

(12) STANDARD PATENT
(19) AUSTRALIAN PATENT OFFICE

(11) Application No. **AU 2011299303 B2**

(54) Title
Integration of a polynucleotide encoding a polypeptide that catalyzes pyruvate to acetolactate conversion

(51) International Patent Classification(s)
C12N 1/21 (2006.01) **C12N 15/63** (2006.01)
C12N 15/52 (2006.01) **C12P 7/16** (2006.01)

(21) Application No: **2011299303** (22) Date of Filing: **2011.09.07**

(87) WIPO No: **WO12/033832**

(30) Priority Data

(31) Number	(32) Date	(33) Country
61/380,563	2010.09.07	US
61/466,557	2011.03.23	US

(43) Publication Date: **2012.03.15**

(44) Accepted Journal Date: **2016.07.07**

(71) Applicant(s)
BUTAMAX(TM) ADVANCED BIOFUELS LLC

(72) Inventor(s)
Anthony, Larry Cameron;Maggio-Hall, Lori Ann;Paul, Brian James

(74) Agent / Attorney
Houlihan², Level 1 70 Doncaster Road, BALWYN NORTH, VIC, 3104

(56) Related Art
US 2009/0305363 A1
US 2007/0092957 A1



(51) International Patent Classification:

C12N 1/21 (2006.01) C12N 15/63 (2006.01)

C12N 15/52 (2006.01) C12P 7/16 (2006.01)

(21) International Application Number:

PCT/US2011/050689

(22) International Filing Date:

7 September 2011 (07.09.2011)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

61/380,563 7 September 2010 (07.09.2010) US

61/466,557 23 March 2011 (23.03.2011) US

(71) Applicant (for all designated States except US):

BUTAMAX(TM) ADVANCED BIOFUELS LLC
[US/US]; 200 Powder Mill Road, Wilmington, DE 19803
(US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): ANTHONY, Larry,
Cameron [US/US]; 501 Cory Lane, Aston, PA 19014
(US). MAGGIO-HALL, Lori, Ann [US/US]; 3413 North
Rockfield Drive, Wilmington, DE 19810 (US). PAUL,
Brian, James [US/US]; 115 Fairfax Boulevard, Wilming-
ton, DE 19803 (US).(74) Agent: LHULIER, Christine, M.; E. I. du Pont de
Nemours and Company, Legal Patent Records Center,
4417 Lancaster Pike, Wilmington, DE 19805 (US).

(81) Designated States (unless otherwise indicated, for every

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

(84) Designated States (unless otherwise indicated, for every

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

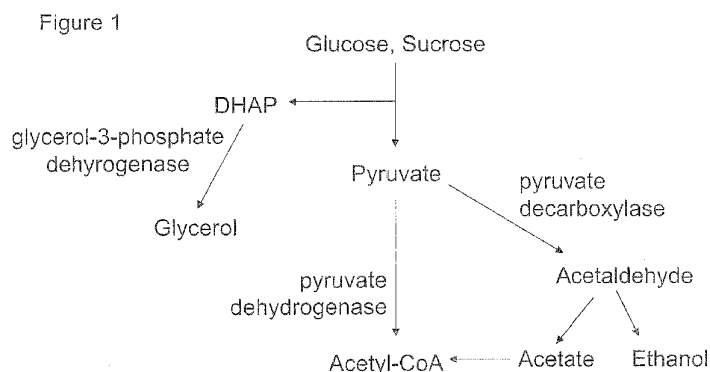
Published:

- with international search report (Art. 21(3))
- before the expiration of the time limit for amending the
claims and to be republished in the event of receipt of
amendments (Rule 48.2(h))
- with sequence listing part of description (Rule 5.2(a))

(88) Date of publication of the international search report:

28 June 2012

(54) Title: INTEGRATION OF A POLYNUCLEOTIDE ENCODING A POLYPEPTIDE THAT CATALYZES PYRUVATE TO ACETOLACTATE CONVERSION



(57) Abstract: The invention relates to recombinant host cells having at least one integrated polynucleotide encoding a polypeptide that catalyzes a step in a pyruvate-utilizing biosynthetic pathway, e.g., pyruvate to acetolactate conversion. The invention also relates to methods of increasing the biosynthetic production of isobutanol, 2,3-butanediol, 2-butanol or 2-butanone using such host cells.

INTEGRATION OF A POLYNUCLEOTIDE ENCODING A POLYPEPTIDE THAT CATALYZES PYRUVATE TO ACETOLACTATE CONVERSION

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application is related to and claims the benefit of priority of U.S. Provisional Patent Application No. 61/380,563, filed September 7, 2010 and U.S. Provisional Patent Application No. 61/466,557, filed March 23, 2011, both of which are herein incorporated by reference in their entireties.

REFERENCE TO SEQUENCE LISTING SUBMITTED ELECTRONICALLY

[0002] The content of the electronically submitted sequence listing in ASCII text file (Name: 20110907_CL5178USNA_SeqList.txt, Size: 669,953 bytes, and Date of Creation: August 31, 2011) filed with the application is incorporated herein by reference in its entirety.

FIELD OF THE INVENTION

[0003] The invention relates to the field of industrial microbiology and the fermentative production of butanol and isomers thereof. More specifically, the invention relates to recombinant host cells having one or more integrated polynucleotide encoding a polypeptide that catalyzes a step in a pyruvate-utilizing biosynthetic pathway, *e.g.*, pyruvate to acetolactate conversion.

BACKGROUND OF THE INVENTION

[0004] Butanol is an important industrial chemical, useful as a fuel additive, as a feedstock chemical in the plastics industry, and as a foodgrade extractant in the food and flavor industry. Each year 10 to 12 billion pounds of butanol are produced by petrochemical means and the need for this commodity chemical will likely increase. 2-butanone, also referred to as methyl ethyl ketone (MEK), is a widely used solvent and is the most important commercially produced ketone, after acetone. It is used as a solvent

for paints, resins, and adhesives, as well as a selective extractant, activator of oxidative reactions, and can be chemically converted to 2-butanol by reacting with hydrogen in the presence of a catalyst (Nystrom *et al.*, *J. Am. Chem. Soc.*, 69:1198, 1947). 2,3-butanediol can be used in the chemical synthesis of butene and butadiene, important industrial chemicals currently obtained from cracked petroleum, and esters of 2,3-butanediol can be used as plasticizers (Voloch *et al.*, "Fermentation Derived 2,3-Butanediol," in: *Comprehensive Biotechnology*, Pergamon Press Ltd., England, Vol. 2, Section 3, pp. 933-947, 1986).

[0005] Microorganisms can be engineered for expression of biosynthetic pathways for the production of products such as 2,3-butanediol, 2-butanone, 2-butanol and isobutanol. U.S. Patent No. 7,851,188 discloses the engineering of recombinant microorganisms for production of isobutanol. U.S. Appl. Pub. Nos. 20070259410 and 20070292927 disclose the engineering of recombinant microorganisms for the production of 2-butanone or 2-butanol. Multiple pathways are known for the biosynthesis of isobutanol and 2-butanol, all of which initiate with cellular pyruvate. Butanediol is an intermediate in the 2-butanol pathway disclosed in U.S. Appl. Pub. No. 20070292927.

[0006] Pyruvate metabolism has been altered in yeast for the production of lactic acid and glycerol. U.S. Appl. Pub. No. 20070031950 discloses a yeast strain with a disruption of one or more pyruvate decarboxylase or pyruvate dehydrogenase genes and expression of a D-lactate dehydrogenase gene, which is used for the production of D-lactic acid. Ishida *et al.* (*Biosci. Biotech. and Biochem.*, 70:1148-1153, 2006) describe *Saccharomyces cerevisiae* with disrupted pyruvate decarboxylase genes and expression of lactate dehydrogenase. U.S. Appl. Pub. No. 2005/0059136 discloses glucose tolerant C2 carbon source-independent (GCSI) yeast strains with no pyruvate decarboxylase activity, which can have an exogenous lactate dehydrogenase gene. Nevoigt *et al.* (*Yeast*, 12:1331-1337, 1996) describe the impact of reduced pyruvate decarboxylase and increased NAD-dependent glycerol-3-phosphate dehydrogenase in *Saccharomyces cerevisiae* on glycerol yield.

[0007] Stable production of polynucleotides by a yeast cell for pyruvate biosynthetic pathways are needed for industrial fermentative production of alcohols or other compounds. Further, there is a need for improved means of isobutanol, 2,3-butanediol, 2-butanol or 2-butanone production in recombinant host cells such as yeast.

BRIEF SUMMARY OF THE INVENTION

- [0008] Provided herein are recombinant host cells having one or more integrated polynucleotides encoding a polypeptide that catalyzes a step in a pyruvate-utilizing biosynthetic pathway, *e.g.*, pyruvate to acetolactate conversion. Such host cells provide a means to stabilize and/or increase product formation of a biosynthetic pathway, such as isobutanol, 2,3-butanediol, 2-butanol or 2-butanone, compared to host cells which do not have an integrated polynucleotide encoding a polypeptide that catalyzes biosynthetic pathway steps such as pyruvate to acetolactate conversion.
- [0009] One aspect of the invention relates to a recombinant host cell comprising a polynucleotide encoding a polypeptide which catalyzes the conversion of pyruvate to acetolactate integrated into the chromosome of the host cell. In another aspect, the host cell comprises a pyruvate-utilizing biosynthetic pathway and a polynucleotide encoding a polypeptide which catalyzes the conversion of pyruvate to acetolactate integrated into the chromosome of the host cell. In another aspect, the polynucleotide is heterologous to the host cell.
- [0010] An aspect of the invention relates to a recombinant host cell comprising an isobutanol biosynthetic pathway wherein said pathway comprises the substrate to product conversion pyruvate to acetolactate catalyzed by a polypeptide encoded by a heterologous polynucleotide integrated into the chromosome and wherein said pathway comprises at least one substrate to product conversion catalyzed by a polypeptide encoded by a polynucleotide on a plasmid; and wherein expression of Fra2 in the host cell is decreased or substantially eliminated. In embodiments, the titer of isobutanol production is increased as compared to a recombinant host cell wherein the polynucleotide encoding a polypeptide that catalyzes the conversion of pyruvate to acetolactate is not integrated into the chromosome.
- [0011] An aspect of the invention relates to a recombinant host cell comprising a 2,3-butanediol, 2-butanol, or 2-butanone biosynthetic pathway wherein said pathway comprises the substrate to product conversion pyruvate to acetolactate catalyzed by a polypeptide encoded by a heterologous polynucleotide integrated into the chromosome and wherein said pathway comprises at least one substrate to product conversion catalyzed by a polypeptide encoded by a polynucleotide on a plasmid. In embodiments,

- 3a -

the titer of 2,3-butanediol, 2-butanol, or 2-butanone production is increased as compared to a recombinant host cell wherein the polynucleotide encoding a polypeptide that

catalyzes the conversion of pyruvate to acetolactate is not integrated into the chromosome.

[0012] In another aspect, the invention relates to a recombinant host cell comprising a first heterologous polynucleotide encoding a first polypeptide which catalyzes the conversion of a step of a pyruvate-utilizing biosynthetic pathway; a second heterologous polynucleotide encoding a second polypeptide which catalyzes the conversion of a step of a pyruvate-utilizing biosynthetic pathway; and a third heterologous polynucleotide encoding a third polypeptide which catalyzes the conversion of a step of a pyruvate-utilizing biosynthetic pathway; wherein the first and second heterologous polynucleotides are integrated into the chromosome of the host cell; wherein the third heterologous polynucleotide is not integrated into the chromosome of the host cell; and wherein the first, second, and third polypeptides catalyze different steps of the pyruvate-utilizing biosynthetic pathway.

[0013] In another aspect, the invention relates to a recombinant host cell comprising (a) a first heterologous polynucleotide encoding a first polypeptide which catalyzes a substrate to product conversion of pyruvate to acetolactate; (b) a second heterologous polynucleotide encoding a second polypeptide which catalyzes the substrate to product conversion of α -ketoisovalerate to isobutyraldehyde; and (c) a third heterologous polynucleotide encoding a third polypeptide which catalyzes the conversion of a step of a isobutanol biosynthetic pathway that is not the conversion of (a) or (b); wherein the first and second heterologous polynucleotides are integrated into the chromosome; wherein the third heterologous polynucleotide is not integrated into the chromosome; and wherein the host cell produces isobutanol.

[0014] In another aspect, the invention relates to a recombinant host cell comprising (a) a first heterologous polynucleotide encoding a first polypeptide which catalyzes a substrate to product conversion of α -ketoisovalerate to isobutyraldehyde; and (b) a second heterologous polynucleotide encoding a second polypeptide which catalyzes the conversion of a step of a isobutanol biosynthetic pathway that is not the conversion of (a); wherein the first heterologous polynucleotide is integrated into the chromosome; wherein the second heterologous polynucleotide is not integrated into the chromosome; and wherein the host cell produces isobutanol.

- [0015] In aspects of the invention, the host cell is a bacterium, a cyanobacterium, a filamentous fungus, or a yeast. In another aspect, the host cell is a member of the genus *Clostridium*, *Zymomonas*, *Escherichia*, *Salmonella*, *Rhodococcus*, *Pseudomonas*, *Bacillus*, *Lactobacillus*, *Enterococcus*, *Alcaligenes*, *Klebsiella*, *Issatchenkia*, *Paenibacillus*, *Arthrobacter*, *Corynebacterium*, *Brevibacterium*, *Pichia*, *Candida*, *Hansenula*, or *Saccharomyces*. In another aspect, the host cell is *Escherichia coli*, *Alcaligenes eutrophus*, *Bacillus licheniformis*, *Paenibacillus macerans*, *Rhodococcus erythropolis*, *Pseudomonas putida*, *Bacillus subtilis*, *Lactobacillus plantarum*, *Enterococcus faecium*, *Enterococcus gallinarum*, *Enterococcus faecalis* or *Saccharomyces cerevisiae*. In another aspect, the host cell is a facultative anaerobe.
- [0016] In another aspect of the invention, the pyruvate-utilizing biosynthetic pathway comprises one or more polynucleotides encoding polypeptides that catalyze a substrate to product conversion of the pathway. In another aspect, one or more of the polynucleotides are integrated into the chromosome. In another aspect, the pyruvate-utilizing biosynthetic pathway forms the product 2,3-butanediol, isobutanol, 2-butanol or 2-butanone.
- [0017] In one aspect of the invention, the pyruvate-utilizing biosynthetic pathway is a butanol biosynthetic pathway. In another aspect, the butanol biosynthetic pathway is a 2-butanol biosynthetic pathway or an isobutanol biosynthetic pathway. In another aspect, the host cell comprises one or more polynucleotides encoding a polypeptide that catalyzes a substrate to product conversion of (i) pyruvate to acetolactate; (ii) acetolactate to 2,3-dihydroxyisovalerate; (iii) 2,3-dihydroxyisovalerate to α -ketoisovalerate; (iv) α -ketoisovalerate to isobutyraldehyde; or (v) isobutyraldehyde to isobutanol. In another aspect, one or more of the polynucleotides of (ii), (iii), (iv), or (v) are on a plasmid. In another aspect, the host cell comprises one or more polynucleotides encoding a polypeptide that catalyzes a substrate to product conversion of (i) pyruvate to acetolactate; (ii) acetolactate to 2,3-dihydroxyisovalerate; (iii) 2,3-dihydroxyisovalerate to α -ketoisovalerate; (iv) α -ketoisovalerate to isobutyryl-CoA; (v) isobutyryl-CoA to isobutyraldehyde; or (vi) isobutyraldehyde to isobutanol. In another aspect, one or more of the polynucleotides of (ii), (iii), (iv), (v), or (vi) are on a plasmid. In another aspect, the host cell comprises one or more polynucleotides encoding a polypeptide that catalyzes a substrate to product of (i) pyruvate to acetolactate; (ii) acetolactate to 2,3-dihydroxyisovalerate; (iii) 2,3-dihydroxyisovalerate to α -ketoisovalerate; (iv) α -

ketoisovalerate to valine; (v) valine to isobutylamine; (vi) isobutylamine to isobutyraldehyde; or (vii) isobutyraldehyde to isobutanol. In another aspect, one or more of the polynucleotides of (ii), (iii), (iv), (v), (vi), or (vii) are on a plasmid.

[0018] In another aspect of the invention, the host cell comprises one or more polynucleotides encoding a polypeptide that catalyzes a substrate to product conversion of (i) pyruvate to acetolactate; (ii) acetolactate to acetoin; (iii) acetoin to 2,3-butanediol; or (iv) 2,3-butanediol to 2-butanone. In another aspect, one or more of the polynucleotides of (ii), (iii), or (iv) are on a plasmid. In another aspect, the host cell comprises one or more polynucleotides encoding a polypeptide that catalyzes a substrate to product conversion of (i) pyruvate to acetolactate; (ii) acetolactate to acetoin; (iii) acetoin to 2,3-butanediol; (iv) 2,3-butanediol to 2-butanone; or (v) 2-butanone to 2-butanol. In another aspect, one or more of the polynucleotides of (ii), (iii), (iv), or (v) are on a plasmid.

[0019] In another aspect of the invention, the host cell comprises one or more polynucleotides encoding a polypeptide that catalyzes a substrate to product conversion of (i) pyruvate to acetolactate; (ii) alpha-acetolactate to acetoin; (iii) acetoin to 3-amino-2-butanol; (iv) 3-amino-2-butanol to 3-amino-2-butanol phosphate; (v) or 3-amino-2-butanol phosphate to 2-butanone. In another aspect, one or more of the polynucleotides of (ii), (iii), (iv), or (v) are on a plasmid. In another aspect, the host cell comprises one or more polynucleotides encoding a polypeptide that catalyzes a substrate to product conversion of (i) pyruvate to acetolactate; (ii) alpha-acetolactate to acetoin; (iii) acetoin to 3-amino-2-butanol; (iv) 3-amino-2-butanol to 3-amino-2-butanol phosphate; (v) 3-amino-2-butanol phosphate to 2-butanone.; or (vi) 2-butanone to 2-butanol. In another aspect, one or more of the polynucleotides of (ii), (iii), (iv), (v), or (vi) are on a plasmid.

[0020] In another aspect, at least one of the polynucleotides encoding a polypeptide that catalyzes a substrate to product conversion is heterologous. In another embodiment, more than one of the polynucleotides encoding a polypeptide that catalyzes a substrate to product conversion are heterologous. In another embodiment, all of the polynucleotides encoding polypeptides for each of the substrate to product conversions of a pyruvate utilizing biosynthetic pathway are heterologous.

[0021] In one aspect of the invention, the polypeptide which catalyzes a substrate to product conversion of pyruvate to acetolactate is acetolactate synthase. In another aspect,

the acetolactate synthase has at least about 80% identity to an amino acid sequence of an acetolactate synthase described in Table 1. In another aspect, the acetolactate synthase has at least about 80% identity to an amino acid sequence with SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, or 18. In another aspect, the polypeptide which catalyzes the conversion of pyruvate to acetolactate corresponds to the Enzyme Commission Number EC 2.2.1.6.

[0022] In another aspect of the invention, the polypeptide which catalyzes the conversion of acetolactate to 2,3-dihydroxyisovalerate corresponds to the Enzyme Commission Number EC 1.1.1.86. In another aspect, the polypeptide which catalyzes the conversion of 2,3-dihydroxyisovalerate to α -ketoisovalerate corresponds to the Enzyme Commission Number EC 4.2.1.9. In another aspect, the polypeptide which catalyzes the conversion of α -ketoisovalerate to isobutyraldehyde corresponds to the Enzyme Commission Number EC 4.1.1.72 or 4.1.1.1. In another aspect, the polypeptide which catalyzes the conversion of isobutyraldehyde to isobutanol corresponds to the Enzyme Commission Number EC 1.1.1.265, 1.1.1.1 or 1.1.1.2.

[0023] In one aspect, the expression of pyruvate decarboxylase in a host cell of the invention is decreased or substantially eliminated. In another aspect, the host cell comprises a deletion, mutation and/or substitution in an endogenous polynucleotide encoding a polypeptide having pyruvate decarboxylase activity.

[0024] In another aspect of the invention, one or more of the polynucleotides encoding a polypeptide which catalyzes a step of biosynthetic pathway described herein are in a plasmid. In another aspect, the plasmid comprises a sequence at least about 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% identical to one or more of SEQ ID NOs: 1 to 89, or is at least about 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% identical to any one of SEQ ID NOs: 129-133 or a coding region thereof.

[0025] In one aspect, the expression of glycerol-3-phosphate dehydrogenase in a host cell of the invention is decreased or substantially eliminated. In another aspect, the expression of FRA2 in a host cell of the invention is decreased or substantially eliminated. In another aspect, one or more of the polynucleotides described herein is integrated into the chromosome of the host cell at the PDC1-TRX1 intergenic region.

[0026] In one aspect, the invention relates to a method of producing a product of a biosynthetic pathway from a host cell of the invention. In another aspect, the invention relates to a method of producing butanol, comprising (a) providing a recombinant host

cell of the invention; and (b) contacting the host cell with a fermentable carbon substrate to form a fermentation broth under conditions whereby butanol is produced. In another aspect, the method further comprises contacting the fermentation broth with an extractant to produce a two-phase fermentation mixture. In another aspect, the extractant comprises fatty acids. In another aspect, the fatty acids are derived from corn oil or soybean oil. In another aspect, the extractant comprises a water immiscible organic extractant such as C₁₂ to C₂₂ fatty alcohols, C₁₂ to C₂₂ fatty acids, esters of C₁₂ to C₂₂ fatty acids, C₁₂ to C₂₂ fatty amides, or C₁₂ to C₂₂ fatty aldehydes. In another aspect, the method further comprises contacting the fermentation broth with an organic acid and an enzyme capable of esterifying the butanol with the organic acid. In another aspect, the method further comprises vaporizing at least a portion of the fermentation broth to form a vapor stream comprising water and butanol.

[0027] In one aspect, the rate of butanol production from a host cell of the invention is increased by at least about 10%, at least about 20%, at least about 30%, at least about 40%, at least about 50%, at least about 60%, at least about 70%, at least about 80%, at least about 90%, at least about 95%, at least about 2-fold, at least about 3-fold, or at least about 4-fold greater as compared to a host cell that does not have a polynucleotide encoding a polypeptide that catalyzes the conversion of pyruvate to acetolactate integrated into the chromosome. In another aspect, the titer of butanol production from a host cell of the invention is increased by at least about 10%, at least about 20%, at least about 30%, at least about 40%, at least about 50%, at least about 60%, at least about 70%, at least about 80%, at least about 90%, at least about 95%, at least about 2-fold, at least about 3-fold, or at least about 4-fold greater as compared to a host cell that does not have a polynucleotide encoding a polypeptide that catalyzes the conversion of pyruvate to acetolactate integrated into the chromosome.

[0028] In another aspect, the invention relates to a method for increasing the copy number or expression of a non-integrated recombinant polynucleotide encoding a polypeptide that catalyzes a step of a biosynthetic pathway described herein, comprising contacting a host cell of the invention with a fermentable carbon substrate to form a fermentation broth under conditions whereby the product of the biosynthetic pathway is produced. In another aspect, the invention relates to a method for increasing the flux in a pyruvate-utilizing biosynthetic pathway comprising: (a) providing a recombinant host cell

of the invention; and (b) contacting the host cell with a fermentable carbon substrate to form a fermentation broth under conditions whereby the flux in the pyruvate-utilizing biosynthetic pathway in the host cell is increased.

[0029] In another aspect, the invention relates to a method of producing a recombinant host cell comprising transforming the host cell with (i) one or more polynucleotides encoding a polypeptide that catalyzes a substrate to product conversion of a pyruvate-utilizing biosynthetic pathway; and (ii) a polynucleotide encoding a peptide that catalyzes the conversion of pyruvate to acetolactate; wherein the polynucleotide of (ii) is integrated into the chromosome. In another aspect, the invention relates to a method of increasing the formation of a product of a pyruvate-utilizing biosynthetic pathway comprising (i) providing a recombinant host cell of the invention; and (ii) growing the host cell under conditions wherein the product of the pyruvate-utilizing pathway is formed at an amount of product greater than the amount of product formed by a host cell comprising a polynucleotide encoding a polypeptide which catalyzes the conversion of pyruvate to acetolactate that is not integrated into the chromosome.

[0030] In another aspect, the invention relates to a composition comprising (i) a host cell of the invention; (ii) butanol; and (iii) an extractant. In another aspect, the invention relates to a composition comprising (i) a host cell of the invention; (ii) butanol; (iii) an extractant; and (iv) an esterification enzyme. In another aspect, the butanol of such composition is isobutanol.

[0031] In another aspect, the invention relates to a method for chromosomally integrating acetolactate synthase (als) into a yeast host cell comprising transforming said host cell with an integration vector comprising SEQ ID NO: 131. In another aspect, the host cell further comprises an isobutanol biosynthetic pathway. In another aspect, the host comprises at least two chromosomally integrated polynucleotides.

BRIEF DESCRIPTION OF THE DRAWINGS/FIGURES

[0032] The accompanying drawings, which are incorporated herein and form a part of the specification, illustrate the present invention and, together with the description, further

serve to explain the principles of the invention and to enable a person skilled in the pertinent art to make and use the invention.

- [0033] FIG. 1 shows pathways and enzymes for pyruvate utilization.
- [0034] FIG. 2 shows three different isobutanol biosynthetic pathways
- [0035] FIG. 3 shows four different 2-butanol biosynthetic pathways.
- [0036] FIGS. 4A-4D show sequence relationships of acetolactate synthase (als) coding regions that were retrieved by BLAST analysis using the sequence of *B. subtilis AlsS*, limiting to the 100 closest neighbors. The als encoding sequence is identified by its source organism.
- [0037] FIG. 5 shows the PNY2204 locus (pdc1Δ::ilvD::pUC19-kan::FBA-alsS::TRX1).
- [0038] FIG. 6 shows the PNY2211 locus (pdc1Δ::ilvD::FBA-alsS::TRX1). The *alsS* gene integration in the pdc1-trx1 intergenic region is considered a "scarless" insertion since vector, marker gene and loxP sequences are lost.
- [0039] The invention can be more fully understood from the following detailed description and the accompanying sequence descriptions which form a part of this application.
- [0040] The following sequences conform with 37 C.F.R. §§ 1.821-1.825 ("Requirements for Patent Applications Containing Nucleotide Sequences and/or Amino Acid Sequence Disclosures - the Sequence Rules") and are consistent with World Intellectual Property Organization (WIPO) Standard ST.25 (2009) and the sequence listing requirements of the EPO and PCT [Rules 5.2 and 49.5(a-bis), and Section 208 and Annex C of the Administrative Instructions]. The symbols and format used for nucleotide and amino acid sequence data comply with the rules set forth in 37 C.F.R. § 1.822.

Table 1: SEQ ID Numbers of Coding Regions and Proteins Referred to Herein

Description	SEQ ID NO :	SEQ ID NO :
	Nucleic acid	Amino acid
<i>Klebsiella pneumoniae budB</i> (acetolactate synthase)	1	2
<i>Bacillus subtilis alsS</i> (acetolactate synthase)	3	4
<i>Lactococcus lactis als</i> (acetolactate synthase)	5	6

<i>Als Staphylococcus aureus</i>	7	8
<i>Als Listeria monocytogenes</i>	9	10
<i>Als Streptococcus mutans</i>	11	12
<i>Als Streptococcus thermophiles</i>	13	14
<i>Als Vibrio angustum</i>	15	16
<i>Als Bacillus cereus</i>	17	18
<i>budA</i> , acetolactate decarboxylase from <i>Klebsiella pneumoniae</i> ATCC 25955	19	20
<i>alsD</i> , acetolactate decarboxylase from <i>Bacillus subtilis</i>	21	22
<i>budA</i> , acetolactate decarboxylase from <i>Klebsiella terrigena</i>	23	24
<i>budC</i> , butanediol dehydrogenase from <i>Klebsiella pneumoniae</i> IAM1063	25	26
butanediol dehydrogenase from <i>Bacillus cereus</i>	27	28
<i>butB</i> , butanediol dehydrogenase from <i>Lactococcus lactis</i>	29	30
<i>RdhtA</i> , B12-indep diol dehydratase from <i>Roseburia inulinivorans</i>	31	32
<i>RdhtB</i> , B12-indep diol dehydratase reactivase from <i>Roseburia inulinivorans</i>	33	34
<i>sadB</i> , butanol dehydrogenase from <i>Achromobacter xylosoxidans</i>	35	36
<i>S. cerevisiae</i> ILV5 (acetohydroxy acid reductoisomerase)	37	38
<i>Vibrio cholerae</i> ketol-acid reductoisomerase	39	40
<i>Pseudomonas aeruginosa</i> ketol-acid reductoisomerase	41	42
<i>Pseudomonas fluorescens</i> ketol-acid reductoisomerase	43	44
<i>S. cerevisiae</i> ILV3 (Dihydroxyacid dehydratase; DHAD)	45	46
<i>Lactococcus lactis kivD</i> (branched-chain α -keto acid decarboxylase), codon optimized	47	48
<i>Lactococcus lactis kivD</i> (branched-chain α -keto acid decarboxylase)	49	48*
Pf5.IlvC-Z4B8 mutant <i>Pseudomonas fluorescens</i> acetohydroxy acid reductoisomerase	82	83

<i>Bacillus subtilis</i> kivD codon optimized for <i>S. cerevisiae</i> expression	84	85
<i>Equus caballus</i> alcohol dehydrogenase codon optimized for <i>S. cerevisiae</i> expression	86	87
<i>Streptococcus mutans</i> ilvD (DHAD)	88	89
K9G9 variant of <i>Anaerostipes caccae</i> KARI	-	225
K9D3 variant of <i>Anaerostipes caccae</i> KARI	-	224
<i>Beijerinckia indica</i> ADH	-	237
Ketoisovalerate decarboxylase from <i>Listeria grayi</i>	-	247
Ketoisovalerate decarboxylase from <i>Macrococcus caseolyticus</i>	-	248

* The same amino acid sequence is encoded by SEQ ID NOs:47 and 49.

Table 2: SEQ ID Numbers of Target Gene Coding Regions and Proteins Referred to Herein

Description	SEQ ID NO: Nucleic acid	SEQ ID NO: Amino acid
<i>PDC1</i> pyruvate decarboxylase from <i>Saccharomyces cerevisiae</i>	50	51
<i>PDC5</i> pyruvate decarboxylase from <i>Saccharomyces cerevisiae</i>	52	53
<i>PDC6</i> pyruvate decarboxylase from <i>Saccharomyces cerevisiae</i>	54	55
pyruvate decarboxylase from <i>Candida glabrata</i>	56	57
<i>PDC1</i> pyruvate decarboxylase from <i>Pichia stipites</i>	58	59
<i>PDC2</i> pyruvate decarboxylase from <i>Pichia stipites</i>	60	61
pyruvate decarboxylase from <i>Kluyveromyces lactis</i>	62	63
pyruvate decarboxylase from <i>Yarrowia lipolytica</i>	64	65
pyruvate decarboxylase from <i>Schizosaccharomyces pombe</i>	66	67

<i>GPD1</i> NAD-dependent glycerol-3-phosphate dehydrogenase from <i>Saccharomyces cerevisiae</i>	68	69
<i>GPD2</i> NAD-dependent glycerol-3-phosphate dehydrogenase from <i>Saccharomyces cerevisiae</i>	70	71
<i>GPD1</i> NAD-dependent glycerol-3-phosphate dehydrogenase from <i>Pichia stipitis</i>	72	73
<i>GPD2</i> NAD-dependent glycerol-3-phosphate dehydrogenase from <i>Pichia stipites</i>	74	75
NAD-dependent glycerol-3-phosphate dehydrogenase from <i>Kluyveromyces thermotolerans</i>	76	77
<i>GPD1</i> NAD-dependent glycerol-3-phosphate dehydrogenase from <i>Schizosaccharomyces pombe</i>	78	79
<i>GPD2</i> NAD-dependent glycerol-3-phosphate dehydrogenase from <i>Schizosaccharomyces pombe</i>	80	81
<i>AFT1</i> from <i>Saccharomyces cerevisiae</i>	227	228
<i>AFT2</i> from <i>Saccharomyces cerevisiae</i>	229	230
<i>FRA2</i> from <i>Saccharomyces cerevisiae</i>	231	232
<i>GRX3</i> from <i>Saccharomyces cerevisiae</i>	233	234
<i>CCC1</i> from <i>Saccharomyces cerevisiae</i>	235	236
<i>ALD6</i> from <i>Saccharomyces cerevisiae</i>	-	223
<i>YMR226C</i> from <i>Saccharomyces cerevisiae</i>	-	226

[0041] SEQ ID NOs:90-222 and 243-246 are sequences used and described in the Examples.

[0042] SEQ ID NOs: 238-242 are hybrid promoter sequences referred to herein.

[0043]

DETAILED DESCRIPTION OF THE INVENTION

- [0044] 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 invention belongs. In case of conflict, the present application including the definitions will control. 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 incorporated by reference in their entireties for all purposes as if each individual publication or patent application were specifically and individually indicated to be incorporated by reference, unless only specific sections of patents or patent publications are indicated to be incorporated by reference.
- [0045] The materials, methods and examples are illustrative only and are not intended to be limiting. Other features and advantages of the invention will be apparent from the detailed description and from the claims.
- [0046] The present invention relates to recombinant microorganisms and methods for the production of butanol. The present invention meets a number of commercial and industrial needs. Butanol is an important industrial commodity chemical with a variety of applications, where its potential as a fuel or fuel additive is particularly significant. Although only a four-carbon alcohol, butanol has energy content similar to that of gasoline and can be blended with any fossil fuel. Butanol is favored as a fuel or fuel additive as it yields only CO₂ and little or no SO_x or NO_x when burned in the standard internal combustion engine. Additionally butanol is less corrosive than ethanol, another fuel additive.
- [0047] In addition to its utility as a biofuel or fuel additive, butanol has the potential of impacting hydrogen distribution problems in the emerging fuel cell industry. Fuel cells today are plagued by safety concerns associated with hydrogen transport and distribution. Butanol can be easily reformed for its hydrogen content and can be distributed through existing gas stations in the purity required for either fuel cells or vehicles.
- [0048] The following definitions and abbreviations are to be used for the interpretation of the claims and the specification.
- [0049] As used herein, the terms "comprises," "comprising," "includes," "including," "has," "having," "contains," or "containing," or any other variation thereof, are intended to be non-exclusive or open-ended. 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).

[0050] As used herein, the term "consisting essentially of" in the context of a claim is intended to represent the intermediate ground between a closed claim written in a "consisting of" format and a fully open claim written in a "comprising" format. *See* M.P.E.P. § 2111.03.

[0051] 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.

[0052] 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 use 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 one embodiment, the term "about" means within 10% of the reported numerical value, preferably within 5% of the reported numerical value.

[0053] 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 specification and the claims.

[0054] The term "butanol" as used herein, refers to 2-butanol, isobutanol or mixtures thereof.

- [0055] The term "isobutanol biosynthetic pathway" refers to an enzyme pathway to produce isobutanol from pyruvate.
- [0056] The term "2-butanol biosynthetic pathway" refers to an enzyme pathway to produce 2-butanol from pyruvate.
- [0057] The term "2-butanone biosynthetic pathway" refers to an enzyme pathway to produce 2-butanone from pyruvate.
- [0058] The term "extractant" as used herein refers to one or more organic solvents which are used to extract butanol and/or other components from a fermentation broth.
- [0059] The terms "acetolactate synthase" and "acetolactate synthetase" are used interchangeably herein to refer to an enzyme that catalyzes the conversion of pyruvate to acetolactate and CO₂. Examples of acetolactate synthases are known by the EC number 2.2.1.6 (*Enzyme Nomenclature 1992*, Academic Press, San Diego). These enzymes are available from a number of sources, including, but not limited to, *Bacillus subtilis* [GenBank Nos: CAB15618 and Z99122, NCBI (National Center for Biotechnology Information) amino acid sequence, NCBI nucleotide sequence, respectively], *Klebsiella pneumoniae* (GenBank Nos: AAA25079 and M73842), and *Lactococcus lactis* (GenBank Nos: AAA25161 and L16975). Additional examples are also provided in Table 1.
- [0060] The term "acetolactate decarboxylase" refers to a polypeptide (or polypeptides) having an enzyme activity that catalyzes the conversion of alpha-acetolactate to acetoin. Examples of acetolactate decarboxylases are known as EC 4.1.1.5 and are available, for example, from *Bacillus subtilis* (DNA: SEQ ID NO:21, Protein: SEQ ID NO:22), *Klebsiella terrigena* (DNA: SEQ ID NO:23, Protein: SEQ ID NO:24) and *Klebsiella pneumoniae* (DNA: SEQ ID NO:19, protein: SEQ ID NO:20).
- [0061] The term "acetoin aminase" 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 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 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 aminase is the amine:pyruvate aminotransferase (also called amine:pyruvate transaminase) described by Shin and Kim (J. Org. Chem. 67:2848-2853 (2002)).

[0062] The term “aminobutanol kinase” refers to a polypeptide (or polypeptides) having an enzyme activity that catalyzes the conversion of 3-amino-2-butanol to 3-amino-2-butanol *O*-phosphate. Aminobutanol kinase may utilize ATP as the phosphate donor. There are reports of enzymes that catalyze the analogous reaction on the similar substrates ethanolamine and 1-amino-2-propanol (Jones et al. (1973) *Biochem. J.* 134:167-182). U.S. Appl. Pub. No. 20070292927 describes, in Example 14, an amino alcohol kinase of *Erwinia carotovora* subsp. *atroseptica*.

[0063] The term “aminobutanol phosphate phospho-lyase”, also called “amino alcohol *O*-phosphate lyase”, refers to a polypeptide (or polypeptides) having an enzyme activity that catalyzes the conversion of 3-amino-2-butanol *O*-phosphate to 2-butanone. Aminobutanol 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. (1973) *Biochem J.* 134:167-182). U.S. Appl. Pub. No. 20070292927 describes, in Example 15, a newly identified aminobutanol phosphate phospho-lyase from the organism *Erwinia carotovora*.

[0064] The term "butanediol dehydrogenase" also known as "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 can have specificity for production of (*R*)- or (*S*)-stereochemistry in the alcohol product. Examples of (*S*)-specific butanediol dehydrogenases are known as EC 1.1.1.76 and are available, for example, from *Klebsiella pneumoniae* (DNA: SEQ ID NO:25, protein: SEQ ID NO:26). Examples of (*R*)-specific butanediol dehydrogenases are known as EC 1.1.1.4 and are available, for example, from *Bacillus cereus* (DNA: SEQ ID NO:27, protein: SEQ ID NO:28), and *Lactococcus lactis* (DNA: SEQ ID NO:29, protein: SEQ ID NO:30).

[0065] The terms "acetohydroxy acid isomeroeductase" and "acetohydroxy acid reductoisomerase" and “ketol-acid reductoisomerase” (KARI) are used interchangeably herein to refer to an enzyme that catalyzes the conversion of acetolactate to 2,3-dihydroxyisovalerate. Suitable enzymes utilize NADH (reduced nicotinamide adenine

dinucleotide) and/or NADPH as electron donor. Examples of acetohydroxy acid isomeroreductases are known by the EC number 1.1.1.86 and sequences are available from a vast array of microorganisms, including, but not limited to, *Escherichia coli* (GenBank Nos: NP_418222 and NC_000913), *Saccharomyces cerevisiae* (GenBank Nos: NP_013459 and NC_001144), *Methanococcus maripaludis* (GenBank Nos: CAF30210 and BX957220), and *Bacillus subtilis* (GenBank Nos: CAB14789 and Z99118).

[0066] The term "acetohydroxy acid dehydratase" refers to an enzyme that catalyzes the conversion of 2,3-dihydroxyisovalerate to α -ketoisovalerate. Examples of acetohydroxy acid dehydratases are known by the EC number 4.2.1.9. These enzymes are available from a vast array of microorganisms, including, but not limited to, *E. coli* (GenBank Nos: YP_026248 and NC_000913), *S. cerevisiae* (GenBank Nos: NP_012550 and NC_001142), *M. maripaludis* (GenBank Nos: CAF29874 and BX957219), and *B. subtilis* (GenBank Nos: CAB14105 and Z99115).

[0067] The term "branched-chain α -keto acid decarboxylase" refers to an enzyme that catalyzes the conversion of α -ketoisovalerate to isobutyraldehyde and CO₂. Examples of branched-chain α -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, AY548760, CAG34226, and AJ746364), *Salmonella typhimurium* (GenBank Nos: NP_461346 and NC_003197), and *Clostridium acetobutylicum* (GenBank Nos: NP_149189 and NC_001988).

[0068] The term "branched-chain alcohol dehydrogenase" refers to an enzyme that catalyzes the conversion of isobutyraldehyde to isobutanol. Examples of 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). These enzymes utilize NADH (reduced nicotinamide adenine dinucleotide) and/or NADPH as electron donor and are available from a number of sources, including, but not limited to, *S. cerevisiae* (GenBank Nos: NP_010656, NC_001136; NP_014051; and NC_001145), *E. coli* (GenBank Nos: NP_417484 and NC_000913) and *C. acetobutylicum* (GenBank Nos: NP_349892, NC_003030; NP_349891, and NC_003030).

[0069] The term "branched-chain keto acid dehydrogenase" refers to an enzyme that catalyzes the conversion of α -ketoisovalerate to isobutyryl-CoA (isobutyryl-coenzyme A), using NAD⁺ (nicotinamide adenine dinucleotide) as electron acceptor. Examples of

branched-chain keto acid dehydrogenases are known by the EC number 1.2.4.4. These 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, CAB14337, and Z99116) and *Pseudomonas putida* (GenBank Nos: AAA65614, M57613, AAA65615, M57613, AAA65617, M57613, AAA65618, and M57613).

[0070] The term "acylating aldehyde dehydrogenase" refers to an enzyme that catalyzes the conversion of isobutyryl-CoA to isobutyraldehyde, using either NADH or NADPH as electron donor. Examples of acylating aldehyde dehydrogenases are known by the EC numbers 1.2.1.10 and 1.2.1.57. These enzymes are available from multiple sources, including, but not limited to, *Clostridium beijerinckii* (GenBank Nos: AAD31841 and AF157306), *C. acetobutylicum* (GenBank Nos: NP_149325, NC_001988, NP_149199, and NC_001988), *P. putida* (GenBank Nos: AAA89106 and U13232), and *Thermus thermophilus* (GenBank Nos: YP_145486 and NC_006461).

[0071] The term "transaminase" refers to an enzyme that catalyzes the conversion of α -ketoisovalerate to L-valine, using either alanine or glutamate as amine donor. Examples of transaminases are known by the EC numbers 2.6.1.42 and 2.6.1.66. These enzymes are available from a number of sources. Examples of sources for alanine-dependent enzymes include, but are not limited to, *E. coli* (GenBank Nos: YP_026231 and NC_000913) and *Bacillus licheniformis* (GenBank Nos: YP_093743 and NC_006322). Examples of sources for glutamate-dependent enzymes include, but are not limited to, *E. coli* (GenBank Nos: YP_026247 and NC_000913), *S. cerevisiae* (GenBank Nos: NP_012682 and NC_001142) and *Methanobacterium thermoautotrophicum* (GenBank Nos: NP_276546 and NC_000916).

[0072] The term "valine dehydrogenase" refers to an enzyme that catalyzes the conversion of α -ketoisovalerate to L-valine, using NAD(P)H as electron donor and ammonia as amine donor. Examples of valine dehydrogenases are known by the EC numbers 1.4.1.8 and 1.4.1.9 and are available from a number of sources, including, but not limited to, *Streptomyces coelicolor* (GenBank Nos: NP_628270 and NC_003888) and *B. subtilis* (GenBank Nos: CAB14339 and Z99116).

- [0073] The term "valine decarboxylase" refers to an enzyme that catalyzes the conversion of L-valine to isobutylamine and CO₂. Examples of valine decarboxylases are known by the EC number 4.1.1.14. These enzymes are found in *Streptomyces*, such as for example, *Streptomyces viridifaciens* (GenBank Nos: AAN10242 and AY116644).
- [0074] The term "omega transaminase" refers to an enzyme that catalyzes the conversion of isobutylamine to isobutyraldehyde using a suitable amino acid as amine donor. Examples of 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* (GenBank Nos: AAP92672 and AY330220), *Ralstonia eutropha* (GenBank Nos: YP_294474 and NC_0073479), *Shewanella oneidensis* (GenBank Nos: NP_719046 and NC_004347), and *P. putida* (GenBank Nos: AAN66223 and AE016776).
- [0075] The term "isobutyryl-CoA mutase" refers to an enzyme that catalyzes the conversion of butyryl-CoA to isobutyryl-CoA. This enzyme uses coenzyme B₁₂ as cofactor. Examples of 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, and AJ246005), *S. coelicolor* (GenBank Nos: CAB70645, AL939123, CAB92663, and L939121), and *Streptomyces avermitilis* (GenBank Nos: NP_824008, NC_003155, NP_824637 and NC_003155).
- [0076] The term "substantially free" when used in reference to the presence or absence of enzyme activities (*e.g.*, pyruvate decarboxylase) in carbon pathways that compete with the present isobutanol pathway means that the level of the enzyme is substantially less than that of the same enzyme in the wildtype host, where less than about 20% of the wildtype level is preferred and less than about 15% or 10% of the wildtype level are more preferred. The activity can be less than about 5%, 4%, 3%, 2% or 1% of wildtype activity.
- [0077] The term "a facultative anaerobe" refers to a microorganism that can grow in both aerobic and anaerobic environments.
- [0078] The term "carbon substrate" or "fermentable carbon substrate" refers to a carbon source capable of being metabolized by host organisms of the present invention and particularly carbon sources selected from the group consisting of monosaccharides, oligosaccharides, polysaccharides, and one-carbon substrates or mixtures thereof.

Sources for carbon substrates can include any feedstock, such as renewable-source feedstocks, including but not limited to any sugar or starch containing biomass such as corn, wheat, sugar cane, wood, algae; any agricultural wastes or residues and any lignocellulosic and/or hemicellulosic materials.

[0079] 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. "Native gene" refers to a gene as found in nature with its own regulatory sequences. "Chimeric gene" refers to any gene that is not a native gene, comprising regulatory and coding sequences that are not found together in nature. Accordingly, a chimeric gene can comprise regulatory sequences and coding sequences that are derived from different sources, or regulatory sequences and coding sequences derived from the same source, but arranged in a manner different than that found in nature. "Endogenous gene" refers to a native gene in its natural location in the genome of an organism. A "foreign gene" or "heterologous gene" refers to a gene not normally found in the host organism, but that is introduced into the host organism, e.g. by gene transfer, or is found or is native to a host organism but is modified in some way to affect its functioning. A polynucleotide integrated (whether a nature or non-native polynucleotide) into a chromosome as described herein is considered a heterologous polynucleotide. Foreign genes can comprise native genes inserted into a non-native organism, or chimeric genes. A "transgene" is a gene that has been introduced into the genome by a transformation procedure.

[0080] As used herein the terms "coding sequence" and "coding region" refer to a DNA sequence that codes for a specific amino acid sequence. "Suitable regulatory sequences" refer to nucleotide sequences located upstream (5' non-coding sequences), within, or downstream (3' non-coding sequences) of a coding sequence, and which influence the transcription, RNA processing or stability, or translation of the associated coding sequence. Regulatory sequences can include promoters, translation leader sequences, introns, polyadenylation recognition sequences, RNA processing site, effector binding site and stem-loop structure.

[0081] The term "promoter" refers to a DNA sequence capable of controlling the expression of a coding sequence or functional RNA. In general, a coding sequence is located 3' to a promoter sequence. Promoters can be derived in their entirety from a

native gene, or be composed of different elements derived from different promoters found in nature, or even comprise synthetic DNA segments. It is understood by those skilled in the art that different promoters may direct the expression of a gene in different tissues or cell types, or at different stages of development, or in response to different environmental or physiological conditions. Promoters which cause a gene to be expressed in most cell types at most times are commonly referred to as "constitutive promoters." It is further recognized that because in most cases the exact boundaries of regulatory sequences have not been completely defined, DNA fragments of different lengths may have identical promoter activity.

[0082] 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.

[0083] 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 can also refer to translation of mRNA into a polypeptide.

[0084] As used herein the term "transformation" refers to the transfer of a nucleic acid fragment into a host organism, resulting in genetically stable inheritance. Host organisms containing the transformed nucleic acid fragments are referred to as "transgenic" or "recombinant" or "transformed" organisms.

[0085] 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 can 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.

[0086] As used herein, the term "codon degeneracy" refers to the nature of the genetic code permitting variation of the nucleotide sequence without affecting the amino acid sequence of an encoded polypeptide. The skilled artisan is well aware of the "codon-bias" exhibited by a specific host cell in usage of nucleotide codons to specify a given amino acid. Therefore, when synthesizing a gene for improved expression in a host cell, it is desirable to design the gene such that its frequency of codon usage approaches the frequency of preferred codon usage of the host cell.

[0087] The term "endogenous" as used herein refers to something that is produced or synthesized by the organism or that is added to the surroundings of the organism.

[0088] The term "codon-optimized" as it refers to genes or 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.

[0089] As used herein, an "isolated nucleic acid fragment" or "isolated nucleic acid molecule" are used interchangeably and mean a polymer of RNA or DNA that is single- or double-stranded, optionally containing synthetic, non-natural or altered nucleotide bases. An isolated nucleic acid fragment in the form of a polymer of DNA can be comprised of one or more segments of cDNA, genomic DNA or synthetic DNA.

[0090] A nucleic acid fragment is "hybridizable" to another nucleic acid fragment, such as a cDNA, genomic DNA, or RNA molecule, when a single-stranded form of the nucleic acid fragment can anneal to the other nucleic acid fragment under the appropriate conditions of temperature and solution ionic strength. Hybridization and washing conditions are well known and exemplified in Sambrook, J., Fritsch, E. F. and Maniatis, T. Molecular Cloning: A Laboratory Manual, 2nd ed., Cold Spring Harbor Laboratory: Cold Spring Harbor, NY (1989), particularly Chapter 11 and Table 11.1 therein (entirely incorporated herein by reference). The conditions of temperature and ionic strength

determine the "stringency" of the hybridization. Stringency conditions can be adjusted to screen for moderately similar fragments (such as homologous sequences from distantly related organisms), to highly similar fragments (such as genes that duplicate functional enzymes from closely related organisms). Post-hybridization washes determine stringency conditions. One set of conditions uses a series of washes starting with 6X SSC, 0.5% SDS at room temperature for 15 min, then repeated with 2X SSC, 0.5% SDS at 45 °C for 30 min, and then repeated twice with 0.2X SSC, 0.5% SDS at 50 °C for 30 min. Another set of stringent conditions uses higher temperatures in which the washes are identical to those above except for the temperature of the final two 30 min washes in 0.2X SSC, 0.5% SDS was increased to 60 °C. Another set of highly stringent conditions uses two final washes in 0.1X SSC, 0.1% SDS at 65 °C. An additional set of stringent conditions include hybridization at 0.1X SSC, 0.1% SDS, 65 °C and washes with 2X SSC, 0.1% SDS followed by 0.1X SSC, 0.1% SDS, for example.

[0091] Hybridization requires that the two nucleic acids contain complementary sequences, although depending on the stringency of the hybridization, mismatches between bases are possible. The appropriate stringency for hybridizing nucleic acids depends on the length of the nucleic acids and the degree of complementation, variables well known in the art. The greater the degree of similarity or homology between two nucleotide sequences, the greater the value of T_m for hybrids of nucleic acids having those sequences. The relative stability (corresponding to higher T_m) of nucleic acid hybridizations decreases in the following order: RNA:RNA, DNA:RNA, DNA:DNA. For hybrids of greater than 100 nucleotides in length, equations for calculating T_m have been derived (*see* Sambrook *et al.*, *supra*, 9.50-9.51). For hybridizations with shorter nucleic acids, *i.e.*, oligonucleotides, the position of mismatches becomes more important, and the length of the oligonucleotide determines its specificity (*see* Sambrook *et al.*, *supra*, 11.7-11.8). In one embodiment the length for a hybridizable nucleic acid is at least about 10 nucleotides. In another embodiment, a minimum length for a hybridizable nucleic acid is at least about 15 nucleotides, at least about 20 nucleotides, or at least about 30 nucleotides. Furthermore, the skilled artisan will recognize that the temperature and wash solution salt concentration can be adjusted as necessary according to factors such as length of the probe.

[0092] A "substantial portion" of an amino acid or nucleotide sequence is that portion comprising enough of the amino acid sequence of a polypeptide or the nucleotide sequence of a gene to putatively identify that polypeptide or gene, either by manual evaluation of the sequence by one skilled in the art, or by computer-automated sequence comparison and identification using algorithms such as BLAST (Altschul *et al.*, *J. Mol. Biol.*, 215:403-410 (1993)). In general, a sequence of ten or more contiguous amino acids or thirty or more nucleotides is necessary in order to putatively identify a polypeptide or nucleic acid sequence as homologous to a known protein or gene. Moreover, with respect to nucleotide sequences, gene specific oligonucleotide probes comprising 20-30 contiguous nucleotides can be used in sequence-dependent methods of gene identification (*e.g.*, Southern hybridization) and isolation (*e.g.*, *in situ* hybridization of bacterial colonies or bacteriophage plaques). In addition, short oligonucleotides of 12-15 bases can be used as amplification primers in PCR in order to obtain a particular nucleic acid fragment comprising the primers. Accordingly, a "substantial portion" of a nucleotide sequence comprises enough of the sequence to specifically identify and/or isolate a nucleic acid fragment comprising the sequence. The instant specification teaches the complete amino acid and nucleotide sequence encoding particular fungal proteins. The skilled artisan, having the benefit of the sequences as reported herein, may now use all or a substantial portion of the disclosed sequences for purposes known to those skilled in this art. Accordingly, the instant invention comprises the complete sequences as reported in the accompanying Sequence Listing, as well as substantial portions of those sequences as defined above.

[0093] The term "complementary" is used to describe the relationship between nucleotide bases that are capable of hybridizing to one another. For example, with respect to DNA, adenosine is complementary to thymine and cytosine is complementary to guanine.

[0094] The term "percent identity," as known in the art, is a relationship between two or more polypeptide sequences or two or more polynucleotide sequences, as determined by comparing the sequences. In the art, "identity" also means the degree of sequence relatedness between polypeptide or polynucleotide sequences, as the case may be, as determined by the match between strings of such sequences. "Identity" and "similarity" can be readily calculated by known methods, including but not limited to those described in: (1) Computational Molecular Biology (Lesk, A. M., Ed.) Oxford University: NY

(1988); (2) Biocomputing: Informatics and Genome Projects (Smith, D. W., Ed.) Academic: NY (1993); (3) Computer Analysis of Sequence Data, Part I (Griffin, A. M., and Griffin, H. G., Eds.) Humana: NJ (1994); (4) Sequence Analysis in Molecular Biology (von Heinje, G., Ed.) Academic (1987); and (5) Sequence Analysis Primer (Gribskov, M. and Devereux, J., Eds.) Stockton: NY (1991).

[0095] Preferred methods to determine identity are designed to give the best match between the sequences tested. Methods to determine identity and similarity are codified in publicly available computer programs. Sequence alignments and percent identity calculations may be performed using the MegAlign™ program of the LASERGENE bioinformatics computing suite (DNASTAR Inc., Madison, WI). Multiple alignment of the sequences is performed using the "Clustal method of alignment" which encompasses several varieties of the algorithm including the "Clustal V method of alignment" corresponding to the alignment method labeled Clustal V (described by Higgins *et al.*, *CABIOS*. 5:151-153, 1989; Higgins *et al.*, *Comput. Appl. Biosci.*, 8:189-191, 1992) and found in the MegAlign™ program of the LASERGENE bioinformatics computing suite (DNASTAR Inc.). For multiple alignments, the default values correspond to GAP PENALTY=10 and GAP LENGTH PENALTY=10. Default parameters for pairwise alignments and calculation of percent identity of protein sequences using the Clustal method are KTUPLE=1, GAP PENALTY=3, WINDOW=5 and DIAGONALS SAVED=5. For nucleic acids these parameters are KTUPLE=2, GAP PENALTY=5, WINDOW=4 and DIAGONALS SAVED=4. After alignment of the sequences using the Clustal V program, it is possible to obtain a "percent identity" by viewing the "sequence distances" table in the same program. Additionally the "Clustal W method of alignment" is available and corresponds to the alignment method labeled Clustal W (described by Higgins *et al.*, *CABIOS*. 5:151-153 (1989); Higgins *et al.*, *Comput. Appl. Biosci.* 8:189-191(1992)) and found in the MegAlign™ v6.1 program of the LASERGENE bioinformatics computing suite (DNASTAR Inc.). Default parameters for multiple alignment (GAP PENALTY=10, GAP LENGTH PENALTY=0.2, Delay Divergen Seqs(%)=30, DNA Transition Weight=0.5, Protein Weight Matrix=Gonnet Series, DNA Weight Matrix=IUB). After alignment of the sequences using the Clustal W program, it is possible to obtain a "percent identity" by viewing the "sequence distances" table in the same program.

[0096] It is well understood by one skilled in the art that many levels of sequence identity are useful in identifying polypeptides, from other species, wherein such polypeptides have the same or similar function or activity. Useful examples of percent identities include, but are not limited to: 24%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, or 95%, or any integer percentage from 24% to 100% may be useful in describing the present invention, such as 25%, 26%, 27%, 28%, 29%, 30%, 31%, 32%, 33%, 34%, 35%, 36%, 37%, 38%, 39%, 40%, 41%, 42%, 43%, 44%, 45%, 46%, 47%, 48%, 49%, 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99%. Suitable nucleic acid fragments not only have the above homologies but typically encode a polypeptide having at least 50 amino acids, at least 100 amino acids, at least 150 amino acids, at least 200 amino acids, or at least 250 amino acids.

[0097] The term "sequence analysis software" refers to any computer algorithm or software program that is useful for the analysis of nucleotide or amino acid sequences. "Sequence analysis software" can be commercially available or independently developed. Typical sequence analysis software will include, but is not limited to: (1) the GCG suite of programs (Wisconsin Package Version 9.0, Genetics Computer Group (GCG), Madison, WI); (2) BLASTP, BLASTN, BLASTX (Altschul *et al.*, *J. Mol. Biol.*, 215:403-410 (1990)); (3) DNASTAR (DNASTAR, Inc. Madison, WI); (4) Sequencher (Gene Codes Corporation, Ann Arbor, MI); and (5) the FASTA program incorporating the Smith-Waterman algorithm (W. R. Pearson, *Comput. Methods Genome Res.*, [Proc. Int. Symp.] (1994), Meeting Date 1992, 111-20. Editor(s): Suhai, Sandor. Plenum: New York, NY). Within the context of this application it will be understood that where sequence analysis software is used for analysis, that the results of the analysis will be based on the "default values" of the program referenced, unless otherwise specified. As used herein "default values" mean any set of values or parameters that originally load with the software when first initialized.

[0098] Standard recombinant DNA and molecular cloning techniques used here 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, NY (1989) (hereinafter "Maniatis"); and by Silhavy, T. J., Bannan, M. L. and Enquist, L. W., *Experiments with Gene Fusions*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY (1984); and by Ausubel, F. M. *et al.*, *Current Protocols in Molecular Biology*, published by Greene Publishing Assoc. and Wiley-Interscience (1987). Additional methods 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, CA). 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; M A Romanos, et al. *Nucleic Acids Res.* 1991 January 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).

Biosynthetic Pathway Production Through Conversion of Pyruvate to Acetolactate

[0099] Microbial cells produce pyruvate from sugars, which is then utilized in a number of pathways of cellular metabolism including those shown in FIG. 1. Microbial host cells can be engineered to produce a number of desirable products with the initial biosynthetic pathway step being conversion of endogenous pyruvate to acetolactate. Engineered biosynthetic pathways for synthesis of isobutanol (FIG. 2) are described in U.S. Appl. Pub. No. 20070092957, which is herein incorporated by reference, and for synthesis of 2-butanol and 2-butanone (FIG. 3) are described in U.S. Appl. Pub. Nos. 20070259410 and 20070292927, which are herein incorporated by reference. The product 2,3-butanediol is an intermediate in the biosynthetic pathway described in U.S. Appl. Pub. No. 20070292927. Each of these pathways has the initial step of converting pyruvate to acetolactate by acetolactate synthase. Therefore, product yield from these biosynthetic pathways will in part depend upon the amount of acetolactate that can be produced from pyruvate and the amount of pyruvate that is available.

[00100] Applicants have discovered that a recombinant host cell comprising a polynucleotide encoding a polypeptide which catalyzes the conversion of pyruvate to acetolactate integrated into the chromosome of the host cell can have improved production of a product of a pyruvate-utilizing biosynthetic pathway (*e.g.*, a butanol such as isobutanol). Applicants found that host cells of the invention can have improved butanol production as shown by increased product titer, increased production rate or increased cell density compared to cells wherein the polynucleotide is not integrated into the chromosome.

[00101] In embodiments, the present invention relates to a recombinant host cell comprising a pyruvate-utilizing biosynthetic pathway and a polynucleotide encoding a polypeptide which catalyzes the conversion of pyruvate to acetolactate integrated into the chromosome of the host cell. In embodiments, the polynucleotide is heterologous to the host cell. In embodiments, the pyruvate-utilizing biosynthetic pathway comprises one or more polynucleotides encoding a polypeptide that catalyzes substrate to product conversions of the pathway. In embodiments, one or more of the polynucleotides are integrated into the chromosome of the host cell.

Expression and integration of acetolactate synthase

[00102] In embodiments of the invention, a polypeptide that catalyzes a substrate to product conversion of pyruvate to acetolactate is an acetolactate synthase. Endogenous acetolactate synthase in a host cell of the invention can be encoded in the mitochondrial genome and expressed in the mitochondria. In embodiments, to prepare a recombinant host cell of the present invention (such as yeast), a genetic modification is made that provides cytosolic expression of acetolactate synthase. In such embodiments, acetolactate synthase is expressed from the nucleus and no mitochondrial targeting signal is included so that the enzyme remains in the cytosol (cytosol-localized). Cytosolic expression of acetolactate synthase is described in US Application Publication No. 20090305363.

[00103] Acetolactate synthase enzymes, which also can be called acetohydroxy acid synthase, belong to EC 2.2.1.6 (switched from 4.1.3.18 in 2002), and are well-known. These enzymes participate in the biosynthetic pathway for the proteinogenic amino acids leucine and valine, as well as in the pathway for fermentative production of 2,3-butanediol and acetoin in a number of organisms.

- [00104] The skilled person will appreciate that polypeptides having acetolactate synthase activity isolated from a variety of sources can be useful in the present invention independent of sequence homology. Suitable acetolactate synthase enzymes are available from a number of sources, as described in the definitions. Acetolactate synthase enzyme activities that have substrate preference for pyruvate over ketobutyrate are of particular utility, such as those encoded by *alsS* of *Bacillus* and *budB* of *Klebsiella* (Gollop *et al.*, *J. Bacteriol.* 172(6):3444-3449, 1990; and Holtzclaw *et al.*, *J. Bacteriol.* 121(3):917-922, 1975).
- [00105] Because acetolactate synthases are well known, and because of the prevalence of genomic sequencing, suitable acetolactate synthases can be readily identified by one skilled in the art on the basis of sequence similarity using bioinformatics approaches. Typically BLAST (described above) searching of publicly available databases with known acetolactate synthase amino acid sequences, such as those provided herein, is used to identify acetolactate synthases, and their encoding sequences, that may be used in the present strains. For example, acetolactate synthases that are the 100 closest neighbors of the *B. subtilis* AlsS sequence are depicted in a phylogenetic tree in FIG. 4. The homology relationships between the sequences identified are shown in this tree. Among these sequences are those having 40% identity, yet these have been verified as acetolactate synthases. Acetolactate synthase proteins having at least about 70-75%, 75%-80%, 80-85%, 85%-90%, 90%-95% or at least about 98% or 99% sequence identity to any of the acetolactate synthase proteins in Table 1, or any of the acetolactate synthase proteins represented in FIG. 4 can be used in the present strains. Identities are based on the Clustal W method of alignment using the default parameters of GAP PENALTY=10, GAP LENGTH PENALTY=0.1, and Gonnet 250 series of protein weight matrix.
- [00106] Examples of sequences encoding acetolactate synthase which can be used to provide cytosolic expression of acetolactate synthase (*als*) activity are listed in Table 1. Additional acetolactate synthase encoding sequences that can be used for yeast cytosolic expression can be identified in the literature and in bioinformatics databases well known to the skilled person, and some coding regions for *als* proteins are represented in FIG. 4 by the source organism. Any *als* having EC number 2.2.1.6 can be identified by one skilled in the art and can be used in the present host cells.

[00107] Additionally, the sequences described herein or those recited in the art can be used to identify other homologs in nature. For example, each of the acetolactate synthase encoding nucleic acid fragments described herein may be used to isolate genes encoding homologous proteins. Isolation of homologous genes using sequence-dependent protocols is well known in the art. Examples of sequence-dependent protocols include, but are not limited to, (1) methods of nucleic acid hybridization; (2) methods of DNA and RNA amplification, as exemplified by various uses of nucleic acid amplification technologies [*e.g.*, polymerase chain reaction (PCR), Mullis *et al.*, U.S. Patent No. 4,683,202; ligase chain reaction (LCR), Tabor *et al.*, *Proc. Acad. Sci. USA* 82:1074, 1985; or strand displacement amplification (SDA), Walker *et al.*, *Proc. Natl. Acad. Sci. U.S.A.*, 89:392, 1992]; and (3) methods of library construction and screening by complementation.

[00108] For example, genes encoding similar proteins or polypeptides to the acetolactate synthase encoding genes described herein can be isolated directly by using all or a portion of the instant nucleic acid fragments as DNA hybridization probes to screen libraries from any desired organism using methodology well known to those skilled in the art. Specific oligonucleotide probes based upon the disclosed nucleic acid sequences can be designed and synthesized by methods known in the art (Maniatis, *supra*). Moreover, the entire sequences can be used directly to synthesize DNA probes by methods known to the skilled artisan (*e.g.*, random primers DNA labeling, nick translation or end-labeling techniques), or RNA probes using available *in vitro* transcription systems. In addition, specific primers can be designed and used to amplify a part of (or full-length of) the instant sequences. The resulting amplification products can be labeled directly during amplification reactions or labeled after amplification reactions, and used as probes to isolate full-length DNA fragments by hybridization under conditions of appropriate stringency.

[0100] Typically, in PCR-type amplification techniques, the primers have different sequences and are not complementary to each other. Depending on the desired test conditions, the sequences of the primers should be designed to provide for both efficient and faithful replication of the target nucleic acid. Methods of PCR primer design are common and well known in the art (Thein *et al.*, "The use of oligonucleotides as specific hybridization probes in the Diagnosis of Genetic Disorders," in *Human Genetic Diseases*:

A Practical Approach, K. E. Davis Ed., 1986, pp. 33-50, IRL: Herndon *et al.*, In *Methods in Molecular Biology*, White, B. A. Ed., (1993) Vol. 15, pp. 31-39, PCR Protocols: Current Methods and Applications. Humana: Totowa, NJ).

- [0101] Generally two short segments of the described sequences can be used in polymerase chain reaction protocols to amplify longer nucleic acid fragments encoding homologous genes from DNA or RNA. The polymerase chain reaction can also be performed on a library of cloned nucleic acid fragments wherein the sequence of one primer is derived from the described nucleic acid fragments, and the sequence of the other primer takes advantage of the presence of the polyadenylic acid tracts to the 3' end of the mRNA precursor encoding microbial genes.
- [0102] Alternatively, the second primer sequence can be based upon sequences derived from the cloning vector. For example, the skilled artisan can follow the RACE protocol (Frohman *et al.*, *PNAS USA* 85:8998 (1988)) to generate cDNAs by using PCR to amplify copies of the region between a single point in the transcript and the 3' or 5' end. Primers oriented in the 3' and 5' directions can be designed from the instant sequences. Using commercially available 3' RACE or 5' RACE systems (*e.g.*, BRL, Gaithersburg, MD), specific 3' or 5' cDNA fragments can be isolated (Ohara *et al.*, *PNAS USA* 86:5673, 1989; Loh *et al.*, *Science* 243:217, 1989).
- [0103] Alternatively, the described acetolactate synthase encoding sequences can be employed as hybridization reagents for the identification of homologs. The basic components of a nucleic acid hybridization test include a probe, a sample suspected of containing the gene or gene fragment of interest, and a specific hybridization method. Probes are typically single-stranded nucleic acid sequences that are complementary to the nucleic acid sequences to be detected. Probes are "hybridizable" to the nucleic acid sequence to be detected. The probe length can vary from 5 bases to tens of thousands of bases, and can depend upon the specific test to be done. Typically a probe length of about 15 bases to about 30 bases is suitable. However, only part of the probe molecule need be complementary to the nucleic acid sequence to be detected. In addition, the complementarity between the probe and the target sequence need not be perfect. Hybridization does occur between imperfectly complementary molecules with the result that a certain fraction of the bases in the hybridized region are not paired with the proper complementary base.

- [0104] Hybridization methods are well defined. Typically the probe and sample must be mixed under conditions that will permit nucleic acid hybridization. This involves contacting the probe and sample in the presence of an inorganic or organic salt under the proper concentration and temperature conditions. The probe and sample nucleic acids must be in contact for a long enough time that any possible hybridization between the probe and sample nucleic acid can occur. The concentration of probe or target in the mixture will determine the time necessary for hybridization to occur. The higher the probe or target concentration, the shorter the hybridization incubation time needed. Optionally, a chaotropic agent may be added. The chaotropic agent stabilizes nucleic acids by inhibiting nuclease activity. Furthermore, the chaotropic agent allows sensitive and stringent hybridization of short oligonucleotide probes at room temperature (Van Ness *et al.*, *Nucl. Acids Res.* 19:5143-5151, 1991). Suitable chaotropic agents include, but are not limited to, guanidinium chloride, guanidinium thiocyanate, sodium thiocyanate, lithium tetrachloroacetate, sodium perchlorate, rubidium tetrachloroacetate, potassium iodide and cesium trifluoroacetate, among others. The chaotropic agent can be present at a final concentration of about 3 M. If desired, one can add formamide to the hybridization mixture, typically 30-50% (v/v).
- [0105] Various hybridization solutions can be employed. Typically, these comprise from about 20 to 60% volume, preferably 30%, of a polar organic solvent. A common hybridization solution employs about 30-50% v/v formamide, about 0.15 to 1 M sodium chloride, about 0.05 to 0.1 M buffers (*e.g.*, sodium citrate, Tris-HCl, PIPES or HEPES (pH range about 6-9)), about 0.05 to 0.2% detergent (*e.g.*, sodium dodecylsulfate), or between 0.5-20 mM EDTA, FICOLL (Pharmacia Inc.) (about 300-500 kdal), polyvinylpyrrolidone (about 250-500 kdal) and serum albumin. Also included in the typical hybridization solution will be unlabeled carrier nucleic acids from about 0.1 to 5 mg/mL, fragmented nucleic DNA (*e.g.*, calf thymus or salmon sperm DNA, or yeast RNA), and optionally from about 0.5 to 2% wt/vol glycine. Other additives can also be included, such as volume exclusion agents that include a variety of polar water-soluble or swellable agents (*e.g.*, polyethylene glycol), anionic polymers (*e.g.*, polyacrylate or polymethylacrylate) and anionic saccharidic polymers (*e.g.*, dextran sulfate).
- [0106] Nucleic acid hybridization is adaptable to a variety of assay formats such as the sandwich assay format. The sandwich assay is particularly adaptable to hybridization

under non-denaturing conditions. A primary component of a sandwich-type assay is a solid support. The solid support has adsorbed to it or covalently coupled to it immobilized nucleic acid probe that is unlabeled and complementary to one portion of the sequence.

[0107] Cytosolic expression of acetolactate synthase can be achieved by transforming with a gene comprising a sequence encoding an acetolactate synthase protein, with no mitochondrial targeting signal sequence. Methods for gene expression in yeasts are known in the art (*see, e.g., 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, CA). Expression of genes in yeast typically requires a promoter, operably linked to a coding region of interest, and a transcriptional terminator. A number of yeast promoters can be used in constructing expression cassettes for genes encoding an acetolactate synthase, including, but not limited to constitutive promoters FBA, GPD1, ADH1, and GPM, and the inducible promoters GAL1, GAL10, and CUP1. Other yeast promoters include hybrid promoters UAS(PGK1)-FBA1p (SEQ ID NO: 238), UAS(PGK1)-ENO2p (SEQ ID NO: 239), UAS(FBA1)-PDC1p (SEQ ID NO: 240), UAS(PGK1)-PDC1p (SEQ ID NO: 241), and UAS(PGK)-OLE1p (SEQ ID NO: 242). Suitable transcriptional terminators include, but are not limited to FBA_t, GPD_t, GPM_t, ERG10_t, GAL1_t, CYC1, and ADH1.

[0108] Suitable promoters, transcriptional terminators, and coding regions can be cloned into a yeast 2 micron plasmid and transformed into yeast cells (Ludwig, et al. *Gene*, 132: 33-40, 1993; US Appl. Pub. No. 20080261861A1).

[0109] Suitable promoters, transcriptional terminators, and coding regions can be cloned into *E. coli*-yeast shuttle vectors, and transformed into yeast cells as described in the Examples. These vectors allow strain propagation in both *E. coli* and yeast strains.

[0110] Typically the vector contains a selectable marker and sequences allowing autonomous replication or chromosomal integration in the desired host. Typically used plasmids in yeast are shuttle vectors pRS423, pRS424, pRS425, and pRS426 (American Type Culture Collection, Rockville, MD), which contain an *E. coli* replication origin (e.g., pMB1), a yeast 2 μ origin of replication, and a marker for nutritional selection. The selection markers for these four vectors are His3 (vector pRS423), Trp1 (vector pRS424), Leu2 (vector pRS425) and Ura3 (vector pRS426). Construction of expression vectors

with a chimeric gene encoding a polypeptide can be performed by either standard molecular cloning techniques in *E. coli* or by the gap repair recombination method in yeast.

[0111] The gap repair cloning approach takes advantage of the highly efficient homologous recombination in yeast. Typically, a yeast vector DNA is digested (e.g., in its multiple cloning site) to create a "gap" in its sequence. A number of insert DNAs of interest are generated that contain a ≥ 21 bp sequence at both the 5' and the 3' ends that sequentially overlap with each other, and with the 5' and 3' terminus of the vector DNA. For example, to construct a yeast expression vector for "Gene X," a yeast promoter and a yeast terminator are selected for the expression cassette. The promoter and terminator are amplified from the yeast genomic DNA, and Gene X is either PCR amplified from its source organism or obtained from a cloning vector comprising Gene X sequence. There is at least a 21 bp overlapping sequence between the 5' end of the linearized vector and the promoter sequence, between the promoter and Gene X, between Gene X and the terminator sequence, and between the terminator and the 3' end of the linearized vector. The "gapped" vector and the insert DNAs are then co-transformed into a yeast strain and plated on the medium containing the appropriate compound mixtures that allow complementation of the nutritional selection markers on the plasmids. The presence of correct insert combinations can be confirmed by PCR mapping using plasmid DNA prepared from the selected cells. The plasmid DNA isolated from yeast (usually low in concentration) can then be transformed into an *E. coli* strain, e.g., *TOP10*, followed by mini preps and restriction mapping to further verify the plasmid construct. Finally the construct can be verified by sequence analysis.

[0112] Like the gap repair technique, integration into the yeast genome also takes advantage of the homologous recombination system in yeast. Typically, a cassette containing a coding region plus control elements (promoter and terminator) and auxotrophic marker is PCR-amplified with a high-fidelity DNA polymerase using primers that hybridize to the cassette and contain 40-70 base pairs of sequence homology to the regions 5' and 3' of the genomic area where insertion is desired. The PCR product is then transformed into yeast and plated on medium containing the appropriate compound mixtures that allow selection for the integrated auxotrophic marker. For example, to integrate "Gene X" into chromosomal location "Y", the promoter-coding regionX-

terminator construct is PCR amplified from a plasmid DNA construct and joined to an autotrophic marker (such as *URA3*) by either SOE (splicing by overlap extension) PCR or by common restriction digests and cloning. The full cassette, containing the promoter-coding regionX-terminator-*URA3* region, is PCR amplified with primer sequences that contain 40-70 bp of homology to the regions 5' and 3' of location "Y" on the yeast chromosome. The PCR product is transformed into yeast and selected on growth media lacking uracil. Transformants can be verified either by colony PCR or by direct sequencing of chromosomal DNA. Alternatively, an integration vector can be constructed and propagated in *E. coli*. Elements necessary for chromosomal integration (at least one host-specific targeting sequence and a yeast selectable marker) can be added to any suitable *E. coli* cloning vector. After preparing the vector from the *E. coli* host, it can be linearized by restriction digestion within the host-specific targeting sequence and transformed into yeast. Homologous recombination between the linearized vector and the native targeting sequence will result in integration of the entire vector (Rothstein, R., *Methods in Enzymology*, Vol 194, pp. 281-301). Transformants are obtained by selection for the auxotrophic marker and confirmed by PCR method or direct sequencing.

- [0113] In embodiments, the present invention is directed to a method of producing a recombinant host cell, comprising transforming a host cell with (i) at least one polynucleotide encoding a polypeptide that catalyzes a substrate to product conversion of a pyruvate-utilizing biosynthetic pathway; and (ii) a polynucleotide encoding a peptide that catalyzes the conversion of pyruvate to acetolactate; wherein the polynucleotide of (ii) is integrated into the chromosome.

Biosynthetic Pathways

- [0114] Suitable biosynthetic pathways for production of butanol are known in the art, and certain suitable pathways are described herein. In some embodiments, the butanol, including isobutanol biosynthetic pathway comprises at least one gene that is heterologous to the host cell. In some embodiments, the butanol biosynthetic pathway comprises more than one gene that is heterologous to the host cell. In some embodiments, the butanol biosynthetic pathway comprises heterologous genes encoding polypeptides corresponding to every step of a biosynthetic pathway. As used herein

heterologous refers to both native and non-native genes that have been modified for the purposes herein.

[0115] Products of pyruvate-utilizing biosynthetic pathway can be advantageously produced in a host cell of the invention. A list of such products includes, but is not limited to, 2,3-butanediol, 2-butanone, 2-butanol, and isobutanol. In embodiments, the pyruvate-utilizing biosynthetic pathway comprises one or more polynucleotides encoding a polypeptide that catalyzes a substrate to product conversion of the pathway. In embodiments, the one or more polynucleotides are integrated into a chromosome of the host cell.

[0116] In some embodiments, the invention relates to a recombinant host cell comprising a first heterologous polynucleotide encoding a first polypeptide which catalyzes the conversion of a step of a pyruvate-utilizing biosynthetic pathway; a second heterologous polynucleotide encoding a second polypeptide which catalyzes the conversion of a step of a pyruvate-utilizing biosynthetic pathway; and a third heterologous polynucleotide encoding a third polypeptide which catalyzes the conversion of a step of a pyruvate-utilizing biosynthetic pathway; wherein the first and second heterologous polynucleotides are integrated into the chromosome of the host cell; wherein the third heterologous polynucleotide is not integrated into the chromosome of the host cell; and wherein the first, second, and third polypeptides catalyze different steps of the pyruvate-utilizing biosynthetic pathway.

[0117] In some embodiments, the invention relates to a recombinant host cell comprising (a) a first heterologous polynucleotide encoding a first polypeptide which catalyzes a substrate to product conversion of pyruvate to acetolactate; (b) a second heterologous polynucleotide encoding a second polypeptide which catalyzes the substrate to product conversion of α -ketoisovalerate to isobutyraldehyde; and (c) a third heterologous polynucleotide encoding a third polypeptide which catalyzes the conversion of a step of a isobutanol biosynthetic pathway that is not the conversion of (a) or (b); wherein the first and second heterologous polynucleotides are integrated into the chromosome; wherein the third heterologous polynucleotide is not integrated into the chromosome; and wherein the host cell produces isobutanol.

[0118] In some embodiments, the invention relates to a recombinant host cell comprising (a) a first heterologous polynucleotide encoding a first polypeptide which catalyzes a

substrate to product conversion of α -ketoisovalerate to isobutyraldehyde; and (b) a second heterologous polynucleotide encoding a second polypeptide which catalyzes the conversion of a step of a isobutanol biosynthetic pathway that is not the conversion of (a); wherein the first heterologous polynucleotide is integrated into the chromosome; wherein the second heterologous polynucleotide is not integrated into the chromosome; and wherein the host cell produces isobutanol.

[0119]

[0120] Biosynthetic pathways starting with a step of converting pyruvate to acetolactate for synthesis of isobutanol are disclosed in U.S. Appl. Pub. No. 20070092957, which is herein incorporated by reference. As described in U.S. U.S. Appl. Pub. No. 20070092957, steps in an example isobutanol biosynthetic pathway using acetolactate include conversion of:

- acetolactate to 2,3-dihydroxyisovalerate (FIG. 2 pathway step b) as catalyzed for example by acetohydroxy acid isomeroreductase;
- 2,3-dihydroxyisovalerate to α -ketoisovalerate (FIG. 2 pathway step c) as catalyzed for example by acetohydroxy acid dehydratase;
- α -ketoisovalerate to isobutyraldehyde (FIG. 2 pathway step d) as catalyzed for example by branched-chain α -keto acid decarboxylase ;and
- isobutyraldehyde to isobutanol (FIG. 2 pathway step e) as catalyzed for example by branched-chain alcohol dehydrogenase.

[0121] Genes and polypeptides that can be used for substrate to product conversions described herein as well as methods of identifying such genes and polypeptides, are described herein and/or in the art, for example, for isobutanol, in the Examples and in U.S. Patent No. 7,851,188. Ketol-acid reductoisomerase enzymes are described in U.S. Patent Appl. Pub. Nos. 20080261230 A1, 20090163376 A1, 20100197519 A1, and PCT Appl. Pub. No. WO/2011/041415. Examples of KARIs disclosed therein are those from *Lactococcus lactis*, *Vibrio cholera*, *Pseudomonas aeruginosa* PAO1, as well as *Pseudomonas fluorescens* PF5 mutants. KARIs include *Anaerostipes caccae* KARI variants “K9G9” and “K9D3” (amino acid sequences SEQ ID NOs: 225 and 224, respectively). US Appl. Pub. No. 20100081154 A1, and U.S. Patent 7,851,188 describe dihydroxyacid dehydratases (DHADs), including a DHAD from *Streptococcus mutans*. Suitable polypeptides to catalyze the conversion of α -ketoisovalerate to isobutyraldehyde

include those from *Listeria grayi*, *Lactococcus lactis*, and *Macrococcus caseolyticus* having SEQ ID NOs: 247, 48, and 248, respectively. U.S. Patent Appl. Publ. No. 20090269823 A1 describes SadB, an alcohol dehydrogenase (ADH) from *Achromobacter xylosoxidans*. Alcohol dehydrogenases also include horse liver ADH and *Beijerinckia indica* ADH (protein SEQ ID NO: 237). Each of the above-referenced applications and patents is herein incorporated by reference.

[0122] Also described in U.S. Appl. Pub. No. 20070092957 is the construction of chimeric genes and genetic engineering of yeast, exemplified by *Saccharomyces cerevisiae*, for isobutanol production using the disclosed biosynthetic pathways.

[0123] In some embodiments, the isobutanol biosynthetic pathway comprises at least one gene, at least two genes, at least three genes, or at least four genes that is/are heterologous to the yeast cell. In some embodiments, the recombinant host cell comprises a heterologous gene for each substrate to product conversion of an isobutanol biosynthetic pathway. 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.

[0124] Biosynthetic pathways starting with a step of converting pyruvate to acetolactate for synthesis of 2-butanone and 2-butanol are disclosed in U.S. Appl. Pub. Nos. 20070259410 and 20070292927, which are herein incorporated by reference. A diagram of the disclosed 2-butanone and 2-butanol biosynthetic pathways is provided in FIG. 3. 2-Butanone is the product made when the last depicted step of converting 2-butanone to 2-butanol is omitted. Production of 2-butanone or 2-butanol in a strain disclosed herein benefits from increased availability of acetolactate. As described in U.S. Appl. Pub. No. 20070292927, steps in an example biosynthetic pathway using acetolactate include conversion of:

- acetolactate to acetoin (FIG. 3 step b) as catalyzed for example by acetolactate decarboxylase;
- acetoin to 2,3-butanediol (FIG. 3 step i) as catalyzed for example by butanediol dehydrogenase;
- 2,3-butanediol to 2-butanone (FIG. 3 step j) as catalyzed for example by diol dehydratase or glycerol dehydratase; and

- 2-butanone to 2-butanol (FIG. 3 step f) as catalyzed for example by butanol dehydrogenase.

[0125] Genes that can be used for expression of these enzymes are described in U.S. Appl. Pub. No. 20070292927. The use in this pathway in yeast of the butanediol dehydratase from *Roseburia inulinivorans*, *RdhtA*, (protein SEQ ID NO:32, coding region SEQ ID NO:31) is disclosed in U.S. Appl. Pub. No. 20090155870. This enzyme is used in conjunction with the butanediol dehydratase reactivase from *Roseburia inulinivorans*, *RdhtB*, (protein SEQ ID NO:34, coding region SEQ ID NO:33).

[0126] As described in U.S. Appl. Pub. No. 20070292927, steps in an example biosynthetic pathway using acetolactate include conversion of:

- alpha-acetolactate to acetoin (FIG. 3 step b) as catalyzed for example by acetolactate decarboxylase;
- acetoin to 3-amino-2-butanol (FIG. 3 step c) as catalyzed for example by acetoin aminase;
- 3-amino-2-butanol to 3-amino-2-butanol phosphate (FIG. 3 step d) as catalyzed for example by aminobutanol kinase;
- 3-amino-2-butanol phosphate to 2-butanone (FIG. 3 step e) as catalyzed for example by aminobutanol phosphate phosphor-lyase; and
- 2-butanone to 2-butanol (FIG. 3 step f) as catalyzed for example by butanol dehydrogenase.

[0127] 2-Butanone is the product made when the last depicted step of converting 2-butanone to 2-butanol is omitted. Production of 2-butanone or 2-butanol in a strain disclosed herein benefits from increased availability of acetolactate.

[0128] Useful for the last step of converting 2-butanone to 2-butanol is a new butanol dehydrogenase isolated from an environmental isolate of a bacterium identified as *Achromobacter xylosoxidans* that is disclosed in U.S. Pub. Appl. No. 20090269823 (DNA: SEQ ID NO:35, protein SEQ ID NO:36).

[0129] Also described in U.S. Pub. Appl. No. 20090155870 is the construction of chimeric genes and genetic engineering of yeast for 2-butanol production using the U.S. Appl. Pub. No. 20070292927 disclosed biosynthetic pathway. 2,3-butanediol is an intermediate in this 2-butanol pathway and the steps in its synthesis are also described above.

[0130] In embodiments of the invention, the pyruvate-utilizing biosynthetic pathway forms a product that includes 2,3-butanediol, isobutanol, 2-butanol or 2-butanone. In embodiments, the pyruvate-utilizing biosynthetic pathway is a butanol biosynthetic pathway. In embodiments, the butanol biosynthetic pathway is a 2-butanol biosynthetic pathway or an isobutanol biosynthetic pathway. In embodiments, the host cell comprises at least one polynucleotide encoding a polypeptide that catalyzes a substrate to product conversion of (i) pyruvate to acetolactate; (ii) acetolactate to 2,3-dihydroxyisovalerate; (iii) 2,3-dihydroxyisovalerate to α -ketoisovalerate; (iv) α -ketoisovalerate to isobutyraldehyde; or (v) isobutyraldehyde to isobutanol. In embodiments, the host cell comprises at least one polynucleotide encoding a polypeptide that catalyzes a substrate to product conversion of (i) pyruvate to acetolactate; (ii) acetolactate to 2,3-dihydroxyisovalerate; (iii) 2,3-dihydroxyisovalerate to α -ketoisovalerate; (iv) α -ketoisovalerate to isobutyryl-CoA; (v) isobutyryl-CoA to isobutyraldehyde; or (vi) isobutyraldehyde to isobutanol. In other embodiments, the host cell comprises at least one polynucleotide encoding a polypeptide that catalyzes a substrate to product conversion of (i) pyruvate to acetolactate; (ii) acetolactate to 2,3-dihydroxyisovalerate; (iii) 2,3-dihydroxyisovalerate to α -ketoisovalerate; (iv) α -ketoisovalerate to valine; (v) valine to isobutylamine; (vi) isobutylamine to isobutyraldehyde; or (vii) isobutyraldehyde to isobutanol.

[0131] In embodiments, the host cell comprises at least one polynucleotide encoding a polypeptide that catalyzes a substrate to product conversion of (i) pyruvate to acetolactate; (ii) acetolactate to acetoin; (iii) acetoin to 2,3-butanediol; or (iv) 2,3-butanediol to 2-butanone. In embodiments, the host cell comprises at least one polynucleotide encoding a polypeptide that catalyzes a substrate to product conversion of (i) pyruvate to acetolactate; (ii) acetolactate to acetoin; (iii) acetoin to 2,3-butanediol; (iv) 2,3-butanediol to 2-butanone; or (v) 2-butanone to 2-butanol.

[0132] In embodiments, the recombinant host cell comprises (a) a heterologous polynucleotide encoding a polypeptide which catalyzes the substrate to product conversion of pyruvate to acetolactate, wherein the polynucleotide is integrated into the chromosome; (b) a heterologous polynucleotide encoding a polypeptide which catalyzes the substrate to product conversion of acetolactate to 2,3-dihydroxyisovalerate; (c) a heterologous polynucleotide encoding a polypeptide which catalyzes the substrate to

product conversion of 2,3-dihydroxyisovalerate to α -ketoisovalerate; and (d) a heterologous polynucleotide encoding a polypeptide which catalyzes the substrate to product conversion of α -ketoisovalerate to isobutyraldehyde, wherein the host cell is substantially free of pyruvate decarboxylase activity; and wherein the host cell produces isobutanol.

[0133] In embodiments, the polypeptide which catalyzes the conversion of acetolactate to 2,3-dihydroxyisovalerate corresponds to the Enzyme Commission Number EC 1.1.1.86. In embodiments, the polypeptide which catalyzes the conversion of 2,3-dihydroxyisovalerate to α -ketoisovalerate corresponds to the Enzyme Commission Number EC 4.2.1.9. In embodiments, the polypeptide which catalyzes the conversion of α -ketoisovalerate to isobutyraldehyde corresponds to the Enzyme Commission Number EC 4.1.1.72 or 4.1.1.1. In other embodiments, the polypeptide which catalyzes the conversion of isobutyraldehyde to isobutanol corresponds to the Enzyme Commission Number EC 1.1.1.265, 1.1.1.1 or 1.1.1.2.

[0134] In other embodiments of the invention, one or more of the polynucleotides encoding a polypeptide which catalyzes the conversion of any of the biosynthetic pathway steps described herein are on a plasmid. In embodiments, one or more polynucleotides encoding a polypeptide which catalyzes the conversion of any of the biosynthetic pathway steps described herein are integrated into the chromosome at the PDC1-TRX1 intergenic region.

