgtcgc	ccttattccc	tttttgcgg	cattttgcct	2220
tcctgtttt	gctcacccag	aaacgctggt	gaaagtaaaa	gatgctgaag	atcagttggg	2280
tgcacgagtg	ggttacatcg	aactggatct	caacagcggt	aagatccttg	agagttttcg	2340
ccccgaagaa	cgttttccaa	tgatgagcac	ttttaaagtt	ctgctatgtg	gcgcggtatt	2400
atcccgtatt	gacgccgggc	aagagcaact	cggtcgccgc	atacactatt	ctcagaatga	2460
cttggttgag	tactcaccag	tcacagaaaa	gcatcttacg	gatggcatga	cagtaagaga	2520
attatgcagt	gctgccataa	ccatgagtga	taacactgcg	gccaacttac	ttctgacaac	2580
gatcggagga	ccgaaggagc	taaccgcttt	tttgcacaac	atgggggatc	atgtaactcg	2640
ccttgatcgt	tgggaaccgg	agctgaatga	agccatacca	aacgacgagc	gtgacaccac	2700
gatgcctgta	gcaatggcaa	caacgttgcg	caaactatta	actggcgaac	tacttactct	2760
agetteeegg	caacaattaa	tagactggat	ggaggcggat	aaagttgcag	gaccacttct	2820
gcgctcggcc	cttccggctg	gctggtttat	tgctgataaa	tctggagccg	gtgagcgtgg	2880
gtctcgcggt	atcattgcag	cactgggggcc	agatggtaag	ccctcccgta	tcgtagttat	2940
ctacacgacg	gggagtcagg	caactatgga	tgaacgaaat	agacagatcg	ctgagatagg	3000
tgcctcactg	attaagcatt	ggtaactgtc	agaccaagtt	tactcatata	tactttagat	3060
tgatttaaaa	cttcattttt	aatttaaaag	gatctaggtg	aagatccttt	ttgataatct	3120
catgaccaaa	atcccttaac	gtgagttttc	gttccactga	gcgtcagacc	ccgtagaaaa	3180
gatcaaagga	tcttcttgag	atcctttttt	tctgcgcgta	atctgctgct	tgcaaacaaa	3240
aaaaccaccg	ctaccagcgg	tggtttgttt	gccggatcaa	gagctaccaa	ctctttttcc	3300
gaaqqtaact	ggetteagea	gagcqcaqat	accaaatact	gteettetag	tgtaqccqta	3360
gttaggedag	cactteaaga	actotatado	accocctaca	tacctedete	tgctaateet	3420
attaccacto	actactacco	ataacaetee	atcatatat	accondition	actraactor	3480
gulaceaglg	geegeegeea	guggugalaa	guugugudut	accyyyttyg	ucccaayacy	5100