[0135] In other embodiments, the host cells of the invention can have reduced or substantially eliminated expression of a polypeptide which catalyzes the conversion of glycerol-3-phosphate into dihydroxyacetone phosphate. In embodiments, the polypeptide which catalyzes the conversion of glycerol-3-phosphate into dihydroxyacetone phosphate is glycerol-3-phosphate dehydrogenase (GPD). In embodiments, the host cell comprises a deletion, mutation, and/or substitution in an endogenous polynucleotide encoding a polypeptide which catalyzes the conversion of glycerol-3-phosphate into dihydroxyacetone phosphate. In embodiments, the polypeptide which catalyzes the conversion of glycerol-3-phosphate into dihydroxyacetone phosphate corresponds to Enzyme Commission Number 1.1.1.8. In embodiments, the polynucleotide encoding a polypeptide which catalyzes the conversion of glycerol-3-phosphate into dihydroxyacetone phosphate is GPD1 or GPD2. In embodiments, the polynucleotide

encoding a polypeptide which catalyzes the conversion of glycerol-3-phosphate into dihydroxyacetone phosphate comprises a GPD sequence of Table 2. Such modifications and others to host cells are described in US Application Publication No. 20090305363.

[0136] In other embodiments, the host cells of the invention can have reduced or substantially eliminated expression of an iron regulatory protein. In embodiments, the host cells of the invention can have reduced or substantially eliminated expression of a polypeptide affecting iron-sulfur (Fe-S) cluster biosynthesis. In embodiments, recombinant host cells 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 endogenous gene encoding a polypeptide affecting Fe-S cluster biosynthesis; and/or (ii) at least one heterologous polynucleotide encoding a polypeptide affecting Fe-S cluster biosynthesis. In embodiments, the polypeptide affecting Fe-S cluster biosynthesis is encoded by *AFT1* (nucleic acid SEQ ID NO: 227, amino acid SEQ ID NO: 228), *AFT2* (SEQ ID NOs: 229 and 230), *FRA2* (SEQ ID NOs: 231 and 232), *GRX3* (SEQ ID NOs: 233 and 234), or *CCCI* (SEQ ID NOs: 235 and 236). In embodiments, the polypeptide affecting Fe-S cluster biosynthesis is constitutive mutant *AFT1* L99A, *AFT1* L102A, *AFT1* C291F, or *AFT1* C293F. In embodiments, the polypeptide affecting Fe-S cluster biosynthesis is selected from AFT1, AFT2, PSE1, FRA2, GRX3, MSN5, or combinations thereof. In embodiments, the host cell comprises a deletion, mutation, and/or substitution in an endogenous polynucleotide encoding an iron regulatory protein. In embodiments, the host cell comprises a deletion, mutation, and/or substitution in an endogenous polynucleotide encoding a polypeptide which affects Fe-S cluster biosynthesis. In embodiments, the polynucleotide encoding a polypeptide which affects Fe-S cluster biosynthesis comprises a sequence as disclosed in WIPO Appl. Pub. No. WO/2011/103300.

It will be appreciated that host cells comprising a butanol biosynthetic pathway such as 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. 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 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 include at least one deletion, mutation, and/or substitution in an endogenous polynucleotide encoding a polypeptide having acetolactate reductase activity. In embodiments, the polypeptide having acetolactate reductase activity is YMR226C (SEQ ID NO: 226) of *Saccharomyces cerevisiae* or a homolog thereof. Additional modifications include a deletion, mutation, and/or substitution in an endogenous polynucleotide encoding a polypeptide having aldehyde dehydrogenase and/or aldehyde oxidase activity. In embodiments, the polypeptide having aldehyde dehydrogenase activity is ALD6 (SEQ ID NO: 223) from *Saccharomyces cerevisiae* or a homolog thereof. A genetic modification which has the effect of reducing glucose repression wherein the yeast production host cell is *pdc-* is described in U.S. Appl. Publication No. 20110124060, incorporated herein by reference. Additionally, host cells may comprise heterologous polynucleotides encoding a polypeptide with phosphoketolase activity and/or a heterologous polynucleotide encoding a polypeptide with phosphotransacetylase activity as described in U.S. Appn. Serial No. 13/161,168, filed on June 15, 2011, incorporated herein by reference. Reduced pyruvate decarboxylase activity

[0137] Endogenous pyruvate decarboxylase activity in microbial cells converts pyruvate to acetaldehyde, which is then converted to ethanol or to acetyl-CoA via acetate (*see* FIG. 1). Microbial cells can have one or more genes encoding pyruvate decarboxylase. For example, in yeast there is one gene encoding pyruvate decarboxylase in *Kluyveromyces lactis*, while there are three isozymes of pyruvate decarboxylase encoded by the *PDC1*, *PDC5*, and *PDC6* genes in *Saccharomyces cerevisiae*, as well as a pyruvate decarboxylase regulatory gene *PDC2*. Expression of pyruvate decarboxylase from *PDC6* is minimal. In embodiments of the invention, host cells can have pyruvate decarboxylase activity that is reduced by disrupting at least one gene encoding a pyruvate decarboxylase, or a gene regulating pyruvate decarboxylase gene expression. For example, in *S. cerevisiae* the *PDC1* and *PDC5* genes, or all three genes, are disrupted. In addition, pyruvate decarboxylase activity can be reduced by disrupting the *PDC2* regulatory gene in *S. cerevisiae*. In other yeasts, genes encoding pyruvate decarboxylase proteins such as

those having at least about 80-85%, 85%-90%, 90%-95%, or at least about 98% sequence identity to *PDC1* or *PDC5* can be disrupted.

[0138] Examples of yeast strains with reduced pyruvate decarboxylase activity due to disruption of pyruvate decarboxylase encoding genes have been reported such as for *Saccharomyces* in Flikweert *et al.* (*Yeast*, 12:247-257, 1996), for *Kluyveromyces* in Bianchi *et al.* (*Mol. Microbiol.*, 19(1):27-36, 1996), and disruption of the regulatory gene in Hohmann (*Mol Gen Genet.*, 241:657-666, 1993). *Saccharomyces* strains having no pyruvate decarboxylase activity are available from the ATCC with Accession #200027 and #200028.

[0139] Expression of pyruvate decarboxylase genes can be reduced in any host cell that is also engineered with acetolactate synthase expression and other biosynthetic pathway enzyme encoding genes for production of a compound derived from acetolactate. Examples of yeast pyruvate decarboxylase genes that may be targeted for disruption are listed in Table 2 (SEQ ID NOs: 50, 52, 54, 56, 58, 60, 62, 64 and 66). Other target genes, such as those encoding pyruvate decarboxylase proteins having at least about 80-85%, 85%-90%, 90%-95%, or at least about 98% or 99% sequence identity to the pyruvate decarboxylases listed in Table 2 (SEQ ID NOs: 51, 53, 55, 57, 59, 61, 63, 65 and 67) can be identified in the literature and in bioinformatics databases well known to the skilled person. Additionally, the sequences described herein or those recited in the art can be used to identify homologs in other yeast strains, as described above for identification of acetolactate synthase encoding genes.

[0140] Alternatively, because pyruvate decarboxylase encoding sequences are well known, and because sequencing of the genomes of yeasts is prevalent, suitable pyruvate decarboxylase gene targets can be identified on the basis of sequence similarity using bioinformatics approaches. Genomes have been completely sequenced and annotated and are publicly available for the following yeast strains: *Ashbya gossypii* ATCC 10895, *Candida glabrata* CBS 138, *Kluyveromyces lactis* NRRL Y-1140, *Pichia stipitis* CBS 6054, *Saccharomyces cerevisiae* S288c, *Schizosaccharomyces pombe* 972h-, and *Yarrowia lipolytica* CLIB122. Typically BLAST (described above) searching of publicly available databases with known pyruvate decarboxylase encoding sequences or pyruvate decarboxylase amino acid sequences, such as those provided herein, is used to identify pyruvate decarboxylase encoding sequences of other yeasts.

- [0141] Accordingly it is within the scope of the invention to provide pyruvate decarboxylase proteins having at least about 70-75%, 75%-80%, 80-85%, 85%- 90%, 90%- 95% or at least about 98% or 99% sequence identity to any of the pyruvate decarboxylase proteins disclosed herein (SEQ ID NOs:51, 53, 55, 57, 59, 61, 63, 65 and 67). Identities are based on the Clustal W method of alignment using the default parameters of GAP PENALTY=10, GAP LENGTH PENALTY=0.1, and Gonnet 250 series of protein weight matrix.
- [0142] In embodiments, the host cell of the invention can have expression of pyruvate decarboxylase, glycerol-3-phosphate dehydrogenase, an iron regulatory protein, and/or a polypeptide affecting iron-sulfur (Fe-S) cluster biosynthesis that is decreased or substantially eliminated. In other embodiments, the host cell comprises a deletion, mutation, and/or substitution in an endogenous polynucleotide encoding a polypeptide having the activity of pyruvate decarboxylase, glycerol-3-phosphate dehydrogenase, an iron regulatory protein, or a polypeptide affecting Fe-S cluster biosynthesis.
- [0143] Genes encoding pyruvate decarboxylase, glycerol-3-phosphate dehydrogenase, an iron regulatory protein, or a polypeptide affecting Fe-S cluster biosynthesis can be disrupted in any host cell using genetic modification. Many methods for genetic modification of target genes are known to one skilled in the art and can be used to create the present yeast strains. Modifications that can be used to reduce or eliminate expression of a target protein are disruptions that include, but are not limited to, deletion of the entire gene or a portion of the gene, inserting a DNA fragment into the gene (in either the promoter or coding region) so that the protein is not expressed or expressed at lower levels, introducing a mutation into the coding region which adds a stop codon or frame shift such that a functional protein is not expressed, and introducing one or more mutations into the coding region to alter amino acids so that a non-functional or a less enzymatically active protein is expressed. In addition, expression of a gene can be blocked by expression of an antisense RNA or an interfering RNA, and constructs can be introduced that result in cosuppression. Moreover, a gene can be synthesized whose expression is low because rare codons are substituted for plentiful ones, and this gene substituted for the endogenous gene. Such a gene will produce the same polypeptide but at a lower rate. In addition, the synthesis or stability of the transcript may be lessened by mutation. Similarly the efficiency by which a protein is translated from mRNA may be

modulated by mutation. All of these methods can be readily practiced by one skilled in the art making use of the known or identified gene sequences.

[0144] DNA sequences surrounding a coding sequence are also useful in some modification procedures and are available for yeasts such as for *Saccharomyces cerevisiae* in the complete genome sequence coordinated by Genome Project ID9518 of Genome Projects coordinated by NCBI (National Center for Biotechnology Information) with identifying GOPID #13838. Additional examples of yeast genomic sequences include that of *Yarrowia lipolytica*, GOPIC #13837, and of *Candida albicans*, which is included in GPID #10771, #10701 and #16373. Other yeast genomic sequences can be readily found by one of skill in the art in publicly available databases.

[0145] In particular, DNA sequences surrounding a gene coding sequence are useful for modification methods using homologous recombination. For example, in this method gene flanking sequences are placed bounding a selectable marker gene to mediate homologous recombination whereby the marker gene replaces the target gene. Also partial gene sequences and gene flanking sequences bounding a selectable marker gene may be used to mediate homologous recombination whereby the marker gene replaces a portion of the target gene. In addition, the selectable marker may be bounded by site-specific recombination sites, so that following expression of the corresponding site-specific recombinase, the resistance gene is excised from the target gene locus without reactivating the latter. The site-specific recombination leaves behind a recombination site which disrupts expression of the protein. A homologous recombination vector can be constructed to also leave a deletion in the target gene following excision of the selectable marker, as is well known to one skilled in the art.

[0146] Deletions can be made using mitotic recombination as described in Wach *et al.* (*Yeast*, 10:1793-1808, 1994). This method involves preparing a DNA fragment that contains a selectable marker between genomic regions that can be as short as 20 bp, and which bound a target DNA sequence. This DNA fragment can be prepared by PCR amplification of the selectable marker gene using as primers oligonucleotides that hybridize to the ends of the marker gene and that include the genomic regions that can recombine with the yeast genome. The linear DNA fragment can be efficiently transformed into yeast and recombined into the genome resulting in gene replacement

including with deletion of the target DNA sequence (as described in "Methods in Enzymology," v194, pp 281-301, 1991).

[0147] In addition, the activity of pyruvate decarboxylase, glycerol-3-phosphate dehydrogenase, an iron regulatory protein, or a polypeptide affecting Fe-S cluster biosynthesis in any host cell of the invention can be disrupted using random mutagenesis, which is followed by screening to identify strains with reduced pyruvate decarboxylase activity. Using this type of method, the DNA sequence of the pyruvate decarboxylase encoding region, or any other region of the genome affecting expression of these activities, need not be known.

[0148] Methods for creating genetic mutations are common and well known in the art and may be applied to the exercise of creating mutants. Commonly used random genetic modification methods (reviewed in *Methods in Yeast Genetics*, 2005, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY) include spontaneous mutagenesis, mutagenesis caused by mutator genes, chemical mutagenesis, irradiation with UV or X-rays, or transposon mutagenesis.

[0149] Chemical mutagenesis of yeast commonly involves treatment of cells with one of the following DNA mutagens: ethyl methanesulfonate (EMS), nitrous acid, diethyl sulfate, or N-methyl-N'-nitro-N-nitroso-guanidine (MNNG). These methods of mutagenesis have been reviewed in Spencer et al (Mutagenesis in Yeast, 1996, *Yeast Protocols: Methods in Cell and Molecular Biology*. Humana Press, Totowa, NJ). Chemical mutagenesis with EMS may be performed as described in *Methods in Yeast Genetics*, 2005, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY. Irradiation with ultraviolet (UV) light or X-rays can also be used to produce random mutagenesis in yeast cells. The primary effect of mutagenesis by UV irradiation is the formation of pyrimidine dimers which disrupt the fidelity of DNA replication. Protocols for UV-mutagenesis of yeast can be found in Spencer *et al.* (Mutagenesis in Yeast, 1996, *Yeast Protocols: Methods in Cell and Molecular Biology*. Humana Press, Totowa, NJ). Introduction of a mutator phenotype can also be used to generate random chromosomal mutations in yeast. Common mutator phenotypes can be obtained through disruption of one or more of the following genes: *PMS1*, *MAG1*, *RAD18* or *RAD51*. Restoration of the non-mutator phenotype can be easily obtained by insertion of the wildtype allele. Collections of modified cells produced from any of these or other known random

mutagenesis processes may be screened for reduced activity of pyruvate decarboxylase, glycerol-3-phosphate decarboxylase, an iron regulatory protein or a polypeptide affecting Fe-S cluster biosynthesis.

Host cells

[0150] The host cells of the invention can be any cell amenable to genetic manipulation. In embodiments, the host cell can be a bacterium, a cyanobacterium, a filamentous fungus, or a yeast. In embodiments, the host cell is a member of the genus *Clostridium*, *Zymomonas*, *Escherichia*, *Salmonella*, *Rhodococcus*, *Pseudomonas*, *Bacillus*, *Lactobacillus*, *Enterococcus*, *Alcaligenes*, *Klebsiella*, *Paenibacillus*, *Arthrobacter*, *Corynebacterium*, *Brevibacterium*, *Pichia*, *Candida*, *Hansenula*, *Kluyveromyces*, or *Saccharomyces*. In other embodiments, the host cell is *Escherichia coli*, *Alcaligenes eutrophus*, *Bacillus licheniformis*, *Paenibacillus macerans*, *Rhodococcus erythropolis*, *Pseudomonas putida*, *Bacillus subtilis*, *Lactobacillus plantarum*, *Enterococcus faecium*, *Enterococcus gallinarum*, or *Enterococcus faecalis*.

[0151] Examples of a yeast include, but are not limited to, *Saccharomyces*, *Schizosaccharomyces*, *Hansenula*, *Issatchenkia*, *Candida*, *Kluyveromyces*, *Yarrowia* and *Pichia*. Examples of yeast strains include, but are not limited to, *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Kluyveromyces lactis*, *Kluyveromyces thermotolerans*, *Candida glabrata*, *Candida albicans*, *Pichia stipitis* and *Yarrowia lipolytica*. 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™ yeast, Bio-Ferm® XR yeast, Gert Strand Prestige Batch Turbo alcohol yeast, Gert Strand Pot Distillers yeast, Gert Strand Distillers Turbo yeast, FerMax™ Green yeast, FerMax™ Gold yeast, Thermosacc® yeast, BG-1, PE-2, CAT-1, CBS7959, CBS7960, and CBS7961.

[0152] In embodiments, the host cell of the invention is a facultative anaerobe. In embodiments, a cell used as a production host preferably has enhanced tolerance to the

produced chemical, and/or can have a high rate of carbohydrate utilization. These characteristics can be conferred by mutagenesis and selection, genetic engineering, or can be natural.

Fermentation Media

[0153] A host cell of the invention can be grown in fermentation media that can contain suitable carbon substrates. Suitable substrates can include, but are not limited to, monosaccharides such as glucose and fructose, oligosaccharides such as lactose or 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. Additionally, the carbon substrate can 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 can utilize a number of other carbon containing compounds such as methylamine, glucosamine and a variety of amino acids for metabolic activity. For example, methylotrophic yeast are known to utilize the carbon from methylamine to form trehalose or glycerol (Bellion *et al.*, *Microb. Growth C1 Compd.*, [Int. Symp.], 7th (1993), 415-32. Editor(s): Murrell, J. Collin; Kelly, Don P. Publisher: Intercept, Andover, UK). Similarly, various species of *Candida* can metabolize alanine or oleic acid (Sulter *et al.*, *Arch. Microbiol.* 153:485-489 (1990)). Hence, it is contemplated that the source of carbon utilized in the present invention can encompass a wide variety of carbon containing substrates.

[0154] In addition to a carbon source, fermentation media can 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 the enzymatic pathway necessary for production of the desired product.

Culture Conditions

[0155] Typically host cells of the invention are grown at a temperature in the range of about 20 °C to about 37 °C in an appropriate medium. Suitable growth media in the present invention are common commercially prepared media such as 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 can 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.

[0156] Suitable pH ranges for the fermentation can be between pH 3.0 to pH 7.5. A pH range of pH 4.5.0 to pH 6.5 can be used in an initial condition.

[0157] Fermentations can be performed under aerobic or anaerobic conditions, where anaerobic or microaerobic conditions are preferred.

[0158] The amount of butanol produced in the fermentation medium can be determined using a number of methods known in the art, for example, high performance liquid chromatography (HPLC) or gas chromatography (GC).

Industrial Batch and Continuous Fermentations

[0159] Methods of the present invention can employ 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. Thus, at the beginning of the fermentation the medium is inoculated with the desired organism or organisms, and fermentation is permitted to occur without adding anything to the system. Typically, however, a "batch" fermentation is batch with respect to the addition of carbon source and attempts are often made at controlling factors such as pH and oxygen concentration. In batch systems, the metabolite and biomass compositions of the system change constantly up to the time the fermentation is stopped. Within batch cultures cells moderate through a static lag phase to a high growth log phase and finally to a stationary phase where growth rate is diminished or halted. If untreated, cells in the stationary phase will eventually die. Cells in log phase generally are responsible for the bulk of production of end product or intermediate.

[0160] 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. Measurement of the actual substrate concentration in Fed-Batch systems is difficult and is therefore estimated on the basis of the changes of measurable factors such as pH, dissolved oxygen and the partial pressure of waste gases such as CO₂. 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, Inc., Sunderland, MA., or Deshpande, Mukund V., *Appl. Biochem. Biotechnol.*, 36:227, (1992), herein incorporated by reference.

[0161] Additionally, the methods of the present invention can be adaptable to 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.

[0162] Continuous fermentation allows for the modulation of one factor or any number of factors that affect cell growth or end product concentration. For example, one method will maintain a limiting nutrient such as the carbon source or nitrogen level at a fixed rate and allow all other parameters to moderate. In other systems a number of factors affecting growth can be altered continuously while the cell concentration, measured by media turbidity, is kept constant. Continuous systems strive to maintain steady state growth conditions and thus the cell loss due to the medium being drawn off must be balanced against the cell growth rate in the fermentation. Methods of modulating nutrients and growth factors for continuous fermentation processes as well as techniques for maximizing the rate of product formation are well known in the art of industrial microbiology and a variety of methods are detailed by Brock, *supra*.

[0163] It is contemplated that the present invention can be practiced using either batch, fed-batch or 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 and subjected to fermentation conditions for butanol production.

Methods for Product Isolation from the Fermentation Medium

- [0164] Products of the biosynthetic pathways of the invention (*e.g.*, butanol) can be isolated from the fermentation medium using methods known in the art. For example, solids can be removed from the fermentation medium by extraction, centrifugation, filtration, decantation, or the like. Then, the product (*e.g.*, butanol) can be isolated from the fermentation medium, which has been treated to remove solids as described above, using methods such as distillation, liquid-liquid extraction, or membrane-based separation. Because butanol forms a low boiling point, azeotropic mixture with water, distillation can only be used to separate the mixture up to its azeotropic composition. Distillation can be used in combination with another separation method to obtain separation around the azeotrope. Methods that can be used in combination with distillation to isolate and purify butanol include, but are not limited to, decantation, liquid-liquid extraction, adsorption, and membrane-based techniques. Additionally, butanol can be isolated using azeotropic distillation using an entrainer (see for example Doherty and Malone, *Conceptual Design of Distillation Systems*, McGraw Hill, New York, 2001).
- [0165] The butanol-water mixture forms a heterogeneous azeotrope so that distillation can be used in combination with decantation to isolate and purify the butanol. In this method, the butanol containing fermentation broth is distilled to near the azeotropic composition. Then, the azeotropic mixture is condensed, and the butanol is separated from the fermentation medium by decantation. The decanted aqueous phase can be returned to the first distillation column as reflux. The butanol-rich decanted organic phase can be further purified by distillation in a second distillation column.
- [0166] The products such as butanol can also be isolated from the fermentation medium using liquid-liquid extraction in combination with distillation. In this method, the butanol is extracted from the fermentation broth using liquid-liquid extraction with a suitable solvent. The butanol-containing organic phase is then distilled to separate the butanol from the solvent.
- [0167] Distillation in combination with adsorption can also be used to isolate butanol from a fermentation medium. In this method, the fermentation broth containing the butanol 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).

- [0168] Additionally, distillation in combination with pervaporation can be used to isolate and purify a product such as butanol from the fermentation medium. In this method, the fermentation broth containing the butanol 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).
- [0169] Methods for producing and recovering a product such as butanol from a fermentation broth using extractive fermentation are described in detail in U.S. Patent Appl. No. 12/478,389 filed on June 4, 2009 and corresponding published U.S. Appn. Publ. No. 20090305370, U.S. Provisional Appl. No. 61/231,699 filed on August 6, 2009, U.S. Provisional Appl. No. 61/368,429 filed on July 28, 2010, and U.S. Appn. Publ. Nos. 20100221802 and 20110097773. Such methods include those which comprise the step of contacting the fermentation broth with a water immiscible organic extractant selected from the group consisting of C₁₂ to C₂₂ fatty alcohols, C₁₂ to C₂₂ fatty acids, esters of C₁₂ to C₂₂ fatty acids, C₁₂ to C₂₂ fatty amides, C₁₂ to C₂₂ fatty aldehydes, and mixtures thereof, to form a two-phase mixture comprising an aqueous phase and a butanol-containing organic phase. "Contacting" means the fermentation medium and the organic extractant are brought into physical contact at any time during the fermentation process.
- [0170] Examples of suitable extractants include, but are not limited to, an extractant comprising at least one solvent selected from the group consisting of oleyl alcohol, behenyl alcohol, cetyl alcohol, lauryl alcohol, myristyl alcohol, stearyl alcohol, oleic acid, lauric acid, myristic acid, stearic acid, methyl myristate, methyl oleate, lauric aldehyde, 1-nonanol, 1-decanol, 1-undecanol, 2-undecanol, 1-nonanal, and mixtures thereof. In one embodiment, the extractant comprises oleyl alcohol. These organic extractants are available commercially from various sources, such as Sigma-Aldrich (St. Louis, MO), in various grades, many of which are suitable for use in extractive fermentation to produce or recover butanol. Technical grades contain a mixture of compounds, including the desired component and higher and lower fatty components. For example, one commercially available technical grade oleyl alcohol contains about 65% oleyl alcohol and a mixture of higher and lower fatty alcohols.

[0171] In embodiments, the present invention is directed to a method of producing butanol, comprising (a) providing a recombinant host cell of the invention; and (b) contacting the host cell with a fermentable carbon substrate to form a fermentation broth under conditions whereby butanol is produced. In other embodiments, the method further comprises contacting the fermentation broth with an extractant to produce a two-phase fermentation mixture. In other embodiments, the extractant comprises fatty acids. In other embodiments, the fatty acids are derived from corn oil or soybean oil. In other embodiments, the extractant comprises a water immiscible organic extractant selected from the group consisting of: C₁₂ to C₂₂ fatty alcohols, C₁₂ to C₂₂ fatty acids, esters of C₁₂ to C₂₂ fatty acids, C₁₂ to C₂₂ fatty aldehydes, C₁₂ to C₂₂ fatty amides. In other embodiments, the method further comprises contacting the fermentation broth with an organic acid and an enzyme capable of esterifying the butanol with the organic acid. In embodiments, the method further comprises vaporizing at least a portion of the fermentation broth to form a vapor stream comprising water and butanol.

[0172] Methods for measuring butanol titer and production are known. For example, butanol titer and production can be measured using gas chromatography (GC) or high performance liquid chromatography (HPLC) as described in the examples. In embodiments, the amount of butanol produced by a host cell of the invention is increased by at least about 10%, at least about 20%, at least about 30%, at least about 40%, at least about 50%, at least about 60%, at least about 70%, at least about 80%, at least about 90%, at least about 95%, at least about 2-fold, at least about 3-fold, or at least about 4-fold greater as compared to the amount of butanol produced by a host cell that does not comprise a polynucleotide encoding a polypeptide that catalyzes the conversion of pyruvate to acetolactate integrated into the chromosome. In embodiments, the titer of butanol produced by a host cell of the invention is increased by at least about 10%, at least about 20%, at least about 30%, at least about 40%, at least about 50%, at least about 60%, at least about 70%, at least about 80%, at least about 90%, at least about 95%, at least about 2-fold, at least about 3-fold, or at least about 4-fold greater as compared to a recombinant host cell wherein the polynucleotide encoding a polypeptide that catalyzes the conversion of pyruvate to acetolactate is not integrated into the chromosome.

[0173] In embodiments, the present invention is directed to a method for increasing the copy number or expression of a non-integrated recombinant polynucleotide encoding a

polypeptide that catalyzes a step of a biosynthetic pathway described herein, comprising contacting a host cell of the invention with a fermentable carbon substrate to form a fermentation broth under conditions whereby the product of the biosynthetic pathway is produced, such as the fermentation conditions described herein. In other embodiments, the present invention is directed to a method for increasing the flux in a pyruvate-utilizing biosynthetic pathway, comprising contacting a host cell of the invention with a fermentable carbon substrate to form a fermentation broth under conditions whereby the flux in the pyruvate-utilizing biosynthetic pathway in the host cell is increased, such as the fermentation conditions described herein.

[0174] In other embodiments, the invention is directed to a method of increasing the formation of a product of a pyruvate-utilizing biosynthetic pathway comprising (i) providing a recombinant host cell of the invention; and (ii) growing the host cell under conditions wherein the product of the pyruvate-utilizing pathway is formed, wherein the amount of product formed by the recombinant host cell is greater than the amount of product formed by a host cell that does not comprise a polynucleotide encoding a polypeptide which catalyzes the conversion of pyruvate to acetolactate integrated into the chromosome. In other embodiments, the pyruvate-utilizing biosynthetic pathway forms 2,3-butanediol, isobutanol, 2-butanol or 2-butanone. In other embodiments, the pyruvate-utilizing biosynthetic pathway is a butanol biosynthetic pathway. In other embodiments, the butanol biosynthetic pathway is (a) a 2-butanol biosynthetic pathway; or (b) an isobutanol biosynthetic pathway.

[0175] In other embodiments, the invention is directed to a composition comprising (i) a host cell of the invention; (ii) butanol; and (iii) an extractant. In other embodiments, the invention is directed to a composition comprising (i) a host cell of the invention; (ii) butanol; (iii) an extractant; and (iv) an esterification enzyme. An esterification enzyme is one that catalyzes the reaction between an acid and an alcohol to generate an ester. In the broadest sense esterification enzymes are hydrolases that act on an ester linkage and often referred to as esterases. As used herein lipases, are a subclass of esterases shown to be effective in forming esters between the fatty acids and isobutanol present in the broth. Such lipases may include one or more esterase enzymes, for example, hydrolase enzymes such as lipase enzymes. Lipase enzymes used may be derived from any source, including, for example, *Absidia*, *Achromobacter*, *Aeromonas*, *Alcaligenes*, *Alternaria*,

Aspergillus, *Achromobacter*, *Aureobasidium*, *Bacillus*, *Beauveria*, *Brochothrix*, *Candida*, *Chromobacter*, *Coprinus*, *Fusarium*, *Geotrichum*, *Hansenula*, *Humicola*, *Hyphozyma*, *Lactobacillus*, *Metarhizium*, *Mucor*, *Nectria*, *Neurospora*, *Paecilomyces*, *Penicillium*, *Pseudomonas*, *Rhizoctonia*, *Rhizomucor*, *Rhizopus*, *Rhodospiridium*, *Rhodotorula*, *Saccharomyces*, *Sus*, *Sporobolomyces*, *Thermomyces*, *Thiarosporella*, *Trichoderma*, *Verticillium*, and/or a strain of *Yarrowia*. In a preferred aspect, the source of the lipase is selected from the group consisting of *Absidia blakesleena*, *Absidia corymbifera*, *Achromobacter iophagus*, *Alcaligenes* sp., *Alternaria brassiciola*, *Aspergillus flavus*, *Aspergillus niger*, *Aureobasidium pullulans*, *Bacillus pumilus*, *Bacillus strearothermophilus*, *Bacillus subtilis*, *Brochothrix thermosohata*, *Candida cylindracea* (*Candida rugosa*), *Candida paralipolytica*, *Candida Antarctica* lipase A, *Candida antartica* lipase B, *Candida ernobii*, *Candida deformans*, *Chromobacter viscosum*, *Coprinus cinerius*, *Fusarium oxysporum*, *Fusarium solani*, *Fusarium solani pisi*, *Fusarium roseum culmorum*, *Geotrichum penicillatum*, *Hansenula anomala*, *Humicola brevispora*, *Humicola brevis* var. *thermoidea*, *Humicola insolens*, *Lactobacillus curvatus*, *Rhizopus oryzae*, *Penicillium cyclopium*, *Penicillium crustosum*, *Penicillium expansum*, *Penicillium* sp. I, *Penicillium* sp. II, *Pseudomonas aeruginosa*, *Pseudomonas alcaligenes*, *Pseudomonas cepacia* (syn. *Burkholderia cepacia*), *Pseudomonas fluorescens*, *Pseudomonas fragi*, *Pseudomonas maltophilia*, *Pseudomonas mendocina*, *Pseudomonas mephitica* lipolytica, *Pseudomonas alcaligenes*, *Pseudomonas plantari*, *Pseudomonas pseudoalcaligenes*, *Pseudomonas putida*, *Pseudomonas stutzeri*, and *Pseudomonas wisconsinensis*, *Rhizoctonia solani*, *Rhizomucor miehei*, *Rhizopus japonicus*, *Rhizopus microsporus*, *Rhizopus nodosus*, *Rhodospiridium toruloides*, *Rhodotorula glutinis*, *Saccharomyces cerevisiae*, *Sporobolomyces shibatanus*, *Sus scrofa*, *Thermomyces lanuginosus* (formerly *Humicola lanuginosa*), *Thiarosporella phaseolina*, *Trichoderma harzianum*, *Trichoderma reesei*, and *Yarrowia lipolytica*. In a further preferred aspect, the lipase is selected from the group consisting of *Thermomyces lanuginosus*, *Aspergillus* sp. lipase, *Aspergillus niger* lipase, *Candida antartica* lipase B, *Pseudomonas* sp. lipase, *Penicillium roqueforti* lipase, *Penicillium camembertii* lipase, *Mucor javanicus* lipase, *Burkholderia cepacia* lipase, *Alcaligenes* sp. lipase, *Candida rugosa* lipase, *Candida parapsilosis* lipase, *Candida deformans* lipase, lipases A and B from *Geotrichum candidum*, *Neurospora crassa* lipase, *Nectria haematococca* lipase,

Fusarium heterosporum lipase *Rhizopus delemar* lipase, *Rhizomucor miehei* lipase, *Rhizopus arrhizus* lipase, and *Rhizopus oryzae* lipase. Suitable commercial lipase preparations suitable as enzyme catalyst 42 include, but are not limited to Lipolase® 100 L, Lipex® 100L, Lipoclean® 2000T, Lipozyme® CALB L, Novozym® CALA L, and Palatase 20000L, available from Novozymes, or from *Pseudomonas fluorescens*, *Pseudomonas cepacia*, *Mucor miehei*, *hog pancreas*, *Candida cylindracea*, *Rhizopus niveus*, *Candida antarctica*, *Rhizopus arrhizus* or *Aspergillus* available from SigmaAldrich.

[0176] In embodiments, the extractant comprises fatty acids. In embodiments, the fatty acids are derived from corn oil or soybean oil. In other embodiments, the extractant is a water immiscible organic extractant. In other embodiments, the extractant is C₁₂ to C₂₂ fatty alcohols, C₁₂ to C₂₂ fatty acids, esters of C₁₂ to C₂₂ fatty acids, or C₁₂ to C₂₂ fatty aldehydes.

EXAMPLES

[0177] The present invention is further defined in the following Examples. It should be understood that these Examples, while indicating preferred embodiments of the invention, are given by way of illustration only. From the above discussion and these Examples, one skilled in the art can ascertain the essential characteristics of this invention, and without departing from the spirit and scope thereof, can make various changes and modifications of the invention to adapt it to various uses and conditions.

General Methods

[0178] Standard recombinant DNA and molecular cloning techniques used in the Examples are well known in the art and are described by Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual*; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY (1989) (Maniatis), Silhavy *et al.*, *Experiments with Gene Fusions*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1984), Ausubel *et al.*, *Current Protocols in Molecular Biology*, pub. by Greene Publishing Assoc. and Wiley-Interscience (1987), and by *Methods in Yeast Genetics*, 2005, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

[0179] 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 Gerhardt *et al.*, eds, American Society for Microbiology, Washington, DC., 1994) or by Thomas D. Brock in *Biotechnology: A Textbook of Industrial Microbiology* (Second Edition, Sinauer Associates, Inc., Sunderland, MA, 1989). All reagents, restriction enzymes and materials used for the growth and maintenance of microbial cells were obtained from Aldrich Chemicals (Milwaukee, WI), BD Diagnostic Systems (Sparks, MD), Life Technologies (Rockville, MD), New England Biolabs (Ipswich, MA) or Sigma Chemical Company (St. Louis, MO) unless otherwise specified. Microbial strains were obtained from The American Type Culture Collection (ATCC), Manassas, VA, unless otherwise noted. All the oligonucleotide primers were synthesized by Sigma-Genosys (Woodlands, TX) or Integrated DNA Technologies (Coralsville, IA). Synthetic complete medium is described in Amberg, Burke and Strathern, 2005, *Methods in Yeast Genetics*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

GC

[0180] The GC method utilized an HP-InnoWax column (30 m x 0.32 mm ID, 0.25 μ m film) from Agilent Technologies (Santa Clara, CA). The carrier gas was helium at a flow rate of 1 ml/min measured at 150 °C with constant head pressure; injector split was 1:10 at 200 °C; oven temperature was 45 °C for 1 min, 45 °C to 230 °C at 10 °C/min, and 230 °C for 30 sec. FID detection was used at 260 °C with 40 ml/min helium makeup gas. Culture broth samples were filtered through 0.2 μ m spin filters before injection. Depending on analytical sensitivity desired, either 0.1 μ l or 0.5 μ l injection volumes were used. Calibrated standard curves were generated for the following compounds: ethanol, isobutanol, acetoin, *meso*-2,3-butanediol, and (2S,3S)-2,3-butanediol. Analytical standards were also utilized to identify retention times for isobutryaldehyde, isobutyric acid, and isoamyl alcohol.

HPLC

[0181] Analysis for fermentation by-product composition is well known to those skilled in the art. For example, one high performance liquid chromatography (HPLC) method

utilizes a Shodex SH-1011 column with a Shodex SH-G guard column (both available from Waters Corporation, Milford, MA), with refractive index (RI) detection. Chromatographic separation is achieved using 0.01 M H₂SO₄ as the mobile phase with a flow rate of 0.5 mL/min and a column temperature of 50 °C. Isobutanol retention time is 47.6 minutes.

Methods for Determining Isobutanol Concentration in Culture Media

[0182] 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, MA), with refractive index (RI) detection. Chromatographic separation was achieved using 0.01 M H₂SO₄ 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 x 0.53 mm id, 1 µm film thickness, Agilent Technologies, Wilmington, DE), 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.

[0183] The meaning of abbreviations is as follows: "s" means second(s), "min" means minute(s), "h" means hour(s), "psi" means pounds per square inch, "nm" means nanometers, "d" means day(s), "µL" means microliter(s), "mL" means milliliter(s), "L" means liter(s), "mm" means millimeter(s), "nm" means nanometers, "mM" means millimolar, "M" means molar, "mmol" means millimole(s), "µmol" means micromole(s), "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" means the gravitation constant, "bp" means base pair(s), "kbp" means kilobase pair(s), "% w/v" means weight/volume percent, "% v/v" means volume/volume percent, "wt %" means percent by weight, "HPLC" means high performance liquid chromatography, and "GC"

means gas chromatography. The term "molar selectivity" is the number of moles of product produced per mole of sugar substrate consumed and is reported as a percent.

Example 1

Construction of *Saccharomyces cerevisiae* strain BP1083 ("NGCI-070"; PNY1504)

- [0184] The strain BP1064 was derived from CEN.PK 113-7D (CBS 8340; Centraalbureau voor Schimmelcultures (CBS) Fungal Biodiversity Centre, Netherlands) and contains deletions of the following genes: URA3, HIS3, PDC1, PDC5, PDC6, and GPD2. BP1064 was transformed with plasmids pYZ090 (SEQ ID NO:134) and pLH468 (SEQ ID NO:135) to create strain NGCI-070 (BP1083, PNY1504).
- [0185] Deletions, which completely removed the entire coding sequence, were created by homologous recombination with PCR fragments containing regions of homology upstream and downstream of the target gene and either a G418 resistance marker or URA3 gene for selection of transformants. The G418 resistance marker, flanked by loxP sites, was removed using Cre recombinase. The URA3 gene was removed by homologous recombination to create a scarless deletion, or if flanked by loxP sites was removed using Cre recombinase.
- [0186] The scarless deletion procedure was adapted from Akada *et al.*, *Yeast*, 23:399, 2006. In general, the PCR cassette for each scarless deletion was made by combining four fragments, A-B-U-C, by overlapping PCR. The PCR cassette contained a selectable/counter-selectable marker, URA3 (Fragment U), consisting of the native CEN.PK 113-7D URA3 gene, along with the promoter (250 bp upstream of the URA3 gene) and terminator (150 bp downstream of the URA3 gene). Fragments A and C, each 500 bp long, corresponded to the 500 bp immediately upstream of the target gene (Fragment A) and the 3' 500 bp of the target gene (Fragment C). Fragments A and C were used for integration of the cassette into the chromosome by homologous recombination. Fragment B (500 bp long) corresponded to the 500 bp immediately downstream of the target gene and was used for excision of the URA3 marker and Fragment C from the chromosome by homologous recombination, as a direct repeat of the sequence corresponding to Fragment B was created upon integration of the cassette into the

chromosome. Using the PCR product ABUC cassette, the URA3 marker was first integrated into and then excised from the chromosome by homologous recombination. The initial integration deleted the gene, excluding the 3' 500 bp. Upon excision, the 3' 500 bp region of the gene was also deleted. For integration of genes using this method, the gene to be integrated was included in the PCR cassette between fragments A and B.

URA3 Deletion

[0187] To delete the endogenous URA3 coding region, a *ura3::loxP-kanMX-loxP* cassette was PCR-amplified from pLA54 template DNA (SEQ ID NO:136). pLA54 contains the *K. lactis* TEF1 promoter and kanMX marker, and is flanked by loxP sites to allow recombination with Cre recombinase and removal of the marker. PCR was done using Phusion DNA polymerase and primers BK505 and BK506 (SEQ ID NOs:137 and 138). The URA3 portion of each primer was derived from the 5' region upstream of the URA3 promoter and 3' region downstream of the coding region such that integration of the loxP-kanMX-loxP marker resulted in replacement of the URA3 coding region. The PCR product was transformed into CEN.PK 113-7D using standard genetic techniques (Methods in Yeast Genetics, 2005, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 201-202) and transformants were selected on YPD containing G418 (100 µg/ml) at 30 °C. Transformants were screened to verify correct integration by PCR using primers LA468 and LA492 (SEQ ID NOs:139 and 140) and designated CEN.PK 113-7D Δ ura3::kanMX.

HIS3 Deletion

[0188] 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, MA) and CEN.PK 113-7D genomic DNA as template, prepared with a Gentra Puregene Yeast/Bact kit (Qiagen; Valencia, CA). HIS3 Fragment A was amplified with primer oBP452 (SEQ ID NO:141) and primer oBP453 (SEQ ID NO:142), 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:143), containing a 5' tail with homology to the 3' end of HIS3 Fragment A, and primer oBP455 (SEQ ID NO:144), 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:145), containing a 5' tail with homology to the 3' end of HIS3 Fragment B, and primer oBP457 (SEQ ID NO:146), 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:147), containing a 5' tail with homology to the 3' end of HIS3 Fragment U, and primer oBP459 (SEQ ID NO:148). 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:141) and oBP455 (SEQ ID NO:144). 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:145) and oBP459 (SEQ ID NO:148). 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:141) and oBP459 (SEQ ID NO:148). The PCR product was purified with a PCR Purification kit (Qiagen).

[0189] Competent cells of CEN.PK 113-7D Δ ura3::kanMX were made and transformed with the HIS3 ABUC PCR cassette using a Frozen-EZ Yeast Transformation II kit (Zymo Research; Orange, CA). Transformation mixtures were plated on synthetic complete media lacking uracil supplemented with 2% glucose at 30 °C. Transformants with a his3 knockout were screened for by PCR with primers oBP460 (SEQ ID NO:149) and oBP461 (SEQ ID NO:150) using genomic DNA prepared with a Gentra Puregene Yeast/Bact kit (Qiagen). A correct transformant was selected as strain CEN.PK 113-7D Δ ura3::kanMX Δ his3::URA3.

KanMX Marker Removal from the Δ ura3 Site and URA3 Marker Removal from the Δ his3 Site

[0190] The KanMX marker was removed by transforming CEN.PK 113-7D Δ ura3::kanMX Δ his3::URA3 with pRS423::PGAL1-cre (SEQ ID NO: 66, described in U.S. Provisional Appl. No. 61/290,639) using a Frozen-EZ Yeast Transformation II kit (Zymo Research) and plating on synthetic complete medium lacking histidine and uracil supplemented with 2% glucose at 30 °C. Transformants were grown in YP supplemented with 1% galactose at 30 °C for ~6 hours to induce the Cre recombinase and KanMX marker excision and plated onto YPD (2% glucose) plates at 30 °C for recovery. An isolate was grown overnight in YPD and plated on synthetic complete medium containing 5-fluoro-oroic acid (0.1%) at 30 °C to select for isolates that lost the URA3 marker. 5-FOA resistant isolates were grown in and plated on YPD for removal of the pRS423::PGAL1-cre plasmid. Isolates were checked for loss of the KanMX marker,

URA3 marker, and pRS423::PGAL1-cre plasmid by assaying growth on YPD+G418 plates, synthetic complete medium lacking uracil plates, and synthetic complete medium lacking histidine plates. A correct isolate that was sensitive to G418 and auxotrophic for uracil and histidine was selected as strain CEN.PK 113-7D Δ ura3::loxP Δ his3 and designated as BP857. The deletions and marker removal were confirmed by PCR and sequencing with primers oBP450 (SEQ ID NO:151) and oBP451 (SEQ ID NO:152) for Δ ura3 and primers oBP460 (SEQ ID NO:149) and oBP461 (SEQ ID NO:150) for Δ his3 using genomic DNA prepared with a Gentra Puregene Yeast/Bact kit (Qiagen).

PDC6 Deletion

[0191] The four fragments for the PCR cassette for the scarless PDC6 deletion were amplified using Phusion High Fidelity PCR Master Mix (New England BioLabs) and CEN.PK 113-7D genomic DNA as template, prepared with a Gentra Puregene Yeast/Bact kit (Qiagen). PDC6 Fragment A was amplified with primer oBP440 (SEQ ID NO:153) and primer oBP441 (SEQ ID NO:154), containing a 5' tail with homology to the 5' end of PDC6 Fragment B. PDC6 Fragment B was amplified with primer oBP442 (SEQ ID NO:155), containing a 5' tail with homology to the 3' end of PDC6 Fragment A, and primer oBP443 (SEQ ID NO:156), containing a 5' tail with homology to the 5' end of PDC6 Fragment U. PDC6 Fragment U was amplified with primer oBP444 (SEQ ID NO:157), containing a 5' tail with homology to the 3' end of PDC6 Fragment B, and primer oBP445 (SEQ ID NO:158), containing a 5' tail with homology to the 5' end of PDC6 Fragment C. PDC6 Fragment C was amplified with primer oBP446 (SEQ ID NO:159), containing a 5' tail with homology to the 3' end of PDC6 Fragment U, and primer oBP447 (SEQ ID NO:160). PCR products were purified with a PCR Purification kit (Qiagen). PDC6 Fragment AB was created by overlapping PCR by mixing PDC6 Fragment A and PDC6 Fragment B and amplifying with primers oBP440 (SEQ ID NO:153) and oBP443 (SEQ ID NO:156). PDC6 Fragment UC was created by overlapping PCR by mixing PDC6 Fragment U and PDC6 Fragment C and amplifying with primers oBP444 (SEQ ID NO:157) and oBP447 (SEQ ID NO:160). The resulting PCR products were purified on an agarose gel followed by a Gel Extraction kit (Qiagen). The PDC6 ABUC cassette was created by overlapping PCR by mixing PDC6 Fragment AB and PDC6 Fragment UC and amplifying with primers oBP440 (SEQ ID NO:153) and

oBP447 (SEQ ID NO:160). The PCR product was purified with a PCR Purification kit (Qiagen).

[0192] Competent cells of CEN.PK 113-7D Δ ura3::loxP Δ his3 were made and transformed with the PDC6 ABUC PCR cassette using a Frozen-EZ Yeast Transformation II kit (Zymo Research). Transformation mixtures were plated on synthetic complete media lacking uracil supplemented with 2% glucose at 30 °C. Transformants with a pdc6 knockout were screened for by PCR with primers oBP448 (SEQ ID NO:161) and oBP449 (SEQ ID NO:162) using genomic DNA prepared with a Gentra Puregene Yeast/Bact kit (Qiagen). A correct transformant was selected as strain CEN.PK 113-7D Δ ura3::loxP Δ his3 Δ pdc6::URA3.

[0193] CEN.PK 113-7D Δ ura3::loxP Δ his3 Δ pdc6::URA3 was grown overnight in YPD and plated on synthetic complete medium containing 5-fluoro-orotic acid (0.1%) at 30 °C to select for isolates that lost the URA3 marker. The deletion and marker removal were confirmed by PCR and sequencing with primers oBP448 (SEQ ID NO:161) and oBP449 (SEQ ID NO:162) using genomic DNA prepared with a Gentra Puregene Yeast/Bact kit (Qiagen). The absence of the PDC6 gene from the isolate was demonstrated by a negative PCR result using primers specific for the coding sequence of PDC6, oBP554 (SEQ ID NO:163) and oBP555 (SEQ ID NO:164). The correct isolate was selected as strain CEN.PK 113-7D Δ ura3::loxP Δ his3 Δ pdc6 and designated as BP891.

PDC1 Deletion ilvDSm Integration

[0194] The PDC1 gene was deleted and replaced with the ilvD coding region from Streptococcus mutans ATCC #700610. The A fragment followed by the ilvD coding region from Streptococcus mutans for the PCR cassette for the PDC1 deletion-ilvDSm integration was amplified using Phusion High Fidelity PCR Master Mix (New England BioLabs) and NYLA83 (described in U.S. Provisional Appl. No. 61/246,709) genomic DNA as template, prepared with a Gentra Puregene Yeast/Bact kit (Qiagen). PDC1 Fragment A-ilvDSm (SEQ ID NO:165) was amplified with primer oBP513 (SEQ ID NO:166) and primer oBP515 (SEQ ID NO:167), containing a 5' tail with homology to the 5' end of PDC1 Fragment B. The B, U, and C fragments for the PCR cassette for the PDC1 deletion-ilvDSm integration were amplified using Phusion High Fidelity PCR Master Mix (New England BioLabs) and CEN.PK 113-7D genomic DNA as template, prepared with a Gentra Puregene Yeast/Bact kit (Qiagen). PDC1 Fragment B was

amplified with primer oBP516 (SEQ ID NO:168) containing a 5' tail with homology to the 3' end of PDC1 Fragment A-ilvDSm, and primer oBP517 (SEQ ID NO:169), containing a 5' tail with homology to the 5' end of PDC1 Fragment U. PDC1 Fragment U was amplified with primer oBP518 (SEQ ID NO:170), containing a 5' tail with homology to the 3' end of PDC1 Fragment B, and primer oBP519 (SEQ ID NO:171), containing a 5' tail with homology to the 5' end of PDC1 Fragment C. PDC1 Fragment C was amplified with primer oBP520 (SEQ ID NO:172), containing a 5' tail with homology to the 3' end of PDC1 Fragment U, and primer oBP521 (SEQ ID NO:173). PCR products were purified with a PCR Purification kit (Qiagen). PDC1 Fragment A-ilvDSm-B was created by overlapping PCR by mixing PDC1 Fragment A-ilvDSm and PDC1 Fragment B and amplifying with primers oBP513 (SEQ ID NO:166) and oBP517 (SEQ ID NO:169). PDC1 Fragment UC was created by overlapping PCR by mixing PDC1 Fragment U and PDC1 Fragment C and amplifying with primers oBP518 (SEQ ID NO:170) and oBP521 (SEQ ID NO:173). The resulting PCR products were purified on an agarose gel followed by a Gel Extraction kit (Qiagen). The PDC1 A-ilvDSm-BUC cassette (SEQ ID NO:174) was created by overlapping PCR by mixing PDC1 Fragment A-ilvDSm-B and PDC1 Fragment UC and amplifying with primers oBP513 (SEQ ID NO:166) and oBP521 (SEQ ID NO:173). The PCR product was purified with a PCR Purification kit (Qiagen).

[0195] Competent cells of CEN.PK 113-7D Δ ura3::loxP Δ his3 Δ pdc6 were made and transformed with the PDC1 A-ilvDSm-BUC PCR cassette using a Frozen-EZ Yeast Transformation II kit (Zymo Research). Transformation mixtures were plated on synthetic complete media lacking uracil supplemented with 2% glucose at 30C. Transformants with a pdc1 knockout ilvDSm integration were screened for by PCR with primers oBP511 (SEQ ID NO:175) and oBP512 (SEQ ID NO:176) using genomic DNA prepared with a Gentra Puregene Yeast/Bact kit (Qiagen). The absence of the PDC1 gene from the isolate was demonstrated by a negative PCR result using primers specific for the coding sequence of PDC1, oBP550 (SEQ ID NO:177) and oBP551 (SEQ ID NO:178). A correct transformant was selected as strain CEN.PK 113-7D Δ ura3::loxP Δ his3 Δ pdc6 Δ pdc1::ilvDSm-URA3.

[0196] CEN.PK 113-7D Δ ura3::loxP Δ his3 Δ pdc6 Δ pdc1::ilvDSm-URA3 was grown overnight in YPD and plated on synthetic complete medium containing 5-fluoro-orotic acid (0.1%) at 30 °C to select for isolates that lost the URA3 marker. The deletion of

PDC1, integration of *ilvDSm*, and marker removal were confirmed by PCR and sequencing with primers oBP511 (SEQ ID NO:175) and oBP512 (SEQ ID NO:176) using genomic DNA prepared with a Gentra Puregene Yeast/Bact kit (Qiagen). The correct isolate was selected as strain CEN.PK 113-7D Δ ura3::loxP Δ his3 Δ pdcc6 Δ pdcc1::ilvDSm and designated as BP907.

PDC5 Deletion *sadB* Integration

- [0197] The PDC5 gene was deleted and replaced with the *sadB* coding region from *Achromobacter xylosoxidans*. A segment of the PCR cassette for the PDC5 deletion-*sadB* integration was first cloned into plasmid pUC19-URA3MCS.
- [0198] pUC19-URA3MCS is pUC19 based and contains the sequence of the URA3 gene from *Saccharomyces cerevisiae* situated within a multiple cloning site (MCS). pUC19 contains the pMB1 replicon and a gene coding for beta-lactamase for replication and selection in *Escherichia coli*. In addition to the coding sequence for URA3, the sequences from upstream and downstream of this gene were included for expression of the URA3 gene in yeast. The vector can be used for cloning purposes and can be used as a yeast integration vector.
- [0199] The DNA encompassing the URA3 coding region along with 250 bp upstream and 150 bp downstream of the URA3 coding region from *Saccharomyces cerevisiae* CEN.PK 113-7D genomic DNA was amplified with primers oBP438 (SEQ ID NO:179), containing BamHI, AscI, PmeI, and FseI restriction sites, and oBP439 (SEQ ID NO:180), containing XbaI, PacI, and NotI restriction sites, using Phusion High-Fidelity PCR Master Mix (New England BioLabs). Genomic DNA was prepared using a Gentra Puregene Yeast/Bact kit (Qiagen). The PCR product and pUC19 (SEQ ID NO:181) were ligated with T4 DNA ligase after digestion with BamHI and XbaI to create vector pUC19-URA3MCS. The vector was confirmed by PCR and sequencing with primers oBP264 (SEQ ID NO:182) and oBP265 (SEQ ID NO:183).
- [0200] The coding sequence of *sadB* and PDC5 Fragment B were cloned into pUC19-URA3MCS to create the *sadB*-BU portion of the PDC5 A-*sadB*-BUC PCR cassette. The coding sequence of *sadB* was amplified using pLH468-*sadB* (SEQ ID NO:184) as template with primer oBP530 (SEQ ID NO:185), containing an AscI restriction site, and primer oBP531 (SEQ ID NO:186), containing a 5' tail with homology to the 5' end of PDC5 Fragment B. PDC5 Fragment B was amplified with primer oBP532 (SEQ ID

NO:187), containing a 5' tail with homology to the 3' end of *sadB*, and primer oBP533 (SEQ ID NO:188), containing a *PmeI* restriction site. PCR products were purified with a PCR Purification kit (Qiagen). *sadB*-PDC5 Fragment B was created by overlapping PCR by mixing the *sadB* and PDC5 Fragment B PCR products and amplifying with primers oBP530 (SEQ ID NO:185) and oBP533 (SEQ ID NO:188). The resulting PCR product was digested with *AscI* and *PmeI* and ligated with T4 DNA ligase into the corresponding sites of pUC19-URA3MCS after digestion with the appropriate enzymes. The resulting plasmid was used as a template for amplification of *sadB*-Fragment B-Fragment U using primers oBP536 (SEQ ID NO:189) and oBP546 (SEQ ID NO:190), containing a 5' tail with homology to the 5' end of PDC5 Fragment C. PDC5 Fragment C was amplified with primer oBP547 (SEQ ID NO:191) containing a 5' tail with homology to the 3' end of PDC5 *sadB*-Fragment B-Fragment U, and primer oBP539 (SEQ ID NO:192). PCR products were purified with a PCR Purification kit (Qiagen). PDC5 *sadB*-Fragment B-Fragment U-Fragment C was created by overlapping PCR by mixing PDC5 *sadB*-Fragment B-Fragment U and PDC5 Fragment C and amplifying with primers oBP536 (SEQ ID NO:189) and oBP539 (SEQ ID NO:192). The resulting PCR product was purified on an agarose gel followed by a Gel Extraction kit (Qiagen). The PDC5 A-*sadB*-BUC cassette (SEQ ID NO:193) was created by amplifying PDC5 *sadB*-Fragment B-Fragment U-Fragment C with primers oBP542 (SEQ ID NO:194), containing a 5' tail with homology to the 50 nucleotides immediately upstream of the native PDC5 coding sequence, and oBP539 (SEQ ID NO:192). The PCR product was purified with a PCR Purification kit (Qiagen).

[0201] Competent cells of CEN.PK 113-7D Δ ura3::loxP Δ his3 Δ pdc6 Δ pdc1::ilvDSm were made and transformed with the PDC5 A-*sadB*-BUC PCR cassette using a Frozen-EZ Yeast Transformation II kit (Zymo Research). Transformation mixtures were plated on synthetic complete media lacking uracil supplemented with 1% ethanol (no glucose) at 30C. Transformants with a *pdc5* knockout *sadB* integration were screened for by PCR with primers oBP540 (SEQ ID NO:195) and oBP541 (SEQ ID NO:196) using genomic DNA prepared with a Gentra Puregene Yeast/Bact kit (Qiagen). The absence of the PDC5 gene from the isolate was demonstrated by a negative PCR result using primers specific for the coding sequence of PDC5, oBP552 (SEQ ID NO:197) and oBP553 (SEQ ID

NO:198). A correct transformant was selected as strain CEN.PK 113-7D Δ ura3::loxP Δ his3 Δ pdc6 Δ pdc1::ilvDSm Δ pdc5::sadB-URA3.

- [0202] CEN.PK 113-7D Δ ura3::loxP Δ his3 Δ pdc6 Δ pdc1::ilvDSm Δ pdc5::sadB-URA3 was grown overnight in YPE (1% ethanol) and plated on synthetic complete medium supplemented with ethanol (no glucose) and containing 5-fluoro-orotic acid (0.1%) at 30 C to select for isolates that lost the URA3 marker. The deletion of PDC5, integration of sadB, and marker removal were confirmed by PCR with primers oBP540 (SEQ ID NO:195) and oBP541 (SEQ ID NO:196) using genomic DNA prepared with a Gentra Puregene Yeast/Bact kit (Qiagen). The correct isolate was selected as strain CEN.PK 113-7D Δ ura3::loxP Δ his3 Δ pdc6 Δ pdc1::ilvDSm Δ pdc5::sadB and designated as BP913.

GPD2 Deletion

- [0203] To delete the endogenous GPD2 coding region, a $gpd2::loxP$ -URA3- $loxP$ cassette (SEQ ID NO:131) was PCR-amplified using $loxP$ -URA3- $loxP$ PCR (SEQ ID NO:200) as template DNA. $loxP$ -URA3- $loxP$ contains the URA3 marker from (ATCC # 77107) flanked by $loxP$ recombinase sites. PCR was done using Phusion DNA polymerase and primers LA512 and LA513 (SEQ ID NOs:201 and 202). The GPD2 portion of each primer was derived from the 5' region upstream of the GPD2 coding region and 3' region downstream of the coding region such that integration of the $loxP$ -URA3- $loxP$ marker resulted in replacement of the GPD2 coding region. The PCR product was transformed into BP913 and transformants were selected on synthetic complete media lacking uracil supplemented with 1% ethanol (no glucose). Transformants were screened to verify correct integration by PCR using primers oBP582 and AA270 (SEQ ID NOs:198 and 203).

- [0204] The URA3 marker was recycled by transformation with pRS423::PGAL1-cre (SEQ ID NO:204) and plating on synthetic complete media lacking histidine supplemented with 1% ethanol at 30 C. Transformants were streaked on synthetic complete medium supplemented with 1% ethanol and containing 5-fluoro-orotic acid (0.1%) and incubated at 30 C to select for isolates that lost the URA3 marker. 5-FOA resistant isolates were grown in YPE (1% ethanol) for removal of the pRS423::PGAL1-cre plasmid. The deletion and marker removal were confirmed by PCR with primers oBP582 (SEQ ID NO:198) and oBP591 (SEQ ID NO:205). The correct isolate was

selected as strain CEN.PK 113-7D Δ ura3::loxP Δ his3 Δ pd₆ Δ pd₁::ilvDSm Δ pd₅::sadB Δ gpd2::loxP and designated as BP1064 (PNY1503).

[0205] BP1064 was transformed with plasmids pYZ090 (SEQ ID NO:134) and pLH468 (SEQ ID NO:135) to create strain NGCI-070 (BP1083; PNY1504).

[0206] pYZ090 is based on the pHR81 (ATCC #87541, Manassas, VA) backbone and was constructed to contain a chimeric gene having the coding region of the *alsS* gene from *Bacillus subtilis* (nt position 457-2172) expressed from the yeast CUP1 promoter (nt 2-449) and followed by the CYC1 terminator (nt 2181-2430) for expression of ALS, and a chimeric gene having the coding region of the *ilvC* gene from *Lactococcus lactis* (nt 3634-4656) expressed from the yeast ILV5 promoter (2433-3626) and followed by the ILV5 terminator (nt 4682-5304) for expression of KARI. The pLH468 plasmid (SEQ ID NO:2) was constructed for expression of DHAD, KivD and HADH in yeast and is described in U.S. Application Publication No. 20090305363, herein incorporated by reference.

Example 2

Construction of *Saccharomyces cerevisiae* strains BP1135 and PNY1507 and Isobutanol-Producing Derivatives

[0207] The purpose of this Example was to construct *Saccharomyces cerevisiae* strains BP1135 and PNY1507. These strains were derived from PNY1503 (BP1064). PNY1503 was derived from CEN.PK 113-7D (CBS 8340; Centraalbureau voor Schimmelcultures (CBS) Fungal Biodiversity Centre, Netherlands). The construction of PNY1503 (BP1064) is described above. BP1135 contains an additional deletion of the *FRA2* gene. PNY1507 was derived from BP1135 with additional deletion of the *ADHI* gene, with integration of the *kivD* gene from *Lactococcus lactis*, codon optimized for expression in *Saccharomyces cerevisiae*, into the *ADHI* locus.

[0208] Deletions, which generally removed the entire coding sequence, were created by homologous recombination with PCR fragments containing regions of homology upstream and downstream of the target gene and the URA3 gene for selection of transformants. The URA3 gene was removed by homologous recombination to create a scarless deletion. Gene integrations were generated in a similar manner.

[0209] The scarless deletion procedure was adapted from Akada *et al.*, *Yeast*, 23:399, 2006. In general, the PCR cassette for each scarless deletion was made by combining four fragments, A-B-U-C, by overlapping PCR. In some instances, the individual fragments were first cloned into a plasmid prior to the entire cassette being amplified by PCR for the deletion/integration procedure. The PCR cassette contained a selectable/counter-selectable marker, *URA3* (Fragment U), consisting of the native CEN.PK 113-7D *URA3* gene, along with the promoter (250 bp upstream of the *URA3* gene) and terminator (150 bp downstream of the *URA3* gene) regions. Fragments A and C, each generally 500 bp long, corresponded to the 500 bp immediately upstream of the target gene (Fragment A) and the 3' 500 bp of the target gene (Fragment C). Fragments A and C were used for integration of the cassette into the chromosome by homologous recombination. Fragment B (500 bp long) corresponded to the 500 bp immediately downstream of the target gene and was used for excision of the *URA3* marker and Fragment C from the chromosome by homologous recombination, as a direct repeat of the sequence corresponding to Fragment B was created upon integration of the cassette into the chromosome.

[0210] Using the PCR product ABUC cassette, the *URA3* marker was first integrated into and then excised from the chromosome by homologous recombination. The initial integration deleted the gene, excluding the 3' 500 bp. Upon excision, the 3' 500 bp region of the gene was also deleted. For integration of genes using this method, the gene to be integrated was included in the PCR cassette between fragments A and B.

FRA2 Deletion

[0211] 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 the scarless *FRA2* deletion were amplified using Phusion High Fidelity PCR Master Mix (New England BioLabs; Ipswich, MA) and CEN.PK 113-7D genomic DNA as template, prepared with a Gentra Puregene Yeast/Bact kit (Qiagen; Valencia, CA). *FRA2* Fragment A was amplified with primer oBP594 (SEQ ID NO:99) and primer oBP595 (SEQ ID NO:100), 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:101), containing a 5' tail with homology to the 3' end of FRA2 Fragment A, and primer oBP597 (SEQ ID NO:102), 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:103), containing a 5' tail with homology to the 3' end of FRA2 Fragment B, and primer oBP599 (SEQ ID NO:104), 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:105), containing a 5' tail with homology to the 3' end of FRA2 Fragment U, and primer oBP601 (SEQ ID NO:106). 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:99) and oBP597 (SEQ ID NO:102). 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:103) and oBP601 (SEQ ID NO:106). 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:99) and oBP601 (SEQ ID NO:106). The PCR product was purified with a PCR Purification kit (Qiagen).

[0212] Competent cells of PNY1503 were made and transformed with the FRA2 ABUC PCR cassette using a Frozen-EZ Yeast Transformation II kit (Zymo Research; Orange, CA). Transformation mixtures were plated on synthetic complete media lacking uracil supplemented with 1% ethanol at 30 °C. Transformants with a *fra2* knockout were screened for by PCR with primers oBP602 (SEQ ID NO:107) and oBP603 (SEQ ID NO:108) using genomic DNA prepared with a Gentra Puregene Yeast/Bact kit (Qiagen). A correct transformant was grown in YPE (yeast extract, peptone, 1% ethanol) and plated on synthetic complete medium containing 5-fluoro-orotic acid (0.1%) at 30 °C to select for isolates that lost the URA3 marker. The deletion and marker removal were confirmed by PCR with primers oBP602 (SEQ ID NO:107) and oBP603 (SEQ ID NO:108) using genomic DNA prepared with a Gentra Puregene Yeast/Bact kit (Qiagen). The absence of the *FRA2* gene from the isolate was demonstrated by a negative PCR result using primers specific for the deleted coding sequence of *FRA2*, oBP605 (SEQ ID NO:109) and oBP606 (SEQ ID NO:110). The correct isolate was selected as strain CEN.PK 113-7D MATa ura3Δ::loxP his3Δ pdc6Δ pdc1Δ::P[PDC1]-DHAD|ilvD_Sm-PDC1t

pdh5Δ::P[PDC5]-ADH|sadB_Ax-PDC5t gpd2Δ::loxP fra2Δ and designated as PNY1505 (BP1135). This strain was transformed with isobutanol pathway plasmids (pYZ090, SEQ ID NO:134) and pLH468 (U.S. Provisional Appl. No. 61/246,709, filed September 29, 2009), and one clone was designated BP1168 (PNY1506).

ADH1 Deletion and *kivD* LI(y) Integration

[0213] The *ADH1* gene was deleted and replaced with the *kivD* coding region from *Lactococcus lactis* codon optimized for expression in *Saccharomyces cerevisiae*. The scarless cassette for the *ADH1* deletion-*kivD*_LI(y) integration was first cloned into plasmid pUC19-URA3MCS, as described in U.S. Appln. No. 61/356379, filed June 18, 2010, incorporated herein by reference.

[0214] The *kivD* coding region from *Lactococcus lactis* codon optimized for expression in *Saccharomyces cerevisiae* was amplified using pLH468 (U.S. Provisional Appl. No. 61/246,709, filed September 29, 2009) as template with primer oBP562 (SEQ ID NO:111), containing a PmeI restriction site, and primer oBP563 (SEQ ID NO:112), containing a 5' tail with homology to the 5' end of ADH1 Fragment B. ADH1 Fragment B was amplified from genomic DNA prepared as above with primer oBP564 (SEQ ID NO:113), containing a 5' tail with homology to the 3' end of *kivD*_LI(y), and primer oBP565 (SEQ ID NO:114), containing a FseI restriction site. PCR products were purified with a PCR Purification kit (Qiagen). *kivD*_LI(y)-ADH1 Fragment B was created by overlapping PCR by mixing the *kivD*_LI(y) and ADH1 Fragment B PCR products and amplifying with primers oBP562 (SEQ ID NO:111) and oBP565 (SEQ ID NO:114). The resulting PCR product was digested with PmeI and FseI and ligated with T4 DNA ligase into the corresponding sites of pUC19-URA3MCS after digestion with the appropriate enzymes. ADH1 Fragment A was amplified from genomic DNA with primer oBP505 (SEQ ID NO:115), containing a SacI restriction site, and primer oBP506 (SEQ ID NO:116), containing an AscI restriction site. The ADH1 Fragment A PCR product was digested with SacI and AscI and ligated with T4 DNA ligase into the corresponding sites of the plasmid containing *kivD*_LI(y)-ADH1 Fragment B. ADH1 Fragment C was amplified from genomic DNA with primer oBP507 (SEQ ID NO:117), containing a PacI restriction site, and primer oBP508 (SEQ ID NO:118), containing a SalI restriction site. The ADH1 Fragment C PCR product was digested with PacI and SalI and ligated with T4

DNA ligase into the corresponding sites of the plasmid containing ADH1 Fragment A-kivD_Ll(y)-ADH1 Fragment B. The hybrid promoter UAS(PGK1)-P_{FBA1} was amplified from vector pRS316-UAS(PGK1)-P_{FBA1}-GUS (described below; SEQ ID NO:206) with primer oBP674 (SEQ ID NO:119), containing an AscI restriction site, and primer oBP675 (SEQ ID NO:120), containing a PmeI restriction site. The UAS(PGK1)-P_{FBA1} PCR product was digested with AscI and PmeI and ligated with T4 DNA ligase into the corresponding sites of the plasmid containing kivD_Ll(y)-ADH1 Fragments ABC. The entire integration cassette was amplified from the resulting plasmid with primers oBP505 (SEQ ID NO:115) and oBP508 (SEQ ID NO:118) and purified with a PCR Purification kit (Qiagen).

[0215] Competent cells of PNY1505 were made and transformed with the ADH1-kivD_Ll(y) PCR cassette constructed above using a Frozen-EZ Yeast Transformation II kit (Zymo Research). Transformation mixtures were plated on synthetic complete media lacking uracil supplemented with 1% ethanol at 30 °C. Transformants were grown in YPE (1% ethanol) and plated on synthetic complete medium containing 5-fluoro-orotic acid (0.1%) at 30 °C to select for isolates that lost the URA3 marker. The deletion of *ADH1* and integration of *kivD_Ll(y)* were confirmed by PCR with external primers oBP495 (SEQ ID NO:121) and oBP496 (SEQ ID NO:122) and with *kivD_Ll(y)* specific primer oBP562 (SEQ ID NO:111) and external primer oBP496 (SEQ ID NO:122) using genomic DNA prepared with a Gentra Puregene Yeast/Bact kit (Qiagen). The correct isolate was selected as strain CEN.PK 113-7D MATa ura3Δ::loxP his3Δ pdc6Δ pdc1Δ::P[PDC1]-DHAD|ilvD_Sm-PDC1tpdc5Δ::P[PDC5]-ADH|sadB_Ax-PDC5t gpd2Δ::loxP fra2Δ adh1Δ::UAS(PGK1)P[FBA1]-kivD_Ll(y)-ADH1t and designated as PNY1507 (BP1201). PNY1507 was transformed with isobutanol pathway plasmids pYZ090 (SEQ ID NO:134) and pBP915 (described below). Isobutanol production by these derivatives is described below.

Construction of the pRS316-UAS(PGK1)-FBA1p-GUS vector

[0216] To clone a cassette UAS(PGK1)-FBA1p (SEQ ID NO:129), first a 602bp FBA1 promoter (FBA1p) was PCR-amplified from genomic DNA of CEN.PK with primers T-FBA1(SalI) (SEQ ID NO:123) and B-FBA1(SpeI) (SEQ ID NO:124), and cloned into SalI and SpeI sites on the plasmid pWS358-PGK1p-GUS (SEQ ID NO:130) after the

PGK1p promoter was removed with a Sall/SpeI digest of the plasmid, yielding pWS358-FBA1p-GUS. The pWS358-PGK1p-GUS plasmid was generated by inserting a PGK1p and beta-glucuronidase gene (GUS) DNA fragments into multiple cloning site of pWS358, which was derived from pRS423 vector (Christianson *et al.*, *Gene*, 110:119-122, 1992). Secondly, the resulting pWS358-FBA1p-GUS plasmid was digested with Sall and SacI, a DNA fragment containing a FBA1p promoter, GUS gene, and FBAt terminator gel-purified, and cloned into Sall/SacI sites on pRS316 to create pRS316-FBA1p-GUS. Thirdly, a 118bp DNA fragment containing an upstream activation sequence (UAS) located between positions -519 and -402 upstream of the 3-phosphoglycerate kinase (PGK1) open reading frame, namely UAS(PGK1), was PCR-amplified from genomic DNA of CEN.PK with primers T-U/PGK1(KpnI) (SEQ ID NO:125) and B-U/PGK1(Sall) (SEQ ID NO:126). The PCR product was digested with KpnI and Sall and cloned into KpnI/Sall sites on pRS316-FBA1p-GUS to create pRS316-UAS(PGK1)-FBA1p-GUS.

Example 3

Construction of PNY2204 and Isobutanol-Producing Derivatives

[0217] The purpose of this example is to describe construction of a vector to enable integration of a gene encoding acetolactate synthase into the naturally occurring intergenic region between the PDC1 and TRX1 coding sequences in Chromosome XII.

Construction of integration vector pUC19-kan::pdc1::FBA-alsS::TRX1

[0218] The FBA-alsS-CYCt cassette was constructed by moving the 1.7kb BbvCI/PacI fragment from pRS426::GPD::alsS::CYC (U.S. Appl. Pub. No. 20070092957) to pRS426::FBA::ILV5::CYC (U.S. Appl. Pub. No. 20070092957, previously digested with BbvCI/PacI to release the ILV5 gene). Ligation reactions were transformed into *E. coli* TOP10 cells and transformants were screened by PCR using primers N98SeqF1 (SEQ ID NO:91) and N99SeqR2 (SEQ ID NO:93). The FBA-alsS-CYCt cassette was isolated from the vector using BglII and NotI for cloning into pUC19-URA3::ilvD-TRX1 (as described in U.S. Appln. No. 61/356379, filed June 18, 2010, incorporated herein by

reference, clone "B"; herein SEQ ID NO: 243) at the AflII site (Klenow fragment was used to make ends compatible for ligation). Transformants containing the *alsS* cassette in both orientations in the vector were obtained and confirmed by PCR using primers N98SeqF4 (SEQ ID NO:92) and N1111 (SEQ ID NO:97) for configuration "A" and N98SeqF4 (SEQ ID NO:92) and N1110 (SEQ ID NO:96) for configuration "B". A geneticin selectable version of the "A" configuration vector was then made by removing the URA3 gene (1.2 kb *NotI/NaeI* fragment) and adding a geneticin cassette (SEQ ID NO: 244 herein; previously described in U.S. Appln. No. 61/356379, filed June 18, 2010, incorporated herein by reference) maintained in a pUC19 vector (cloned at the *SmaI* site). The *kan* gene was isolated from pUC19 by first digesting with *KpnI*, removal of 3' overhanging DNA using Klenow Fragment (NEB, Cat. No. M212), digesting with *HincII* and then gel purifying the 1.8 kb gene fragment (Zymoclean™ Gel DNA Recovery Kit, Cat. No. D4001, Zymo Research, Orange, CA; SEQ ID NO: 245). Klenow fragment was used to make all ends compatible for ligation, and transformants were screened by PCR to select a clone with the geneticin resistance gene in the same orientation as the previous URA3 marker using primers BK468 (SEQ ID NO:90) and N160SeqF5 (SEQ ID NO:94). The resulting clone was called pUC19-*kan::pdc1::FBA-alsS::TRX1* (clone A)(SEQ ID NO:131).

Construction of *alsS* integrant strains and isobutanol-producing derivatives

[0219] The pUC19-*kan::pdc1::FBA-alsS* integration vector described above was linearized with *PmeI* and transformed into PNY1507 (described above in Example 1). *PmeI* cuts the vector within the cloned *pdc1*-*TRX1* intergenic region and thus leads to targeted integration at that location (Rodney Rothstein, *Methods in Enzymology*, 1991, volume 194, pp. 281-301). Transformants were selected on YPE plus 50 µg/ml G418. Patched transformants were screened by PCR for the integration event using primers N160SeqF5 (SEQ ID NO:94) and oBP512 (SEQ ID NO:98). Two transformants were tested indirectly for acetolactate synthase function by evaluating the strains ability to make isobutanol. To do this, additional isobutanol pathway genes were supplied on *E. coli*-yeast shuttle vectors (pYZ090Δ*alsS* and pBP915, described below). One clone, strain MATa *ura3Δ::loxP his3Δ pdc6Δ pdc1Δ::P[PDC1]-DHAD|ilvD_Sm-PDC1t-pUC19-loxP-kanMX-loxP-P[FBA1]-ALS|alsS_Bs-CYC1t pdc5Δ::P[PDC5]-*

ADH|sadB_Ax-PDC5t gpd2Δ::loxP fra2Δ adh1Δ::UAS(PGK1)P[FBA1]-kivD_Ll(y)-ADH1t was designated as PNY2204. The plasmid-free parent strain was designated PNY2204. The PNY2204 locus (pdc1Δ::ilvD::pUC19-kan::FBA-alsS::TRX1) is depicted in FIG. 5.

Isobutanol pathway plasmids (pYZ090ΔalsS and pBP915)

- [0220] pYZ090 (SEQ ID NO:134) was digested with *SpeI* and *NotI* to remove most of the CUP1 promoter and all of the *alsS* coding sequence and CYC terminator. The vector was then self-ligated after treatment with Klenow fragment and transformed into *E. coli* Stbl3 cells, selecting for ampicillin resistance. Removal of the DNA region was confirmed for two independent clones by DNA sequencing across the ligation junction by PCR using primer N191 (SEQ ID NO:95). The resulting plasmid was named pYZ090ΔalsS (SEQ ID NO:132).
- [0221] pBP915 was constructed from pLH468 (SEQ ID NO:124) by deleting the *kivD* gene and 957 base pairs of the *TDH3* promoter upstream of *kivD*. pLH468 was digested with *SwaI* and the large fragment (12896 bp) was purified on an agarose gel followed by a Gel Extraction kit (Qiagen; Valencia, CA). The isolated fragment of DNA was self-ligated with T4 DNA ligase and used to transform electrocompetent TOP10 *Escherichia coli* (Invitrogen; Carlsbad, CA). Plasmids from transformants were isolated and checked for the proper deletion by restriction analysis with the *SwaI* restriction enzyme. Isolates were also sequenced across the deletion site with primers oBP556 (SEQ ID NO:127) and oBP561 (SEQ ID NO:128). A clone with the proper deletion was designated pBP915 (pLH468ΔkivD)(SEQ ID NO:133).

Example 4

Isobutanol Production in Strains with an Integrated Copy of the *kivD* Gene

- [0222] The purpose of this example is to show isobutanol production in strains with an integrated copy of the *kivD* gene compared to strains with plasmid-borne *kivD*. Strains without the *kivD* integration, carrying plasmids pYZ090 and pLH468, were compared to the integration strain, PNY1507, carrying plasmid pYZ090 and pBP915. All media

components were from Sigma-Aldrich, St. Louis, MO. Strains were grown in synthetic medium (Yeast Nitrogen Base Without Amino Acids and Yeast Synthetic Drop-Out Media Supplement without uracil, histidine, tryptophan, and leucine) supplemented with 76 mg/L tryptophan, 38 0mg/L leucine, 100 mM MES pH5.5, 20 mg/L nicotinic acid, 20 mg/L thiamine hydrochloride, 0.2% glucose, and 0.2% ethanol. Overnight cultures were grown in 8 ml of medium in 125 ml vented Erlenmeyer flasks at 30 °C, 250 RPM in a New Brunswick Scientific I24 shaker. 19 mL of medium in 125 mL tightly-capped Erlenmeyer flasks was inoculated with overnight culture to an OD600 0.5 and grown for 8 hours at 30 °C, 250 RPM in a New Brunswick Scientific I24 shaker. Glucose was added to 2% (time 0 hours). After 48 hours, culture supernatants (collected using Spin-X centrifuge tube filter units, Costar Cat. No. 8169) were analyzed by HPLC per methods described in U.S. Appl. Pub. No. 20070092957. Results are shown in Table 3. The strains with an integrated copy of the *kivD* gene has a similar isobutanol titer compared to strains with plasmid-borne *kivD*.

Table 3: Isobutanol titer in strains with an integrated or plasmid-borne *kivD* gene

Strain	Isobutanol Titer (g/L)
PNY1506 (BP1168)	1.7 +/- 0.3 (n=2*)
PNY1507 / pYZ090 / pBP915	1.8 +/- 0.1 (n=2 [#])

*Biological replicates

[#]Independent transformants

Example 5

Isobutanol Production in Strains with an Integrated Copy of the *alsS* Gene

[0223] The purpose of this example is to show increased production of isobutanol when the acetolactate synthase was removed from a plasmid and integrated into the yeast genome. Strains without *alsS* integration (PNY1507 carrying plasmids pYZ090 and pBP915) were compared to the integration strains (PNY2204 carrying plasmid pYZ090DalsS and pBP915). All strains were grown in synthetic complete medium, minus histidine and uracil containing 0.3 % glucose and 0.3 % ethanol as carbon sources (10 mL medium in 125 mL vented Erlenmeyer flasks (VWR Cat. No. 89095-260). After overnight incubation (30 °C, 250 rpm in an Innova®40 New Brunswick Scientific

Shaker), cultures were diluted back to 0.2 OD (Eppendorf BioPhotometer measurement) in synthetic complete medium containing 2% glucose and 0.05% ethanol (20 ml medium in 125 mL tightly-capped Erlenmeyer flasks (VWR Cat. No. 89095-260)). After 48 hours incubation (30 °C, 250 rpm in an Innova®40 New Brunswick Scientific Shaker), culture supernatants (collected using Spin-X centrifuge tube filter units, Costar Cat. No. 8169) were analyzed by HPLC per methods described in U.S. Appl. Pub. No. 20070092957. Results are shown below in Table 4. The isobutanol titer from strains with an integrated copy of the *alsS* gene were significantly greater than the isobutanol titer without *alsS* integration.

Table 4: Isobutanol titer in strains with or without *alsS* gene integration

Strain	Isobutanol Titer (g/L)
PNY1507 / pYZ090 / pBP915	1.5 +/- 0.2 (n=3*)
PNY2204 / pYZ090Δ <i>alsS</i> / pBP915 (PNY2205)	2.6 +/- 0.1 (n=3*)

*Biological replicates

Example 6

Isobutanol Production in Strains with an Integrated Copy of the *alsS* Gene

[0224] The purpose of this Example is to show increased cell density and production of isobutanol when the acetolactate synthase was removed from a plasmid and integrated into the yeast genome. Strains without *alsS* integration (PNY1504 as described in U.S. Appln. No 61/379,546, filed September 2, 2010, incorporated herein by reference, and PNY1506) were compared to the integration strain PNY2205 (PNY2204 transformed with pYZ090Δ*alsS* and pBP915 plasmids and having *alsS* integration).

Inoculum and bioreactor media

[0225] A yeast inoculum media (1 L) was prepared containing 6.7 g of Yeast Nitrogen Base w/o amino acids (Difco 0919-15-3); 2.8 g Yeast Synthetic Drop-out Medium Supplement Without Histidine, Leucine, Tryptophan and Uracil (Sigma Y2001); 20 mL of 1% (w/v) L-Leucine; 4 mL of 1% (w/v) L-Tryptophan; 0.8 mL of Ergosterol & Tween solution; 3 g of ethanol; and 3 g of glucose. For 10 mL Ergosterol & Tween solution, 100

mg of Ergosterol was dissolved in 5 mL 100% ethanol and 5 mL Tween 80. The solution was heated for 10 min at 70 °C.

[0226] A 125 mL shake flask was inoculated directly from a frozen vial by pipetting the whole vial culture (approx. 1 mL) into 10 mL of the inoculum medium. The flask was incubated at 260 rpm and 30 °C. The strain was grown overnight until OD about 1.0. OD at $\lambda = 600$ nm was determined in a HELIOS α spectrophotometer (Thermo Electron Corporation, USA). At this point, a 2L shake flask containing 110 mL of the inoculum medium were inoculated from the overnight culture. The starting OD in the 2L flask was 0.1. The flask was incubated at 260 rpm and 30 °C. When OD in the shake flask reached about 1.0, 20 mL of 1M MES buffer, 20 mL of 10x yeast extract and peptone (YEP), glucose up to final concentration of 30 g/L and about 160 mL of oleyl alcohol (90-95 %, Cognis, Cincinnati OH, USA) were added to the shake flask. 24 hours afterwards, the oleyl alcohol was removed and bioreactors inoculated.

[0227] A 10x YEP solution was prepared by dissolving 100 g of yeast extract and 200 of peptone in water to a final volume of 1 L.

[0228] A bioreactor medium (1 L) was prepared containing:

- (i) salts: ammonium sulfate 5.0 g, potassium phosphate monobasic 2.8 g, magnesium sulfate heptahydrate 1.9 g, zinc sulfate heptahydrate 0.2 g;
- (ii) vitamins: biotin (D-) 0.40 mg, Ca D(+) panthotenate 8.00 mg, myo-inositol 200.00 mg, pyridoxol hydrochloride 8.00 mg, p-aminobenzoic acid 1.60 mg, riboflavin 1.60 mg, folic acid 0.02 mg, niacin 30.0 mg, and thiamine 30 mg;
- (iii) amino acids: yeast synthetic drop-out medium supplement without histidine, leucine, tryptophan and uracil (Sigma Y2001) 2.8 g, 1% (w/v) L-leucine 20 mL, and 1% (w/v) L-tryptophan 4 mL; and
- (iv) trace elements: EDTA (Titrplex III7) 99.38 mg, zinc sulphate heptahydrate 29.81 mg, manganese chloride dehydrate 5.57 mg, cobalt(II)chloride hexahydrate 1.99 mg, copper(II)sulphate pentahydrate 1.99 mg, Di-sodium molybdenum dehydrate 2.65 mg, calcium chloride dehydrate 29.81 mg, iron sulphate heptahydrate 19.88 mg, boric acid.

Bioreactor experimental design

[0229] Experiments were executed in 2 L BIORSTAT B-DCU Tween2L bioreactors from Sartorius (USA). The fermentors are connected to mass-spec from Thermo Electron

Corporation (USA) Directly after inoculation with 80 mL of the inoculum the volume in fermentors was about 800 mL, dissolved oxygen tension (DOT) was controlled at 10%, pH was controlled at 5.25, aeration was controlled at 0.5 L/min, 0.8 L of oleyl alcohol was added. Oleyl alcohol was used in order to extract isobutanol from culture broth.

Methods for analyzing cultivation experiments

- [0230] Optical density (OD) at $\lambda = 600$ nm was determined using a spectrophotometer by pipetting a well mixed broth sample into an appropriate cuvette (CS500 VWR International, Germany). If the biomass concentration of the sample exceeded the linear absorption range of the spectrophotometer (typically OD values from 0.000 to 0.600), the sample was diluted with 0.9% NaCl solution to yield values in the linear range.
- [0231] Metabolites and products in medium were analyzed and quantified using a GC method and an ZB-WAXplus column (30 m x 0.25 mm ID, 0.25 μ m film) from Phenomenex (Torrance, CA). A helium carrier gas was used at a constant flow rate of 2.3 mL/min; an injector split of 1:20 at 250 °C; an oven temperature of 70 °C for 1 min, followed by 70 °C to 160 °C at 10°C/min, and 160 °C to 240°C at 30°C/min. Flame Ionization Detection (FID) was used at 260 °C with 40 mL/min helium makeup gas. Culture broth samples were filtered through 0.2 μ m spin filters before injection. 0.5 μ L injection volumes were used. Calibrated standard curves were generated for isobutanol.
- [0232] Glucose and fermentation by-product analysis were carried out by high performance liquid chromatography (HPLC) using methods known in the art. The HPLC method utilized a Shodex SH-1011 column with a Shodex SH-G guard column (both available from Waters Corporation, Milford, MA), with refractive index (RI) detection. Chromatographic separation was achieved using 0.01 N H₂SO₄ as the mobile phase with a flow rate of 0.5 mL/min and a column temperature of 50 °C. Isobutanol retention time was 47.6 minutes.
- [0233] Isobutanol concentration in the culture supernatant was determined by the HPLC method. Isobutanol concentration in the oleyl alcohol phase was determined by the GC method. Isobutanol concentration in off-gas samples was determined by mass-spec as mentioned above.

Results

[0234] The measured values of optical density (OD), isobutanol production rate (R), produced isobutanol per liter of culture broth (isobutanol titer, T), and isobutanol yield per consumed glucose (Y) at about 46 hours of fermentation time are presented in Table 5. The PNY2205 strain compared to PNY1504 and PNY1506 strains grow to higher cell density and resulted in higher titer and rate but similar yield.

Table 5: Optical density, isobutanol production rate, titer and yield in PNY1504, PNY1506 and PNY2205 strains

Strain	OD	R (g/L/h)	T (g/L)	Y (g/g)
PNY1504	26.5	0.45	20.7	0.27
PNY1506	27.2	0.55	25.2	0.28
PNY2205	34.6	0.84	39.7	0.27

Example 7

Comparing the performance of strains PNY1504 and PNY2205 under the same reactive liquid extraction conditions

Stock Solutions Used

Pre-Seed Media

[0235] The following reagents were mixed with gentle agitation at room temperature: 6.7 g of Yeast Nitrogen Base without amino acids (Difco 0919-15-3); 2.8 g of Yeast Synthetic Drop-out Medium Supplement Without Histidine, Leucine, Tryptophan and Uracil (Sigma Y2001); 20 mL of 1% (w/v) L-Leucine; 4 mL of 1% (w/v) L-Tryptophan; 3 g of ethanol; 3 g of glucose and enough water to make a total of 1L of solution.

Seed Flask Media

[0236] The following reagents were mixed with gentle agitation at room temperature: 6.7 g of Yeast Nitrogen Base without amino acids (Difco 0919-15-3); 2.8 g of Yeast Synthetic Drop-out Medium Supplement Without Histidine, Leucine, Tryptophan and

Uracil (Sigma Y2001); 20 mL of 1% (w/v) L-Leucine; 4 mL of 1% (w/v) L-Tryptophan; 3 g of ethanol; 30 g of glucose; 38 g of MES buffer (Sigma-Aldrich YXXX) and enough water to make a total of 1L of solution. After mixing, the solution was filter sterilized.

Ergosterol Solution

- [0237] A solution of 0.2 g of Ergosterol, 10 mL of 200 proof ethanol and 10 ml of Tween 80 was mixed and heated to 70 °C for 10 minutes.

Distillase Stock Solution

- [0238] A solution of 0.9 mL of Distillase L 400 and 49.1 mL of filter sterilized tap water was mixed.

Lipolase 100L Stock Solution

- [0239] A solution of 2.12 mL of Lipolase 100L (Sigma Aldrich L0777) and 40 g of phosphate buffer solution at pH 6.8 were mixed and filter sterilized.

Vitamin Stock Solution

- [0240] 5 g of nicotinic acid and 1 g of thiamine were mixed in 500 mL of filter sterilized Deionized water.

Corn Mash

- [0241] Corn mash was added to a 30L liquefaction tank. Next, 16910 g of tap water was added to the 30L liquefaction tank with agitation at 120 rpm. The tank was outfitted with a dual-blade pitched-blade turbine with $D_B/D_T = \sim 0.25$. Next, 14091 g of ground corn (ground in a Hammer Mill with a 1 micron screen) was added, and the mash was heated to 55°C and held there for 30 minutes. The pH was adjusted to 5.8 by adding 5.4 g of 17% NaOH solution in water. An alpha-amylase enzyme solution was prepared by mixing 1986 g of tap water and 19.5 g of Spezyme Fred L from Genencor and sterile filtered the resulting solution through a 0.2 micron filter. 2004 g of this solution was added to the 30L liquefaction tank and held at 55 °C for an additional 60 minutes. The solution was then heated to 95°C and held at that temperature for 120 minutes. The solution was cooled to 30 °C before using in fermentation.

PNY1504 Process

Pre-Seed Growth

- [0242] 30 mL of Pre-Seed Media was added to a 250 mL baffled, vented shake flask. Next, 2 Frozen Seed Vials of Strain PNY1504, ca. 1.5 mL of total volume, were added to

the same flask. The culture was then incubated for 24 hours at 30 °C at 250 rpm on an incubator shaker.

Seed Flask Stage 1

- [0243] 15 mL of the pre-seed culture was added to 300 mL of the Seed Flask media in a 2L baffled, vented shake flask. The flask was incubated for 24 hours at 30 °C and 250 rpm on an incubator shaker.

Seed Flask Stage 2

- [0244] 30 mL of yeast extract peptone and 300 mL of sterile oleyl alcohol were then added to the flask. The flask was incubated for 24 hours at 30 °C at 250 rpm on an incubator shaker.

1L Production Fermentor

- [0245] A 1L fermentor with water covering the probes was sterilized for 30 min at 121 °C. The water was drained and 520 mL of sterile corn mash media was added. Next, the following aseptic additions to the corn mash were made in the fermentor: 3.8 mL of ethanol, 0.6 mL of 1% ergosterol solution, 6 mL of nicotinic acid/thiamine solution and 4.8 mL of Lipolase 100L stock solution. Next, 60 mL of the aqueous phase of Seed Flask Stage 2 was added, followed by 2 mL of the Distillase stock solution. Directly thereafter, 96 mL of corn oil fatty acid was added. After 12 hours, 2 mL of the Distillase Stock solution was added. At 24 hours post inoculation, another 2 mL of Distillase Stock solution was added. The solution was then incubated at pH 5.2, temperature 30 °C and a pO₂ (partial pressure of dissolved oxygen) setpoint of 3 %. Airflow was set at 0.2 slpm (standard liters per minute) and the pO₂ was controlled via agitation. pH was controlled with 20 % w/v KOH solution and no acid was required throughout the fermentation. Samples were taken and analyzed over the course of the fermentation.

PNY2205 Process

Pre-Seed Growth

- [0246] 30 mL of Pre-Seed Media was added to a 250 mL baffled, vented shake flask. Next, 2 Frozen Seed Vials of Strain PNY2205, ca. 1.5 mL of total volume, were added to the same flask. We then held the flask for 24 hours at 30 °C at 250 rpm on an incubator shaker.

Seed Flask Stage 1

- [0247] 300 mL of the Seed Flask media was added to a 2 L baffled, vented shake flask. 15 mL of pre-seed culture was added flask and incubated for 24 hours at 30 °C and 250 rpm on an incubator shaker.

Seed Flask Stage 2

- [0248] 30 mL of yeast extract peptone and 300 mL of sterile oleyl alcohol was added to the flask and the flask was incubated for 24 hours at 30 °C and 250 rpm on an incubator shaker.

1L Production Fermentor

- [0249] A 1L fermentor with water covering the probes was sterilized for 30 min at 121 °C. The water was drained and 520 mL of sterile corn mash media was added. Next, the following aseptic additions were made to the corn mash in the fermentor: 3.8 mL of ethanol, 0.6 mL of 1% ergosterol solution, 6 mL of nicotinic acid/thiamine solution and 4.8 mL of Lipolase 100L stock solution. Next, 60 mL of the aqueous phase of Seed Flask Stage 2 was added, followed by 2 mL of the Distillase stock solution. Directly thereafter, 96 mL of corn oil fatty acid was added. 12 hours post inoculation, 2 mL of the Distillase Stock solution was added. At 24 hours post inoculation, 2 mL of Distillase Stock solution was also added. The solution was incubated at pH 5.2, temperature 30 °C and the pO₂ setpoint of 3 %. Airflow was set at 0.2 slpm and the pO₂ was controlled via agitation. pH was controlled with 20 % w/v KOH solution and no acid was required throughout the fermentation. Samples were taken and analyzed over the course of the fermentation.

Methods for analyzing cultivation experiments

- [0250] Optical density (OD) of the resulting cultures was measured at $\lambda = 600$ nm using a spectrophotometer. First, a well mixed broth sample was pipetted into an appropriate cuvette. When the biomass concentration of the sample exceeded the linear absorption range of the spectrophotometer (typically OD values from 0.000 to 0.600), the sample was diluted with 0.9% NaCl solution to yield values in the linear range. Dry weight of the cell suspension was determined by centrifuging 5 mL of cell broth in a pre-weighed centrifuge tube, followed by washing with distilled water, drying to constant weight at 80 °C in an oven and determining the weight difference.

[0251] Metabolites and products in medium were analyzed and quantified by a GC method utilizing a ZB-WAXplus column (30 m x 0.25 mm ID, 0.25 μ m film) from Phenomenex (Torrance, CA). The carrier gas was helium at a constant flow rate of 2.3 mL/min; injector split was 1:20 at 250 °C; oven temperature is 70 °C for 1 min, 70 °C to 160°C at 10 °C/min, and 160 °C to 240 °C at 30°C/min. FID detection was used at 260 °C with 40 ml/min helium makeup gas. Culture broth samples were filtered through 0.2 μ m spin filters before injection. A calibrated standard curve for isobutanol (w-methyl-1-propanol) was used.

[0252] Glucose analysis was carried out by YSI (YSI 2700 Select biochemistry analyzer that uses enzyme electrode technology to generate rapid measurement of glucose concentration.

Results

[0253] Isobutanol production rate, isobutanol per liter of culture broth (effective titer), and isobutanol yield per consumed glucose are presented in Table 6. The PNY2205 strain compared to PNY1504 strains resulted in higher production rate and titer but similar yield.

Table 6: Optical density and isobutanol production of PNY2205 compared to PNY1504

52-56 hr result	PNY1504	PNY2205
rate, g/L-h	0.50	0.64
effective titer (g/L)	26.3	35.5
g/g glu yield	0.27	0.27

Example 8

Comparing the performance of strains PNY1504 and PNY2205 under the same reactive liquid extraction conditions

Stock Solutions Used

Pre-Seed Media

- [0254] The following reagents were mixed with gentle agitation at room temperature: 6.7 g of Yeast Nitrogen Base without amino acids (Difco 0919-15-3); 2.8 g of Yeast Synthetic Drop-out Medium Supplement Without Histidine, Leucine, Tryptophan and Uracil (Sigma Y2001); 20 mL of 1% (w/v) L-Leucine; 4 mL of 1% (w/v) L-Tryptophan; 3 g of ethanol; 3 g of glucose and enough water to make a total of 1L of solution.

Seed Flask Media

- [0255] The following reagents were mixed with gentle agitation at room temperature: 6.7 g of Yeast Nitrogen Base without amino acids (Difco 0919-15-3); 2.8 g of Yeast Synthetic Drop-out Medium Supplement Without Histidine, Leucine, Tryptophan and Uracil (Sigma Y2001); 20 mL of 1% (w/v) L-Leucine; 4 mL of 1% (w/v) L-Tryptophan; 3 g of ethanol; 30 g of glucose; 38 g of MES buffer (Sigma-Aldrich YXXX) and enough water to make a total of 1L of solution. After mixing, the solution was filter sterilized.

Ergosterol Solution

- [0256] A solution of 0.2 g of Ergosterol, 10 mL of 200 proof ethanol and 10 mL of Tween 80 was mixed and heated to 70 °C for 10 minutes.

Distillase Stock Solution

- [0257] A solution of 0.9 mL of Distillase L 400 and 49.1 mL of filter sterilized tap water was mixed.

Lipolase 100L Stock Solution

- [0258] A solution of 2.12 mL of Lipolase 100L (Sigma Aldrich L0777) and 40 g of phosphate buffer solution at pH 6.8 was mixed and filter sterilized.

Vitamin Stock Solution

- [0259] A solution of 5 g of nicotinic acid and 1 g of thiamine in was mixed in 500 mL of filter sterilized Deionized water.

Corn Mash

- [0260] Corn mash was added to a 30L liquefaction tank. Next, 16910 g of tap water was added to the 30L liquefaction tank with agitation at 120 rpm. The tank was outfitted with a dual-blade pitched-blade turbine with $D_{B/DT} = \sim 0.25$. Next, 14091 g of ground corn (ground in a Hammer Mill with a 1 micron screen) was added and the mash heated to 55°C and incubated for 30 minutes. The pH was adjusted to 5.8 by adding 5.4 g of 17% NaOH solution in water. An alpha-amylase enzyme solution was prepared by mixing

1986 g of tap water and 19.5 g of Spezyme Fred L from Genencor and sterile filtering through a 0.2 micron filter. 2004 g of this solution was added to the 30L liquefaction tank and incubated at 55 °C for an additional 60 minutes. Then, the solution was heated to 95 °C and held there for 120 minutes. The solution was then cooled to 30 °C before using in fermentation.

PNY1504 Process

Pre-Seed Growth

- [0261] 30 mL of Pre-Seed Media was added to a 250 mL baffled, vented shake flask. Next, 2 Frozen Seed Vials of Strain PNY1504, ca. 1.5 ml of total volume, were added to the same flask. The culture was incubated for 24 hours at 30 °C at 250 rpm on an incubator shaker.

Seed Flask Stage 1

- [0262] 300 mL of the Seed Flask media was added to a 2L baffled, vented shake flask. 15 mL of pre-seed was then transferred to flask. The flask was then incubated for 24 hours at 30 °C and 250 rpm on an incubator shaker.

Seed Flask Stage 2

- [0263] 30 mL of yeast extract peptone and 300 mL of sterile oleyl alcohol were added to the flask and the flask incubated for 24 hours at 30 °C at 250 rpm on an incubator shaker.

1L Production Fermentor

- [0264] A 1L fermentor with water covering the probes was sterilized for 30 min at 121 °C. The water was drained and 520 mL of sterile corn mash media was added. Next, the following aseptic additions were made to the corn mash in the fermentor: 3.8 mL of ethanol, 0.6 mL of 1% ergosterol solution, 6 mL of nicotinic acid/thiamine solution and 4.8 mL of Lipolase 100L stock solution. Next, 60 mL of the aqueous phase of Seed Flask Stage 2 was added, followed by 2 mL of the Distillase stock solution. Directly thereafter, 141 mL of corn oil fatty acid was added. At 12 hours post inoculation, 2 mL of the Distillase Stock solution was added. At 24 hours post inoculation, 2 mL of Distillase Stock solution was also added. The solution was then incubated at pH 5.2, temperature 30 °C and pO₂ setpoint of 3 %. Airflow was set at 0.2 slpm and pO₂ was controlled via agitation. pH was controlled with 20 % w/v KOH solution and no acid was

required throughout the fermentation. Samples were taken and analyzed over the course of the fermentation.

PNY2205 Process

Pre-Seed Growth

- [0265] 30 mL of Pre-Seed Media was added to a 250 mL baffled, vented shake flask. Next, 2 Frozen Seed Vials of Strain PNY2205, ca. 1.5 ml of total volume, were added to the same flask. The flask was then incubated for 24 hours at 30 °C at 250 rpm on an incubator shaker.

Seed Flask Stage 1

- [0266] 300 mL of the Seed Flask media was added to a 2L baffled, vented shake flask. 15 mL of the pre-seed growth was then added to the flask and incubated for 24 hours at 30 °C and 250 rpm on an incubator shaker.

Seed Flask Stage 2

- [0267] 30 mL of yeast extract peptone and 300 mL of sterile oleyl alcohol were added to the flask and incubated for 24 hours at 30 °C and 250 rpm on an incubator shaker.

1L Production Fermentor

- [0268] A 1L fermentor with water covering the probes was sterilized for 30 min at 121 °C. The water was drained and 520 mL of sterile corn mash media was added. Next, the following aseptic additions were made to the corn mash in the fermentor: 3.8 mL of ethanol, 0.6 mL of 1% ergosterol solution, 6 mL of nicotinic acid/thiamine solution and 4.8 mL of Lipolase 100L stock solution. Next, 60 mL of the aqueous phase of Seed Flask Stage 2 was added followed by 2 mL of the Distillase stock solution. Directly thereafter, 96 mL of corn oil fatty acid was added. At 12 hours post inoculation, 2 mL of the Distillase Stock solution was added. At 24 hours post inoculation, 2 mL of Distillase Stock solution was also added and the solution was incubated at pH 5.2, temperature 30 °C and pO₂ setpoint of 3 %. Airflow was set at 0.2 slpm and pO₂ was controlled via agitation. pH was controlled with 20 % w/v KOH solution and no acid was required throughout the fermentation. Samples were taken and analyzed over the course of the fermentation.

Results

[0269] Isobutanol production rate, isobutanol per liter of culture broth (effective titer), and isobutanol yield per consumed glucose are presented in Table 7. The PNY2205 strain compared to PNY1504 strains resulted in higher production rate and titer but similar yield.

Table 7: Optical density and isobutanol production of PNY2205 compared to PNY1504

52-56 hr result	PNY1504	PNY2205
rate, g/l-h	0.48	0.54
effective titer (g/l)	25.2	30.1
g/g glu yield	0.27	0.27

Example 9

Comparing the performance of strains PNY1504 and PNY2205 under the same reactive liquid extraction conditions

Stock Solutions Used

Pre-Seed Media

[0270] The following reagents were mixed with gentle agitation at room temperature: 6.7 g of Yeast Nitrogen Base without amino acids (Difco 0919-15-3); 2.8 g of Yeast Synthetic Drop-out Medium Supplement Without Histidine, Leucine, Tryptophan and Uracil (Sigma Y2001); 20 mL of 1% (w/v) L-Leucine; 4 mL of 1% (w/v) L-Tryptophan; 3 g of ethanol; 3 g of glucose and enough water to make a total of 1L of solution.

Seed Flask Media

[0271] The following reagents were mixed with gentle agitation at room temperature: 6.7 g of Yeast Nitrogen Base without amino acids (Difco 0919-15-3); 2.8 g of Yeast Synthetic Drop-out Medium Supplement Without Histidine, Leucine, Tryptophan and Uracil (Sigma Y2001); 20 mL of 1% (w/v) L-Leucine; 4 mL of 1% (w/v) L-Tryptophan; 3 g of ethanol; 30 g of glucose; 38 g of MES buffer (Sigma-Aldrich YXXX) and enough water to make a total of 1L of solution. After mixing, the solution was filter sterilized.

Ergosterol Solution

[0272] A solution of 0.2 g of Ergosterol, 10 mL of 200 proof ethanol and 10 mL of Tween 80 was mixed and heated to 70 °C for 10 minutes.

Distillase Stock Solution

- [0273] A solution of 0.9 mL of Distillase L 400 and 49.1 mL of filter sterilized tap water was mixed.

Lipolase 100L Stock Solution

- [0274] A solution of 2.12 mL of Lipolase 100L (Sigma Aldrich L0777) and 40 g of phosphate buffer solution at pH 6.8 was mixed and filter sterilized.

Vitamin Stock Solution

- [0275] A solution of 5 g of nicotinic acid and 1 g of thiamine in was mixed in 500 mL of filter sterilized Deionized water.

Corn Mash

- [0276] Corn mash was added to a 30L liquefaction tank. Next, 16910 g of tap water was added to the 30L liquefaction tank with agitation at 120 rpm. The tank was outfitted with a dual-blade pitched-blade turbine with $D_B/D_T = \sim 0.25$. Next, 14091 g of ground corn (ground in a Hammer Mill with a 1 micron screen) was added and the mash heated to 55°C and incubated for 30 minutes. The pH was adjusted to 5.8 by adding 5.4 g of 17% NaOH solution in water. An alpha-amylase enzyme solution was prepared by mixing 1986 g of tap water and 19.5 g of Spezyme Fred L from Genencor and sterile filtering through a 0.2 micron filter. 2004 g of this solution was added to the 30L liquefaction tank and incubated at 55 °C for an additional 60 minutes. Then, the solution was heated to 95 °C and held there for 120 minutes. The solution was then cooled to 30 °C before using in fermentation.

PNY1504 ProcessPre-Seed Growth

- [0277] 30 mL of Pre-Seed Media was added to a 250 mL baffled, vented shake flask. Next, 2 Frozen Seed Vials of Strain PNY1504, ca. 1.5 ml of total volume, were added to the same flask. The culture was incubated for 24 hours at 30 °C at 250 rpm on an incubator shaker.

Seed Flask Stage 1

- [0278] 300 mL of the Seed Flask media was added to a 2L baffled, vented shake flask. 15 mL of pre-seed was then transferred to flask. The flask was then incubated for 24 hours at 30 °C and 250 rpm on an incubator shaker.

Seed Flask Stage 2

- [0279] 30 mL of yeast extract peptone and 300 mL of sterile oleyl alcohol were added to the flask and the flask incubated for 24 hours at 30 °C at 250 rpm on an incubator shaker.

1L Production Fermentor

- [0280] A 1L fermentor with water covering the probes was sterilized for 30 min at 121 °C. The water was drained and 520 mL of sterile corn mash media was added. Next, the following aseptic additions were made to the corn mash in the fermentor: 3.8 mL of ethanol, 0.6 mL of 1% ergosterol solution, 6 mL of nicotinic acid/thiamine solution and 4.8 mL of Lipolase 100L stock solution. Next, 60 mL of the aqueous phase of Seed Flask Stage 2 was added, followed by 2 mL of the Distillase stock solution. Directly thereafter, 141 mL of corn oil fatty acid was added. At 12 hours post inoculation, 2 mL of the Distillase Stock solution was added. At 24 hours post inoculation, 2 mL of Distillase Stock solution was also added. The solution was then incubated at pH 5.2, temperature 30 °C and pO₂ setpoint of 3 %. Airflow was set at 0.2 slpm and pO₂ was controlled via agitation. pH was controlled with 20 % w/v KOH solution and no acid was required throughout the fermentation. Samples were taken and analyzed over the course of the fermentation.

PNY2205 Process

Pre-Seed Growth

- [0281] 30 mL of Pre-Seed Media was added to a 250 mL baffled, vented shake flask. Next, 2 Frozen Seed Vials of Strain PNY2205, ca. 1.5 ml of total volume, were added to the same flask. The flask was then incubated for 24 hours at 30 °C at 250 rpm on an incubator shaker.

Seed Flask Stage 1

- [0282] 300 mL of the Seed Flask media was added to a 2L baffled, vented shake flask. 15 mL of the pre-seed growth was then added to the flask and incubated for 24 hours at 30 °C and 250 rpm on an incubator shaker.

Seed Flask Stage 2

[0283] 30 mL of yeast extract peptone and 300 mL of sterile oleyl alcohol were added to the flask and incubated for 24 hours at 30 °C and 250 rpm on an incubator shaker.

1L Production Fermentor

[0284] A 1L fermentor with water covering the probes was sterilized for 30 min at 121 °C. The water was drained and 520 mL of sterile corn mash media was added. Next, the following aseptic additions were made to the corn mash in the fermentor: 3.8 mL of ethanol, 0.6 mL of 1% ergosterol solution, 6 mL of nicotinic acid/thiamine solution and 4.8 mL of Lipolase 100L stock solution. Next, 60 mL of the aqueous phase of Seed Flask Stage 2 was added followed by 2 mL of the Distillase stock solution. Directly thereafter, 96 mL of corn oil fatty acid was added. At 12 hours post inoculation, 2 mL of the Distillase Stock solution was added. At 24 hours post inoculation, 2 mL of Distillase Stock solution was also added and the solution was incubated at pH 5.2, temperature 30 °C and pO₂ setpoint of 3 %. Airflow was set at 0.2 slpm and pO₂ was controlled via agitation. pH was controlled with 20 % w/v KOH solution and no acid was required throughout the fermentation. Samples were taken and analyzed over the course of the fermentation.

Results

[0285] Isobutanol production rate, isobutanol per liter of culture broth (effective titer), and isobutanol yield per consumed glucose are presented in Table 8. The PNY2205 strain compared to PNY1504 strains resulted in higher production rate and titer but similar yield.

Table 8: Isobutanol production of PNY2205 compared to PNY1504

52-56 hr result	PNY1504	PNY2205
rate, g/l-h	0.51	0.58
effective titer (g/l)	26.7	32.6
g/g glu yield	0.27	0.27

Example 10

Construction of *S. cerevisiae* strain PNY2211

[0286] PNY2211 was constructed in several steps from *S. cerevisiae* strain PNY1507 (Example 2) as described in the following paragraphs. First, the strain was modified to contain a phosphoketolase gene. Construction of phosphoketolase gene cassettes and integration strains was previously described in U.S. Appl. No. 61/356,379, filed June 18, 2010. Next, an acetolactate synthase gene (*alsS*) was added to the strain, using an integration vector described in Example 3. Finally, homologous recombination was used to remove the phosphoketolase gene and integration vector sequences, resulting in a scarless insertion of *alsS* in the intergenic region between *pdc1Δ::ilvD* (a previously described deletion/insertion of the PDC1 ORF, U.S. Appl. No. 61/356,379, filed June 18, 2010; see Example 1 herein) and the native TRX1 gene of chromosome XII. The resulting genotype of PNY2211 is MATa *ura3Δ::loxP* *his3Δ* *pdc6Δ* *pdc1Δ::P[PDC1]-DHAD|ilvD_Sm-PDC1t-P[FBA1]-ALS|alsS_Bs-CYC1t* *pdc5Δ::P[PDC5]-ADH|sadB_Ax-PDC5t* *gpd2Δ::loxP* *fra2Δ* *adh1Δ::UAS(PGK1)P[FBA1]-kivD_LI(y)-ADH1t*.

[0287] A phosphoketolase gene cassette was introduced into PNY1507 by homologous recombination. The integration construct was generated as follows. The plasmid *pRS423::CUP1-alsS+FBA-budA* (described in U.S. Pub. No. 2009/0305363 A1) was digested with *NotI* and *XmaI* to remove the 1.8 kb FBA-budA sequence, and the vector was religated after treatment with Klenow fragment. Next, the CUP1 promoter was replaced with a TEF1 promoter variant (M4 variant described by Nevoigt *et al. Appl. Environ. Microbiol.* 2006. 72(8): 5266-5273) via DNA synthesis and vector construction service from DNA2.0 (Menlo Park, CA). The resulting plasmid, *pRS423::TEF(M4)-alsS* was cut with *StuI* and *MluI* (removes 1.6 kb portion containing part of the *alsS* gene and CYC1 terminator), combined with the 4 kb PCR product generated from *pRS426::GPD-xpk1+ADH-eutD* (described in U.S. Appl. No. 61/356,379, filed June 18, 2010; SEQ ID NO: 246 herein) with primers N1176 (SEQ ID NO:207) and N1177 (SEQ ID NO:208) and an 0.8 kb PCR product DNA generated from yeast genomic DNA (ENO1 promoter region) with primers N822 (SEQ ID NO:209) and N1178 (SEQ ID NO:210) and transformed into *S. cerevisiae* strain BY4741 (ATCC 201388; gap repair cloning methodology, *see* Ma and Botstein). Transformants were obtained by plating cells on synthetic complete medium without histidine. Proper assembly of the expected plasmid (*pRS423::TEF(M4)-xpk1+ENO1-eutD*, SEQ ID NO:211) was confirmed by PCR

(primers N821 (SEQ ID NO:212) and N1115 (SEQ ID NO:213)) and by restriction digest (*Bgl*I). Two clones were subsequently sequenced. The 3.1 kb TEF(M4)-xpk1 gene was isolated by digestion with *Sac*I and *Not*I and cloned into the pUC19-URA3::ilvD-TRX1 vector (described in U.S. Appl. No. 61/356,379, filed June 18, 2010 SEQ ID NO: 243, herein) Clone A, cut with *Afl*III). Cloning fragments were treated with Klenow fragment to generate blunt ends for ligation. Ligation reactions were transformed into *E. coli* Stbl3 cells, selecting for ampicillin resistance. Insertion of TEF(M4)-xpk1 was confirmed by PCR (primers N1110 (SEQ ID NO:214) and N1114 (SEQ ID NO:215)). The vector was linearized with *Afl*III and treated with Klenow fragment. The 1.8 kb KpnI-HincII geneticin resistance cassette (described in U.S. Appl. No. 61/356,379, filed June 18, 2010; SEQ ID NO: 245 herein), was cloned by ligation after Klenow fragment treatment. Ligation reactions were transformed into *E. coli* Stbl3 cells, selecting for ampicillin resistance. Insertion of the geneticin cassette was confirmed by PCR (primers N160SeqF5 (SEQ ID NO:216) and BK468 (SEQ ID NO:217)). The plasmid sequence is provided as SEQ ID NO:218 (pUC19-URA3::pdc1::TEF(M4)-xpk1::kan).

[0288] The resulting integration cassette (pdc1::TEF(M4)-xpk1::KanMX::TRX1) was isolated (*Asc*I and *Nae*I digestion generated a 5.3 kb band that was gel purified) and transformed into PNY1507 (Example 2) using the Zymo Research Frozen-EZ Yeast Transformation Kit (Cat. No. T2001). Transformants were selected by plating on YPE plus 50 µg/ml G418. Integration at the expected locus was confirmed by PCR (primers N886 (SEQ ID NO:219) and N1214 (SEQ ID NO:220)). Next, plasmid pRS423::GAL1p-Cre, encoding Cre recombinase, was used to remove the loxP-flanked KanMX cassette (vector and methods described herein). Proper removal of the cassette was confirmed by PCR (primers oBP512 (SEQ ID NO:221) and N160SeqF5 (SEQ ID NO:222)). Finally, the *alsS* integration plasmid described herein (pUC19-kan::pdc1::FBA-alsS::TRX1, clone A) was transformed into this strain using the included geneticin selection marker. Two integrants were tested for acetolactate synthase activity by transformation with plasmids pYZ090ΔalsS and pBP915 (plasmids described herein, transformed using Protocol #2 in "Methods in Yeast Genetics" 2005. Amberg, Burke and Strathern) and evaluation of growth and isobutanol production in glucose-containing media (methods for growth and isobutanol measurement are described herein and U.S. Appl. No. 60/730,290, filed October 26, 2005 and U.S. Pub. No. 2007/0092957 A1). One

of the two clones was positive and was named PNY2218. An isolate of PNY2218 containing the plasmids pYZ090 Δ alsS and pBP915 was designated PNY2209.

[0289] PNY2218 was treated with Cre recombinase and resulting clones were screened for loss of the *xpk1* gene and pUC19 integration vector sequences by PCR (primers N886 (SEQ ID NO:219) and N160SeqR5 (SEQ ID NO:222)). This leaves only the *alsS* gene integrated in the *pdcl*-TRX1 intergenic region after recombination the DNA upstream of *xpk1* and the homologous DNA introduced during insertion of the integration vector (a "scarless" insertion since vector, marker gene and loxP sequences are lost, FIG. 6). Although this recombination could have occurred at any point, the vector integration appeared to be stable even without geneticin selection and the recombination event was only observed after introduction of the Cre recombinase. One clone was designated PNY2211.

Example 11

Comparing the performance of strains PNY2205 and PNY2211 under the same reactive liquid extraction conditions

[0290] Isolates with the scarless integration (in particular, two clones "B" and "M") were transformed with pYZ090 Δ alsS and pBP915 in order to compare isobutanol production with PNY2205. Integrants were selected on synthetic complete medium (minus histidine and uracil) containing 1% ethanol as the carbon source. Integrants were patched to the same medium, and patched cells were patched again to plates containing 2% glucose plus 0.05% ethanol as carbon sources. After two days, patches were used to inoculate liquid medium (10 mL synthetic complete, minus histidine and uracil, with 2% glucose and 0.05% ethanol in 125 mL vented flasks). After overnight incubation (30°C, 250 rpm) cultures were diluted back to OD 0.2 (20 mL medium in 125 mL tightly capped flasks). After 48 hours, samples were taken to determine isobutanol production. The new strain backgrounds supported similar isobutanol production to PNY2205. Clone M was selected for further engineering was named PNY2211. Clone M7 transformed with plasmids pYZ090 Δ alsS and pBP915 was designated PNY2213.

[0291] The production of isobutanol per liter of culture broth (effective titer in g/L) of strains PNY2205, Clone B and Clone M is presented in Table 9. Clone B and M strains had a similar isobutanol titer compared to PNY2205.

Table 9: Isobutanol production of PNY2205 compared to PNY2211

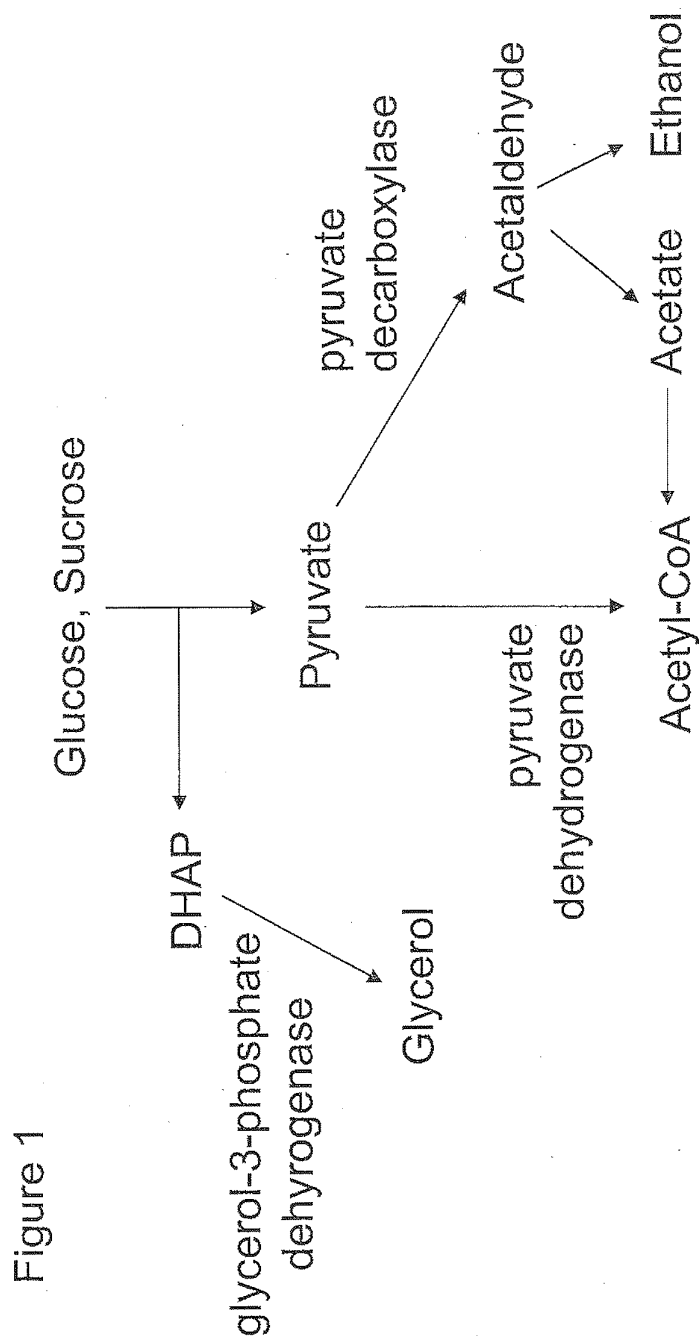
Strains	Isobutanol titer (g/L)
PNY2205	4.0
Clone B strains (n=3)	3.8 +/- 0.5
Clone M strains (n=3)(PNY2213)	4.0 +/-0.5

The Claims defining the invention are as follows:

1. A recombinant host cell comprising an isobutanol biosynthetic pathway wherein said pathway comprises the substrate to product conversion pyruvate to acetolactate catalyzed by a polypeptide encoded by a heterologous polynucleotide integrated into the chromosome and wherein said pathway comprises at least one substrate to product conversion catalyzed by a polypeptide encoded by a polynucleotide on a plasmid; and wherein expression of Fra2 in the host cell is decreased or substantially eliminated.
2. The recombinant host cell of claim 1, wherein said pathway comprises the substrate to product conversion acetolactate to 2,3-dihydroxyisovalerate catalyzed by a polypeptide encoded by a polynucleotide on a plasmid.
3. The recombinant host cell of either claim 1 or claim 2, wherein the host cell is a member of a genus selected from the group consisting of: *Clostridium*, *Zymomonas*, *Escherichia*, *Salmonella*, *Rhodococcus*, *Pseudomonas*, *Bacillus*, *Lactobacillus*, *Enterococcus*, *Alcaligenes*, *Klebsiella*, *Paenibacillus*, *Arthrobacter*, *Corynebacterium*, *Brevibacterium*, *Pichia*, *Candida*, *Hansenula*, *Issatchenkia*, *Kluyveromyces*, and *Saccharomyces*.
4. The recombinant host cell of any one of claims 1 to 2, wherein the polypeptide is expressed in the host cell cytosol.
5. The recombinant host cell of any one of claims 1 to 4, wherein the host cell further comprises polynucleotides encoding polypeptides that catalyze the substrate to product conversions: 2,3-dihydroxyisovalerate to α -ketoisovalerate; α -ketoisovalerate to isobutyraldehyde; and isobutyraldehyde to isobutanol.
6. The recombinant host cell of any one of claims 1 to 4, wherein the host cell further comprises polynucleotides encoding polypeptides that catalyze the substrate to product conversions: 2,3-dihydroxyisovalerate to α -ketoisovalerate; α -ketoisovalerate to isobutyryl-CoA; isobutyryl-CoA to isobutyraldehyde; and isobutyraldehyde to isobutanol.
7. The recombinant host cell of any one of claims 1 to 4, wherein the host cell further comprises polynucleotides encoding polypeptides that catalyze the substrate to product conversions: 2,3-dihydroxyisovalerate to α -ketoisovalerate; α -ketoisovalerate to valine;

- valine to isobutylamine; isobutylamine to isobutyraldehyde; and isobutyraldehyde to isobutanol.
8. The recombinant host cell of any one of claims 1 to 4, wherein the polypeptide that catalyzes a substrate to product conversion of pyruvate to acetolactate is acetolactate synthase.
 9. The recombinant host of claim 8, wherein the acetolactate synthase has at least about 80% identity to an amino acid sequence selected from SEQ ID NOs: 2, 4, 6, 8, 10, 12, 14, 16, or 18.
 10. The recombinant host cell of any one of claims 1 to 9, wherein expression of pyruvate decarboxylase in the host cell is decreased or substantially eliminated.
 11. The recombinant host cell of any one of claims 1 to 9, wherein the polynucleotide encoding a polypeptide which catalyzes the conversion of pyruvate to acetolactate is integrated into an endogenous polynucleotide encoding a polypeptide having pyruvate decarboxylase activity.
 12. The recombinant host cell of any one of claims 1 to 11, wherein expression of glycerol-3-phosphate dehydrogenase in the host cell is decreased or substantially eliminated.
 13. A method comprising:
 - (a) providing the recombinant host cell of any one of claims 1 to 12, or any combination thereof; and
 - (b) contacting the host cell with a fermentable carbon substrate to form a fermentation broth under conditions whereby the biosynthetic pathway product is produced.
 14. The method of claim 13, further comprising contacting the fermentation broth with an extractant to produce a two-phase fermentation mixture.
 15. The method of claim 14, wherein the extractant comprises a water immiscible organic extractant selected from the group consisting of: C₁₂ to C₂₂ fatty alcohols, C₁₂ to C₂₂ fatty acids, esters of C₁₂ to C₂₂ fatty acids, C₁₂ to C₂₂ fatty aldehydes, and C₁₂ to C₂₂ fatty amides.

16. The method of any one of claims 13 to 15, wherein the biosynthetic pathway product is an alcohol, and wherein the method further comprises contacting the fermentation broth with an organic acid and an enzyme capable of esterifying the alcohol with the organic acid.
17. The method of any one of claims 13 to 16, wherein the biosynthetic pathway product is an alcohol, and wherein the method further comprises vaporizing at least a portion of the fermentation broth to form a vapor stream comprising water and alcohol.
18. The method of any one of claims 13 to 17 wherein the biosynthetic pathway product is isobutanol.
19. A composition comprising:
 - (i) the host cell of any one of claims 1 to 12, or any combination thereof;
 - (ii) butanol;
 - (iii) an extractant; and
 - (iv) optionally an esterification enzyme.
20. Use of a host cell of any one of claims 1 to 12 or a composition of claim 19 in a process for the production of isobutanol, or for the production of a fuel composition comprising isobutanol.



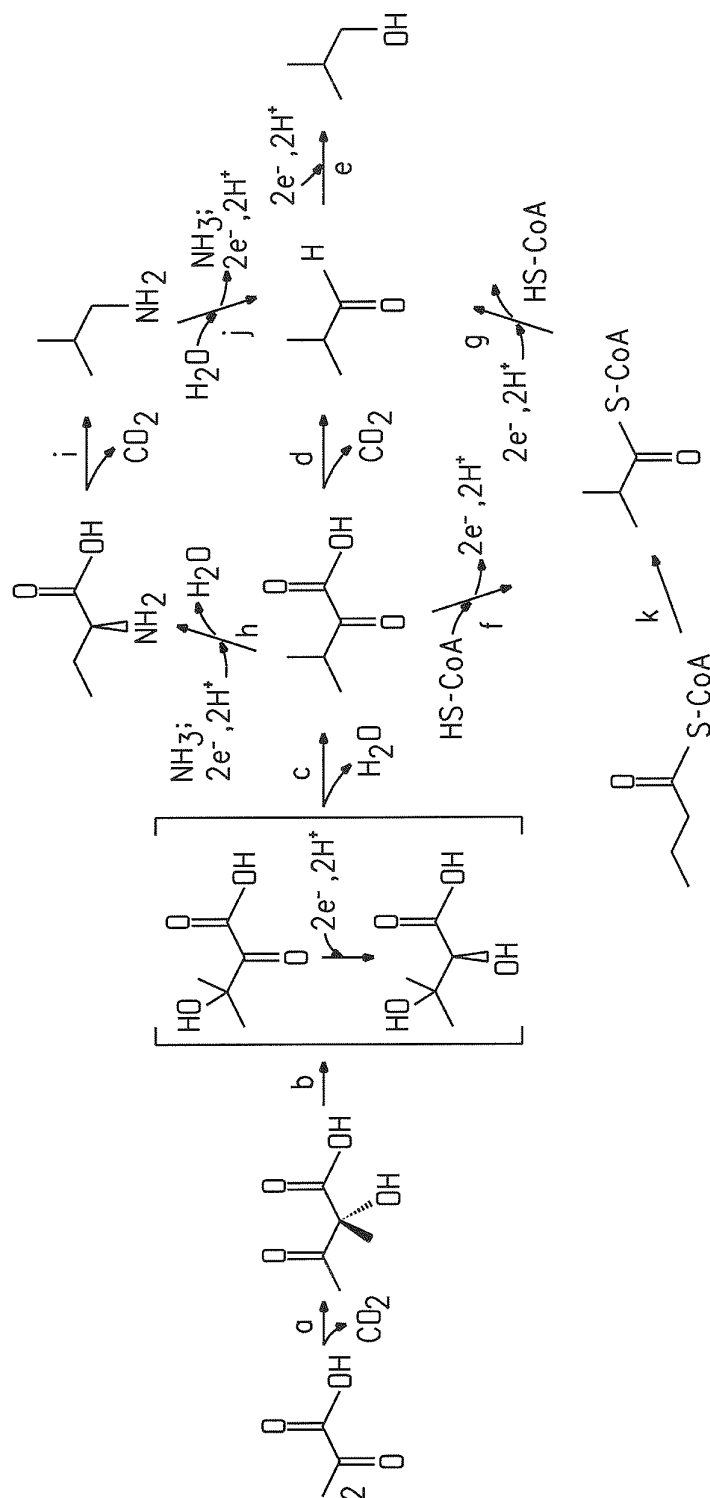


FIG. 2

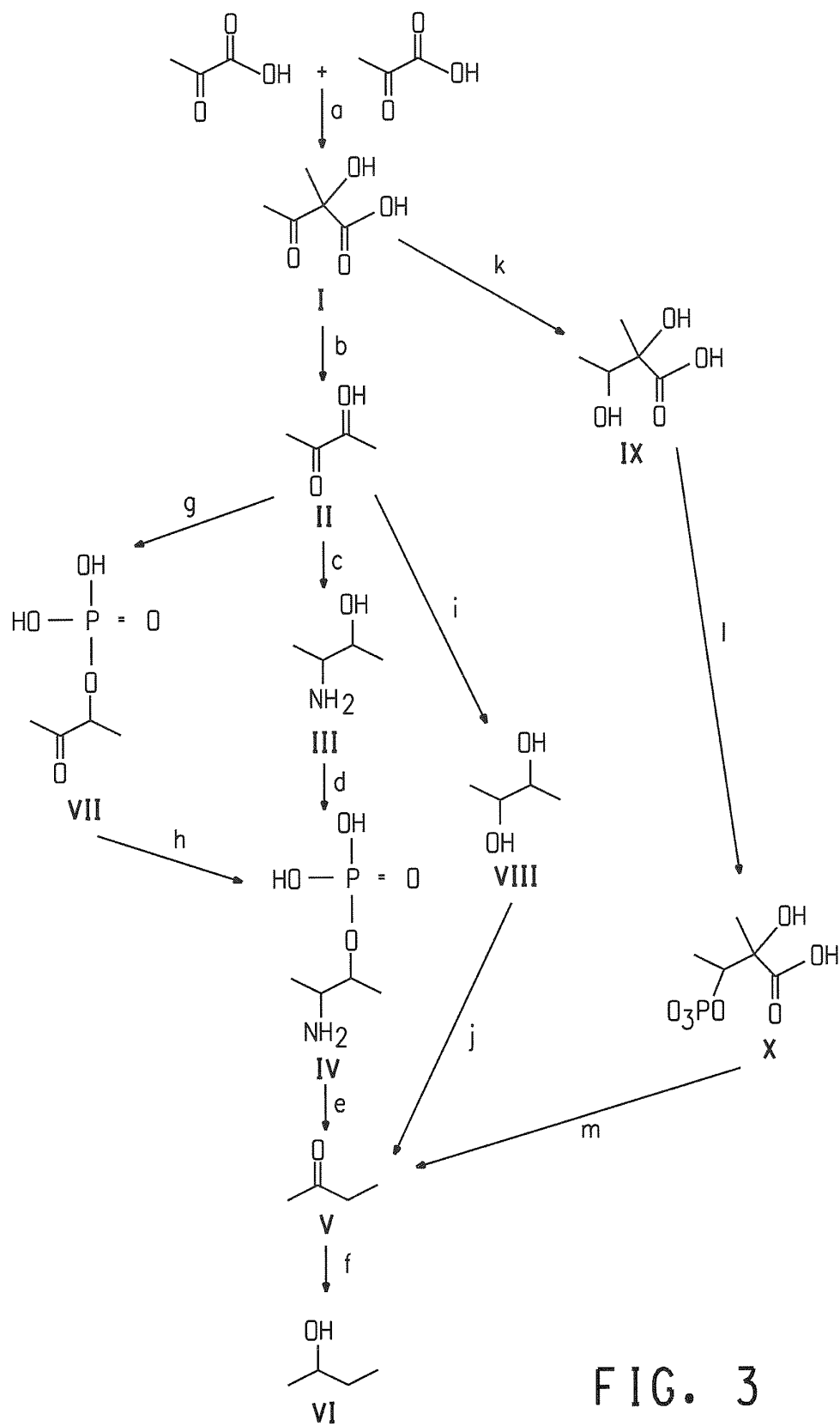
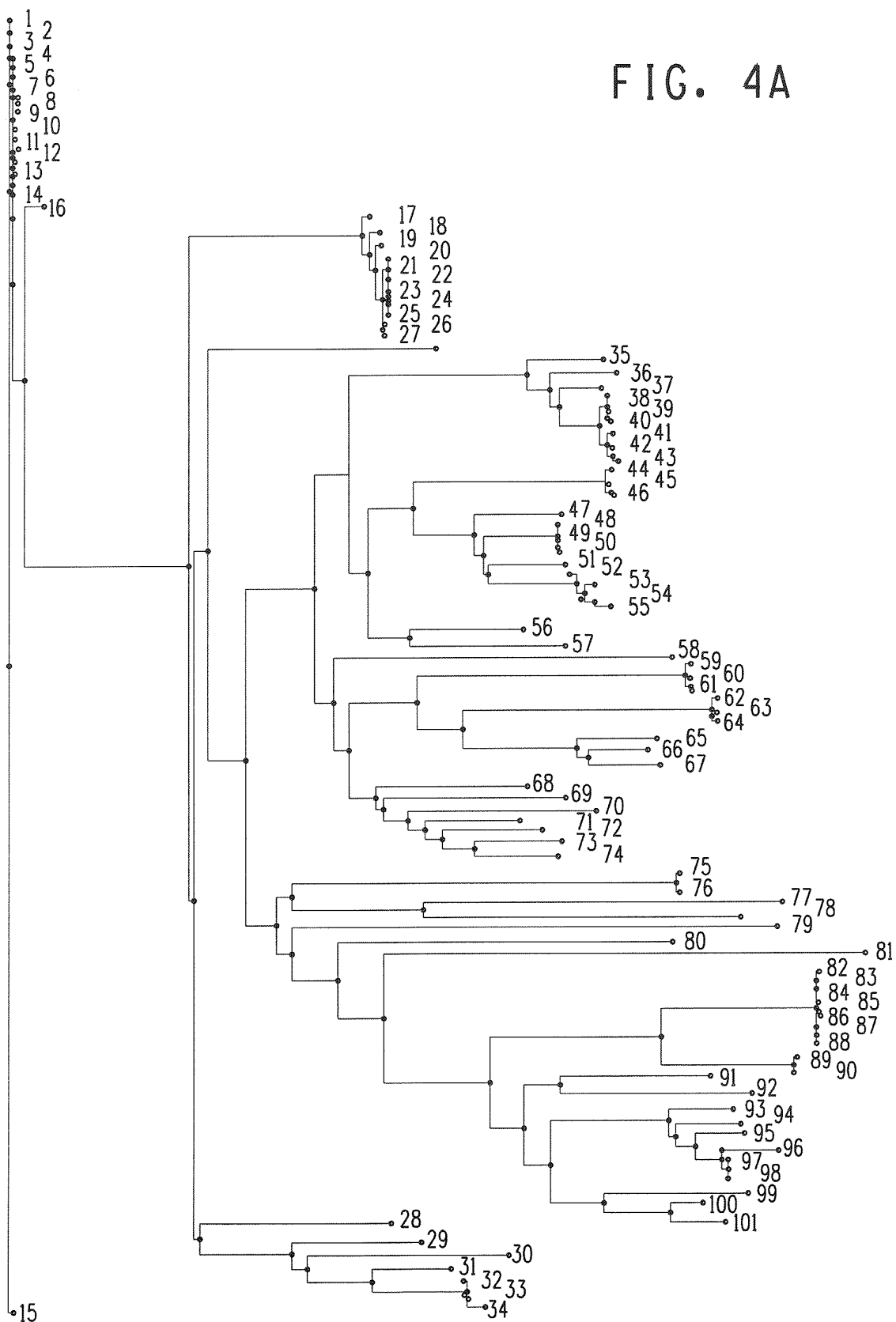


FIG. 3

FIG. 4A



- 1 *Bacillus cereus* E33L
- 2 *Bacillus thuringiensis* serovar konkukian str. 97-27
- 3 *Bacillus cereus* G9241
- 4 *Bacillus cereus* B4264
- 5 *Bacillus cereus* AH1134
- 6 *Bacillus cereus* ATCC 14579
- 7 *Bacillus cereus* G9842
- 8 *Bacillus thuringiensis* serovar israelensis ATCC 35646
- 9 *Bacillus cereus* 03BB108
- 10 *Bacillus cereus* NVH0597-99
- 11 *Bacillus thuringiensis* str. AlHakam
- 12 *Bacillus anthracis* str. Ames
- 13 *Bacillus cereus* W
- 14 *Bacillus cereus* AH187
- 15 *Bacillus cereus* ATCC 10987
- 16 *Bacillus weihenstephanensis* KBAB4
- 17 *Listeria monocytogenes* F5L J1-208
- 18 *Listeria welshimeri* serovar 6b str. 5LCC5334
- 19 *Listeria innocua* clip 11262
- 20 *Listeria monocytogenes* F5L J1-175
- 21 *Listeria monocytogenes* str. 4b H7858
- 22 *Listeria monocytogenes* str. 4b F2365
- 23 *Listeria monocytogenes* HP B2262
- 24 *Listeria monocytogenes* F5L N1-017
- 25 *Listeria monocytogenes* str. 1/2a F6854
- 26 *Listeria monocytogenes* EDG-e
- 27 *Exiguobacterium sibiricum* 255-15
- 28 *Bacillus coagulans* 36D1
- 29 *Bacillus licheniformis* ATCC 14580
- 30 *Bacillus pumilus* 5A FR-032
- 31 *Bacillus amyloliquefaciens* FZB42
- 32 Unknown
- 33 *Bacillus subtilis* subsp. *subtilis* str. 168
- 34 *Bacillus subtilis*
- 35 *Staphylococcus haemolyticus* JC5C1435
- 36 *Staphylococcus epidermidis* ATCC 12228
- 37 *Staphylococcus saprophyticus* subsp. *saprophyticus* ATCC 15305
- 38 *Staphylococcus aureus* subsp. *aureus* Mu 50
- 39 *Staphylococcus aureus* subsp. *aureus* MR5A252
- 40 *Staphylococcus aureus* subsp. *aureus* NCTC 8325
- 41 *Staphylococcus aureus*
- 42 *Staphylococcus aureus* subsp. *aureus* MW2
- 43 *Staphylococcus aureus* RF122
- 44 *Lactococcus lactis* subsp. *lactis* il1403
- 45 *Lactococcus lactis* subsp. *cremoris* MG 1363

FIG. 4B

- 46 *Lactococcus lactis* subsp. *cremoris* 5K11
- 47 *Streptococcus infantarius* subsp. *infantarius* ATCC BAA-102
- 48 *Streptococcus agalactiae* CDH1
- 49 *Streptococcus agalactiae* 2603V/R
- 50 *Streptococcus agalactiae* NEM316
- 51 *Streptococcus mutans* UA 159
- 52 *Streptococcus thermophilus* LMD-9
- 53 *Streptococcus thermophilus* LMG 18311
- 54 *Streptococcus thermophilus* CNRZ1066
- 55 *Streptococcus thermophilus*
- 56 *Enterococcus faecium* DQ
- 57 *Enterococcus faecalis* V583
- 58 *Lactobacillus brevis* ATCC 367
- 59 *Oenococcus oeni*
- 60 *Oenococcus oeni* P 5U-1
- 61 *Oenococcus oeni* ATCC BAA-1163
- 62 *Lactobacillus reuteri* F275
- 63 *Lactobacillus reuteri*
- 64 *Lactobacillus reuteri* 100-23
- 65 *Leuconostoc mesenteroides* subsp. *mesenteroides* ATCC 8293
- 66 *Leuconostoc lactis*
- 67 *Leuconostoc citreum* KM20
- 68 *Lactobacillus casei* ATCC 334
- 69 *Pediococcus pentosaceus* ATCC 25745
- 70 *Lactobacillus fermentum* IFO 3956
- 71 *Lactobacillus plantarum* WCF51
- 72 *Lactobacillus sakei* subsp. *sakei* 23K
- 73 *Lactobacillus salivarius* UCC118
- 74 *Lactobacillus johnsonii* NCC 533
- 75 *Mycobacterium marinum*
- 76 *Mycobacterium ulcerans* Agy99
- 77 *Magnaporthe grisea* 70-15
- 78 *Phaeosphaeria nodorum* 5N15
- 79 *Methylococcus capsulatus* str. Bath
- 80 *Vibrio angustum* 514
- 81 *Synechococcus* sp. CC9605
- 82 *Vibrio cholerae* 1587
- 83 *Vibrio cholerae* AM-19226
- 84 *Vibrio cholerae* 623-39
- 85 *Vibrio cholerae* O1 biovar *el Tor* str. N1...
- 86 *Vibrio cholerae* V51
- 87 *Vibrio cholerae* 2740-80
- 88 *Vibrio cholerae* V52
- 89 *Vibrio alginolyticus* 12G01
- 90 *Vibrio* sp. Ex25

FIG. 4C

- 91 *Serratia protoemaculans* 568
- 92 *Aeromonas hydrophila* subsp. *hydrophila* ATCC 7966
- 93 *Enterobacter* sp. 638
- 94 *Enterobacter sakazakii* ATCC BAA-894
- 95 *Raoultella terrigena*
- 96 *Klebsiella pneumoniae* subsp. *pneumoniae* MGH 78578
- 97 *Klebsiella pneumoniae*
- 98 *Klebsiella pneumoniae*
- 99 *Pectobacterium atrosepticum* 5CRI1043
- 100 *Versinia intermedia* ATCC 29909
- 101 *Versinia enterocolitica* subsp. *enterocolitica* 8081

Fig. 5

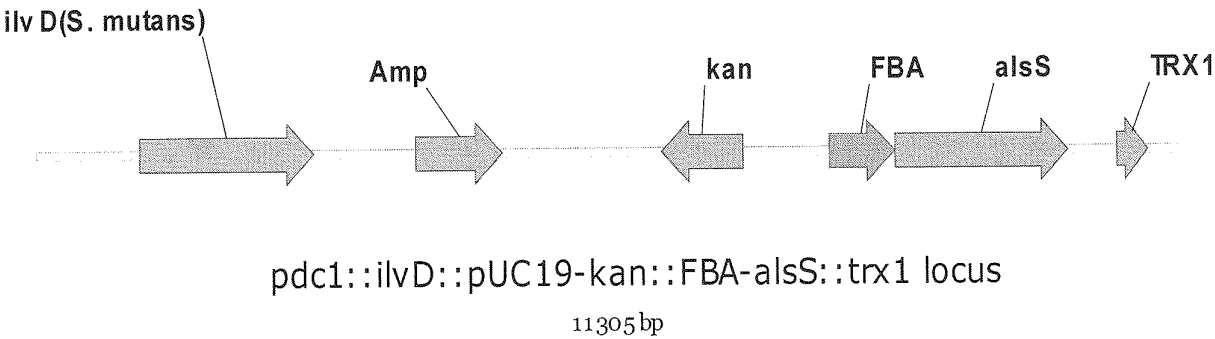
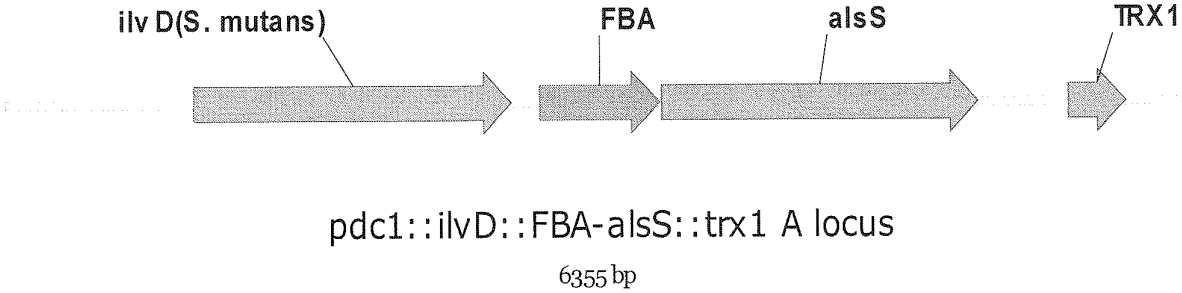


Fig. 6



20110907_CL5178USNA_SeqList.txt
SEQUENCE LISTING

<110> Butamax (TM) Advanced Biofuels
Maggio-Hall, Lori

<120> Integration of a Polynucleotide Encoding a Polypeptide That
Catalyzes Pyruvate to Acetolactate Conversion

<130> CL5178

<150> US 61/380,563
<151> 2010-09-07

<150> US 61/466,557
<151> 2011-03-23

<160> 248

<170> PatentIn version 3.5

<210> 1
<211> 1680
<212> DNA
<213> K. pneumoniae

<400> 1
atggacaaac agtatccggt acgccagtgg gcgcacggcg ccgatctcgt cgtcagtcag 60
ctggaagctc agggagtacg ccaggtgttc ggcattcccc gcgccaaaat cgacaaggtc 120
tttgattcac tgctggattc ctccattcgc attattccgg tacgccacga agccaacgcc 180
gcatttatgg ccgccgccgt cggacgcatt accggcaaag cgggcgtggc gctggtcacc 240
tccggtccgg gctgttccaa cctgatcacc ggcattggcca ccgcgaacag cgaaggcgac 300
ccggtggtgg ccctgggcgg cgcggtaaaa cgcgccgata aagcgaagca ggtccaccag 360
agtatggata cgggtggcgt gttcagcccc gtcaccaa atcgccatcga ggtgacggcg 420
ccggatgctc tggcggaagt ggtctccaac gccttccgcg ccgccgagca gggccggccg 480
ggcagcgcgt tcgttagcct gccgcaggat gtggtcgatg gcccggtcag cggcaaagtg 540
ctgccggcca gcggggcccc gcagatgggc gccgcgccgg atgatgccat cgaccagggtg 600
gcgaagctta tcgcccaggc gaagaacccg atcttctcgc tcggcctgat ggccagccag 660
ccggaaaaca gcaaggcgct gcgccgtttg ctggagacca gccatattcc agtcaccagc 720
acctatcagg ccgccggagc ggtgaatcag gataacttct ctgccttcgc cggccggggtt 780
gggctgttta acaaccaggc cggggaccgt ctgctgcagc tcgccgacct ggtgatctgc 840
atcggctaca gcccggtgga atacgaaccg gcgatgtgga acagcggcaa cgcgacgctg 900
gtgcacatcg acgtgctgcc cgcctatgaa gagcgcaact acaccccgga tgtcgagctg 960
gtgggcgata tcgccggcac tctcaacaag ctggcgcaaa atatcgatca tcggctgggtg 1020
ctctccccgc aggcggcgga gatcctccgc gaccgccagc accagcgcga gctgctggac 1080
cgccgcggcg cgcagctcaa ccagtttgcc ctgcatcccc tgcgcatcgt tcgcgccatg 1140
caggatatcg tcaacagcga cgtcacgttg accgtggaca tgggcagctt ccatatctgg 1200
attgcccgtt acctgtacac gttccgcgcc cgtcaggtga tgatctccaa cggccagcag 1260

20110907_CL5178USNA_SeqList.txt

accatgggcg tcgccctgcc ctgggctatc ggcgcttggc tggatcaatcc tgagcgcaaa 1320
 gtggtctccg tctccggcga cggcggcttc ctgcagtcga gcatggagct ggagaccgcc 1380
 gtccgcctga aagccaacgt gctgcatctt atctgggtcg ataacggcta caacatggtc 1440
 gctatccagg aagagaaaaa atatcagcgc ctgtccggcg tcgagtttgg gccgatggat 1500
 tttaaagcct atgccgaatc cttcggcgcg aaagggtttg ccgtggaaag cgccgaggcg 1560
 ctggagccga ccctgcgcgc gccgatggac gtcgacggcc cggcggtagt ggccatcccg 1620
 gtggattatc gcgataaccc gctgctgatg ggccagctgc atctgagtca gattctgtaa 1680

<210> 2
 <211> 559
 <212> PRT
 <213> K. pneumoniae

<400> 2

Met Asp Lys Gln Tyr Pro Val Arg Gln Trp Ala His Gly Ala Asp Leu
1 5 10 15

Val Val Ser Gln Leu Glu Ala Gln Gly Val Arg Gln Val Phe Gly Ile
20 25 30

Pro Gly Ala Lys Ile Asp Lys Val Phe Asp Ser Leu Leu Asp Ser Ser
35 40 45

Ile Arg Ile Ile Pro Val Arg His Glu Ala Asn Ala Ala Phe Met Ala
50 55 60

Ala Ala Val Gly Arg Ile Thr Gly Lys Ala Gly Val Ala Leu Val Thr
65 70 75 80

Ser Gly Pro Gly Cys Ser Asn Leu Ile Thr Gly Met Ala Thr Ala Asn
85 90 95

Ser Glu Gly Asp Pro Val Val Ala Leu Gly Gly Ala Val Lys Arg Ala
100 105 110

Asp Lys Ala Lys Gln Val His Gln Ser Met Asp Thr Val Ala Met Phe
115 120 125

Ser Pro Val Thr Lys Tyr Ala Ile Glu Val Thr Ala Pro Asp Ala Leu
130 135 140

Ala Glu Val Val Ser Asn Ala Phe Arg Ala Ala Glu Gln Gly Arg Pro
145 150 155 160

Gly Ser Ala Phe Val Ser Leu Pro Gln Asp Val Val Asp Gly Pro Val
165 170 175

Ser Gly Lys Val Leu Pro Ala Ser Gly Ala Pro Gln Met Gly Ala Ala
180 185 190

20110907_CL5178USNA_SeqList.txt

Pro Asp Asp Ala Ile Asp Gln Val Ala Lys Leu Ile Ala Gln Ala Lys
195 200 205

Asn Pro Ile Phe Leu Leu Gly Leu Met Ala Ser Gln Pro Glu Asn Ser
210 215 220

Lys Ala Leu Arg Arg Leu Leu Glu Thr Ser His Ile Pro Val Thr Ser
225 230 235 240

Thr Tyr Gln Ala Ala Gly Ala Val Asn Gln Asp Asn Phe Ser Arg Phe
245 250 255

Ala Gly Arg Val Gly Leu Phe Asn Asn Gln Ala Gly Asp Arg Leu Leu
260 265 270

Gln Leu Ala Asp Leu Val Ile Cys Ile Gly Tyr Ser Pro Val Glu Tyr
275 280 285

Glu Pro Ala Met Trp Asn Ser Gly Asn Ala Thr Leu Val His Ile Asp
290 295 300

Val Leu Pro Ala Tyr Glu Glu Arg Asn Tyr Thr Pro Asp Val Glu Leu
305 310 315 320

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

His Arg Leu Val Leu Ser Pro Gln Ala Ala Glu Ile Leu Arg Asp Arg
340 345 350

Gln His Gln Arg Glu Leu Leu Asp Arg Arg Gly Ala Gln Leu Asn Gln
355 360 365

Phe Ala Leu His Pro Leu Arg Ile Val Arg Ala Met Gln Asp Ile Val
370 375 380

Asn Ser Asp Val Thr Leu Thr Val Asp Met Gly Ser Phe His Ile Trp
385 390 395 400

Ile Ala Arg Tyr Leu Tyr Thr Phe Arg Ala Arg Gln Val Met Ile Ser
405 410 415

Asn Gly Gln Gln Thr Met Gly Val Ala Leu Pro Trp Ala Ile Gly Ala
420 425 430

Trp Leu Val Asn Pro Glu Arg Lys Val Val Ser Val Ser Gly Asp Gly
435 440 445

Gly Phe Leu Gln Ser Ser Met Glu Leu Glu Thr Ala Val Arg Leu Lys
450 455 460

20110907_CL5178USNA_SeqList.txt

Ala Asn Val Leu His Leu Ile Trp Val Asp Asn Gly Tyr Asn Met Val
465 470 475 480

Ala Ile Gln Glu Glu Lys Lys Tyr Gln Arg Leu Ser Gly Val Glu Phe
485 490 495

Gly Pro Met Asp Phe Lys Ala Tyr Ala Glu Ser Phe Gly Ala Lys Gly
500 505 510

Phe Ala Val Glu Ser Ala Glu Ala Leu Glu Pro Thr Leu Arg Ala Ala
515 520 525

Met Asp Val Asp Gly Pro Ala Val Val Ala Ile Pro Val Asp Tyr Arg
530 535 540

Asp Asn Pro Leu Leu Met Gly Gln Leu His Leu Ser Gln Ile Leu
545 550 555

<210> 3
<211> 1716
<212> DNA
<213> Bacillus subtilis

<400> 3
atgttgacaa aagcaacaaa agaacaaaaa tcccttgtga aaaacagagg ggcggagctt 60
gttgttgatt gcttagtgga gcaaggtgtc acacatgtat ttggcattcc aggtgcaaaa 120
attgatgcgg tatttgacgc tttaacaagat aaaggacctg aaattatcgt tgcccggcac 180
gaacaaaacg cagcattcat ggccaagca gtcggccgtt taactggaaa accgggagtc 240
gtgttagtca catcaggacc ggggtgcctct aacttggcaa caggcctgct gacagcgaac 300
actgaaggag accctgtcgt tgcgcttgct ggaaacgtga tccgtgcaga tcgtttaaaa 360
cggacacatc aatcttttga taatgcggcg ctattccagc cgattacaaa atacagtgtg 420
gaagttcaag atgtaaaaaa tataccggaa gctgtttaca atgcatttag gatagcgtca 480
gcagggcagg ctggggccgc ttttgtgagc tttccgcaag atgttgtgaa tgaagtcaca 540
aatacgaaaa acgtgcgtgc tgttgacagc caaaactcg gtcctgcagc agatgatgca 600
atcagtgcgg ccatagcaaa aatccaaaca gcaaaacttc ctgtcgtttt ggtcggcatg 660
aaaggcggaa gaccggaagc aattaaagcg gttcgcaagc ttttgaaaaa ggttcagctt 720
ccatttgttg aaacatatca agctgccggt acccttttcta gagatttaga ggatcaatat 780
tttggccgta tcggtttgtt ccgcaaccag cctggcgatt tactgctaga gcaggcagat 840
gttgttctga cgatcggcta tgacccgatt gaatatgatc cgaaattctg gaatatcaat 900
ggagaccgga caattatcca tttagacgag attatcgctg acattgatca tgcttaccag 960
cctgatcttg aattgatcgg tgacattccg tccacgatca atcatatcga acacgatgct 1020
gtgaaagtgg aatttgcaga gcgtgagcag aaaatccttt ctgattttaa acaatatatg 1080

20110907_CL5178USNA_SeqList.txt

catgaaggtg agcaggtgcc tgcagattgg aaatcagaca gagcgcaccc tcttgaaatc	1140
gttaaagagt tgcgtaatgc agtcgatgat catgttacag taacttgcca tatcggttcg	1200
cacgccattt ggatgtcacg ttatttccgc agctacgagc cgттаacatt aatgatcagt	1260
aacggtatgc aaacactcgg cgttgcgctt ccttgggcaa tcggcgcttc attggtgaaa	1320
ccgggagaaa aagtggtttc tgtctctggt gacggcggtt tcttattctc agcaatggaa	1380
ttagagacag cagttcgact aaaagcacca attgtacaca ttgtatggaa cgacagcaca	1440
tatgacatgg ttgcattcca gcaattgaaa aaatataacc gtacatctgc ggtcgatttc	1500
ggaaatatcg atatcgtgaa atatcgcgaa agcttcggag caactggctt gcgcgtagaa	1560
tcaccagacc agctggcaga tgttctgcgt caaggcatga acgctgaagg tcctgtcatc	1620
atcgatgtcc cggttgacta cagtgataac attaatttag caagtgacaa gcttccgaaa	1680
gaattcgggg aactcatgaa aacgaaagct ctctag	1716

<210> 4
 <211> 571
 <212> PRT
 <213> Bacillus subtilis

<400> 4

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

Gly	Ala	Glu	Leu	Val	Val	Asp	Cys	Leu	Val	Glu	Gln	Gly	Val	Thr	His
			20					25					30		

Val	Phe	Gly	Ile	Pro	Gly	Ala	Lys	Ile	Asp	Ala	Val	Phe	Asp	Ala	Leu
		35					40					45			

Gln	Asp	Lys	Gly	Pro	Glu	Ile	Ile	Val	Ala	Arg	His	Glu	Gln	Asn	Ala
	50					55					60				

Ala	Phe	Met	Ala	Gln	Ala	Val	Gly	Arg	Leu	Thr	Gly	Lys	Pro	Gly	Val
65					70					75					80

Val	Leu	Val	Thr	Ser	Gly	Pro	Gly	Ala	Ser	Asn	Leu	Ala	Thr	Gly	Leu
				85					90					95	

Leu	Thr	Ala	Asn	Thr	Glu	Gly	Asp	Pro	Val	Val	Ala	Leu	Ala	Gly	Asn
			100					105					110		

Val	Ile	Arg	Ala	Asp	Arg	Leu	Lys	Arg	Thr	His	Gln	Ser	Leu	Asp	Asn
		115					120					125			

Ala	Ala	Leu	Phe	Gln	Pro	Ile	Thr	Lys	Tyr	Ser	Val	Glu	Val	Gln	Asp
	130					135					140				

Val	Lys	Asn	Ile	Pro	Glu	Ala	Val	Thr	Asn	Ala	Phe	Arg	Ile	Ala	Ser
145					150					155					160

20110907_CL5178USNA_SeqList.txt

Ala Gly Gln Ala Gly Ala Ala Phe Val Ser Phe Pro Gln Asp Val Val
165 170 175

Asn Glu Val Thr Asn Thr Lys Asn Val Arg Ala Val Ala Ala Pro Lys
180 185 190

Leu Gly Pro Ala Ala Asp Asp Ala Ile Ser Ala Ala Ile Ala Lys Ile
195 200 205

Gln Thr Ala Lys Leu Pro Val Val Leu Val Gly Met Lys Gly Gly Arg
210 215 220

Pro Glu Ala Ile Lys Ala Val Arg Lys Leu Leu Lys Lys Val Gln Leu
225 230 235 240

Pro Phe Val Glu Thr Tyr Gln Ala Ala Gly Thr Leu Ser Arg Asp Leu
245 250 255

Glu Asp Gln Tyr Phe Gly Arg Ile Gly Leu Phe Arg Asn Gln Pro Gly
260 265 270

Asp Leu Leu Leu Glu Gln Ala Asp Val Val Leu Thr Ile Gly Tyr Asp
275 280 285

Pro Ile Glu Tyr Asp Pro Lys Phe Trp Asn Ile Asn Gly Asp Arg Thr
290 295 300

Ile Ile His Leu Asp Glu Ile Ile Ala Asp Ile Asp His Ala Tyr Gln
305 310 315 320

Pro Asp Leu Glu Leu Ile Gly Asp Ile Pro Ser Thr Ile Asn His Ile
325 330 335

Glu His Asp Ala Val Lys Val Glu Phe Ala Glu Arg Glu Gln Lys Ile
340 345 350

Leu Ser Asp Leu Lys Gln Tyr Met His Glu Gly Glu Gln Val Pro Ala
355 360 365

Asp Trp Lys Ser Asp Arg Ala His Pro Leu Glu Ile Val Lys Glu Leu
370 375 380

Arg Asn Ala Val Asp Asp His Val Thr Val Thr Cys Asp Ile Gly Ser
385 390 395 400

His Ala Ile Trp Met Ser Arg Tyr Phe Arg Ser Tyr Glu Pro Leu Thr
405 410 415

Leu Met Ile Ser Asn Gly Met Gln Thr Leu Gly Val Ala Leu Pro Trp
420 425 430

20110907_CL5178USNA_SeqList.txt

Ala Ile Gly Ala Ser Leu Val Lys Pro Gly Glu Lys Val Val Ser Val
435 440 445

Ser Gly Asp Gly Gly Phe Leu Phe Ser Ala Met Glu Leu Glu Thr Ala
450 455 460

Val Arg Leu Lys Ala Pro Ile Val His Ile Val Trp Asn Asp Ser Thr
465 470 475 480

Tyr Asp Met Val Ala Phe Gln Gln Leu Lys Lys Tyr Asn Arg Thr Ser
485 490 495

Ala Val Asp Phe Gly Asn Ile Asp Ile Val Lys Tyr Ala Glu Ser Phe
500 505 510

Gly Ala Thr Gly Leu Arg Val Glu Ser Pro Asp Gln Leu Ala Asp Val
515 520 525

Leu Arg Gln Gly Met Asn Ala Glu Gly Pro Val Ile Ile Asp Val Pro
530 535 540

Val Asp Tyr Ser Asp Asn Ile Asn Leu Ala Ser Asp Lys Leu Pro Lys
545 550 555 560

Glu Phe Gly Glu Leu Met Lys Thr Lys Ala Leu
565 570

<210> 5
<211> 1665
<212> DNA
<213> Lactococcus lactis

<400> 5
atgtctgaga aacaatttgg ggcgaacttg gttgtcgata gtttgattaa ccataaagtg 60
aagtatgtat ttgggattcc aggagcaaaa attgaccggg tttttgattt attagaaaat 120
gaagaaggcc ctcaaattgg cgtgactcgt catgagcaag gagctgcttt catggctcaa 180
gctgtcggtc gtttaactgg cgaacctggg ttagtagttg ttacgagtgg gcctgggtgta 240
tcaaaccttg cgactccgct tttgaccgcg acatcagaag gtgatgctat tttggctatc 300
ggtggacaag ttaaacgaag tgaccgtctt aaacgtgcgc accaatcaat ggataatgct 360
ggaatgatgc aatcagcaac aaaatattca gcagaagttc ttgaccctaa tacactttct 420
gaatcaattg ccaacgctta tcgtattgca aaatcaggac atccagggtgc aactttctta 480
tcaatcccc aagatgtaac ggatgccgaa gtatcaatca aagccattca accactttca 540
gaccctaaaa tggggaatgc ctctattgat gacattaatt atttagcaca agcaattaa 600
aatgctgtat tgccagtaat tttggttggg gctggtgctt cagatgctaa agtcgcttca 660
tccttgcgta atctattgac tcatgttaat attcctgtcg ttgaaacatt ccaagggtgca 720

20110907_CL5178USNA_SeqList.txt

```

gggggttattt cacatgattt agaacatact ttttatggac gtatcgggtct tttccgcaat      780
caaccaggcg atatgcttct gaaacgttct gaccttggtta ttgctgttgg ttatgaccca      840
attgaatatg aagctcgtaa ctggaatgca gaaattgata gtcgaattat cgttattgat      900
aatgccattg ctgaaattga tacttactac caaccagagc gtgaattaat tggatgatatc      960
gcagcaacat tggataatct tttaccagct gttcgtggct acaaaattcc aaaaggaaca     1020
aaagattatc tcgatggcct tcatgaagtt gctgagcaac acgaatttga tactgaaaat     1080
actgaagaag gtagaatgca ccctcttgat ttggtcagca ctttccaaga aatcgtcaag     1140
gatgatgaaa cagtaaccgt tgacgtaggt tcaactctaca tttggatggc acgtcatttc     1200
aatcatatcg aaccacgtca tctcctcttc tcaaacggaa tgcaaact cggagttgca     1260
cttccttggg caattacagc cgcattgttg cgcccaggta aaaaagttaa ttcacactct     1320
ggtgatggag gcttcctttt cacagggcaa gaattggaaa cagctgtacg tttgaatctt     1380
ccaatcggtc aaattatctg gaatgacggc cattatgata tggttaaatt ccaagaagaa     1440
atgaaatatg gtcgttcagc agccgttgat tttggctatg ttgattacgt aaaatatgct     1500
gaagcaatga gagcaaaagg ttaccgtgca cacagcaaag aagaacttgc tgaaattctc     1560
aatcaatcc cagatactac tggaccggtg gtaattgacg ttcctttgga ctattctgat     1620
aacattaaat tagcagaaaa attattgcct gaagagtttt attga                       1665

```

```

<210> 6
<211> 554
<212> PRT
<213> Lactococcus lactis

```

<400> 6

```

Met Ser Glu Lys Gln Phe Gly Ala Asn Leu Val Val Asp Ser Leu Ile
1          5          10          15

```

```

Asn His Lys Val Lys Tyr Val Phe Gly Ile Pro Gly Ala Lys Ile Asp
20          25          30

```

```

Arg Val Phe Asp Leu Leu Glu Asn Glu Glu Gly Pro Gln Met Val Val
35          40          45

```

```

Thr Arg His Glu Gln Gly Ala Ala Phe Met Ala Gln Ala Val Gly Arg
50          55          60

```

```

Leu Thr Gly Glu Pro Gly Val Val Val Val Thr Ser Gly Pro Gly Val
65          70          75          80

```

```

Ser Asn Leu Ala Thr Pro Leu Leu Thr Ala Thr Ser Glu Gly Asp Ala
85          90          95

```

```

Ile Leu Ala Ile Gly Gly Gln Val Lys Arg Ser Asp Arg Leu Lys Arg
100          105          110

```


20110907_CL5178USNA_SeqList.txt

Ala His Gln Ser Met Asp Asn Ala Gly Met Met Gln Ser Ala Thr Lys
115 120 125

Tyr Ser Ala Glu Val Leu Asp Pro Asn Thr Leu Ser Glu Ser Ile Ala
130 135 140

Asn Ala Tyr Arg Ile Ala Lys Ser Gly His Pro Gly Ala Thr Phe Leu
145 150 155 160

Ser Ile Pro Gln Asp Val Thr Asp Ala Glu Val Ser Ile Lys Ala Ile
165 170 175

Gln Pro Leu Ser Asp Pro Lys Met Gly Asn Ala Ser Ile Asp Asp Ile
180 185 190

Asn Tyr Leu Ala Gln Ala Ile Lys Asn Ala Val Leu Pro Val Ile Leu
195 200 205

Val Gly Ala Gly Ala Ser Asp Ala Lys Val Ala Ser Ser Leu Arg Asn
210 215 220

Leu Leu Thr His Val Asn Ile Pro Val Val Glu Thr Phe Gln Gly Ala
225 230 235 240

Gly Val Ile Ser His Asp Leu Glu His Thr Phe Tyr Gly Arg Ile Gly
245 250 255

Leu Phe Arg Asn Gln Pro Gly Asp Met Leu Leu Lys Arg Ser Asp Leu
260 265 270

Val Ile Ala Val Gly Tyr Asp Pro Ile Glu Tyr Glu Ala Arg Asn Trp
275 280 285

Asn Ala Glu Ile Asp Ser Arg Ile Ile Val Ile Asp Asn Ala Ile Ala
290 295 300

Glu Ile Asp Thr Tyr Tyr Gln Pro Glu Arg Glu Leu Ile Gly Asp Ile
305 310 315 320

Ala Ala Thr Leu Asp Asn Leu Leu Pro Ala Val Arg Gly Tyr Lys Ile
325 330 335

Pro Lys Gly Thr Lys Asp Tyr Leu Asp Gly Leu His Glu Val Ala Glu
340 345 350

Gln His Glu Phe Asp Thr Glu Asn Thr Glu Glu Gly Arg Met His Pro
355 360 365

Leu Asp Leu Val Ser Thr Phe Gln Glu Ile Val Lys Asp Asp Glu Thr
370 375 380

20110907_CL5178USNA_SeqList.txt

Val Thr Val Asp Val Gly Ser Leu Tyr Ile Trp Met Ala Arg His Phe
385 390 395 400

Lys Ser Tyr Glu Pro Arg His Leu Leu Phe Ser Asn Gly Met Gln Thr
405 410 415

Leu Gly Val Ala Leu Pro Trp Ala Ile Thr Ala Ala Leu Leu Arg Pro
420 425 430

Gly Lys Lys Val Tyr Ser His Ser Gly Asp Gly Gly Phe Leu Phe Thr
435 440 445

Gly Gln Glu Leu Glu Thr Ala Val Arg Leu Asn Leu Pro Ile Val Gln
450 455 460

Ile Ile Trp Asn Asp Gly His Tyr Asp Met Val Lys Phe Gln Glu Glu
465 470 475 480

Met Lys Tyr Gly Arg Ser Ala Ala Val Asp Phe Gly Tyr Val Asp Tyr
485 490 495

Val Lys Tyr Ala Glu Ala Met Arg Ala Lys Gly Tyr Arg Ala His Ser
500 505 510

Lys Glu Glu Leu Ala Glu Ile Leu Lys Ser Ile Pro Asp Thr Thr Gly
515 520 525

Pro Val Val Ile Asp Val Pro Leu Asp Tyr Ser Asp Asn Ile Lys Leu
530 535 540

Ala Glu Lys Leu Leu Pro Glu Glu Phe Tyr
545 550

<210> 7
<211> 1665
<212> DNA
<213> Staphylococcus aureus

<220>
<221> CDS
<222> (1)..(1665)

<400> 7
atg act gat aaa aag tac act gca gcc gat atg gtt att gat act ttg 48
Met Thr Asp Lys Lys Tyr Thr Ala Ala Asp Met Val Ile Asp Thr Leu
1 5 10 15

aaa aat aat ggg gta gaa tat gtt ttt ggt att ccg ggt gca aag ata 96
Lys Asn Asn Gly Val Glu Tyr Val Phe Gly Ile Pro Gly Ala Lys Ile
20 25 30

gac tat cta ttt aat gct tta att gat gat ggt cct gaa ctt att gtc 144
Asp Tyr Leu Phe Asn Ala Leu Ile Asp Asp Gly Pro Glu Leu Ile Val
35 40 45

act cgt cat gaa caa aat gct gca atg atg gca caa ggt att gga aga 192

20110907_CL5178USNA_SeqList.txt

Thr	Arg	His	Glu	Gln	Asn	Ala	Ala	Met	Met	Ala	Gln	Gly	Ile	Gly	Arg	
50						55					60					
tta	aca	ggg	aaa	ccg	ggg	gta	gta	ctt	gtt	aca	agt	ggc	cct	ggg	gta	240
Leu	Thr	Gly	Lys	Pro	Gly	Val	Val	Leu	Val	Thr	Ser	Gly	Pro	Gly	Val	
65					70					75					80	
agt	aat	tta	acg	act	gga	cta	tta	aca	gct	aca	tct	gaa	ggg	gat	cct	288
Ser	Asn	Leu	Thr	Thr	Gly	Leu	Leu	Thr	Ala	Thr	Ser	Glu	Gly	Asp	Pro	
				85					90					95		
gta	tta	gcg	tta	ggg	ggc	caa	gtg	aaa	cgt	aat	gat	tta	tta	cga	tta	336
Val	Leu	Ala	Leu	Gly	Gly	Gln	Val	Lys	Arg	Asn	Asp	Leu	Leu	Arg	Leu	
			100					105					110			
acg	cat	caa	agt	att	gat	aat	gct	gcg	cta	tta	aaa	tat	tca	tca	aaa	384
Thr	His	Gln	Ser	Ile	Asp	Asn	Ala	Ala	Leu	Leu	Lys	Tyr	Ser	Ser	Lys	
		115					120					125				
tac	agt	gaa	gaa	gta	caa	gat	cct	gaa	tca	tta	tca	gaa	gtt	atg	aca	432
Tyr	Ser	Glu	Glu	Val	Gln	Asp	Pro	Glu	Ser	Leu	Ser	Glu	Val	Met	Thr	
	130					135					140					
aat	gca	att	cga	att	gct	act	tca	gga	aaa	aat	ggc	gca	agt	ttt	att	480
Asn	Ala	Ile	Arg	Ile	Ala	Thr	Ser	Gly	Lys	Asn	Gly	Ala	Ser	Phe	Ile	
145					150					155					160	
agt	att	ccg	caa	gac	gtt	att	tct	tca	cca	gtt	gaa	tct	aaa	gct	ata	528
Ser	Ile	Pro	Gln	Asp	Val	Ile	Ser	Ser	Pro	Val	Glu	Ser	Lys	Ala	Ile	
				165					170					175		
tca	ctt	tcg	caa	aaa	cca	aat	tta	gga	gta	ccg	agt	gaa	caa	gat	att	576
Ser	Leu	Cys	Gln	Lys	Pro	Asn	Leu	Gly	Val	Pro	Ser	Glu	Gln	Asp	Ile	
			180					185					190			
aat	gat	gtc	att	gaa	gcg	att	aaa	aat	gca	tca	ttt	cct	gtt	tta	tta	624
Asn	Asp	Val	Ile	Glu	Ala	Ile	Lys	Asn	Ala	Ser	Phe	Pro	Val	Leu	Leu	
		195					200					205				
gct	ggg	atg	aga	agt	tca	agt	gca	gaa	gaa	aca	aat	gcc	att	cgc	aaa	672
Ala	Gly	Met	Arg	Ser	Ser	Ser	Ala	Glu	Glu	Thr	Asn	Ala	Ile	Arg	Lys	
	210					215					220					
tta	gtt	gag	cgc	acg	aat	tta	cca	gtt	gta	gaa	aca	ttc	caa	ggg	gca	720
Leu	Val	Glu	Arg	Thr	Asn	Leu	Pro	Val	Val	Glu	Thr	Phe	Gln	Gly	Ala	
225					230					235					240	
ggg	gta	att	agt	cgt	gaa	tta	gaa	aat	cat	ttc	ttc	ggg	cgt	gtg	ggc	768
Gly	Val	Ile	Ser	Arg	Glu	Leu	Glu	Asn	His	Phe	Phe	Gly	Arg	Val	Gly	
				245					250					255		
tta	ttc	cgc	aat	caa	gtt	ggg	gat	gaa	tta	tta	cgt	aaa	agt	gat	tta	816
Leu	Phe	Arg	Asn	Gln	Val	Gly	Asp	Glu	Leu	Leu	Arg	Lys	Ser	Asp	Leu	
			260					265					270			
gtt	gtt	aca	atc	ggg	tat	gat	cca	att	gaa	tac	gaa	gct	agt	aac	tgg	864
Val	Val	Thr	Ile	Gly	Tyr	Asp	Pro	Ile	Glu	Tyr	Glu	Ala	Ser	Asn	Trp	
		275					280					285				
aat	aaa	gaa	tta	gaa	aca	caa	att	atc	aat	att	gac	gaa	gtt	caa	gct	912
Asn	Lys	Glu	Leu	Glu	Thr	Gln	Ile	Ile	Asn	Ile	Asp	Glu	Val	Gln	Ala	
	290					295					300					
gaa	att	act	aat	tat	atg	caa	ccg	aaa	aaa	gag	ttg	att	ggg	aat	att	960
Glu	Ile	Thr	Asn	Tyr	Met	Gln	Pro	Lys	Lys	Glu	Leu	Ile	Gly	Asn	Ile	
305					310					315					320	
gct	aaa	acg	att	gaa	atg	att	tct	gaa	aaa	gtg	gat	gag	cca	ttt	ata	1008

20110907_CL5178USNA_SeqList.txt

Ala	Lys	Thr	Ile	Glu	Met	Ile	Ser	Glu	Lys	Val	Asp	Glu	Pro	Phe	Ile		
				325					330					335			
aat	caa	caa	cat	tta	gac	gaa	tta	gaa	caa	tta	aga	aca	cat	att	gat		1056
Asn	Gln	Gln	His	Leu	Asp	Glu	Leu	Glu	Gln	Leu	Arg	Thr	His	Ile	Asp		
			340					345					350				
gaa	gaa	act	ggt	att	aaa	gcg	acg	cat	gaa	gaa	gga	att	cta	cat	cca		1104
Glu	Glu	Thr	Gly	Ile	Lys	Ala	Thr	His	Glu	Glu	Gly	Ile	Leu	His	Pro		
		355					360					365					
gtg	gaa	att	att	gaa	tct	atg	caa	aag	gta	tta	act	gat	gat	act	act		1152
Val	Glu	Ile	Ile	Glu	Ser	Met	Gln	Lys	Val	Leu	Thr	Asp	Asp	Thr	Thr		
	370					375					380						
gta	aca	gtt	gat	gtt	gga	agt	cac	tat	att	tgg	atg	gca	cgt	aat	ttc		1200
Val	Thr	Val	Asp	Val	Gly	Ser	His	Tyr	Ile	Trp	Met	Ala	Arg	Asn	Phe		
	385				390					395					400		
aga	agt	tac	aat	cca	aga	cat	tta	tta	ttt	agc	aat	ggt	atg	caa	acg		1248
Arg	Ser	Tyr	Asn	Pro	Arg	His	Leu	Leu	Phe	Ser	Asn	Gly	Met	Gln	Thr		
				405					410					415			
ctt	ggt	gta	gca	tta	ccg	tgg	gca	att	tca	gct	gca	ctt	gtg	cgc	cct		1296
Leu	Gly	Val	Ala	Leu	Pro	Trp	Ala	Ile	Ser	Ala	Ala	Leu	Val	Arg	Pro		
			420					425					430				
aat	acg	caa	gtt	gtg	tcc	gtt	gct	ggc	gat	ggt	ggc	ttt	tta	ttt	tca		1344
Asn	Thr	Gln	Val	Val	Ser	Val	Ala	Gly	Asp	Gly	Gly	Phe	Leu	Phe	Ser		
		435					440					445					
tca	caa	gat	tta	gaa	acg	gcc	gta	cgt	aaa	aat	tta	aat	atc	atc	cag		1392
Ser	Gln	Asp	Leu	Glu	Thr	Ala	Val	Arg	Lys	Asn	Leu	Asn	Ile	Ile	Gln		
	450					455					460						
ctt	att	tgg	aat	gat	gga	aaa	tat	aac	atg	gtt	gaa	ttc	caa	gaa	gaa		1440
Leu	Ile	Trp	Asn	Asp	Gly	Lys	Tyr	Asn	Met	Val	Glu	Phe	Gln	Glu	Glu		
	465				470					475					480		
atg	aaa	tat	aaa	cgt	tcg	tca	ggt	gta	gac	ttc	ggt	cct	gta	gat	ttt		1488
Met	Lys	Tyr	Lys	Arg	Ser	Ser	Gly	Val	Asp	Phe	Gly	Pro	Val	Asp	Phe		
				485					490					495			
gta	aaa	tat	gca	gaa	tca	ttt	ggc	gcg	aaa	ggt	tta	cga	gtt	act	aat		1536
Val	Lys	Tyr	Ala	Glu	Ser	Phe	Gly	Ala	Lys	Gly	Leu	Arg	Val	Thr	Asn		
			500				505						510				
caa	gaa	gaa	tta	gaa	gca	gca	att	aaa	gag	ggc	tat	gaa	aca	gat	ggt		1584
Gln	Glu	Glu	Leu	Glu	Ala	Ala	Ile	Lys	Glu	Gly	Tyr	Glu	Thr	Asp	Gly		
		515					520					525					
cca	gta	tta	att	gat	ata	cct	gta	aat	tac	aaa	gat	aat	atc	aaa	ctt		1632
Pro	Val	Leu	Ile	Asp	Ile	Pro	Val	Asn	Tyr	Lys	Asp	Asn	Ile	Lys	Leu		
	530					535					540						
tca	aca	aat	atg	tta	cct	gac	gta	ttt	aac	taa							1665
Ser	Thr	Asn	Met	Leu	Pro	Asp	Val	Phe	Asn								
	545				550												