441

442

atagttaccg	gataaggcgc	agcggtcggg	ctgaacgggg	ggttcgtgca	cacagcccag	3540
cttggagcga	acgacctaca	ccgaactgag	atacctacag	cgtgagctat	gagaaagcgc	3600
cacgetteee	gaagggagaa	aggcggacag	gtatccggta	agcggcaggg	tcggaacagg	3660
agagcgcacg	agggagcttc	caggggggaaa	cgcctggtat	ctttatagtc	ctgtcgggtt	3720
tcgccacctc	tgacttgagc	gtcgattttt	gtgatgctcg	tcagggggggc	ggagcctatg	3780
gaaaaacgcc	agcaacgcgg	cctttttacg	gttcctggcc	ttttgctggc	cttttgctca	3840
catgttcttt	cctgcgttat	cccctgattc	tgtggataac	cgtattaccg	cctttgagtg	3900
agctgatacc	gctcgccgca	gccgaacgac	cgagcgcagc	gagtcagtga	gcgaggaagc	3960
ggaagagcgc	ccaatacgca	aaccgcctct	ccccgcgcgt	tggccgattc	attaatgcag	4020
ctggcacgac	aggtttcccg	actggaaagc	gggcagtgag	cgcaacgcaa	ttaatgtgag	4080
ttagctcact	cattaggcac	cccaggcttt	acactttatg	cttccggctc	gtatgttgtg	4140
tggaattgtg	agcggataac	aatttcacac	aggaaacagc	tatgaccatg	attacgccaa	4200
gcttacctgg	taaaacctct	agtggagtag	tagatgtaat	caatgaagcg	gaagccaaaa	4260
gaccagagta	gaggcctata	gaagaaactg	cgataccttt	tgtgatggct	aaacaaacag	4320
acatctttt	atatgttttt	acttctgtat	atcgtgaagt	agtaagtgat	aagcgaattt	4380
ggctaagaac	gttgtaagtg	aacaagggac	ctcttttgcc	tttcaaaaaa	ggattaaatg	4440
gagttaatca	ttgagattta	gttttcgtta	gattctgtat	ccctaaataa	ctcccttacc	4500
cgacgggaag	gcacaaaaga	cttgaataat	agcaaacggc	cagtagccaa	gaccaaataa	4560
tactagagtt	aactgatggt	cttaaacagg	cattacgtgg	tgaactccaa	gaccaatata	4620
caaaatatcg	ataagttatt	cttgcccacc	aatttaagga	gcctacatca	ggacagtagt	4680
accattcctc	agagaagagg	tatacataac	aagaaaatcg	cgtgaacacc	ttatataact	4740
tagcccgtta	ttgagctaaa	aaaccttgca	aaatttccta	tgaataagaa	tacttcagac	4800
gtgataaaaa	tttactttct	aactcttctc	acgctgcccc	tatctgttct	tccgctctac	4860
cgtgagaaat	aaagcatcga	gtacggcagt	tcgctgtcac	tgaactaaaa	caataaggct	4920
agttcgaatg	atgaacttgc	ttgctgtcaa	acttctgagt	tgccgctgat	gtgacactgt	4980
gacaataaat	tcaaaccggt	tatagcggtc	tcctccggta	ccggttctgc	cacctccaat	5040
agagctcagt	aggagtcaga	acctctgcgg	tggctgtcag	tgactcatcc	gcgtttcgta	5100
agttgtgcgc	gtgcacattt	cgcccgttcc	cgctcatctt	gcagcaggcg	gaaattttca	5160
tcacgctgta	ggacgcaaaa	aaaaaataat	taatcgtaca	agaatcttgg	aaaaaaatt	5220
gaaaaatttt	gtataaaagg	gatgacctaa	cttgactcaa	tggcttttac	acccagtatt	5280
ttccctttcc	ttgtttgtta	caattataga	agcaagacaa	aaacatatag	acaacctatt	5340
cctaggagtt	atatttttt	accctaccag	caatataagt	aaaaaactgt	ttatgaaagc	5400
attagtgtat	aggggcccag	gccagaagtt	ggtggaagag	agacagaagc	cagagcttaa	5460
ggaacctggt	gacgctatag	tgaaggtaac	aaagactaca	atttgcggaa	ccgatctaca	5520
cattcttaaa	ggtgacgttg	cgacttgtaa	acccggtcgt	gtattagggc	atgaaggagt	5580
gggggttatt	gaatcagtcg	gatctggggt	tactgctttc	caaccaggcg	atagagtttt	5640
gatatcatgt	atatcgagtt	gcggaaagtg	ctcattttgt	agaagaggaa	tgttcagtca	5700
ctgtacgacc	gggggttgga	ttctgggcaa	cgaaattgat	ggtacccaag	cagagtacgt	5760
aagagtacca	catgctgaca	catcccttta	tcgtattccg	gcaggtgcgg	atgaagaggc	5820

cttagtcatg ttatcagata ttctaccaac gggttttgag tgcggagtcc taaacggcaa	5880
agtcgcacct ggttcttcgg tggctatagt aggtgctggt cccgttggtt tggccgcctt	5940
actgacagca caattctact ccccagctga aatcataatg atcgatcttg atgataacag	6000
gctgggatta gccaaacaat ttggtgccac cagaacagta aactccacgg gtggtaacgc	6060
cgcagccgaa gtgaaagctc ttactgaagg cttaggtgtt gatactgcga ttgaagcagt	6120
tgggatacct gctacatttg aattgtgtca gaatatcgta gctcccggtg gaactatcgc	6180
taatgtegge gtteaeggta geaaagttga tttgeatett gaaagtttat ggteeeataa	6240
tgtcacgatt actacaaggt tggttgacac ggctaccacc ccgatgttac tgaaaactgt	6300
tcaaagtcac aagctagatc catctagatt gataacacat agattcagcc tggaccagat	6360
cttggacgca tatgaaactt ttggccaagc tgcgtctact caagcactaa aagtcatcat	6420
ttcgatggag gcttgattaa ttaagagtaa gcgaatttct tatgatttat gattttatt	6480
attaaataag ttataaaaaa aataagtgta tacaaatttt aaagtgactc ttaggtttta	6540
aaacgaaaat tettattett gagtaactet tteetgtagg teaggttget tteteaggta	6600
tagcatgagg tegetettat tgaccacace tetaceggea tgeegageaa atgeetgeaa	6660
atcgctcccc atttcaccca attgtagata tgctaactcc agcaatgagt tgatgaatct	6720
cggtgtgtat tttatgtcct cagaggacaa cacctgtggt g	6761
<pre><li><lo> SEQ ID NO 212 <li>&lt;211&gt; LENGTH: 90 <li>&lt;212&gt; TYPE: DNA <li>&lt;213&gt; ORGANISM: Artificial Sequence <li>&lt;220&gt; FEATURE: <li>&lt;223&gt; OTHER INFORMATION: 895 <li>&lt;400&gt; SEQUENCE: 212</li></li></li></li></li></li></lo></li></pre>	
	40
atottorea approactae actual acadaguda atoaatuaa	90 90
alyttyavaa aayoaavaaa ayaavaaaaa	90
<210> SEQ ID NO 213 <211> LENGTH: 81 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: 679	
gtggagcatc gaagactggc aacatgattt caatcattct gatcttagag caccttggct	60
aactcgttgt atcatcactg g	81
<210> SEQ ID NO 214 <211> LENGTH: 20 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: 681	
<400> SEQUENCE: 214	
ttattgctta gcgttggtag	20
<210> SEQ ID NO 215 <211> LENGTH: 22 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: 92	