<210> 8

<211> 554

<212> PRT

<213> Staphylococcus aureus

<400> 8

Met Thr Asp Lys Lys Tyr Thr Ala Ala Asp Met Val Ile Asp Thr Leu

1 5 10 15
 Lys Asn Asn Gly Val Glu Tyr Val Phe Gly Ile Pro Gly Ala Lys Ile
 20 25 30
 Asp Tyr Leu Phe Asn Ala Leu Ile Asp Asp Gly Pro Glu Leu Ile Val
 35 40 45
 Thr Arg His Glu Gln Asn Ala Ala Met Met Ala Gln Gly Ile Gly Arg
 50 55 60
 Leu Thr Gly Lys Pro Gly Val Val Leu Val Thr Ser Gly Pro Gly Val
 65 70 75 80
 Ser Asn Leu Thr Thr Gly Leu Leu Thr Ala Thr Ser Glu Gly Asp Pro
 85 90 95
 Val Leu Ala Leu Gly Gly Gln Val Lys Arg Asn Asp Leu Leu Arg Leu
 100 105 110
 Thr His Gln Ser Ile Asp Asn Ala Ala Leu Leu Lys Tyr Ser Ser Lys
 115 120 125
 Tyr Ser Glu Glu Val Gln Asp Pro Glu Ser Leu Ser Glu Val Met Thr
 130 135 140
 Asn Ala Ile Arg Ile Ala Thr Ser Gly Lys Asn Gly Ala Ser Phe Ile
 145 150 155 160
 Ser Ile Pro Gln Asp Val Ile Ser Ser Pro Val Glu Ser Lys Ala Ile
 165 170 175
 Ser Leu Cys Gln Lys Pro Asn Leu Gly Val Pro Ser Glu Gln Asp Ile
 180 185 190
 Asn Asp Val Ile Glu Ala Ile Lys Asn Ala Ser Phe Pro Val Leu Leu
 195 200 205
 Ala Gly Met Arg Ser Ser Ser Ala Glu Glu Thr Asn Ala Ile Arg Lys
 210 215 220
 Leu Val Glu Arg Thr Asn Leu Pro Val Val Glu Thr Phe Gln Gly Ala
 225 230 235 240
 Gly Val Ile Ser Arg Glu Leu Glu Asn His Phe Phe Gly Arg Val Gly
 245 250 255
 Leu Phe Arg Asn Gln Val Gly Asp Glu Leu Leu Arg Lys Ser Asp Leu
 260 265 270
 Val Val Thr Ile Gly Tyr Asp Pro Ile Glu Tyr Glu Ala Ser Asn Trp

275

280

285

Asn Lys Glu Leu Glu Thr Gln Ile Ile Asn Ile Asp Glu Val Gln Ala
 290 295 300

Glu Ile Thr Asn Tyr Met Gln Pro Lys Lys Glu Leu Ile Gly Asn Ile
 305 310 315 320

Ala Lys Thr Ile Glu Met Ile Ser Glu Lys Val Asp Glu Pro Phe Ile
 325 330 335

Asn Gln Gln His Leu Asp Glu Leu Glu Gln Leu Arg Thr His Ile Asp
 340 345 350

Glu Glu Thr Gly Ile Lys Ala Thr His Glu Glu Gly Ile Leu His Pro
 355 360 365

Val Glu Ile Ile Glu Ser Met Gln Lys Val Leu Thr Asp Asp Thr Thr
 370 375 380

Val Thr Val Asp Val Gly Ser His Tyr Ile Trp Met Ala Arg Asn Phe
 385 390 395 400

Arg Ser Tyr Asn Pro Arg His Leu Leu Phe Ser Asn Gly Met Gln Thr
 405 410 415

Leu Gly Val Ala Leu Pro Trp Ala Ile Ser Ala Ala Leu Val Arg Pro
 420 425 430

Asn Thr Gln Val Val Ser Val Ala Gly Asp Gly Gly Phe Leu Phe Ser
 435 440 445

Ser Gln Asp Leu Glu Thr Ala Val Arg Lys Asn Leu Asn Ile Ile Gln
 450 455 460

Leu Ile Trp Asn Asp Gly Lys Tyr Asn Met Val Glu Phe Gln Glu Glu
 465 470 475 480

Met Lys Tyr Lys Arg Ser Ser Gly Val Asp Phe Gly Pro Val Asp Phe
 485 490 495

Val Lys Tyr Ala Glu Ser Phe Gly Ala Lys Gly Leu Arg Val Thr Asn
 500 505 510

Gln Glu Glu Leu Glu Ala Ala Ile Lys Glu Gly Tyr Glu Thr Asp Gly
 515 520 525

Pro Val Leu Ile Asp Ile Pro Val Asn Tyr Lys Asp Asn Ile Lys Leu
 530 535 540

Ser Thr Asn Met Leu Pro Asp Val Phe Asn

545

550

<210> 9
 <211> 1698
 <212> DNA
 <213> *Listeria monocytogenes*

<220>
 <221> CDS
 <222> (1)..(1698)

<400> 9
 atg gcg aaa cta gaa aaa gac caa gaa aaa gta ata aca caa ggg aaa 48
 Met Ala Lys Leu Glu Lys Asp Gln Glu Lys Val Ile Thr Gln Gly Lys
 1 5 10 15
 tca gga gcg gat tta gtt gta gac agc tta att aat caa ggt gtt acg 96
 Ser Gly Ala Asp Leu Val Val Asp Ser Leu Ile Asn Gln Gly Val Thr
 20 25 30
 cat gta ttc ggg att ccg gga gcg aaa att gat aaa gtt ttt gat gtg 144
 His Val Phe Gly Ile Pro Gly Ala Lys Ile Asp Lys Val Phe Asp Val
 35 40 45
 atg gaa gaa cgt gga cca gaa tta att gtc agt cgt cat gaa caa aat 192
 Met Glu Glu Arg Gly Pro Glu Leu Ile Val Ser Arg His Glu Gln Asn
 50 55 60
 gcg gcg ttt atg gct gct gct atc ggt cgt cta acc ggg aaa cct ggt 240
 Ala Ala Phe Met Ala Ala Ala Ile Gly Arg Leu Thr Gly Lys Pro Gly
 65 70 75 80
 gtt gta ctt gta act agt gga cct ggc gca tcg aat ctt gca aca ggg 288
 Val Val Leu Val Thr Ser Gly Pro Gly Ala Ser Asn Leu Ala Thr Gly
 85 90 95
 ctt gta acc gca act gca gaa gga gat cca gtc gtt gcg att gct ggt 336
 Leu Val Thr Ala Thr Ala Glu Gly Asp Pro Val Val Ala Ile Ala Gly
 100 105 110
 aac gta aca agg caa gac cgc tta aaa aga acc cac caa tca atg gat 384
 Asn Val Thr Arg Gln Asp Arg Leu Lys Arg Thr His Gln Ser Met Asp
 115 120 125
 aat gca gca ctt ttc cgt ccg att aca aaa tac agc gaa gaa gta gtt 432
 Asn Ala Ala Leu Phe Arg Pro Ile Thr Lys Tyr Ser Glu Glu Val Val
 130 135 140
 cac gcc gaa agt att cca gaa gca atc act aac gct ttt cgc tcg gca 480
 His Ala Glu Ser Ile Pro Glu Ala Ile Thr Asn Ala Phe Arg Ser Ala
 145 150 155 160
 aca gaa cca aac caa ggc gct gct ttt gtc agt ttg cca caa gat atc 528
 Thr Glu Pro Asn Gln Gly Ala Ala Phe Val Ser Leu Pro Gln Asp Ile
 165 170 175
 gtg aac gaa cca aac gta cca gta aaa gcg att cgc cca ctt gct aaa 576
 Val Asn Glu Pro Asn Val Pro Val Lys Ala Ile Arg Pro Leu Ala Lys
 180 185 190
 cca gaa aat ggt cct gct tcc aaa gaa caa gtt gca aaa ctt gtt aca 624
 Pro Glu Asn Gly Pro Ala Ser Lys Glu Gln Val Ala Lys Leu Val Thr
 195 200 205
 cgt ttg aaa aaa gcg aaa tta ccg gta ttg cta ttg ggt atg cga gca 672
 Arg Leu Lys Lys Ala Lys Leu Pro Val Leu Leu Leu Gly Met Arg Ala

210

215

220

tct Ser 225	agt Ser	cca Pro	gaa Glu	gta Val	act Thr 230	ggg Gly	gca Ala	att Ile	cgt Arg	cgc Arg 235	tta Leu	ctc Leu	caa Gln	aaa Lys	aca Thr 240	720
agt Ser	atc Ile	cca Pro	gta Val	gta Val 245	gaa Glu	act Thr	ttc Phe	caa Gln	gca Ala 250	gct Ala	ggc Gly	gtc Val	att Ile	tca Ser 255	cgc Arg	768
gac Asp	tta Leu	gaa Glu	gat Asp 260	aac Asn	ttc Phe	ttt Phe	gga Gly	cgt Arg 265	ggt Val	ggg Gly	ctg Leu	ttc Phe	cgc Arg 270	aac Asn	caa Gln	816
cca Pro	ggg Gly	gat Asp 275	att Ile	ttg Leu	tta Leu	aat Asn	aaa Lys 280	gct Ala	gat Asp	tta Leu	ggt Val	att Ile 285	aca Thr	gtg Val	ggg Gly	864
tat Tyr	gat Asp 290	cca Pro	att Ile	gaa Glu	tac Tyr	gat Asp 295	cca Pro	aaa Lys	gct Ala	tgg Trp	aat Asn 300	gcc Ala	tct Ser	ggg Gly	gat Asp	912
aga Arg 305	acg Thr	att Ile	gtc Val	cat His	tta Leu 310	gac Asp	gac Asp	att Ile	cgc Arg	gct Ala 315	gat Asp	att Ile	gat Asp	cat His	tat Tyr 320	960
tac Tyr	caa Gln	cca Pro	gtg Val	aca Thr 325	gag Glu	cta Leu	gtc Val	gga Gly	aac Asn 330	atc Ile	gcg Ala	ctt Leu	act Thr	tta Leu 335	gac Asp	1008
cga Arg	gtg Val	aat Asn	gcg Ala 340	aaa Lys	ttc Phe	agc Ser	ggg Gly	tta Leu 345	gaa Glu	tta Leu	gcg Ala	gaa Glu	aaa Lys 350	gaa Glu	ctt Leu	1056
gaa Glu	aca Thr 355	tta Leu	aaa Lys	gaa Glu	ctt Leu	cat His	gct Ala 360	caa Gln	tta Leu	gaa Glu	gag Glu	cga Arg 365	gat Asp	gtt Val	ccg Pro	1104
cca Pro	gaa Glu 370	agt Ser	gat Asp	gaa Glu	act Thr	aac Asn 375	cga Arg	gta Val	cat His	cca Pro	ttg Leu 380	tcg Ser	gtc Val	att Ile	caa Gln	1152
aca Thr 385	cta Leu	cgt Arg	tcg Ser	gca Ala	att Ile 390	gat Asp	gac Asp	aac Asn	gta Val	act Thr 395	gtg Val	aca Thr	gtc Val	gac Asp	gtt Val 400	1200
ggg Gly	tca Ser	cat His	tat Tyr	att Ile 405	tgg Trp	atg Met	gca Ala	cgt Arg	cat His 410	ttc Phe	cgc Arg	tcc Ser	tat Tyr	gaa Glu 415	cca Pro	1248
cgc Arg	cgt Arg	ctg Leu	ctt Leu 420	ttc Phe	agt Ser	aac Asn	ggg Gly	atg Met 425	caa Gln	acg Thr	ctt Leu	ggg Gly	gtt Val 430	gcg Ala	ctt Leu	1296
cct Pro	tgg Trp	gga Gly 435	att Ile	gct Ala	gca Ala	aca Thr	ctt Leu 440	gta Val	cat His	ccg Pro	ggg Gly	gaa Glu 445	aaa Lys	gtg Val	gtt Val	1344
tcg Ser	att Ile 450	tct Ser	ggg Gly	gac Asp	ggg Gly	ggg Gly 455	ttc Phe	tta Leu	ttt Phe	tcc Ser	gcg Ala 460	atg Met	gaa Glu	tta Leu	gaa Glu	1392
aca Thr 465	gct Ala	gtc Val	cgc Arg	ttg Leu	cgt Arg 470	gcg Ala	cca Pro	ctt Leu	gta Val	cac His 475	cta Leu	gta Val	tgg Trp	aat Asn	gac Asp 480	1440
gga Gly	agc Ser	tat Tyr	gac Asp	atg Met	gtt Val	gct Ala	ttc Phe	caa Gln	caa Gln	aaa Lys	atg Met	aaa Lys	tac Tyr	ggc Gly	aaa Lys	1488

20110907_CL5178USNA_SeqList.txt

485										490					495					
gaa	gca	gct	gtt	cgt	ttt	ggc	gat	gtt	gat	atc	gta	aaa	ttt	gca	gaa	1536				
Glu	Ala	Ala	Val	Arg	Phe	Gly	Asp	Val	Asp	Ile	Val	Lys	Phe	Ala	Glu					
			500					505					510							
agt	ttc	gga	gca	aaa	ggc	ctt	cgc	gta	aca	aat	cca	gca	gaa	ctt	tct	1584				
Ser	Phe	Gly	Ala	Lys	Gly	Leu	Arg	Val	Thr	Asn	Pro	Ala	Glu	Leu	Ser					
		515					520					525								
gat	gtg	tta	aaa	gaa	gca	ctt	gaa	aca	gaa	gga	ccc	gtc	gtt	gta	gat	1632				
Asp	Val	Leu	Lys	Glu	Ala	Leu	Glu	Thr	Glu	Gly	Pro	Val	Val	Val	Asp					
	530					535					540									
att	cca	att	gat	tac	cgt	gat	aac	atc	aaa	ctt	ggc	gaa	act	tta	cta	1680				
Ile	Pro	Ile	Asp	Tyr	Arg	Asp	Asn	Ile	Lys	Leu	Gly	Glu	Thr	Leu	Leu					
545					550					555					560					
cct	gac	caa	ttt	tat	taa											1698				
Pro	Asp	Gln	Phe	Tyr																
				565																

<210> 10
 <211> 565
 <212> PRT
 <213> *Listeria monocytogenes*

<400> 10

Met	Ala	Lys	Leu	Glu	Lys	Asp	Gln	Glu	Lys	Val	Ile	Thr	Gln	Gly	Lys	
1				5					10					15		
Ser	Gly	Ala	Asp	Leu	Val	Val	Asp	Ser	Leu	Ile	Asn	Gln	Gly	Val	Thr	
			20					25					30			
His	Val	Phe	Gly	Ile	Pro	Gly	Ala	Lys	Ile	Asp	Lys	Val	Phe	Asp	Val	
		35					40					45				
Met	Glu	Glu	Arg	Gly	Pro	Glu	Leu	Ile	Val	Ser	Arg	His	Glu	Gln	Asn	
	50					55					60					
Ala	Ala	Phe	Met	Ala	Ala	Ala	Ile	Gly	Arg	Leu	Thr	Gly	Lys	Pro	Gly	
65				70						75					80	
Val	Val	Leu	Val	Thr	Ser	Gly	Pro	Gly	Ala	Ser	Asn	Leu	Ala	Thr	Gly	
				85					90					95		
Leu	Val	Thr	Ala	Thr	Ala	Glu	Gly	Asp	Pro	Val	Val	Ala	Ile	Ala	Gly	
			100					105					110			
Asn	Val	Thr	Arg	Gln	Asp	Arg	Leu	Lys	Arg	Thr	His	Gln	Ser	Met	Asp	
		115					120					125				
Asn	Ala	Ala	Leu	Phe	Arg	Pro	Ile	Thr	Lys	Tyr	Ser	Glu	Glu	Val	Val	
	130					135					140					
His	Ala	Glu	Ser	Ile	Pro	Glu	Ala	Ile	Thr	Asn	Ala	Phe	Arg	Ser	Ala	
145					150					155					160	

20110907_CL5178USNA_SeqList.txt

Thr Glu Pro Asn Gln Gly Ala Ala Phe Val Ser Leu Pro Gln Asp Ile
165 170 175

Val Asn Glu Pro Asn Val Pro Val Lys Ala Ile Arg Pro Leu Ala Lys
180 185 190

Pro Glu Asn Gly Pro Ala Ser Lys Glu Gln Val Ala Lys Leu Val Thr
195 200 205

Arg Leu Lys Lys Ala Lys Leu Pro Val Leu Leu Leu Gly Met Arg Ala
210 215 220

Ser Ser Pro Glu Val Thr Gly Ala Ile Arg Arg Leu Leu Gln Lys Thr
225 230 235 240

Ser Ile Pro Val Val Glu Thr Phe Gln Ala Ala Gly Val Ile Ser Arg
245 250 255

Asp Leu Glu Asp Asn Phe Phe Gly Arg Val Gly Leu Phe Arg Asn Gln
260 265 270

Pro Gly Asp Ile Leu Leu Asn Lys Ala Asp Leu Val Ile Thr Val Gly
275 280 285

Tyr Asp Pro Ile Glu Tyr Asp Pro Lys Ala Trp Asn Ala Ser Gly Asp
290 295 300

Arg Thr Ile Val His Leu Asp Asp Ile Arg Ala Asp Ile Asp His Tyr
305 310 315 320

Tyr Gln Pro Val Thr Glu Leu Val Gly Asn Ile Ala Leu Thr Leu Asp
325 330 335

Arg Val Asn Ala Lys Phe Ser Gly Leu Glu Leu Ala Glu Lys Glu Leu
340 345 350

Glu Thr Leu Lys Glu Leu His Ala Gln Leu Glu Glu Arg Asp Val Pro
355 360 365

Pro Glu Ser Asp Glu Thr Asn Arg Val His Pro Leu Ser Val Ile Gln
370 375 380

Thr Leu Arg Ser Ala Ile Asp Asp Asn Val Thr Val Thr Val Asp Val
385 390 395 400

Gly Ser His Tyr Ile Trp Met Ala Arg His Phe Arg Ser Tyr Glu Pro
405 410 415

Arg Arg Leu Leu Phe Ser Asn Gly Met Gln Thr Leu Gly Val Ala Leu
420 425 430

20110907_CL5178USNA_SeqList.txt

Pro Trp Gly Ile Ala Ala Thr Leu Val His Pro Gly Glu Lys Val Val
435 440 445

Ser Ile Ser Gly Asp Gly Gly Phe Leu Phe Ser Ala Met Glu Leu Glu
450 455 460

Thr Ala Val Arg Leu Arg Ala Pro Leu Val His Leu Val Trp Asn Asp
465 470 475 480

Gly Ser Tyr Asp Met Val Ala Phe Gln Gln Lys Met Lys Tyr Gly Lys
485 490 495

Glu Ala Ala Val Arg Phe Gly Asp Val Asp Ile Val Lys Phe Ala Glu
500 505 510

Ser Phe Gly Ala Lys Gly Leu Arg Val Thr Asn Pro Ala Glu Leu Ser
515 520 525

Asp Val Leu Lys Glu Ala Leu Glu Thr Glu Gly Pro Val Val Val Asp
530 535 540

Ile Pro Ile Asp Tyr Arg Asp Asn Ile Lys Leu Gly Glu Thr Leu Leu
545 550 555 560

Pro Asp Gln Phe Tyr
565

<210> 11
<211> 1680
<212> DNA
<213> Streptococcus mutans

<220>
<221> CDS
<222> (1)..(1680)

<400> 11	
atg acc gaa ata aat aag gaa ggc tat ggg gct gac ctg att gta gac	48
Met Thr Glu Ile Asn Lys Glu Gly Tyr Gly Ala Asp Leu Ile Val Asp	
1 5 10 15	
agc ctc att aat cat gat gtc aac tat gtt ttt gga atc cct ggt gca	96
Ser Leu Ile Asn His Asp Val Asn Tyr Val Phe Gly Ile Pro Gly Ala	
20 25 30	
aaa att gat cgt gtc ttt gat acc tta gaa gat aag ggg cca gaa ctt	144
Lys Ile Asp Arg Val Phe Asp Thr Leu Glu Asp Lys Gly Pro Glu Leu	
35 40 45	
att gta gca cgc cat gag caa aat gct gct ttt atg gct caa gga att	192
Ile Val Ala Arg His Glu Gln Asn Ala Ala Phe Met Ala Gln Gly Ile	
50 55 60	
ggc cgt att act ggt gag cct ggt gtt gtg att aca acc agc ggt ccc	240
Gly Arg Ile Thr Gly Glu Pro Gly Val Val Ile Thr Thr Ser Gly Pro	
65 70 75 80	

20110907_CL5178USNA_SeqList.txt

ggt gtt tcc aat ctg gtg act ggt ctt gtt act gcg aca gct gag gga Gly Val Ser Asn Leu Val Thr Gly Leu Val Thr Ala Thr Ala Glu Gly 85 90 95	288
gat cct gtc ctt gct att ggt ggt cag gtt aaa cgt gct gat ttg ctc Asp Pro Val Leu Ala Ile Gly Gly Gln Val Lys Arg Ala Asp Leu Leu 100 105 110	336
aaa cgg gct cac cag tca atg aat aat gtt gct atg ctc gat ccc att Lys Arg Ala His Gln Ser Met Asn Asn Val Ala Met Leu Asp Pro Ile 115 120 125	384
acc aaa tat tca gca gaa att cag gat ccc gca aca ctt tca gaa aat Thr Lys Tyr Ser Ala Glu Ile Gln Asp Pro Ala Thr Leu Ser Glu Asn 130 135 140	432
att gct aat gcc tat cgt ttg gct aaa gca gga aag ccg gga gct agt Ile Ala Asn Ala Tyr Arg Leu Ala Lys Ala Gly Lys Pro Gly Ala Ser 145 150 155 160	480
ttc tta tct att cct caa gat ata act gat agt cct gtt act gtc aag Phe Leu Ser Ile Pro Gln Asp Ile Thr Asp Ser Pro Val Thr Val Lys 165 170 175	528
gcg att aag ccc ttg aca gat cct aaa cta ggt tca gcg tca gtt gct Ala Ile Lys Pro Leu Thr Asp Pro Lys Leu Gly Ser Ala Ser Val Ala 180 185 190	576
gat att aat tat ttg gca cag gcc ata aaa aat gcg gtc ctt cct gtc Asp Ile Asn Tyr Leu Ala Gln Ala Ile Lys Asn Ala Val Leu Pro Val 195 200 205	624
tta ctt tta gga aat ggt gcg tca acg gct gca gtt aca gct tct att Leu Leu Leu Gly Asn Gly Ala Ser Thr Ala Ala Val Thr Ala Ser Ile 210 215 220	672
cgc cgt ttg tta gga gct gtc aag ctg cca gtc gtt gaa act ttc caa Arg Arg Leu Leu Gly Ala Val Lys Leu Pro Val Val Glu Thr Phe Gln 225 230 235 240	720
gga gct ggt att gtt tca aga gat tta gaa gag gac act ttt ttt ggt Gly Ala Gly Ile Val Ser Arg Asp Leu Glu Glu Asp Thr Phe Phe Gly 245 250 255	768
cgt gtg ggg ctt ttt cgt aat cag ccc gga gat atg ttg ctg aag cgt Arg Val Gly Leu Phe Arg Asn Gln Pro Gly Asp Met Leu Leu Lys Arg 260 265 270	816
tct gac tta gtt atc gct att ggc tat gat cct att gaa tat gaa gcg Ser Asp Leu Val Ile Ala Ile Gly Tyr Asp Pro Ile Glu Tyr Glu Ala 275 280 285	864
cgc aat tgg aat gct gaa att tcg gct cgc att atc gtt att gat gtt Arg Asn Trp Asn Ala Glu Ile Ser Ala Arg Ile Ile Val Ile Asp Val 290 295 300	912
gct cca gct gaa att gat act tat ttc caa cct gaa cgt gaa tta att Ala Pro Ala Glu Ile Asp Thr Tyr Phe Gln Pro Glu Arg Glu Leu Ile 305 310 315 320	960
ggt gat ata gct gaa aca ctt gat tta ctc cta cct gct att agt ggc Gly Asp Ile Ala Glu Thr Leu Asp Leu Leu Leu Pro Ala Ile Ser Gly 325 330 335	1008
tac tca ctt cca aaa ggt tct ctt gac tat ctc aaa ggc ctt cgt gat Tyr Ser Leu Pro Lys Gly Ser Leu Asp Tyr Leu Lys Gly Leu Arg Asp 340 345 350	1056

20110907_CL5178USNA_SeqList.txt

aat gta gta gaa gat gtc aaa ttt gat aag aca gtc aaa tcc ggt ctg	1104
Asn Val Val Glu Asp Val Lys Phe Asp Lys Thr Val Lys Ser Gly Leu	
355 360 365	
gtt cat ccg ctt gat gtg att gat gtc ctt caa aag caa acg act gat	1152
Val His Pro Leu Asp Val Ile Asp Val Leu Gln Lys Gln Thr Thr Asp	
370 375 380	
gat atg aca gta acg gtt gat gtt ggc agc cat tat att tgg atg gct	1200
Asp Met Thr Val Thr Val Asp Val Val Gly Ser His Tyr Ile Trp Met Ala	
385 390 395 400	
cgt tat ttt aaa agc tat gaa gca cgg cac tta ctt ttc tca aat ggt	1248
Arg Tyr Phe Lys Ser Tyr Glu Ala Arg His Leu Leu Phe Ser Asn Gly	
405 410 415	
atg caa acc tta ggt gtt gct ttg cct tgg gca att tcg gca gct ctt	1296
Met Gln Thr Leu Gly Val Ala Leu Pro Trp Ala Ile Ser Ala Ala Leu	
420 425 430	
gta cgg cca aat gag aag att att tct att tca ggt gat ggt ggt ttc	1344
Val Arg Pro Asn Glu Lys Ile Ile Ser Ile Ser Gly Asp Gly Gly Phe	
435 440 445	
ctc ttt tct ggc caa gaa ttg gaa aca gct gtt cgt tta cat tta cca	1392
Leu Phe Ser Gly Gln Glu Leu Glu Thr Ala Val Arg Leu His Leu Pro	
450 455 460	
att gtt cat atc att tgg aat gat ggt aaa tat aat atg gtt gaa ttc	1440
Ile Val His Ile Ile Trp Asn Asp Gly Lys Tyr Asn Met Val Glu Phe	
465 470 475 480	
caa gaa gaa atg aaa tac ggc cgt tca gca ggt gtt gat ttt ggt cct	1488
Gln Glu Glu Met Lys Tyr Gly Arg Ser Ala Gly Val Asp Phe Gly Pro	
485 490 495	
gtt gat ttt gtc aag tat gct gat agt ttc ggt gct aaa ggt tac cgt	1536
Val Asp Phe Val Lys Tyr Ala Asp Ser Phe Gly Ala Lys Gly Tyr Arg	
500 505 510	
gct gat agt aaa gaa aag ttt gat caa gtt ctt caa aca gca ctc aag	1584
Ala Asp Ser Lys Glu Lys Phe Asp Gln Val Leu Gln Thr Ala Leu Lys	
515 520 525	
gaa gct gca aat ggc cca gtt ctc att gat gtt cca atg gac tat aaa	1632
Glu Ala Ala Asn Gly Pro Val Leu Ile Asp Val Pro Met Asp Tyr Lys	
530 535 540	
gat aat gta aaa ttg ggt gaa act att ttg cct gat gaa ttc tac taa	1680
Asp Asn Val Lys Leu Gly Glu Thr Ile Leu Pro Asp Glu Phe Tyr	
545 550 555	

<210> 12
 <211> 559
 <212> PRT
 <213> Streptococcus mutans

<400> 12

Met Thr Glu Ile Asn Lys Glu Gly Tyr Gly Ala Asp Leu Ile Val Asp
1 5 10 15

Ser Leu Ile Asn His Asp Val Asn Tyr Val Phe Gly Ile Pro Gly Ala
20 25 30

20110907_CL5178USNA_SeqList.txt

Lys Ile Asp Arg Val Phe Asp Thr Leu Glu Asp Lys Gly Pro Glu Leu
 35 40 45

 Ile Val Ala Arg His Glu Gln Asn Ala Ala Phe Met Ala Gln Gly Ile
 50 55 60

 Gly Arg Ile Thr Gly Glu Pro Gly Val Val Ile Thr Thr Ser Gly Pro
 65 70 75 80

 Gly Val Ser Asn Leu Val Thr Gly Leu Val Thr Ala Thr Ala Glu Gly
 85 90 95

 Asp Pro Val Leu Ala Ile Gly Gly Gln Val Lys Arg Ala Asp Leu Leu
 100 105 110

 Lys Arg Ala His Gln Ser Met Asn Asn Val Ala Met Leu Asp Pro Ile
 115 120 125

 Thr Lys Tyr Ser Ala Glu Ile Gln Asp Pro Ala Thr Leu Ser Glu Asn
 130 135 140

 Ile Ala Asn Ala Tyr Arg Leu Ala Lys Ala Gly Lys Pro Gly Ala Ser
 145 150 155 160

 Phe Leu Ser Ile Pro Gln Asp Ile Thr Asp Ser Pro Val Thr Val Lys
 165 170 175

 Ala Ile Lys Pro Leu Thr Asp Pro Lys Leu Gly Ser Ala Ser Val Ala
 180 185 190

 Asp Ile Asn Tyr Leu Ala Gln Ala Ile Lys Asn Ala Val Leu Pro Val
 195 200 205

 Leu Leu Leu Gly Asn Gly Ala Ser Thr Ala Ala Val Thr Ala Ser Ile
 210 215 220

 Arg Arg Leu Leu Gly Ala Val Lys Leu Pro Val Val Glu Thr Phe Gln
 225 230 235 240

 Gly Ala Gly Ile Val Ser Arg Asp Leu Glu Glu Asp Thr Phe Phe Gly
 245 250 255

 Arg Val Gly Leu Phe Arg Asn Gln Pro Gly Asp Met Leu Leu Lys Arg
 260 265 270

 Ser Asp Leu Val Ile Ala Ile Gly Tyr Asp Pro Ile Glu Tyr Glu Ala
 275 280 285

 Arg Asn Trp Asn Ala Glu Ile Ser Ala Arg Ile Ile Val Ile Asp Val
 290 295 300

20110907_CL5178USNA_SeqList.txt

Ala Pro Ala Glu Ile Asp Thr Tyr Phe Gln Pro Glu Arg Glu Leu Ile
305 310 315 320

Gly Asp Ile Ala Glu Thr Leu Asp Leu Leu Leu Pro Ala Ile Ser Gly
325 330 335

Tyr Ser Leu Pro Lys Gly Ser Leu Asp Tyr Leu Lys Gly Leu Arg Asp
340 345 350

Asn Val Val Glu Asp Val Lys Phe Asp Lys Thr Val Lys Ser Gly Leu
355 360 365

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

Asp Met Thr Val Thr Val Asp Val Gly Ser His Tyr Ile Trp Met Ala
385 390 395 400

Arg Tyr Phe Lys Ser Tyr Glu Ala Arg His Leu Leu Phe Ser Asn Gly
405 410 415

Met Gln Thr Leu Gly Val Ala Leu Pro Trp Ala Ile Ser Ala Ala Leu
420 425 430

Val Arg Pro Asn Glu Lys Ile Ile Ser Ile Ser Gly Asp Gly Gly Phe
435 440 445

Leu Phe Ser Gly Gln Glu Leu Glu Thr Ala Val Arg Leu His Leu Pro
450 455 460

Ile Val His Ile Ile Trp Asn Asp Gly Lys Tyr Asn Met Val Glu Phe
465 470 475 480

Gln Glu Glu Met Lys Tyr Gly Arg Ser Ala Gly Val Asp Phe Gly Pro
485 490 495

Val Asp Phe Val Lys Tyr Ala Asp Ser Phe Gly Ala Lys Gly Tyr Arg
500 505 510

Ala Asp Ser Lys Glu Lys Phe Asp Gln Val Leu Gln Thr Ala Leu Lys
515 520 525

Glu Ala Ala Asn Gly Pro Val Leu Ile Asp Val Pro Met Asp Tyr Lys
530 535 540

Asp Asn Val Lys Leu Gly Glu Thr Ile Leu Pro Asp Glu Phe Tyr
545 550 555

<210> 13
<211> 1683
<212> DNA

<213> Streptococcus thermophilus

<220>

<221> CDS

<222> (1)..(1683)

<400> 13

gtg ttc atg tca gaa gaa aag caa ttg tat ggt gca gat tta gtg gtt	48
Val Phe Met Ser Glu Glu Lys Gln Leu Tyr Gly Ala Asp Leu Val Val	
1 5 10 15	
gat agt ttg atc aac cat gat gtt gag tat gtc ttt ggg att cca ggc	96
Asp Ser Leu Ile Asn His Asp Val Glu Tyr Val Phe Gly Ile Pro Gly	
20 25 30	
gca aaa atc gat agg gtt ttt gat acc ttg gaa gat aag gga cct gaa	144
Ala Lys Ile Asp Arg Val Phe Asp Thr Leu Glu Asp Lys Gly Pro Glu	
35 40 45	
ttg att gtt gcc cgt cat gag caa aat gct gct ttt atg gct caa ggt	192
Leu Ile Val Ala Arg His Glu Gln Asn Ala Ala Phe Met Ala Gln Gly	
50 55 60	
gtt gga cgt att act ggg aaa cca ggt gta gta ttg gta aca tct ggt	240
Val Gly Arg Ile Thr Gly Lys Pro Gly Val Val Leu Val Thr Ser Gly	
65 70 75 80	
cca ggt gtc tcc aat ttg gct act ggt ttg gta aca gcg acg gat gaa	288
Pro Gly Val Ser Asn Leu Ala Thr Gly Leu Val Thr Ala Thr Asp Glu	
85 90 95	
gga gac cct gtt ctt gct att ggt ggt cag gtt aag cgt gca gat ctc	336
Gly Asp Pro Val Leu Ala Ile Gly Gly Gln Val Lys Arg Ala Asp Leu	
100 105 110	
ttg aaa cgt gcc cac caa tca atg aat aac gtt gct atg ctt gag cca	384
Leu Lys Arg Ala His Gln Ser Met Asn Asn Val Ala Met Leu Glu Pro	
115 120 125	
att acc aaa tat gct gct gaa gta cat gat gct aac acc ctt tct gaa	432
Ile Thr Lys Tyr Ala Ala Glu Val His Asp Ala Asn Thr Leu Ser Glu	
130 135 140	
acg gtt gct aat gcc tat cgt cac gct aag tca ggg aaa cca ggt gca	480
Thr Val Ala Asn Ala Tyr Arg His Ala Lys Ser Gly Lys Pro Gly Ala	
145 150 155 160	
agc ttc att tca att cct caa gac gtg acg gat gct ccg gtc agt gtt	528
Ser Phe Ile Ser Ile Pro Gln Asp Val Thr Asp Ala Pro Val Ser Val	
165 170 175	
aag gct att aag cct atg aca gat cca aaa ctt ggt tca gca tct gtt	576
Lys Ala Ile Lys Pro Met Thr Asp Pro Lys Leu Gly Ser Ala Ser Val	
180 185 190	
tct gat att aac tat cta gca caa gcc att aaa aat gca gtg ttg cca	624
Ser Asp Ile Asn Tyr Leu Ala Gln Ala Ile Lys Asn Ala Val Leu Pro	
195 200 205	
gtc ttt ctt ttg ggg aat ggt gcc tca tca gaa gcc gta act tac tct	672
Val Phe Leu Leu Gly Asn Gly Ala Ser Ser Glu Ala Val Thr Tyr Ser	
210 215 220	
att cgc caa att ttg aag cat gtt aaa ttg cca gtt gtt gaa act ttc	720
Ile Arg Gln Ile Leu Lys His Val Lys Leu Pro Val Val Glu Thr Phe	
225 230 235 240	

20110907_CL5178USNA_SeqList.txt

caa ggt gcc ggt atc gtg tca cgt gac ctt gaa gaa gat act ttc ttt	768
Gln Gly Ala Gly Ile Val Ser Arg Asp Leu Glu Glu Asp Thr Phe Phe	
	245 250 255
ggt cgt gta ggt ctt ttc cgt aac caa ccc gga gac atg ttg ctt aaa	816
Gly Arg Val Gly Leu Phe Arg Asn Gln Pro Gly Asp Met Leu Leu Lys	
	260 265 270
aaa tcc gac tta gtt att gcc att ggt tat gat cca atc gaa tat gaa	864
Lys Ser Asp Leu Val Ile Ala Ile Gly Tyr Asp Pro Ile Glu Tyr Glu	
	275 280 285
gca cgt aac tgg aat gct gaa att tca gca cgt atc atc gtt att gat	912
Ala Arg Asn Trp Asn Ala Glu Ile Ser Ala Arg Ile Val Ile Asp	
	290 295 300
gtc gag ccg gcc gag gtg gac act tac ttc caa ccg gaa cgt gaa ttg	960
Val Glu Pro Ala Glu Val Asp Thr Tyr Phe Gln Pro Glu Arg Glu Leu	
	305 310 315 320
att ggt aat gta gaa gcg agc tta gac ttg ctt ttg ccc gct att caa	1008
Ile Gly Asn Val Glu Ala Ser Leu Asp Leu Leu Leu Pro Ala Ile Gln	
	325 330 335
ggt tat aaa ttg cct gaa ggt gcg gtt gaa tat ctt aaa ggt ttg aaa	1056
Gly Tyr Lys Leu Pro Glu Gly Ala Val Glu Tyr Leu Lys Gly Leu Lys	
	340 345 350
aac aat gtt gtt gag gat gtt aag ttt gac cgt cag cct gat gaa ggt	1104
Asn Asn Val Val Glu Asp Val Lys Phe Asp Arg Gln Pro Asp Glu Gly	
	355 360 365
acg gtg cat ccg cta gat ttc atc gaa aat ttg caa gaa cac aca gat	1152
Thr Val His Pro Leu Asp Phe Ile Glu Asn Leu Gln Glu His Thr Asp	
	370 375 380
gat gat atg act gtt acg ttt gat gtt ggt agt cac tat att tgg atg	1200
Asp Asp Met Thr Val Thr Phe Asp Val Gly Ser His Tyr Ile Trp Met	
	385 390 395 400
gca cgt tat ctc aaa tcg tat gaa cca cgt cat ttg ctt ttc tca aat	1248
Ala Arg Tyr Leu Lys Ser Tyr Glu Pro Arg His Leu Leu Phe Ser Asn	
	405 410 415
ggg atg caa acg ata ggt att gct att aca tgg gct atc tct gca gca	1296
Gly Met Gln Thr Ile Gly Ile Ala Ile Thr Trp Ala Ile Ser Ala Ala	
	420 425 430
ttg gtt cgt cct aag aca aaa gtg att tct gta tct ggt gat ggt ggt	1344
Leu Val Arg Pro Lys Thr Lys Val Ile Ser Val Ser Gly Asp Gly Gly	
	435 440 445
ttc ctc ttc tca gca caa gaa ttg gaa aca gca gtt cgt ttg aaa ttg	1392
Phe Leu Phe Ser Ala Gln Glu Leu Glu Thr Ala Val Arg Leu Lys Leu	
	450 455 460
cca att gtc cat att atc tgg aac gat ggt cat tac aat atg gtg gaa	1440
Pro Ile Val His Ile Ile Trp Asn Asp Gly His Tyr Asn Met Val Glu	
	465 470 475 480
ttc cag gaa gaa atg aag tac ggt cgt tca tct ggg gtt gac ttt ggt	1488
Phe Gln Glu Glu Met Lys Tyr Gly Arg Ser Ser Gly Val Asp Phe Gly	
	485 490 495
cct gta gat ttt gta aaa tat gct gag agc ttt gga gcc aaa ggt tat	1536
Pro Val Asp Phe Val Lys Tyr Ala Glu Ser Phe Gly Ala Lys Gly Tyr	
	500 505 510

20110907_CL5178USNA_SeqList.txt

cgt gca aca agt aaa gca gcg ttt gct agc ttg ctt caa gag gct ttg	1584
Arg Ala Thr Ser Lys Ala Ala Phe Ala Ser Leu Leu Gln Glu Ala Leu	
515 520 525	
act cag gct gta gat gga cca gtc ctt att gat gtt cca att gac tat	1632
Thr Gln Ala Val Asp Gly Pro Val Leu Ile Asp Val Pro Ile Asp Tyr	
530 535 540	
aaa gat aac att aaa ctc ggc gaa act att ttg cca gat gaa ttt tac	1680
Lys Asp Asn Ile Lys Leu Gly Glu Thr Ile Leu Pro Asp Glu Phe Tyr	
545 550 555 560	
taa	1683

<210> 14
 <211> 560
 <212> PRT
 <213> Streptococcus thermophilus
 <400> 14

Val Phe Met Ser Glu Glu Lys Gln Leu Tyr Gly Ala Asp Leu Val Val	
1 5 10 15	
Asp Ser Leu Ile Asn His Asp Val Glu Tyr Val Phe Gly Ile Pro Gly	
20 25 30	
Ala Lys Ile Asp Arg Val Phe Asp Thr Leu Glu Asp Lys Gly Pro Glu	
35 40 45	
Leu Ile Val Ala Arg His Glu Gln Asn Ala Ala Phe Met Ala Gln Gly	
50 55 60	
Val Gly Arg Ile Thr Gly Lys Pro Gly Val Val Leu Val Thr Ser Gly	
65 70 75 80	
Pro Gly Val Ser Asn Leu Ala Thr Gly Leu Val Thr Ala Thr Asp Glu	
85 90 95	
Gly Asp Pro Val Leu Ala Ile Gly Gly Gln Val Lys Arg Ala Asp Leu	
100 105 110	
Leu Lys Arg Ala His Gln Ser Met Asn Asn Val Ala Met Leu Glu Pro	
115 120 125	
Ile Thr Lys Tyr Ala Ala Glu Val His Asp Ala Asn Thr Leu Ser Glu	
130 135 140	
Thr Val Ala Asn Ala Tyr Arg His Ala Lys Ser Gly Lys Pro Gly Ala	
145 150 155 160	
Ser Phe Ile Ser Ile Pro Gln Asp Val Thr Asp Ala Pro Val Ser Val	
165 170 175	
Lys Ala Ile Lys Pro Met Thr Asp Pro Lys Leu Gly Ser Ala Ser Val	
180 185 190	

20110907_CL5178USNA_SeqList.txt

Ser Asp Ile Asn Tyr Leu Ala Gln Ala Ile Lys Asn Ala Val Leu Pro
195 200 205

Val Phe Leu Leu Gly Asn Gly Ala Ser Ser Glu Ala Val Thr Tyr Ser
210 215 220

Ile Arg Gln Ile Leu Lys His Val Lys Leu Pro Val Val Glu Thr Phe
225 230 235 240

Gln Gly Ala Gly Ile Val Ser Arg Asp Leu Glu Glu Asp Thr Phe Phe
245 250 255

Gly Arg Val Gly Leu Phe Arg Asn Gln Pro Gly Asp Met Leu Leu Lys
260 265 270

Lys Ser Asp Leu Val Ile Ala Ile Gly Tyr Asp Pro Ile Glu Tyr Glu
275 280 285

Ala Arg Asn Trp Asn Ala Glu Ile Ser Ala Arg Ile Ile Val Ile Asp
290 295 300

Val Glu Pro Ala Glu Val Asp Thr Tyr Phe Gln Pro Glu Arg Glu Leu
305 310 315 320

Ile Gly Asn Val Glu Ala Ser Leu Asp Leu Leu Leu Pro Ala Ile Gln
325 330 335

Gly Tyr Lys Leu Pro Glu Gly Ala Val Glu Tyr Leu Lys Gly Leu Lys
340 345 350

Asn Asn Val Val Glu Asp Val Lys Phe Asp Arg Gln Pro Asp Glu Gly
355 360 365

Thr Val His Pro Leu Asp Phe Ile Glu Asn Leu Gln Glu His Thr Asp
370 375 380

Asp Asp Met Thr Val Thr Phe Asp Val Gly Ser His Tyr Ile Trp Met
385 390 395 400

Ala Arg Tyr Leu Lys Ser Tyr Glu Pro Arg His Leu Leu Phe Ser Asn
405 410 415

Gly Met Gln Thr Ile Gly Ile Ala Ile Thr Trp Ala Ile Ser Ala Ala
420 425 430

Leu Val Arg Pro Lys Thr Lys Val Ile Ser Val Ser Gly Asp Gly Gly
435 440 445

Phe Leu Phe Ser Ala Gln Glu Leu Glu Thr Ala Val Arg Leu Lys Leu
450 455 460

20110907_CL5178USNA_SeqList.txt

Pro Ile Val His Ile Ile Trp Asn Asp Gly His Tyr Asn Met Val Glu
465 470 475 480

Phe Gln Glu Glu Met Lys Tyr Gly Arg Ser Ser Gly Val Asp Phe Gly
485 490 495

Pro Val Asp Phe Val Lys Tyr Ala Glu Ser Phe Gly Ala Lys Gly Tyr
500 505 510

Arg Ala Thr Ser Lys Ala Ala Phe Ala Ser Leu Leu Gln Glu Ala Leu
515 520 525

Thr Gln Ala Val Asp Gly Pro Val Leu Ile Asp Val Pro Ile Asp Tyr
530 535 540

Lys Asp Asn Ile Lys Leu Gly Glu Thr Ile Leu Pro Asp Glu Phe Tyr
545 550 555 560

<210> 15
<211> 1665
<212> DNA
<213> Vibrio angustum

<220>
<221> CDS
<222> (1)..(1665)

<400> 15
atg tcg gat aaa acc gtc tct ggt gct gaa ctg gtt gtt gaa act tta 48
Met Ser Asp Lys Thr Val Ser Gly Ala Glu Leu Val Val Glu Thr Leu
1 5 10 15
aat gca cat aac gtt cca cac att ttt ggt att cct gga gca aag gtg 96
Asn Ala His Asn Val Pro His Ile Phe Gly Ile Pro Gly Ala Lys Val
20 25 30
gat gct gtt ttc gat gct gtt tgt gat aac gga cca gaa atc att att 144
Asp Ala Val Phe Asp Ala Val Cys Asp Asn Gly Pro Glu Ile Ile Ile
35 40 45
tgt cat cat gaa caa aat gca gcg ttt atg gca gca gca act ggg cgt 192
Cys His His Glu Gln Asn Ala Ala Phe Met Ala Ala Ala Thr Gly Arg
50 55 60
tta acg ggt aaa gca ggc att tgt tta gca acc tct gga cca ggc gca 240
Leu Thr Gly Lys Ala Gly Ile Cys Leu Ala Thr Ser Gly Pro Gly Ala
65 70 75 80
tca aac ctt gtc aca ggc gtt gca aca gcg aat agt gaa ggt gat cct 288
Ser Asn Leu Val Thr Gly Val Ala Thr Ala Asn Ser Glu Gly Asp Pro
85 90 95
gtg gtt gca ctt gca ggt gct gta cct ctt tct atg tat tct cac aat 336
Val Val Ala Leu Ala Gly Ala Val Pro Leu Ser Met Tyr Ser His Asn
100 105 110
act cat caa tcc atg gat acc cgt tca ctg ttt act cct atc acc aag 384
Thr His Gln Ser Met Asp Thr Arg Ser Leu Phe Thr Pro Ile Thr Lys
115 120 125

20110907_CL5178USNA_SeqList.txt

ttt tca gca gaa gtg atg gat agc agc tcg gta tct gat gtt gta cat	432
Phe Ser Ala Glu Val Met Asp Ser Ser Ser Val Ser Asp Val Val His	
130 135 140	
aaa gct ttt cgt att gca gag caa cct acc caa ggt gct agc ttt gtt	480
Lys Ala Phe Arg Ile Ala Glu Gln Pro Thr Gln Gly Ala Ser Phe Val	
145 150 155 160	
agt cta ccg caa gat att cta act aac cgt att cct tac cag cca gta	528
Ser Leu Pro Gln Asp Ile Leu Thr Asn Arg Ile Pro Tyr Gln Pro Val	
165 170 175	
caa cag cct aat cca att ttg ttc ggt ggt gca cac cca caa gct att	576
Gln Gln Pro Asn Pro Ile Leu Phe Gly Gly Ala His Pro Gln Ala Ile	
180 185 190	
cgt cag gct gct gat cgc att aat gct gca aaa aat ccg gtg tta tta	624
Arg Gln Ala Ala Asp Arg Ile Asn Ala Ala Lys Asn Pro Val Leu Leu	
195 200 205	
ctg ggc atg gat gca agc cag cct ttt gtt gct gat gct att cgc caa	672
Leu Gly Met Asp Ala Ser Gln Pro Phe Val Ala Asp Ala Ile Arg Gln	
210 215 220	
cta ctc aaa caa aca cca att gcc gtt gtg aat acg ttt gcc gca gct	720
Leu Leu Lys Gln Thr Pro Ile Ala Val Val Asn Thr Phe Ala Ala Ala	
225 230 235 240	
ggg gtt att tct cat gat tta tac aac tgc ttt tta ggt cgt gtt ggc	768
Gly Val Ile Ser His Asp Leu Tyr Asn Cys Phe Leu Gly Arg Val Gly	
245 250 255	
tta ttt aaa aat caa ccc ggt gat att gca tta aac agt gca gat tta	816
Leu Phe Lys Asn Gln Pro Gly Asp Ile Ala Leu Asn Ser Ala Asp Leu	
260 265 270	
atc att acc att ggc tac agc cca att gaa tac gat ccg att ctt tgg	864
Ile Ile Thr Ile Gly Tyr Ser Pro Ile Glu Tyr Asp Pro Ile Leu Trp	
275 280 285	
aat aaa gat gca aac aca cca att att cat att ggt tat caa caa gca	912
Asn Lys Asp Ala Asn Thr Pro Ile Ile His Ile Gly Tyr Gln Gln Ala	
290 295 300	
gat tta gaa att agc tat aac cct gtt tgt gaa gtt gtg ggt gac tta	960
Asp Leu Glu Ile Ser Tyr Asn Pro Val Cys Glu Val Val Gly Asp Leu	
305 310 315 320	
gcg gtg tct gtc acg tct att gct tct gaa tta gat aag cga gaa tca	1008
Ala Val Ser Val Thr Ser Ile Ala Ser Glu Leu Asp Lys Arg Glu Ser	
325 330 335	
tta gaa aat aac caa caa atc caa tta tta cgc cac gat tta caa cat	1056
Leu Glu Asn Asn Gln Gln Ile Gln Leu Leu Arg His Asp Leu Gln His	
340 345 350	
att atg cag atg ggg gta aat aaa acc tca aca aac ggc gtt cac ccg	1104
Ile Met Gln Met Gly Val Asn Lys Thr Ser Thr Asn Gly Val His Pro	
355 360 365	
ctt cgt ttt gtt cat gag tta cgt cgc ttt gtt agt gac gac acc act	1152
Leu Arg Phe Val His Glu Leu Arg Arg Phe Val Ser Asp Asp Thr Thr	
370 375 380	
gta tgt tgt gat gta ggc tct att tat att tgg atg gca cgt tac ttc	1200
Val Cys Cys Asp Val Gly Ser Ile Tyr Ile Trp Met Ala Arg Tyr Phe	
385 390 395 400	

20110907_CL5178USNA_SeqList.txt

cac agc ttt gaa cct cgt cgt tta ttg ttc agc aat ggc caa caa aca His Ser Phe Glu Pro Arg Arg Leu Leu Phe Ser Asn Gly Gln Gln Thr 405 410 415	1248
ttg ggc gta gct tta cct tgg gca att gca gct tcc ctt ctt cac cct Leu Gly Val Ala Leu Pro Trp Ala Ile Ala Ala Ser Leu Leu His Pro 420 425 430	1296
aat gaa aaa gta att tcc atg tct ggt gat ggt ggc ttc cta ttc tca Asn Glu Lys Val Ile Ser Met Ser Gly Asp Gly Gly Phe Leu Phe Ser 435 440 445	1344
tca atg gaa tta gcc acg gcc gtt cgc cat aaa tgt aat atc gtt cac Ser Met Glu Leu Ala Thr Ala Val Arg His Lys Cys Asn Ile Val His 450 455 460	1392
ttt gtt tgg aca gat cac agt tat gac atg gtt aag atc caa cag ctt Phe Val Trp Thr Asp His Ser Tyr Asp Met Val Lys Ile Gln Gln Leu 465 470 475 480	1440
aaa aag tat ggt cga gag agt gcc gtc agc ttt ata ggt cct gat att Lys Lys Tyr Gly Arg Glu Ser Ala Val Ser Phe Ile Gly Pro Asp Ile 485 490 495	1488
ggt aag tac gca gaa agc ttc ggc gca cat ggt tta gcg atc aat act Val Lys Tyr Ala Glu Ser Phe Gly Ala His Gly Leu Ala Ile Asn Thr 500 505 510	1536
gcc gat gat att gag cct gtt atg cga aaa gct atg agc tta agt ggc Ala Asp Asp Ile Glu Pro Val Met Arg Lys Ala Met Ser Leu Ser Gly 515 520 525	1584
cca gta ttg gtc aac gtc aat gtt gat tat agc gat aac agt cgc cta Pro Val Leu Val Asn Val Asn Val Asp Tyr Ser Asp Asn Ser Arg Leu 530 535 540	1632
ctt gat caa ctt cat cca tgc caa caa gat taa Leu Asp Gln Leu His Pro Cys Gln Gln Asp 545 550	1665

<210> 16
 <211> 554
 <212> PRT
 <213> Vibrio angustum

<400> 16

Met Ser Asp Lys Thr Val Ser Gly Ala Glu Leu Val Val Glu Thr Leu 1 5 10 15
Asn Ala His Asn Val Pro His Ile Phe Gly Ile Pro Gly Ala Lys Val 20 25 30
Asp Ala Val Phe Asp Ala Val Cys Asp Asn Gly Pro Glu Ile Ile Ile 35 40 45
Cys His His Glu Gln Asn Ala Ala Phe Met Ala Ala Ala Thr Gly Arg 50 55 60
Leu Thr Gly Lys Ala Gly Ile Cys Leu Ala Thr Ser Gly Pro Gly Ala 65 70 75 80

20110907_CL5178USNA_SeqList.txt

Ser Asn Leu Val Thr Gly Val Ala Thr Ala Asn Ser Glu Gly Asp Pro
 85 90 95
 Val Val Ala Leu Ala Gly Ala Val Pro Leu Ser Met Tyr Ser His Asn
 100 105 110
 Thr His Gln Ser Met Asp Thr Arg Ser Leu Phe Thr Pro Ile Thr Lys
 115 120 125
 Phe Ser Ala Glu Val Met Asp Ser Ser Ser Val Ser Asp Val Val His
 130 135 140
 Lys Ala Phe Arg Ile Ala Glu Gln Pro Thr Gln Gly Ala Ser Phe Val
 145 150 155 160
 Ser Leu Pro Gln Asp Ile Leu Thr Asn Arg Ile Pro Tyr Gln Pro Val
 165 170 175
 Gln Gln Pro Asn Pro Ile Leu Phe Gly Gly Ala His Pro Gln Ala Ile
 180 185 190
 Arg Gln Ala Ala Asp Arg Ile Asn Ala Ala Lys Asn Pro Val Leu Leu
 195 200 205
 Leu Gly Met Asp Ala Ser Gln Pro Phe Val Ala Asp Ala Ile Arg Gln
 210 215 220
 Leu Leu Lys Gln Thr Pro Ile Ala Val Val Asn Thr Phe Ala Ala Ala
 225 230 235 240
 Gly Val Ile Ser His Asp Leu Tyr Asn Cys Phe Leu Gly Arg Val Gly
 245 250 255
 Leu Phe Lys Asn Gln Pro Gly Asp Ile Ala Leu Asn Ser Ala Asp Leu
 260 265 270
 Ile Ile Thr Ile Gly Tyr Ser Pro Ile Glu Tyr Asp Pro Ile Leu Trp
 275 280 285
 Asn Lys Asp Ala Asn Thr Pro Ile Ile His Ile Gly Tyr Gln Gln Ala
 290 295 300
 Asp Leu Glu Ile Ser Tyr Asn Pro Val Cys Glu Val Val Gly Asp Leu
 305 310 315 320
 Ala Val Ser Val Thr Ser Ile Ala Ser Glu Leu Asp Lys Arg Glu Ser
 325 330 335
 Leu Glu Asn Asn Gln Gln Ile Gln Leu Leu Arg His Asp Leu Gln His
 340 345 350

20110907_CL5178USNA_SeqList.txt

Ile Met Gln Met Gly Val Asn Lys Thr Ser Thr Asn Gly Val His Pro
355 360 365

Leu Arg Phe Val His Glu Leu Arg Arg Phe Val Ser Asp Asp Thr Thr
370 375 380

Val Cys Cys Asp Val Gly Ser Ile Tyr Ile Trp Met Ala Arg Tyr Phe
385 390 395 400

His Ser Phe Glu Pro Arg Arg Leu Leu Phe Ser Asn Gly Gln Gln Thr
405 410 415

Leu Gly Val Ala Leu Pro Trp Ala Ile Ala Ala Ser Leu Leu His Pro
420 425 430

Asn Glu Lys Val Ile Ser Met Ser Gly Asp Gly Gly Phe Leu Phe Ser
435 440 445

Ser Met Glu Leu Ala Thr Ala Val Arg His Lys Cys Asn Ile Val His
450 455 460

Phe Val Trp Thr Asp His Ser Tyr Asp Met Val Lys Ile Gln Gln Leu
465 470 475 480

Lys Lys Tyr Gly Arg Glu Ser Ala Val Ser Phe Ile Gly Pro Asp Ile
485 490 495

Val Lys Tyr Ala Glu Ser Phe Gly Ala His Gly Leu Ala Ile Asn Thr
500 505 510

Ala Asp Asp Ile Glu Pro Val Met Arg Lys Ala Met Ser Leu Ser Gly
515 520 525

Pro Val Leu Val Asn Val Asn Val Asp Tyr Ser Asp Asn Ser Arg Leu
530 535 540

Leu Asp Gln Leu His Pro Cys Gln Gln Asp
545 550

<210> 17
<211> 1689
<212> DNA
<213> Bacillus cereus

<400> 17
ttgagtacag gtgtaaaagc aaacgacgtg aagacaaaaa caaaaggagc agatcttggt 60
gttgattggt taattaaaca aggtgttaca catgttttcg gtattccagg agcaaagatt 120
gactctgtat ttgatgtact gcaagaaaga ggaccagagt taattgtttg tcgtcatgaa 180
caaaatgcag catttatggc agctgctatt ggtagattaa caggaaaacc gggcgtatgt 240
cttgtaactt caggaccagg gacatcaaat ttagcgacag gtcttggttac tgcgaatgcg 300

20110907_CL5178USNA_SeqList.txt

gagagtgatc ccgttggtgc tttagctggt gcagttccgc gtacggatcg attaaaacgt 360
acacatcaat ctatggataa tgctgcacta ttcgaaccaa tcacaaaata tagcgtagaa 420
gtagagcatc ctgataatgt gccagaagca ttatcaaagt cattccgaag tgcgacttct 480
acaaatccag gagcaacttt agtaagtttg ccgcaagacg ttatgactgc ggaaacgact 540
gtagagtcta tcggtgcgct ttctaagcca cagcttggaa tcgctccac acatgatatt 600
acatatgtag tagataaaat aaaagcagcg aaattaccag ttattttact cggtatgaga 660
gcgagcacia atgaagtgc gaaagccgtt cgtaaattaa ttgcggatac agaacttcct 720
gtcgttgaia catatcaagc ggctggtgcc atttcacgtg agttagaaga tcatttcttc 780
ggccgtgttg gactattccg taaccaacca ggtgatattt tactagaaga ggcagatctt 840
gttatttcta tcggttatga tccaattgag tatgatccaa agttctggaa taaacttgga 900
gacagaacga ttattcatct tgatgacct caagcagata tagatcatga ttaccaacca 960
gagcgtgaat taattggtga tattgcctta acagtaaata gcatcgcaga aaagttaccg 1020
aaacttgtgt taagtacgaa atcagaagca gtgttagaac gattacgcgc gaaattatca 1080
gaacaagcag aagttccaaa tcgtccttca gaaggtgtta cacatccgct tcaagtgatt 1140
cgtacacttc gttctttaat tagtgacgac acaaccgtta catgtgacat cggttcccat 1200
tctatttga tggcgagatg tttccgttct tatgaaccac gtagattatt atttagtaac 1260
ggtatgcaga cgtaggtgt tgcaattcct tgggcaattg ctgctacttt agtagaacca 1320
ggtaaaaaag tagtttccgt atcaggtgac ggtggtttct tattctcagc gatggagtta 1380
gaaacggcgg tacgtttaaa ttctccaatc gtccatcttg tttggagaga cggcacatat 1440
gatatggttg cattccaaca aatgatgaaa tacggcagaa catcagctac agagtttggg 1500
gatgttgatc ttgttaaata tgcggaaagt ttcggggcgt taggtcttcg tggttaacacg 1560
cctgatgaat tagaaggggt attgaaagaa gactagcag cagacggccc tgtcattatt 1620
gatattccaa ttgactatcg tgacaacatt aaattaagcg aaaaattatt accaaaccaa 1680
ttaaactaa 1689

<210> 18
<211> 562
<212> PRT
<213> Bacillus cereus

<400> 18

Met Ser Thr Gly Val Lys Ala Asn Asp Val Lys Thr Lys Thr Lys Gly
1 5 10 15

Ala Asp Leu Val Val Asp Cys Leu Ile Lys Gln Gly Val Thr His Val
20 25 30

Phe Gly Ile Pro Gly Ala Lys Ile Asp Ser Val Phe Asp Val Leu Gln
35 40 45

20110907_CL5178USNA_SeqList.txt

Glu Arg Gly Pro Glu Leu Ile Val Cys Arg His Glu Gln Asn Ala Ala
 50 55 60
 Phe Met Ala Ala Ala Ile Gly Arg Leu Thr Gly Lys Pro Gly Val Cys
 65 70 75 80
 Leu Val Thr Ser Gly Pro Gly Thr Ser Asn Leu Ala Thr Gly Leu Val
 85 90 95
 Thr Ala Asn Ala Glu Ser Asp Pro Val Val Ala Leu Ala Gly Ala Val
 100 105 110
 Pro Arg Thr Asp Arg Leu Lys Arg Thr His Gln Ser Met Asp Asn Ala
 115 120 125
 Ala Leu Phe Glu Pro Ile Thr Lys Tyr Ser Val Glu Val Glu His Pro
 130 135 140
 Asp Asn Val Pro Glu Ala Leu Ser Asn Ala Phe Arg Ser Ala Thr Ser
 145 150 155 160
 Thr Asn Pro Gly Ala Thr Leu Val Ser Leu Pro Gln Asp Val Met Thr
 165 170 175
 Ala Glu Thr Thr Val Glu Ser Ile Gly Ala Leu Ser Lys Pro Gln Leu
 180 185 190
 Gly Ile Ala Pro Thr His Asp Ile Thr Tyr Val Val Asp Lys Ile Lys
 195 200 205
 Ala Ala Lys Leu Pro Val Ile Leu Leu Gly Met Arg Ala Ser Thr Asn
 210 215 220
 Glu Val Thr Lys Ala Val Arg Lys Leu Ile Ala Asp Thr Glu Leu Pro
 225 230 235 240
 Val Val Glu Thr Tyr Gln Ala Ala Gly Ala Ile Ser Arg Glu Leu Glu
 245 250 255
 Asp His Phe Phe Gly Arg Val Gly Leu Phe Arg Asn Gln Pro Gly Asp
 260 265 270
 Ile Leu Leu Glu Glu Ala Asp Leu Val Ile Ser Ile Gly Tyr Asp Pro
 275 280 285
 Ile Glu Tyr Asp Pro Lys Phe Trp Asn Lys Leu Gly Asp Arg Thr Ile
 290 295 300
 Ile His Leu Asp Asp His Gln Ala Asp Ile Asp His Asp Tyr Gln Pro
 305 310 315 320

20110907_CL5178USNA_SeqList.txt

Glu Arg Glu Leu Ile Gly Asp Ile Ala Leu Thr Val Asn Ser Ile Ala
325 330 335

Glu Lys Leu Pro Lys Leu Val Leu Ser Thr Lys Ser Glu Ala Val Leu
340 345 350

Glu Arg Leu Arg Ala Lys Leu Ser Glu Gln Ala Glu Val Pro Asn Arg
355 360 365

Pro Ser Glu Gly Val Thr His Pro Leu Gln Val Ile Arg Thr Leu Arg
370 375 380

Ser Leu Ile Ser Asp Asp Thr Thr Val Thr Cys Asp Ile Gly Ser His
385 390 395 400

Ser Ile Trp Met Ala Arg Cys Phe Arg Ser Tyr Glu Pro Arg Arg Leu
405 410 415

Leu Phe Ser Asn Gly Met Gln Thr Leu Gly Val Ala Leu Pro Trp Ala
420 425 430

Ile Ala Ala Thr Leu Val Glu Pro Gly Lys Lys Val Val Ser Val Ser
435 440 445

Gly Asp Gly Gly Phe Leu Phe Ser Ala Met Glu Leu Glu Thr Ala Val
450 455 460

Arg Leu Asn Ser Pro Ile Val His Leu Val Trp Arg Asp Gly Thr Tyr
465 470 475 480

Asp Met Val Ala Phe Gln Gln Met Met Lys Tyr Gly Arg Thr Ser Ala
485 490 495

Thr Glu Phe Gly Asp Val Asp Leu Val Lys Tyr Ala Glu Ser Phe Gly
500 505 510

Ala Leu Gly Leu Arg Val Asn Thr Pro Asp Glu Leu Glu Gly Val Leu
515 520 525

Lys Glu Ala Leu Ala Ala Asp Gly Pro Val Ile Ile Asp Ile Pro Ile
530 535 540

Asp Tyr Arg Asp Asn Ile Lys Leu Ser Glu Lys Leu Leu Pro Asn Gln
545 550 555 560

Leu Asn

<210> 19
<211> 780
<212> DNA

<213> Klebsiella pneumoniae

<400> 19

```

atgaatcatt ctgctgaatg cacctgcgaa gagagtctat gcgaaaccct gcgggcgttt      60
tccgcgcagc atcccgagag cgtgctctat cagacatcgc tcatgagcgc cctgctgagc      120
ggggtttacg aaggcagcac caccatcgcg gacctgctga aacacggcga tttcggcctc      180
ggcaccttta atgagctgga cggggagctg atcgcccttca gcagtcaggt ctatcagctg      240
cgcgccgacg gcagcgcgcg caaagcccag ccggagcaga aaacgccgtt cgcggtgatg      300
acctggttcc agccgcagta ccggaacc tttgaccatc cggtagaccg ccagcagctg      360
cacgaggtga tcgaccagca aatcccctct gacaacctgt tctgcgccct gcgcatcgac      420
ggccatttcc gccatgcccc taccgcacc gtgccgcgcc agacgccgcc gtaccgggcg      480
atgaccgacg tcctcgacga tcagccggtg ttccgcttta accagcgcga aggggtgctg      540
gtcggcttcc ggaccccga gcatatgcag gggatcaacg tcgccgggta tcacgagcac      600
tttattaccg atgaccgcaa aggcggcggt cacctgctgg attaccagct cgaccatggg      660
gtgctgacct tcggcgaaat tcacaagctg atgatcgacc tgcccgccga cagcgcgttc      720
ctgcaggcta atctgcatcc cgataatctc gatgccgcca tccgttccgt agaaagttaa      780

```

<210> 20

<211> 259

<212> PRT

<213> Klebsiella pneumoniae

<400> 20

```

Met Asn His Ser Ala Glu Cys Thr Cys Glu Glu Ser Leu Cys Glu Thr
1          5          10          15

```

```

Leu Arg Ala Phe Ser Ala Gln His Pro Glu Ser Val Leu Tyr Gln Thr
          20          25          30

```

```

Ser Leu Met Ser Ala Leu Leu Ser Gly Val Tyr Glu Gly Ser Thr Thr
          35          40          45

```

```

Ile Ala Asp Leu Leu Lys His Gly Asp Phe Gly Leu Gly Thr Phe Asn
          50          55          60

```

```

Glu Leu Asp Gly Glu Leu Ile Ala Phe Ser Ser Gln Val Tyr Gln Leu
65          70          75          80

```

```

Arg Ala Asp Gly Ser Ala Arg Lys Ala Gln Pro Glu Gln Lys Thr Pro
          85          90          95

```

```

Phe Ala Val Met Thr Trp Phe Gln Pro Gln Tyr Arg Lys Thr Phe Asp
          100          105          110

```

```

His Pro Val Ser Arg Gln Gln Leu His Glu Val Ile Asp Gln Gln Ile
          115          120          125

```

20110907_CL5178USNA_SeqList.txt

Pro Ser Asp Asn Leu Phe Cys Ala Leu Arg Ile Asp Gly His Phe Arg
130 135 140

His Ala His Thr Arg Thr Val Pro Arg Gln Thr Pro Pro Tyr Arg Ala
145 150 155 160

Met Thr Asp Val Leu Asp Asp Gln Pro Val Phe Arg Phe Asn Gln Arg
165 170 175

Glu Gly Val Leu Val Gly Phe Arg Thr Pro Gln His Met Gln Gly Ile
180 185 190

Asn Val Ala Gly Tyr His Glu His Phe Ile Thr Asp Asp Arg Lys Gly
195 200 205

Gly Gly His Leu Leu Asp Tyr Gln Leu Asp His Gly Val Leu Thr Phe
210 215 220

Gly Glu Ile His Lys Leu Met Ile Asp Leu Pro Ala Asp Ser Ala Phe
225 230 235 240

Leu Gln Ala Asn Leu His Pro Asp Asn Leu Asp Ala Ala Ile Arg Ser
245 250 255

Val Glu Ser

<210> 21
<211> 768
<212> DNA
<213> Bacillus subtilis

<400> 21
atgaaacgag aaagcaacat tcaagtgtct agccgtggct aaaaagatca gcctgtgagc 60
cagatttatc aagtatcaac aatgacttct ctattagacg gagtatatga cggagatttt 120
gaactgtcag agattccgaa atatggagac ttcggtatcg gaacctttta caagcttgac 180
ggagagctga ttgggtttga cggcgaattt taccgtcttc gctcagacgg aaccgcgaca 240
ccggtccaaa atggagaccg ttcaccgttc tgttcattta cgttctttac accggacatg 300
acgcacaaaa ttgatgcaa aatgacacgc gaagactttg aaaaagagat caacagcatg 360
ctgccaagca gaaacttatt ttatgcaatt cgcattgacg gattgtttta aaaggtgcag 420
acaagaacag tagaacttca agaaaaacct tacgtgccaa tggttgaagc ggtcaaaaca 480
cagccgattt tcaacttcga caacgtgaga ggaacgattg taggtttctt gacaccagct 540
tatgcaaacg gaatcgccgt ttctggctat cacctgcact tcattgacga aggacgcaat 600
tcaggcggac acgtttttga ctatgtgctt gaggattgca cggttacgat ttctcaaaaa 660
atgaacatga atctcagact tccgaacaca gcggatttct ttaatgcaa tctggataac 720
cctgattttg cgaaagatat cgaaacaact gaaggaagcc ctgaataa 768

<210> 22
 <211> 255
 <212> PRT
 <213> Bacillus subtilis

<400> 22

Met Lys Arg Glu Ser Asn Ile Gln Val Leu Ser Arg Gly Gln Lys Asp
 1 5 10 15

Gln Pro Val Ser Gln Ile Tyr Gln Val Ser Thr Met Thr Ser Leu Leu
 20 25 30

Asp Gly Val Tyr Asp Gly Asp Phe Glu Leu Ser Glu Ile Pro Lys Tyr
 35 40 45

Gly Asp Phe Gly Ile Gly Thr Phe Asn Lys Leu Asp Gly Glu Leu Ile
 50 55 60

Gly Phe Asp Gly Glu Phe Tyr Arg Leu Arg Ser Asp Gly Thr Ala Thr
 65 70 75 80

Pro Val Gln Asn Gly Asp Arg Ser Pro Phe Cys Ser Phe Thr Phe Phe
 85 90 95

Thr Pro Asp Met Thr His Lys Ile Asp Ala Lys Met Thr Arg Glu Asp
 100 105 110

Phe Glu Lys Glu Ile Asn Ser Met Leu Pro Ser Arg Asn Leu Phe Tyr
 115 120 125

Ala Ile Arg Ile Asp Gly Leu Phe Lys Lys Val Gln Thr Arg Thr Val
 130 135 140

Glu Leu Gln Glu Lys Pro Tyr Val Pro Met Val Glu Ala Val Lys Thr
 145 150 155 160

Gln Pro Ile Phe Asn Phe Asp Asn Val Arg Gly Thr Ile Val Gly Phe
 165 170 175

Leu Thr Pro Ala Tyr Ala Asn Gly Ile Ala Val Ser Gly Tyr His Leu
 180 185 190

His Phe Ile Asp Glu Gly Arg Asn Ser Gly Gly His Val Phe Asp Tyr
 195 200 205

Val Leu Glu Asp Cys Thr Val Thr Ile Ser Gln Lys Met Asn Met Asn
 210 215 220

Leu Arg Leu Pro Asn Thr Ala Asp Phe Phe Asn Ala Asn Leu Asp Asn
 225 230 235 240

20110907_CL5178USNA_SeqList.txt

Pro Asp Phe Ala Lys Asp Ile Glu Thr Thr Glu Gly Ser Pro Glu
245 250 255

<210> 23
<211> 780
<212> DNA
<213> Klebsiella terrigena

<400> 23
gtgaatcatt atcctgaatg cacctgccag gagagcctgt gcgaaaccgt acgcggcttc 60
tccgcccacc accctgatag cgttatctat cagacctctc tgatgagcgc gctgctgagc 120
ggggtctatg agggtagcac caccatcgcc gacctgctga cccacggcga cttcgggtctc 180
ggcaccttta acgaactcga tggcgaactg attgccttta gcagcgaggt ctaccagctg 240
cgcgctgacg gcagcgcgcg taaagcccgg gcggatcaaa aaacgccctt cgcggtgatg 300
acctgggttca gaccgcagta ccgtaaaacc tttgaccacc cggtcagccg ccagcagctg 360
cacgacgtta tcgaccagca aatcccctcc gataacctgt tctgcgccct gcatattgat 420
ggtcactttc gccacgccc aacccgcacc gtgccgcggc agacgccgcc ctatcgggcg 480
atgaccgacg tgctcgatga ccagccggtt ttccgcttca accagcgcaa ggggacgctg 540
gtcggctttc gcaccccga gcatatgcag ggccttaacg ttgccggcta ccacgagcac 600
tttattaccg acgatcgcca gggcggcggc catctgctgg actaccagct cgatagcggc 660
gtgctgacct tcggcgagat ccacaagctg atgattgacc tcccggccga cagcgctttc 720
ctgcaggccg acctgcatcc tgacaatctc gatgccgcta ttcgtgcggt agaaaactaa 780

<210> 24
<211> 259
<212> PRT
<213> Klebsiella terrigena

<400> 24

Met Asn His Tyr Pro Glu Cys Thr Cys Gln Glu Ser Leu Cys Glu Thr
1 5 10 15

Val Arg Gly Phe Ser Ala His His Pro Asp Ser Val Ile Tyr Gln Thr
20 25 30

Ser Leu Met Ser Ala Leu Leu Ser Gly Val Tyr Glu Gly Ser Thr Thr
35 40 45

Ile Ala Asp Leu Leu Thr His Gly Asp Phe Gly Leu Gly Thr Phe Asn
50 55 60

Glu Leu Asp Gly Glu Leu Ile Ala Phe Ser Ser Glu Val Tyr Gln Leu
65 70 75 80

Arg Ala Asp Gly Ser Ala Arg Lys Ala Arg Ala Asp Gln Lys Thr Pro
85 90 95

20110907_CL5178USNA_SeqList.txt

Phe Ala Val Met Thr Trp Phe Arg Pro Gln Tyr Arg Lys Thr Phe Asp
100 105 110

His Pro Val Ser Arg Gln Gln Leu His Asp Val Ile Asp Gln Gln Ile
115 120 125

Pro Ser Asp Asn Leu Phe Cys Ala Leu His Ile Asp Gly His Phe Arg
130 135 140

His Ala His Thr Arg Thr Val Pro Arg Gln Thr Pro Pro Tyr Arg Ala
145 150 155 160

Met Thr Asp Val Leu Asp Asp Gln Pro Val Phe Arg Phe Asn Gln Arg
165 170 175

Lys Gly Thr Leu Val Gly Phe Arg Thr Pro Gln His Met Gln Gly Leu
180 185 190

Asn Val Ala Gly Tyr His Glu His Phe Ile Thr Asp Asp Arg Gln Gly
195 200 205

Gly Gly His Leu Leu Asp Tyr Gln Leu Asp Ser Gly Val Leu Thr Phe
210 215 220

Gly Glu Ile His Lys Leu Met Ile Asp Leu Pro Ala Asp Ser Ala Phe
225 230 235 240

Leu Gln Ala Asp Leu His Pro Asp Asn Leu Asp Ala Ala Ile Arg Ala
245 250 255

Val Glu Asn

<210> 25
<211> 771
<212> DNA
<213> Klebsiella pneumoniae

<400> 25
atgaaaaaag tcgcacttgt taccggcgcc ggccagggga ttggtaaagc tatcgccctt 60
cgtctggtga aggatggatt tgccgtggcc attgccgatt ataacgacgc caccgccaaa 120
gcggtcgcct cggaatcaa ccaggccggc ggacacgccg tggcggtgaa agtggatgtc 180
tccgaccgcg atcaggtatt tgccgccgtt gaacaggcgc gcaaaacgct gggcggcttc 240
gacgtcatcg tcaataacgc cgggtgtggca ccgtctacgc cgatcgagtc cattaccccg 300
gagattgtcg acaaagtcta caacatcaac gtcaaagggg tgatctgggg tattcaggcg 360
gcggtcgagg ctttaagaa agaggggcac ggcgggaaaa tcatcaacgc ctgttcccag 420
gccggccacg tcggcaaccc ggagctggcg gtgtatagct ccagtaaatt cgcggtacgc 480
ggcttaaccc agaccgccgc tcgcgacctc gcgccgtgg gcatcacggt caacggctac 540

20110907_CL5178USNA_SeqList.txt

tgcccgggga ttgtcaaaac gccaatgtgg gccgaaattg accgccaggt gtccgaagcc	600
gccggtaaac cgctgggcta cggtaccgcc gagttcgcca aacgcatcac tctcggtcgt	660
ctgtccgagc cggaagatgt cgccgcctgc gtctcctatc ttgccagccc ggattctgat	720
tacatgaccg gtcagtcgtt gctgatcgac ggcgggatgg tatttaacta a	771

<210> 26
 <211> 256
 <212> PRT
 <213> Klebsiella pneumoniae

<400> 26

Met	Lys	Lys	Val	Ala	Leu	Val	Thr	Gly	Ala	Gly	Gln	Gly	Ile	Gly	Lys
1				5				10						15	

Ala	Ile	Ala	Leu	Arg	Leu	Val	Lys	Asp	Gly	Phe	Ala	Val	Ala	Ile	Ala
			20					25					30		

Asp	Tyr	Asn	Asp	Ala	Thr	Ala	Lys	Ala	Val	Ala	Ser	Glu	Ile	Asn	Gln
		35					40					45			

Ala	Gly	Gly	His	Ala	Val	Ala	Val	Lys	Val	Asp	Val	Ser	Asp	Arg	Asp
	50					55					60				

Gln	Val	Phe	Ala	Ala	Val	Glu	Gln	Ala	Arg	Lys	Thr	Leu	Gly	Gly	Phe
65					70					75					80

Asp	Val	Ile	Val	Asn	Asn	Ala	Gly	Val	Ala	Pro	Ser	Thr	Pro	Ile	Glu
				85					90					95	

Ser	Ile	Thr	Pro	Glu	Ile	Val	Asp	Lys	Val	Tyr	Asn	Ile	Asn	Val	Lys
			100					105					110		

Gly	Val	Ile	Trp	Gly	Ile	Gln	Ala	Ala	Val	Glu	Ala	Phe	Lys	Lys	Glu
		115					120					125			

Gly	His	Gly	Gly	Lys	Ile	Ile	Asn	Ala	Cys	Ser	Gln	Ala	Gly	His	Val
	130					135					140				

Gly	Asn	Pro	Glu	Leu	Ala	Val	Tyr	Ser	Ser	Ser	Lys	Phe	Ala	Val	Arg
145					150					155					160

Gly	Leu	Thr	Gln	Thr	Ala	Ala	Arg	Asp	Leu	Ala	Pro	Leu	Gly	Ile	Thr
				165					170					175	

Val	Asn	Gly	Tyr	Cys	Pro	Gly	Ile	Val	Lys	Thr	Pro	Met	Trp	Ala	Glu
			180					185					190		

Ile	Asp	Arg	Gln	Val	Ser	Glu	Ala	Ala	Gly	Lys	Pro	Leu	Gly	Tyr	Gly
		195					200					205			

20110907_CL5178USNA_SeqList.txt

Thr Ala Glu Phe Ala Lys Arg Ile Thr Leu Gly Arg Leu Ser Glu Pro
210 215 220

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

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

<210> 27
<211> 1053
<212> DNA
<213> Bacillus cereus

<400> 27
atgaaagcac tactttggca taatcaacgt gatgtacgag tagaagaagt accagaacca 60
acagtaaaac caggaacagt gaaaatcaaa gttaaattgt gtggtatttg tgggacagac 120
ttgcatgaat atttagcagg gcctatTTTT attccaacag aagaacatcc attaacacat 180
gtgaaagcac ctgttatttt aggtcatgag tttagtgggtg aggtaataga gattgggtgaa 240
ggagttacat ctcataaagt gggagaccgc gttgtttagtag agccaattta ttcttgtggt 300
aaatgtgaag cttgtaaaca tggacattac aatgtttgtg aacaacttgt tttccacggt 360
cttggcggag aaggcggcgg tttctctgaa tatacagtag taccagaaga tatggttcat 420
cacattccag atgaaatgac gtatgaacaa ggtgcgcttg tagaaccagc agcagtagca 480
gttcatgcag tacgtcaaag taaattaaaa gaaggggaag ctgtagcgggt atttggttgc 540
ggccaattg gacttcttgt tatccaagca gctaaagcag caggagcaac tcctgttatt 600
gcagttgaac tttctaaaga acgtcaagag ttagcgaaat tagcaggtgc ggattatgta 660
ttaaatccag caactcaaga tgtgttagct gaaattcgta acttaacaaa tggtttaggt 720
gtaaatgtta gctttgaagt aacagggtgtt gaagttgtac tacgccaagc gattgaaagt 780
acaagcttcg aaggacaaac tgtaattgtt agtgtatggg aaaaagacgc aacaattact 840
ccaaataact tagtattaaa agaaaaagaa gttattggta ttttaggata ccgtcacatc 900
ttcccagctg ttattaaatt gattagctcc ggtcaaattc aagcagagaa attaattacg 960
aaaaaaatta cagtggatca agttgttgaa gaaggatttg aagcacttgt aaaagataaa 1020
acacaagtga aaattcttgt ttcacctaaa taa 1053

<210> 28
<211> 350
<212> PRT
<213> Bacillus cereus

<400> 28

Met Lys Ala Leu Leu Trp His Asn Gln Arg Asp Val Arg Val Glu Glu
1 5 10 15

Val Pro Glu Pro Thr Val Lys Pro Gly Thr Val Lys Ile Lys Val Lys

20

25

30

Trp Cys Gly Ile Cys Gly Thr Asp Leu His Glu Tyr Leu Ala Gly Pro
 35 40 45

Ile Phe Ile Pro Thr Glu Glu His Pro Leu Thr His Val Lys Ala Pro
 50 55 60

Val Ile Leu Gly His Glu Phe Ser Gly Glu Val Ile Glu Ile Gly Glu
 65 70 75 80

Gly Val Thr Ser His Lys Val Gly Asp Arg Val Val Val Glu Pro Ile
 85 90 95

Tyr Ser Cys Gly Lys Cys Glu Ala Cys Lys His Gly His Tyr Asn Val
 100 105 110

Cys Glu Gln Leu Val Phe His Gly Leu Gly Gly Glu Gly Gly Gly Phe
 115 120 125

Ser Glu Tyr Thr Val Val Pro Glu Asp Met Val His His Ile Pro Asp
 130 135 140

Glu Met Thr Tyr Glu Gln Gly Ala Leu Val Glu Pro Ala Ala Val Ala
 145 150 155 160

Val His Ala Val Arg Gln Ser Lys Leu Lys Glu Gly Glu Ala Val Ala
 165 170 175

Val Phe Gly Cys Gly Pro Ile Gly Leu Leu Val Ile Gln Ala Ala Lys
 180 185 190

Ala Ala Gly Ala Thr Pro Val Ile Ala Val Glu Leu Ser Lys Glu Arg
 195 200 205

Gln Glu Leu Ala Lys Leu Ala Gly Ala Asp Tyr Val Leu Asn Pro Ala
 210 215 220

Thr Gln Asp Val Leu Ala Glu Ile Arg Asn Leu Thr Asn Gly Leu Gly
 225 230 235 240

Val Asn Val Ser Phe Glu Val Thr Gly Val Glu Val Val Leu Arg Gln
 245 250 255

Ala Ile Glu Ser Thr Ser Phe Glu Gly Gln Thr Val Ile Val Ser Val
 260 265 270

Trp Glu Lys Asp Ala Thr Ile Thr Pro Asn Asn Leu Val Leu Lys Glu
 275 280 285

Lys Glu Val Ile Gly Ile Leu Gly Tyr Arg His Ile Phe Pro Ala Val
 Page 43

290

295

300

Ile Lys Leu Ile Ser Ser Gly Gln Ile Gln Ala Glu Lys Leu Ile Thr
 305 310 315 320

Lys Lys Ile Thr Val Asp Gln Val Val Glu Glu Gly Phe Glu Ala Leu
 325 330 335

Val Lys Asp Lys Thr Gln Val Lys Ile Leu Val Ser Pro Lys
 340 345 350

<210> 29
 <211> 1113
 <212> DNA
 <213> Lactococcus lactis

<400> 29
 ttgcctgaaa cgacaacat cctatataga ggaggcgttt ttatgcgcgc agcacgtttt 60
 tacgaccgcg gggatatccg cattgatgaa attaatgaac caatagtaaa agctggccaa 120
 gttggcattg atgtggcttg gtgtggaatt tgtggaacag atctccatga atttttagat 180
 ggcccaattt tttgtccgtc agcagaacat cctaatacaa ttactggaga agtaccacca 240
 gtcactcttg gacatgaaat gtctgggggt gtaaatttta taggtgaagg agtaagcgga 300
 cttaaagtag gtgaccatgt cgttgtcgaa ccttatatcg ttcccgaagg gactgataca 360
 agtgaaactg gacattataa cctctcagaa ggctcaaact ttattgggtt gggcggaat 420
 ggtggagggt tggctgaaaa aatttctgtt gatgaacgtt ggggttcacaa aattcctgat 480
 aacttaccat tggatgaagc tgctctaatt gagccactat cagtcggcta tcacgctgtt 540
 gaacgagcaa atttaagtga aaagagtacg gtattagttg ttggtgctgg accaattgga 600
 ctattaactg ctgccgttgc aaaagcgcaa ggacatactg ttatcatcag tgaacctagt 660
 ggacttcgtc gtaaaaaagc acaagaagca caagttgctg attatttctt caatccaatt 720
 gaagatgaca ttcaagctaa agttcatgaa attaatgaaa aaggagtgga cgcagccttt 780
 gaatgtacct ctgtccaacc gggatttgac gcttgtctag atgcgattcg tatgggtgga 840
 acagttgtca ttgtcgcaat ttggggcaag cctgctagtg ttgatatggc aaaattagta 900
 atcaaagaag ctaacctttt aggaacgatt gcttataata aactcatcc aaaaacaatt 960
 gatttagtat caacaggtaa aataaaattg gaccaattca tcacagctaa aatcggtttg 1020
 gatgatttga ttgataaagg attcgatacg ctgattcatc ataataaaac agctgttaaa 1080
 attttagttt caccaactgg taaaggtcta taa 1113

<210> 30
 <211> 370
 <212> PRT
 <213> Lactococcus lactis

<400> 30

Met Pro Glu Thr Thr Thr Ile Leu Tyr Arg Gly Gly Val Phe Met Arg
 Page 44

```

1           5           10           15
Ala Ala Arg Phe Tyr Asp Arg Gly Asp Ile Arg Ile Asp Glu Ile Asn
    20           25           30
Glu Pro Ile Val Lys Ala Gly Gln Val Gly Ile Asp Val Ala Trp Cys
    35           40           45
Gly Ile Cys Gly Thr Asp Leu His Glu Phe Leu Asp Gly Pro Ile Phe
    50           55           60
Cys Pro Ser Ala Glu His Pro Asn Pro Ile Thr Gly Glu Val Pro Pro
    65           70           75           80
Val Thr Leu Gly His Glu Met Ser Gly Val Val Asn Phe Ile Gly Glu
    85           90           95
Gly Val Ser Gly Leu Lys Val Gly Asp His Val Val Val Glu Pro Tyr
    100          105          110
Ile Val Pro Glu Gly Thr Asp Thr Ser Glu Thr Gly His Tyr Asn Leu
    115          120          125
Ser Glu Gly Ser Asn Phe Ile Gly Leu Gly Gly Asn Gly Gly Gly Leu
    130          135          140
Ala Glu Lys Ile Ser Val Asp Glu Arg Trp Val His Lys Ile Pro Asp
    145          150          155          160
Asn Leu Pro Leu Asp Glu Ala Ala Leu Ile Glu Pro Leu Ser Val Gly
    165          170          175
Tyr His Ala Val Glu Arg Ala Asn Leu Ser Glu Lys Ser Thr Val Leu
    180          185          190
Val Val Gly Ala Gly Pro Ile Gly Leu Leu Thr Ala Ala Val Ala Lys
    195          200          205
Ala Gln Gly His Thr Val Ile Ile Ser Glu Pro Ser Gly Leu Arg Arg
    210          215          220
Lys Lys Ala Gln Glu Ala Gln Val Ala Asp Tyr Phe Phe Asn Pro Ile
    225          230          235          240
Glu Asp Asp Ile Gln Ala Lys Val His Glu Ile Asn Glu Lys Gly Val
    245          250          255
Asp Ala Ala Phe Glu Cys Thr Ser Val Gln Pro Gly Phe Asp Ala Cys
    260          265          270
Leu Asp Ala Ile Arg Met Gly Gly Thr Val Val Ile Val Ala Ile Trp

```

275

280

285

Gly Lys Pro Ala Ser Val Asp Met Ala Lys Leu Val Ile Lys Glu Ala
 290 295 300

Asn Leu Leu Gly Thr Ile Ala Tyr Asn Asn Thr His Pro Lys Thr Ile
 305 310 315 320

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

Lys Ile Gly Leu Asp Asp Leu Ile Asp Lys Gly Phe Asp Thr Leu Ile
 340 345 350

His His Asn Glu Thr Ala Val Lys Ile Leu Val Ser Pro Thr Gly Lys
 355 360 365

Gly Leu
 370

<210> 31

<211> 2532

<212> DNA

<213> Roseburia inulinivorans

<400> 31

atgggcaatt acgattcaac accgatagct aaaagtgata ggattaaaag attggttgat 60
 catttgatg ctaaaatgcc tgaaattgag gccgctagag cagagctaata tactgaatcc 120
 tttaaggcca ccgaagggtca acctgttggt atgagaaagg ctagagcttt tgaacatata 180
 ctaaagaatt tgccaattat cataagacca gaagaactga ttgttggtc aactacaatt 240
 gccctagag gttgccaaac gtatccagaa ttctcatag agtggttaga ggctgaattt 300
 gaaactgtcg aaacgcgttc agctgaccca ttttatattt cagaagaaac gaagaaacgt 360
 ttgctggctg ccgatgctta ttggaaagg aaacaacct cagagttggc aacttcatat 420
 atggccccag aaactctaag agccatgaag cataacttct tcaccctgg aaactacttc 480
 tacaatggtg tcggtcatgt cacagttcaa tatgaaacag tattagcaat cggcttgaat 540
 ggagtaaaag agaaggtag gaaagagatg gagaattgtc attttggtga tgccgattat 600
 agtacaaga tgtgtttctt ggagagcatt ttaatatcgt gtgatgccgt aatcacttat 660
 gctaatagat atgccaagat ggccgaggaa atggctgaaa aagaaacaga tgctgcaagg 720
 aggcaagaac tattaacaat cgccagggtt tgcaaaaacg ttcctgaatt ccagccgaa 780
 agcttccagg aggcctgcca atccttttgg ttcatacaac aagtgttca aattgaatcc 840
 agtggctcatt caatttccc aggtagattt gatcaatata tgtatcctta ttacgaaaag 900
 gatttaaagg aaggtagctt aactagggaa tatgctcagg aactgatcga ttgtatctgg 960
 gttaagttaa atgatctgaa taagtgcagg gatgctgcct ctgctgaggg ctttgcagga 1020
 tattccttat ttcaaaactt aatcgttggg ggccaaacgg ttcaaggaag ggacgccacc 1080

20110907_CL5178USNA_SeqList.txt

```

aatgatttga gttttatgtg tatcacggca tctgaacacg tctttttacc gatgccgtcg      1140
ttgtctataa gagtttggca tggtagttcc aaagcactgc ttatgagagc agctgaattg      1200
actagaaccg gtataggctt acctgcttat tacaatgatg aagtcatcat accagctttg      1260
gtgcataggg gtgctactat ggatgaagca agaaattaca acataatagg atgtgtcgaa      1320
ccgcaggttc ctggtaaaac tgatggctgg cacgatgcag cattctttta catgtgcaga      1380
cctttggaaa tgggtgtttag taatggttat gataacggtg aaattgcatc tatacaaact      1440
ggtaacgtag aatcttttca gagttttgat gagtttatgg aagcttacag aaaacaaatg      1500
ctatataaca tagaacttat ggtaaattgcc gacaacgcga tagattatgc ccacgcaaag      1560
ttggcccat tgccatttga gtcattgtttg gttgatgact gtataaagag aggaatgtcc      1620
gctcaggaag gcggcgcaat ctataatttc actggtccac agggcttttg tattgcaaac      1680
gttgctgata gcttgatac gattaagaaa ttggtgttcg aggagaagag aattacgatg      1740
ggtgaattaa agaaagcgtt ggaaatgaat tatggtaagg gtttggtatgc cacaaccgct      1800
ggtgacatcg caatgcaggt cgcgaggga ctaaaagatg ccggacagga agtgggtccc      1860
gacgtgatcg ctaatacaat ccgtcaagtt cttgaaatgg aattaccaga agatgtaaga      1920
aagagatatg aagagatcca tgaaatgata cttgagttac caaagtatgg taatgatata      1980
gatgaagttg atgaattagc tagagaagca gcttactttt acacaagacc attagaaact      2040
tttaagaatc caaggggtgg catgtatcaa gccggccttt atcccgtgtc cgctaattgtg      2100
ccactaggcg ctcaaacggg ggccacaccc gatggacggt tggcgcatat acccgtggcg      2160
gatggcggtg gtccgacatc aggcttcgat atatccggac caacagcttc ttgcaattct      2220
gtcgccaagt tggatcatgc tatagcctct aatggtacct tatttaatat gaagatgcac      2280
ccaaccgcaa tggcagggtga aaagggctta gaatccttca tatcgttgat ccgtggttat      2340
ttcgatcaac aaggatatga catgcaattt aacgtagtag acagggctac actgcttgat      2400
gcgcaggccc accctgaaaa gtattcaggc ttaattgtca gagtggcagg ttattctgcc      2460
ctttttacca cattgtccaa gtcattacaa gatgatataa tcaaacgtac cgaacaagca      2520
gacaatagat ag                                                                2532

```

<210> 32
 <211> 843
 <212> PRT
 <213> Roseburia inulinivorans

<400> 32

Met Gly Asn Tyr Asp Ser Thr Pro Ile Ala Lys Ser Asp Arg Ile Lys
 1 5 10 15

Arg Leu Val Asp His Leu Tyr Ala Lys Met Pro Glu Ile Glu Ala Ala
 20 25 30

Arg Ala Glu Leu Ile Thr Glu Ser Phe Lys Ala Thr Glu Gly Gln Pro

35

40

45

Val Val Met Arg Lys Ala Arg Ala Phe Glu His Ile Leu Lys Asn Leu
 50 55 60
 Pro Ile Ile Ile Arg Pro Glu Glu Leu Ile Val Gly Ser Thr Thr Ile
 65 70 75 80
 Ala Pro Arg Gly Cys Gln Thr Tyr Pro Glu Phe Ser Tyr Glu Trp Leu
 85 90 95
 Glu Ala Glu Phe Glu Thr Val Glu Thr Arg Ser Ala Asp Pro Phe Tyr
 100 105 110
 Ile Ser Glu Glu Thr Lys Lys Arg Leu Leu Ala Ala Asp Ala Tyr Trp
 115 120 125
 Lys Gly Lys Thr Thr Ser Glu Leu Ala Thr Ser Tyr Met Ala Pro Glu
 130 135 140
 Thr Leu Arg Ala Met Lys His Asn Phe Phe Thr Pro Gly Asn Tyr Phe
 145 150 155 160
 Tyr Asn Gly Val Gly His Val Thr Val Gln Tyr Glu Thr Val Leu Ala
 165 170 175
 Ile Gly Leu Asn Gly Val Lys Glu Lys Val Arg Lys Glu Met Glu Asn
 180 185 190
 Cys His Phe Gly Asp Ala Asp Tyr Ser Thr Lys Met Cys Phe Leu Glu
 195 200 205
 Ser Ile Leu Ile Ser Cys Asp Ala Val Ile Thr Tyr Ala Asn Arg Tyr
 210 215 220
 Ala Lys Met Ala Glu Glu Met Ala Glu Lys Glu Thr Asp Ala Ala Arg
 225 230 235 240
 Arg Gln Glu Leu Leu Thr Ile Ala Arg Val Cys Lys Asn Val Pro Glu
 245 250 255
 Phe Pro Ala Glu Ser Phe Gln Glu Ala Cys Gln Ser Phe Trp Phe Ile
 260 265 270
 Gln Gln Val Leu Gln Ile Glu Ser Ser Gly His Ser Ile Ser Pro Gly
 275 280 285
 Arg Phe Asp Gln Tyr Met Tyr Pro Tyr Tyr Glu Lys Asp Leu Lys Glu
 290 295 300
 Gly Ser Leu Thr Arg Glu Tyr Ala Gln Glu Leu Ile Asp Cys Ile Trp

305 310 315 320
 Val Lys Leu Asn Asp₃₂₅ Leu Asn Lys Cys Arg₃₃₀ Asp Ala Ala Ser Ala₃₃₅ Glu
 Gly Phe Ala Gly₃₄₀ Tyr Ser Leu Phe Gln₃₄₅ Asn Leu Ile Val Gly₃₅₀ Gly Gln
 Thr Val Gln₃₅₅ Gly Arg Asp Ala Thr₃₆₀ Asn Asp Leu Ser Phe₃₆₅ Met Cys Ile
 Thr Ala₃₇₀ Ser Glu His Val Phe₃₇₅ Leu Pro Met Pro Ser₃₈₀ Leu Ser Ile Arg
 Val₃₈₅ Trp His Gly Ser Ser₃₉₀ Lys Ala Leu Leu Met₃₉₅ Arg Ala Ala Glu Leu₄₀₀
 Thr Arg Thr Gly Ile₄₀₅ Gly Leu Pro Ala Tyr₄₁₀ Tyr Asn Asp Glu Val₄₁₅ Ile
 Ile Pro Ala Leu₄₂₀ Val His Arg Gly Ala₄₂₅ Thr Met Asp Glu Ala₄₃₀ Arg Asn
 Tyr Asn Ile₄₃₅ Ile Gly Cys Val Glu₄₄₀ Pro Gln Val Pro Gly₄₄₅ Lys Thr Asp
 Gly Trp₄₅₀ His Asp Ala Ala Phe₄₅₅ Phe Asn Met Cys Arg₄₆₀ Pro Leu Glu Met
 Val Phe Ser Asn Gly Tyr₄₇₀ Asp Asn Gly Glu Ile₄₇₅ Ala Ser Ile Gln Thr₄₈₀
 Gly Asn Val Glu Ser₄₈₅ Phe Gln Ser Phe Asp₄₉₀ Glu Phe Met Glu Ala₄₉₅ Tyr
 Arg Lys Gln Met₅₀₀ Leu Tyr Asn Ile Glu₅₀₅ Leu Met Val Asn Ala₅₁₀ Asp Asn
 Ala Ile Asp₅₁₅ Tyr Ala His Ala Lys₅₂₀ Leu Ala Pro Leu Pro₅₂₅ Phe Glu Ser
 Cys Leu₅₃₀ Val Asp Asp Cys Ile₅₃₅ Lys Arg Gly Met Ser₅₄₀ Ala Gln Glu Gly
 Gly Ala Ile Tyr Asn Phe₅₅₀ Thr Gly Pro Gln Gly₅₅₅ Phe Gly Ile Ala Asn₅₆₀
 Val Ala Asp Ser Leu₅₆₅ Tyr Thr Ile Lys Lys₅₇₀ Leu Val Phe Glu Glu₅₇₅ Lys
 Arg Ile Thr Met Gly Glu Leu Lys Lys Ala Leu Glu Met Asn Tyr Gly
 Page 49

580

585

590

Lys Gly Leu Asp Ala Thr Thr Ala Gly Asp Ile Ala Met Gln Val Ala
595 600 605

Lys Gly Leu Lys Asp Ala Gly Gln Glu Val Gly Pro Asp Val Ile Ala
610 615 620

Asn Thr Ile Arg Gln Val Leu Glu Met Glu Leu Pro Glu Asp Val Arg
625 630 635 640

Lys Arg Tyr Glu Glu Ile His Glu Met Ile Leu Glu Leu Pro Lys Tyr
645 650 655

Gly Asn Asp Ile Asp Glu Val Asp Glu Leu Ala Arg Glu Ala Ala Tyr
660 665 670

Phe Tyr Thr Arg Pro Leu Glu Thr Phe Lys Asn Pro Arg Gly Gly Met
675 680 685

Tyr Gln Ala Gly Leu Tyr Pro Val Ser Ala Asn Val Pro Leu Gly Ala
690 695 700

Gln Thr Gly Ala Thr Pro Asp Gly Arg Leu Ala His Thr Pro Val Ala
705 710 715 720

Asp Gly Val Gly Pro Thr Ser Gly Phe Asp Ile Ser Gly Pro Thr Ala
725 730 735

Ser Cys Asn Ser Val Ala Lys Leu Asp His Ala Ile Ala Ser Asn Gly
740 745 750

Thr Leu Phe Asn Met Lys Met His Pro Thr Ala Met Ala Gly Glu Lys
755 760 765

Gly Leu Glu Ser Phe Ile Ser Leu Ile Arg Gly Tyr Phe Asp Gln Gln
770 775 780

Gly Met His Met Gln Phe Asn Val Val Asp Arg Ala Thr Leu Leu Asp
785 790 795 800

Ala Gln Ala His Pro Glu Lys Tyr Ser Gly Leu Ile Val Arg Val Ala
805 810 815

Gly Tyr Ser Ala Leu Phe Thr Thr Leu Ser Lys Ser Leu Gln Asp Asp
820 825 830

Ile Ile Lys Arg Thr Glu Gln Ala Asp Asn Arg
835 840

<210> 33

<211> 794

<212> DNA

<213> Roseburia inulinivorans

<400> 33

```

atgaaagaat atcttaatac ttcaggtaga atatttgata tccagaggta ttctattcac      60
gatggccctg gtgtgcgtac aattgtgttt ctaaaagggt gtgcccttag atgcagatgg      120
tgctgtaatc ctgaaagcca aagcttcgaa gttgaaacaa tgacgattaa tggaaaacct      180
aaagtcatgg gtaaagatgt tacagtcgcc gaggttatga agacggtaga aagagacatg      240
ccttattacc ttcaatcagg tgggtggtatc accttatcgg gtggcgaaatg tactttgcaa      300
ccagaatttt cccttggcct attgagagct gcaaaggatt tgggcatatc cacggcaata      360
gagagcatgg cgtacgcaaa gtacgaagta atagaaactc ttcttccgta tttggatacg      420
tatttaatgg acatcaaaca tatgaatcct gagaaacata aagaatacac tggatcatgat      480
aacttgagga tgtagaaaa cgccttaaga gtcgcgcatt ctggtcagac cgaactgatc      540
atcagagtac ctgtcatccc aggattcaac gcaactgagc aggaactact agatattgca      600
aaattcgagc atacactgcc tggagttaga caaatacaca tcttgccata tcataatttt      660
ggtcagggta aatacgaagg attgaacagg gactatccga tggggggacac tgagaaaccc      720
tctaataaac agatgaaagc ttttcaagaa atgattcaaa agaacacttc cctacattgc      780
caaatcggtg gtta                                     794

```

<210> 34

<211> 264

<212> PRT

<213> roseburia inulinivorans

<400> 34

```

Met Lys Glu Tyr Leu Asn Thr Ser Gly Arg Ile Phe Asp Ile Gln Arg
1           5           10          15

```

```

Tyr Ser Ile His Asp Gly Pro Gly Val Arg Thr Ile Val Phe Leu Lys
          20          25          30

```

```

Gly Cys Ala Leu Arg Cys Arg Trp Cys Cys Asn Pro Glu Ser Gln Ser
          35          40          45

```

```

Phe Glu Val Glu Thr Met Thr Ile Asn Gly Lys Pro Lys Val Met Gly
          50          55          60

```

```

Lys Asp Val Thr Val Ala Glu Val Met Lys Thr Val Glu Arg Asp Met
65          70          75          80

```

```

Pro Tyr Tyr Leu Gln Ser Gly Gly Gly Ile Thr Leu Ser Gly Gly Glu
          85          90          95

```

```

Cys Thr Leu Gln Pro Glu Phe Ser Leu Gly Leu Leu Arg Ala Ala Lys
          100          105          110

```

20110907_CL5178USNA_SeqList.txt

Asp Leu Gly Ile Ser Thr Ala Ile Glu Ser Met Ala Tyr Ala Lys Tyr
115 120 125

Glu Val Ile Glu Thr Leu Leu Pro Tyr Leu Asp Thr Tyr Leu Met Asp
130 135 140

Ile Lys His Met Asn Pro Glu Lys His Lys Glu Tyr Thr Gly His Asp
145 150 155 160

Asn Leu Arg Met Leu Glu Asn Ala Leu Arg Val Ala His Ser Gly Gln
165 170 175

Thr Glu Leu Ile Ile Arg Val Pro Val Ile Pro Gly Phe Asn Ala Thr
180 185 190

Glu Gln Glu Leu Leu Asp Ile Ala Lys Phe Ala Asp Thr Leu Pro Gly
195 200 205

Val Arg Gln Ile His Ile Leu Pro Tyr His Asn Phe Gly Gln Gly Lys
210 215 220

Tyr Glu Gly Leu Asn Arg Asp Tyr Pro Met Gly Asp Thr Glu Lys Pro
225 230 235 240

Ser Asn Glu Gln Met Lys Ala Phe Gln Glu Met Ile Gln Lys Asn Thr
245 250 255

Ser Leu His Cys Gln Ile Gly Gly
260

<210> 35
<211> 1047
<212> DNA
<213> Achromobacter xylosoxidans

<400> 35
atgaaagctc tggtttatca cggtgaccac aagatctcgc ttgaagacaa gccaagccc 60
acccttcaaa agccacgga tgtagtagta cgggttttga agaccacgat ctgcggcacg 120
gatctcggca tctacaaagg caagaatcca gaggtcgccg acgggcgcat cctgggccat 180
gaaggggtag gcgtcatcga ggaagtgggc gagagtgtca cgcagttcaa gaaaggcgac 240
aaggtcctga tttcctgcgt cacttcttgc ggctcgtgcg actactgcaa gaagcagctt 300
tactccatt gccgcgacgg cgggtggatc ctgggttaca tgatcgatgg cgtgcaggcc 360
gaatacgtcc gcatcccgca tgccgacaac agcctctaca agatcccca gacaattgac 420
gacgaaatcg ccgtcctgct gagcgacatc ctgccaccg gccacgaaat cggcgtccag 480
tatgggaatg tccagccggg cgatgcggtg gctattgtcg gcgcgggccc cgtcggcatg 540
tccgtactgt tgaccgcca gttctactcc ccctcgacca tcatcgtgat cgacatggac 600
gagaatcgcc tccagctcgc caaggagctc ggggcaacgc acaccatcaa ctccggcacg 660

20110907_CL5178USNA_SeqList.txt

gagaacgttg tcgaagccgt gcataggatt gcggcagagg gagtcgatgt tgcgatcgag	720
gcggtgggca taccggcgac ttgggacatc tgccaggaga tcgtcaagcc cggcgcgcac	780
atcgccaacg tcggcgtgca tggcgtcaag gttgacttcg agattcagaa gctctggatc	840
aagaacctga cgatcaccac gggactggtg aacacgaaca cgacgcccat gctgatgaag	900
gtcgcctcga cgcacaagct tccgttgaag aagatgatta cccatcgctt cgagctggcc	960
gagatcgagc acgcctatca ggtattcctc aatggcgcca aggagaaggc gatgaagatc	1020
atcctctcga acgcaggcgc tgcctga	1047

<210> 36
 <211> 348
 <212> PRT
 <213> Achromobacter xylosoxidans
 <400> 36

Met	Lys	Ala	Leu	Val	Tyr	His	Gly	Asp	His	Lys	Ile	Ser	Leu	Glu	Asp
1				5					10					15	

Lys	Pro	Lys	Pro	Thr	Leu	Gln	Lys	Pro	Thr	Asp	Val	Val	Val	Arg	Val
			20					25					30		

Leu	Lys	Thr	Thr	Ile	Cys	Gly	Thr	Asp	Leu	Gly	Ile	Tyr	Lys	Gly	Lys
		35					40					45			

Asn	Pro	Glu	Val	Ala	Asp	Gly	Arg	Ile	Leu	Gly	His	Glu	Gly	Val	Gly
	50					55					60				

Val	Ile	Glu	Glu	Val	Gly	Glu	Ser	Val	Thr	Gln	Phe	Lys	Lys	Gly	Asp
65					70					75					80

Lys	Val	Leu	Ile	Ser	Cys	Val	Thr	Ser	Cys	Gly	Ser	Cys	Asp	Tyr	Cys
				85					90					95	

Lys	Lys	Gln	Leu	Tyr	Ser	His	Cys	Arg	Asp	Gly	Gly	Trp	Ile	Leu	Gly
			100					105					110		

Tyr	Met	Ile	Asp	Gly	Val	Gln	Ala	Glu	Tyr	Val	Arg	Ile	Pro	His	Ala
		115					120					125			

Asp	Asn	Ser	Leu	Tyr	Lys	Ile	Pro	Gln	Thr	Ile	Asp	Asp	Glu	Ile	Ala
	130					135					140				

Val	Leu	Leu	Ser	Asp	Ile	Leu	Pro	Thr	Gly	His	Glu	Ile	Gly	Val	Gln
145					150					155					160

Tyr	Gly	Asn	Val	Gln	Pro	Gly	Asp	Ala	Val	Ala	Ile	Val	Gly	Ala	Gly
				165					170					175	

Pro	Val	Gly	Met	Ser	Val	Leu	Leu	Thr	Ala	Gln	Phe	Tyr	Ser	Pro	Ser

Page 53

180

185

190

Thr Ile Ile Val Ile Asp Met Asp Glu Asn Arg Leu Gln Leu Ala Lys
 195 200 205

Glu Leu Gly Ala Thr His Thr Ile Asn Ser Gly Thr Glu Asn Val Val
 210 215 220

Glu Ala Val His Arg Ile Ala Ala Glu Gly Val Asp Val Ala Ile Glu
 225 230 235 240

Ala Val Gly Ile Pro Ala Thr Trp Asp Ile Cys Gln Glu Ile Val Lys
 245 250 255

Pro Gly Ala His Ile Ala Asn Val Gly Val His Gly Val Lys Val Asp
 260 265 270

Phe Glu Ile Gln Lys Leu Trp Ile Lys Asn Leu Thr Ile Thr Thr Gly
 275 280 285

Leu Val Asn Thr Asn Thr Thr Pro Met Leu Met Lys Val Ala Ser Thr
 290 295 300

Asp Lys Leu Pro Leu Lys Lys Met Ile Thr His Arg Phe Glu Leu Ala
 305 310 315 320

Glu Ile Glu His Ala Tyr Gln Val Phe Leu Asn Gly Ala Lys Glu Lys
 325 330 335

Ala Met Lys Ile Ile Leu Ser Asn Ala Gly Ala Ala
 340 345

<210> 37

<211> 1188

<212> DNA

<213> *Saccharomyces cerevisiae*

<400> 37

atgttgagaa ctcaagccgc cagattgatc tgcaactccc gtgtcatcac tgctaagaga 60

acctttgctt tggccaccgc tgctgctgct tacagcagac cagctgcccg tttcgttaag 120

ccaatgatca ctaccgctgg tttgaagcaa atcaacttcg gtggtactgt tgaaaccgtc 180

tacgaaagag ctgactggcc aagagaaaag ttgttgact acttcaagaa cgacactttt 240

gctttgatcg gttacggttc ccaagggttac ggtcaagggt tgaacttgag agacaacggt 300

ttgaacgtta tcattggtgt ccgtaaagat ggtgcttctt ggaaggctgc catcgaagac 360

ggttgggttc caggcaagaa cttgttcact gttgaagatg ctatcaagag aggtagttac 420

gttatgaact tgttgtccga tgccgctcaa tcagaaacct ggcctgctat caagccattg 480

ttgaccaagg gtaagacttt gtacttctcc cacggtttct cccagtcctt caaggacttg 540

actcacgttg aaccaccaa ggacttagat gttatcttgg ttgctccaaa gggttccggt 600

20110907_CL5178USNA_SeqList.txt

```

agaactgtca gatctttgtt caaggaaggt cgtggtatta actcttctta cgccgtctgg      660
aacgatgtca ccggttaaggc tcacgaaaag gcccaagctt tggccgttgc cattggttcc      720
ggttacgttt accaaaccac tttcgaaaga gaagtcaact ctgacttgta cggtgaaaga      780
ggttgtttaa tgggtggtat ccacggtatg ttcttggtc aatacgacgt cttgagagaa      840
aacggtcact ccccatctga agctttcaac gaaaccgtcg aagaagctac ccaatctcta      900
taccattga tcggttaagta cggtatggat tacatgtacg atgcttggtc caccaccgcc      960
agaagaggtg ctttgactg gtaccaatc ttcaagaatg ctttgaagcc tgttttccaa     1020
gacttgtagc aatctacaa gaacggtacc gaaaccaaga gatctttgga attcaactct     1080
caacctgact acagagaaaa gctagaaaag gaattagaca ccatcagaaa catggaaatc     1140
tggaagggtg gtaaggaagt cagaaagttg agaccagaaa accaataa                     1188

```

```

<210> 38
<211> 395
<212> PRT
<213> Saccharomyces cerevisiae

```

<400> 38

```

Met Leu Arg Thr Gln Ala Ala Arg Leu Ile Cys Asn Ser Arg Val Ile
1           5           10          15

```

```

Thr Ala Lys Arg Thr Phe Ala Leu Ala Thr Arg Ala Ala Ala Tyr Ser
          20          25          30

```

```

Arg Pro Ala Ala Arg Phe Val Lys Pro Met Ile Thr Thr Arg Gly Leu
          35          40          45

```

```

Lys Gln Ile Asn Phe Gly Gly Thr Val Glu Thr Val Tyr Glu Arg Ala
          50          55          60

```

```

Asp Trp Pro Arg Glu Lys Leu Leu Asp Tyr Phe Lys Asn Asp Thr Phe
65          70          75          80

```

```

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
          100          105          110

```

```

Ser Trp Lys Ala Ala Ile Glu Asp Gly Trp Val Pro Gly Lys Asn Leu
          115          120          125

```

```

Phe Thr Val Glu Asp Ala Ile Lys Arg Gly Ser Tyr Val Met Asn Leu
          130          135          140

```

```

Leu Ser Asp Ala Ala Gln Ser Glu Thr Trp Pro Ala Ile Lys Pro Leu
145          150          155          160

```

20110907_CL5178USNA_SeqList.txt

Leu Thr Lys Gly Lys Thr Leu Tyr Phe Ser His Gly Phe Ser Pro Val
165 170 175

Phe Lys Asp Leu Thr His Val Glu Pro Pro Lys Asp Leu Asp Val Ile
180 185 190

Leu Val Ala Pro Lys Gly Ser Gly Arg Thr Val Arg Ser Leu Phe Lys
195 200 205

Glu Gly Arg Gly Ile Asn Ser Ser Tyr Ala Val Trp Asn Asp Val Thr
210 215 220

Gly Lys Ala His Glu Lys Ala Gln Ala Leu Ala Val Ala Ile Gly Ser
225 230 235 240

Gly Tyr Val Tyr Gln Thr Thr Phe Glu Arg Glu Val Asn Ser Asp Leu
245 250 255

Tyr Gly Glu Arg Gly Cys Leu Met Gly Gly Ile His Gly Met Phe Leu
260 265 270

Ala Gln Tyr Asp Val Leu Arg Glu Asn Gly His Ser Pro Ser Glu Ala
275 280 285

Phe Asn Glu Thr Val Glu Glu Ala Thr Gln Ser Leu Tyr Pro Leu Ile
290 295 300

Gly Lys Tyr Gly Met Asp Tyr Met Tyr Asp Ala Cys Ser Thr Thr Ala
305 310 315 320

Arg Arg Gly Ala Leu Asp Trp Tyr Pro Ile Phe Lys Asn Ala Leu Lys
325 330 335

Pro Val Phe Gln Asp Leu Tyr Glu Ser Thr Lys Asn Gly Thr Glu Thr
340 345 350

Lys Arg Ser Leu Glu Phe Asn Ser Gln Pro Asp Tyr Arg Glu Lys Leu
355 360 365

Glu Lys Glu Leu Asp Thr Ile Arg Asn Met Glu Ile Trp Lys Val Gly
370 375 380

Lys Glu Val Arg Lys Leu Arg Pro Glu Asn Gln
385 390 395

<210> 39
<211> 1485
<212> DNA
<213> Vibrio cholerae

<400> 39
atggcgcaatt atttcaatac gctgaatctg cgtgaacagt tggatcaact tggtcgttgc

20110907_CL5178USNA_SeqList.txt

cgttttatgg cgcgagaaga gtttgcaacc gaagctgatt acctaaaagg taagaaagtg 120
 gtgatcgtag gttgtggggc tcaaggccta aaccaaggcc tcaatatgcg tgattcaggt 180
 ttggatgttt cttacgctct gcgtcaggct gcgattgatg aacagcgtca gtcatttaag 240
 aatgccaaga ataatggctt caacgtgggt agttatgaac aactcatccc aaccgcagat 300
 ttggtgatta acttgacgcc agacaagcag cacaccagtg tggtaaatgc ggtgatgcct 360
 ctgatgaagc aaggtgctgc cttgggttac tcacacgggt ttaatatcgt tgaagagggc 420
 atgcagatcc gtaaagacat cacggttgtg atggtggcac caaaatgtcc gggtagcgaa 480
 gttcgtgaag agtataagcg cggtttcggc gttcctactc ttatcgcggt acaccctgaa 540
 aacgatccac aaggtgaagg ttgggaaatt gctaaagcgt gggctgcggc aacgggtggc 600
 catcgtgcgg gctgtttagc ttcttctttt gtggcggaag tgaaatccga tttgatgggt 660
 gagcaaacca ttctctgcgg tatgctgcaa gcgggctcta tcgtttgtta cgagaaaatg 720
 gttgctgatg gcatcgaccc tggttatgcg ggcaagcttt tgcaatttg tttgggaaacc 780
 attacgaag cactcaagtt tggcgggtatt actcatatga tggatcgctt gtctaaccct 840
 gcaaaaatca aagcgtttga gctgtctgaa gagttgaaag atctgatgcg cccactgtac 900
 aacaagcata tggatgacat ctttctggc cacttctcta gcaccatgat ggcggattgg 960
 gcgaatgatg ataaagactt attcggctgg cgtgcagaaa ccgctgagac gacctttgaa 1020
 aactatccaa caaccgacgt aaaaattgct gagcaagaat actttgataa cggtatatttg 1080
 atgattgcca tgggtgcgtgc tgggggttag tttggcgtttg aagcgatgac ggcttcaggc 1140
 atcatcgatg agtcggctta ctatgaatca ctgcacgaac tcccactgat tgccaatacg 1200
 gtacgcgta agcgtctgta tgaaatgaac gtggtaatct ctgacactgc tgagtacggt 1260
 aactatctgt ttgccaatgt ggcgggtacca ctattgcgtg aaaagtttat gccgaaagtg 1320
 ggcactgatg tgattggtaa aggattaggc gtggtctcta atcaagttga taacgcaacg 1380
 cttatcgaag taaacagcat catccgtaac catccggttg agtatatcgg tgaagagcta 1440
 cgcggttaca tgaaagacat gaagcgcac gccgtgggtg attaa 1485

<210> 40
 <211> 494
 <212> PRT
 <213> *Vibrio cholerae*
 <400> 40

Met Ala Asn Tyr Phe Asn Thr Leu Asn Leu Arg Glu Gln Leu Asp Gln
 1 5 10 15
 Leu Gly Arg Cys Arg Phe Met Ala Arg Glu Glu Phe Ala Thr Glu Ala
 20 25 30
 Asp Tyr Leu Lys Gly Lys Lys Val Val Ile Val Gly Cys Gly Ala Gln
 35 40 45

20110907_CL5178USNA_SeqList.txt

Gly Leu Asn Gln Gly Leu Asn Met Arg Asp Ser Gly Leu Asp Val Ser
 50 55 60
 Tyr Ala Leu Arg Gln Ala Ala Ile Asp Glu Gln Arg Gln Ser Phe Lys
 65 70 75 80
 Asn Ala Lys Asn Asn Gly Phe Asn Val Gly Ser Tyr Glu Gln Leu Ile
 85 90 95
 Pro Thr Ala Asp Leu Val Ile Asn Leu Thr Pro Asp Lys Gln His Thr
 100 105 110
 Ser Val Val Asn Ala Val Met Pro Leu Met Lys Gln Gly Ala Ala Leu
 115 120 125
 Gly Tyr Ser His Gly Phe Asn Ile Val Glu Glu Gly Met Gln Ile Arg
 130 135 140
 Lys Asp Ile Thr Val Val Met Val Ala Pro Lys Cys Pro Gly Thr Glu
 145 150 155 160
 Val Arg Glu Glu Tyr Lys Arg Gly Phe Gly Val Pro Thr Leu Ile Ala
 165 170 175
 Val His Pro Glu Asn Asp Pro Gln Gly Glu Gly Trp Glu Ile Ala Lys
 180 185 190
 Ala Trp Ala Ala Ala Thr Gly Gly His Arg Ala Gly Cys Leu Ala Ser
 195 200 205
 Ser Phe Val Ala Glu Val Lys Ser Asp Leu Met Gly Glu Gln Thr Ile
 210 215 220
 Leu Cys Gly Met Leu Gln Ala Gly Ser Ile Val Cys Tyr Glu Lys Met
 225 230 235 240
 Val Ala Asp Gly Ile Asp Pro Gly Tyr Ala Gly Lys Leu Leu Gln Phe
 245 250 255
 Gly Trp Glu Thr Ile Thr Glu Ala Leu Lys Phe Gly Gly Ile Thr His
 260 265 270
 Met Met Asp Arg Leu Ser Asn Pro Ala Lys Ile Lys Ala Phe Glu Leu
 275 280 285
 Ser Glu Glu Leu Lys Asp Leu Met Arg Pro Leu Tyr Asn Lys His Met
 290 295 300
 Asp Asp Ile Ile Ser Gly His Phe Ser Ser Thr Met Met Ala Asp Trp
 305 310 315 320

20110907_CL5178USNA_SeqList.txt

Ala Asn Asp Asp Lys Asp Leu Phe Gly Trp Arg Ala Glu Thr Ala Glu
325 330 335

Thr Thr Phe Glu Asn Tyr Pro Thr Thr Asp Val Lys Ile Ala Glu Gln
340 345 350

Glu Tyr Phe Asp Asn Gly Ile Leu Met Ile Ala Met Val Arg Ala Gly
355 360 365

Val Glu Leu Ala Phe Glu Ala Met Thr Ala Ser Gly Ile Ile Asp Glu
370 375 380

Ser Ala Tyr Tyr Glu Ser Leu His Glu Leu Pro Leu Ile Ala Asn Thr
385 390 395 400

Val Ala Arg Lys Arg Leu Tyr Glu Met Asn Val Val Ile Ser Asp Thr
405 410 415

Ala Glu Tyr Gly Asn Tyr Leu Phe Ala Asn Val Ala Val Pro Leu Leu
420 425 430

Arg Glu Lys Phe Met Pro Lys Val Gly Thr Asp Val Ile Gly Lys Gly
435 440 445

Leu Gly Val Val Ser Asn Gln Val Asp Asn Ala Thr Leu Ile Glu Val
450 455 460

Asn Ser Ile Ile Arg Asn His Pro Val Glu Tyr Ile Gly Glu Glu Leu
465 470 475 480

Arg Gly Tyr Met Lys Asp Met Lys Arg Ile Ala Val Gly Asp
485 490

<210> 41
<211> 1014
<212> DNA
<213> Pseudomonas

<400> 41
atgcgcgttt tctacgataa agactgtgac ctctcgatca tccagggcaa gaaagttgcc 60
atcatcggct acggctccca gggccacgcc catgcctgca acctgaagga ctccggcgctc 120
gacgtcaccg tgggcctgcg tagcggctcc gccaccgtgg ccaaggccga agcgcacggt 180
ctgaaggttg ccgacgtgaa gaccgccgtc gccgcagccg acgtggtcat gatcctcacc 240
ccggacgagt tccagggccg cctgtacaag gaagagatcg agccgaacct gaagaagggc 300
gccaccctgg ccttcgctca cggcttctcc atccactaca accaggctcgt cccgcgcgcc 360
gacctcgacg tgatcatgat cgcgccgaag gcaccgggtc acaccgtgcg ttccgagttc 420
gtcaagggcg gtggcatccc tgacctgac gccatctacc aggacgcttc cggcaacgcc 480
aagaacgtcg ccctgtccta cgcctgcggc gtcggcgcg gtcgtaccgg tatcatcgaa 540

20110907_CL5178USNA_SeqList.txt

accaccttca aggacgagac cgaaaccgac ctgttcggtg agcaggccgt tctctgcggt 600
 gggtgcgtcg agctgggtcaa ggccggtttc gaaaccctgg tcgaagccgg ttacgcgccg 660
 gaaatggcct acttcgagtg cctgcacgag ctgaagctga tcgtcgacct gatgtacgaa 720
 ggcggcacgc ccaacatgaa ctactccatc tccaacaatg ccgaatacgg tgagtacgta 780
 accggtccgg aggtgatcaa cgccgagtcg cgtgctgccg tgcgcaacgc cctgaagcgc 840
 atccaggacg gcgagtacgc gaaaatgttc attaccgaag gtgcggccaa ctaccgctcg 900
 atgactgcct accgccgcaa caacgccgct caccgcatcg agcagatcgg cgagaagctg 960
 cgcgcgatga tgccgtggat cgcagccaac aagatcgtcg acaagagcaa gaac 1014

<210> 42
 <211> 338
 <212> PRT
 <213> Pseudomonas

<400> 42

Met Arg 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 Tyr Gly Ser Gln Gly His Ala His Ala
 20 25 30

Cys Asn Leu Lys Asp Ser Gly Val Asp Val Thr Val Gly Leu Arg Ser
 35 40 45

Gly Ser Ala Thr Val Ala Lys Ala Glu Ala His Gly Leu Lys Val Ala
 50 55 60

Asp Val Lys Thr Ala Val Ala Ala Ala Asp Val Val Met Ile Leu Thr
 65 70 75 80

Pro Asp Glu Phe Gln Gly Arg Leu Tyr Lys Glu Glu Ile Glu Pro Asn
 85 90 95

Leu Lys Lys Gly Ala Thr Leu Ala Phe Ala His Gly Phe Ser 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 Ala Ser Gly Asn Ala
 145 150 155 160

Lys Asn Val Ala Leu Ser Tyr Ala Cys Gly Val Gly Gly Gly Arg Thr
 165 170 175

20110907_CL5178USNA_SeqList.txt

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 Cys 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

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 Ala
260 265 270

Ala Met Arg Asn Ala Leu Lys Arg Ile Gln Asp Gly Glu Tyr Ala Lys
275 280 285

Met Phe Ile Thr Glu Gly Ala Ala Asn Tyr Pro Ser Met Thr Ala Tyr
290 295 300

Arg Arg Asn Asn Ala Ala His Pro Ile Glu Gln Ile Gly Glu Lys Leu
305 310 315 320

Arg Ala Met Met Pro Trp Ile Ala Ala Asn Lys Ile Val Asp Lys Ser
325 330 335

Lys Asn

<210> 43
<211> 1014
<212> DNA
<213> Pseudomonas fluorescens

<400> 43
atgaaagttt tctacgataa agactgcgac ctgtcgatca tccaaggtaa gaaagttgcc 60
atcatcggct acggttccca gggccacgct caagcatgca acctgaagga ttccggcgta 120
gacgtgactg ttggcctgcg taaaggctcg gctaccgttg ccaaggctga agcccacggc 180
ttgaaagtga cgcacgttgc tgcagccgtt gccggtgccg acttggtcat gatcctgacc 240
ccggacgagt tccagtccca gctgtacaag aacgaaatcg agccgaacat caagaagggc 300
gccactctgg ctttctccca cggcttcgcg atccactaca accaggttgt gcctcgtgcc 360
gacctcgacg tgatcatgat cgcgccgaag gctccaggcc acaccgtacg ttccgagttc 420
gtcaagggcg gtggtattcc tgacctgacg gcgatctacc aggacgcttc cggcaacgcc 480
aagaacgttg ccctgtccta cgccgcaggc gtgggcggcg gccgtaccgg catcatcgaa 540

20110907_CL5178USNA_SeqList.txt

```

accaccttca aggacgagac tgaaaccgac ctgttcggtg agcaggctgt tctgtgtggc      600
ggtagcgtcg agctgggtcaa agccgggtttc gaaaccctgg ttgaagctgg ctacgctcca      660
gaaatggcct acttcgagtg cctgcacgaa ctgaagctga tcgttgacct catgtacgaa      720
ggcggtatcg ccaacatgaa ctactcgatc tccaacaacg ctgaatacgg cgagtacgtg      780
actgggtccag aagtcaccaa cgccgaatcc cgtcaggcca tgcgcaatgc tctgaagcgc      840
atccaggacg gcgaatacgc gaagatgttc atcagcgaag gcgctaccgg ctacccatcg      900
atgaccgcca agcgtcgtaa caacgtgct caggtatcg aaatcatcgg cgagcaactg      960
cgctcgatga tgccttggat cgggtgccaac aaaatcgtcg acaaagccaa gaac      1014

```

```

<210> 44
<211> 338
<212> PRT
<213> Pseudomonas

```

<400> 44

```

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 Tyr Gly Ser Gln Gly His Ala Gln Ala
          20          25          30

```

```

Cys Asn Leu Lys Asp Ser Gly Val Asp Val Thr Val Gly Leu Arg Lys
          35          40          45

```

```

Gly Ser Ala Thr Val Ala Lys Ala Glu Ala His Gly Leu Lys Val Thr
          50          55          60

```

```

Asp Val Ala Ala Ala Val Ala Gly Ala Asp Leu Val Met Ile Leu Thr
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 Ala 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

```

20110907_CL5178USNA_SeqList.txt

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

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

Ala Met Arg Asn Ala Leu Lys Arg Ile Gln Asp Gly Glu Tyr Ala Lys
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 Leu
305 310 315 320

Arg Ser Met Met Pro Trp Ile Gly Ala Asn Lys Ile Val Asp Lys Ala
325 330 335

Lys Asn

<210> 45
<211> 1758
<212> DNA
<213> Saccharomyces cerevisiae

<400> 45
atgggcttgt taacgaaagt tgctacatct agacaattct ctacaacgag atgcgttgca 60
aagaagctca acaagtactc gtatatcatc actgaaccta agggccaagg tgcgtcccag 120
gccatgcttt atgccaccgg tttcaagaag gaagatttca agaagcctca agtcgggggtt 180
ggttcctgtt ggtggtccgg taacctatgt aacatgcatc tattggactt gaataacaga 240
tgttctcaat ccattgaaaa agcgggtttg aaagctatgc agttcaacac catcggtgtt 300
tcagacggta tctctatggg tactaaaggt atgagatact cgttacaaag tagagaaatc 360
attgcagact ctttgaac catcatgatg gcacaacact acgatgctaa catcgccatc 420
ccatcatgtg acaaaaacat gcccggtgtc atgatggcca tgggtagaca taacagacct 480
tccatcatgg tatatggtgg tactatcttg cccggtcatc caacatgtgg ttcttcgaag 540

20110907_CL5178USNA_SeqList.txt

```

atctctaaaa acatcgatat cgtctctgcg ttccaatcct acggtgaata tatttccaag      600
caattcactg aagaagaaag agaagatggt gtggaacatg catgcccagg tcctgggttct      660
tgtgggtgga tgtatactgc caacacaatg gcttctgccg ctgaagtgct aggtttgacc      720
attccaaact cctcttcctt cccagccggt tccaaggaga agttagctga gtgtgacaac      780
attggtgaat acatcaagaa gacaatggaa ttgggtatatt tacctcgtga tatcctcaca      840
aaagaggctt ttgaaaacgc cattacttat gtcgttgcaa ccggtgggtc cactaatgct      900
gttttgcat tgggtggctgt tgctcactct gcgggtgtca agttgtcacc agatgatttc      960
caaagaatca gtgatactac accattgatc ggtgacttca aaccttctgg taaatacgtc     1020
atggccgatt tgattaacgt tgggtgtacc caatctgtga ttaagtatct atatgaaaac     1080
aacatgttgc acggtaacac aatgactggt accggtgaca ctttggcaga acgtgcaaag     1140
aaagcaccaa gcctacctga aggacaagag attattaagc cactctccca cccaatcaag     1200
gccaacggtc acttgcaaat tctgtacggt tcattggcac caggtggagc tgtgggtaaa     1260
attaccggta aggaaggtag ttacttcaag ggtagagcac gtgtgttcga agaggaagggt     1320
gcctttattg aagccttgga aagagggtgaa atcaagaagg gtgaaaaaac cgttgttggt     1380
atcagatatg aagggtccaag aggtgcacca ggtatgcctg aaatgctaaa gccttcctct     1440
gctctgatgg gttacgggtt gggtaaagat gttgcattgt tgactgatgg tagattctct     1500
gggtggttctc acgggttctt aatcggccac attgttcccg aagccgctga aggtggctct     1560
atcgggttgg tcagagacgg cgatgagatt atcattgatg ctgataataa caagattgac     1620
ctattagtct ctgataagga aatggctcaa cgtaaacaaa gttgggttgc acctccacct     1680
cgttacacaa gaggtactct atccaagtat gctaagttgg tttccaacgc ttccaacggt     1740
tgtgttttag atgcttga                                     1758

```

<210> 46
 <211> 585
 <212> PRT
 <213> *Saccharomyces cerevisiae*

<400> 46

Met Gly Leu Leu Thr Lys Val Ala Thr Ser Arg Gln Phe Ser Thr Thr
 1 5 10 15

Arg Cys Val Ala Lys Lys Leu Asn Lys Tyr Ser Tyr Ile Ile Thr Glu
 20 25 30

Pro Lys Gly Gln Gly Ala Ser Gln Ala Met Leu Tyr Ala Thr Gly Phe
 35 40 45

Lys Lys Glu Asp Phe Lys Lys Pro Gln Val Gly Val Gly Ser Cys Trp
 50 55 60

Trp Ser Gly Asn Pro Cys Asn Met His Leu Leu Asp Leu Asn Asn Arg
 Page 64

65					70					75					80
Cys	Ser	Gln	Ser	Ile	Glu	Lys	Ala	Gly	Leu	Lys	Ala	Met	Gln	Phe	Asn
				85					90					95	
Thr	Ile	Gly	Val	Ser	Asp	Gly	Ile	Ser	Met	Gly	Thr	Lys	Gly	Met	Arg
			100					105					110		
Tyr	Ser	Leu	Gln	Ser	Arg	Glu	Ile	Ile	Ala	Asp	Ser	Phe	Glu	Thr	Ile
		115					120					125			
Met	Met	Ala	Gln	His	Tyr	Asp	Ala	Asn	Ile	Ala	Ile	Pro	Ser	Cys	Asp
	130					135					140				
Lys	Asn	Met	Pro	Gly	Val	Met	Met	Ala	Met	Gly	Arg	His	Asn	Arg	Pro
145					150					155					160
Ser	Ile	Met	Val	Tyr	Gly	Gly	Thr	Ile	Leu	Pro	Gly	His	Pro	Thr	Cys
				165					170					175	
Gly	Ser	Ser	Lys	Ile	Ser	Lys	Asn	Ile	Asp	Ile	Val	Ser	Ala	Phe	Gln
			180					185					190		
Ser	Tyr	Gly	Glu	Tyr	Ile	Ser	Lys	Gln	Phe	Thr	Glu	Glu	Glu	Arg	Glu
		195					200					205			
Asp	Val	Val	Glu	His	Ala	Cys	Pro	Gly	Pro	Gly	Ser	Cys	Gly	Gly	Met
	210					215					220				
Tyr	Thr	Ala	Asn	Thr	Met	Ala	Ser	Ala	Ala	Glu	Val	Leu	Gly	Leu	Thr
225					230					235					240
Ile	Pro	Asn	Ser	Ser	Ser	Phe	Pro	Ala	Val	Ser	Lys	Glu	Lys	Leu	Ala
				245					250					255	
Glu	Cys	Asp	Asn	Ile	Gly	Glu	Tyr	Ile	Lys	Lys	Thr	Met	Glu	Leu	Gly
			260					265					270		
Ile	Leu	Pro	Arg	Asp	Ile	Leu	Thr	Lys	Glu	Ala	Phe	Glu	Asn	Ala	Ile
		275					280					285			
Thr	Tyr	Val	Val	Ala	Thr	Gly	Gly	Ser	Thr	Asn	Ala	Val	Leu	His	Leu
	290					295					300				
Val	Ala	Val	Ala	His	Ser	Ala	Gly	Val	Lys	Leu	Ser	Pro	Asp	Asp	Phe
305					310					315					320
Gln	Arg	Ile	Ser	Asp	Thr	Thr	Pro	Leu	Ile	Gly	Asp	Phe	Lys	Pro	Ser
				325					330					335	
Gly	Lys	Tyr	Val	Met	Ala	Asp	Leu	Ile	Asn	Val	Gly	Gly	Thr	Gln	Ser

340

345

350

Val Ile Lys Tyr Leu Tyr Glu Asn Asn Met Leu His Gly Asn Thr Met
 355 360 365

Thr Val Thr Gly Asp Thr Leu Ala Glu Arg Ala Lys Lys Ala Pro Ser
 370 375 380

Leu Pro Glu Gly Gln Glu Ile Ile Lys Pro Leu Ser His Pro Ile Lys
 385 390 395 400

Ala Asn Gly His Leu Gln Ile Leu Tyr Gly Ser Leu Ala Pro Gly Gly
 405 410 415

Ala Val Gly Lys Ile Thr Gly Lys Glu Gly Thr Tyr Phe Lys Gly Arg
 420 425 430

Ala Arg Val Phe Glu Glu Glu Gly Ala Phe Ile Glu Ala Leu Glu Arg
 435 440 445

Gly Glu Ile Lys Lys Gly Glu Lys Thr Val Val Val Ile Arg Tyr Glu
 450 455 460

Gly Pro Arg Gly Ala Pro Gly Met Pro Glu Met Leu Lys Pro Ser Ser
 465 470 475 480

Ala Leu Met Gly Tyr Gly Leu Gly Lys Asp Val Ala Leu Leu Thr Asp
 485 490 495

Gly Arg Phe Ser Gly Gly Ser His Gly Phe Leu Ile Gly His Ile Val
 500 505 510

Pro Glu Ala Ala Glu Gly Gly Pro Ile Gly Leu Val Arg Asp Gly Asp
 515 520 525

Glu Ile Ile Ile Asp Ala Asp Asn Asn Lys Ile Asp Leu Leu Val Ser
 530 535 540

Asp Lys Glu Met Ala Gln Arg Lys Gln Ser Trp Val Ala Pro Pro Pro
 545 550 555 560

Arg Tyr Thr Arg Gly Thr Leu Ser Lys Tyr Ala Lys Leu Val Ser Asn
 565 570 575

Ala Ser Asn Gly Cys Val Leu Asp Ala
 580 585

<210> 47
 <211> 1662
 <212> DNA
 <213> Lactococcus lactis

20110907_CL5178USNA_SeqList.txt

```

<400> 47
tctagacata tgtatactgt gggggattac ctgctggatc gcctgcacga actggggatt      60
gaagaaattht tccgtgtgcc aggcgattat aacctgcagt tcctggacca gattatctcg      120
caciaagata tgaagtgggt cggtaacgcc aacgaactga acgcgagcta tatggcagat      180
ggttatgccc gtacaaaaaa agctgctgctg tttctgacga cttttggcgt tggcgaactg      240
agcgcctgca acggactggc aggaagctac gccgagaacc tgccagttgt cgaaattggt      300
gggtcgccta cttctaaggt tcagaatgaa ggcaaatttg tgcaccatac tctggctgat      360
ggggatttta aacattttat gaaaatgcat gaaccggtta ctgcggcccg cacgctgctg      420
acagcagaga atgctacggt tgagatcgac cgcgtcctgt ctgcgctgct gaaagagcgc      480
aagccggtat atatcaatct gcctgtcgat gttgccgcag cgaaagccga aaagccgctg      540
ctgccactga aaaaagaaaa cagcacctcc aatacatcgg accaggaaat tctgaataaa      600
atccaggaat cactgaagaa tgcgaagaaa ccgatcgtca tcaccggaca tgagatcatc      660
tcttttggcc tggaaaaaac ggtcacgcag ttcattttcta agaccaaact gcctatcacc      720
accctgaact tcggcaaata tagcgtcgat gaagcgtgc cgagttttct gggatatctat      780
aatggtagcc tgtccgaacc gaacctgaaa gaattcgtcg aaagcgcgga ctttatcctg      840
atgctgggctg tgaactgac ggatagctcc acaggcgcag ttaccacca tctgaacgag      900
aataaaatga tttccctgaa tatcgacgaa ggcaaaatct ttaacgagcg catccagaac      960
ttcgattttg aatctctgat tagttcgtg ctggatctgt ccgaaattga gtataaagggt     1020
aaatatattg ataaaaaaca ggaggatttt gtgccgtcta atgcgctgct gagtcaggat     1080
cgtctgtggc aagccgtaga aaacctgaca cagtctaata aaacgattgt tgcggaacag     1140
ggaacttcat ttttcggcgc ctcattcatt tttctgaaat ccaaaagcca tttcattggc     1200
caaccgctgt gggggagtat tggttatacc tttccggcgg cgctgggttc acagattgca     1260
gataaggaat cagccatct gctgtttatt ggtgacggca gcctgcagct gactgtccag     1320
gaactggggc tggcgatccg tgaaaaaatc aatccgattt gctttatcat caataacgac     1380
ggctacaccg tcgaacgcga aattcatgga ccgaatcaaa gttacaatga catcccgatg     1440
tggaactata gcaaactgcc ggaatccttt ggcgcgacag aggatcgctg ggtgagtaaa     1500
attgtgcgta cggaaaacga atttgtgtcg gttatgaaag aagcgcaggc tgaccggaat     1560
cgcatgtatt ggattgaact gatcctggca aaagaaggcg caccgaaagt tctgaaaaag     1620
atggggaaac tgtttgcgga gcaaaataaa agctaaggat cc                        1662

```

```

<210> 48
<211> 548
<212> PRT
<213> Lactococcus lactis

```

```
<400> 48
```

```

Met Tyr Thr Val Gly Asp Tyr Leu Leu Asp Arg Leu His Glu Leu Gly
1          5          10          15

```

20110907_CL5178USNA_SeqList.txt

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
 50 55 60
 Ala Ala Ala Phe Leu Thr Thr Phe Gly Val Gly Glu Leu Ser Ala Val
 65 70 75 80
 Asn Gly Leu Ala Gly Ser Tyr Ala Glu Asn Leu Pro Val Val Glu Ile
 85 90 95
 Val Gly Ser Pro Thr Ser Lys Val Gln Asn Glu Gly Lys Phe Val His
 100 105 110
 His Thr Leu Ala Asp Gly Asp Phe Lys His Phe Met Lys Met His Glu
 115 120 125
 Pro Val Thr Ala Ala Arg Thr Leu Leu Thr Ala Glu Asn Ala Thr Val
 130 135 140
 Glu Ile Asp Arg Val Leu Ser Ala Leu Leu Lys Glu Arg Lys Pro Val
 145 150 155 160
 Tyr Ile Asn Leu Pro Val Asp Val Ala Ala Lys Ala Glu Lys Pro
 165 170 175
 Ser Leu Pro Leu Lys Lys Glu Asn Ser Thr Ser Asn Thr Ser Asp Gln
 180 185 190
 Glu Ile Leu Asn Lys Ile Gln Glu Ser Leu Lys Asn Ala Lys Lys Pro
 195 200 205
 Ile Val Ile Thr Gly His Glu Ile Ile Ser Phe Gly Leu Glu Lys Thr
 210 215 220
 Val Thr Gln Phe Ile Ser Lys Thr Lys Leu Pro Ile Thr Thr Leu Asn
 225 230 235 240
 Phe Gly Lys Ser Ser Val Asp Glu Ala Leu Pro Ser Phe Leu Gly Ile
 245 250 255
 Tyr Asn Gly Thr Leu Ser Glu Pro Asn Leu Lys Glu Phe Val Glu Ser
 260 265 270
 Ala Asp Phe Ile Leu Met Leu Gly Val Lys Leu Thr Asp Ser Ser Thr
 275 280 285

20110907_CL5178USNA_SeqList.txt

Gly Ala Phe Thr His His Leu Asn Glu Asn Lys Met Ile Ser Leu Asn
290 295 300

Ile Asp Glu Gly Lys Ile Phe Asn Glu Arg Ile Gln Asn Phe Asp Phe
305 310 315 320

Glu Ser Leu Ile Ser Ser Leu Leu Asp Leu Ser Glu Ile Glu Tyr Lys
325 330 335

Gly Lys Tyr Ile Asp Lys Lys Gln Glu Asp Phe Val Pro Ser Asn Ala
340 345 350

Leu Leu Ser Gln Asp Arg Leu Trp Gln Ala Val Glu Asn Leu Thr Gln
355 360 365

Ser Asn Glu Thr Ile Val Ala Glu Gln Gly Thr Ser Phe Phe Gly Ala
370 375 380

Ser Ser Ile Phe Leu Lys Ser Lys Ser His Phe Ile Gly Gln Pro Leu
385 390 395 400

Trp Gly Ser Ile Gly Tyr Thr Phe Pro Ala Ala Leu Gly Ser Gln Ile
405 410 415

Ala Asp Lys Glu Ser Arg His Leu Leu Phe Ile Gly Asp Gly Ser Leu
420 425 430

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

Pro Ile Cys Phe Ile Ile Asn Asn Asp Gly Tyr Thr Val Glu Arg Glu
450 455 460

Ile His Gly Pro Asn Gln Ser Tyr Asn Asp Ile Pro Met Trp Asn Tyr
465 470 475 480

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

20110907_CL5178USNA_SeqList.txt

<210> 49
 <211> 1647
 <212> DNA
 <213> *Lactococcus lactis*

<400> 49
 atgtatacag taggagatta cctattagac cgattacacg agttaggaat tgaagaaatt 60
 tttggagtcc ctggagacta taacttacaa tttttagatc aaattatttc ccacaaggat 120
 atgaaatggg tcggaaatgc taatgaatta aatgcttcat atatggctga tggctatgct 180
 cgtactaaaa aagctgccgc atttcttaca acctttggag taggtgaatt gagtgcagtt 240
 aatggattag caggaagtta cgccgaaaat ttaccagtag tagaaatagt gggatcacct 300
 acatcaaaaag ttcaaaatga aggaaaattht gttcatcata cgctggctga cggtgatttt 360
 aaacacttta tgaaaatgca cgaacctgtt acagcagctc gaactttact gacagcagaa 420
 aatgcaaccg ttgaaattga ccgagtactt tctgcactat taaaagaaag aaaacctgtc 480
 tatatcaact taccagttga tgttgctgct gcaaaagcag agaaaccctc actccctttg 540
 aaaaaggaaa actcaacttc aaatacaagt gaccaagaaa ttttgaacaa aattcaagaa 600
 agcttgaaaa atgccaaaa accaatcgtg attacaggac atgaaataat tagttttggc 660
 ttagaaaaaa cagtcactca atttatttca aagacaaaac tacctattac gacattaaac 720
 tttggtaaaa gttcagttga tgaagccctc ccttcatttt taggaatcta taatggtaca 780
 ctctcagagc ctaatcttaa agaattcgtg gaatcagccg acttcatctt gatgcttgga 840
 gttaaactca cagactcttc aacaggagcc ttcactcatc atttaaataa aaataaaatg 900
 atttactga atatagatga aggaaaaata tttaacgaaa gaatccaaaa ttttgatttt 960
 gaatccctca tctcctctct cttagaccta agcgaaatag aatacaaagg aaaatatatc 1020
 gataaaaagc aagaagactt tgttccatca aatgcgcttt tatcacaaga ccgcctatgg 1080
 caagcagttg aaaacctaac tcaaagcaat gaaacaatcg ttgctgaaca aggacatca 1140
 ttctttggcg cttcatcaat tttcttaaaa tcaaagagtc attttatttg tcaaccctta 1200
 tggggatcaa ttggatatac attcccagca gcattaggaa gccaaattgc agataaagaa 1260
 agcagacacc ttttatttat tgggtgatgg tcaattcaac ttacagtga agaattagga 1320
 ttagcaatca gagaaaaaat taatccaatt tgctttatta tcaataatga tggttatata 1380
 gtcgaaagag aaattcatgg accaaatcaa agctacaatg atattccaat gtggaattac 1440
 tcaaaattac cagaatcgtt tggagcaaca gaagatcgag tagtctcaaa aatcgttaga 1500
 actgaaaatg aatttggtgc tgtcatgaaa gaagctcaag cagatccaaa tagaatgtac 1560
 tggattgagt taattttggc aaaagaaggt gcaccaaag tactgaaaaa aatgggcaaa 1620
 ctatttgctg aacaaaataa atcataa 1647

<210> 50
 <211> 1692
 <212> DNA
 <213> *Saccharomyces cerevisiae*

<220>

<221> CDS

<222> (1)..(1692)

<400> 50

atg	tct	gaa	att	act	ttg	ggt	aaa	tat	ttg	ttc	gaa	aga	tta	aag	caa		48
Met	Ser	Glu	Ile	Thr	Leu	Gly	Lys	Tyr	Leu	Phe	Glu	Arg	Leu	Lys	Gln		
1				5					10					15			

gtc	aac	ggt	aac	acc	ggt	ttc	ggt	ttg	cca	ggt	gac	ttc	aac	ttg	tcc		96
Val	Asn	Val	Asn	Thr	Val	Phe	Gly	Leu	Pro	Gly	Asp	Phe	Asn	Leu	Ser		
			20					25					30				

ttg	ttg	gac	aag	atc	tac	gaa	ggt	gaa	ggt	atg	aga	tgg	gct	ggt	aac		144
Leu	Leu	Asp	Lys	Ile	Tyr	Glu	Val	Glu	Gly	Met	Arg	Trp	Ala	Gly	Asn		
		35					40					45					

gcc	aac	gaa	ttg	aac	gct	gct	tac	gcc	gct	gat	ggt	tac	gct	cgt	atc		192
Ala	Asn	Glu	Leu	Asn	Ala	Ala	Tyr	Ala	Ala	Asp	Gly	Tyr	Ala	Arg	Ile		
	50					55					60						

aag	ggt	atg	tct	tgt	atc	atc	acc	acc	ttc	ggt	gtc	ggt	gaa	ttg	tct		240
Lys	Gly	Met	Ser	Cys	Ile	Ile	Thr	Thr	Phe	Gly	Val	Gly	Glu	Leu	Ser		
65				70						75				80			

gct	ttg	aac	ggt	att	gcc	ggt	tct	tac	gct	gaa	cac	gtc	ggt	ggt	ttg		288
Ala	Leu	Asn	Gly	Ile	Ala	Gly	Ser	Tyr	Ala	Glu	His	Val	Gly	Val	Leu		
				85					90					95			

cac	ggt	ggt	ggt	gtc	cca	tcc	atc	tct	gct	caa	gct	aag	caa	ttg	ttg		336
His	Val	Val	Gly	Val	Pro	Ser	Ile	Ser	Ala	Gln	Ala	Lys	Gln	Leu	Leu		
			100					105					110				

ttg	cac	cac	acc	ttg	ggt	aac	ggt	gac	ttc	act	ggt	ttc	cac	aga	atg		384
Leu	His	His	Thr	Leu	Gly	Asn	Gly	Asp	Phe	Thr	Val	Phe	His	Arg	Met		
		115					120					125					

tct	gcc	aac	att	tct	gaa	acc	act	gct	atg	atc	act	gac	att	gct	acc		432
Ser	Ala	Asn	Ile	Ser	Glu	Thr	Thr	Ala	Met	Ile	Thr	Asp	Ile	Ala	Thr		
	130					135					140						

gcc	cca	gct	gaa	att	gac	aga	tgt	atc	aga	acc	act	tac	gtc	acc	caa		480
Ala	Pro	Ala	Glu	Ile	Asp	Arg	Cys	Ile	Arg	Thr	Thr	Tyr	Val	Thr	Gln		
145					150					155					160		

aga	cca	gtc	tac	tta	ggt	ttg	cca	gct	aac	ttg	gtc	gac	ttg	aac	gtc		528
Arg	Pro	Val	Tyr	Leu	Gly	Leu	Pro	Ala	Asn	Leu	Val	Asp	Leu	Asn	Val		
				165					170					175			

cca	gct	aag	ttg	ttg	caa	act	cca	att	gac	atg	tct	ttg	aag	cca	aac		576
Pro	Ala	Lys	Leu	Leu	Gln	Thr	Pro	Ile	Asp	Met	Ser	Leu	Lys	Pro	Asn		
			180					185					190				

gat	gct	gaa	tcc	gaa	aag	gaa	gtc	att	gac	acc	atc	ttg	gct	ttg	gtc		624
Asp	Ala	Glu	Ser	Glu	Lys	Glu	Val	Ile	Asp	Thr	Ile	Leu	Ala	Leu	Val		
		195				200						205					

aag	gat	gct	aag	aac	cca	ggt	atc	ttg	gct	gat	gct	tgt	tgt	tcc	aga		672
Lys	Asp	Ala	Lys	Asn	Pro	Val	Ile	Leu	Ala	Asp	Ala	Cys	Cys	Ser	Arg		
	210					215					220						

cac	gac	gtc	aag	gct	gaa	act	aag	aag	ttg	att	gac	ttg	act	caa	ttc		720
His	Asp	Val	Lys	Ala	Glu	Thr	Lys	Lys	Leu	Ile	Asp	Leu	Thr	Gln	Phe		
225				230						235					240		

cca	gct	ttc	gtc	acc	cca	atg	ggt	aag	ggt	tcc	att	gac	gaa	caa	cac		768

20110907_CL5178USNA_SeqList.txt

Pro	Ala	Phe	Val	Thr 245	Pro	Met	Gly	Lys	Gly 250	Ser	Ile	Asp	Glu	Gln 255	His	
cca	aga	tac	ggt	ggt	ggt	tac	gtc	ggt	acc	ttg	tcc	aag	cca	gaa	ggt	816
Pro	Arg	Tyr	Gly 260	Gly	Val	Tyr	Val	Gly 265	Thr	Leu	Ser	Lys	Pro 270	Glu	Val	
aag	gaa	gcc	ggt	gaa	tct	gct	gac	ttg	att	ttg	tct	gtc	ggt	gct	ttg	864
Lys	Glu	Ala 275	Val	Glu	Ser	Ala	Asp 280	Leu	Ile	Leu	Ser	Val 285	Gly	Ala	Leu	
ttg	tct	gat	ttc	aac	acc	ggt	tct	ttc	tct	tac	tct	tac	aag	acc	aag	912
Leu	Ser 290	Asp	Phe	Asn	Thr	Gly 295	Ser	Phe	Ser	Tyr	Ser 300	Tyr	Lys	Thr	Lys	
aac	att	gtc	gaa	ttc	cac	tcc	gac	cac	atg	aag	atc	aga	aac	gcc	act	960
Asn 305	Ile	Val	Glu	Phe	His 310	Ser	Asp	His	Met	Lys 315	Ile	Arg	Asn	Ala	Thr 320	
ttc	cca	ggt	gtc	caa	atg	aaa	ttc	ggt	ttg	caa	aag	ttg	ttg	acc	act	1008
Phe	Pro	Gly	Val	Gln 325	Met	Lys	Phe	Val	Leu 330	Gln	Lys	Leu	Leu	Thr 335	Thr	
att	gct	gac	gcc	gct	aag	ggt	tac	aag	cca	ggt	gct	gtc	cca	gct	aga	1056
Ile	Ala	Asp	Ala 340	Ala	Lys	Gly	Tyr	Lys 345	Pro	Val	Ala	Val	Pro 350	Ala	Arg	
act	cca	gct	aac	gct	gct	gtc	cca	gct	tct	acc	cca	ttg	aag	caa	gaa	1104
Thr	Pro	Ala 355	Asn	Ala	Ala	Val	Pro 360	Ala	Ser	Thr	Pro	Leu 365	Lys	Gln	Glu	
tgg	atg	tgg	aac	caa	ttg	ggt	aac	ttc	ttg	caa	gaa	ggt	gat	ggt	gtc	1152
Trp	Met 370	Trp	Asn	Gln	Leu	Gly 375	Asn	Phe	Leu	Gln	Glu 380	Gly	Asp	Val	Val	
att	gct	gaa	acc	ggt	acc	tcc	gct	ttc	ggt	atc	aac	caa	acc	act	ttc	1200
Ile 385	Ala	Glu	Thr	Gly	Thr 390	Ser	Ala	Phe	Gly	Ile 395	Asn	Gln	Thr	Thr	Phe 400	
cca	aac	aac	acc	tac	ggt	atc	tct	caa	gtc	tta	tgg	ggt	tcc	att	ggt	1248
Pro	Asn	Asn	Thr	Tyr 405	Gly	Ile	Ser	Gln	Val 410	Leu	Trp	Gly	Ser	Ile 415	Gly	
ttc	acc	act	ggt	gct	acc	ttg	ggt	gct	gct	ttc	gct	gct	gaa	gaa	att	1296
Phe	Thr	Thr	Gly 420	Ala	Thr	Leu	Gly	Ala 425	Ala	Phe	Ala	Ala	Glu 430	Glu	Ile	
gat	cca	aag	aag	aga	ggt	atc	tta	ttc	att	ggt	gac	ggt	tct	ttg	caa	1344
Asp	Pro	Lys 435	Lys	Arg	Val	Ile	Leu 440	Phe	Ile	Gly	Asp	Gly 445	Ser	Leu	Gln	
ttg	act	ggt	caa	gaa	atc	tcc	acc	atg	atc	aga	tgg	ggc	ttg	aag	cca	1392
Leu	Thr 450	Val	Gln	Glu	Ile	Ser 455	Thr	Met	Ile	Arg	Trp 460	Gly	Leu	Lys	Pro	
tac	ttg	ttc	gtc	ttg	aac	aac	gat	ggt	tac	acc	att	gaa	aag	ttg	att	1440
Tyr 465	Leu	Phe	Val	Leu	Asn 470	Asn	Asp	Gly	Tyr	Thr 475	Ile	Glu	Lys	Leu	Ile 480	
cac	ggt	cca	aag	gct	caa	tac	aac	gaa	att	caa	ggt	tgg	gac	cac	cta	1488
His	Gly	Pro	Lys	Ala 485	Gln	Tyr	Asn	Glu	Ile 490	Gln	Gly	Trp	Asp	His 495	Leu	
tcc	ttg	ttg	cca	act	ttc	ggt	gct	aag	gac	tat	gaa	acc	cac	aga	gtc	1536
Ser	Leu	Leu	Pro 500	Thr	Phe	Gly	Ala	Lys 505	Asp	Tyr	Glu	Thr	His 510	Arg	Val	
gct	acc	acc	ggt	gaa	tgg	gac	aag	ttg	acc	caa	gac	aag	tct	ttc	aac	1584

20110907_CL5178USNA_SeqList.txt

Ala	Thr	Thr	Gly	Glu	Trp	Asp	Lys	Leu	Thr	Gln	Asp	Lys	Ser	Phe	Asn		
		515					520					525					
gac	aac	tct	aag	atc	aga	atg	att	gaa	atc	atg	ttg	cca	gtc	ttc	gat		1632
Asp	Asn	Ser	Lys	Ile	Arg	Met	Ile	Glu	Ile	Met	Leu	Pro	Val	Phe	Asp		
		530				535					540						
gct	cca	caa	aac	ttg	gtt	gaa	caa	gct	aag	ttg	act	gct	gct	acc	aac		1680
Ala	Pro	Gln	Asn	Leu	Val	Glu	Gln	Ala	Lys	Leu	Thr	Ala	Ala	Thr	Asn		
					550					555					560		
gct	aag	caa	taa														1692
Ala	Lys	Gln															

<210> 51
 <211> 563
 <212> PRT
 <213> Saccharomyces cerevisiae
 <400> 51

Met	Ser	Glu	Ile	Thr	Leu	Gly	Lys	Tyr	Leu	Phe	Glu	Arg	Leu	Lys	Gln		
1				5					10					15			
Val	Asn	Val	Asn	Thr	Val	Phe	Gly	Leu	Pro	Gly	Asp	Phe	Asn	Leu	Ser		
			20					25					30				
Leu	Leu	Asp	Lys	Ile	Tyr	Glu	Val	Glu	Gly	Met	Arg	Trp	Ala	Gly	Asn		
		35					40					45					
Ala	Asn	Glu	Leu	Asn	Ala	Ala	Tyr	Ala	Ala	Asp	Gly	Tyr	Ala	Arg	Ile		
	50					55					60						
Lys	Gly	Met	Ser	Cys	Ile	Ile	Thr	Thr	Phe	Gly	Val	Gly	Glu	Leu	Ser		
65					70					75					80		
Ala	Leu	Asn	Gly	Ile	Ala	Gly	Ser	Tyr	Ala	Glu	His	Val	Gly	Val	Leu		
				85					90					95			
His	Val	Val	Gly	Val	Pro	Ser	Ile	Ser	Ala	Gln	Ala	Lys	Gln	Leu	Leu		
			100					105					110				
Leu	His	His	Thr	Leu	Gly	Asn	Gly	Asp	Phe	Thr	Val	Phe	His	Arg	Met		
		115					120					125					
Ser	Ala	Asn	Ile	Ser	Glu	Thr	Thr	Ala	Met	Ile	Thr	Asp	Ile	Ala	Thr		
	130					135					140						
Ala	Pro	Ala	Glu	Ile	Asp	Arg	Cys	Ile	Arg	Thr	Thr	Tyr	Val	Thr	Gln		
145					150					155					160		
Arg	Pro	Val	Tyr	Leu	Gly	Leu	Pro	Ala	Asn	Leu	Val	Asp	Leu	Asn	Val		
				165					170					175			
Pro	Ala	Lys	Leu	Leu	Gln	Thr	Pro	Ile	Asp	Met	Ser	Leu	Lys	Pro	Asn		

180

185

190

Asp Ala Glu Ser Glu Lys Glu Val Ile Asp Thr Ile Leu Ala Leu Val
 195 200 205

Lys Asp Ala Lys Asn Pro Val Ile Leu Ala Asp Ala Cys Cys Ser Arg
 210 215 220

His Asp Val Lys Ala Glu Thr Lys Lys Leu Ile Asp Leu Thr Gln Phe
 225 230 235 240

Pro Ala Phe Val Thr Pro Met Gly Lys Gly Ser Ile Asp Glu Gln His
 245 250 255

Pro Arg Tyr Gly Gly Val Tyr Val Gly Thr Leu Ser Lys Pro Glu Val
 260 265 270

Lys Glu Ala Val Glu Ser Ala Asp Leu Ile Leu Ser Val Gly Ala Leu
 275 280 285

Leu Ser Asp Phe Asn Thr Gly Ser Phe Ser Tyr Ser Tyr Lys Thr Lys
 290 295 300

Asn Ile Val Glu Phe His Ser Asp His Met Lys Ile Arg Asn Ala Thr
 305 310 315 320

Phe Pro Gly Val Gln Met Lys Phe Val Leu Gln Lys Leu Leu Thr Thr
 325 330 335

Ile Ala Asp Ala Ala Lys Gly Tyr Lys Pro Val Ala Val Pro Ala Arg
 340 345 350

Thr Pro Ala Asn Ala Ala Val Pro Ala Ser Thr Pro Leu Lys Gln Glu
 355 360 365

Trp Met Trp Asn Gln Leu Gly Asn Phe Leu Gln Glu Gly Asp Val Val
 370 375 380

Ile Ala Glu Thr Gly Thr Ser Ala Phe Gly Ile Asn Gln Thr Thr Phe
 385 390 395 400

Pro Asn Asn Thr Tyr Gly Ile Ser Gln Val Leu Trp Gly Ser Ile Gly
 405 410 415

Phe Thr Thr Gly Ala Thr Leu Gly Ala Ala Phe Ala Ala Glu Glu Ile
 420 425 430

Asp Pro Lys Lys Arg Val Ile Leu Phe Ile Gly Asp Gly Ser Leu Gln
 435 440 445

Leu Thr Val Gln Glu Ile Ser Thr Met Ile Arg Trp Gly Leu Lys Pro

450

455

460

Tyr Leu Phe Val Leu Asn Asn Asp Gly Tyr Thr Ile Glu Lys Leu Ile
 465 470 475 480

His Gly Pro Lys Ala Gln Tyr Asn Glu Ile Gln Gly Trp Asp His Leu
 485 490 495

Ser Leu Leu Pro Thr Phe Gly Ala Lys Asp Tyr Glu Thr His Arg Val
 500 505 510

Ala Thr Thr Gly Glu Trp Asp Lys Leu Thr Gln Asp Lys Ser Phe Asn
 515 520 525

Asp Asn Ser Lys Ile Arg Met Ile Glu Ile Met Leu Pro Val Phe Asp
 530 535 540

Ala Pro Gln Asn Leu Val Glu Gln Ala Lys Leu Thr Ala Ala Thr Asn
 545 550 555 560

Ala Lys Gln

<210> 52
 <211> 1692
 <212> DNA
 <213> Saccharomyces cerevisiae

<220>
 <221> CDS
 <222> (1)..(1692)

<400> 52
 atg tct gaa ata acc tta ggt aaa tat tta ttt gaa aga ttg agc caa 48
 Met Ser Glu Ile Thr Leu Gly Lys Tyr Leu Phe Glu Arg Leu Ser Gln
 1 5 10 15
 gtc aac tgt aac acc gtc ttc ggt ttg cca ggt gac ttt aac ttg tct 96
 Val Asn Cys Asn Thr Val Phe Gly Leu Pro Gly Asp Phe Asn Leu Ser
 20 25 30
 ctt ttg gat aag ctt tat gaa gtc aaa ggt atg aga tgg gct ggt aac 144
 Leu Leu Asp Lys Leu Tyr Glu Val Lys Gly Met Arg Trp Ala Gly Asn
 35 40 45
 gct aac gaa ttg aac gct gcc tat gct gct gat ggt tac gct cgt atc 192
 Ala Asn Glu Leu Asn Ala Ala Tyr Ala Ala Asp Gly Tyr Ala Arg Ile
 50 55 60
 aag ggt atg tcc tgt att att acc acc ttc ggt gtt ggt gaa ttg tct 240
 Lys Gly Met Ser Cys Ile Ile Thr Thr Phe Gly Val Gly Glu Leu Ser
 65 70 75 80
 gct ttg aat ggt att gcc ggt tct tac gct gaa cat gtc ggt gtt ttg 288
 Ala Leu Asn Gly Ile Ala Gly Ser Tyr Ala Glu His Val Gly Val Leu
 85 90 95
 cac gtt gtt ggt gtt cca tcc atc tct tct caa gct aag caa ttg ttg 336
 His Val Val Gly Val Pro Ser Ile Ser Ser Gln Ala Lys Gln Leu Leu

20110907_CL5178USNA_SeqList.txt

100				105				110								
ttg	cat	cat	acc	ttg	ggt	aac	ggt	gac	ttc	act	ggt	ttc	cac	aga	atg	384
Leu	His	His	Thr	Leu	Gly	Asn	Gly	Asp	Phe	Thr	Val	Phe	His	Arg	Met	
		115					120					125				
tct	gcc	aac	att	tct	gaa	acc	act	gcc	atg	atc	act	gat	att	gct	aac	432
Ser	Ala	Asn	Ile	Ser	Glu	Thr	Thr	Ala	Met	Ile	Thr	Asp	Ile	Ala	Asn	
	130					135					140					
gct	cca	gct	gaa	att	gac	aga	tgt	atc	aga	acc	acc	tac	act	acc	caa	480
Ala	Pro	Ala	Glu	Ile	Asp	Arg	Cys	Ile	Arg	Thr	Thr	Tyr	Thr	Thr	Gln	
145					150					155					160	
aga	cca	gtc	tac	ttg	ggt	ttg	cca	gct	aac	ttg	ggt	gac	ttg	aac	gtc	528
Arg	Pro	Val	Tyr	Leu	Gly	Leu	Pro	Ala	Asn	Leu	Val	Asp	Leu	Asn	Val	
				165					170					175		
cca	gcc	aag	tta	ttg	gaa	act	cca	att	gac	ttg	tct	ttg	aag	cca	aac	576
Pro	Ala	Lys	Leu	Leu	Glu	Thr	Pro	Ile	Asp	Leu	Ser	Leu	Lys	Pro	Asn	
			180					185					190			
gac	gct	gaa	gct	gaa	gct	gaa	ggt	ggt	aga	act	ggt	ggt	gaa	ttg	atc	624
Asp	Ala	Glu	Ala	Glu	Ala	Glu	Val	Val	Arg	Thr	Val	Val	Glu	Leu	Ile	
	195						200					205				
aag	gat	gct	aag	aac	cca	ggt	atc	ttg	gct	gat	gct	tgt	gct	tct	aga	672
Lys	Asp	Ala	Lys	Asn	Pro	Val	Ile	Leu	Ala	Asp	Ala	Cys	Ala	Ser	Arg	
	210					215					220					
cat	gat	gtc	aag	gct	gaa	act	aag	aag	ttg	atg	gac	ttg	act	caa	ttc	720
His	Asp	Val	Lys	Ala	Glu	Thr	Lys	Lys	Leu	Met	Asp	Leu	Thr	Gln	Phe	
225					230					235					240	
cca	ggt	tac	gtc	acc	cca	atg	ggt	aag	ggt	gct	att	gac	gaa	caa	cac	768
Pro	Val	Tyr	Val	Thr	Pro	Met	Gly	Lys	Gly	Ala	Ile	Asp	Glu	Gln	His	
				245					250					255		
cca	aga	tac	ggt	ggt	ggt	tac	ggt	acc	ttg	tct	aga	cca	gaa	ggt		816
Pro	Arg	Tyr	Gly	Gly	Val	Tyr	Val	Thr	Leu	Ser	Arg	Pro	Glu	Val		
			260				265					270				
aag	aag	gct	gta	gaa	tct	gct	gat	ttg	ata	ttg	tct	atc	ggt	gct	ttg	864
Lys	Lys	Ala	Val	Glu	Ser	Ala	Asp	Leu	Ile	Leu	Ser	Ile	Gly	Ala	Leu	
		275					280					285				
ttg	tct	gat	ttc	aat	acc	ggt	tct	ttc	tct	tac	tcc	tac	aag	acc	aaa	912
Leu	Ser	Asp	Phe	Asn	Thr	Gly	Ser	Phe	Ser	Tyr	Ser	Tyr	Lys	Thr	Lys	
	290					295					300					
aat	atc	ggt	gaa	ttc	cac	tct	gac	cac	atc	aag	atc	aga	aac	gcc	acc	960
Asn	Ile	Val	Glu	Phe	His	Ser	Asp	His	Ile	Lys	Ile	Arg	Asn	Ala	Thr	
305					310					315					320	
ttc	cca	ggt	ggt	caa	atg	aaa	ttt	gcc	ttg	caa	aaa	ttg	ttg	gat	gct	1008
Phe	Pro	Gly	Val	Gln	Met	Lys	Phe	Ala	Leu	Gln	Lys	Leu	Leu	Asp	Ala	
				325					330					335		
att	cca	gaa	gtc	gtc	aag	gac	tac	aaa	cct	ggt	gct	gtc	cca	gct	aga	1056
Ile	Pro	Glu	Val	Val	Lys	Asp	Tyr	Lys	Pro	Val	Ala	Val	Pro	Ala	Arg	
			340					345					350			
ggt	cca	att	acc	aag	tct	act	cca	gct	aac	act	cca	atg	aag	caa	gaa	1104
Val	Pro	Ile	Thr	Lys	Ser	Thr	Pro	Ala	Asn	Thr	Pro	Met	Lys	Gln	Glu	
		355					360					365				
tgg	atg	tgg	aac	cat	ttg	ggt	aac	ttc	ttg	aga	gaa	ggt	gat	att	ggt	1152
Trp	Met	Trp	Asn	His	Leu	Gly	Asn	Phe	Leu	Arg	Glu	Gly	Asp	Ile	Val	

370	375	380	
att gct gaa acc ggt act tcc gcc ttc ggt att aac caa act act ttc Ile Ala Glu Thr Gly Thr Ser Ala Phe Gly Ile Asn Gln Thr Thr Phe 385 390 395 400			1200
cca aca gat gta tac gct atc gtc caa gtc ttg tgg ggt tcc att ggt Pro Thr Asp Val Tyr Ala Ile Val Gln Val Leu Trp Gly Ser Ile Gly 405 410 415			1248
ttc aca gtc ggc gct cta ttg ggt gct act atg gcc gct gaa gaa ctt Phe Thr Val Gly Ala Leu Leu Gly Ala Thr Met Ala Ala Glu Glu Leu 420 425 430			1296
gat cca aag aag aga gtt att tta ttc att ggt gac ggt tct cta caa Asp Pro Lys Lys Arg Val Ile Leu Phe Ile Gly Asp Gly Ser Leu Gln 435 440 445			1344
ttg act gtt caa gaa atc tct acc atg att aga tgg ggt ttg aag cca Leu Thr Val Gln Glu Ile Ser Thr Met Ile Arg Trp Gly Leu Lys Pro 450 455 460			1392
tac att ttt gtc ttg aat aac aac ggt tac acc att gaa aaa ttg att Tyr Ile Phe Val Leu Asn Asn Asn Gly Tyr Thr Ile Glu Lys Leu Ile 465 470 475 480			1440
cac ggt cct cat gcc gaa tat aat gaa att caa ggt tgg gac cac ttg His Gly Pro His Ala Glu Tyr Asn Glu Ile Gln Gly Trp Asp His Leu 485 490 495			1488
gcc tta ttg cca act ttt ggt gct aga aac tac gaa acc cac aga gtt Ala Leu Leu Pro Thr Phe Gly Ala Arg Asn Tyr Glu Thr His Arg Val 500 505 510			1536
gct acc act ggt gaa tgg gaa aag ttg act caa gac aag gac ttc caa Ala Thr Thr Gly Glu Trp Glu Lys Leu Thr Gln Asp Lys Asp Phe Gln 515 520 525			1584
gac aac tct aag att aga atg att gaa gtt atg ttg cca gtc ttt gat Asp Asn Ser Lys Ile Arg Met Ile Glu Val Met Leu Pro Val Phe Asp 530 535 540			1632
gct cca caa aac ttg gtt aaa caa gct caa ttg act gcc gct act aac Ala Pro Gln Asn Leu Val Lys Gln Ala Gln Leu Thr Ala Ala Thr Asn 545 550 555 560			1680
gct aaa caa taa Ala Lys Gln			1692

<210> 53
 <211> 563
 <212> PRT
 <213> Saccharomyces cerevisiae

<400> 53

Met Ser Glu Ile Thr Leu Gly Lys Tyr Leu Phe Glu Arg Leu Ser Gln
1 5 10 15

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

Leu Leu Asp Lys Leu Tyr Glu Val Lys Gly Met Arg Trp Ala Gly Asn
35 40 45

20110907_CL5178USNA_SeqList.txt

Ala Asn Glu Leu Asn Ala Ala Tyr Ala Ala Asp Gly Tyr Ala Arg Ile
50 55 60

Lys Gly Met Ser Cys Ile Ile Thr Thr Phe Gly Val Gly Glu Leu Ser
65 70 75 80

Ala Leu Asn Gly Ile Ala Gly Ser Tyr Ala Glu His Val Gly Val Leu
85 90 95

His Val Val Gly Val Pro Ser Ile Ser Ser Gln Ala Lys Gln Leu Leu
100 105 110

Leu His His Thr Leu Gly Asn Gly Asp Phe Thr Val Phe His Arg Met
115 120 125

Ser Ala Asn Ile Ser Glu Thr Thr Ala Met Ile Thr Asp Ile Ala Asn
130 135 140

Ala Pro Ala Glu Ile Asp Arg Cys Ile Arg Thr Thr Tyr Thr Thr Gln
145 150 155 160

Arg Pro Val Tyr Leu Gly Leu Pro Ala Asn Leu Val Asp Leu Asn Val
165 170 175

Pro Ala Lys Leu Leu Glu Thr Pro Ile Asp Leu Ser Leu Lys Pro Asn
180 185 190

Asp Ala Glu Ala Glu Ala Glu Val Val Arg Thr Val Val Glu Leu Ile
195 200 205

Lys Asp Ala Lys Asn Pro Val Ile Leu Ala Asp Ala Cys Ala Ser Arg
210 215 220

His Asp Val Lys Ala Glu Thr Lys Lys Leu Met Asp Leu Thr Gln Phe
225 230 235 240

Pro Val Tyr Val Thr Pro Met Gly Lys Gly Ala Ile Asp Glu Gln His
245 250 255

Pro Arg Tyr Gly Gly Val Tyr Val Gly Thr Leu Ser Arg Pro Glu Val
260 265 270

Lys Lys Ala Val Glu Ser Ala Asp Leu Ile Leu Ser Ile Gly Ala Leu
275 280 285

Leu Ser Asp Phe Asn Thr Gly Ser Phe Ser Tyr Ser Tyr Lys Thr Lys
290 295 300

Asn Ile Val Glu Phe His Ser Asp His Ile Lys Ile Arg Asn Ala Thr
305 310 315 320

20110907_CL5178USNA_SeqList.txt

Phe Pro Gly Val Gln Met Lys Phe Ala Leu Gln Lys Leu Leu Asp Ala
325 330 335

Ile Pro Glu Val Val Lys Asp Tyr Lys Pro Val Ala Val Pro Ala Arg
340 345 350

Val Pro Ile Thr Lys Ser Thr Pro Ala Asn Thr Pro Met Lys Gln Glu
355 360 365

Trp Met Trp Asn His Leu Gly Asn Phe Leu Arg Glu Gly Asp Ile Val
370 375 380

Ile Ala Glu Thr Gly Thr Ser Ala Phe Gly Ile Asn Gln Thr Thr Phe
385 390 395 400

Pro Thr Asp Val Tyr Ala Ile Val Gln Val Leu Trp Gly Ser Ile Gly
405 410 415

Phe Thr Val Gly Ala Leu Leu Gly Ala Thr Met Ala Ala Glu Glu Leu
420 425 430

Asp Pro Lys Lys Arg Val Ile Leu Phe Ile Gly Asp Gly Ser Leu Gln
435 440 445

Leu Thr Val Gln Glu Ile Ser Thr Met Ile Arg Trp Gly Leu Lys Pro
450 455 460

Tyr Ile Phe Val Leu Asn Asn Asn Gly Tyr Thr Ile Glu Lys Leu Ile
465 470 475 480

His Gly Pro His Ala Glu Tyr Asn Glu Ile Gln Gly Trp Asp His Leu
485 490 495

Ala Leu Leu Pro Thr Phe Gly Ala Arg Asn Tyr Glu Thr His Arg Val
500 505 510

Ala Thr Thr Gly Glu Trp Glu Lys Leu Thr Gln Asp Lys Asp Phe Gln
515 520 525

Asp Asn Ser Lys Ile Arg Met Ile Glu Val Met Leu Pro Val Phe Asp
530 535 540

Ala Pro Gln Asn Leu Val Lys Gln Ala Gln Leu Thr Ala Ala Thr Asn
545 550 555 560

Ala Lys Gln

<210> 54
<211> 1599

<212> DNA

<213> *Saccharomyces cerevisiae*

<220>

<221> CDS

<222> (1)..(1599)