-continued	
<400> SEQUENCE: 215	
gagaagatgc ggccagcaaa ac	22
<210> SEQ ID NO 216 <211> LENGTH: 25	
<212> TYPE: DNA	
<213> ORGANISM: ATTIFICIAL Sequence <220> FEATURE: <223> OTHER INFORMATION: N245	
<400> SEQUENCE: 216	
agggtagcct ccccataaca taaac	25
<210> SEQ ID NO 217 <211> LENGTH: 25 <212> TYPE: DNA	
<213> ORGANISM: Artificial Sequence <220> FEATURE:	
<223> OTHER INFORMATION: N246	
<400> SEQUENCE: 217	
tetecaaata tatacetett gtgtg	25
<210> SEQ ID NO 218 <211> LENGTH: 90 <212> TYPE: DNA	
<213> ORGANISM: Artificial Sequence	
<220> FEATURE: <223> OTHER INFORMATION: 896	
<400> SEQUENCE: 218	
ttttatatac agtataaata aaaaacccac gtaatatagc aaaaacatat tgccaacaaa	60
aattaccgtc gctcgtgatt tgtttgcaaa	90
<210> SEQ ID NO 219	
<211> LENGTH: 90 <212> TYPE: DNA	
<213> ORGANISM: Artificial Sequence	
<220> FEATURE: <223> OTHER INFORMATION: 897	
<400> SEQUENCE: 219	
caaactgtgt aagtttattt atttgcaaca ataattcgtt tgagtacact actaatggcc	60
accttggcta actogttgta toatoactgg	90
<210> SEQ ID NO 220 <211> LENGTH: 28 <212> TYPE: DNA	
<213> ORGANISM: Artificial Sequence	
<220> FEATURE: <223> OTHER INFORMATION: 365	
<400> SEQUENCE: 220	
ctctatctcc gctcaggcta agcaattg	28
<210> SEQ ID NO 221	
<211> LENGTH: 26 <212> TYPE: DNA	
<213> ORGANISM: Artificial Sequence	
<220> FEATURE:	
<223> OTHER INFORMATION: 300	

US 9,771,602 B2

4	1	7
_	_	

-continued

cagccgactc aacggcctgt ttcacg	26
<pre>&lt;210&gt; SEQ ID NO 222 &lt;211&gt; LENGTH: 28 &lt;212&gt; TYPE: DNA &lt;213&gt; ORGANISM: Artificial Sequence &lt;220&gt; FEATURE: &lt;223&gt; OTHER INFORMATION: N638</pre>	
<400> SEQUENCE: 222	
aaaagatagt gtagtagtga taaactgg	28
<210> SEQ ID NO 223 <211> LENGTH: 22 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: 740	
<400> SEQUENCE: 223	
cgataatcct gctgtcatta tc	22
<210> SEQ ID NO 224 <211> LENGTH: 83 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: 856	
<400> SEQUENCE: 224	
gcttatttag aagtgtcaac aacgtatcta ccaacgattt gacccttttc cacaccttgg	60
ctaactcgtt gtatcatcac tgg	83
<210> SEQ ID NO 225 <211> LENGTH: 80 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: 857	
<400> SEQUENCE: 225	
gcacaatatt tcaagctata ccaagcatac aatcaactat ctcatataca atgaaagcat	60
tagtgtatag gggcccaggc	80
<210> SEQ ID NO 226 <211> LENGTH: 25 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: BK415	
<400> SEQUENCE: 226	
gcctcattga tggtggtaca taacg	25
<210> SEQ ID NO 227 <211> LENGTH: 26 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: N1092	
<400> SEQUENCE: 22/	26
ayayılılıya lalcalylal alcyay	20