<400> 54

atg	tct	gaa	att	act	ctt	gga	aaa	tac	tta	ttt	gaa	aga	ttg	aag	caa	48
Met	Ser	Glu	Ile	Thr	Leu	Gly	Lys	Tyr	Leu	Phe	Glu	Arg	Leu	Lys	Gln	
1				5					10					15		

ggt	aat	ggt	aac	acc	att	ttt	ggg	cta	cca	ggc	gac	ttc	aac	ttg	tcc	96
Val	Asn	Val	Asn	Thr	Ile	Phe	Gly	Leu	Pro	Gly	Asp	Phe	Asn	Leu	Ser	
			20					25					30			

cta	ttg	gac	aag	att	tac	gag	gta	gat	gga	ttg	aga	tgg	gct	ggt	aat	144
Leu	Leu	Asp	Lys	Ile	Tyr	Glu	Val	Asp	Gly	Leu	Arg	Trp	Ala	Gly	Asn	
		35				40					45					

gca	aat	gag	ctg	aac	gcc	gcc	tat	gcc	gcc	gat	ggt	tac	gca	cgc	atc	192
Ala	Asn	Glu	Leu	Asn	Ala	Ala	Tyr	Ala	Ala	Asp	Gly	Tyr	Ala	Arg	Ile	
	50				55					60						

aag	ggt	tta	tct	gtg	ctg	gta	act	act	ttt	ggc	gta	ggt	gaa	tta	tcc	240
Lys	Gly	Leu	Ser	Val	Leu	Val	Thr	Thr	Phe	Gly	Val	Gly	Glu	Leu	Ser	
65				70					75						80	

gcc	ttg	aat	ggt	att	gca	gga	tcg	tat	gca	gaa	cac	gtc	ggt	gta	ctg	288
Ala	Leu	Asn	Gly	Ile	Ala	Gly	Ser	Tyr	Ala	Glu	His	Val	Gly	Val	Leu	
				85					90					95		

cat	ggt	ggt	ggt	gtc	ccc	tct	atc	tcc	gct	cag	gct	aag	caa	ttg	ttg	336
His	Val	Val	Gly	Val	Pro	Ser	Ile	Ser	Ala	Gln	Ala	Lys	Gln	Leu	Leu	
			100					105					110			

ttg	cat	cat	acc	ttg	ggt	aac	ggt	gat	ttt	acc	ggt	ttt	cac	aga	atg	384
Leu	His	His	Thr	Leu	Gly	Asn	Gly	Asp	Phe	Thr	Val	Phe	His	Arg	Met	
		115					120					125				

tcc	gcc	aat	atc	tca	gaa	act	aca	tca	atg	att	aca	gac	att	gct	aca	432
Ser	Ala	Asn	Ile	Ser	Glu	Thr	Thr	Ser	Met	Ile	Thr	Asp	Ile	Ala	Thr	
	130					135					140					

gcc	cct	tca	gaa	atc	gat	agg	ttg	atc	agg	aca	aca	ttt	ata	aca	caa	480
Ala	Pro	Ser	Glu	Ile	Asp	Arg	Leu	Ile	Arg	Thr	Thr	Phe	Ile	Thr	Gln	
145					150					155					160	

agg	cct	agc	tac	ttg	ggg	ttg	cca	gcg	aat	ttg	gta	gat	cta	aag	ggt	528
Arg	Pro	Ser	Tyr	Leu	Gly	Leu	Pro	Ala	Asn	Leu	Val	Asp	Leu	Lys	Val	
				165					170					175		

cct	ggt	tct	ctt	ttg	gaa	aaa	ccg	att	gat	cta	tca	tta	aaa	cct	aac	576
Pro	Gly	Ser	Leu	Leu	Glu	Lys	Pro	Ile	Asp	Leu	Ser	Leu	Lys	Pro	Asn	
			180					185					190			

gat	ccc	gaa	gct	gaa	aag	gaa	ggt	att	gat	acc	gta	cta	gaa	ttg	atc	624
Asp	Pro	Glu	Ala	Glu	Lys	Glu	Val	Ile	Asp	Thr	Val	Leu	Glu	Leu	Ile	
		195				200					205					

cag	aat	tcg	aaa	aac	cct	ggt	ata	cta	tcg	gat	gcc	tgt	gct	tct	agg	672
Gln	Asn	Ser	Lys	Asn	Pro	Val	Ile	Leu	Ser	Asp	Ala	Cys	Ala	Ser	Arg	
	210					215					220					

cac	aac	ggt	aaa	aaa	gaa	acc	cag	aag	tta	att	gat	ttg	acg	caa	ttc	720
His	Asn	Val	Lys	Lys	Glu	Thr	Gln	Lys	Leu	Ile	Asp	Leu	Thr	Gln	Phe	
225					230					235					240	

20110907_CL5178USNA_SeqList.txt

cca	gct	ttt	gtg	aca	cct	cta	ggg	aaa	ggg	tca	ata	gat	gaa	cag	cat	768
Pro	Ala	Phe	Val	Thr	Pro	Leu	Gly	Lys	Gly	Ser	Ile	Asp	Glu	Gln	His	
				245					250					255		
ccc	aga	tat	ggc	ggg	gtt	tat	gtg	gga	acg	ctg	tcc	aaa	caa	gac	gtg	816
Pro	Arg	Tyr	Gly	Gly	Val	Tyr	Val	Gly	Thr	Leu	Ser	Lys	Gln	Asp	Val	
			260					265					270			
aaa	cag	gcc	gtt	gag	tcg	gct	gat	ttg	atc	ctt	tcg	gtc	ggg	gct	ttg	864
Lys	Gln	Ala	Val	Glu	Ser	Ala	Asp	Leu	Ile	Leu	Ser	Val	Gly	Ala	Leu	
		275					280					285				
ctc	tct	gat	ttt	aac	aca	ggg	tcg	ttt	tcc	tac	tcc	tac	aag	act	aaa	912
Leu	Ser	Asp	Phe	Asn	Thr	Gly	Ser	Phe	Ser	Tyr	Ser	Tyr	Lys	Thr	Lys	
	290					295					300					
aat	gta	gtg	gag	ttt	cat	tcc	gat	tac	gta	aag	gtg	aag	aac	gct	acg	960
Asn	Val	Val	Glu	Phe	His	Ser	Asp	Tyr	Val	Lys	Val	Lys	Asn	Ala	Thr	
305					310					315					320	
ttc	ctc	ggg	gta	caa	atg	aaa	ttt	gca	cta	caa	aac	tta	ctg	aag	gtt	1008
Phe	Leu	Gly	Val	Gln	Met	Lys	Phe	Ala	Leu	Gln	Asn	Leu	Leu	Lys	Val	
				325					330					335		
att	ccc	gat	gtt	gtt	aag	ggc	tac	aag	agc	gtt	ccc	gta	cca	acc	aaa	1056
Ile	Pro	Asp	Val	Val	Lys	Gly	Tyr	Lys	Ser	Val	Pro	Val	Pro	Thr	Lys	
			340					345					350			
act	ccc	gca	aac	aaa	ggg	gta	cct	gct	agc	acg	ccc	ttg	aaa	caa	gag	1104
Thr	Pro	Ala	Asn	Lys	Gly	Val	Pro	Ala	Ser	Thr	Pro	Leu	Lys	Gln	Glu	
		355					360					365				
tgg	ttg	tgg	aac	gaa	ttg	tcc	aaa	ttc	ttg	caa	gaa	ggg	gat	gtt	atc	1152
Trp	Leu	Trp	Asn	Glu	Leu	Ser	Lys	Phe	Leu	Gln	Glu	Gly	Asp	Val	Ile	
	370					375					380					
att	tcc	gag	acc	ggc	acg	tct	gcc	ttc	ggg	atc	aat	caa	act	atc	ttt	1200
Ile	Ser	Glu	Thr	Gly	Thr	Ser	Ala	Phe	Gly	Ile	Asn	Gln	Thr	Ile	Phe	
385					390					395					400	
cct	aag	gac	gcc	tac	ggg	atc	tcg	cag	gtg	ttg	tgg	ggg	tcc	atc	ggg	1248
Pro	Lys	Asp	Ala	Tyr	Gly	Ile	Ser	Gln	Val	Leu	Trp	Gly	Ser	Ile	Gly	
				405					410					415		
ttt	aca	aca	gga	gca	act	tta	ggg	gct	gcc	ttt	gcc	gct	gag	gag	att	1296
Phe	Thr	Thr	Gly	Ala	Thr	Leu	Gly	Ala	Ala	Phe	Ala	Ala	Glu	Glu	Ile	
			420					425					430			
gac	ccc	aac	aag	aga	gtc	atc	tta	ttc	ata	ggg	gac	ggg	tct	ttg	cag	1344
Asp	Pro	Asn	Lys	Arg	Val	Ile	Leu	Phe	Ile	Gly	Asp	Gly	Ser	Leu	Gln	
		435					440					445				
tta	acc	gtc	caa	gaa	atc	tcc	acc	atg	atc	aga	tgg	ggg	tta	aag	ccg	1392
Leu	Thr	Val	Gln	Glu	Ile	Ser	Thr	Met	Ile	Arg	Trp	Gly	Leu	Lys	Pro	
	450					455					460					
tat	ctt	ttt	gtc	ctt	aac	aac	gac	ggc	tac	act	atc	gaa	aag	ctg	att	1440
Tyr	Leu	Phe	Val	Leu	Asn	Asn	Asp	Gly	Tyr	Thr	Ile	Glu	Lys	Leu	Ile	
465					470					475					480	
cat	ggg	cct	cac	gca	gag	tac	aac	gaa	atc	cag	acc	tgg	gat	cac	ctc	1488
His	Gly	Pro	His	Ala	Glu	Tyr	Asn	Glu	Ile	Gln	Thr	Trp	Asp	His	Leu	
				485					490					495		
gcc	ctg	ttg	ccc	gca	ttt	ggg	gcg	aaa	aag	tac	gaa	aat	cac	aag	atc	1536
Ala	Leu	Leu	Pro	Ala	Phe	Gly	Ala	Lys	Lys	Tyr	Glu	Asn	His	Lys	Ile	
			500					505					510			

20110907_CL5178USNA_SeqList.txt

gcc act acg ggt gag tgg gat gcc tta acc act gat tca gag ttc cag 1584
 Ala Thr Thr Gly Glu Trp Asp Ala Leu Thr Thr Asp Ser Glu Phe Gln
 515 520 525

aaa aac tcg gtg atc 1599
 Lys Asn Ser Val Ile
 530

<210> 55
 <211> 533
 <212> PRT
 <213> Saccharomyces cerevisiae

<400> 55

Met Ser Glu Ile Thr Leu Gly Lys Tyr Leu Phe Glu Arg Leu Lys Gln
 1 5 10 15

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

Leu Leu Asp Lys Ile Tyr Glu Val Asp Gly Leu Arg Trp Ala Gly Asn
 35 40 45

Ala Asn Glu Leu Asn Ala Ala Tyr Ala Ala Asp Gly Tyr Ala Arg Ile
 50 55 60

Lys Gly Leu Ser Val Leu Val Thr Thr Phe Gly Val Gly Glu Leu Ser
 65 70 75 80

Ala Leu Asn Gly Ile Ala Gly Ser Tyr Ala Glu His Val Gly Val Leu
 85 90 95

His Val Val Gly Val Pro Ser Ile Ser Ala Gln Ala Lys Gln Leu Leu
 100 105 110

Leu His His Thr Leu Gly Asn Gly Asp Phe Thr Val Phe His Arg Met
 115 120 125

Ser Ala Asn Ile Ser Glu Thr Thr Ser Met Ile Thr Asp Ile Ala Thr
 130 135 140

Ala Pro Ser Glu Ile Asp Arg Leu Ile Arg Thr Thr Phe Ile Thr Gln
 145 150 155 160

Arg Pro Ser Tyr Leu Gly Leu Pro Ala Asn Leu Val Asp Leu Lys Val
 165 170 175

Pro Gly Ser Leu Leu Glu Lys Pro Ile Asp Leu Ser Leu Lys Pro Asn
 180 185 190

Asp Pro Glu Ala Glu Lys Glu Val Ile Asp Thr Val Leu Glu Leu Ile
 195 200 205

20110907_CL5178USNA_SeqList.txt

Gln Asn Ser Lys Asn Pro Val Ile Leu Ser Asp Ala Cys Ala Ser Arg
210 215 220

His Asn Val Lys Lys Glu Thr Gln Lys Leu Ile Asp Leu Thr Gln Phe
225 230 235 240

Pro Ala Phe Val Thr Pro Leu Gly Lys Gly Ser Ile Asp Glu Gln His
245 250 255

Pro Arg Tyr Gly Gly Val Tyr Val Gly Thr Leu Ser Lys Gln Asp Val
260 265 270

Lys Gln Ala Val Glu Ser Ala Asp Leu Ile Leu Ser Val Gly Ala Leu
275 280 285

Leu Ser Asp Phe Asn Thr Gly Ser Phe Ser Tyr Ser Tyr Lys Thr Lys
290 295 300

Asn Val Val Glu Phe His Ser Asp Tyr Val Lys Val Lys Asn Ala Thr
305 310 315 320

Phe Leu Gly Val Gln Met Lys Phe Ala Leu Gln Asn Leu Leu Lys Val
325 330 335

Ile Pro Asp Val Val Lys Gly Tyr Lys Ser Val Pro Val Pro Thr Lys
340 345 350

Thr Pro Ala Asn Lys Gly Val Pro Ala Ser Thr Pro Leu Lys Gln Glu
355 360 365

Trp Leu Trp Asn Glu Leu Ser Lys Phe Leu Gln Glu Gly Asp Val Ile
370 375 380

Ile Ser Glu Thr Gly Thr Ser Ala Phe Gly Ile Asn Gln Thr Ile Phe
385 390 395 400

Pro Lys Asp Ala Tyr Gly Ile Ser Gln Val Leu Trp Gly Ser Ile Gly
405 410 415

Phe Thr Thr Gly Ala Thr Leu Gly Ala Ala Phe Ala Ala Glu Glu Ile
420 425 430

Asp Pro Asn Lys Arg Val Ile Leu Phe Ile Gly Asp Gly Ser Leu Gln
435 440 445

Leu Thr Val Gln Glu Ile Ser Thr Met Ile Arg Trp Gly Leu Lys Pro
450 455 460

Tyr Leu Phe Val Leu Asn Asn Asp Gly Tyr Thr Ile Glu Lys Leu Ile
465 470 475 480

20110907_CL5178USNA_SeqList.txt

His Gly Pro His Ala Glu Tyr Asn Glu Ile Gln Thr Trp Asp His Leu
485 490 495

Ala Leu Leu Pro Ala Phe Gly Ala Lys Lys Tyr Glu Asn His Lys Ile
500 505 510

Ala Thr Thr Gly Glu Trp Asp Ala Leu Thr Thr Asp Ser Glu Phe Gln
515 520 525

Lys Asn Ser Val Ile
530

<210> 56
<211> 1695
<212> DNA
<213> Candida glabrata

<220>
<221> CDS
<222> (1)..(1695)

<400> 56	
atg tct gag att act ttg ggt aga tac ttg ttc gag aga ttg aac caa	48
Met Ser Glu Ile Thr Leu Gly Arg Tyr Leu Phe Glu Arg Leu Asn Gln	
1 5 10 15	
gtc gac gtt aag acc atc ttc ggt ttg cca ggt gac ttc aac ttg tcc	96
Val Asp Val Lys Thr Ile Phe Gly Leu Pro Gly Asp Phe Asn Leu Ser	
20 25 30	
cta ttg gac aag atc tac gaa gtt gaa ggt atg aga tgg gct ggt aac	144
Leu Leu Asp Lys Ile Tyr Glu Val Glu Gly Met Arg Trp Ala Gly Asn	
35 40 45	
gct aac gaa ttg aac gct gct tac gct gct gac ggt tac gct aga atc	192
Ala Asn Glu Leu Asn Ala Ala Tyr Ala Ala Asp Gly Tyr Ala Arg Ile	
50 55 60	
aag ggt atg tcc tgt atc atc acc acc ttc ggt gtc ggt gaa ttg tct	240
Lys Gly Met Ser Cys Ile Ile Thr Thr Phe Gly Val Gly Glu Leu Ser	
65 70 75 80	
gcc ttg aac ggt att gcc ggt tct tac gct gaa cac gtc ggt gtc ttg	288
Ala Leu Asn Gly Ile Ala Gly Ser Tyr Ala Glu His Val Gly Val Leu	
85 90 95	
cac gtc gtc ggt gtc cca tcc atc tcc tct caa gct aag caa ttg ttg	336
His Val Val Gly Val Pro Ser Ile Ser Ser Gln Ala Lys Gln Leu Leu	
100 105 110	
ttg cac cac acc ttg ggt aac ggt gac ttc act gtc ttc cac aga atg	384
Leu His His Thr Leu Gly Asn Gly Asp Phe Thr Val Phe His Arg Met	
115 120 125	
tcc gct aac atc tct gag acc acc gct atg gtc act gac atc gct acc	432
Ser Ala Asn Ile Ser Glu Thr Thr Ala Met Val Thr Asp Ile Ala Thr	
130 135 140	
gct cca gct gag atc gac aga tgt atc aga acc acc tac atc acc caa	480
Ala Pro Ala Glu Ile Asp Arg Cys Ile Arg Thr Thr Tyr Ile Thr Gln	
145 150 155 160	

20110907_CL5178USNA_SeqList.txt

aga	cca	gtc	tac	ttg	ggt	cta	cca	gct	aac	ttg	gtc	gac	cta	aag	gtc	528
Arg	Pro	Val	Tyr	Leu	Gly	Leu	Pro	Ala	Asn	Leu	Val	Asp	Leu	Lys	Val	
				165					170					175		
cca	gcc	aag	ctt	ttg	gaa	acc	cca	att	gac	ttg	tcc	ttg	aag	cca	aac	576
Pro	Ala	Lys	Leu	Leu	Glu	Thr	Pro	Ile	Asp	Leu	Ser	Leu	Lys	Pro	Asn	
			180					185					190			
gac	cca	gaa	gcc	gaa	act	gaa	gtc	ggt	gac	acc	gtc	ttg	gaa	ttg	atc	624
Asp	Pro	Glu	Ala	Glu	Thr	Glu	Val	Val	Asp	Thr	Val	Leu	Glu	Leu	Ile	
		195					200					205				
aag	gct	gct	aag	aac	cca	ggt	atc	ttg	gct	gat	gct	tgt	gct	tcc	aga	672
Lys	Ala	Ala	Lys	Asn	Pro	Val	Ile	Leu	Ala	Asp	Ala	Cys	Ala	Ser	Arg	
	210					215				220						
cac	gac	gtc	aag	gct	gaa	acc	aag	aag	ttg	att	gac	gcc	act	caa	ttc	720
His	Asp	Val	Lys	Ala	Glu	Thr	Lys	Lys	Leu	Ile	Asp	Ala	Thr	Gln	Phe	
					230					235					240	
cca	tcc	ttc	gtt	acc	cca	atg	ggt	aag	ggt	tcc	atc	gac	gaa	caa	cac	768
Pro	Ser	Phe	Val	Thr	Pro	Met	Gly	Lys	Gly	Ser	Ile	Asp	Glu	Gln	His	
				245					250					255		
cca	aga	ttc	ggt	ggt	gtc	tac	gtc	ggt	acc	ttg	tcc	aga	cca	gaa	gtt	816
Pro	Arg	Phe	Gly	Gly	Val	Tyr	Val	Gly	Thr	Leu	Ser	Arg	Pro	Glu	Val	
			260					265					270			
aag	gaa	gct	gtt	gaa	tcc	gct	gac	ttg	atc	ttg	tct	gtc	ggt	gct	ttg	864
Lys	Glu	Ala	Val	Glu	Ser	Ala	Asp	Leu	Ile	Leu	Ser	Val	Gly	Ala	Leu	
		275					280					285				
ttg	tcc	gat	ttc	aac	act	ggt	tct	ttc	tct	tac	tct	tac	aag	acc	aag	912
Leu	Ser	Asp	Phe	Asn	Thr	Gly	Ser	Phe	Ser	Tyr	Ser	Tyr	Lys	Thr	Lys	
	290					295				300						
aac	atc	gtc	gaa	ttc	cac	tct	gac	tac	atc	aag	atc	aga	aac	gct	acc	960
Asn	Ile	Val	Glu	Phe	His	Ser	Asp	Tyr	Ile	Lys	Ile	Arg	Asn	Ala	Thr	
					310					315					320	
ttc	cca	ggt	gtc	caa	atg	aag	ttc	gct	ttg	caa	aag	ttg	ttg	aac	gcc	1008
Phe	Pro	Gly	Val	Gln	Met	Lys	Phe	Ala	Leu	Gln	Lys	Leu	Leu	Asn	Ala	
				325					330					335		
gtc	cca	gaa	gct	atc	aag	ggt	tac	aag	cca	gtc	cct	gtc	cca	gct	aga	1056
Val	Pro	Glu	Ala	Ile	Lys	Gly	Tyr	Lys	Pro	Val	Pro	Val	Pro	Ala	Arg	
			340					345					350			
gtc	cca	gaa	aac	aag	tcc	tgt	gac	cca	gct	acc	cca	ttg	aag	caa	gaa	1104
Val	Pro	Glu	Asn	Lys	Ser	Cys	Asp	Pro	Ala	Thr	Pro	Leu	Lys	Gln	Glu	
		355					360					365				
tgg	atg	tgg	aac	caa	gtt	tcc	aag	ttc	ttg	caa	gaa	ggt	gat	gtt	gtt	1152
Trp	Met	Trp	Asn	Gln	Val	Ser	Lys	Phe	Leu	Gln	Glu	Gly	Asp	Val	Val	
	370					375					380					
atc	act	gaa	acc	ggt	acc	tcc	gct	ttt	ggt	atc	aac	caa	acc	cca	ttc	1200
Ile	Thr	Glu	Thr	Gly	Thr	Ser	Ala	Phe	Gly	Ile	Asn	Gln	Thr	Pro	Phe	
					390					395					400	
cca	aac	aac	gct	tac	ggt	atc	tcc	caa	ggt	cta	tgg	ggt	tcc	atc	ggt	1248
Pro	Asn	Asn	Ala	Tyr	Gly	Ile	Ser	Gln	Val	Leu	Trp	Gly	Ser	Ile	Gly	
				405					410					415		
ttc	acc	acc	ggt	gct	tgt	ttg	ggt	gcc	gct	ttc	gct	gct	gaa	gaa	atc	1296
Phe	Thr	Thr	Gly	Ala	Cys	Leu	Gly	Ala	Ala	Phe	Ala	Ala	Glu	Glu	Ile	
			420					425					430			

20110907_CL5178USNA_SeqList.txt

gac cca aag aag aga gtt atc ttg ttc att ggt gac ggt tct ttg caa	1344
Asp Pro Lys Lys Arg Val Ile Leu Phe Ile Gly Asp Gly Ser Leu Gln	
435 440 445	
ttg act gtc caa gaa atc tcc acc atg atc aga tgg ggc ttg aag cca	1392
Leu Thr Val Gln Glu Ile Ser Thr Met Ile Arg Trp Gly Leu Lys Pro	
450 455 460	
tac ttg ttc gtc ttg aac aac gac ggt tac acc atc gaa aga ttg att	1440
Tyr Leu Phe Val Leu Asn Asn Asp Gly Tyr Thr Ile Glu Arg Leu Ile	
465 470 475 480	
cac ggt gaa aag gct ggt tac aac gac atc caa aac tgg gac cac ttg	1488
His Gly Glu Lys Ala Gly Tyr Asn Asp Ile Gln Asn Trp Asp His Leu	
485 490 495	
gct cta ttg cca acc ttc ggt gct aag gac tac gaa aac cac aga gtc	1536
Ala Leu Leu Pro Thr Phe Gly Ala Lys Asp Tyr Glu Asn His Arg Val	
500 505 510	
gcc acc acc ggt gaa tgg gac aag ttg acc caa gac aag gaa ttc aac	1584
Ala Thr Thr Gly Glu Trp Asp Lys Leu Thr Gln Asp Lys Glu Phe Asn	
515 520 525	
aag aac tcc aag atc aga atg atc gaa gtt atg ttg cca gtt atg gac	1632
Lys Asn Ser Lys Ile Arg Met Ile Glu Val Met Leu Pro Val Met Asp	
530 535 540	
gct cca act tcc ttg att gaa caa gct aag ttg acc gct tcc atc aac	1680
Ala Pro Thr Ser Leu Ile Glu Gln Ala Lys Leu Thr Ala Ser Ile Asn	
545 550 555 560	
gct aag caa gaa taa	1695
Ala Lys Gln Glu	

<210> 57
 <211> 564
 <212> PRT
 <213> Candida glabrata

<400> 57

Met Ser Glu Ile Thr Leu Gly Arg Tyr Leu Phe Glu Arg Leu Asn Gln	
1 5 10 15	
Val Asp Val Lys Thr Ile Phe Gly Leu Pro Gly Asp Phe Asn Leu Ser	
20 25 30	
Leu Leu Asp Lys Ile Tyr Glu Val Glu Gly Met Arg Trp Ala Gly Asn	
35 40 45	
Ala Asn Glu Leu Asn Ala Ala Tyr Ala Ala Asp Gly Tyr Ala Arg Ile	
50 55 60	
Lys Gly Met Ser Cys Ile Ile Thr Thr Phe Gly Val Gly Glu Leu Ser	
65 70 75 80	
Ala Leu Asn Gly Ile Ala Gly Ser Tyr Ala Glu His Val Gly Val Leu	
85 90 95	

20110907_CL5178USNA_SeqList.txt

His Val Val Gly Val Pro Ser Ile Ser Ser Gln Ala Lys Gln Leu Leu
100 105 110

Leu His His Thr Leu Gly Asn Gly Asp Phe Thr Val Phe His Arg Met
115 120 125

Ser Ala Asn Ile Ser Glu Thr Thr Ala Met Val Thr Asp Ile Ala Thr
130 135 140

Ala Pro Ala Glu Ile Asp Arg Cys Ile Arg Thr Thr Tyr Ile Thr Gln
145 150 155 160

Arg Pro Val Tyr Leu Gly Leu Pro Ala Asn Leu Val Asp Leu Lys Val
165 170 175

Pro Ala Lys Leu Leu Glu Thr Pro Ile Asp Leu Ser Leu Lys Pro Asn
180 185 190

Asp Pro Glu Ala Glu Thr Glu Val Val Asp Thr Val Leu Glu Leu Ile
195 200 205

Lys Ala Ala Lys Asn Pro Val Ile Leu Ala Asp Ala Cys Ala Ser Arg
210 215 220

His Asp Val Lys Ala Glu Thr Lys Lys Leu Ile Asp Ala Thr Gln Phe
225 230 235 240

Pro Ser Phe Val Thr Pro Met Gly Lys Gly Ser Ile Asp Glu Gln His
245 250 255

Pro Arg Phe Gly Gly Val Tyr Val Gly Thr Leu Ser Arg Pro Glu Val
260 265 270

Lys Glu Ala Val Glu Ser Ala Asp Leu Ile Leu Ser Val Gly Ala Leu
275 280 285

Leu Ser Asp Phe Asn Thr Gly Ser Phe Ser Tyr Ser Tyr Lys Thr Lys
290 295 300

Asn Ile Val Glu Phe His Ser Asp Tyr Ile Lys Ile Arg Asn Ala Thr
305 310 315 320

Phe Pro Gly Val Gln Met Lys Phe Ala Leu Gln Lys Leu Leu Asn Ala
325 330 335

Val Pro Glu Ala Ile Lys Gly Tyr Lys Pro Val Pro Val Pro Ala Arg
340 345 350

Val Pro Glu Asn Lys Ser Cys Asp Pro Ala Thr Pro Leu Lys Gln Glu
355 360 365

Trp Met Trp Asn Gln Val Ser Lys Phe Leu Gln Glu Gly Asp Val Val
 370 375 380

Ile Thr Glu Thr Gly Thr Ser Ala Phe Gly Ile Asn Gln Thr Pro Phe
 385 390 395 400

Pro Asn Asn Ala Tyr Gly Ile Ser Gln Val Leu Trp Gly Ser Ile Gly
 405 410 415

Phe Thr Thr Gly Ala Cys Leu Gly Ala Ala Phe Ala Ala Glu Glu Ile
 420 425 430

Asp Pro Lys Lys Arg Val Ile Leu Phe Ile Gly Asp Gly Ser Leu Gln
 435 440 445

Leu Thr Val Gln Glu Ile Ser Thr Met Ile Arg Trp Gly Leu Lys Pro
 450 455 460

Tyr Leu Phe Val Leu Asn Asn Asp Gly Tyr Thr Ile Glu Arg Leu Ile
 465 470 475 480

His Gly Glu Lys Ala Gly Tyr Asn Asp Ile Gln Asn Trp Asp His Leu
 485 490 495

Ala Leu Leu Pro Thr Phe Gly Ala Lys Asp Tyr Glu Asn His Arg Val
 500 505 510

Ala Thr Thr Gly Glu Trp Asp Lys Leu Thr Gln Asp Lys Glu Phe Asn
 515 520 525

Lys Asn Ser Lys Ile Arg Met Ile Glu Val Met Leu Pro Val Met Asp
 530 535 540

Ala Pro Thr Ser Leu Ile Glu Gln Ala Lys Leu Thr Ala Ser Ile Asn
 545 550 555 560

Ala Lys Gln Glu

<210> 58
 <211> 1791
 <212> DNA
 <213> *Pichia stipitis*

<220>
 <221> CDS
 <222> (1)..(1791)

<400> 58
 atg gct gaa gtc tca tta gga aga tat ctc ttc gag aga ttg tac caa 48
 Met Ala Glu Val Ser Leu Gly Arg Tyr Leu Phe Glu Arg Leu Tyr Gln
 1 5 10 15

ttg caa gtg cag acc atc ttc ggt gtc cct ggt gat ttc aac ttg tcg 96
 Page 88

20110907_CL5178USNA_SeqList.txt

Leu	Gln	Val	Gln	Thr	Ile	Phe	Gly	Val	Pro	Gly	Asp	Phe	Asn	Leu	Ser	
			20					25					30			
ctt	ttg	gac	aag	atc	tac	gaa	gtg	gaa	gat	gcc	cat	ggc	aag	aat	tcg	144
Leu	Leu	Asp	Lys	Ile	Tyr	Glu	Val	Glu	Asp	Ala	His	Gly	Lys	Asn	Ser	
		35					40					45				
ttt	aga	tgg	gct	ggc	aat	gcc	aac	gaa	ttg	aat	gca	tcg	tac	gct	gct	192
Phe	Arg	Trp	Ala	Gly	Asn	Ala	Asn	Glu	Leu	Asn	Ala	Ser	Tyr	Ala	Ala	
	50					55					60					
gac	ggc	tac	tcg	aga	gtc	aag	cgt	tta	ggg	tgt	ttg	gtc	act	acc	ttt	240
Asp	Gly	Tyr	Ser	Arg	Val	Lys	Arg	Leu	Gly	Cys	Leu	Val	Thr	Thr	Phe	
65				70					75						80	
ggc	gtc	ggc	gaa	ttg	tct	gct	ttg	aat	ggc	att	gcc	ggc	tct	tat	gcc	288
Gly	Val	Gly	Glu	Leu	Ser	Ala	Leu	Asn	Gly	Ile	Ala	Gly	Ser	Tyr	Ala	
			85					90						95		
gaa	cat	gtt	ggc	ttg	ctt	cat	gtc	gta	ggc	gtt	cca	tcg	att	tcc	tcg	336
Glu	His	Val	Gly	Leu	Leu	His	Val	Val	Gly	Val	Pro	Ser	Ile	Ser	Ser	
			100					105					110			
caa	gct	aag	caa	ttg	tta	ctt	cac	cac	act	ttg	ggc	aat	ggc	gat	ttc	384
Gln	Ala	Lys	Gln	Leu	Leu	Leu	His	His	Thr	Leu	Gly	Asn	Gly	Asp	Phe	
		115					120					125				
act	gtt	ttc	cat	aga	atg	tcc	aac	aac	att	tct	cag	acc	aca	gcc	ttt	432
Thr	Val	Phe	His	Arg	Met	Ser	Asn	Asn	Ile	Ser	Gln	Thr	Thr	Ala	Phe	
	130					135					140					
atc	tcc	gat	atc	aac	tcg	gct	cca	gct	gaa	att	gat	aga	tgt	atc	aga	480
Ile	Ser	Asp	Ile	Asn	Ser	Ala	Pro	Ala	Glu	Ile	Asp	Arg	Cys	Ile	Arg	
145					150					155					160	
gag	gcc	tac	gtc	aaa	caa	aga	cca	gtt	tat	atc	ggg	tta	cca	gct	aac	528
Glu	Ala	Tyr	Val	Lys	Gln	Arg	Pro	Val	Tyr	Ile	Gly	Leu	Pro	Ala	Asn	
				165					170					175		
tta	gtt	gat	ttg	aat	gtt	ccg	gcc	tct	ttg	ctt	gag	tct	cca	atc	aac	576
Leu	Val	Asp	Leu	Asn	Val	Pro	Ala	Ser	Leu	Leu	Glu	Ser	Pro	Ile	Asn	
			180					185					190			
ttg	tcg	ttg	gaa	aag	aac	gac	cca	gag	gct	caa	gat	gaa	gtc	att	gac	624
Leu	Ser	Leu	Glu	Lys	Asn	Asp	Pro	Glu	Ala	Gln	Asp	Glu	Val	Ile	Asp	
		195					200					205				
tct	gtc	tta	gac	ttg	atc	aaa	aag	tcg	ctg	aac	cca	atc	atc	ttg	gtc	672
Ser	Val	Leu	Asp	Leu	Ile	Lys	Lys	Ser	Leu	Asn	Pro	Ile	Ile	Leu	Val	
	210					215					220					
gat	gcc	tgt	gcc	tcg	aga	cat	gac	tgt	aag	gct	gaa	gtt	act	cag	ttg	720
Asp	Ala	Cys	Ala	Ser	Arg	His	Asp	Cys	Lys	Ala	Glu	Val	Thr	Gln	Leu	
225					230				235						240	
att	gaa	caa	acc	caa	ttc	cca	gta	ttt	gtc	act	cca	atg	ggc	aaa	ggc	768
Ile	Glu	Gln	Thr	Gln	Phe	Pro	Val	Phe	Val	Thr	Pro	Met	Gly	Lys	Gly	
				245					250					255		
acc	gtt	gat	gag	ggc	ggc	gta	gac	gga	gaa	ttg	tta	gaa	gat	gat	cct	816
Thr	Val	Asp	Glu	Gly	Gly	Val	Asp	Gly	Glu	Leu	Leu	Glu	Asp	Asp	Pro	
			260					265					270			
cat	ttg	att	gcc	aag	gtc	gct	gct	agg	ttg	tct	gct	ggc	aag	aac	gct	864
His	Leu	Ile	Ala	Lys	Val	Ala	Ala	Arg	Leu	Ser	Ala	Gly	Lys	Asn	Ala	
		275				280						285				
gcc	tct	aga	ttc	gga	ggc	gtt	tat	gtc	gga	acc	ttg	tcg	aag	ccc	gaa	912

20110907_CL5178USNA_SeqList.txt

Ala	Ser	Arg	Phe	Gly	Gly	Val	Tyr	Val	Gly	Thr	Leu	Ser	Lys	Pro	Glu	
290						295					300					
gtc	aag	gac	gct	gta	gag	agt	gca	gat	ttg	att	ttg	tct	gtc	ggg	gcc	960
Val	Lys	Asp	Ala	Val	Glu	Ser	Ala	Asp	Leu	Ile	Leu	Ser	Val	Gly	Ala	
305					310					315					320	
ctt	ttg	tct	gat	ttc	aac	act	ggg	tca	ttt	tcc	tac	tcc	tac	aga	acc	1008
Leu	Leu	Ser	Asp	Phe	Asn	Thr	Gly	Ser	Phe	Ser	Tyr	Ser	Tyr	Arg	Thr	
				325					330					335		
aag	aac	atc	gtc	gaa	ttc	cat	tct	gat	tac	act	aag	att	aga	caa	gcc	1056
Lys	Asn	Ile	Val	Glu	Phe	His	Ser	Asp	Tyr	Thr	Lys	Ile	Arg	Gln	Ala	
			340					345					350			
act	ttc	cca	ggg	gtg	cag	atg	aag	gaa	gcc	ttg	caa	gaa	ttg	aac	aag	1104
Thr	Phe	Pro	Gly	Val	Gln	Met	Lys	Glu	Ala	Leu	Gln	Glu	Leu	Asn	Lys	
		355					360					365				
aaa	gtt	tca	tct	gct	gct	agt	cac	tat	gaa	gtc	aag	cct	gtg	ccc	aag	1152
Lys	Val	Ser	Ser	Ala	Ala	Ser	His	Tyr	Glu	Val	Lys	Pro	Val	Pro	Lys	
	370					375					380					
atc	aag	ttg	gcc	aat	aca	cca	gcc	acc	aga	gaa	gtc	aag	tta	act	cag	1200
Ile	Lys	Leu	Ala	Asn	Thr	Pro	Ala	Thr	Arg	Glu	Val	Lys	Leu	Thr	Gln	
385					390					395					400	
gaa	tgg	ttg	tgg	acc	aga	gtg	tct	tcg	tgg	ttc	aga	gaa	ggg	gat	att	1248
Glu	Trp	Leu	Trp	Thr	Arg	Val	Ser	Ser	Trp	Phe	Arg	Glu	Gly	Asp	Ile	
				405					410					415		
att	atc	acc	gaa	acc	ggg	aca	tcc	tcc	ttc	ggg	ata	gtt	caa	tcc	aga	1296
Ile	Ile	Thr	Glu	Thr	Gly	Thr	Ser	Ser	Phe	Gly	Ile	Val	Gln	Ser	Arg	
			420					425					430			
ttc	cca	aac	aac	acc	atc	ggg	atc	tcc	caa	gta	ttg	tgg	ggg	tct	att	1344
Phe	Pro	Asn	Asn	Thr	Ile	Gly	Ile	Ser	Gln	Val	Leu	Trp	Gly	Ser	Ile	
		435					440					445				
ggg	ttc	tct	gtt	ggg	gcc	act	ttg	ggg	gct	gcc	atg	gct	gcc	caa	gaa	1392
Gly	Phe	Ser	Val	Gly	Ala	Thr	Leu	Gly	Ala	Ala	Met	Ala	Ala	Gln	Glu	
	450					455					460					
ctc	gac	cct	aac	aag	aga	acc	atc	ttg	ttt	gtt	gga	gat	ggg	tct	ttg	1440
Leu	Asp	Pro	Asn	Lys	Arg	Thr	Ile	Leu	Phe	Val	Gly	Asp	Gly	Ser	Leu	
465					470				475						480	
caa	ttg	acc	gtt	cag	gaa	atc	tcc	acc	ata	atc	aga	tgg	ggg	acc	aca	1488
Gln	Leu	Thr	Val	Gln	Glu	Ile	Ser	Thr	Ile	Ile	Arg	Trp	Gly	Thr	Thr	
				485					490					495		
cct	tac	ctt	ttc	gtg	ttg	aac	aat	gac	ggg	tac	acc	atc	gag	cgt	ttg	1536
Pro	Tyr	Leu	Phe	Val	Leu	Asn	Asn	Asp	Gly	Tyr	Thr	Ile	Glu	Arg	Leu	
			500					505					510			
atc	cac	ggg	gta	aat	gcc	tca	tat	aat	gac	atc	caa	cca	tgg	caa	aac	1584
Ile	His	Gly	Val	Asn	Ala	Ser	Tyr	Asn	Asp	Ile	Gln	Pro	Trp	Gln	Asn	
		515					520					525				
ttg	gaa	atc	ttg	cct	act	ttc	tcg	gcc	aag	aac	tac	gac	gct	gtg	aga	1632
Leu	Glu	Ile	Leu	Pro	Thr	Phe	Ser	Ala	Lys	Asn	Tyr	Asp	Ala	Val	Arg	
	530					535					540					
atc	tcc	aac	atc	gga	gaa	gca	gaa	gat	atc	ttg	aaa	gac	aag	gaa	ttc	1680
Ile	Ser	Asn	Ile	Gly	Glu	Ala	Glu	Asp	Ile	Leu	Lys	Asp	Lys	Glu	Phe	
545					550					555					560	
gga	aag	aac	tcc	aag	att	aga	ttg	ata	gaa	gtc	atg	tta	cca	aga	ttg	1728

20110907_CL5178USNA_SeqList.txt

Gly	Lys	Asn	Ser	Lys	Ile	Arg	Leu	Ile	Glu	Val	Met	Leu	Pro	Arg	Leu		
				565					570					575			
gat	gca	cca	tct	aac	ctt	gcc	aaa	caa	gct	gcc	att	aca	gct	gcc	acc		1776
Asp	Ala	Pro	Ser	Asn	Leu	Ala	Lys	Gln	Ala	Ala	Ile	Thr	Ala	Ala	Thr		
			580					585					590				
aac	gcc	gaa	gct	tag													1791
Asn	Ala	Glu	Ala														
		595															
<210>	59																
<211>	596																
<212>	PRT																
<213>	Pichia stipitis																
<400>	59																
Met	Ala	Glu	Val	Ser	Leu	Gly	Arg	Tyr	Leu	Phe	Glu	Arg	Leu	Tyr	Gln		
1				5					10					15			
Leu	Gln	Val	Gln	Thr	Ile	Phe	Gly	Val	Pro	Gly	Asp	Phe	Asn	Leu	Ser		
			20					25					30				
Leu	Leu	Asp	Lys	Ile	Tyr	Glu	Val	Glu	Asp	Ala	His	Gly	Lys	Asn	Ser		
		35					40					45					
Phe	Arg	Trp	Ala	Gly	Asn	Ala	Asn	Glu	Leu	Asn	Ala	Ser	Tyr	Ala	Ala		
	50					55					60						
Asp	Gly	Tyr	Ser	Arg	Val	Lys	Arg	Leu	Gly	Cys	Leu	Val	Thr	Thr	Phe		
65					70					75					80		
Gly	Val	Gly	Glu	Leu	Ser	Ala	Leu	Asn	Gly	Ile	Ala	Gly	Ser	Tyr	Ala		
				85					90					95			
Glu	His	Val	Gly	Leu	Leu	His	Val	Val	Gly	Val	Pro	Ser	Ile	Ser	Ser		
			100					105					110				
Gln	Ala	Lys	Gln	Leu	Leu	Leu	His	His	Thr	Leu	Gly	Asn	Gly	Asp	Phe		
		115					120					125					
Thr	Val	Phe	His	Arg	Met	Ser	Asn	Asn	Ile	Ser	Gln	Thr	Thr	Ala	Phe		
	130					135					140						
Ile	Ser	Asp	Ile	Asn	Ser	Ala	Pro	Ala	Glu	Ile	Asp	Arg	Cys	Ile	Arg		
145					150					155					160		
Glu	Ala	Tyr	Val	Lys	Gln	Arg	Pro	Val	Tyr	Ile	Gly	Leu	Pro	Ala	Asn		
				165					170					175			
Leu	Val	Asp	Leu	Asn	Val	Pro	Ala	Ser	Leu	Leu	Glu	Ser	Pro	Ile	Asn		
			180					185					190				
Leu	Ser	Leu	Glu	Lys	Asn	Asp	Pro	Glu	Ala	Gln	Asp	Glu	Val	Ile	Asp		

195

200

205

Ser Val Leu Asp Leu Ile Lys Lys Ser Leu Asn Pro Ile Ile Leu Val
 210 215 220

Asp Ala Cys Ala Ser Arg His Asp Cys Lys Ala Glu Val Thr Gln Leu
 225 230 235 240

Ile Glu Gln Thr Gln Phe Pro Val Phe Val Thr Pro Met Gly Lys Gly
 245 250 255

Thr Val Asp Glu Gly Gly Val Asp Gly Glu Leu Leu Glu Asp Asp Pro
 260 265 270

His Leu Ile Ala Lys Val Ala Ala Arg Leu Ser Ala Gly Lys Asn Ala
 275 280 285

Ala Ser Arg Phe Gly Gly Val Tyr Val Gly Thr Leu Ser Lys Pro Glu
 290 295 300

Val Lys Asp Ala Val Glu Ser Ala Asp Leu Ile Leu Ser Val Gly Ala
 305 310 315 320

Leu Leu Ser Asp Phe Asn Thr Gly Ser Phe Ser Tyr Ser Tyr Arg Thr
 325 330 335

Lys Asn Ile Val Glu Phe His Ser Asp Tyr Thr Lys Ile Arg Gln Ala
 340 345 350

Thr Phe Pro Gly Val Gln Met Lys Glu Ala Leu Gln Glu Leu Asn Lys
 355 360 365

Lys Val Ser Ser Ala Ala Ser His Tyr Glu Val Lys Pro Val Pro Lys
 370 375 380

Ile Lys Leu Ala Asn Thr Pro Ala Thr Arg Glu Val Lys Leu Thr Gln
 385 390 395 400

Glu Trp Leu Trp Thr Arg Val Ser Ser Trp Phe Arg Glu Gly Asp Ile
 405 410 415

Ile Ile Thr Glu Thr Gly Thr Ser Ser Phe Gly Ile Val Gln Ser Arg
 420 425 430

Phe Pro Asn Asn Thr Ile Gly Ile Ser Gln Val Leu Trp Gly Ser Ile
 435 440 445

Gly Phe Ser Val Gly Ala Thr Leu Gly Ala Ala Met Ala Ala Gln Glu
 450 455 460

Leu Asp Pro Asn Lys Arg Thr Ile Leu Phe Val Gly Asp Gly Ser Leu
 Page 92

465 470 475 480

Gln Leu Thr Val Gln Glu Ile Ser Thr Ile Ile Arg Trp Gly Thr Thr
485 490 495

Pro Tyr Leu Phe Val Leu Asn Asn Asp Gly Tyr Thr Ile Glu Arg Leu
500 505 510

Ile His Gly Val Asn Ala Ser Tyr Asn Asp Ile Gln Pro Trp Gln Asn
515 520 525

Leu Glu Ile Leu Pro Thr Phe Ser Ala Lys Asn Tyr Asp Ala Val Arg
530 535 540

Ile Ser Asn Ile Gly Glu Ala Glu Asp Ile Leu Lys Asp Lys Glu Phe
545 550 555 560

Gly Lys Asn Ser Lys Ile Arg Leu Ile Glu Val Met Leu Pro Arg Leu
565 570 575

Asp Ala Pro Ser Asn Leu Ala Lys Gln Ala Ala Ile Thr Ala Ala Thr
580 585 590

Asn Ala Glu Ala
595

<210> 60
<211> 1710
<212> DNA
<213> Pichia stipitis

<220>
<221> CDS
<222> (1)..(1710)

<400> 60
atg gta tca acc tac cca gaa tca gag gtt act cta gga agg tac ctc 48
Met Val Ser Thr Tyr Pro Glu Ser Glu Val Thr Leu Gly Arg Tyr Leu
1 5 10 15

ttt gag cga ctc cac caa ttg aaa gtg gac acc att ttc ggc ttg ccg 96
Phe Glu Arg Leu His Gln Leu Lys Val Asp Thr Ile Phe Gly Leu Pro
20 25 30

ggt gac ttc aac ctt tcc tta ttg gac aaa gtg tat gaa gtt ccg gat 144
Gly Asp Phe Asn Leu Ser Leu Leu Asp Lys Val Tyr Glu Val Pro Asp
35 40 45

atg agg tgg gct gga aat gcc aac gaa ttg aat gct gcc tat gct gcc 192
Met Arg Trp Ala Gly Asn Ala Asn Glu Leu Asn Ala Ala Tyr Ala Ala
50 55 60

gat ggt tac tcc aga ata aag gga ttg tct tgc ttg gtc aca act ttt 240
Asp Gly Tyr Ser Arg Ile Lys Gly Leu Ser Cys Leu Val Thr Thr Phe
65 70 75 80

ggt gtt ggt gaa ttg tct gct tta aac gga gtt ggt ggt gcc tat gct 288
Gly Val Gly Glu Leu Ser Ala Leu Asn Gly Val Gly Gly Ala Tyr Ala

20110907_CL5178USNA_SeqList.txt

85															90					95					
gaa	cac	gta	gga	ctt	cta	cat	gtc	ggt	gga	ggt	cca	tcc	ata	tcg	tca	336									
Glu	His	Val	Gly	Leu	Leu	His	Val	Val	Gly	Val	Pro	Ser	Ile	Ser	Ser										
			100					105					110												
cag	gct	aaa	cag	ttg	ttg	ctc	cac	cat	acc	ttg	ggt	aat	ggt	gac	ttc	384									
Gln	Ala	Lys	Gln	Leu	Leu	Leu	His	His	Thr	Leu	Gly	Asn	Gly	Asp	Phe										
		115					120					125													
act	ggt	ttt	cac	aga	atg	tcc	aat	agc	att	tct	caa	act	aca	gca	ttt	432									
Thr	Val	Phe	His	Arg	Met	Ser	Asn	Ser	Ile	Ser	Gln	Thr	Thr	Ala	Phe										
	130					135					140														
ctc	tca	gat	atc	tct	att	gca	cca	ggt	caa	ata	gat	aga	tgc	atc	aga	480									
Leu	Ser	Asp	Ile	Ser	Ile	Ala	Pro	Gly	Gln	Ile	Asp	Arg	Cys	Ile	Arg										
145					150					155					160										
gaa	gca	tat	ggt	cat	cag	aga	cca	ggt	tat	ggt	ggt	tta	ccg	gca	aat	528									
Glu	Ala	Tyr	Val	His	Gln	Arg	Pro	Val	Tyr	Val	Gly	Leu	Pro	Ala	Asn										
				165					170					175											
atg	ggt	gat	ctc	aag	ggt	cct	tct	agt	ctc	tta	gaa	act	cca	att	gat	576									
Met	Val	Asp	Leu	Lys	Val	Pro	Ser	Ser	Leu	Leu	Glu	Thr	Pro	Ile	Asp										
			180					185					190												
ttg	aaa	ttg	aaa	caa	aat	gat	cct	gaa	gct	caa	gaa	ggt	ggt	gaa	aca	624									
Leu	Lys	Leu	Lys	Gln	Asn	Asp	Pro	Glu	Ala	Gln	Glu	Val	Val	Glu	Thr										
	195						200					205													
gtc	ctg	aag	ttg	gtg	tcc	caa	gct	aca	aac	ccc	att	atc	ttg	gta	gac	672									
Val	Leu	Lys	Leu	Val	Ser	Gln	Ala	Thr	Asn	Pro	Ile	Ile	Leu	Val	Asp										
	210					215					220														
gct	tgt	gcc	ctc	aga	cac	aat	tgc	aaa	gag	gaa	gtc	aaa	caa	ttg	ggt	720									
Ala	Cys	Ala	Leu	Arg	His	Asn	Cys	Lys	Glu	Glu	Val	Lys	Gln	Leu	Val										
225					230				235					240											
gat	gcc	act	aat	ttt	caa	gtc	ttt	aca	act	cca	atg	ggt	aaa	tct	ggt	768									
Asp	Ala	Thr	Asn	Phe	Gln	Val	Phe	Thr	Thr	Pro	Met	Gly	Lys	Ser	Gly										
				245				250						255											
atc	tcc	gaa	tct	cat	cca	aga	ttg	ggc	ggt	gtc	tat	gtc	ggg	aca	atg	816									
Ile	Ser	Glu	Ser	His	Pro	Arg	Leu	Gly	Gly	Val	Tyr	Val	Gly	Thr	Met										
			260					265					270												
tcg	agt	cct	caa	gtc	aaa	aaa	gcc	ggt	gaa	aat	gcc	gat	ctt	ata	cta	864									
Ser	Ser	Pro	Gln	Val	Lys	Lys	Ala	Val	Glu	Asn	Ala	Asp	Leu	Ile	Leu										
		275					280					285													
tct	ggt	ggt	tcg	ttg	tta	tcg	gac	ttc	aat	aca	ggt	tca	ttt	tca	tac	912									
Ser	Val	Gly	Ser	Leu	Leu	Ser	Asp	Phe	Asn	Thr	Gly	Ser	Phe	Ser	Tyr										
	290					295					300														
tcc	tac	aag	acg	aag	aat	ggt	ggt	gaa	ttc	cac	tct	gac	tat	atg	aaa	960									
Ser	Tyr	Lys	Thr	Lys	Asn	Val	Val	Glu	Phe	His	Ser	Asp	Tyr	Met	Lys										
305					310				315						320										
atc	aga	cag	gcc	acc	ttc	cca	gga	ggt	caa	atg	aaa	gaa	gcc	ttg	caa	1008									
Ile	Arg	Gln	Ala	Thr	Phe	Pro	Gly	Val	Gln	Met	Lys	Glu	Ala	Leu	Gln										
				325					330					335											
cag	ttg	ata	aaa	agg	gtc	tct	tct	tac	atc	aat	cca	agc	tac	att	cct	1056									
Gln	Leu	Ile	Lys	Arg	Val	Ser	Ser	Tyr	Ile	Asn	Pro	Ser	Tyr	Ile	Pro										
			340					345					350												
act	cga	ggt	cct	aaa	agg	aaa	cag	cca	ttg	aaa	gct	cca	tca	gaa	gct	1104									
Thr	Arg	Val	Pro	Lys	Arg	Lys	Gln	Pro	Leu	Lys	Ala	Pro	Ser	Glu	Ala										

20110907_CL5178USNA_SeqList.txt

355	360	365	
cct ttg acc caa gaa tat ttg tgg tct aaa gta tcc ggc tgg ttt aga Pro Leu Thr Gln Glu Tyr Leu Trp Ser Lys Val Ser Gly Trp Phe Arg 370 375 380			1152
gag ggt gat att atc gta acc gaa act ggt aca tct gct ttc gga att Glu Gly Asp Ile Ile Val Thr Glu Thr Gly Thr Ser Ala Phe Gly Ile 385 390 395 400			1200
att caa tcc cat ttt ccc agc aac act atc ggt ata tcc caa gtc ttg Ile Gln Ser His Phe Pro Ser Asn Thr Ile Gly Ile Ser Gln Val Leu 405 410 415			1248
tgg ggc tca att ggt ttc aca gta ggt gca aca gtt ggt gct gcc atg Trp Gly Ser Ile Gly Phe Thr Val Gly Ala Thr Val Gly Ala Ala Met 420 425 430			1296
gca gcc cag gaa atc gac cct agc agg aga gta att ttg ttc gtc ggt Ala Ala Gln Glu Ile Asp Pro Ser Arg Arg Val Ile Leu Phe Val Gly 435 440 445			1344
gat ggt tca ttg cag ttg acg gtt cag gaa atc tct acg ttg tgt aaa Asp Gly Ser Leu Gln Leu Thr Val Gln Glu Ile Ser Thr Leu Cys Lys 450 455 460			1392
tgg gat tgt aac aat act tat ctt tac gtg ttg aac aat gat ggt tac Trp Asp Cys Asn Asn Thr Tyr Leu Tyr Val Leu Asn Asn Asp Gly Tyr 465 470 475 480			1440
act ata gaa agg ttg atc cac ggc aaa agt gcc agc tac aac gat ata Thr Ile Glu Arg Leu Ile His Gly Lys Ser Ala Ser Tyr Asn Asp Ile 485 490 495			1488
cag cct tgg aac cat tta tcc ttg ctt cgc tta ttc aat gct aag aaa Gln Pro Trp Asn His Leu Ser Leu Leu Arg Leu Phe Asn Ala Lys Lys 500 505 510			1536
tac caa aat gtc aga gta tcg act gct gga gaa ttg gac tct ttg ttc Tyr Gln Asn Val Arg Val Ser Thr Ala Gly Glu Leu Asp Ser Leu Phe 515 520 525			1584
tct gat aag aaa ttt gct tct cca gat agg ata aga atg att gag gtg Ser Asp Lys Lys Phe Ala Ser Pro Asp Arg Ile Arg Met Ile Glu Val 530 535 540			1632
atg tta tcg aga ttg gat gca cca gca aat ctt gtt gct caa gca aag Met Leu Ser Arg Leu Asp Ala Pro Ala Asn Leu Val Ala Gln Ala Lys 545 550 555 560			1680
ttg tct gaa cgg gta aac ctt gaa aat tga Leu Ser Glu Arg Val Asn Leu Glu Asn 565			1710

<210> 61
 <211> 569
 <212> PRT
 <213> Pichia stipitis

<400> 61

Met Val Ser Thr Tyr Pro Glu Ser Glu Val Thr Leu Gly Arg Tyr Leu
1 5 10 15

Phe Glu Arg Leu His Gln Leu Lys Val Asp Thr Ile Phe Gly Leu Pro
20 25 30

20110907_CL5178USNA_SeqList.txt

Gly Asp Phe Asn Leu Ser Leu Leu Asp Lys Val Tyr Glu Val Pro Asp
35 40 45

Met Arg Trp Ala Gly Asn Ala Asn Glu Leu Asn Ala Ala Tyr Ala Ala
50 55 60

Asp Gly Tyr Ser Arg Ile Lys Gly Leu Ser Cys Leu Val Thr Thr Phe
65 70 75 80

Gly Val Gly Glu Leu Ser Ala Leu Asn Gly Val Gly Gly Ala Tyr Ala
85 90 95

Glu His Val Gly Leu Leu His Val Val Gly Val Pro Ser Ile Ser Ser
100 105 110

Gln Ala Lys Gln Leu Leu Leu His His Thr Leu Gly Asn Gly Asp Phe
115 120 125

Thr Val Phe His Arg Met Ser Asn Ser Ile Ser Gln Thr Thr Ala Phe
130 135 140

Leu Ser Asp Ile Ser Ile Ala Pro Gly Gln Ile Asp Arg Cys Ile Arg
145 150 155 160

Glu Ala Tyr Val His Gln Arg Pro Val Tyr Val Gly Leu Pro Ala Asn
165 170 175

Met Val Asp Leu Lys Val Pro Ser Ser Leu Leu Glu Thr Pro Ile Asp
180 185 190

Leu Lys Leu Lys Gln Asn Asp Pro Glu Ala Gln Glu Val Val Glu Thr
195 200 205

Val Leu Lys Leu Val Ser Gln Ala Thr Asn Pro Ile Ile Leu Val Asp
210 215 220

Ala Cys Ala Leu Arg His Asn Cys Lys Glu Glu Val Lys Gln Leu Val
225 230 235 240

Asp Ala Thr Asn Phe Gln Val Phe Thr Thr Pro Met Gly Lys Ser Gly
245 250 255

Ile Ser Glu Ser His Pro Arg Leu Gly Gly Val Tyr Val Gly Thr Met
260 265 270

Ser Ser Pro Gln Val Lys Lys Ala Val Glu Asn Ala Asp Leu Ile Leu
275 280 285

Ser Val Gly Ser Leu Leu Ser Asp Phe Asn Thr Gly Ser Phe Ser Tyr
290 295 300

20110907_CL5178USNA_SeqList.txt

Ser Tyr Lys Thr Lys Asn Val Val Glu Phe His Ser Asp Tyr Met Lys
305 310 315 320

Ile Arg Gln Ala Thr Phe Pro Gly Val Gln Met Lys Glu Ala Leu Gln
325 330 335

Gln Leu Ile Lys Arg Val Ser Ser Tyr Ile Asn Pro Ser Tyr Ile Pro
340 345 350

Thr Arg Val Pro Lys Arg Lys Gln Pro Leu Lys Ala Pro Ser Glu Ala
355 360 365

Pro Leu Thr Gln Glu Tyr Leu Trp Ser Lys Val Ser Gly Trp Phe Arg
370 375 380

Glu Gly Asp Ile Ile Val Thr Glu Thr Gly Thr Ser Ala Phe Gly Ile
385 390 395 400

Ile Gln Ser His Phe Pro Ser Asn Thr Ile Gly Ile Ser Gln Val Leu
405 410 415

Trp Gly Ser Ile Gly Phe Thr Val Gly Ala Thr Val Gly Ala Ala Met
420 425 430

Ala Ala Gln Glu Ile Asp Pro Ser Arg Arg Val Ile Leu Phe Val Gly
435 440 445

Asp Gly Ser Leu Gln Leu Thr Val Gln Glu Ile Ser Thr Leu Cys Lys
450 455 460

Trp Asp Cys Asn Asn Thr Tyr Leu Tyr Val Leu Asn Asn Asp Gly Tyr
465 470 475 480

Thr Ile Glu Arg Leu Ile His Gly Lys Ser Ala Ser Tyr Asn Asp Ile
485 490 495

Gln Pro Trp Asn His Leu Ser Leu Leu Arg Leu Phe Asn Ala Lys Lys
500 505 510

Tyr Gln Asn Val Arg Val Ser Thr Ala Gly Glu Leu Asp Ser Leu Phe
515 520 525

Ser Asp Lys Lys Phe Ala Ser Pro Asp Arg Ile Arg Met Ile Glu Val
530 535 540

Met Leu Ser Arg Leu Asp Ala Pro Ala Asn Leu Val Ala Gln Ala Lys
545 550 555 560

Leu Ser Glu Arg Val Asn Leu Glu Asn
565

<210> 62
 <211> 1692
 <212> DNA
 <213> Kluyveromyces lactis

<220>
 <221> CDS
 <222> (1)..(1692)

<400> 62
 atg tct gaa att aca tta ggt cgt tac ttg ttc gaa aga tta aag caa 48
 Met Ser Glu Ile Thr Leu Gly Arg Tyr Leu Phe Glu Arg Leu Lys Gln
 1 5 10 15
 gtc gaa gtt caa acc atc ttt ggt cta cca ggt gat ttc aac ttg tcc 96
 Val Glu Val Gln Thr Ile Phe Gly Leu Pro Gly Asp Phe Asn Leu Ser
 20 25 30
 cta ttg gac aat atc tac gaa gtc cca ggt atg aga tgg gct ggt aat 144
 Leu Leu Asp Asn Ile Tyr Glu Val Pro Gly Met Arg Trp Ala Gly Asn
 35 40 45
 gcc aac gaa ttg aac gct gct tac gct gct gat ggt tac gcc aga tta 192
 Ala Asn Glu Leu Asn Ala Ala Tyr Ala Ala Asp Gly Tyr Ala Arg Leu
 50 55 60
 aag ggt atg tcc tgt atc atc acc acc ttc ggt gtc ggt gaa ttg tct 240
 Lys Gly Met Ser Cys Ile Ile Thr Thr Phe Gly Val Gly Glu Leu Ser
 65 70 75 80
 gct ttg aac ggt att gcc ggt tct tac gct gaa cac gtt ggt gtc ttg 288
 Ala Leu Asn Gly Ile Ala Gly Ser Tyr Ala Glu His Val Gly Val Leu
 85 90 95
 cac gtt gtc ggt gtt cca tcc gtc tct tct caa gct aag caa ttg ttg 336
 His Val Val Gly Val Pro Ser Val Ser Ser Gln Ala Lys Gln Leu Leu
 100 105 110
 ttg cac cac acc ttg ggt aac ggt gac ttc act gtt ttc cac aga atg 384
 Leu His His Thr Leu Gly Asn Gly Asp Phe Thr Val Phe His Arg Met
 115 120 125
 tcc tcc aac att tct gaa acc act gct atg atc acc gat atc aac act 432
 Ser Ser Asn Ile Ser Glu Thr Thr Ala Met Ile Thr Asp Ile Asn Thr
 130 135 140
 gcc cca gct gaa atc gac aga tgt atc aga acc act tac gtt tcc caa 480
 Ala Pro Ala Glu Ile Asp Arg Cys Ile Arg Thr Thr Tyr Val Ser Gln
 145 150 155 160
 aga cca gtc tac ttg ggt ttg cca gct aac ttg gtc gac ttg act gtc 528
 Arg Pro Val Tyr Leu Gly Leu Pro Ala Asn Leu Val Asp Leu Thr Val
 165 170 175
 cca gct tct ttg ttg gac act cca att gat ttg agc ttg aag cca aat 576
 Pro Ala Ser Leu Leu Asp Thr Pro Ile Asp Leu Ser Leu Lys Pro Asn
 180 185 190
 gac cca gaa gcc gaa gaa gaa gtc atc gaa aac gtc ttg caa ctg atc 624
 Asp Pro Glu Ala Glu Glu Glu Val Ile Glu Asn Val Leu Gln Leu Ile
 195 200 205
 aag gaa gct aag aac cca gtt atc ttg gct gat gct tgt tgt tcc aga 672
 Lys Glu Ala Lys Asn Pro Val Ile Leu Ala Asp Ala Cys Cys Ser Arg
 210 215 220

20110907_CL5178USNA_SeqList.txt

cac His 225	gat Asp	gcc Ala	aag Lys	gct Ala	gag Glu 230	acc Thr	aag Lys	aag Lys	ttg Leu	atc Ile 235	gac Asp	ttg Leu	act Thr	caa Gln	ttc Phe 240	720
cca Pro	gcc Ala	ttc Phe	gtt Val	acc Thr 245	cca Pro	atg Met	ggg Gly	aag Lys	ggg Gly 250	tcc Ser	att Ile	gac Asp	gaa Glu	aag Lys 255	cac His	768
cca Pro	aga Arg	ttc Phe	ggg Gly 260	ggg Gly	gtc Val	tac Tyr	gtc Val	ggg Gly 265	acc Thr	cta Leu	tct Ser	tct Ser	cca Pro 270	gct Ala	gtc Val	816
aag Lys	gaa Glu	gcc Ala 275	gtt Val	gaa Glu	tct Ser	gct Ala	gac Asp 280	ttg Leu	gtt Val	cta Leu	tgc Ser	gtc Val 285	ggg Gly	gct Ala	cta Leu	864
ttg Leu	tcc Ser 290	gat Asp	ttc Phe	aac Asn	act Thr	ggg Gly 295	tct Ser	ttc Phe	tct Ser	tac Tyr	tct Ser 300	tac Tyr	aag Lys	acc Thr	aag Lys	912
aac Asn 305	att Ile	gtc Val	gaa Glu	ttc Phe	cac His 310	tct Ser	gac Asp	tac Tyr	acc Thr	aag Lys 315	atc Ile	aga Arg	agc Ser	gct Ala	acc Thr 320	960
ttc Phe	cca Pro	ggg Gly	gtc Val	caa Gln 325	atg Met	aag Lys	ttc Phe	gct Ala	tta Leu 330	caa Gln	aaa Lys	ttg Leu	ttg Leu	act Thr 335	aag Lys	1008
gtt Val	gcc Ala	gat Asp	gct Ala 340	gct Ala	aag Lys	ggg Gly	tac Tyr	aag Lys 345	cca Pro	gtt Val	cca Pro	gtt Val	cca Pro 350	tct Ser	gaa Glu	1056
cca Pro	gaa Glu	cac His 355	aac Asn	gaa Glu	gct Ala	gtc Val	gct Ala 360	gac Asp	tcc Ser	act Thr	cca Pro	ttg Leu 365	aag Lys	caa Gln	gaa Glu	1104
tgg Trp	gtc Val 370	tgg Trp	act Thr	caa Gln	gtc Val	ggg Gly 375	gaa Glu	ttc Phe	ttg Leu	aga Arg	gaa Glu 380	ggg Gly	gat Asp	gtt Val	gtt Val	1152
atc Ile 385	act Thr	gaa Glu	acc Thr	ggg Gly	acc Thr 390	tct Ser	gcc Ala	ttc Phe	ggg Gly	atc Ile 395	aac Asn	caa Gln	act Thr	cat His	ttc Phe 400	1200
cca Pro	aac Asn	aac Asn	aca Thr	tac Tyr 405	ggg Gly	atc Ile	tct Ser	caa Gln	gtt Val 410	tta Leu	tgg Trp	ggg Gly	tcc Ser	att Ile 415	ggg Gly	1248
ttc Phe	acc Thr	act Thr	ggg Gly 420	gct Ala	acc Thr	ttg Leu	ggg Gly	gct Ala 425	gcc Ala	ttc Phe	gct Ala	gcc Ala	gaa Glu 430	gaa Glu	att Ile	1296
gat Asp	cca Pro	aag Lys 435	aag Lys	aga Arg	gtt Val	atc Ile	tta Leu 440	ttc Phe	att Ile	ggg Gly	gac Asp	ggg Gly 445	tct Ser	ttg Leu	caa Gln	1344
ttg Leu	act Thr 450	gtt Val	caa Gln	gaa Glu	atc Ile	tcc Ser 455	acc Thr	atg Met	atc Ile	aga Arg	tgg Trp 460	ggc Gly	ttg Leu	aag Lys	cca Pro	1392
tac Tyr 465	ttg Leu	ttc Phe	gta Val	ttg Leu	aac Asn 470	aac Asn	gac Asp	ggg Gly	tac Tyr	acc Thr 475	att Ile	gaa Glu	aga Arg	ttg Leu	att Ile 480	1440
cac His	ggg Gly	gaa Glu	acc Thr	gct Ala 485	caa Gln	tac Tyr	aac Asn	tgt Cys	atc Ile 490	caa Gln	aac Asn	tgg Trp	caa Gln	cac His 495	ttg Leu	1488

20110907_CL5178USNA_SeqList.txt

gaa tta ttg cca act ttc ggt gcc aag gac tac gaa gct gtc aga gtt 1536
 Glu Leu Leu Pro Thr Phe Gly Ala Lys Asp Tyr Glu Ala Val Arg Val
 500 505 510

tcc acc act ggt gaa tgg aac aag ttg acc act gac gaa aag ttc caa 1584
 Ser Thr Thr Gly Glu Trp Asn Lys Leu Thr Thr Asp Glu Lys Phe Gln
 515 520 525

gac aac acc aga atc aga ttg atc gaa gtt atg ttg cca act atg gat 1632
 Asp Asn Thr Arg Ile Arg Leu Ile Glu Val Met Leu Pro Thr Met Asp
 530 535 540

gct cca tct aac ttg gtt aag caa gct caa ttg act gct gct acc aac 1680
 Ala Pro Ser Asn Leu Val Lys Gln Ala Gln Leu Thr Ala Ala Thr Asn
 545 550 555 560

gct aag aac taa 1692
 Ala Lys Asn

<210> 63
 <211> 563
 <212> PRT
 <213> Kluyveromyces lactis

<400> 63

Met Ser Glu Ile Thr Leu Gly Arg Tyr Leu Phe Glu Arg Leu Lys Gln
 1 5 10 15

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

Leu Leu Asp Asn Ile Tyr Glu Val Pro Gly Met Arg Trp Ala Gly Asn
 35 40 45

Ala Asn Glu Leu Asn Ala Ala Tyr Ala Ala Asp Gly Tyr Ala Arg Leu
 50 55 60

Lys Gly Met Ser Cys Ile Ile Thr Thr Phe Gly Val Gly Glu Leu Ser
 65 70 75 80

Ala Leu Asn Gly Ile Ala Gly Ser Tyr Ala Glu His Val Gly Val Leu
 85 90 95

His Val Val Gly Val Pro Ser Val Ser Ser Gln Ala Lys Gln Leu Leu
 100 105 110

Leu His His Thr Leu Gly Asn Gly Asp Phe Thr Val Phe His Arg Met
 115 120 125

Ser Ser Asn Ile Ser Glu Thr Thr Ala Met Ile Thr Asp Ile Asn Thr
 130 135 140

Ala Pro Ala Glu Ile Asp Arg Cys Ile Arg Thr Thr Tyr Val Ser Gln
 145 150 155 160

20110907_CL5178USNA_SeqList.txt

Arg Pro Val Tyr Leu Gly Leu Pro Ala Asn Leu Val Asp Leu Thr Val
 165 170 175
 Pro Ala Ser Leu Leu Asp Thr Pro Ile Asp Leu Ser Leu Lys Pro Asn
 180 185 190
 Asp Pro Glu Ala Glu Glu Glu Val Ile Glu Asn Val Leu Gln Leu Ile
 195 200 205
 Lys Glu Ala Lys Asn Pro Val Ile Leu Ala Asp Ala Cys Cys Ser Arg
 210 215 220
 His Asp Ala Lys Ala Glu Thr Lys Lys Leu Ile Asp Leu Thr Gln Phe
 225 230 235 240
 Pro Ala Phe Val Thr Pro Met Gly Lys Gly Ser Ile Asp Glu Lys His
 245 250 255
 Pro Arg Phe Gly Gly Val Tyr Val Gly Thr Leu Ser Ser Pro Ala Val
 260 265 270
 Lys Glu Ala Val Glu Ser Ala Asp Leu Val Leu Ser Val Gly Ala Leu
 275 280 285
 Leu Ser Asp Phe Asn Thr Gly Ser Phe Ser Tyr Ser Tyr Lys Thr Lys
 290 295 300
 Asn Ile Val Glu Phe His Ser Asp Tyr Thr Lys Ile Arg Ser Ala Thr
 305 310 315 320
 Phe Pro Gly Val Gln Met Lys Phe Ala Leu Gln Lys Leu Leu Thr Lys
 325 330 335
 Val Ala Asp Ala Ala Lys Gly Tyr Lys Pro Val Pro Val Pro Ser Glu
 340 345 350
 Pro Glu His Asn Glu Ala Val Ala Asp Ser Thr Pro Leu Lys Gln Glu
 355 360 365
 Trp Val Trp Thr Gln Val Gly Glu Phe Leu Arg Glu Gly Asp Val Val
 370 375 380
 Ile Thr Glu Thr Gly Thr Ser Ala Phe Gly Ile Asn Gln Thr His Phe
 385 390 395 400
 Pro Asn Asn Thr Tyr Gly Ile Ser Gln Val Leu Trp Gly Ser Ile Gly
 405 410 415
 Phe Thr Thr Gly Ala Thr Leu Gly Ala Ala Phe Ala Ala Glu Glu Ile
 420 425 430

20110907_CL5178USNA_SeqList.txt

Asp Pro Lys Lys Arg Val Ile Leu Phe Ile Gly Asp Gly Ser Leu Gln
435 440 445

Leu Thr Val Gln Glu Ile Ser Thr Met Ile Arg Trp Gly Leu Lys Pro
450 455 460

Tyr Leu Phe Val Leu Asn Asn Asp Gly Tyr Thr Ile Glu Arg Leu Ile
465 470 475 480

His Gly Glu Thr Ala Gln Tyr Asn Cys Ile Gln Asn Trp Gln His Leu
485 490 495

Glu Leu Leu Pro Thr Phe Gly Ala Lys Asp Tyr Glu Ala Val Arg Val
500 505 510

Ser Thr Thr Gly Glu Trp Asn Lys Leu Thr Thr Asp Glu Lys Phe Gln
515 520 525

Asp Asn Thr Arg Ile Arg Leu Ile Glu Val Met Leu Pro Thr Met Asp
530 535 540

Ala Pro Ser Asn Leu Val Lys Gln Ala Gln Leu Thr Ala Ala Thr Asn
545 550 555 560

Ala Lys Asn

<210> 64
<211> 1716
<212> DNA
<213> Yarrowia lipolytica

<220>
<221> CDS
<222> (1)..(1716)

<400> 64	
atg agc gac tcc gaa ccc caa atg gtc gac ctg ggc gac tat ctc ttt	48
Met Ser Asp Ser Glu Pro Gln Met Val Asp Leu Gly Asp Tyr Leu Phe	
1 5 10 15	
gcc cga ttc aag cag cta ggc gtg gac tcc gtc ttt gga gtg ccc ggc	96
Ala Arg Phe Lys Gln Leu Gly Val Asp Ser Val Phe Gly Val Pro Gly	
20 25 30	
gac ttc aac ctc acc ctg ttg gac cac gtg tac aat gtc gac atg cgg	144
Asp Phe Asn Leu Thr Leu Leu Asp His Val Tyr Asn Val Asp Met Arg	
35 40 45	
tgg gtt ggg aac aca aac gag ctg aat gcc ggc tac tcg gcc gac ggc	192
Trp Val Gly Asn Thr Asn Glu Leu Asn Ala Gly Tyr Ser Ala Asp Gly	
50 55 60	
tac tcc cgg gtc aag cgg ctg gca tgt ctt gtc acc acc ttt ggc gtg	240
Tyr Ser Arg Val Lys Arg Leu Ala Cys Leu Val Thr Thr Phe Gly Val	
65 70 75 80	

20110907_CL5178USNA_SeqList.txt

gga Gly	gag Glu	ctg Leu	tct Ser	gcc Ala 85	gtg Val	gct Ala	gct Ala	gtg Val	gca Ala 90	ggc Gly	tcg Ser	tac Tyr	gcc Ala	gag Glu 95	cat His	288
gtg Val	ggc Gly	gtg Val	gtg Val 100	cat His	ggt Val	ggc Gly	ggt Val 105	ccc Pro	agc Ser	acc Thr	tct Ser	gct Ala 110	gag Glu	aac Asn	336	
aag Lys	cat His	ctg Leu 115	ctg Leu	ctg Leu	cac His	cac His	aca Thr 120	ctc Leu	ggt Gly	aac Asn	ggc Gly	gac Asp 125	ttc Phe	cgg Arg	gtc Val	384
ttt Phe	gcc Ala 130	cag Gln	atg Met	tcc Ser	aaa Lys	ctc Leu 135	atc Ile	tcc Ser	gag Glu	tac Tyr	acc Thr 140	cac His	cat His	att Ile	gag Glu	432
gac Asp 145	ccc Pro	agc Ser	gag Glu	gct Ala 150	gcc Ala	gac Asp	gta Val	atc Ile	gac Asp 155	acc Thr 155	gcc Ala	atc Ile	cga Arg	atc Ile	gcc Ala 160	480
tac Tyr	acc Thr	cac His	cag Gln	cgg Arg 165	ccc Pro	ggt Val	tac Tyr	att Ile	gct Ala 170	gtg Val	ccc Pro	tcc Ser	aac Asn	ttc Phe 175	tcc Ser	528
gag Glu	gtc Val	gat Asp	att Ile 180	gcc Ala	gac Asp	cag Gln	gct Ala	aga Arg 185	ctg Leu	gat Asp	acc Thr	ccc Pro	ctg Leu 190	gac Asp	ctt Leu	576
tcg Ser	ctg Leu	cag Gln 195	ccc Pro	aac Asn	gac Asp	ccc Pro	gag Glu 200	agc Ser	cag Gln	tac Tyr	gag Glu 205	gtg Val 205	att Ile	gag Glu	gag Glu	624
att Ile	tgc Cys 210	tcg Ser	cgt Arg	atc Ile	aag Lys	gcc Ala 215	gcc Ala	aag Lys	aag Lys	ccc Pro	gtg Val 220	att Ile	ctc Leu	gtc Val	gac Asp	672
gcc Ala 225	tgc Cys	gct Ala	tcg Ser	cga Arg	tac Tyr 230	aga Arg	tgt Cys	gtg Val	gac Asp	gag Glu 235	acc Thr	aag Lys	gag Glu	ctg Leu	gcc Ala 240	720
aag Lys	atc Ile	acc Thr	aac Asn	ttt Phe 245	gcc Ala	tac Tyr	ttt Phe	gtc Val	act Thr 250	ccc Pro	atg Met	ggt Gly	aag Lys	ggt Gly 255	tct Ser	768
gtg Val	gac Asp	gag Glu	gat Asp 260	act Thr	gac Asp	cgg Arg	tac Tyr	gga Gly 265	gga Gly	aca Thr	tac Tyr	gtc Val	gga Gly 270	tcg Ser	ctg Leu	816
act Thr	gct Ala	cct Pro 275	gct Ala	act Thr	gcc Ala	gag Glu	gtg Val 280	ggt Val	gag Glu	aca Thr	gct Ala	gat Asp 285	ctc Leu	atc Ile	atc Ile	864
tcc Ser	gta Val 290	gga Gly	gct Ala	ctt Leu	ctg Leu	tcg Ser 295	gac Asp	ttc Phe	aac Asn	acc Thr	ggt Gly 300	tcc Ser	ttc Phe	tcg Ser	tac Tyr	912
tcc Ser 305	tac Tyr	tcc Ser	acc Thr	aaa Lys	aac Asn 310	gtg Val	gtg Val	gaa Glu	ttg Leu	cat His 315	tcg Ser	gac Asp	cac His	gtc Val	aaa Lys 320	960
atc Ile	aag Lys	tcc Ser	gcc Ala	acc Thr 325	tac Tyr	aac Asn	aac Asn	gtc Val	ggc Gly 330	atg Met	aaa Lys	atg Met	ctg Leu	ttc Phe 335	ccg Pro	1008
ccc Pro	ctg Leu	ctc Leu	gaa Glu 340	gcc Ala	gtc Val	aag Lys	aaa Lys	ctg Leu 345	ggt Val	gcc Ala	gag Glu	acc Thr	cct Pro 350	gac Asp	ttt Phe	1056

20110907_CL5178USNA_SeqList.txt

gca tcc aag gct ctg gct gtt ccc gac acc act ccc aag atc ccc gag	1104
Ala Ser Lys Ala Leu Ala Val Pro Asp Thr Thr Pro Lys Ile Pro Glu	
355 360 365	
gta ccc gat gat cac att acg acc cag gca tgg ctg tgg cag cgt ctc	1152
Val Pro Asp Asp His Ile Thr Thr Gln Ala Trp Leu Trp Gln Arg Leu	
370 375 380	
agt tac ttt ctg agg ccc acc gac atc gtg gtc acc gag acc gga acc	1200
Ser Tyr Phe Leu Arg Pro Thr Asp Ile Val Val Thr Glu Thr Gly Thr	
385 390 395 400	
tcg tcc ttt gga atc atc cag acc aag ttc ccc cac aac gtc cga ggt	1248
Ser Ser Phe Gly Ile Ile Gln Thr Lys Phe Pro His Asn Val Arg Gly	
405 410 415	
atc tcg cag gtg ctg tgg ggc tct att gga tac tcg gtg gga gca gcc	1296
Ile Ser Gln Val Leu Trp Gly Ser Ile Gly Tyr Ser Val Gly Ala Ala	
420 425 430	
tgt gga gcc tcc att gct gca cag gag att gac ccc cag cag cga gtg	1344
Cys Gly Ala Ser Ile Ala Ala Gln Glu Ile Asp Pro Gln Gln Arg Val	
435 440 445	
att ctg ttt gtg ggc gac ggc tct ctt cag ctg acg gtg acc gag atc	1392
Ile Leu Phe Val Gly Asp Gly Ser Leu Gln Leu Thr Val Thr Glu Ile	
450 455 460	
tcg tgc atg atc cgc aac aac gtc aag ccg tac att ttt gtg ctc aac	1440
Ser Cys Met Ile Arg Asn Asn Val Lys Pro Tyr Ile Phe Val Leu Asn	
465 470 475 480	
aac gac ggc tac acc atc gag agg ctc att cac ggc gaa aac gcc tcg	1488
Asn Asp Gly Tyr Thr Ile Glu Arg Leu Ile His Gly Glu Asn Ala Ser	
485 490 495	
tac aac gat gtg cac atg tgg aag tac tcc aag att ctc gac acg ttc	1536
Tyr Asn Asp Val His Met Trp Lys Tyr Ser Lys Ile Leu Asp Thr Phe	
500 505 510	
aac gcc aag gcc cac gag tcg att gtg gtc aac acc aag ggc gag atg	1584
Asn Ala Lys Ala His Glu Ser Ile Val Val Asn Thr Lys Gly Glu Met	
515 520 525	
gac gct ctg ttc gac aac gaa gag ttt gcc aag ccc gac aag atc cgg	1632
Asp Ala Leu Phe Asp Asn Glu Glu Phe Ala Lys Pro Asp Lys Ile Arg	
530 535 540	
ctc att gag gtc atg tgc gac aag atg gac gcg cct gcc tcg ttg atc	1680
Leu Ile Glu Val Met Cys Asp Lys Met Asp Ala Pro Ala Ser Leu Ile	
545 550 555 560	
aag cag gct gag ctc tct gcc aag acc aac gtt tag	1716
Lys Gln Ala Glu Leu Ser Ala Lys Thr Asn Val	
565 570	

<210> 65

<211> 571

<212> PRT

<213> Yarrowia lipolytica

<400> 65

Met Ser Asp Ser Glu Pro Gln Met Val Asp Leu Gly Asp Tyr Leu Phe
1 5 10 15

20110907_CL5178USNA_SeqList.txt

Ala Arg Phe Lys Gln Leu Gly Val Asp Ser Val Phe Gly Val Pro Gly
20 25 30

Asp Phe Asn Leu Thr Leu Leu Asp His Val Tyr Asn Val Asp Met Arg
35 40 45

Trp Val Gly Asn Thr Asn Glu Leu Asn Ala Gly Tyr Ser Ala Asp Gly
50 55 60

Tyr Ser Arg Val Lys Arg Leu Ala Cys Leu Val Thr Thr Phe Gly Val
65 70 75 80

Gly Glu Leu Ser Ala Val Ala Ala Val Ala Gly Ser Tyr Ala Glu His
85 90 95

Val Gly Val Val His Val Val Gly Val Pro Ser Thr Ser Ala Glu Asn
100 105 110

Lys His Leu Leu Leu His His Thr Leu Gly Asn Gly Asp Phe Arg Val
115 120 125

Phe Ala Gln Met Ser Lys Leu Ile Ser Glu Tyr Thr His His Ile Glu
130 135 140

Asp Pro Ser Glu Ala Ala Asp Val Ile Asp Thr Ala Ile Arg Ile Ala
145 150 155 160

Tyr Thr His Gln Arg Pro Val Tyr Ile Ala Val Pro Ser Asn Phe Ser
165 170 175

Glu Val Asp Ile Ala Asp Gln Ala Arg Leu Asp Thr Pro Leu Asp Leu
180 185 190

Ser Leu Gln Pro Asn Asp Pro Glu Ser Gln Tyr Glu Val Ile Glu Glu
195 200 205

Ile Cys Ser Arg Ile Lys Ala Ala Lys Lys Pro Val Ile Leu Val Asp
210 215 220

Ala Cys Ala Ser Arg Tyr Arg Cys Val Asp Glu Thr Lys Glu Leu Ala
225 230 235 240

Lys Ile Thr Asn Phe Ala Tyr Phe Val Thr Pro Met Gly Lys Gly Ser
245 250 255

Val Asp Glu Asp Thr Asp Arg Tyr Gly Gly Thr Tyr Val Gly Ser Leu
260 265 270

Thr Ala Pro Ala Thr Ala Glu Val Val Glu Thr Ala Asp Leu Ile Ile
275 280 285

20110907_CL5178USNA_SeqList.txt

Ser Val Gly Ala Leu Leu Ser Asp Phe Asn Thr Gly Ser Phe Ser Tyr
290 295 300

Ser Tyr Ser Thr Lys Asn Val Val Glu Leu His Ser Asp His Val Lys
305 310 315 320

Ile Lys Ser Ala Thr Tyr Asn Asn Val Gly Met Lys Met Leu Phe Pro
325 330 335

Pro Leu Leu Glu Ala Val Lys Lys Leu Val Ala Glu Thr Pro Asp Phe
340 345 350

Ala Ser Lys Ala Leu Ala Val Pro Asp Thr Thr Pro Lys Ile Pro Glu
355 360 365

Val Pro Asp Asp His Ile Thr Thr Gln Ala Trp Leu Trp Gln Arg Leu
370 375 380

Ser Tyr Phe Leu Arg Pro Thr Asp Ile Val Val Thr Glu Thr Gly Thr
385 390 395 400

Ser Ser Phe Gly Ile Ile Gln Thr Lys Phe Pro His Asn Val Arg Gly
405 410 415

Ile Ser Gln Val Leu Trp Gly Ser Ile Gly Tyr Ser Val Gly Ala Ala
420 425 430

Cys Gly Ala Ser Ile Ala Ala Gln Glu Ile Asp Pro Gln Gln Arg Val
435 440 445

Ile Leu Phe Val Gly Asp Gly Ser Leu Gln Leu Thr Val Thr Glu Ile
450 455 460

Ser Cys Met Ile Arg Asn Asn Val Lys Pro Tyr Ile Phe Val Leu Asn
465 470 475 480

Asn Asp Gly Tyr Thr Ile Glu Arg Leu Ile His Gly Glu Asn Ala Ser
485 490 495

Tyr Asn Asp Val His Met Trp Lys Tyr Ser Lys Ile Leu Asp Thr Phe
500 505 510

Asn Ala Lys Ala His Glu Ser Ile Val Val Asn Thr Lys Gly Glu Met
515 520 525

Asp Ala Leu Phe Asp Asn Glu Glu Phe Ala Lys Pro Asp Lys Ile Arg
530 535 540

Leu Ile Glu Val Met Cys Asp Lys Met Asp Ala Pro Ala Ser Leu Ile
545 550 555 560

Lys Gln Ala Glu Leu Ser Ala Lys Thr Asn Val
565 570

<210> 66
<211> 1716
<212> DNA
<213> Schizosaccharomyces pombe

<220>
<221> CDS
<222> (1)..(1716)

<400> 66
atg agt ggg gat att tta gtc ggt gaa tat cta ttc aaa agg ctt gaa 48
Met Ser Gly Asp Ile Leu Val Gly Glu Tyr Leu Phe Lys Arg Leu Glu
1 5 10 15
caa tta ggg gtc aag tcc att ctt ggt gtt cca gga gat ttc aat tta 96
Gln Leu Gly Val Lys Ser Ile Leu Gly Val Pro Gly Asp Phe Asn Leu
20 25 30
gct cta ctt gac tta att gag aaa gtt gga gat gag aaa ttt cgt tgg 144
Ala Leu Leu Asp Leu Ile Glu Lys Val Gly Asp Glu Lys Phe Arg Trp
35 40 45
gtt ggc aat acc aat gag ttg aat ggt gct tat gcc gct gat ggt tat 192
Val Gly Asn Thr Asn Glu Leu Asn Gly Ala Tyr Ala Ala Asp Gly Tyr
50 55 60
gct cgt gtt aat ggt ctt tca gcc att gtt aca acg ttc ggc gtg gga 240
Ala Arg Val Asn Gly Leu Ser Ala Ile Val Thr Thr Phe Gly Val Gly
65 70 75 80
gag ctt tcc gct att aat gga gtg gca ggt tct tat gcg gag cat gtc 288
Glu Leu Ser Ala Ile Asn Gly Val Ala Gly Ser Tyr Ala Glu His Val
85 90 95
cca gta gtt cat att gtt gga atg cct tcc aca aag gtg caa gat act 336
Pro Val Val His Ile Val Gly Met Pro Ser Thr Lys Val Gln Asp Thr
100 105 110
gga gct ttg ctt cat cat act tta gga gat gga gac ttt cgc act ttc 384
Gly Ala Leu Leu His His Thr Leu Gly Asp Gly Asp Phe Arg Thr Phe
115 120 125
atg gat atg ttt aag aaa gtt tct gcc tac agt ata atg atc gat aac 432
Met Asp Met Phe Lys Lys Val Ser Ala Tyr Ser Ile Met Ile Asp Asn
130 135 140
gga aac gat gca gct gaa aag atc gat gaa gcc ttg tcg att tgt tat 480
Gly Asn Asp Ala Ala Glu Lys Ile Asp Glu Ala Leu Ser Ile Cys Tyr
145 150 155 160
aaa aag gct agg cct gtt tac att ggt att cct tct gat gct ggc tac 528
Lys Lys Ala Arg Pro Val Tyr Ile Gly Ile Pro Ser Asp Ala Gly Tyr
165 170 175
ttc aaa gca tct tca tca aat ctt ggg aaa aga cta aag ctc gag gag 576
Phe Lys Ala Ser Ser Ser Asn Leu Gly Lys Arg Leu Lys Leu Glu Glu
180 185 190
gat act aac gat cca gca gtt gag caa gaa gtc atc aat cat atc tcg 624
Asp Thr Asn Asp Pro Ala Val Glu Gln Glu Val Ile Asn His Ile Ser
195 200 205
gaa atg gtt gtc aat gca aag aaa cca gtg att tta att gac gct tgt 672

20110907_CL5178USNA_SeqList.txt

Glu	Met	Val	Val	Asn	Ala	Lys	Lys	Pro	Val	Ile	Leu	Ile	Asp	Ala	Cys		
210						215					220						
gct	gta	aga	cat	cgt	gtc	ggt	cca	gaa	gta	cat	gag	ctg	att	aaa	ttg	720	
Ala	Val	Arg	His	Arg	Val	Val	Pro	Glu	Val	His	Glu	Leu	Ile	Lys	Leu	240	
225					230					235							
acc	cat	ttc	cct	aca	tat	gta	act	ccc	atg	ggg	aaa	tct	gca	att	gac	768	
Thr	His	Phe	Pro	Thr	Tyr	Val	Thr	Pro	Met	Gly	Lys	Ser	Ala	Ile	Asp	255	
				245					250								
gaa	act	tcg	caa	ttt	ttt	gac	ggc	ggt	tat	ggt	ggg	tca	att	tca	gat	816	
Glu	Thr	Ser	Gln	Phe	Phe	Asp	Gly	Val	Tyr	Val	Gly	Ser	Ile	Ser	Asp	270	
			260					265									
cct	gaa	ggt	aaa	gac	aga	att	gaa	tcc	act	gat	ctg	ttg	cta	tcc	atc	864	
Pro	Glu	Val	Lys	Asp	Arg	Ile	Glu	Ser	Thr	Asp	Leu	Leu	Leu	Ser	Ile	285	
		275					280										
ggg	gct	ctc	aaa	tca	gac	ttt	aac	acg	ggg	tcc	ttc	tct	tac	cac	ctc	912	
Gly	Ala	Leu	Lys	Ser	Asp	Phe	Asn	Thr	Gly	Ser	Phe	Ser	Tyr	His	Leu	300	
	290					295					300						
agc	caa	aag	aat	gcc	ggt	gag	ttt	cat	tca	gac	cac	atg	cgc	att	cga	960	
Ser	Gln	Lys	Asn	Ala	Val	Glu	Phe	His	Ser	Asp	His	Met	Arg	Ile	Arg	320	
305				310						315							
tat	gct	ctt	tat	cca	aat	gta	gcc	atg	aag	tat	att	ctt	cgc	aaa	ctg	1008	
Tyr	Ala	Leu	Tyr	Pro	Asn	Val	Ala	Met	Lys	Tyr	Ile	Leu	Arg	Lys	Leu	335	
				325					330								
ttg	aaa	gta	ctt	gat	gct	tct	atg	tgt	cat	tcc	aag	gct	gct	cct	acc	1056	
Leu	Lys	Val	Leu	Asp	Ala	Ser	Met	Cys	His	Ser	Lys	Ala	Ala	Pro	Thr	345	
			340					345					350				
att	ggc	tac	aac	atc	aag	cct	aag	cat	gcg	gaa	gga	tat	tct	tcc	aac	1104	
Ile	Gly	Tyr	Asn	Ile	Lys	Pro	Lys	His	Ala	Glu	Gly	Tyr	Ser	Ser	Asn	365	
		355					360					365					
gag	att	act	cat	tgc	tgg	ttt	tgg	cct	aaa	ttt	agt	gaa	ttt	ttg	aag	1152	
Glu	Ile	Thr	His	Cys	Trp	Phe	Trp	Pro	Lys	Phe	Ser	Glu	Phe	Leu	Lys	380	
	370					375					380						
ccc	cga	gat	ggt	ttg	atc	acc	gag	act	gga	act	gca	aac	ttt	ggg	gtc	1200	
Pro	Arg	Asp	Val	Leu	Ile	Thr	Glu	Thr	Gly	Thr	Ala	Asn	Phe	Gly	Val	400	
385					390					395							
ctt	gat	tgc	agg	ttt	cca	aag	gat	gta	aca	gcc	att	tcc	cag	gta	tta	1248	
Leu	Asp	Cys	Arg	Phe	Pro	Lys	Asp	Val	Thr	Ala	Ile	Ser	Gln	Val	Leu	410	
				405					410					415			
tgg	gga	tct	att	gga	tac	tcc	ggt	ggt	gca	atg	ttt	ggg	gct	ggt	ttg	1296	
Trp	Gly	Ser	Ile	Gly	Tyr	Ser	Val	Gly	Ala	Met	Phe	Gly	Ala	Val	Leu	425	
			420				425						430				
gcc	gtc	cac	gat	tct	aaa	gag	ccc	gat	cgt	cgt	acc	att	ctt	gta	gta	1344	
Ala	Val	His	Asp	Ser	Lys	Glu	Pro	Asp	Arg	Arg	Thr	Ile	Leu	Val	Val	440	
		435					440					445					
ggg	gat	gga	tcc	tta	caa	ctg	acg	att	aca	gag	att	tca	acc	tgc	att	1392	
Gly	Asp	Gly	Ser	Leu	Gln	Leu	Thr	Ile	Thr	Glu	Ile	Ser	Thr	Cys	Ile	450	
	450					455					460						
cgc	cat	aac	ctc	aaa	cca	att	att	ttc	ata	att	aac	aac	gac	ggg	tac	1440	
Arg	His	Asn	Leu	Lys	Pro	Ile	Ile	Phe	Ile	Ile	Asn	Asn	Asp	Gly	Tyr	470	
465					470					475					480		
acc	att	gag	cgt	tta	att	cat	ggg	ttg	cat	gct	agc	tat	aac	gaa	att	1488	

20110907_CL5178USNA_SeqList.txt

Thr	Ile	Glu	Arg	Leu	Ile	His	Gly	Leu	His	Ala	Ser	Tyr	Asn	Glu	Ile		
				485					490					495			
aac	act	aaa	tgg	ggc	tac	caa	cag	att	ccc	aag	ttt	ttc	gga	gct	gct		1536
Asn	Thr	Lys	Trp	Gly	Tyr	Gln	Gln	Ile	Pro	Lys	Phe	Phe	Gly	Ala	Ala		
			500					505					510				
gaa	aac	cac	ttc	cgc	act	tac	tgt	ggt	aaa	act	cct	act	gac	ggt	gaa		1584
Glu	Asn	His	Phe	Arg	Thr	Tyr	Cys	Val	Lys	Thr	Pro	Thr	Asp	Val	Glu		
		515					520					525					
aag	ttg	ttt	agc	gac	aag	gag	ttt	gca	aat	gca	gat	gtc	att	caa	gta		1632
Lys	Leu	Phe	Ser	Asp	Lys	Glu	Phe	Ala	Asn	Ala	Asp	Val	Ile	Gln	Val		
	530					535					540						
gtt	gag	ctt	gta	atg	cct	atg	ttg	gat	gca	cct	cgt	gtc	cta	gtt	gag		1680
Val	Glu	Leu	Val	Met	Pro	Met	Leu	Asp	Ala	Pro	Arg	Val	Leu	Val	Glu		
	545				550					555					560		
caa	gcc	aag	ttg	acg	tct	aag	atc	aat	aag	caa	tga						1716
Gln	Ala	Lys	Leu	Thr	Ser	Lys	Ile	Asn	Lys	Gln							
				565					570								

<210> 67
 <211> 571
 <212> PRT
 <213> Schizosaccharomyces pombe

<400> 67

Met	Ser	Gly	Asp	Ile	Leu	Val	Gly	Glu	Tyr	Leu	Phe	Lys	Arg	Leu	Glu		
1				5					10					15			

Gln	Leu	Gly	Val	Lys	Ser	Ile	Leu	Gly	Val	Pro	Gly	Asp	Phe	Asn	Leu		
			20					25					30				

Ala	Leu	Leu	Asp	Leu	Ile	Glu	Lys	Val	Gly	Asp	Glu	Lys	Phe	Arg	Trp		
		35					40					45					

Val	Gly	Asn	Thr	Asn	Glu	Leu	Asn	Gly	Ala	Tyr	Ala	Ala	Asp	Gly	Tyr		
	50					55					60						

Ala	Arg	Val	Asn	Gly	Leu	Ser	Ala	Ile	Val	Thr	Thr	Phe	Gly	Val	Gly		
65					70					75					80		

Glu	Leu	Ser	Ala	Ile	Asn	Gly	Val	Ala	Gly	Ser	Tyr	Ala	Glu	His	Val		
				85					90					95			

Pro	Val	Val	His	Ile	Val	Gly	Met	Pro	Ser	Thr	Lys	Val	Gln	Asp	Thr		
			100					105					110				

Gly	Ala	Leu	Leu	His	His	Thr	Leu	Gly	Asp	Gly	Asp	Phe	Arg	Thr	Phe		
		115					120					125					

Met	Asp	Met	Phe	Lys	Lys	Val	Ser	Ala	Tyr	Ser	Ile	Met	Ile	Asp	Asn		
	130					135					140						

Gly	Asn	Asp	Ala	Ala	Glu	Lys	Ile	Asp	Glu	Ala	Leu	Ser	Ile	Cys	Tyr		

145 150 155 160
 Lys Lys Ala Arg Pro Val Tyr Ile Gly Ile Pro Ser Asp Ala Gly Tyr
 165 170 175
 Phe Lys Ala Ser Ser Ser Asn Leu Gly Lys Arg Leu Lys Leu Glu Glu
 180 185 190
 Asp Thr Asn Asp Pro Ala Val Glu Gln Glu Val Ile Asn His Ile Ser
 195 200 205
 Glu Met Val Val Asn Ala Lys Lys Pro Val Ile Leu Ile Asp Ala Cys
 210 215 220
 Ala Val Arg His Arg Val Val Pro Glu Val His Glu Leu Ile Lys Leu
 225 230 235 240
 Thr His Phe Pro Thr Tyr Val Thr Pro Met Gly Lys Ser Ala Ile Asp
 245 250 255
 Glu Thr Ser Gln Phe Phe Asp Gly Val Tyr Val Gly Ser Ile Ser Asp
 260 265 270
 Pro Glu Val Lys Asp Arg Ile Glu Ser Thr Asp Leu Leu Leu Ser Ile
 275 280 285
 Gly Ala Leu Lys Ser Asp Phe Asn Thr Gly Ser Phe Ser Tyr His Leu
 290 295 300
 Ser Gln Lys Asn Ala Val Glu Phe His Ser Asp His Met Arg Ile Arg
 305 310 315 320
 Tyr Ala Leu Tyr Pro Asn Val Ala Met Lys Tyr Ile Leu Arg Lys Leu
 325 330 335
 Leu Lys Val Leu Asp Ala Ser Met Cys His Ser Lys Ala Ala Pro Thr
 340 345 350
 Ile Gly Tyr Asn Ile Lys Pro Lys His Ala Glu Gly Tyr Ser Ser Asn
 355 360 365
 Glu Ile Thr His Cys Trp Phe Trp Pro Lys Phe Ser Glu Phe Leu Lys
 370 375 380
 Pro Arg Asp Val Leu Ile Thr Glu Thr Gly Thr Ala Asn Phe Gly Val
 385 390 395 400
 Leu Asp Cys Arg Phe Pro Lys Asp Val Thr Ala Ile Ser Gln Val Leu
 405 410 415
 Trp Gly Ser Ile Gly Tyr Ser Val Gly Ala Met Phe Gly Ala Val Leu
 Page 110

420

425

430

Ala Val His Asp Ser Lys Glu Pro Asp Arg Arg Thr Ile Leu Val Val
 435 440 445

Gly Asp Gly Ser Leu Gln Leu Thr Ile Thr Glu Ile Ser Thr Cys Ile
 450 455 460

Arg His Asn Leu Lys Pro Ile Ile Phe Ile Ile Asn Asn Asp Gly Tyr
 465 470 475 480

Thr Ile Glu Arg Leu Ile His Gly Leu His Ala Ser Tyr Asn Glu Ile
 485 490 495

Asn Thr Lys Trp Gly Tyr Gln Gln Ile Pro Lys Phe Phe Gly Ala Ala
 500 505 510

Glu Asn His Phe Arg Thr Tyr Cys Val Lys Thr Pro Thr Asp Val Glu
 515 520 525

Lys Leu Phe Ser Asp Lys Glu Phe Ala Asn Ala Asp Val Ile Gln Val
 530 535 540

Val Glu Leu Val Met Pro Met Leu Asp Ala Pro Arg Val Leu Val Glu
 545 550 555 560

Gln Ala Lys Leu Thr Ser Lys Ile Asn Lys Gln
 565 570

<210> 68
 <211> 1176
 <212> DNA
 <213> Saccharomyces cerevisiae

<220>
 <221> CDS
 <222> (1)..(1176)

<400> 68	
atg tct gct gct gct gat aga tta aac tta act tcc ggc cac ttg aat	48
Met Ser Ala Ala Ala Asp Arg Leu Asn Leu Thr Ser Gly His Leu Asn	
1 5 10 15	
gct ggt aga aag aga agt tcc tct tct gtt tct ttg aag gct gcc gaa	96
Ala Gly Arg Lys Arg Ser Ser Ser Val Ser Leu Lys Ala Ala Glu	
20 25 30	
aag cct ttc aag gtt act gtg att gga tct ggt aac tgg ggt act act	144
Lys Pro Phe Lys Val Thr Val Ile Gly Ser Gly Asn Trp Gly Thr Thr	
35 40 45	
att gcc aag gtg gtt gcc gaa aat tgt aag gga tac cca gaa gtt ttc	192
Ile Ala Lys Val Val Ala Glu Asn Cys Lys Gly Tyr Pro Glu Val Phe	
50 55 60	
gct cca ata gta caa atg tgg gtg ttc gaa gaa gag atc aat ggt gaa	240
Ala Pro Ile Val Gln Met Trp Val Phe Glu Glu Glu Ile Asn Gly Glu	

20110907_CL5178USNA_SeqList.txt

65											70											75											80	
aaa	ttg	act	gaa	atc	ata	aat	act	aga	cat	caa	aac	gtg	aaa	tac	ttg		288																	
Lys	Leu	Thr	Glu	Ile	Ile	Asn	Thr	Arg	His	Gln	Asn	Val	Lys	Tyr	Leu																			
				85					90					95																				
cct	ggc	atc	act	cta	ccc	gac	aat	ttg	ggt	gct	aat	cca	gac	ttg	att		336																	
Pro	Gly	Ile	Thr	Leu	Pro	Asp	Asn	Leu	Val	Ala	Asn	Pro	Asp	Leu	Ile																			
			100					105					110																					
gat	tca	gtc	aag	gat	gtc	gac	atc	atc	ggt	ttc	aac	att	cca	cat	caa		384																	
Asp	Ser	Val	Lys	Asp	Val	Asp	Ile	Ile	Val	Phe	Asn	Ile	Pro	His	Gln																			
		115					120					125																						
ttt	ttg	ccc	cgt	atc	tgt	agc	caa	ttg	aaa	ggg	cat	ggt	gat	tca	cac		432																	
Phe	Leu	Pro	Arg	Ile	Cys	Ser	Gln	Leu	Lys	Gly	His	Val	Asp	Ser	His																			
	130					135					140																							
gtc	aga	gct	atc	tcc	tgt	cta	aag	ggg	ttt	gaa	ggt	ggt	gct	aaa	ggg		480																	
Val	Arg	Ala	Ile	Ser	Cys	Leu	Lys	Gly	Phe	Glu	Val	Gly	Ala	Lys	Gly																			
					150					155					160																			
gtc	caa	ttg	cta	tcc	tct	tac	atc	act	gag	gaa	cta	ggg	att	caa	tgt		528																	
Val	Gln	Leu	Leu	Ser	Ser	Tyr	Ile	Thr	Glu	Glu	Leu	Gly	Ile	Gln	Cys																			
				165					170					175																				
ggg	gct	cta	tct	ggg	gct	aac	att	gcc	acc	gaa	gtc	gct	caa	gaa	cac		576																	
Gly	Ala	Leu	Ser	Gly	Ala	Asn	Ile	Ala	Thr	Glu	Val	Ala	Gln	Glu	His																			
			180					185					190																					
tgg	tct	gaa	aca	aca	ggt	gct	tac	cac	att	cca	aag	gat	ttc	aga	ggc		624																	
Trp	Ser	Glu	Thr	Thr	Val	Ala	Tyr	His	Ile	Pro	Lys	Asp	Phe	Arg	Gly																			
		195					200					205																						
gag	ggc	aag	gac	gtc	gac	cat	aag	ggt	cta	aag	gcc	ttg	ttc	cac	aga		672																	
Glu	Gly	Lys	Asp	Val	Asp	His	Lys	Val	Leu	Lys	Ala	Leu	Phe	His	Arg																			
	210					215					220																							
cct	tac	ttc	cac	ggt	agt	gtc	atc	gaa	gat	ggt	gct	ggg	atc	tcc	atc		720																	
Pro	Tyr	Phe	His	Val	Ser	Val	Ile	Glu	Asp	Val	Ala	Gly	Ile	Ser	Ile																			
					230					235					240																			
tgt	ggg	gct	ttg	aag	aac	ggt	ggt	gcc	tta	ggg	tgt	ggg	ttc	gtc	gaa		768																	
Cys	Gly	Ala	Leu	Lys	Asn	Val	Val	Ala	Leu	Gly	Cys	Gly	Phe	Val	Glu																			
				245					250					255																				
ggg	cta	ggc	tgg	ggg	aac	aac	gct	tct	gct	gcc	atc	caa	aga	gtc	ggg		816																	
Gly	Leu	Gly	Trp	Gly	Asn	Asn	Ala	Ser	Ala	Ala	Ile	Gln	Arg	Val	Gly																			
			260					265					270																					
ttg	ggg	gag	atc	atc	aga	ttc	ggg	caa	atg	ttt	ttc	cca	gaa	tct	aga		864																	
Leu	Gly	Glu	Ile	Ile	Arg	Phe	Gly	Gln	Met	Phe	Phe	Pro	Glu	Ser	Arg																			
		275					280					285																						
gaa	gaa	aca	tac	tac	caa	gag	tct	gct	ggg	ggt	gct	gat	ttg	atc	acc		912																	
Glu	Glu	Thr	Tyr	Tyr	Gln	Glu	Ser	Ala	Gly	Val	Ala	Asp	Leu	Ile	Thr																			
	290					295					300																							
acc	tgc	gct	ggg	ggg	aga	aac	gtc	aag	ggt	gct	agg	cta	atg	gct	act		960																	
Thr	Cys	Ala	Gly	Gly	Arg	Asn	Val	Lys	Val	Ala	Arg	Leu	Met	Ala	Thr																			
					310					315					320																			
tct	ggg	aag	gac	gcc	tgg	gaa	tgt	gaa	aag	gag	ttg	ttg	aat	ggc	caa		1008																	
Ser	Gly	Lys	Asp	Ala	Trp	Glu	Cys	Glu	Lys	Glu	Leu	Leu	Asn	Gly	Gln																			
				325					330					335																				
tcc	gct	caa	ggg	tta	att	acc	tgc	aaa	gaa	ggt	cac	gaa	tgg	ttg	gaa		1056																	
Ser	Ala	Gln	Gly	Leu	Ile	Thr	Cys	Lys	Glu	Val	His	Glu	Trp	Leu	Glu																			

20110907_CL5178USNA_SeqList.txt

	340	345	350	
aca tgt ggc tct gtc gaa gac ttc cca tta ttt gaa gcc gta tac caa				1104
Thr Cys Gly Ser Val Glu Asp Phe Pro Leu Phe Glu Ala Val Tyr Gln				
	355	360	365	
atc gtt tac aac aac tac cca atg aag aac ctg ccg gac atg att gaa				1152
Ile Val Tyr Asn Asn Tyr Pro Met Lys Asn Leu Pro Asp Met Ile Glu				
	370	375	380	
gaa tta gat cta cat gaa gat tag				1176
Glu Leu Asp Leu His Glu Asp				
	385	390		

<210> 69
 <211> 391
 <212> PRT
 <213> Saccharomyces cerevisiae

<400> 69

Met Ser Ala Ala Ala Asp Arg Leu Asn Leu Thr Ser Gly His Leu Asn
1 5 10 15

Ala Gly Arg Lys Arg Ser Ser Ser Ser Val Ser Leu Lys Ala Ala Glu
20 25 30

Lys Pro Phe Lys Val Thr Val Ile Gly Ser Gly Asn Trp Gly Thr Thr
35 40 45

Ile Ala Lys Val Val Ala Glu Asn Cys Lys Gly Tyr Pro Glu Val Phe
50 55 60

Ala Pro Ile Val Gln Met Trp Val Phe Glu Glu Glu Ile Asn Gly Glu
65 70 75 80

Lys Leu Thr Glu Ile Ile Asn Thr Arg His Gln Asn Val Lys Tyr Leu
85 90 95

Pro Gly Ile Thr Leu Pro Asp Asn Leu Val Ala Asn Pro Asp Leu Ile
100 105 110

Asp Ser Val Lys Asp Val Asp Ile Ile Val Phe Asn Ile Pro His Gln
115 120 125

Phe Leu Pro Arg Ile Cys Ser Gln Leu Lys Gly His Val Asp Ser His
130 135 140

Val Arg Ala Ile Ser Cys Leu Lys Gly Phe Glu Val Gly Ala Lys Gly
145 150 155 160

Val Gln Leu Leu Ser Ser Tyr Ile Thr Glu Glu Leu Gly Ile Gln Cys
165 170 175

Gly Ala Leu Ser Gly Ala Asn Ile Ala Thr Glu Val Ala Gln Glu His
180 185 190

20110907_CL5178USNA_SeqList.txt

Trp Ser Glu Thr Thr Val Ala Tyr His Ile Pro Lys Asp Phe Arg Gly
195 200 205

Glu Gly Lys Asp Val Asp His Lys Val Leu Lys Ala Leu Phe His Arg
210 215 220

Pro Tyr Phe His Val Ser Val Ile Glu Asp Val Ala Gly Ile Ser Ile
225 230 235 240

Cys Gly Ala Leu Lys Asn Val Val Ala Leu Gly Cys Gly Phe Val Glu
245 250 255

Gly Leu Gly Trp Gly Asn Asn Ala Ser Ala Ala Ile Gln Arg Val Gly
260 265 270

Leu Gly Glu Ile Ile Arg Phe Gly Gln Met Phe Phe Pro Glu Ser Arg
275 280 285

Glu Glu Thr Tyr Tyr Gln Glu Ser Ala Gly Val Ala Asp Leu Ile Thr
290 295 300

Thr Cys Ala Gly Gly Arg Asn Val Lys Val Ala Arg Leu Met Ala Thr
305 310 315 320

Ser Gly Lys Asp Ala Trp Glu Cys Glu Lys Glu Leu Leu Asn Gly Gln
325 330 335

Ser Ala Gln Gly Leu Ile Thr Cys Lys Glu Val His Glu Trp Leu Glu
340 345 350

Thr Cys Gly Ser Val Glu Asp Phe Pro Leu Phe Glu Ala Val Tyr Gln
355 360 365

Ile Val Tyr Asn Asn Tyr Pro Met Lys Asn Leu Pro Asp Met Ile Glu
370 375 380

Glu Leu Asp Leu His Glu Asp
385 390

<210> 70
<211> 1323
<212> DNA
<213> Saccharomyces cerevisiae

<220>
<221> CDS
<222> (1)..(1323)

<400> 70
atg ctt gct gtc aga aga tta aca aga tac aca ttc ctt aag cga acg
Met Leu Ala Val Arg Arg Leu Thr Arg Tyr Thr Phe Leu Lys Arg Thr
1 5 10 15

48

20110907_CL5178USNA_SeqList.txt

cat His	ccg Pro	gtg Val	tta Leu 20	tat Tyr	act Thr	cgt Arg	cgt Arg	gca Ala 25	tat Tyr	aaa Lys	att Ile	ttg Leu 30	cct Pro	tca Ser	aga Arg	96
tct Ser	act Thr	ttc Phe 35	cta Leu	aga Arg	aga Arg	tca Ser	tta Leu 40	tta Leu	caa Gln	aca Thr	caa Gln	ctg Leu 45	cac His	tca Ser	aag Lys	144
atg Met	act Thr 50	gct Ala	cat His	act Thr	aat Asn	atc Ile 55	aaa Lys	cag Gln	cac His	aaa Lys	cac His 60	tgt Cys	cat His	gag Glu	gac Asp	192
cat His 65	cct Pro	atc Ile	aga Arg	aga Arg	tcg Ser 70	gac Asp	tct Ser	gcc Ala	gtg Val	tca Ser 75	att Ile	gta Val	cat His	ttg Leu	aaa Lys 80	240
cgt Arg	gcg Ala	ccc Pro	ttc Phe	aag Lys 85	gtt Val	aca Thr	gtg Val	att Ile	ggg Gly 90	tct Ser	ggg Gly	aac Asn	tgg Trp	ggg Gly 95	acc Thr	288
acc Thr	atc Ile	gcc Ala	aaa Lys 100	gtc Val	att Ile	gcg Ala	gaa Glu	aac Asn 105	aca Thr	gaa Glu	ttg Leu	cat His	tcc Ser 110	cat His	atc Ile	336
ttc Phe	gag Glu	cca Pro 115	gag Glu	gtg Val	aga Arg	atg Met	tgg Trp 120	gtt Val	ttt Phe	gat Asp	gaa Glu	aag Lys 125	atc Ile	ggc Gly	gac Asp	384
gaa Glu	aat Asn 130	ctg Leu	acg Thr	gat Asp	atc Ile	ata Ile 135	aat Asn	aca Thr	aga Arg	cac His	cag Gln 140	aac Asn	gtt Val	aaa Lys	tat Tyr	432
cta Leu 145	ccc Pro	aat Asn	att Ile	gac Asp	ctg Leu 150	ccc Pro	cat His	aat Asn	cta Leu	gtg Val 155	gcc Ala	gat Asp	cct Pro	gat Asp	ctt Leu 160	480
tta Leu	cac His	tcc Ser	atc Ile	aag Lys 165	ggg Gly	gct Ala	gac Asp	atc Ile	ctt Leu 170	gtt Val	ttc Phe	aac Asn	atc Ile	cct Pro 175	cat His	528
caa Gln	ttt Phe	tta Leu 180	cca Pro	aac Asn	ata Ile	gtc Val	aaa Lys	caa Gln 185	ttg Leu	caa Gln	ggc Gly	cac His	gtg Val 190	gcc Ala	cct Pro	576
cat His	gta Val	agg Arg 195	gcc Ala	atc Ile	tcg Ser	tgt Cys	cta Leu 200	aaa Lys	ggg Gly	ttc Phe	gag Glu	ttg Leu 205	ggc Gly	tcc Ser	aag Lys	624
ggg Gly	gtg Val 210	caa Gln	ttg Leu	cta Leu	tcc Ser	tcc Ser 215	tat Tyr	gtt Val	act Thr	gat Asp	gag Glu 220	tta Leu	gga Gly	atc Ile	caa Gln	672
tgt Cys 225	ggc Gly	gca Ala	cta Leu	tct Ser	ggg Gly 230	gca Ala	aac Asn	ttg Leu	gca Ala	ccg Pro 235	gaa Glu	gtg Val	gcc Ala	aag Lys	gag Glu 240	720
cat His	tgg Trp	tcc Ser	gaa Glu	acc Thr 245	acc Thr	gtg Val	gct Ala	tac Tyr	caa Gln 250	cta Leu	cca Pro	aag Lys	gat Asp	tat Tyr 255	caa Gln	768
ggg Gly	gat Asp	ggc Gly	aag Lys 260	gat Asp	gta Val	gat Asp	cat His	aag Lys 265	att Ile	ttg Leu	aaa Lys	ttg Leu	ctg Leu 270	ttc Phe	cac His	816
aga Arg	cct Pro	tac Tyr 275	ttc Phe	cac His	gtc Val	aat Asn	gtc Val 280	atc Ile	gat Asp	gat Asp	gtt Val	gct Ala 285	ggg Gly	ata Ile	tcc Ser	864

20110907_CL5178USNA_SeqList.txt

att gcc ggt gcc ttg aag aac gtc gtg gca ctt gca tgt ggt ttc gta Ile Ala Gly Ala Leu Lys Asn Val Val Ala Leu Ala Cys Gly Phe Val 290 295 300	912
gaa ggt atg gga tgg ggt aac aat gcc tcc gca gcc att caa agg ctg Glu Gly Met Gly Trp Gly Asn Asn Ala Ser Ala Ala Ile Gln Arg Leu 305 310 315 320	960
ggt tta ggt gaa att atc aag ttc ggt aga atg ttt ttc cca gaa tcc Gly Leu Gly Glu Ile Ile Lys Phe Gly Arg Met Phe Phe Pro Glu Ser 325 330 335	1008
aaa gtc gag acc tac tat caa gaa tcc gct ggt gtt gca gat ctg atc Lys Val Glu Thr Tyr Tyr Gln Glu Ser Ala Gly Val Ala Asp Leu Ile 340 345 350	1056
acc acc tgc tca ggc ggt aga aac gtc aag gtt gcc aca tac atg gcc Thr Thr Cys Ser Gly Gly Arg Asn Val Lys Val Ala Thr Tyr Met Ala 355 360 365	1104
aag acc ggt aag tca gcc ttg gaa gca gaa aag gaa ttg ctt aac ggt Lys Thr Gly Lys Ser Ala Leu Glu Ala Glu Lys Glu Leu Leu Asn Gly 370 375 380	1152
caa tcc gcc caa ggg ata atc aca tgc aga gaa gtt cac gag tgg cta Gln Ser Ala Gln Gly Ile Ile Thr Cys Arg Glu Val His Glu Trp Leu 385 390 395 400	1200
caa aca tgt gag ttg acc caa gaa ttc cca tta ttc gag gca gtc tac Gln Thr Cys Glu Leu Thr Gln Glu Phe Pro Leu Phe Glu Ala Val Tyr 405 410 415	1248
cag ata gtc tac aac aac gtc cgc atg gaa gac cta ccg gag atg att Gln Ile Val Tyr Asn Asn Val Arg Met Glu Asp Leu Pro Glu Met Ile 420 425 430	1296
gaa gag cta gac atc gat gac gaa tag Glu Glu Leu Asp Ile Asp Asp Glu 435 440	1323

<210> 71
 <211> 440
 <212> PRT
 <213> Saccharomyces cerevisiae

<400> 71

Met Leu Ala Val Arg Arg Leu Thr Arg Tyr Thr Phe Leu Lys Arg Thr 1 5 10 15
His Pro Val Leu Tyr Thr Arg Arg Ala Tyr Lys Ile Leu Pro Ser Arg 20 25 30
Ser Thr Phe Leu Arg Arg Ser Leu Leu Gln Thr Gln Leu His Ser Lys 35 40 45
Met Thr Ala His Thr Asn Ile Lys Gln His Lys His Cys His Glu Asp 50 55 60
His Pro Ile Arg Arg Ser Asp Ser Ala Val Ser Ile Val His Leu Lys 65 70 75 80

20110907_CL5178USNA_SeqList.txt

Arg Ala Pro Phe Lys Val Thr Val Ile Gly Ser Gly Asn Trp Gly Thr
85 90 95

Thr Ile Ala Lys Val Ile Ala Glu Asn Thr Glu Leu His Ser His Ile
100 105 110

Phe Glu Pro Glu Val Arg Met Trp Val Phe Asp Glu Lys Ile Gly Asp
115 120 125

Glu Asn Leu Thr Asp Ile Ile Asn Thr Arg His Gln Asn Val Lys Tyr
130 135 140

Leu Pro Asn Ile Asp Leu Pro His Asn Leu Val Ala Asp Pro Asp Leu
145 150 155 160

Leu His Ser Ile Lys Gly Ala Asp Ile Leu Val Phe Asn Ile Pro His
165 170 175

Gln Phe Leu Pro Asn Ile Val Lys Gln Leu Gln Gly His Val Ala Pro
180 185 190

His Val Arg Ala Ile Ser Cys Leu Lys Gly Phe Glu Leu Gly Ser Lys
195 200 205

Gly Val Gln Leu Leu Ser Ser Tyr Val Thr Asp Glu Leu Gly Ile Gln
210 215 220

Cys Gly Ala Leu Ser Gly Ala Asn Leu Ala Pro Glu Val Ala Lys Glu
225 230 235 240

His Trp Ser Glu Thr Thr Val Ala Tyr Gln Leu Pro Lys Asp Tyr Gln
245 250 255

Gly Asp Gly Lys Asp Val Asp His Lys Ile Leu Lys Leu Leu Phe His
260 265 270

Arg Pro Tyr Phe His Val Asn Val Ile Asp Asp Val Ala Gly Ile Ser
275 280 285

Ile Ala Gly Ala Leu Lys Asn Val Val Ala Leu Ala Cys Gly Phe Val
290 295 300

Glu Gly Met Gly Trp Gly Asn Asn Ala Ser Ala Ala Ile Gln Arg Leu
305 310 315 320

Gly Leu Gly Glu Ile Ile Lys Phe Gly Arg Met Phe Phe Pro Glu Ser
325 330 335

Lys Val Glu Thr Tyr Tyr Gln Glu Ser Ala Gly Val Ala Asp Leu Ile
340 345 350

20110907_CL5178USNA_SeqList.txt

Thr Thr Cys Ser Gly Gly Arg Asn Val Lys Val Ala Thr Tyr Met Ala
355 360 365

Lys Thr Gly Lys Ser Ala Leu Glu Ala Glu Lys Glu Leu Leu Asn Gly
370 375 380

Gln Ser Ala Gln Gly Ile Ile Thr Cys Arg Glu Val His Glu Trp Leu
385 390 395 400

Gln Thr Cys Glu Leu Thr Gln Glu Phe Pro Leu Phe Glu Ala Val Tyr
405 410 415

Gln Ile Val Tyr Asn Asn Val Arg Met Glu Asp Leu Pro Glu Met Ile
420 425 430

Glu Glu Leu Asp Ile Asp Asp Glu
435 440

<210> 72
<211> 1113
<212> DNA
<213> Pichia stipitis

<220>
<221> CDS
<222> (1)..(1113)

<400> 72
atg acc gct act act att cct tac aac atc cct tcg cgt ttc aga att 48
Met Thr Ala Thr Thr Ile Pro Tyr Asn Ile Pro Ser Arg Phe Arg Ile
1 5 10 15
gca att atc ggc tcc ggg aac tgg ggt acg gcc gtc gcc aaa atc gta 96
Ala Ile Ile Gly Ser Gly Asn Trp Gly Thr Ala Val Ala Lys Ile Val
20 25 30
tct gaa aac aca gct gaa aaa tcg gat atc ttc gag ccc atc gtt aag 144
Ser Glu Asn Thr Ala Glu Lys Ser Asp Ile Phe Glu Pro Ile Val Lys
35 40 45
atg tgg gtt ttt gaa gaa gat gtc cag gga aga aaa ttg acg gaa atc 192
Met Trp Val Phe Glu Glu Asp Val Gln Gly Arg Lys Leu Thr Glu Ile
50 55 60
ata aac aat gac cac gaa aat gtc cgt tat ttg cca gga gtg caa ttg 240
Ile Asn Asn Asp His Glu Asn Val Arg Tyr Leu Pro Gly Val Gln Leu
65 70 75 80
cca gaa aac ttg gta gcc gtc cca gac atc gtc gac aca gtt aaa gat 288
Pro Glu Asn Leu Val Ala Val Pro Asp Ile Val Asp Thr Val Lys Asp
85 90 95
gct gat ttg ctc att ttc aac gtt cca cac cag ttt ttg ggc aga atc 336
Ala Asp Leu Leu Ile Phe Asn Val Pro His Gln Phe Leu Gly Arg Ile
100 105 110
tgc aag caa ttg atc ggc aaa gtt tct cca tct gtg aga gct atc tcg 384
Cys Lys Gln Leu Ile Gly Lys Val Ser Pro Ser Val Arg Ala Ile Ser
115 120 125

20110907_CL5178USNA_SeqList.txt

tgt	ttg	aaa	ggt	ttg	gaa	gtc	aat	tcc	gac	ggc	tgc	aaa	ttg	ttg	tct		432
Cys	Leu	Lys	Gly	Leu	Glu	Val	Asn	Ser	Asp	Gly	Cys	Lys	Leu	Leu	Ser		
	130					135					140						
caa	gta	gtc	aca	gac	act	ttg	gga	atc	tac	tgt	ggg	gtc	ttg	tct	gga		480
Gln	Val	Val	Thr	Asp	Thr	Leu	Gly	Ile	Tyr	Cys	Gly	Val	Leu	Ser	Gly		
	145				150					155					160		
gcc	aac	atc	gcc	aat	gaa	ggt	gct	aga	caa	aga	tgg	tcc	gaa	acc	tca		528
Ala	Asn	Ile	Ala	Asn	Glu	Val	Ala	Arg	Gln	Arg	Trp	Ser	Glu	Thr	Ser		
				165					170					175			
atc	gca	tat	aca	gct	cca	aag	gac	ttc	cgt	ggg	cct	gga	ctc	gac	att		576
Ile	Ala	Tyr	Thr	Ala	Pro	Lys	Asp	Phe	Arg	Gly	Pro	Gly	Leu	Asp	Ile		
			180					185					190				
gac	gat	ttc	gtc	ttg	aaa	caa	gct	ttc	cac	aga	cct	tac	ttc	cac	ggt		624
Asp	Asp	Phe	Val	Leu	Lys	Gln	Ala	Phe	His	Arg	Pro	Tyr	Phe	His	Val		
		195					200					205					
aga	ggt	att	gaa	gat	gtc	att	ggg	gct	tct	ata	gct	ggg	gct	ctc	aag		672
Arg	Val	Ile	Glu	Asp	Val	Ile	Gly	Ala	Ser	Ile	Ala	Gly	Ala	Leu	Lys		
	210					215					220						
aat	gtc	gtg	gcc	att	gca	gtg	ggg	ctc	gta	gaa	ggc	gct	gga	tgg	ggg		720
Asn	Val	Val	Ala	Ile	Ala	Val	Gly	Leu	Val	Glu	Gly	Ala	Gly	Trp	Gly		
	225				230					235					240		
gac	aat	gcc	aag	gct	gct	atc	atg	aga	att	ggg	atc	aag	gaa	act	atc		768
Asp	Asn	Ala	Lys	Ala	Ala	Ile	Met	Arg	Ile	Gly	Ile	Lys	Glu	Thr	Ile		
				245					250					255			
cgg	ttt	gct	tct	tac	tac	aag	aag	ttt	ggc	atc	aga	ggc	gca	gct	cca		816
Arg	Phe	Ala	Ser	Tyr	Tyr	Lys	Lys	Phe	Gly	Ile	Arg	Gly	Ala	Ala	Pro		
			260					265					270				
gag	ccc	act	aca	ttt	aca	gaa	gaa	agt	gcc	ggg	gtg	gcc	gat	ctt	atc		864
Glu	Pro	Thr	Thr	Phe	Thr	Glu	Glu	Ser	Ala	Gly	Val	Ala	Asp	Leu	Ile		
		275					280					285					
act	aca	tgt	tcc	ggg	ggg	aga	aac	gtc	aag	gtc	gcc	aga	tac	atg	gtc		912
Thr	Thr	Cys	Ser	Gly	Gly	Arg	Asn	Val	Lys	Val	Ala	Arg	Tyr	Met	Val		
		290				295					300						
gaa	aat	ggg	gtc	gat	gcc	tgg	gaa	gct	gaa	aag	ata	ttg	ttg	aac	ggg		960
Glu	Asn	Gly	Val	Asp	Ala	Trp	Glu	Ala	Glu	Lys	Ile	Leu	Leu	Asn	Gly		
	305				310					315					320		
cag	agt	tca	caa	ggg	atc	ttg	act	gca	aag	gaa	gtg	cac	gag	ctc	ttg		1008
Gln	Ser	Ser	Gln	Gly	Ile	Leu	Thr	Ala	Lys	Glu	Val	His	Glu	Leu	Leu		
				325					330					335			
gag	aac	ttt	gac	ttg	aag	gag	gaa	ttc	ccc	ttg	ttc	gaa	gca	act	tat		1056
Glu	Asn	Phe	Asp	Leu	Lys	Glu	Glu	Phe	Pro	Leu	Phe	Glu	Ala	Thr	Tyr		
			340					345					350				
gcg	gta	att	tac	caa	aac	cat	tct	gtg	gac	gac	ttc	cca	gca	ttg	ttg		1104
Ala	Val	Ile	Tyr	Gln	Asn	His	Ser	Val	Asp	Asp	Phe	Pro	Ala	Leu	Leu		
		355					360					365					
gag	tgt	taa															1113
Glu	Cys																
	370																

<210> 73
 <211> 370
 <212> PRT

<213> Pichia stipitis

<400> 73

Met Thr Ala Thr Thr Ile Pro Tyr Asn Ile Pro Ser Arg Phe Arg Ile
1 5 10 15

Ala Ile Ile Gly Ser Gly Asn Trp Gly Thr Ala Val Ala Lys Ile Val
20 25 30

Ser Glu Asn Thr Ala Glu Lys Ser Asp Ile Phe Glu Pro Ile Val Lys
35 40 45

Met Trp Val Phe Glu Glu Asp Val Gln Gly Arg Lys Leu Thr Glu Ile
50 55 60

Ile Asn Asn Asp His Glu Asn Val Arg Tyr Leu Pro Gly Val Gln Leu
65 70 75 80

Pro Glu Asn Leu Val Ala Val Pro Asp Ile Val Asp Thr Val Lys Asp
85 90 95

Ala Asp Leu Leu Ile Phe Asn Val Pro His Gln Phe Leu Gly Arg Ile
100 105 110

Cys Lys Gln Leu Ile Gly Lys Val Ser Pro Ser Val Arg Ala Ile Ser
115 120 125

Cys Leu Lys Gly Leu Glu Val Asn Ser Asp Gly Cys Lys Leu Leu Ser
130 135 140

Gln Val Val Thr Asp Thr Leu Gly Ile Tyr Cys Gly Val Leu Ser Gly
145 150 155 160

Ala Asn Ile Ala Asn Glu Val Ala Arg Gln Arg Trp Ser Glu Thr Ser
165 170 175

Ile Ala Tyr Thr Ala Pro Lys Asp Phe Arg Gly Pro Gly Leu Asp Ile
180 185 190

Asp Asp Phe Val Leu Lys Gln Ala Phe His Arg Pro Tyr Phe His Val
195 200 205

Arg Val Ile Glu Asp Val Ile Gly Ala Ser Ile Ala Gly Ala Leu Lys
210 215 220

Asn Val Val Ala Ile Ala Val Gly Leu Val Glu Gly Ala Gly Trp Gly
225 230 235 240

Asp Asn Ala Lys Ala Ala Ile Met Arg Ile Gly Ile Lys Glu Thr Ile
245 250 255

20110907_CL5178USNA_SeqList.txt

Arg Phe Ala Ser Tyr Tyr Lys Lys Phe Gly Ile Arg Gly Ala Ala Pro
260 265 270

Glu Pro Thr Thr Phe Thr Glu Glu Ser Ala Gly Val Ala Asp Leu Ile
275 280 285

Thr Thr Cys Ser Gly Gly Arg Asn Val Lys Val Ala Arg Tyr Met Val
290 295 300

Glu Asn Gly Val Asp Ala Trp Glu Ala Glu Lys Ile Leu Leu Asn Gly
305 310 315 320

Gln Ser Ser Gln Gly Ile Leu Thr Ala Lys Glu Val His Glu Leu Leu
325 330 335

Glu Asn Phe Asp Leu Lys Glu Glu Phe Pro Leu Phe Glu Ala Thr Tyr
340 345 350

Ala Val Ile Tyr Gln Asn His Ser Val Asp Asp Phe Pro Ala Leu Leu
355 360 365

Glu Cys
370

<210> 74
<211> 1176
<212> DNA
<213> Pichia stipitis

<220>
<221> CDS
<222> (1)..(1176)

<400> 74
atg tct acg cct gtt gcc cgt ttg gcc caa cta gcc aac att ttg gca 48
Met Ser Thr Pro Val Ala Arg Leu Ala Gln Leu Ala Asn Ile Leu Ala
1 5 10 15
cct aat ccc acc gca gca tac gct gca cat ctg cac cac cca gag gcc 96
Pro Asn Pro Thr Ala Ala Tyr Ala His Leu His His Pro Glu Ala
20 25 30
tcc ttg tct cca gaa cac cct ttc cgt gtt gct gtg atc ggc tcg ggt 144
Ser Leu Ser Pro Glu His Pro Phe Arg Val Ala Val Ile Gly Ser Gly
35 40 45
aac tgg ggg acg acc ata gct aaa gtg atc gcc gaa aac gct gcg gca 192
Asn Trp Gly Thr Thr Ile Ala Lys Val Ile Ala Glu Asn Ala Ala Ala
50 55 60
aga ccc cgt ttg ttc cgt cac caa gtc aat atg tgg gtg cac gat gaa 240
Arg Pro Arg Leu Phe Arg His Gln Val Asn Met Trp Val His Asp Glu
65 70 75 80
atc atc gat ggt gaa aaa ttg act cac atc atc aac acc aga cac gaa 288
Ile Ile Asp Gly Glu Lys Leu Thr His Ile Ile Asn Thr Arg His Glu
85 90 95
aac gtc aag tac ttg ccc ggt gtg atc ttg ccg aga aac atc aga gcc 336

20110907_CL5178USNA_SeqList.txt

Asn	Val	Lys	Tyr	Leu	Pro	Gly	Val	Ile	Leu	Pro	Arg	Asn	Ile	Arg	Ala	
			100					105					110			
gaa	gca	gat	atc	ggc	aat	gta	gtc	cat	gac	gcc	gac	ttg	att	gtc	ttt	384
Glu	Ala	Asp	Ile	Gly	Asn	Val	Val	His	Asp	Ala	Asp	Leu	Ile	Val	Phe	
		115					120					125				
aac	ttg	cct	cat	cag	ttc	ttg	ccc	cgt	gta	gtc	aaa	tcg	ttg	aag	ggc	432
Asn	Leu	Pro	His	Gln	Phe	Leu	Pro	Arg	Val	Val	Lys	Ser	Leu	Lys	Gly	
	130					135					140					
aag	ata	aag	cac	ggc	gcc	aga	gcc	atc	tcg	tgc	ttg	aag	ggc	ttg	gaa	480
Lys	Ile	Lys	His	Gly	Ala	Arg	Ala	Ile	Ser	Cys	Leu	Lys	Gly	Leu	Glu	
145					150					155					160	
gtg	act	cca	gaa	ggc	tgc	aaa	ttg	ttg	tct	acc	tac	atc	acg	gaa	gag	528
Val	Thr	Pro	Glu	Gly	Cys	Lys	Leu	Leu	Ser	Thr	Tyr	Ile	Thr	Glu	Glu	
				165					170					175		
tta	ggc	atc	gtt	tgc	ggc	gct	ctc	agt	ggc	gcc	aac	att	gct	cca	gaa	576
Leu	Gly	Ile	Val	Cys	Gly	Ala	Leu	Ser	Gly	Ala	Asn	Ile	Ala	Pro	Glu	
			180					185					190			
gtt	gct	aga	tgc	aaa	tgg	tca	gag	act	acc	gta	gcc	tac	aag	ctc	cca	624
Val	Ala	Arg	Cys	Lys	Trp	Ser	Glu	Thr	Thr	Val	Ala	Tyr	Lys	Leu	Pro	
		195					200					205				
gag	gac	ttc	aga	ggc	gca	ggc	aag	gac	atc	gac	aag	ttc	gtg	ttg	aga	672
Glu	Asp	Phe	Arg	Gly	Ala	Gly	Lys	Asp	Ile	Asp	Lys	Phe	Val	Leu	Arg	
	210					215					220					
gca	tgt	ttc	cac	aga	ccc	tac	ttc	cac	gtc	aat	gtt	atc	gag	gat	gtt	720
Ala	Cys	Phe	His	Arg	Pro	Tyr	Phe	His	Val	Asn	Val	Ile	Glu	Asp	Val	
225					230					235					240	
gcc	ggc	gta	tct	gta	gct	gga	gct	ttg	aag	aat	gtg	gtt	gct	tta	gct	768
Ala	Gly	Val	Ser	Val	Ala	Gly	Ala	Leu	Lys	Asn	Val	Val	Ala	Leu	Ala	
				245					250					255		
gta	gga	ttt	gta	gag	ggc	ttg	ggc	tgg	ggc	gac	aac	gct	aag	gct	gct	816
Val	Gly	Phe	Val	Glu	Gly	Leu	Gly	Trp	Gly	Asp	Asn	Ala	Lys	Ala	Ala	
			260					265					270			
gtc	atg	aga	gta	gga	ttg	ttg	gag	acg	atc	aag	ttc	tca	gaa	aca	ttc	864
Val	Met	Arg	Val	Gly	Leu	Leu	Glu	Thr	Ile	Lys	Phe	Ser	Glu	Thr	Phe	
		275					280					285				
ttc	cct	gag	tcc	att	gca	tcc	act	ttc	aca	gcc	gag	tcg	gct	ggc	gta	912
Phe	Pro	Glu	Ser	Ile	Ala	Ser	Thr	Phe	Thr	Ala	Glu	Ser	Ala	Gly	Val	
	290					295					300					
gct	gat	ttg	atc	aca	tca	tgc	tct	ggc	ggc	aga	aat	gtg	aaa	gtc	ggc	960
Ala	Asp	Leu	Ile	Thr	Ser	Cys	Ser	Gly	Gly	Arg	Asn	Val	Lys	Val	Gly	
305					310					315					320	
aga	tac	atg	gcc	caa	acc	gga	tct	tcg	gcc	gaa	aac	gct	gaa	aag	atg	1008
Arg	Tyr	Met	Ala	Gln	Thr	Gly	Ser	Ser	Ala	Glu	Asn	Ala	Glu	Lys	Met	
				325					330					335		
tta	ttg	aac	ggc	cag	agc	tcc	cag	ggc	atc	gta	act	gtt	cgt	gaa	gtt	1056
Leu	Leu	Asn	Gly	Gln	Ser	Ser	Gln	Gly	Ile	Val	Thr	Val	Arg	Glu	Val	
			340					345					350			
cac	gac	ttg	ttg	acc	aac	gta	ggc	atg	ctc	gac	aaa	ttc	ccc	ttg	ttt	1104
His	Asp	Leu	Leu	Thr	Asn	Val	Gly	Met	Leu	Asp	Lys	Phe	Pro	Leu	Phe	
		355					360					365				
gaa	gct	acc	tat	caa	atc	atc	tat	ggc	tca	gag	tcc	atc	gaa	aac	ttg	1152

20110907_CL5178USNA_SeqList.txt

Glu Ala Thr Tyr Gln Ile Ile Tyr Gly Ser Glu Ser Ile Glu Asn Leu
370 375 380

cct ctc ttg tta gta gga gaa taa
Pro Leu Leu Leu Val Gly Glu
385 390

1176

<210> 75
<211> 391
<212> PRT
<213> Pichia stipitis

<400> 75

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

Pro Asn Pro Thr Ala Ala Tyr Ala Ala His Leu His His Pro Glu Ala
20 25 30

Ser Leu Ser Pro Glu His Pro Phe Arg Val Ala Val Ile Gly Ser Gly
35 40 45

Asn Trp Gly Thr Thr Ile Ala Lys Val Ile Ala Glu Asn Ala Ala Ala
50 55 60

Arg Pro Arg Leu Phe Arg His Gln Val Asn Met Trp Val His Asp Glu
65 70 75 80

Ile Ile Asp Gly Glu Lys Leu Thr His Ile Ile Asn Thr Arg His Glu
85 90 95

Asn Val Lys Tyr Leu Pro Gly Val Ile Leu Pro Arg Asn Ile Arg Ala
100 105 110

Glu Ala Asp Ile Gly Asn Val Val His Asp Ala Asp Leu Ile Val Phe
115 120 125

Asn Leu Pro His Gln Phe Leu Pro Arg Val Val Lys Ser Leu Lys Gly
130 135 140

Lys Ile Lys His Gly Ala Arg Ala Ile Ser Cys Leu Lys Gly Leu Glu
145 150 155 160

Val Thr Pro Glu Gly Cys Lys Leu Leu Ser Thr Tyr Ile Thr Glu Glu
165 170 175

Leu Gly Ile Val Cys Gly Ala Leu Ser Gly Ala Asn Ile Ala Pro Glu
180 185 190

Val Ala Arg Cys Lys Trp Ser Glu Thr Thr Val Ala Tyr Lys Leu Pro
195 200 205

Glu Asp Phe Arg Gly Ala Gly Lys Asp Ile Asp Lys Phe Val Leu Arg
Page 123

210

215

220

Ala Cys Phe His Arg Pro Tyr Phe His Val Asn Val Ile Glu Asp Val
225 230 235 240

Ala Gly Val Ser Val Ala Gly Ala Leu Lys Asn Val Val Ala Leu Ala
245 250 255

Val Gly Phe Val Glu Gly Leu Gly Trp Gly Asp Asn Ala Lys Ala Ala
260 265 270

Val Met Arg Val Gly Leu Leu Glu Thr Ile Lys Phe Ser Glu Thr Phe
275 280 285

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

Ala Asp Leu Ile Thr Ser Cys Ser Gly Gly Arg Asn Val Lys Val Gly
305 310 315 320

Arg Tyr Met Ala Gln Thr Gly Ser Ser Ala Glu Asn Ala Glu Lys Met
325 330 335

Leu Leu Asn Gly Gln Ser Ser Gln Gly Ile Val Thr Val Arg Glu Val
340 345 350

His Asp Leu Leu Thr Asn Val Gly Met Leu Asp Lys Phe Pro Leu Phe
355 360 365

Glu Ala Thr Tyr Gln Ile Ile Tyr Gly Ser Glu Ser Ile Glu Asn Leu
370 375 380

Pro Leu Leu Leu Val Gly Glu
385 390

<210> 76
<211> 1236
<212> DNA
<213> Kluyveromyces thermotolerans

<220>
<221> CDS
<222> (1)..(1236)

<400> 76
atg ttt tca atc tcc aga atc act aga act tct agt ttc act aca cag 48
Met Phe Ser Ile Ser Arg Ile Thr Arg Thr Ser Ser Phe Thr Thr Gln
1 5 10 15

ttt aga gcg cta tac cgt ttt aaa cac tca gca agg aaa ctc cag agc 96
Phe Arg Ala Leu Tyr Arg Phe Lys His Ser Ala Arg Lys Leu Gln Ser
20 25 30

atc cct ttc agc ata tac aag aaa atg tcc gct gcc gac aga ctg aat 144
Ile Pro Phe Ser Ile Tyr Lys Lys Met Ser Ala Ala Asp Arg Leu Asn

35					40					45						
cag Gln	acc Thr 50	cac His	gac Asp	atc Ile	cta Leu	tcc Ser 55	gaa Glu	tcc Ser	gtg Val	caa Gln	gcc Ala 60	gtg Val	gaa Glu	aac Asn	cct Pro	192
ttc Phe 65	aaa Lys	gtc Val	acc Thr	gtg Val	att Ile 70	ggg Gly	tcc Ser	ggg Gly	aac Asn	tgg Trp 75	ggg Gly	acc Thr	acc Thr	atc Ile	tcc Ser 80	240
aag Lys	gtt Val	gtg Val	gcc Ala	gag Glu 85	aac Asn	gct Ala	gcg Ala	cta Leu	aga Arg 90	cca Pro	cac His	ttg Leu	ttc Phe	gtc Val 95	aag Lys	288
cgt Arg	gtc Val	gac Asp	atg Met 100	tgg Trp	gtg Val	ttt Phe	gag Glu	gag Glu 105	acc Thr	gtc Val	gac Asp	ggc Gly	cag Gln 110	aag Lys	ttg Leu	336
acc Thr	gag Glu	atc Ile 115	atc Ile	aac Asn	acc Thr	aag Lys	cac His 120	cag Gln	aac Asn	gtc Val	aag Lys	tac Tyr 125	ctg Leu	cct Pro	aac Asn	384
atc Ile	gac Asp 130	ctg Leu	cca Pro	gag Glu	aac Asn	ctg Leu 135	gtc Val	gcc Ala	aac Asn	cca Pro	gac Asp 140	ttg Leu	gtc Val	tct Ser	gcc Ala	432
gtc Val 145	aag Lys	gac Asp	gcc Ala	gac Asp	atc Ile 150	ctg Leu	gtc Val	ttc Phe	aac Asn	atc Ile 155	cct Pro	cac His	cag Gln	ttc Phe	ttg Leu 160	480
cca Pro	cgc Arg	att Ile	gtc Val	tcc Ser 165	cag Gln	ttg Leu	cag Gln	ggc Gly	aac Asn 170	atc Ile	aag Lys	aag Lys	gac Asp	gcc Ala 175	cgt Arg	528
gcc Ala	atc Ile	tcc Ser	tgt Cys 180	ttg Leu	aag Lys	ggg Gly	ttc Phe	gac Asp 185	gtg Val	tcc Ser	aag Lys	gac Asp	ggg Gly 190	gtc Val	aag Lys	576
ctg Leu	cta Leu	tcc Ser 195	acc Thr	tac Tyr	gtc Val	acc Thr	gag Glu 200	aag Lys	ctc Leu	gga Gly	atc Ile	acc Thr 205	tgt Cys	ggg Gly	gcc Ala	624
ctt Leu	tcc Ser 210	ggg Gly	gct Ala	aac Asn	ttg Leu	gcc Ala 215	cca Pro	gag Glu	gtt Val	gcc Ala 220	aag Lys	gag Glu	aac Asn	tgg Trp	tcc Ser	672
gag Glu 225	acc Thr	act Thr	gtt Val	gcc Ala	tac Tyr 230	gag Glu	ttg Leu	cca Pro	aag Lys	gac Asp 235	ttc Phe	aag Lys	ggc Gly	gag Glu 240	ggc Gly	720
aag Lys	gac Asp	gtc Val	gac Asp	cac His 245	gcc Ala	gtc Val	ttg Leu	aag Lys	gct Ala 250	ttg Leu	ttc Phe	cac His	aga Arg	cca Pro 255	tac Tyr	768
ttc Phe	cac His	gtc Val	aac Asn 260	gtc Val	att Ile	gac Asp	gat Asp	gtc Val 265	gct Ala	ggg Gly	atc Ile	tct Ser	gtt Val 270	gcc Ala	ggg Gly	816
gcc Ala	ttg Leu	aag Lys 275	aac Asn	gtt Val	gtc Val	gct Ala	cta Leu 280	ggg Gly	tgc Cys	ggg Gly	ttc Phe	gtc Val 285	gag Glu	ggg Gly	cta Leu	864
ggc Gly	tgg Trp 290	ggg Gly	aac Asn	aac Asn	gcc Ala	tcc Ser 295	gct Ala	gcc Ala	atc Ile	cag Gln	aga Arg 300	gtt Val	ggg Gly	ttg Leu	ggg Gly	912
gag Glu	atc Ile	atc Ile	aag Lys	ttc Phe	ggg Gly	cag Gln	atg Met	ttc Phe	ttc Phe	cca Pro	gac Asp	tcc Ser	cgt Arg	gtc Val	gag Glu	960

20110907_CL5178USNA_SeqList.txt

```

305          310          315          320
acc tac tac cag gag tcc gcc ggt gtt gct gac ttg atc acc acc tgc      1008
Thr Tyr Tyr Gln Glu Ser Ala Gly Val Ala Asp Leu Ile Thr Thr Cys
325          330          335

tct ggt ggt aga aac gtc aga gtc gcc acc cac atg gcc aag act ggt      1056
Ser Gly Gly Arg Asn Val Arg Val Ala Thr His Met Ala Lys Thr Gly
340          345          350

aag tct gcc gag gag tgc gag aag gag ctg ttg aac ggc cag tcc gcc      1104
Lys Ser Ala Glu Glu Cys Glu Lys Glu Leu Leu Asn Gly Gln Ser Ala
355          360          365

cag gta ttc aca cat gta agg agg tcc acg agt ggt tgg ccg agt gcg      1152
Gln Val Phe Thr His Val Arg Arg Ser Thr Ser Gly Trp Pro Ser Ala
370          375          380

gta aga ccg atg aat tcg ttc tgt tcg agg ccg ttt acc aga ttg tct      1200
Val Arg Pro Met Asn Ser Phe Cys Ser Arg Pro Phe Thr Arg Leu Ser
385          390          395

acg aga acg ctc cta tgg aca ctt tgc cag aca tga      1236
Thr Arg Thr Leu Leu Trp Thr Leu Cys Gln Thr
405          410

```

```

<210> 77
<211> 411
<212> PRT
<213> Kluyveromyces thermotolerans
<400> 77

```

```

Met Phe Ser Ile Ser Arg Ile Thr Arg Thr Ser Ser Phe Thr Thr Gln
1          5          10          15

```

```

Phe Arg Ala Leu Tyr Arg Phe Lys His Ser Ala Arg Lys Leu Gln Ser
20          25          30

```

```

Ile Pro Phe Ser Ile Tyr Lys Lys Met Ser Ala Ala Asp Arg Leu Asn
35          40          45

```

```

Gln Thr His Asp Ile Leu Ser Glu Ser Val Gln Ala Val Glu Asn Pro
50          55          60

```

```

Phe Lys Val Thr Val Ile Gly Ser Gly Asn Trp Gly Thr Thr Ile Ser
65          70          75          80

```

```

Lys Val Val Ala Glu Asn Ala Ala Leu Arg Pro His Leu Phe Val Lys
85          90          95

```

```

Arg Val Asp Met Trp Val Phe Glu Glu Thr Val Asp Gly Gln Lys Leu
100          105          110

```

```

Thr Glu Ile Ile Asn Thr Lys His Gln Asn Val Lys Tyr Leu Pro Asn
115          120          125

```

```

Ile Asp Leu Pro Glu Asn Leu Val Ala Asn Pro Asp Leu Val Ser Ala
130          135          140

```

20110907_CL5178USNA_SeqList.txt

Val Lys Asp Ala Asp Ile Leu Val Phe Asn Ile Pro His Gln Phe Leu
145 150 155 160

Pro Arg Ile Val Ser Gln Leu Gln Gly Asn Ile Lys Lys Asp Ala Arg
165 170 175

Ala Ile Ser Cys Leu Lys Gly Phe Asp Val Ser Lys Asp Gly Val Lys
180 185 190

Leu Leu Ser Thr Tyr Val Thr Glu Lys Leu Gly Ile Thr Cys Gly Ala
195 200 205

Leu Ser Gly Ala Asn Leu Ala Pro Glu Val Ala Lys Glu Asn Trp Ser
210 215 220

Glu Thr Thr Val Ala Tyr Glu Leu Pro Lys Asp Phe Lys Gly Glu Gly
225 230 235 240

Lys Asp Val Asp His Ala Val Leu Lys Ala Leu Phe His Arg Pro Tyr
245 250 255

Phe His Val Asn Val Ile Asp Asp Val Ala Gly Ile Ser Val Ala Gly
260 265 270

Ala Leu Lys Asn Val Val Ala Leu Gly Cys Gly Phe Val Glu Gly Leu
275 280 285

Gly Trp Gly Asn Asn Ala Ser Ala Ala Ile Gln Arg Val Gly Leu Gly
290 295 300

Glu Ile Ile Lys Phe Gly Gln Met Phe Phe Pro Asp Ser Arg Val Glu
305 310 315 320

Thr Tyr Tyr Gln Glu Ser Ala Gly Val Ala Asp Leu Ile Thr Thr Cys
325 330 335

Ser Gly Gly Arg Asn Val Arg Val Ala Thr His Met Ala Lys Thr Gly
340 345 350

Lys Ser Ala Glu Glu Cys Glu Lys Glu Leu Leu Asn Gly Gln Ser Ala
355 360 365

Gln Val Phe Thr His Val Arg Arg Ser Thr Ser Gly Trp Pro Ser Ala
370 375 380

Val Arg Pro Met Asn Ser Phe Cys Ser Arg Pro Phe Thr Arg Leu Ser
385 390 395 400

Thr Arg Thr Leu Leu Trp Thr Leu Cys Gln Thr
405 410

<210> 78
 <211> 1158
 <212> DNA
 <213> Schizosaccharomyces pombe

<220>
 <221> CDS
 <222> (1)..(1158)

<400> 78
 atg tct gga tat ggt caa caa ggt gtt tct gct gcc aac atc gac agc 48
 Met Ser Gly Tyr Gly Gln Gln Gly Val Ser Ala Ala Asn Ile Asp Ser
 1 5 10 15
 atc cgc ccc aag aaa cgt ttg tca att ggt gta gtt ggc tcc ggt aac 96
 Ile Arg Pro Lys Lys Arg Leu Ser Ile Gly Val Val Gly Ser Gly Asn
 20 25 30
 tgg ggt act gcc att gcc aag att tgc ggt gaa aat gcc cgt gcc cac 144
 Trp Gly Thr Ala Ile Ala Lys Ile Cys Gly Glu Asn Ala Arg Ala His
 35 40 45
 ggt cac cat ttc aga agt aag gtc cgc atg tgg gtc ttt gag gag gag 192
 Gly His His Phe Arg Ser Lys Val Arg Met Trp Val Phe Glu Glu Glu
 50 55 60
 att gag tac aag ggt gag aag aga aag ctc acc gaa gta ttc aac gaa 240
 Ile Glu Tyr Lys Gly Glu Lys Arg Lys Leu Thr Thr Glu Val Phe Asn Glu
 65 70 75 80
 gct cac gag aat gtc aaa tac tta ccc ggc atc gaa tgc cct ccc aac 288
 Ala His Glu Asn Val Lys Tyr Leu Pro Gly Ile Glu Cys Pro Pro Asn
 85 90 95
 gtt att gcc gtc ccc gat gtt cgt gag gtc gct aga cgt gcc gac atc 336
 Val Ile Ala Val Pro Asp Val Arg Glu Val Ala Arg Arg Ala Asp Ile
 100 105 110
 ctt gtc ttt gtc gtt cct cat caa ttt att gaa cgc gtt tgc gac caa 384
 Leu Val Phe Val Val Pro His Gln Phe Ile Glu Arg Val Cys Asp Gln
 115 120 125
 atg gtc ggt ctc att cgc cct ggt gcc gtt ggt att tcc tgt atc aag 432
 Met Val Gly Leu Ile Arg Pro Gly Ala Val Gly Ile Ser Cys Ile Lys
 130 135 140
 ggt gtt gct gtc agc aag gaa ggc gtc cgc ctt tac tct gag gtt atc 480
 Gly Val Ala Val Ser Lys Glu Gly Val Arg Leu Tyr Ser Glu Val Ile
 145 150 155 160
 agc gag aaa ctc ggt att tac tgt ggt gtt ctt tct ggt gct aac gtt 528
 Ser Glu Lys Leu Gly Ile Tyr Cys Gly Val Leu Ser Gly Ala Asn Val
 165 170 175
 gca aac gaa gtt gcc cgt gag caa ttc tgt gag act act att ggt ttc 576
 Ala Asn Glu Val Ala Arg Glu Gln Phe Cys Glu Thr Thr Ile Gly Phe
 180 185 190
 aac cct cct aat gaa gtt gat atc cct cgc gag caa atc gcc gcc gtc 624
 Asn Pro Pro Asn Glu Val Asp Ile Pro Arg Glu Gln Ile Ala Ala Val
 195 200 205
 ttt gat cgc cct tac ttc tca gtt gtc tcc gtt gac gac gtt gcc ggt 672
 Phe Asp Arg Pro Tyr Phe Ser Val Val Ser Val Asp Asp Val Ala Gly
 210 215 220

20110907_CL5178USNA_SeqList.txt

gtc gcc ttg ggt ggt gct ttg aag aac gta gtt gcc atg gcc gtt ggt Val Ala Leu Gly Gly Ala Leu Lys Asn Val Val Ala Met Ala Val Gly 225 230 235 240	720
ttc gct gat ggt ttg gaa tgg ggc ggt aat acc aag gcc gct att atg Phe Ala Asp Gly Leu Glu Trp Gly Gly Asn Thr Lys Ala Ala Ile Met 245 250 255	768
cgt cgt ggt ttg ttg gag atg caa aag ttt gct act acc ttc ttc gac Arg Arg Gly Leu Leu Glu Met Gln Lys Phe Ala Thr Thr Phe Phe Asp 260 265 270	816
tct gat cct cgt acc atg gtt gag caa tct tgc ggt atc gct gac ttg Ser Asp Pro Arg Thr Met Val Glu Gln Ser Cys Gly Ile Ala Asp Leu 275 280 285	864
gtc act tct tgt ttg ggt ggc cgt aac aat cgt tgt gct gaa gca ttt Val Thr Ser Cys Leu Gly Gly Arg Asn Asn Arg Cys Ala Glu Ala Phe 290 295 300	912
gtc aag act ggt aaa tct tta gag acg ctt gaa aaa gag ctc tta ggt Val Lys Thr Gly Lys Ser Leu Glu Thr Leu Glu Lys Glu Leu Leu Gly 305 310 315 320	960
ggt caa ctt ctt caa gga gct gcc act tcc aag gat gtt cat gaa ttc Gly Gln Leu Leu Gln Gly Ala Ala Thr Ser Lys Asp Val His Glu Phe 325 330 335	1008
ctt ctc acc aag gat atg gtc aag gat ttc ccc ttg ttc act gcc gtt Leu Leu Thr Lys Asp Met Val Lys Asp Phe Pro Leu Phe Thr Ala Val 340 345 350	1056
tat aac att tcc tat gaa gac atg gat ccc aag gat ttg atc atc gtc Tyr Asn Ile Ser Tyr Glu Asp Met Asp Pro Lys Asp Leu Ile Ile Val 355 360 365	1104
ctt caa ccc ctt aag gag gac tct gag aac gag ggc ggt act gaa acc Leu Gln Pro Leu Lys Glu Asp Ser Glu Asn Glu Gly Gly Thr Glu Thr 370 375 380	1152
gag taa Glu 385	1158

<210> 79
 <211> 385
 <212> PRT
 <213> Schizosaccharomyces pombe

<400> 79

Met Ser Gly Tyr Gly Gln Gln Gly Val Ser Ala Ala Asn Ile Asp Ser 1 5 10 15
Ile Arg Pro Lys Lys Arg Leu Ser Ile Gly Val Val Gly Ser Gly Asn 20 25 30
Trp Gly Thr Ala Ile Ala Lys Ile Cys Gly Glu Asn Ala Arg Ala His 35 40 45
Gly His His Phe Arg Ser Lys Val Arg Met Trp Val Phe Glu Glu Glu 50 55 60

20110907_CL5178USNA_SeqList.txt

Ile Glu Tyr Lys Gly Glu Lys Arg Lys Leu Thr Glu Val Phe Asn Glu
65 70 75 80

Ala His Glu Asn Val Lys Tyr Leu Pro Gly Ile Glu Cys Pro Pro Asn
85 90 95

Val Ile Ala Val Pro Asp Val Arg Glu Val Ala Arg Arg Ala Asp Ile
100 105 110

Leu Val Phe Val Val Pro His Gln Phe Ile Glu Arg Val Cys Asp Gln
115 120 125

Met Val Gly Leu Ile Arg Pro Gly Ala Val Gly Ile Ser Cys Ile Lys
130 135 140

Gly Val Ala Val Ser Lys Glu Gly Val Arg Leu Tyr Ser Glu Val Ile
145 150 155 160

Ser Glu Lys Leu Gly Ile Tyr Cys Gly Val Leu Ser Gly Ala Asn Val
165 170 175

Ala Asn Glu Val Ala Arg Glu Gln Phe Cys Glu Thr Thr Ile Gly Phe
180 185 190

Asn Pro Pro Asn Glu Val Asp Ile Pro Arg Glu Gln Ile Ala Ala Val
195 200 205

Phe Asp Arg Pro Tyr Phe Ser Val Val Ser Val Asp Asp Val Ala Gly
210 215 220

Val Ala Leu Gly Gly Ala Leu Lys Asn Val Val Ala Met Ala Val Gly
225 230 235 240

Phe Ala Asp Gly Leu Glu Trp Gly Gly Asn Thr Lys Ala Ala Ile Met
245 250 255

Arg Arg Gly Leu Leu Glu Met Gln Lys Phe Ala Thr Thr Phe Phe Asp
260 265 270

Ser Asp Pro Arg Thr Met Val Glu Gln Ser Cys Gly Ile Ala Asp Leu
275 280 285

Val Thr Ser Cys Leu Gly Gly Arg Asn Asn Arg Cys Ala Glu Ala Phe
290 295 300

Val Lys Thr Gly Lys Ser Leu Glu Thr Leu Glu Lys Glu Leu Leu Gly
305 310 315 320

Gly Gln Leu Leu Gln Gly Ala Ala Thr Ser Lys Asp Val His Glu Phe
325 330 335

20110907_CL5178USNA_SeqList.txt

Leu Leu Thr Lys Asp Met Val Lys Asp Phe Pro Leu Phe Thr Ala Val
 340 345 350

Tyr Asn Ile Ser Tyr Glu Asp Met Asp Pro Lys Asp Leu Ile Ile Val
 355 360 365

Leu Gln Pro Leu Lys Glu Asp Ser Glu Asn Glu Gly Gly Thr Glu Thr
 370 375 380

Glu
 385

<210> 80
 <211> 1122
 <212> DNA
 <213> Schizosaccharomyces pombe

<220>
 <221> CDS
 <222> (1)..(1122)

<400> 80
 atg act gtg gct gct ttg aac aaa ctc agc gct ctc tcc gga agt att 48
 Met Thr Val Ala Ala Leu Asn Lys Leu Ser Ala Leu Ser Gly Ser Ile
 1 5 10 15
 caa aaa tct ttt tca cct aaa ctt att tct gtt ggt atc atc gga tca 96
 Gln Lys Ser Phe Ser Pro Lys Leu Ile Ser Val Gly Ile Ile Gly Ser
 20 25 30
 gga aat tgg gga acc gct att gct aaa ata tgt ggt gaa aat gcc aag 144
 Gly Asn Trp Gly Thr Ala Ile Ala Lys Ile Cys Gly Glu Asn Ala Lys
 35 40 45
 gct cat cct gat att ttc cat cct caa gta cac atg tgg atg tat gaa 192
 Ala His Pro Asp Ile Phe His Pro Gln Val His Met Trp Met Tyr Glu
 50 55 60
 gag aag att caa cat gag gga aaa gag tgc aat ctc acg gaa gtt ttt 240
 Glu Lys Ile Gln His Glu Gly Lys Glu Cys Asn Leu Thr Glu Val Phe
 65 70 75 80
 aac act act cat gaa aac gtt aaa tat ctc aaa ggt atc aaa tgc cct 288
 Asn Thr Thr His Glu Asn Val Lys Tyr Leu Lys Gly Ile Lys Cys Pro
 85 90 95
 tct aac gtc ttc gca aac ccg gac att cgt gat gta ggt tca cgt agc 336
 Ser Asn Val Phe Ala Asn Pro Asp Ile Arg Asp Val Gly Ser Arg Ser
 100 105 110
 gac att ctg gta tgg gtt ctc cct cac cag ttc gtt gtg cgt att tgc 384
 Asp Ile Leu Val Trp Val Leu Pro His Gln Phe Val Val Arg Ile Cys
 115 120 125
 aat caa ttg aag gga tgc cta aag aag gat gct gtt gca att tca tgc 432
 Asn Gln Leu Lys Gly Cys Leu Lys Lys Asp Ala Val Ala Ile Ser Cys
 130 135 140
 atc aaa ggt gta tct gtc acc aag gac cgt gtt cgc ctc ttt tct gat 480
 Ile Lys Gly Val Ser Val Thr Lys Asp Arg Val Arg Leu Phe Ser Asp
 145 150 155 160

20110907_CL5178USNA_SeqList.txt

att atc gaa gaa aac acg gga atg tat tgt ggc gtt ctc tct ggc gcc Ile Ile Glu Glu Asn Thr Gly Met Tyr Cys Gly Val Leu Ser Gly Ala 165 170 175	528
aac att gcc agc gaa gtt gct caa gag aag ttt tgc gaa act act atc Asn Ile Ala Ser Glu Val Ala Gln Glu Lys Phe Cys Glu Thr Thr Ile 180 185 190	576
gga tat ttg cct aat agt tct gtt aat ccc cgc tat act cct aag act Gly Tyr Leu Pro Asn Ser Ser Val Asn Pro Arg Tyr Thr Pro Lys Thr 195 200 205	624
atc caa gct ttg ttt aac cgt ccc tac ttc cgt gtc aac att gtt gag Ile Gln Ala Leu Phe Asn Arg Pro Tyr Phe Arg Val Asn Ile Val Glu 210 215 220	672
gat gtt cct ggt gtt gct ttg ggc ggt gca ctc aag aat atc gtc gct Asp Val Pro Gly Val Ala Leu Gly Gly Ala Leu Lys Asn Ile Val Ala 225 230 235 240	720
gtc gct gcc ggt att att gat gga ctt gaa ttg gga gat aat acc aaa Val Ala Ala Gly Ile Ile Asp Gly Leu Glu Leu Gly Asp Asn Thr Lys 245 250 255	768
tct gct gtt atg cgc att ggc ctt ctg gaa atg cag aaa ttc ggc agg Ser Ala Val Met Arg Ile Gly Leu Leu Glu Met Gln Lys Phe Gly Arg 260 265 270	816
atg ttt ttc gat tgt aag cct ctt act atg agc gag gaa tct tgt ggc Met Phe Phe Asp Cys Lys Pro Leu Thr Met Ser Glu Glu Ser Cys Gly 275 280 285	864
ata gcc gat tta att aca act tgc tta ggc ggc cgt aac cac aaa tgc Ile Ala Asp Leu Ile Thr Thr Cys Leu Gly Gly Arg Asn His Lys Cys 290 295 300	912
gct gtg gca ttt gtc aag aca gga aag ccc atg cat gtt gtt gaa caa Ala Val Ala Phe Val Lys Thr Gly Lys Pro Met His Val Val Glu Gln 305 310 315 320	960
gaa ctt ctt gat ggt cag aag ttg caa ggt gca gct acc gcg aag gag Glu Leu Leu Asp Gly Gln Lys Leu Gln Gly Ala Ala Thr Ala Lys Glu 325 330 335	1008
gtt tat gag ttc ctt gat aac cag aat aag gta agc gaa ttc cca ttg Val Tyr Glu Phe Leu Asp Asn Gln Asn Lys Val Ser Glu Phe Pro Leu 340 345 350	1056
ttt aca gct gtt tat cgc att gtt tat gag gga ctt cca cct aat aag Phe Thr Ala Val Tyr Arg Ile Val Tyr Glu Gly Leu Pro Pro Asn Lys 355 360 365	1104
ctt ctg gag gct att taa Leu Leu Glu Ala Ile 370	1122

<210> 81
 <211> 373
 <212> PRT
 <213> Schizosaccharomyces pombe

<400> 81

Met Thr Val Ala Ala Leu Asn Lys Leu Ser Ala Leu Ser Gly Ser Ile
 1 5 10 15

20110907_CL5178USNA_SeqList.txt

Gln Lys Ser Phe Ser Pro Lys Leu Ile Ser Val Gly Ile Ile Gly Ser
20 25 30

Gly Asn Trp Gly Thr Ala Ile Ala Lys Ile Cys Gly Glu Asn Ala Lys
35 40 45

Ala His Pro Asp Ile Phe His Pro Gln Val His Met Trp Met Tyr Glu
50 55 60

Glu Lys Ile Gln His Glu Gly Lys Glu Cys Asn Leu Thr Glu Val Phe
65 70 75 80

Asn Thr Thr His Glu Asn Val Lys Tyr Leu Lys Gly Ile Lys Cys Pro
85 90 95

Ser Asn Val Phe Ala Asn Pro Asp Ile Arg Asp Val Gly Ser Arg Ser
100 105 110

Asp Ile Leu Val Trp Val Leu Pro His Gln Phe Val Val Arg Ile Cys
115 120 125

Asn Gln Leu Lys Gly Cys Leu Lys Lys Asp Ala Val Ala Ile Ser Cys
130 135 140

Ile Lys Gly Val Ser Val Thr Lys Asp Arg Val Arg Leu Phe Ser Asp
145 150 155 160

Ile Ile Glu Glu Asn Thr Gly Met Tyr Cys Gly Val Leu Ser Gly Ala
165 170 175

Asn Ile Ala Ser Glu Val Ala Gln Glu Lys Phe Cys Glu Thr Thr Ile
180 185 190

Gly Tyr Leu Pro Asn Ser Ser Val Asn Pro Arg Tyr Thr Pro Lys Thr
195 200 205

Ile Gln Ala Leu Phe Asn Arg Pro Tyr Phe Arg Val Asn Ile Val Glu
210 215 220

Asp Val Pro Gly Val Ala Leu Gly Gly Ala Leu Lys Asn Ile Val Ala
225 230 235 240

Val Ala Ala Gly Ile Ile Asp Gly Leu Glu Leu Gly Asp Asn Thr Lys
245 250 255

Ser Ala Val Met Arg Ile Gly Leu Leu Glu Met Gln Lys Phe Gly Arg
260 265 270

Met Phe Phe Asp Cys Lys Pro Leu Thr Met Ser Glu Glu Ser Cys Gly
275 280 285

20110907_CL5178USNA_SeqList.txt

Ile Ala Asp Leu Ile Thr Thr Cys Leu Gly Gly Arg Asn His Lys Cys
290 295 300

Ala Val Ala Phe Val Lys Thr Gly Lys Pro Met His Val Val Glu Gln
305 310 315 320

Glu Leu Leu Asp Gly Gln Lys Leu Gln Gly Ala Ala Thr Ala Lys Glu
325 330 335

Val Tyr Glu Phe Leu Asp Asn Gln Asn Lys Val Ser Glu Phe Pro Leu
340 345 350

Phe Thr Ala Val Tyr Arg Ile Val Tyr Glu Gly Leu Pro Pro Asn Lys
355 360 365

Leu Leu Glu Ala Ile
370

<210> 82
<211> 1014
<212> DNA
<213> artificial sequence

<220>
<223> Mutant of Pseudomonas fluorescens ilvC coding region

<400> 82
atgaagtggt tttacgataa agactgcat ctgagcatca tccagggaaa gaaggttgct 60
attataggat atggttccca aggacacgca caagccttga acttgaaaga ttctggggctc 120
gacgtgacag taggtctgta taaaggtgct gctgatgcag caaaggctga agcacatggc 180
tttaaagtca cagatgttgc agcggctgtt gctggcgctg atttagtcat gattttaatt 240
ccagatgaat ttcaatcgca attgtacaaa aatgaaatag aaccaaacad taagaagggc 300
gctaccttgg ccttcagtca tggatttgcc attcattaca atcaagtagt cccaggggca 360
gatttggacg ttattatgat tgcacctaag gctccggggc atactgttag gagcgaattt 420
gttaaggggtg gtggtattcc agatttgatc gctatatacc aagacgttag cggaaacgct 480
aagaatgtag ctttaagcta cgcagcagga gttggtggcg ggagaacggg tataatagaa 540
accactttta aagacgagac tgagacagat ttatttggag aacaagcggg tctgtgcgga 600
ggaactgttg aattgggtta agcaggcttt gagacgcttg tcgaagcagg gtacgctccc 660
gaaatggcat acttcgaatg tctacatgaa ttgaagttag tagtagactt aatgtatgaa 720
ggtggtatag ctaatatgaa ctattccatt tcaaataatg cagaatatgg tgagtatgtc 780
accggacctg aagtcattaa cgcagaatca agacaagcca tgagaaatgc cttgaaacgt 840
atccaggacg gtgaatacgc taagatgttc ataagtgaag gcgctacggg ttaccggagt 900
atgactgcta aaagaagaaa caatgcagca catggtatcg aaattattgg tgaacagtta 960
aggtctatga tgccctggat cggtgcta ataatcgtag acaaggcgaa aaat 1014

<210> 83

<211> 338

<212> PRT

<213> artificial sequence

<220>

<223> mutant of Pseudomonas fluorescens protein

<400> 83

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 Tyr 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 Tyr Lys
35 40 45

Gly Ala Ala Asp Ala 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 Ala 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

20110907_CL5178USNA_SeqList.txt

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

Ala Met Arg Asn Ala Leu Lys Arg Ile Gln Asp Gly Glu Tyr Ala Lys
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 Leu
305 310 315 320

Arg Ser Met Met Pro Trp Ile Gly Ala Asn Lys Ile Val Asp Lys Ala
325 330 335

Lys Asn

<210> 84
<211> 1644
<212> DNA
<213> artificial sequence

<220>
<223> Bacillus subtilis kivD coding region codon optimized for
expression in S. cerevisiae

<400> 84
atgtatacag taggtgacta tctgttggac agattacacg aattaggtat agaagaaata 60
ttcggagtac caggtgacta caatttgcaa tttctagatc aaattatttc acacaaagat 120
atgaaatggg tgggaaatgc taatgagttaaatgcctcct atatggccga cgggtacgca 180
agaacgaaaa aggctgcggc attcttgact acatttggtg ttggcgaatt atccgcagtt 240
aatggcttag cgggctccta tgctgagaac ctgcctgttg ttgagatcgt gggatctcct 300
acctcgaaag tgcagaacga aggtaagttt gttcaccata cgttggctga tgggtatttc 360
aagcacttta tgaagatgca cgaaccggtt actgctgccg ggactttatt gacagccgag 420
aatgcaactg ttgaaattga tagagtgttg tctgccttac taaaggaaag aaagccggtt 480
tacatcaatt tacctgtaga tgtagctgcc gctaaggctg aaaaaccatc cttgcctcct 540
aagaaggaaa attccacgtc gaatacatct gatcaagaga ttctgaacaa aatacaggaa 600
agtctgaaga atgccaagaa accaattgta atcacaggcc atgaaattat atcggttcggc 660
ctagagaaga ctgttactca gtttatttca aagactaagt tacctattac tactttgaac 720
tttggtaaat catctgttga tgaagcattg ccctcatttt tggggattta caacggtact 780
ctgtcagagc caaacttgaa ggaatttgtg gaatctgctg attttattct tatgttgggt 840


```

gtaaagctta ccgattctag tacgggtgca ttactcacc atcttaatga aaataaaatg      900
atttccttga atatcgatga aggtaaaatt ttcaacgaaa gaatccaaaa tttcgacttc      960
gaatccctga tatcatctct tcttgacttg tccgaaattg aatataaagg caagtacata     1020
gataaaaagc aagaagattt tgtaccttct aacgcgctgt tgtcacaaga tagactgtgg     1080
caagctgtcg aaaatttgac ccaaagtaat gagacgatcg tggctgaaca aggcaattct     1140
ttcttcggtg cctcatctat atttctgaaa tcgaaatcac attttatttg tcaacccttg     1200
tggggatcta taggatacac tttccccgca gctctaggca gccaaattgc agataaagaa     1260
tctagacatt tattgtttat cggagatgga tcattgcaac tgactgtcca agaattagga     1320
ctagcatta gagagaagat aaaccaatc tgctttatca ttaataacga tggttacacg      1380
gttgagaggg aaattcatgg tccgaaccag agttataatg acattcctat gtggaattac     1440
tcaaaactgc cagaaagttt cggggcaacg gaagacagag ttgtgtccaa aattgtgaga     1500
acagaaaatg aattcgtatc cgtgatgaaa gaagctcaag cagatccaaa taggatgtat     1560
tggatagaac ttattctagc aaaggagggt gcacctaaag ttttgaaaaa gatgggtaag     1620
ttatttgcag aacaaaacaa gagc                                           1644

```

```

<210> 85
<211> 548
<212> PRT
<213> Bacillus subtilis

```

```
<400> 85
```

```
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
50          55          60
```

```
Ala Ala Ala Phe Leu Thr Thr Phe Gly Val Gly Glu Leu Ser Ala Val
65          70          75          80
```

```
Asn Gly Leu Ala Gly Ser Tyr Ala Glu Asn Leu Pro Val Val Glu Ile
85          90          95
```

```
Val Gly Ser Pro Thr Ser Lys Val Gln Asn Glu Gly Lys Phe Val His
100         105         110
```

```
His Thr Leu Ala Asp Gly Asp Phe Lys His Phe Met Lys Met His Glu
115         120         125
```

20110907_CL5178USNA_SeqList.txt

Pro Val Thr Ala Ala Arg Thr Leu Leu Thr Ala Glu Asn Ala Thr Val
130 135 140

Glu Ile Asp Arg Val Leu Ser Ala Leu Leu Lys Glu Arg Lys Pro Val
145 150 155 160

Tyr Ile Asn Leu Pro Val Asp Val Ala Ala Ala Lys Ala Glu Lys Pro
165 170 175

Ser Leu Pro Leu Lys Lys Glu Asn Ser Thr Ser Asn Thr Ser Asp Gln
180 185 190

Glu Ile Leu Asn Lys Ile Gln Glu Ser Leu Lys Asn Ala Lys Lys Pro
195 200 205

Ile Val Ile Thr Gly His Glu Ile Ile Ser Phe Gly Leu Glu Lys Thr
210 215 220

Val Thr Gln Phe Ile Ser Lys Thr Lys Leu Pro Ile Thr Thr Leu Asn
225 230 235 240

Phe Gly Lys Ser Ser Val Asp Glu Ala Leu Pro Ser Phe Leu Gly Ile
245 250 255

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

Ala Asp Phe Ile Leu Met Leu Gly Val Lys Leu Thr Asp Ser Ser Thr
275 280 285

Gly Ala Phe Thr His His Leu Asn Glu Asn Lys Met Ile Ser Leu Asn
290 295 300

Ile Asp Glu Gly Lys Ile Phe Asn Glu Arg Ile Gln Asn Phe Asp Phe
305 310 315 320

Glu Ser Leu Ile Ser Ser Leu Leu Asp Leu Ser Glu Ile Glu Tyr Lys
325 330 335

Gly Lys Tyr Ile Asp Lys Lys Gln Glu Asp Phe Val Pro Ser Asn Ala
340 345 350

Leu Leu Ser Gln Asp Arg Leu Trp Gln Ala Val Glu Asn Leu Thr Gln
355 360 365

Ser Asn Glu Thr Ile Val Ala Glu Gln Gly Thr Ser Phe Phe Gly Ala
370 375 380

Ser Ser Ile Phe Leu Lys Ser Lys Ser His Phe Ile Gly Gln Pro Leu
385 390 395 400

20110907_CL5178USNA_SeqList.txt

Trp Gly Ser Ile Gly Tyr Thr Phe Pro Ala Ala Leu Gly Ser Gln Ile
405 410 415

Ala Asp Lys Glu Ser Arg His Leu Leu Phe Ile Gly Asp Gly Ser Leu
420 425 430

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

Pro Ile Cys Phe Ile Ile Asn Asn Asp Gly Tyr Thr Val Glu Arg Glu
450 455 460

Ile His Gly Pro Asn Gln Ser Tyr Asn Asp Ile Pro Met Trp Asn Tyr
465 470 475 480

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> 86
<211> 1125
<212> DNA
<213> artificial sequence