-cont	inued	

<210> SEQ ID NO 228	
<211> LENGTH: 92	
<212> TYPE: DNA	
<213> ORGANISM: Artificial Sequence	
<220> FEATURE:	
<223> OTHER INFORMATION: 906	
<400> SEQUENCE: 228	
atgacaggtg aaagaattga aaaggtgaaa ataaatgacg aatttgcaaa atcacatttc	60
	0.2
acciggiaaa acciciagig gagiagiaga ig	92
<210> SEO ID NO 229	
<211> LENGTH: 87	
<212> TYPE: DNA	
<213> ORGANISM: Artificial Sequence	
<220> FEATURE:	
<223> OTHER INFORMATION: 907	
<400> SEQUENCE: 229	
aaaaagattc aatgccgtct cctttcgaaa cttaataata gaacaatatc atccttcacc	60
	0.7
liggelaael egilgialea leaelgg	87
<210> SEO ID NO 230	
<211> LENGTH: 70	
<212> TYPE: DNA	
213> ORGANISM: Artificial Sequence	
220> FEATURE	
223> OTHER INFORMATION: 667	
<400> SEQUENCE: 230	
teteettteg aaaettaata atagaacaat ateateettt tgtaaaaega eggeeagtga	60
attcaccttg	70

What is claimed:

**1**. A method for production of isobutanol in a fermentation process comprising:

- providing a fermentation mix comprising a recombinant yeast production microorganism which comprises an engineered isobutanol biosynthetic pathway, a heterologous polynucleotide encoding a polypeptide having acetolactate synthase activity, wherein said polypeptide has the amino acid sequence of SEQ ID NO: 9 and confers resistant to sulfonylureas, and a heterologous polynucleotide encoding a polypeptide having 3-phosphoshikimate 1-carboxylvinyltransferase activity; and
- contacting the fermentation mix with at least one sulfonylurea which preferentially inhibits at least one contaminant yeast microorganism, wherein a fermentation product of the contaminant yeast microorganism is ethanol;
- wherein production competitiveness of the recombinant 55 yeast production microorganism is associated with a higher isobutanol-to-ethanol ratio as compared to a fermentation process without addition of one or more inhibitors, antibiotics, or combinations thereof.

**2**. The method of claim **1**, wherein the specific growth rate <sup>60</sup> of the at least one contaminant yeast microorganism is reduced more than the specific growth rate of the recombinant yeast production microorganism.

**3**. The method of claim **1**, wherein production of the fermentation product of the at least one contaminant yeast 65 microorganism is reduced more than the isobutanol production of the recombinant yeast production microorganism.

4. The method of claim 1, wherein the contaminant yeast microorganism is *Saccharomyces cerevisiae*.

**5**. The method of claim **1**, wherein the sulfonylurea is an inhibitor of an ethanol biosynthesis pathway.

- **6**. The method of claim **1**, wherein the sulfonylurea is an inhibitor of an amino acid biosynthesis pathway.
- 7. The method of claim 1, wherein the sulfonylurea is selected from a group consisting of: nicosulfuron methyl, metsulfuron methyl, chlorimuron ethyl, sulfometuron methyl, chlorsulfuron, thifensulfuron methyl, and mixtures thereof.

8. The method of claim 1, wherein the recombinant yeast production microorganism is selected from *Schizosaccharomyces*, *Issatchenkia*, *Kluyveromyces*, *Yarrowia*, *Pichia*, *Candida*, *Hansenula*, *Aspergillus*, *Pachysolen*, *Rhodotorula*, *Zygosaccharomyces*, *Galactomyces*, *Torulaspora*, *Debayo-myces*, *Williopsis*, *Dekkera*, *Kloeckera*, *Metschnikowia*, and *Saccharomyces*.

**9**. The method of claim **1**, wherein the isobutanol biosynthetic pathway comprises the following substrate to product conversions:

- a) pyruvate to acetolactate;
- b) acetolactate to 2,3-dihydroxyisovalerate;
- c) 2,3-dihydroxyisovalerate to  $\alpha$ -ketoisovalerate;
- d)  $\alpha$ -ketoisovalerate to isobutyraldehyde; and
- e) isobutyraldehyde to isobutanol.

10. The method of claim 1, wherein the recombinant yeast production microorganism further comprises one or more of the following modifications:

a deletion in one or more endogenous polynucleotides encoding a polypeptide having pyruvate decarboxylase activity;

- a deletion, mutation, or substitution in an endogenous polynucleotide encoding a polypeptide having acetolactate reductase activity;
- a deletion, mutation, or substitution in an endogenous polynucleotide encoding a polypeptide having aldehyde dehydrogenase activity;
- a deletion in an endogenous polynucleotide encoding a polypeptide having hexokinase activity;
- a deletion in an endogenous polynucleotide encoding a polypeptide having glycerol-3-phosphate dehydroge- 10 nase activity; or
- a deletion in an endogenous gene encoding a polypeptide affecting Fe-S cluster biosynthesis, wherein the polypeptide is FRA2.

\* \* \* \* \*