<220>
<223> horse ADH coding region codon optimized for *S. cerevisiae*
expression

<400> 86
atgtcaacag ccggtaaagt tattaagtgt aaagcggcag ttttgtggga agagaaaaag 60
ccgttttagca 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
ttagagtct gcaaacatcc tgagggtaat ttctgcctta aaaatgattt gtctatgcct 360
agaggtaacta tgcaggatgg tacaagcaga ttacatgca gagggaaacc tatacaccat 420
ttccttggtta cttctacatt ttccaatac acagtgggtg acgagatatc tgctgctaaa 480
atcgatgcag cttcaccact ggaaaaagtt tgcttgatag ggtgcggatt ttccaccggt 540

20110907_CL5178USNA_SeqList.txt

tacgggtccg cagttaaagt tgcaaagggt acacaggggt cgacttggtc agtattcgggt	600
ttaggaggag taggactaag cggtattatg ggggtgtaaag ctgcaggcgc agcgaggatt	660
ataggtgtag acatcaataa ggacaaattht gcaaaagcta aggaggtcgg ggctactgaa	720
tgtgttaacc ctcaagatta taagaaacca atacaagaag tccttactga aatgtcaaac	780
gggtggagttg atttctcttt tgaagttata ggccgtcttg atactatgggt aactgcgttg	840
tcctgctgtc aagaggcata tggagtcagt gtgatcgtag gtgttcctcc tgattcacia	900
aatttgtcga tgaatcctat gctgttgcta agcggtcgta catggaaggg agctatattht	960
ggcggthttta agagcaagga tagtgthtcca aaacttgtht ccgactthtat ggcaagaag	1020
thtgctcttg atcctthtat tacacatgta ttgccattcg agaaaatcaa tgaagggttht	1080
gatttgthta gaagtggthga atctattcgt acaattthta cthtt	1125

<210> 87
 <211> 375
 <212> PRT
 <213> Equus caballus

<400> 87

Met Ser Thr Ala Gly Lys Val Ile Lys Cys Lys Ala Ala Val Leu Trp
1 5 10 15

Glu Glu Lys Lys Pro Phe Ser Ile Glu Glu Val Glu Val Ala Pro Pro
20 25 30

Lys Ala His Glu Val Arg Ile Lys Met Val Ala Thr Gly Ile Cys Arg
35 40 45

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

Ile Ala Gly His Glu Ala Ala Gly Ile Val Glu Ser Ile Gly Glu Gly
65 70 75 80

Val Thr Thr Val Arg Pro Gly Asp Lys Val Ile Pro Leu Phe Thr Pro
85 90 95

Gln Cys Gly Lys Cys Arg Val Cys Lys His Pro Glu Gly Asn Phe Cys
100 105 110

Leu Lys Asn Asp Leu Ser Met Pro Arg Gly Thr Met Gln Asp Gly Thr
115 120 125

Ser Arg Phe Thr Cys Arg Gly Lys Pro Ile His His Phe Leu Gly Thr
130 135 140

Ser Thr Phe Ser Gln Tyr Thr Val Val Asp Glu Ile Ser Val Ala Lys
145 150 155 160

20110907_CL5178USNA_SeqList.txt

Ile Asp Ala Ala Ser Pro Leu Glu Lys Val Cys Leu Ile Gly Cys Gly
165 170 175

Phe Ser Thr Gly Tyr Gly Ser Ala Val Lys Val Ala Lys Val Thr Gln
180 185 190

Gly Ser Thr Cys Ala Val Phe Gly Leu Gly Gly Val Gly Leu Ser Val
195 200 205

Ile Met Gly Cys Lys Ala Ala Gly Ala Ala Arg Ile Ile Gly Val Asp
210 215 220

Ile Asn Lys Asp Lys Phe Ala Lys Ala Lys Glu Val Gly Ala Thr Glu
225 230 235 240

Cys Val Asn Pro Gln Asp Tyr Lys Lys Pro Ile Gln Glu Val Leu Thr
245 250 255

Glu Met Ser Asn Gly Gly Val Asp Phe Ser Phe Glu Val Ile Gly Arg
260 265 270

Leu Asp Thr Met Val Thr Ala Leu Ser Cys Cys Gln Glu Ala Tyr Gly
275 280 285

Val Ser Val Ile Val Gly Val Pro Pro Asp Ser Gln Asn Leu Ser Met
290 295 300

Asn Pro Met Leu Leu Leu Ser Gly Arg Thr Trp Lys Gly Ala Ile Phe
305 310 315 320

Gly Gly Phe Lys Ser Lys Asp Ser Val Pro Lys Leu Val Ala Asp Phe
325 330 335

Met Ala Lys Lys Phe Ala Leu Asp Pro Leu Ile Thr His Val Leu Pro
340 345 350

Phe Glu Lys Ile Asn Glu Gly Phe Asp Leu Leu Arg Ser Gly Glu Ser
355 360 365

Ile Arg Thr Ile Leu Thr Phe
370 375

<210> 88
<211> 1713
<212> DNA
<213> Streptococcus mutans

<400> 88
atgactgaca aaaaaactct taaagactta agaaatcgta gttctgttta cgattcaatg 60
gttaaatacac ctaatcgtgc tatgttgcgt gcaactggta tgcaagatga agactttgaa 120
aaacctatcg tcggtgtcat ttcaacttgg gctgaaaaca caccttgtaa tatccactta 180

20110907_CL5178USNA_SeqList.txt

catgactttg gtaaactagc caaagtcggg gtttaaggaag ctgggtgcttg gccagttcag	240
ttcggaaacaa tcacgggtttc tgatggaatc gccatgggaa cccaaggaat gcgttttctcc	300
ttgacatctc gtgatattat tgcagattct attgaagcag ccatgggagg tcataatgcg	360
gatgcttttg tagccattgg cggtttgatg aaaaacatgc ccggttctgt ttcgctatg	420
gctaacatgg atatcccagc catttttgct tacggcggaa caattgcacc tggtaattta	480
gacggcaaag atatcgattt agtctctgtc tttgaagggtg 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 ctgttgctaa aatgctcgaa atgggcttaa aaccttctga cattttaacg	780
cgtgaagctt ttgaagatgc tattactgta actatggctc tgggagggtt aaccaactca	840
accttcacc tcttagctat tgcccatgct gctaattgtg aattgacact tgatgatttc	900
aatactttcc aagaaaaagt tcctcatttg gctgatttga aaccttctgg tcaatatgta	960
ttccaagacc tttacaaggt cggaggggta ccagcagtta tgaaatatct ccttaaaaat	1020
ggcttccttc atggtgaccg tatcacttgt actggcaaaa cagtcgctga aaatttgaag	1080
gcttttgatg atttaacacc tgggtcaaaag gttattatgc cgcttgaaaa tcctaaacgt	1140
gaagatggtc cgctcattat tctccatggg aacttggctc cagacgggtc cgttgccaaa	1200
gtttctggtg taaaagtgcg tcgtcatgtc ggtcctgcta aggtctttta ttctgaagaa	1260
gaagccattg aagctgtctt gaatgatgat attgttgatg gtgatgttgt tgtcgtacgt	1320
ttttagtagc caaagggcgg tcctgggtatg cctgaaatgc tttccctttc atcaatgatt	1380
gttggtaaaag ggcaagggtg aaaagttgcc cttctgacag atggccgctt ctcagggtgt	1440
acttatggtc ttgtcgtggg tcatatcgct cctgaagcac aagatggcgg tccaatcgcc	1500
tacctgcaaa caggagacat agtcactatt gaccaagaca ctaaggaatt acactttgat	1560
atctccgatg aagagttaaa acatcgtaaa gagaccattg aattgccacc gctctattca	1620
cgcggtatcc ttggtaaata tgctcacatc gtttcgtctg cttctagggg agccgtaaca	1680
gacttttgga agcctgaaga aactggcaaa aaa	1713

<210> 89

<211> 571

<212> PRT

<213> Streptococcus mutans

<400> 89

Met	Thr	Asp	Lys	Lys	Thr	Leu	Lys	Asp	Leu	Arg	Asn	Arg	Ser	Ser	Val
1				5					10					15	

Tyr	Asp	Ser	Met	Val	Lys	Ser	Pro	Asn	Arg	Ala	Met	Leu	Arg	Ala	Thr
			20					25					30		

20110907_CL5178USNA_SeqList.txt

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

Cys Asp Lys Asn Met Pro Gly Ser Val Ile Ala Met Ala Asn Met Asp
130 135 140

Ile Pro Ala Ile Phe Ala Tyr Gly Gly Thr Ile Ala Pro Gly Asn Leu
145 150 155 160

Asp Gly Lys Asp Ile Asp Leu Val Ser Val Phe Glu Gly Val Gly His
165 170 175

Trp Asn His Gly Asp Met Thr Lys Glu Glu Val Lys Ala Leu Glu Cys
180 185 190

Asn Ala Cys Pro Gly Pro Gly Gly Cys Gly Gly Met Tyr Thr Ala Asn
195 200 205

Thr Met Ala Thr Ala Ile Glu Val Leu Gly Leu Ser Leu Pro Gly Ser
210 215 220

Ser Ser His Pro Ala Glu Ser Ala Glu Lys Lys Ala Asp Ile Glu Glu
225 230 235 240

Ala Gly Arg Ala Val Val Lys Met Leu Glu Met Gly Leu Lys Pro Ser
245 250 255

Asp Ile Leu Thr Arg Glu Ala Phe Glu Asp Ala Ile Thr Val Thr Met
260 265 270

Ala Leu Gly Gly Ser Thr Asn Ser Thr Leu His Leu Leu Ala Ile Ala
275 280 285

His Ala Ala Asn Val Glu Leu Thr Leu Asp Asp Phe Asn Thr Phe Gln
290 295 300

20110907_CL5178USNA_SeqList.txt

Glu Lys Val Pro His Leu Ala Asp Leu Lys Pro Ser Gly Gln Tyr Val
305 310 315 320

Phe Gln Asp Leu Tyr Lys Val Gly Gly Val Pro Ala Val Met Lys Tyr
325 330 335

Leu Leu Lys Asn Gly Phe Leu His Gly Asp Arg Ile Thr Cys Thr Gly
340 345 350

Lys Thr Val Ala Glu Asn Leu Lys Ala Phe Asp Asp Leu Thr Pro Gly
355 360 365

Gln Lys Val Ile Met Pro Leu Glu Asn Pro Lys Arg Glu Asp Gly Pro
370 375 380

Leu Ile Ile Leu His Gly Asn Leu Ala Pro Asp Gly Ala Val Ala Lys
385 390 395 400

Val Ser Gly Val Lys Val Arg Arg His Val Gly Pro Ala Lys Val Phe
405 410 415

Asn Ser Glu Glu Glu Ala Ile Glu Ala Val Leu Asn Asp Asp Ile Val
420 425 430

Asp Gly Asp Val Val Val Val Arg Phe Val Gly Pro Lys Gly Gly Pro
435 440 445

Gly Met Pro Glu Met Leu Ser Leu Ser Ser Met Ile Val Gly Lys Gly
450 455 460

Gln Gly Glu Lys Val Ala Leu Leu Thr Asp Gly Arg Phe Ser Gly Gly
465 470 475 480

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
565 570

<210> 90
 <211> 38
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> primer BK468

 <400> 90
 gcctcgagtt ttaatgttac ttctcttgca gttagggga 38

<210> 91
 <211> 20
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> primer N98SeqF1

 <400> 91
 cgtgtagtc acatcaggac 20

<210> 92
 <211> 20
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> primer N98SeqF4

 <400> 92
 ggtttctgtc tctggtgacg 20

<210> 93
 <211> 22
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> primer N99SeqR2

 <400> 93
 catcgactgc attacgcaac tc 22

<210> 94
 <211> 22
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> primer N160SeqF5

 <400> 94
 cctgaagtct aggtccctat tt 22

<210> 95
 <211> 38
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> primer N191

<400> 95
atccgcggat agatctccca ttaccgacat ttggg'gcgc 38

<210> 96
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> primer N1110

<400> 96
gcgatttaat ctctaattat tagttaaagt 30

<210> 97
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> primer N1111

<400> 97
tatttgatc gaggtgtcta gtcttctatt 30

<210> 98
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> primer oBP512

<400> 98
aaagttggca tagcggaaac tt 22

<210> 99
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> primer oBP594

<400> 99
agctgtctcg tggtgtgggt tt 22

<210> 100
<211> 49
<212> DNA
<213> Artificial Sequence

<220>
<223> primer oBP595

<400> 100
cttaataata gaacaatatc atcctttacg ggcattctat agtgtcggt 49

<210> 101
<211> 49
<212> DNA
<213> Artificial Sequence

```

<220>
<223> primer oBP596

<400> 101
gcgccaacga cactataaga tgcccgtaaa ggatgatatt gttctatta 49

<210> 102
<211> 49
<212> DNA
<213> Artificial Sequence

<220>
<223> primer oBP597

<400> 102
tatggaccct gaaaccacag ccacattgca acgacgacaa tgccaaacc 49

<210> 103
<211> 49
<212> DNA
<213> Artificial Sequence

<220>
<223> primer oBP598

<400> 103
tccttggttt ggcattgtcg tcgttgcaat gtggctgtgg tttcagggt 49

<210> 104
<211> 49
<212> DNA
<213> Artificial Sequence

<220>
<223> primer oBP599

<400> 104
atcctctcgc ggagtccttg ttcagtaaag gccatgaagc tttttcttt 49

<210> 105
<211> 49
<212> DNA
<213> Artificial Sequence

<220>
<223> primer oBP600

<400> 105
attggaaaga aaaagcttca tggcctttac tgaacagggg ctccgag 49

<210> 106
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> primer oBP601

<400> 106
tcataccaca atcttagacc at 22

```

<210> 107
 <211> 21
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> primer oBP602

 <400> 107
 tgttcaaacc cctaaccaac c 21

<210> 108
 <211> 22
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> primer oBP603

 <400> 108
 tgttcccaca atctattacc ta 22

<210> 109
 <211> 21
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> primer oBP605

 <400> 109
 tactgaacag ggactccgcg a 21

<210> 110
 <211> 21
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> primer oBP606

 <400> 110
 tcataccaca atcttagacc a 21

<210> 111
 <211> 34
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> primer oBP562

 <400> 111
 aattgtttaa acatgtatac agtaggtgac tatc 34

<210> 112
 <211> 49
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> primer oBP563

<400> 112
aatcataaat cataagaaat tcgcttatca gctcttgttt tgttctgca 49

<210> 113
<211> 49
<212> DNA
<213> Artificial Sequence

<220>
<223> primer oBP564

<400> 113
ttatttgag aacaaaaaa gagctgataa gcgaatttct tatgattta 49

<210> 114
<211> 34
<212> DNA
<213> Artificial Sequence

<220>
<223> primer oBP565

<400> 114
aattggccgg ccaaaaaaag catgcacgta taca 34

<210> 115
<211> 32
<212> DNA
<213> Artificial Sequence

<220>
<223> primer oBP505

<400> 115
aattgagctc actgtagccc tagacttgat ag 32

<210> 116
<211> 35
<212> DNA
<213> Artificial Sequence

<220>
<223> primer oBP506

<400> 116
aattggcgcg cctgtatatg agatagttga ttgta 35

<210> 117
<211> 35
<212> DNA
<213> Artificial Sequence

<220>
<223> primer oBP507

<400> 117
aattttaatt aagtctaggt tctttggctg ttcaa 35

<210> 118
<211> 33
<212> DNA
<213> Artificial Sequence

```

<220>
<223> primer oBP508

<400> 118
aattgtcgcac tttagaagtg tcaacaacgt atc
33

<210> 119
<211> 34
<212> DNA
<213> Artificial Sequence

<220>
<223> primer oBP674

<400> 119
aattggcgcg ccaattaccg tcgctcgtga ttg
34

<210> 120
<211> 34
<212> DNA
<213> Artificial Sequence

<220>
<223> primer oBP675

<400> 120
aattgtttaa acttgaatat gtattacttg gtta
34

<210> 121
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> primer oBP495

<400> 121
ggagatatat aatagaacag at
22

<210> 122
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> primer oBP496

<400> 122
tagcaatggg gtttttttca gt
22

<210> 123
<211> 47
<212> DNA
<213> Artificial Sequence

<220>
<223> primer T-FBA1(SalI)

<400> 123
attctacgta cgtcgacgcc tacttggtt cacatacgtt gcatacg
47

```

<210> 124
 <211> 49
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> primer B-FBA1(SpeI)

 <400> 124
 gtatcaaata ctagtttgaa tatgtattac ttggttatgg ttatatatg 49

 <210> 125
 <211> 43
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> primer T-U/PGK1(KpnI)

 <400> 125
 actacagatg gtaccaatta ccgtcgctcg tgatttgttt gca 43

 <210> 126
 <211> 45
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> primer B-U/PGK1(SalI)

 <400> 126
 agcatccttg tcgacaggac cttgttgtgt gacgaaattg gaagc 45

 <210> 127
 <211> 22
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> primer oBP556

 <400> 127
 tattttcgag gaccttgtca cc 22

 <210> 128
 <211> 25
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> primer oBP561

 <400> 128
 tggccattaa tctttcccat attag 25

 <210> 129
 <211> 725
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> UAS(PGK1)-FBA1p

20110907_CL5178USNA_SeqList.txt

<400> 129
aattaccgtc gctcgtgatt tgtttgcaaa aagaacaaaa ctgaaaaaac ccagacacgc 60
tcgacttcct gtcttcctat tgattgcagc ttccaatttc gtcacacaac aaggctcctgt 120
cgacgcctac ttggcttcac atacgttgca tacgtcgata tagataataa tgataatgac 180
agcaggatta tcgtaatacg taatagttga aaatctcaaa aatgtgtggg tcattacgta 240
aataatgata ggaatgggat tcttctatctt ttcctttttc cattctagca gccgtcggga 300
aaacgtggca tcctctcttt cgggctcaat tggagtcacg ctgccgtgag catcctctct 360
ttccatatct aacaactgag cacgtaacca atggaaaagc atgagcttag cgttgctcca 420
aaaaagtatt ggatgggttaa taccatttgt ctgttctctt ctgactttga ctctcaaaa 480
aaaaaaaaatc tacaatcaac agatcgcttc aattacgccc tcacaaaaac ttttttcctt 540
cttcttcgcc cacgttaaat tttatccctc atgttgtcta acggatttct gcacttgatt 600
tattataaaa agacaaagac ataatacttc tctatcaatt tcagttattg ttcttccttg 660
cgttattctt ctgttcttct ttttcttttg tcatatataa ccataaccaa gtaatacata 720
ttcaa 725

<210> 130
<211> 10494
<212> DNA
<213> Artificial Sequence

<220>
<223> pWS358-PGK1p-GUS

<400> 130
tcgcgcgttt cggatgatgac ggtgaaaacc tctgacacat gcagctcccg gagacggtca 60
cagcttgtct gtaagcggat gccgggagca gacaagcccg tcagggcgcg tcagcgggtg 120
ttggcgggtg tcggggctgg cttactatg cggcatcaga gcagattgta ctgagagtgc 180
accataaatt cccgttttaa gagcttggtg agcgctagga gtcactgcca ggtatcgttt 240
gaacacggca ttagtcaggg aagtcataac acagtccttt cccgcaattt tctttttcta 300
ttactcttgg cctcctctag tacactctat atttttttat gcctcggtaa tgattttcat 360
tttttttttt cccctagcgg atgactcttt ttttttctta gcgattggca ttatcacata 420
atgaattata cattatataa agtaatgtga tttcttcgaa gaataacta aaaaatgagc 480
aggcaagata aacgaaggca aagatgacag agcagaaaagc cctagtaaag cgtattacaa 540
atgaaaccaa gattcagatt gcgatctctt taaaggggtg tcccctagcg atagagcact 600
cgatcttccc agaaaaagag gcagaagcag tagcagaaca ggccacacaa tcgcaagtga 660
ttaacgtcca cacaggtata gggtttcttg accatatgct agggattcat aaccattttc 720
tcaatcgaat tacacagaac acaccgtaca aacctctcta tcataactac ttaatagtca 780
cacacgtact cgtctaaata cacatcatcg tcctacaagt tcatcaaagt gttggacaga 840
caactatacc agcatggatc tcttgatcg gttcttttct cccgctctct cgcaataaca 900
atgaacactg ggtcaatcat agcctacaca ggtgaacaga gtagcgttta tacagggttt 960

20110907_CL5178USNA_SeqList.txt

atacggtgat	tcctacggca	aaaatTTTTc	atttctaaaa	aaaaaaagaa	aaatTTTTct	1020
ttccaacgct	agaaggaaaa	gaaaaatcta	attaaattga	tttggtgatt	ttctgagagt	1080
tccctTTTTc	atataatcgaa	ttttgaatat	aaaaggagat	cgaaaaaatt	tttctattca	1140
atctgTTTTc	tggttttatt	tgatagtttt	tttgtgtatt	attattatgg	attagtactg	1200
gtttatatgg	gtttttctgt	ataacttctt	tttatttttag	tttgtttaat	cttattttga	1260
gttacattat	agttccctaa	ctgcaagaga	agtaacatta	aaactcgaga	tgggtaagga	1320
aaagactcac	gtttcgaggc	cgcgattaaa	ttccaacatg	gatgctgatt	tatatgggta	1380
taaatgggct	cgcgataatg	tcgggcaatc	aggtgcgaca	atctatcgat	tgtatgggaa	1440
gcccgatgcg	ccagagttgt	ttctgaaaca	tggtcaaagg	agcgttgcca	atgatgttac	1500
agatgagatg	gtcagactaa	actggctgac	ggaatttatg	cctcttccga	ccatcaagca	1560
ttttatccgt	actcctgatg	atgcatgggt	actcaccact	gcatccccg	gcaaaacagc	1620
attccaggta	ttagaagaat	atcctgattc	aggtgaaaat	attgttgatg	cgctggcagt	1680
gttcctgcg	cggttgcat	cgattcctgt	ttgtaattgt	ccttttaaca	gcatcgcg	1740
atttcgtctc	gtcaggcg	aatcacgaat	gaataacggt	ttggttgatg	cgagtgttt	1800
tgatgacgag	cgtaatggct	ggcctgttga	acaagtctgg	aaagaaatgc	ataagctttt	1860
gccatttctca	ccggattcag	tcgtcactca	tggtgatttc	tcacttgata	accttatttt	1920
tgacgagggg	aaattaatag	gttgatttga	tggttgacga	gtcggaatcg	cagaccgata	1980
ccaggatctt	gccatcctat	ggaactgcct	cggtgagttt	tctccttcat	tacagaaacg	2040
gctttttcaa	aaatatggta	ttgataatcc	tgatatgaat	aaattgcagt	ttcatttgat	2100
gctcgatgag	tttttctaag	tttaacttga	tactactaga	ttttttctct	tcatttataa	2160
aatttttggt	tataattgaa	gctttagaag	tatgaaaaaa	tccttttttt	tcattctttg	2220
caaccaaagt	aagaagcttc	ttttattcat	tgaaatgatg	aatataaacc	taacaaaaga	2280
aaaagactcg	aatatcaaac	attaaaaaaa	aataaaagag	gttatctgtt	ttcccattta	2340
gttgaggttt	gcattttcta	atagatagaa	ctctcaatta	atgtggattt	agtttctctg	2400
ttcgctgcag	catacgatat	atatacatgt	gtatatatgt	ataacctatga	atgtcagtaa	2460
gtatgtatac	gaacagtatg	atactgaaga	tgacaaggta	atgcatcatt	ctatacgtgt	2520
cattctgaac	gaggcgcgct	ttcctttttt	ctttttgctt	tttctttttt	tttctcttga	2580
actcgacgga	tctatgcggt	gtgaaatacc	gcacagatgc	gtaaggagaa	aataccgcat	2640
caggaaattg	taaacgttaa	tattttgtta	aaattcgctg	taaatttttg	ttaaatcagc	2700
tcatttttta	accaataggc	cgaaatcggc	aaaatccctt	ataaatcaaa	agaatagacc	2760
gagatagggg	tgagtgttgt	tccagtttgg	aacaagagtc	cactattaaa	gaacgtggac	2820
tccaacgtca	aaggcgcaaa	aaccgtctat	cagggcgatg	gccactacg	tgaaccatca	2880
ccctaataca	gttttttggg	gtcgaggtgc	cgtaaagcac	taaatcggaa	ccctaaaggg	2940
agcccccgat	ttagagcttg	acgggggaaag	ccggcgaacg	tggtcgagaaa	ggaagggaag	3000

20110907_CL5178USNA_SeqList.txt

aaagcgaaag	gagcggggcg	tagggcgctg	gcaagtgtag	cggtcacgct	gcgcgtaacc	3060
accacacccg	ccgcgcgttaa	tgcgccgcta	cagggcgcgct	cgcgccattc	gccattcagg	3120
ctgcgcaact	gttggaag	gcgatcgggtg	cgggcctctt	cgctattacg	ccagctggcg	3180
aaaggggggat	gtgctgcaag	gcgattaagt	tgggtaacgc	caggggttttc	ccagtcacga	3240
cgttgtaaaa	cgacggccag	tgagcgcgcg	taatacgact	cactataggg	cgaattgggt	3300
accgggcccc	ccctcgaggt	cgacgtgagt	aaggaaagag	tgaggaacta	tcgcatacct	3360
gcattttaag	atgccgattt	gggcgcgaat	cctttatttt	ggcttcaccc	tcatactatt	3420
atcagggcca	gaaaaaggaa	gtgtttccct	ccttcttgaa	ttgatgttac	cctcataaag	3480
cacgtggcct	cttatcgaga	aagaaattac	cgtcgctcgt	gatttgtttg	caaaaagaac	3540
aaaactgaaa	aaaccagac	acgctcgact	tcctgtcttc	ctattgattg	cagcttccaa	3600
tttcgtcaca	caacaaggtc	ctagcgacgg	ctcacagggt	ttgtaacaag	caatcgaagg	3660
ttctggaatg	gcgggaaagg	gtttagtacc	acatgctatg	atgccactg	tgatctccag	3720
agcaaagttc	gttcgatcgt	actgttactc	tctctctttc	aaacagaatt	gtccgaatcg	3780
tgtgacaaca	acagcctgtt	ctcacacact	cttttcttct	aaccaagggg	gtggtttagt	3840
ttagtagaac	ctcgtgaaac	ttacatttac	atatatataa	acttgcataa	attgggtcaat	3900
gcaagaaata	catatttggt	cttttcta	tcgtagtttt	tcaagttctt	agatgctttc	3960
tttttctctt	ttttacagat	catcaaggaa	gtaattatct	actttttaca	acaaatataa	4020
aacaactagt	atggtacgtc	ctgtagaaac	cccaaccggt	gaaatcaaaa	aactcgacgg	4080
cctgtgggca	ttcagtcctgg	atcgcgaaaa	ctgtggaatt	gatcagcggt	ggtagggaaa	4140
cgcgttacaa	gaaagccggg	caattgctgt	gccaggcagt	tttaacgatc	agttcgccga	4200
tgcagatatt	cgtaattatg	cgggcaacgt	ctggtatcag	cgcgaagtct	ttataccgaa	4260
aggttgggca	ggccagcgta	tcgtgctgcg	tttcgatgcg	gtcactcatt	acggcaaaagt	4320
gtgggtcaat	aatcaggaag	tgatggagca	tcagggcggc	tatactccat	ttgaagccga	4380
tgtcacgccg	tatgttattg	ccgggaaaag	tgtacgtatc	accgtttgtg	tgaacaacga	4440
actgaactgg	cagactatcc	cgccgggaat	ggtgattacc	gacgaaaacg	gcaagaaaaa	4500
gcagtcttac	ttccatgatt	tctttaacta	tgccggaatc	catcgcagcg	taatgctcta	4560
caccacgccg	aacacctggg	tggacgatat	caccgtgggtg	acgcatgtcg	cgcaagactg	4620
taaccacgcg	tctgttgact	ggcaggtgggt	ggccaatgggt	gatgtcagcg	ttgaactgcg	4680
tgatgcggat	caacaggtgg	ttgcaactgg	acaaggcact	agcgggactt	tgcaagtgggt	4740
gaatccgcac	ctctggcaac	cgggtgaagg	ttatctctat	gaactgtgcg	tcacagccaa	4800
aagccagaca	gagtgtgata	tctaccgcgt	tcgcgtcggc	atccggtcag	tggcagtgaa	4860
gggcgaacag	ttcctgatta	accacaaacc	gttctacttt	actggctttg	gtcgtcatga	4920
agatgcggac	ttgcgtggca	aaggattcga	taacgtgctg	atggtgcacg	accacgcatt	4980
aatggactgg	attggggcca	actcctaccg	tacctcgc	taccttacg	ctgaagagat	5040

20110907_CL5178USNA_SeqList.txt

gctcgactgg	gcagatgaac	atggcatcgt	ggtgattgat	gaaactgctg	ctgtcggctt	5100
taacctctct	ttaggcattg	gtttcgaagc	gggcaacaag	ccgaaagaac	tgtacagcga	5160
agaggcagtc	aacggggaaa	ctcagcaagc	gcacttacag	gcgattaaag	agctgatagc	5220
gcgtgacaaa	aaccacccaa	gcgtggtgat	gtggagtatt	gccaacgaac	cggatacccg	5280
tccgcaaggt	gcacgggaat	atttcgcgcc	actggcggaa	gcaacgcgta	aactcgaccc	5340
gacgcgtccg	atcacctgcg	tcaatgtaat	gttctgcgac	gctcacaccg	ataccatcag	5400
cgatctcttt	gatgtgctgt	gcctgaaccg	ttattacgga	tggtatgtcc	aaagcggcga	5460
tttgaaacg	gcagagaagg	tactggaaaa	agaacttctg	gcctggcagg	agaaactgca	5520
tcagccgatt	atcatcaccg	aatacggcgt	ggatacgtta	gccgggctgc	actcaatgta	5580
caccgacatg	tggagtgaag	agtatcagtg	tgcatggctg	gatatgtatc	accgcgtctt	5640
tgatcgcgtc	agcgccgtcg	tcggtgaaca	ggtatggaat	ttcgccgatt	ttgcgacctc	5700
gcaaggcata	ttgcgcgttg	gcggtaacaa	gaaagggatc	ttcactcgcg	accgcaaacc	5760
gaagtcggcg	gcttttctgc	tgcaaaaacg	ctggactggc	atgaacttcg	gtgaaaaacc	5820
gcagcaggga	ggcaaacaat	gattaattaa	ctagagcggc	cgcgtaatt	caaattaatt	5880
gatatagttt	tttaatgagt	attgaatctg	tttagaaata	atggaatatt	atttttat	5940
atttat	attattggtc	ggctcttttc	ttctgaagg	caatgacaaa	atgatatgaa	6000
ggaaataatg	atttctaaaa	ttttacaacg	taagatat	ttacaaaagc	ctagctcatc	6060
ttttgtcatg	cactat	ctcacgcttg	aaattaacg	ccagtcact	gcggagtc	6120
ttcaaagtca	tccta	tctatcgttt	ttgatagctc	at	tttgag	6180
cttctgttat	tcacaactgt	ttta	atttcattct	ggaactcttc	gagttctttg	6240
taaagtcttt	catagtagct	tactttatcc	tccaacatat	ttaa	gtcaat	6300
gctcttaaat	tttccacatc	atcaagttca	acatcatctt	ttaa	tttattctct	6360
agctcttcca	accaagcctc	attgctcctt	gatttactgg	tgaaaagtga	tacactttgc	6420
gcgcaatcca	ggtcaaaact	ttcctgcaaa	gaattcacca	at	atcatagtac	6480
aatttg	gttctcccat	cacaatttaa	tatacctgat	ggattcttat	gaagcgtgg	6540
gtaatggacg	tgtcactcta	cttcgccttt	ttccctactc	cttttagtac	ggaagacaat	6600
gctaataaat	aagagggtaa	taataatatt	attaatcggc	aaaaaagatt	aaacgccaa	6660
cgtttaatta	tcagaaagca	aacgtcgtac	caatccttga	atgcttcca	attgtatatt	6720
aagagtc	acagcaacat	attcttg	ttaaattaat	tattattgat	ttttgatatt	6780
gtataaaaaa	accaa	tataaaaaa	gtgaataaaa	aataccaagt	atggagaa	6840
atattagaag	tctatacgtt	aaaccaccgc	ggtggagctc	cagctttt	tccttttagt	6900
gagggtta	tgcgcgcttg	gcgtaatcat	ggtcatagct	gtttcctgtg	tgaaattgtt	6960
atccgctcac	aattccacac	aacataggag	ccggaagcat	aaagtgtaaa	gcctggggtg	7020
ccta	gaggt	acattaattg	cgttgcgctc	actgcccgct	ttccagtcgg	7080

20110907_CL5178USNA_SeqList.txt

gaaacctgtc	gtgccagctg	cattaatgaa	tcggccaacg	cgcggggaga	ggcggtttgc	7140
gtattgggcg	ctcttcgct	tcctcgctca	ctgactcgct	gcgctcggtc	gttcgggctgc	7200
ggcgagcggg	atcagctcac	tcaaaggcgg	taatacgggt	atccacagaa	tcaggggata	7260
acgcaggaaa	gaacatgtga	gcaaaaggcc	agcaaaaggc	caggaaccgt	aaaaaggccg	7320
cgttgctggc	gtttttccat	aggctccgcc	cccctgacga	gcatacaaaa	aatcgacgct	7380
caagtcagag	gtggcgaaac	ccgacaggac	tataaagata	ccaggcgttt	ccccctggaa	7440
gctccctcgt	gcgctctcct	gttccgaccc	tgccgcttac	cggataacctg	tccgcctttc	7500
tcccttcggg	aagcgtggcg	ctttctcata	gctcacgctg	taggtatctc	agttcgggtg	7560
aggctcgttcg	ctccaagctg	ggctgtgtgc	acgaaccccc	cgttcagccc	gaccgctgcg	7620
ccttatccgg	taactatcgt	cttgagtcca	acccggtaag	acacgactta	tcgccactgg	7680
cagcagccac	tggtaacagg	attagcagag	cgaggtatgt	aggcggtgct	acagagttct	7740
tgaagtgggtg	gcctaactac	ggctacacta	gaaggacagt	atttggtatc	tgcgctctgc	7800
tgaagccagt	taccttcgga	aaaagagttg	gtagctcttg	atccggcaaa	caaaccaccg	7860
ctggtagcgg	tggttttttt	gtttgcaagc	agcagattac	gcgcagaaaa	aaaggatctc	7920
aagaagatcc	tttgatcttt	tctacggggg	ctgacgctca	gtggaacgaa	aactcacggt	7980
aagggatttt	ggcatgaga	ttatcaaaaa	ggatcttcac	ctagatcctt	ttaaattaaa	8040
aatgaagttt	taaatcaatc	taaagtatat	atgagtaaac	ttggtctgac	agttaccaat	8100
gcttaatcag	tgaggcacct	atctcagcga	tctgtctatt	tcgttcaccc	atagttgcct	8160
gactccccgt	cgtgtagata	actacgatac	gggagggctt	accatctggc	cccagtgctg	8220
caatgatacc	gcgagacca	cgctcaccgg	ctccagattt	atcagcaata	aaccagccag	8280
ccggaagggc	cgagcgcaga	agtggctctg	caactttatc	cgctccatc	cagtctatta	8340
attgttgccg	ggaagctaga	gtaagtagtt	cgccagttaa	tagtttgcg	aacgttggtg	8400
ccattgctac	aggcatcgtg	gtgtcacgct	cgctggtttg	tatggcttca	ttcagctccg	8460
gttcccaacg	atcaaggcga	gttacatgat	cccccatggt	gtgcaaaaaa	gcggttagct	8520
ccttcggtcc	tccgatcgtt	gtcagaagta	agttggccgc	agtgttatca	ctcatggtta	8580
tggcagcact	gcataattct	cttactgtca	tgccatccgt	aagatgcttt	tctgtgactg	8640
gtgagtactc	aaccaagtca	ttctgagaat	agtgtatgcg	gcgaccgagt	tgctcttgcc	8700
cggcgtcaat	acgggataat	accgcgccac	atagcagaac	tttaaaagtg	ctcatcattg	8760
gaaaacgttc	ttcggggcga	aaactctcaa	ggatcttacc	gctgttgaga	tccagttcga	8820
tgaaccac	tcgtgcaccc	aactgatctt	cagcatcttt	tactttcacc	agcgtttctg	8880
ggtgagcaaa	aacaggaagg	caaatgccg	caaaaaagg	aataagggcg	acacggaaat	8940
gttgaatact	catactcttc	ctttttcaat	attattgaag	catttatcag	ggttattgtc	9000
tcatgagcgg	atacatattt	gaatgtattt	agaaaaataa	acaaatagg	gttccgcgca	9060
catttccccg	aaaagtgcc	cctgaacgaa	gcatactgtg	ttcattttgt	agaacaaaaa	9120

20110907_CL5178USNA_SeqList.txt

tgcaacgcga	gagcgcta	at	tttcaaaca	aagaatctga	gctgcatttt	tacagaacag	9180
aaatgcaacg	cgaaagcgct	at	ttttaccaa	cgaagaatct	gtgcttcatt	tttgtaaaac	9240
aaaaatgcaa	cgcgagagcg	cta	at	tttttc	aaacaaagaa	tctgagctgc	9300
aacagaaatg	caacgcgaga	gcgctat	tttt	accaacaaag	aatctatact	tcttttttgt	9360
tctacaaaaa	tgcattccga	gagcgctat	tttctaaca	agcatcttag	attacttttt		9420
ttctcctttg	tgcgctctat	aatgcagtct	cttgataact	ttttgactg	taggtccgtt		9480
aaggtagaa	gaaggctact	ttggtgtcta	ttttctcttc	cataaaaaaa	gcctgactcc		9540
acttcccgcg	tttactgatt	actagcgaa	ctgcgggtgc	at	ttttttcaa	gataaaggca	9600
tccccgatta	tattctatac	cgatgtggat	tgcgcatact	ttgtgaacag	aaagtgatag		9660
cgttgatgat	tcttcattgg	tcagaaaatt	atgaacggtt	tcttctat	tt	tgtctctata	9720
tactacgtat	aggaaatgtt	tacatttttcg	tattgttttc	gattcactct	atgaatagtt		9780
cttactacaa	tttttttgtc	taaagagtaa	tactagagat	aaacataaaa	aatgtagagg		9840
tcgagtttag	atgcaagttc	aaggagcgaa	aggtggatgg	gtaggttata	tagggatata		9900
gcacagagat	atatagcaaa	gagatacttt	tgagcaatgt	ttgtggaagc	ggtattcgca		9960
atatttttagt	agctcggtac	agtccggtgc	gtttttgggt	ttttgaaagt	gcgtcttcag		10020
agcgcttttg	gttttcaaaa	gcgctctgaa	gttcctatac	tttctagaga	ataggaactt		10080
cggaatagga	acttcaaagc	gtttccgaaa	acgagcgctt	ccgaaaatgc	aacgcgagct		10140
gcgcacatac	agctcactgt	tcacgtcgca	cctatatctg	cgtgttgcc	gtatatatat		10200
atacatgaga	agaacggcat	agtgcgtgtt	tatgcttaaa	tgcgactta	tatgcgtcta		10260
tttatgtagg	atgaaaggta	gtctagtacc	tcctgtgata	ttatccatt	ccatgcgggg		10320
tatcgtatgc	ttccttcagc	actacccttt	agctgttcta	tatgctgcca	ctcctcaatt		10380
ggattagtct	catccttcaa	tgctatcatt	tcctttgata	ttggatcatc	taagaaacca		10440
ttattatcat	gacattaacc	tataaaaata	ggcgtatcac	gaggcccttt	cgtc		10494

<210> 131

<211> 7589

<212> DNA

<213> Artificial Sequence

<220>

<223> pUC19-kan::pdc1::FBA-alsS::TRX1 (clone A)

<400> 131

tatttgatc	gaggtgtcta	gtcttctatt	acactaatgc	agtttcaggg	ttttgaaac	60
cacactgtt	aaacagtgtt	ccttaatcaa	ggatacctct	ttttttttcc	ttggttcac	120
taattcatcg	gttttttttt	tggaagacat	cttttccaac	gaaaagaata	tacatatcgt	180
ttaagagaaa	ttctccaaat	ttgtaaagaa	gcggaccag	acttaaggcc	gcccgcaaat	240
taaagccttc	gagcgtccca	aaaccttctc	aagcaagggt	ttcagtataa	tgttacatgc	300
gtacacgcgt	ctgtacagaa	aaaaaagaaa	aatttgaaat	ataaataacg	ttcttaatac	360

20110907_CL5178USNA_SeqList.txt

taacataact	ataaaaaaat	aaatagggac	ctagacttca	ggttgtctaa	ctccttcctt	420
ttcgggttaga	gcggatgtgg	ggggagggcg	tgaatgtaag	cgtgacataa	ctaattacat	480
gattaattaa	ctagagagct	ttcgttttca	tgagttcccc	gaattctttc	ggaagcttgt	540
cacttgctaa	attaatgtta	tcactgtagt	caaccgggac	atcgatgatg	acaggacctt	600
cagcgttcat	gccttgacgc	agaacatctg	ccagctggtc	tgggtgattct	acgcgcaagc	660
cagttgctcc	gaagctttcc	gcatatttca	cgatatcgat	atttccgaaa	tcgaccgcag	720
atgtacgggt	atattttttc	aattgctgga	atgcaaccat	gtcatatgtg	ctgtcgttcc	780
atacaatgtg	tacaattgggt	gcttttagtc	gaactgctgt	ctctaattcc	attgctgaga	840
ataagaaacc	gccgtcacca	gagacagaaa	ccactttttc	tcccggtttc	accaatgaag	900
cgccgattgc	ccaaggaagc	gcaacgccga	gtgtttgcat	accgttactg	atcattaatg	960
ttaacggctc	gtagctgcgg	aaataacgtg	acatccaaat	ggcgtgcgaa	ccgatatcgc	1020
aagttactgt	aacatgatca	tcgactgcat	tacgcaactc	tttaacgatt	tcaagaggggt	1080
gcgctctgtc	tgatttccaa	tctgcaggca	cctgctcacc	ttcatgcata	tattgtttta	1140
aatcagaaaag	gattttctgc	tcacgctctg	caaattccac	tttcacagca	tcgtgttcga	1200
tatgattgat	cgtggacgga	atgtcaccga	tcaattcaag	atcaggctgg	taagcatgat	1260
caatgtcagc	gataatctcg	tctaaatgga	taattgtccg	gtctccattg	atattccaga	1320
atttcggatc	atattcaatc	gggtcatagc	cgatcgtcag	aacaacatct	gcctgctcta	1380
gcagtaaadc	gccaggctgg	ttgcggaaca	aaccgatacg	gccaaaatat	tgatcctcta	1440
aatctctaga	aagggtaccg	gcagcttgat	atgtttcaac	aaatggaagc	tgaacctttt	1500
tcaaaagctt	gcgaaccgct	ttaattgctt	ccggtcttcc	gcctttcatg	ccgacaaaaa	1560
cgacaggaag	ttttgctgtt	tggatttttg	ctatggccgc	actgattgca	tcattctgctg	1620
caggaccgag	ttttggcgct	gcaacagcac	gcacgttttt	cgtattttgtg	acttcattca	1680
caacatcttg	cggaaagctc	acaaaagcgg	ccccagcctg	ccctgctgac	gctatcctaa	1740
atgcatttgt	aacagcttcc	ggtatatatt	ttacatcttg	aacttctaca	ctgtattttg	1800
taatcggctg	gaatagcgcc	gcattatcca	aagattgatg	tgtccgtttt	aaacgatctg	1860
cacggatcac	gtttccagca	agcgcaacga	cagggtctcc	ttcagtgttc	gctgtcagca	1920
ggcctgttgc	caagttagag	gcacccggtc	ctgatgtgac	taacacgact	cccggttttc	1980
cagttaaacg	gccgactgct	tgggccatga	atgctgcgtt	ttgttcgtgc	cgggcaacga	2040
taatttcagg	tcctttatct	tgtaaagcgt	caaataccgc	atcaattttt	gcacctggaa	2100
tgccaaatac	atgtgtgaca	ccttgctcca	ctaagcaatc	aacaacaagc	tccgcccctc	2160
tgtttttcac	aagggatttt	tgttcttttg	ttgcttttgt	caacatcctc	agctctagat	2220
ttgaatatgt	attacttgggt	tatggttata	tatgacaaaa	gaaaaagaag	aacagaagaa	2280
taacgcaagg	aagaacaata	actgaaattg	atagagaagt	attatgtctt	tgtcttttta	2340
taataaatca	agtgcagaaa	tccgttagac	aacatgaggg	ataaaattta	acgtgggcga	2400

20110907_CL5178USNA_SeqList.txt

agaagaagga	aaaaagtttt	tgtgagggcg	taattgaagc	gatctgttga	ttgtagattt	2460
tttttttttg	aggagtcaaa	gtcagaagag	aacagacaaa	tggtattaac	catccaatac	2520
ttttttggag	caacgctaag	ctcatgcttt	tccattgggt	acgtgctcag	ttgttagata	2580
tggaaagaga	ggatgctcac	ggcagcgtga	ctccaattga	gcccgaagaa	gaggatgcca	2640
cgttttcccg	acggctgcta	gaatggaaaa	aggaaaaata	gaagaatccc	attcctatca	2700
ttatttacgt	aatgaccac	acatttttga	gattttcaac	tattacgtat	tacgataatc	2760
ctgctgtcat	tatcattatt	atctatatcg	acgtatgcaa	cgtatgtgaa	gccaaagtagg	2820
caattattta	gtactgtcag	tattgttatt	catttcagat	cttaagccta	accaggccaa	2880
ttcaacagac	tgtcggcaac	ttcttgtctg	gtctttccat	ggtaagtga	agtgcagtaa	2940
taatatgaac	caatttattt	ttcgttacat	aaaaatgctt	ataaaacttt	aactaataat	3000
tagagattaa	atcgcaaacg	gccgactcta	gaggatcccc	caccttggct	aactcgttgt	3060
atcatcactg	gataacttcg	tatagcatac	attatacgaa	gttatctagg	gattcataac	3120
cattttctca	atcgaattac	acagaacaca	ccgtacaaac	ctctctatca	taactactta	3180
atagtcacac	acgtactcgt	ctaaatacac	atcatcgtcc	tacaagttca	tcaaagtgtt	3240
ggacagacaa	ctataccagc	atggatctct	tgtatcgggt	cttttctccc	gctctctcgc	3300
aataacaatg	aacactgggt	caatcatagc	ctacacaggt	gaacagagta	gcgtttatac	3360
agggtttata	cggtgattcc	tacggcaaaa	atttttcatt	tctaaaaaaa	aaaagaaaaa	3420
tttttctttc	caacgctaga	aggaaaagaa	aaatctaatt	aaattgattt	ggtgattttc	3480
tgagagtcc	ctttttcata	tatcgaattt	tgaatataaa	aggagatcga	aaaaattttt	3540
ctattcaatc	tgttttctgg	ttttatttga	tagttttttt	gtgtattatt	attatggatt	3600
agtactgggt	tatatgggtt	tttctgtata	acttcttttt	attttagttt	gtttaatctt	3660
attttgagtt	acattatagt	tccctaactg	caagagaagt	aacattaaaa	ctcgagatgg	3720
gtaaggaaaa	gactcacgtt	tcgaggccgc	gattaaattc	caacatggat	gctgattttat	3780
atgggtataa	atgggctcgc	gataatgtcg	ggcaatcagg	tgcgacaatc	tatcgattgt	3840
atgggaagcc	cgatgcgcca	gagttgtttc	tgaacatgg	caaaggtagc	gttgccaatg	3900
atgttacaga	tgagatggtc	agactaaact	ggctgacgga	atttatgcct	cttccgacca	3960
tcaagcattt	tatccgtact	cctgatgatg	catggttact	caccactgcg	atccccggca	4020
aaacagcatt	ccaggattta	gaagaatatc	ctgattcagg	tgaaaatatt	gttgatgcgc	4080
tggcagtgtt	cctgcgccgg	ttgcattcga	ttcctgtttg	taattgtcct	tttaacagcg	4140
atcgcgtatt	tcgtctcgtc	caggcgcaat	cacgaatgaa	taacggtttg	gttgatgcga	4200
gtgattttga	tgacgagcgt	aatggctggc	ctgttgaaca	agtctggaaa	gaaatgcata	4260
agcttttgcc	attctcaccg	gattcagtcg	tcactcatgg	tgattttctca	cttgataacc	4320
ttatttttga	cgaggggaaa	ttaatagggt	gtattgatgt	tggacgagtc	ggaatcgag	4380
accgatacca	ggatcttgcc	atcctatgga	actgcctcgg	tgagttttct	ccttcattac	4440

20110907_CL5178USNA_SeqList.txt

agaaacggct	ttttcaaaaa	tatggtattg	ataatcctga	tatgaataaa	ttgcagtttc	4500
atttgatgct	cgatgagttt	ttctaagttt	aacttgatac	tactagattt	tttctcttca	4560
tttataaaat	ttttggttat	aattgaagct	ttagaagtat	gaaaaaatcc	ttttttttca	4620
ttctttgcaa	ccaaaataag	aagcttcttt	tattcattga	aatgatgaat	ataaacctaa	4680
caaaagaaaa	agactcgaat	atcaaacatt	aaaaaaaaat	aaaagagggt	atctgttttc	4740
ccatttagtt	ggagtttgca	ttttctaata	gatagaactc	tcaattaatg	tggatttagt	4800
ttctctgttc	gataacttcg	tatagcatac	attatacgaa	gttatctgaa	cattagaata	4860
cgtaatccgc	aatgcggggc	cgcttaatta	atctagagtc	gacctgcagg	catgcaagct	4920
tggcgtaatc	atggtcatag	ctgtttcctg	tgtgaaattg	ttatccgctc	acaattccac	4980
acaacatacg	agccggaagc	ataaagtgtg	aagcctgggg	tgcctaatag	gtgagctaac	5040
tcacattaat	tgcgttcgcg	tcactgcccg	ctttccagtc	gggaaacctg	tcgtgccagc	5100
tgcattaatg	aatcggccaa	cgcgcgggga	gaggcggttt	gcgtattggg	cgctcttccg	5160
cttcctcgct	cactgactcg	ctgcgctcgg	tcgttcgggt	gcggcgagcg	gtatcagctc	5220
actcaaaggc	ggtaatacgg	ttatccacag	aatcagggga	taacgcagga	aagaacatgt	5280
gagcaaaagg	ccagcaaaag	gccaggaacc	gtaaaaaggc	cgcgttgctg	gcgtttttcc	5340
ataggctccg	ccccctgac	gagcatcaca	aaaatcgacg	ctcaagtcag	aggtggcgaa	5400
acccgacagg	actataaaga	taccaggcgt	ttccccctgg	aagctccctc	gtgcgctctc	5460
ctgttccgac	cctgccgctt	accggatacc	tgtccgcctt	tctcccttcg	ggaagcgtgg	5520
cgctttctca	tagctcacgc	tgtaggtatc	tcagttcggg	gtaggtcggt	cgctccaagc	5580
tgggctgtgt	gcacgaacct	cccgttcagc	ccgaccgctg	cgccttatcc	ggtaactatc	5640
gtcttgagtc	caaccgggta	agacacgact	tatcgccact	ggcagcagcc	actggtaaca	5700
ggattagcag	agcgagggtat	gtaggcgggtg	ctacagagtt	cttgaagtgg	tggcctaact	5760
acggctacac	tagaaggaca	gtatttggtg	tctgcgctct	gctgaagcca	gttaccttcg	5820
gaaaaagagt	tggtagctct	tgatccggca	aacaaaccac	cgctggtagc	ggtgggtttt	5880
ttgtttgcaa	gcagcagatt	acgcgcagaa	aaaaaggatc	tcaagaagat	cctttgatct	5940
tttctacggg	gtctgacgct	cagtggaacg	aaaactcacg	ttaagggtat	ttgggtcatga	6000
gattatcaaa	aaggatcttc	acctagatcc	ttttaaatga	aaaatgaagt	tttaaatcaa	6060
tctaaagtat	atatgagtaa	acttggtctg	acagttacca	atgcttaatc	agtgaggcac	6120
ctatctcagc	gatctgtcta	tttcgttcct	ccatagttgc	ctgactcccc	gtcgtgtaga	6180
taactacgat	acgggagggc	ttaccatctg	gccccagtg	tgcaatgata	ccgcgagacc	6240
cacgtccacc	ggctccagat	ttatcagcaa	taaaccagcc	agccggaagg	gccgagcgca	6300
gaagtgggtc	tgcaacttta	tccgcctcca	tccagtcctat	taattgttgc	cggaagcta	6360
gagtaagtag	ttcgccagtt	aatagtttgc	gcaacgttgt	tgccattgct	acaggcatcg	6420
tgggtgtcacg	ctcgtcgttt	ggatatggctt	cattcagctc	cggttcccaa	cgatcaaggc	6480

20110907_CL5178USNA_SeqList.txt

gagttacatg atcccccatg ttgtgcaaaa aagcggttag ctccttcggt cctccgatcg	6540
ttgtcagaag taagttggcc gcagtgttat cactcatggt tatggcagca ctgcataatt	6600
ctcttactgt catgccatcc gtaagatgct tttctgtgac tgggtgagtac tcaaccaagt	6660
cattctgaga atagtgtatg cggcgaccga gttgctcttg cccggcggtca atacgggata	6720
ataccgcgcc acatagcaga actttaaaag tgctcatcat tggaaaacgt tcttcggggc	6780
gaaaactctc aaggatctta ccgctgttga gatccagttc gatgtaaccc actcgtgcac	6840
ccaactgatc ttcagcatct tttactttca ccagcgtttc tgggtgagca aaaacaggaa	6900
ggcaaaatgc cgcaaaaaag ggaataaggg cgacacggaa atgttgaata ctcatactct	6960
tcctttttca atattattga agcatttatc agggttattg tctcatgagc ggatacatat	7020
ttgaatgtat ttagaaaaat aaacaaatag gggttccgcg cacatttccc cgaaaagtgc	7080
cacctgacgt ctaagaaacc attattatca tgacattaac ctataaaaaat aggcgtatca	7140
cgaggccctt tcgtctcgcg cgtttcggtg atgacggtga aaacctctga cacatgcagc	7200
tcccggagac ggtcacagct tgtctgtaag cggatgccgg gagcagacaa gcccgtcagg	7260
gcgcgtcagc ggggtgttggc ggggtgtcggg gctggcctta ctatgcggca tcagagcaga	7320
ttgtactgag agtgcaccat atgcggtgtg aaataccgca cagatgcgta aggagaaaat	7380
accgcatcag gcgccattcg ccattcaggc tgcgcaactg ttgggaaggg cgatcgggtgc	7440
gggcctcttc gctattacgc cagctggcga aagggggatg tgctgcaagg cgattaagtt	7500
gggtaacgcc agggttttcc cagtcacgac gttgtaaaac gacggccagt gaattcgagc	7560
tcggtacccg gggatccggc gcgccgttt	7589

<210> 132
 <211> 9593
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> pYZ090 alsS

<400> 132	
ggccgcacct ggtaaacct ctagtggagt agtagatgta atcaatgaag cggaagccaa	60
aagaccagag tagaggccta tagaagaaac tgcgatacct tttgtgatgg ctaaacaac	120
agacatcttt ttatatgttt ttacttctgt atatcgtgaa gtagtaagtg ataagcgaat	180
ttggctaaga acgttgtaag tgaacaaggg acctcttttg cttttcaaaa aaggattaaa	240
tggagttaat cattgagatt tagttttcgt tagattctgt atccctaaat aactccctta	300
cccgcggga aggcacaaaa gacttgaata atagcaaacg gccagtagcc aagaccaaat	360
aatactagag ttaactgatg gtcttaacaa ggcattacgt ggtgaactcc aagaccaata	420
tacaaaatat cgataagtta ttcttgccca ccaatttaag gagcctacat caggacagta	480
gtaccattcc tcagagaaga ggtatacata acaagaaaat cgcgtgaaca cttatataa	540
cttagcccgt tattgagcta aaaaaccttg caaaatttcc tatgaataag aatacttcag	600

20110907_CL5178USNA_SeqList.txt

acgtgataaa	aatttacttt	ctaactcttc	tcacgctgcc	cctatctgtt	cttccgctct	660
accgtgagaa	ataaagcatc	gagtacggca	gttcgctgtc	actgaactaa	aacaataagg	720
ctagttcgaa	tgatgaactt	gcttgctgtc	aaacttctga	gttgccgctg	atgtgacact	780
gtgacaataa	attcaaaccg	gttatagcgg	tctcctccgg	taccggttct	gccacctcca	840
atagagctca	gtaggagtca	gaacctctgc	ggtggctgtc	agtgactcat	ccgcgtttcg	900
taagttgtgc	gcgtgcacat	ttcgcccgtt	cccgtctatc	ttgcagcagg	cgaaaatttt	960
catcacgctg	taggacgcaa	aaaaaaaaata	attaatcgta	caagaatctt	ggaaaaaaaaa	1020
ttgaaaaatt	ttgtataaaa	gggatgacct	aacttgactc	aatggctttt	acaccagta	1080
ttttcccttt	ccttgtttgt	tacaattata	gaagcaagac	aaaaacatat	agacaaccta	1140
ttcctaggag	ttatatTTTT	ttaccctacc	agcaatataa	gtaaaaaact	gtttaaacag	1200
tatggcagtt	acaatgtatt	atgaagatga	tgtagaagta	tcagcacttg	ctggaaagca	1260
aattgcagta	atcggttatg	gttcacaagg	acatgctcac	gcacagaatt	tgcgtgattc	1320
tggtcacaac	gttatcattg	gtgtgcgcca	cggaaaatct	tttgataaag	caaaagaaga	1380
tggctttgaa	acatttgaag	taggagaagc	agtagctaaa	gctgatgtta	ttatggtttt	1440
ggcaccagat	gaacttcaac	aatccattta	tgaagaggac	atcaaaccaa	acttgaaagc	1500
aggttcagca	cttggttttg	ctcacggatt	taatatccat	tttggtata	ttaaagtacc	1560
agaagacgtt	gacgtcttta	tggttgcgcc	taaggctcca	ggtcaccttg	tccgtcggac	1620
ttatactgaa	ggttttggta	caccagcttt	gtttgtttca	cacaaaaatg	caagtgggtca	1680
tgcgcgtgaa	atcgcaatgg	attgggccaa	aggaattggg	tgtgctcgag	tgggaattat	1740
tgaacaact	tttaaagaag	aaacagaaga	agatttgttt	ggagaacaag	ctgttctatg	1800
tggaggtttg	acagcacttg	ttgaagccgg	ttttgaaaca	ctgacagaag	ctggatacgc	1860
tggcgaattg	gcttactttg	aagttttgca	cgaaatgaaa	ttgattgttg	acctcatgta	1920
tgaaggtggg	ttactaaaa	tgcgtcaatc	catctcaaat	actgctgagt	ttggcgatta	1980
tgtgactggg	ccacggatta	ttactgacga	agttaaaaag	aatatgaagc	ttgttttggc	2040
tgatattcaa	tctggaaaat	ttgctcaaga	tttcgttgat	gacttcaaag	cggggcgctcc	2100
aaaattaata	gcctatcgcg	aagctgcaaa	aaatcttgaa	attgaaaaaa	ttggggcaga	2160
gctacgtcaa	gcaatgccat	tcacacaatc	tggtgatgac	gatgccttta	aaatctatca	2220
gtaaggccct	gcaggccaga	ggaaaataat	atcaagtgtc	ggaaactttt	tctcttggaa	2280
tttttgaac	atcaagtcac	agtcaattga	attgacccaa	tttcacattt	aagatttttt	2340
ttttttcatc	cgacatacat	ctgtacacta	ggaagccctg	tttttctgaa	gcagcttcaa	2400
atatatatat	tttttacata	tttattatga	ttcaatgaac	aatctaatta	aatcgaaaac	2460
aagaaccgaa	acgcgaataa	ataatttatt	tagatgggtga	caagtgtata	agtcctcatc	2520
gggacagcta	cgatttctct	ttcggttttg	gctgagctac	tggttgctgt	gacgcagcgg	2580
cattagcgcg	gcgttatgag	ctaccctcgt	ggcctgaaag	atggcgggaa	taaagcggaa	2640

20110907_CL5178USNA_SeqList.txt

ctaaaaatta	ctgactgagc	catattgagg	tcaatttgtc	aactcgtcaa	gtcacgtttg	2700
gtggacggcc	cctttccaac	gaatcgtata	tactaacatg	cgcgcgcttc	ctatatacac	2760
atatacatat	atatatatat	atatatgtgt	gcgtgtatgt	gtacacctgt	atttaatttc	2820
cttactcgcg	ggtttttctt	ttttctcaat	tcttggttc	ctctttctcg	agcggaccgg	2880
atcctccgcg	gtgccggcag	atctatttaa	atggcgcgcc	gacgtcaggt	ggcacttttc	2940
ggggaaatgt	gcgcggaacc	cctatttggt	tatttttcta	aatacattca	aatatgtatc	3000
cgctcatgag	acaataaccc	tgataaatgc	ttcaataata	ttgaaaaagg	aagagtatga	3060
gtattcaaca	tttccgtgtc	gcccttattc	ccttttttgc	ggcattttgc	cttcctgttt	3120
ttgctcacc	agaaacgctg	gtgaaagtaa	aagatgctga	agatcagttg	gggtgcacgag	3180
tgggttacat	cgaactggat	ctcaacagcg	gtaagatcct	tgagagtttt	cgccccgaag	3240
aacgttttcc	aatgatgagc	acttttaaa	ttctgctatg	tggcgcggta	ttatcccgtat	3300
ttgacgccgg	gcaagagcaa	ctcggtcgcc	gcatacacta	ttctcagaat	gacttggttg	3360
agtactcacc	agtcacagaa	aagcatctta	cggatggcat	gacagtaaga	gaattatgca	3420
gtgctgccat	aaccatgagt	gataacactg	cggccaactt	acttctgaca	acgatcggag	3480
gaccgaagga	gctaaccgct	tttttgcaca	acatggggga	tcatgtaact	cgccttgatc	3540
gttggaacc	ggagctgaat	gaagccatac	caaacgacga	gcgtgacacc	acgatgcctg	3600
tagcaatggc	aacaacgttg	cgaaactat	taactggcga	actacttact	ctagcttccc	3660
ggcaacaatt	aatagactgg	atggaggcgg	ataaagttgc	aggaccactt	ctgcgctcgg	3720
cccttccggc	tggctggttt	attgctgata	aatctggagc	cgggtgagcgt	gggtctcgcg	3780
gtatcattgc	agcactgggg	ccagatggta	agccctccc	tatcgtagtt	atctacacga	3840
cggggagtca	ggcaactatg	gatgaacgaa	atagacagat	cgctgagata	gggtgcctcac	3900
tgattaagca	ttggtaactg	tcagaccaag	tttactcata	tatacttttag	attgatttaa	3960
aacttcattt	ttaatttaaa	aggatctagg	tgaagatcct	ttttgataat	ctcatgacca	4020
aaatccctta	acgtgagttt	tcgttccact	gagcgtcaga	ccccgtagaa	aagatcaaag	4080
gatcttcttg	agatcctttt	tttctgcgcg	taatctgctg	cttgcaaaca	aaaaaaccac	4140
cgctaccagc	ggtggtttgt	ttgccggatc	aagagctacc	aactcttttt	ccgaaggtaa	4200
ctggcttcag	cagagcgag	ataccaaata	ctgttcttct	agtgtagccg	tagttaggcc	4260
accacttcaa	gaactctgta	gcaccgccta	catacctcgc	tctgctaata	ctgttaccag	4320
tggctgctgc	cagtggcgat	aagtcgtgtc	ttaccgggtt	ggactcaaga	cgatagttac	4380
cggataaggc	gcagcggtcg	ggctgaacgg	ggggttcgtg	cacacagccc	agcttgagac	4440
gaacgaccta	caccgaactg	agatacctac	agcgtgagct	atgagaaagc	gccacgcttc	4500
ccgaagggag	aaaggcggac	aggtatccgg	taagcggcag	ggtcggaaca	ggagagcgca	4560
cgagggagct	tccaggggga	aacgcctggg	atctttatag	tcctgtcggg	tttcgccacc	4620
tctgacttga	gcgtcgattt	ttgtgatgct	cgtcaggggg	gcggagccta	tggaaaaacg	4680

20110907_CL5178USNA_SeqList.txt

ccagcaacgc	ggccttttta	cggttcctgg	ccttttgctg	gccttttgct	cacatgttct	4740
ttcctgcgtt	atccccctgat	tctgtggata	accgtattac	cgcctttgag	tgagctgata	4800
ccgctcgccg	cagccgaacg	accgagcgca	gcgagtcagt	gagcgaggaa	gcggaagagc	4860
gccaatacgc	caaaccgcct	ctccccgcgc	gttggccgat	tcattaatgc	agctggcacg	4920
acaggtttcc	cgactggaaa	gcgggcagtg	agcgcaacgc	aattaatgtg	agttagctca	4980
ctcattaggc	accccaggct	ttacacttta	tgcttccggc	tcgtatgttg	tgtggaattg	5040
tgagcggata	acaatttcac	acaggaaaca	gctatgacca	tgattacgcc	aagctttttc	5100
tttccaattt	tttttttttc	gtcattataa	aaatcattac	gaccgagatt	cccgggtaat	5160
aactgatata	attaaattga	agctctaatt	tgtgagttta	gtatacatgc	atttacttat	5220
aatacagttt	tttagttttg	ctggccgcat	cttctcaaat	atgcttccca	gcctgctttt	5280
ctgtaacggt	caccctctac	cttagcatcc	cttccctttg	caaatagtcc	tcttccaaca	5340
ataataatgt	cagatcctgt	agagaccaca	tcattccacg	ttctatactg	ttgacccaat	5400
gcgtctccct	tgtcatctaa	accacacacc	ggtgtcataa	tcaaccaatc	gtaaccttca	5460
tctcttccac	ccatgtctct	ttgagcaata	aagccgataa	caaaatcttt	gtcgtctttc	5520
gcaatgtcaa	cagtaccctt	agtatattct	ccagtagata	gggagccctt	gcatgacaat	5580
tctgctaaca	tcaaaaggcc	tctaggttcc	tttgttactt	cttctgccgc	ctgcttcaaa	5640
ccgctaacaa	tacctgggcc	caccacaccg	tgtgcattcg	taatgtctgc	ccattctgct	5700
attctgtata	caccgcgaga	gtactgcaat	ttgactgtat	taccaatgtc	agcaaatttt	5760
ctgtcttcga	agagtaaaaa	attgtacttg	gcggataatg	ccttttagcg	cttaactgtg	5820
ccctccatgg	aaaaatcagt	caagatatcc	acatgtgttt	ttagtaaaca	aattttggga	5880
cctaattgct	caactaactc	cagtaattcc	ttggtggtac	gaacatccaa	tgaagcacac	5940
aagtttggtt	gcttttcgtg	catgatatta	aatagcttgg	cagcaacagg	actaggatga	6000
gtagcagcac	gttccttata	tgtagctttc	gacatgattt	atcttcgttt	cctgcagggt	6060
tttgttctgt	gcagttgggt	taagaatact	gggcaatttc	atgtttcttc	aacactacat	6120
atgcgtatat	ataccaatct	aagtctgtgc	tccttccttc	gttcttcctt	ctgttcggag	6180
attaccgaat	caaaaaaatt	tcaaggaaac	cgaaatcaaa	aaaaagaata	aaaaaaaaat	6240
gatgaattga	aaagcttgca	tgccctgcag	tcgactctag	tatactccgt	ctactgtacg	6300
atacacttcc	gctcaggtcc	ttgtccttta	acgaggcctt	accactcttt	tgttactcta	6360
ttgatccagc	tcagcaaagg	cagtgtgatc	taagattcta	tcttcgcat	gtagtaaaac	6420
tagctagacc	gagaaagaga	ctagaaatgc	aaaaggcact	tctacaatgg	ctgccatcat	6480
tattatccga	tgtgacgctg	catttttttt	tttttttttt	tttttttttt	tttttttttt	6540
tttttttttt	ttttgtacaa	atatcataaa	aaaagagaat	ctttttaagc	aaggattttc	6600
tttaacttct	cggcgacagc	atcaccgact	tcggtggtac	tgttggaacc	acctaataca	6660
ccagttctga	tacctgcac	caaaaccttt	tttaactgcat	cttcaatggc	tttaccttct	6720

20110907_CL5178USNA_SeqList.txt

tcaggcaagt	tcaatgacaa	tttcaacatc	attgcagcag	acaagatagt	ggcgataggg	6780
ttgaccttat	tctttggcaa	atctggagcg	gaaccatggc	atggttcgta	caaaccaa	6840
gcggtgttct	tgtctggcaa	agaggccaag	gacgcagatg	gcaacaaacc	caaggagcct	6900
gggataacgg	aggcttcac	ggagatgata	tcaccaaaca	tgttgctggt	gattataata	6960
ccatttaggt	gggttgggtt	cttaactagg	atcatggcgg	cagaatcaat	caattgatgt	7020
tgaactttca	atgtagggaa	ttcgttcttg	atggtttcct	ccacagtttt	tctccataat	7080
cttgaagagg	ccaaaacatt	agctttatcc	aaggaccaa	taggcaatgg	tggctcatgt	7140
tgtagggcc	tgaagcggc	cattcttgtg	attctttgca	cttctggaac	ggtgtattgt	7200
tcactatccc	aagcgacacc	atcaccatcg	tcttcctttc	tcttaccaa	gtaaatacct	7260
cccactaatt	ctctaacaac	aacgaagtca	gtacctttag	caaattgtgg	cttgattgga	7320
gataagtcta	aaagagagtc	ggatgcaaag	ttacatggtc	ttaagttggc	gtacaattga	7380
agttctttac	ggatttttag	taaaccttgt	tcaggtctaa	cactaccggt	accccattta	7440
ggaccacca	cagcacctaa	caaacggca	tcagccttct	tggaggcttc	cagcgccctca	7500
tctggaagt	gaacacctgt	agcatcgata	gcagcaccac	caattaaatg	attttcgaaa	7560
tcgaacttga	cattggaacg	aacatcagaa	atagctttta	gaaccttaat	ggcttcggct	7620
gtgatttctt	gaccaacgtg	gtcacctggc	aaaacgacga	tcttcttagg	ggcagacatt	7680
acaatgggat	atccttgaaa	tatatataaa	aaaaaaaaaa	aaaaaaaaaa	aaaaaatgc	7740
agcttctcaa	tgatattcga	atacgctttg	aggagataca	gcctaatatc	cgacaaactg	7800
ttttacagat	ttacgatcgt	acttgttacc	catcattgaa	ttttgaacat	ccgaacctgg	7860
gagttttccc	tgaacagat	agtatatattg	aacctgtata	ataatatata	gtctagcgct	7920
ttacggaaga	caatgtatgt	atttcggttc	ctggagaaac	tattgcatct	attgcatagg	7980
taatcttgca	cgtcgcatcc	ccggttcatt	ttctgcgttt	ccatcttgca	cttcaatagc	8040
atatctttgt	taacgaagca	tctgtgcttc	attttgtaga	acaaaaatgc	aacgcgagag	8100
cgctaatttt	tcaacaaag	aatctgagct	gcatttttac	agaacagaaa	tgcaacgcga	8160
aagcgctatt	ttaccaacga	agaatctgtg	cttcattttt	gtaaaacaaa	aatgcaacgc	8220
gagagcgcta	atttttcaaa	caaagaatct	gagctgcatt	tttacagaac	agaaatgcaa	8280
cgcgagagcg	ctattttacc	aacaaagaat	ctatacttct	tttttgttct	acaaaaatgc	8340
atcccagagag	cgctattttt	ctaacaagc	atcttagatt	actttttttc	tcctttgtgc	8400
gctctataat	gcagtctctt	gataactttt	tgcactgtag	gtccgttaag	gttagaagaa	8460
ggctactttg	gtgtctat	tctcttccat	aaaaaaagcc	tgactccact	tcccgcgttt	8520
actgattact	agcgaagctg	cgggtgcatt	ttttcaagat	aaaggcatcc	ccgattatat	8580
tctataccga	tgtggattgc	gcatactttg	tgaacagaaa	gtgatagcgt	tgatgattct	8640
tcattgggtca	gaaaattatg	aacggtttct	tctattttgt	ctctatatac	tacgtatagg	8700
aaatgtttac	attttcgtat	tgttttcgat	tcactctatg	aatagttctt	actacaattt	8760

20110907_CL5178USNA_SeqList.txt

ttttgtctaa agagtaatac tagagataaa cataaaaaat gtagaggtcg agtttagatg	8820
caagttcaag gagcgaaagg tggatgggta ggttatatag ggatatagca cagagatata	8880
tagcaaagag atacttttga gcaatgtttg tggaagcggg attcgcaata ttttagtagc	8940
tcgttacagt ccggtgctgt tttggttttt tgaaagtgcg tcttcagagc gcttttggtt	9000
ttcaaaagcg ctctgaagtt cctatacttt ctagagaata ggaacttcgg aataggaact	9060
tcaaagcgtt tccgaaaacg agcgcttccg aaaatgcaac gcgagctgcg cacatacagc	9120
tcactgttca cgtcgcacct atatctgcgt gttgcctgta tatatatata catgagaaga	9180
acggcatagt gcgtgtttat gcttaaagtc gtacttatat gcgtctatct atgtaggatg	9240
aaaggtagtc tagtacctcc tgtgatatta tcccattcca tgcgggggtat cgtatgcttc	9300
cttcagcact acccttttagc tgttctatat gctgccactc ctcaattgga ttagtctcat	9360
ccttcaatgc tatcatttcc tttgatattg gatcatatgc atagtaccga gaaactagag	9420
gatctcccat taccgacatt tgggcgctat acgtgcatat gttcatgtat gtatctgtat	9480
ttaaaacact tttgtattat ttttcctcat atatgtgtat aggtttatac ggatgattta	9540
attattactt caccaccctt tatttcaggc tgatatctta gccttggtac tag	9593

<210> 133
 <211> 12896
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> pBP915

<400> 133	
tcgcgctgtt cggtgatgac ggtgaaaacc tctgacacat gcagctcccg gagacgggtca	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 atttttttat gcctcggtaa tgattttcat	360
tttttttttt ccacctagcg gatgactctt tttttttctt agcgattggc attatcacat	420
aatgaattat acattatata aagtaatgtg atttcttcga agaataact aaaaaatgag	480
caggcaagat aaacgaaggc aaagatgaca gagcagaaag ccctagtaaa gcgtattaca	540
aatgaaacca agattcagat tgcgatctct ttaaagggtg gtcccctagc gatagagcac	600
tcgatcttcc cagaaaaaga ggcagaagca gtagcagaac aggccacaca atcgcaagtg	660
attaacgtcc acacaggtat agggtttctg gaccatatga tacatgctct ggccaagcat	720
tccggctggg cgctaactgt tgagtgcatt ggtgacttac acatagacga ccatcacacc	780
actgaagact gcgggattgc tctcgttcaa gcttttaaaag aggccctagg ggccgtgcgt	840
ggagtaaaaa ggtttggatc aggatttgcg ctttggatg aggcactttc cagagcgggtg	900

20110907_CL5178USNA_SeqList.txt

gtagatcttt	cgaacaggcc	gtacgcagtt	gtcgaacttg	gtttgcaaag	ggagaaagta	960
ggagatctct	cttgcgagat	gatcccgcat	tttcttgaaa	gctttgcaga	ggctagcaga	1020
attaccctcc	acgttgattg	tctgcgaggc	aagaatgata	atcaccgtag	tgagagtgcg	1080
ttcaaggctc	ttgcggttgc	cataagagaa	gccacctcgc	ccaatggtac	caacgatgtt	1140
ccctccacca	aagggtgttct	tatgtagtga	caccgattat	ttaaagctgc	agcatacgat	1200
atatatacat	gtgtatatat	gtatacctat	gaatgtcagt	aagtatgtat	acgaacagta	1260
tgatactgaa	gatgacaagg	taatgcatca	ttctatacgt	gtcattctga	acgaggcgcg	1320
ctttcctttt	ttctttttgc	tttttctttt	tttttctctt	gaactcgacg	gatctatgcg	1380
gtgtgaaata	ccgcacagat	gcgtaaggag	aaaataccgc	atcaggaaat	tgtaagcggt	1440
aatatTTTTgt	taaaattcgc	gttaaatttt	tgtaaataca	gtcatttttt	taaccaatag	1500
gccgaaatcg	gcaaaatccc	ttataaatca	aaagaataga	ccgagatagg	gttgagtgtt	1560
gttccagttt	ggaacaagag	tccactatta	aagaacgtgg	actccaacgt	caaagggcga	1620
aaaaccgtct	atcaggggcg	tggcccacta	cgtgaaccat	caccctaata	aagttttttg	1680
gggtcgaggt	gccgtaaagc	actaaatcgg	aaccctaaag	ggagcccccg	atttagagct	1740
tgacggggaa	agccggcgaa	cgtggcgaga	aaggaagggg	agaaagcgaa	aggagcgggc	1800
gctagggcg	tggcaagtgt	agcggtcacg	ctgcgcgtaa	ccaccacacc	cgccgcgctt	1860
aatgcgccgc	tacagggcg	gtccattcgc	cattcaggct	gcgcaactgt	tgggaagggc	1920
gcggtgcggg	cctcttcgct	attacgccag	ctggcgaaag	ggggatgtgc	tgcaaggcga	1980
ttaagttggg	taacgccagg	gttttcccag	tcacgacgtt	gtaaaacgac	ggccagttag	2040
cgcgcgtaat	acgactcact	atagggcgaa	ttgggtaccg	ggccccccct	cgaggctcgc	2100
ggcgcgccac	tggtagagag	cgactttgta	tgccccaatt	gcgaaaccgc	cgatatcctt	2160
ctcgattctt	tagtaccgga	ccaggacaag	gaaaaggagg	tcgaaacgtt	tttgaagaaa	2220
caagaggaac	tacacggaag	ctctaaagat	ggcaaccagc	cagaaactaa	gaaaatgaag	2280
ttgatggatc	caactggcac	cgctggcttg	aacaacaata	ccagccttcc	aacttctgta	2340
aataacggcg	gtacgccagt	gccaccagta	ccgttacctt	tcggtatacc	tcctttcccc	2400
atgtttccaa	tgcccttcat	gcctccaacg	gctactatca	caaatcctca	tcaagctgac	2460
gcaagcccta	agaaatgaat	aacaatactg	acagtactaa	ataattgcct	acttggcttc	2520
acatacgttg	catacgtcga	tatagataat	aatgataatg	acagcaggat	tatcgtaata	2580
cgtaatagct	gaaaatctca	aaaatgtgtg	ggtcattacg	taaataatga	taggaatggg	2640
attcttctat	ttttcctttt	tccattctag	cagccgtcgg	gaaaacgtgg	catcctctct	2700
ttcggggtca	attggagtca	cgctgccgtg	agcatcctct	ctttccatat	ctaacaactg	2760
agcacgtaac	caatggaaaa	gcatgagctt	agcgttgctc	caaaaaagta	ttggatgggt	2820
aataccattt	gtctgttctc	ttctgacttt	gactcctcaa	aaaaaaaaat	ctacaatcaa	2880
cagatcgctt	caattacgcc	ctcacaaaaa	cttttttctt	tcttcttcgc	ccacgttaaa	2940

20110907_CL5178USNA_SeqList.txt

ttttatccct	catgtttgtct	aacggatttc	tgcacttgat	ttattataaa	aagacaaaga	3000
cataatactt	ctctatcaat	ttcagttatt	gttcttcctt	gcgttattct	tctgtttcttc	3060
tttttctttt	gtcatatata	accataacca	agtaatacat	attcaaacta	gtatgactga	3120
caaaaaaact	cttaaagact	taagaaatcg	tagttctggt	tacgattcaa	tggttaaadc	3180
acctaatacgt	gctatgttgc	gtgcaactgg	tatgcaagat	gaagactttg	aaaaacctat	3240
cgtcggtgtc	atttcaactt	gggctgaaaa	cacaccttgt	aatatccact	tacatgactt	3300
tggtaaacta	gccaaagtcg	gtgttaagga	agctgggtgt	tggccagttc	agttcggaac	3360
aatcacgggt	tctgatggaa	tcgccatggg	aaccaagga	atgcgtttct	ccttgacatc	3420
tcgtgatatt	attgcagatt	ctattgaagc	agccatggga	ggtcataatg	cggatgcttt	3480
tgtagccatt	ggcggttgtg	ataaaaacat	gcccggttct	gttatcgcta	tggctaacat	3540
ggatatccca	gccatttttg	cttacggcgg	aacaattgca	cctggtaatt	tagacggcaa	3600
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	aagctggctg	3840
cgctgttgtc	aaaatgctcg	aatggggctt	aaaaccttct	gacattttta	cgctgaagc	3900
ttttgaagat	gctattactg	taactatggc	tctgggaggt	tcaaccaact	caacccttca	3960
cctcttagct	attgccccatg	ctgctaattg	ggaattgaca	cctgatgatt	tcaatacttt	4020
ccaagaaaaa	gttcctcatt	tggctgattt	gaaaccttct	ggtcataatg	tattccaaga	4080
cctttacaag	gtcggagggg	taccagcagt	tatgaaatat	ctccttaaaa	atggcttctt	4140
tcatggtgac	cgtatcactt	gtactggcaa	aacagtcgct	gaaaatttga	aggcttttga	4200
tgatttaaca	cctgggtcaaa	aggttattat	gccgcttgaa	aatcctaaac	gtgaagatgg	4260
tccgctcatt	attctccatg	gtaacttggc	tccagacggt	gccgttgcca	aagtttctgg	4320
tgtaaaagt	cgtcgtcatg	tcggtcctgc	taaggctctt	aattctgaag	aagaagccat	4380
tgaagctgtc	ttgaatgatg	atattgttga	tggatgatgt	gttgctgtac	gttttgtagg	4440
accaaagggc	ggtcctggta	tgcctgaaat	gctttccctt	tcatcaatga	ttgttggtaa	4500
agggaaggt	gaaaaagtgt	cccttctgac	agatggccgc	ttctcaggtg	gtacttatgg	4560
tcttgctgtg	ggcatatcgc	ctcctgaagc	acaagatggc	ggccaatcg	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	tcctgggtgc	tgtgggttaag	cggccgcgtt	4860
aattcaaatt	aattgatata	gttttttaat	gagtattgaa	tctgtttaga	aataatggaa	4920
tattattttt	atttatattt	ttatattatt	ggtcggctct	tttcttctga	aggatcaatga	4980

20110907_CL5178USNA_SeqList.txt

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	atattttaact	5280
tcatgtcaat	ttcggctctt	aaattttcca	catcatcaag	ttcaacatca	tctttttaact	5340
tgaatttatt	ctctagctct	tccaaccaag	cctcattgct	ccttgattta	ctggtgaaaa	5400
gtgatacact	ttgcgcgcaa	tccagggtcaa	aactttcctg	caaagaattc	accaattttct	5460
cgacatcata	gtacaatttg	ttttgttctc	ccatcacaa	ttaatatacc	tgatggattc	5520
ttatgaagcg	ctgggtaatg	gacgtgtcac	tctacttcgc	ctttttccct	actcctttta	5580
gtacggaaga	caatgcta	aaataagagg	gtaataataa	tattattaat	cggcaaaaaa	5640
gattaaacgc	caagcgttta	attatcagaa	agcaaacgct	gtaccaatcc	ttgaatgctt	5700
cccaattgta	tattaagagt	catcacagca	acatattctt	gttattaaat	taattattat	5760
tgatttttga	tattgtataa	aaaaaccaa	tatgtataaa	aaaagtgaat	aaaaaatacc	5820
aagtatggag	aaatatatta	gaagtctata	cgttaaacca	cccgggcccc	ccctcgaggt	5880
cgacggtatc	gataagcttg	atatcgaatt	cctgcagccc	gggggatcca	ctagttctag	5940
agcggccgct	ctagaactag	taccacaggt	gttgctctct	gaggacataa	aatacacacc	6000
gagattcatc	aactcattgc	tggagtttagc	atatctacaa	ttgggtgaaa	tggggagcga	6060
tttgcaggca	tttgctcggc	atgccggtag	aggtgtggtc	aataagagcg	acctcatgct	6120
atacctgaga	aagcaacctg	acctacagga	aagagttact	caagaataag	aattttcgtt	6180
ttaaaaccta	agagtcactt	taaaatttgt	atacacttat	tttttttata	acttatttta	6240
taataaaaaat	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	tcacgcacaa	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	gcaactttta	ctgcggaacc	gtaaccggtg	gaaaatccgc	accctatcaa	6900
gcaaaccttt	tccagtgggtg	aagctgcac	gatttttagcg	acagatatct	cgtccaccac	6960
tgtgtattgg	gaaaatgtag	aagtaccaag	gaaatggtgt	ataggtttcc	ctctgcatgt	7020

20110907_CL5178USNA_SeqList.txt

aaatctgctt	gtaccatcct	gcatagtacc	tctaggcata	gacaaatcat	ttttaaggca	7080
gaaattaccc	tcaggatggt	tgcagactct	acacttacca	cattgaggag	tgaacagtgg	7140
gatcacttta	tcaccaggac	gaacagtggg	aacaccttca	cctatggatt	caacgattcc	7200
ggcagcctcg	tgtcccgcga	ttactggcaa	aggagtaact	agagtgccac	tcaccacatg	7260
gtcgtcggat	ctacagattc	cgggtggcaac	catcttgatt	ctaacctcgt	gtgcttttgg	7320
tggcgctact	tctacttctt	ctatgctaaa	cggctttttc	tcttcccaca	aaactgccgc	7380
tttacactta	ataactttac	cggctgttga	catcctcagc	tagctattgt	aatatgtgtg	7440
tttgtttgga	ttattaagaa	gaataattac	aaaaaaaaatt	acaaaggaag	gtaattacaa	7500
cagaattaag	aaaggacaag	aaggaggaag	agaatcagtt	cattatcttct	tctttgttat	7560
ataacaaacc	caagtagcga	tttggccata	cattaaaagt	tgagaaccac	cctccctggc	7620
aacagccaca	actcgttacc	attgttcac	acgatcatga	aactcgtgtg	cagctgaaat	7680
ttcacctcag	tggatctctc	tttttattct	tcacgtttcc	actaaccttt	ttccatcagc	7740
tggcagggaa	cggaaagtgg	aatcccattt	agcgagcttc	ctcttttctt	caagaaaaga	7800
cgaagcttgt	gtgtgggtgc	gcgcgctagt	atctttccac	attaagaaat	ataccataaa	7860
ggttacttag	acatcactat	ggctatatat	atatatatat	atatatgtaa	cttagcacca	7920
tcgcgctgc	atcactgcat	gtgttaaccg	aaaagtgttg	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	ggtctttaga	cgaattaccc	ttcatttctc	aaactggcgt	caagggatcc	8340
tggtatgggt	ttatcgtttt	atttctgggt	cttatagcat	cgttttggac	ttctctgttc	8400
ccattaggcg	gttcaggagc	cagcgagaa	tcattctttg	aaggatactt	atcctttcca	8460
attttgattg	tctgttacgt	tggacataaa	ctgtatacta	gaaattggac	tttgatgggtg	8520
aaactagaag	atatggatct	tgataccggc	agaaaacaag	tagatttgac	tcttcgtagg	8580
gaagaaatga	ggattgagcg	agaaacatta	gcaaaaagat	ccttcgtaac	aagattttta	8640
catttctggg	gttgaaggga	aagatatgag	ctatacagcg	gaatttccat	atcactcaga	8700
ttttgttatc	taattttttc	cttcccacgt	ccgcgggaat	ctgtgtatat	tactgcatct	8760
agatatatgt	tatcttatct	tggcgctac	atttaatttt	caacgtattc	tataagaaat	8820
tgcgggagtt	tttttcatgt	agatgatact	gactgcacgc	aaatataggc	atgattttata	8880
ggcatgattt	gatggctgta	ccgataggaa	cgctaagagt	aacttcagaa	tcgttatcct	8940
ggcggaaaaa	attcatttgt	aaacttttaa	aaaaaaagcc	aatatcccca	aaattattaa	9000
gagcgcctcc	attattaact	aaaatttcac	tcagcatcca	caatgtatca	ggtatctact	9060

20110907_CL5178USNA_SeqList.txt

acagatatta	catgtggcga	aaaagacaag	aacaatgcaa	tagcgcatca	agaaaaaaca	9120
caaagctttc	aatcaatgaa	tcgaaaatgt	cattaaaata	gtatataaat	tgaaactaag	9180
tcataaagct	ataaaaagaa	aatttatatta	aatcttggct	ctcttgggct	caaggtgaca	9240
aggtcctcga	aaatagggcg	cgccccaccg	cggtggagct	ccagcttttg	ttccctttag	9300
tgaggggttaa	ttgcgcgctt	ggcgtaatca	tggtcatagc	tgtttcctgt	gtgaaattgt	9360
tatccgctca	caattccaca	caacatacga	gccggaagca	taaagtgtaa	agcctggggg	9420
gcctaattgag	tgagctaact	cacattaatt	gcgttgcgct	cactgcccgc	ttccagtcg	9480
ggaaacctgt	cgtgccagct	gcattaatga	atcggccaac	gcgcggggag	aggcggtttg	9540
cgtattgggc	gctcttccgc	ttcctcgctc	actgactcgc	tgcgctcggt	cgttcggctg	9600
cggcgagcgg	tatcagctca	ctcaaaggcg	gtaatacgg	tatccacaga	atcaggggat	9660
aacgcaggaa	agaacatgtg	agcaaaaggc	cagcaaaagg	ccaggaaccg	taaaaaggcc	9720
gcgttgctgg	cgtttttcca	taggctccgc	ccccctgacg	agcatcacia	aaatcgacgc	9780
tcaagtcaga	ggtggcgaaa	cccgacagga	ctataaagat	accaggcggt	tccccctgga	9840
agctccctcg	tgcgctctcc	tgttccgacc	ctgccgctta	ccggatacct	gtccgccttt	9900
ctcccttcgg	gaagcgtggc	gctttctcat	agctcacgct	gtaggtatct	cagttcgggtg	9960
taggtcggtc	gctccaagct	gggctgtgtg	cacgaacccc	ccgttcagcc	cgaccgctgc	10020
gccttatccg	gtaactatcg	tcttgagtcc	aaccgggtaa	gacacgactt	atcgccactg	10080
gcagcagcca	ctggtaacag	gattagcaga	gcgaggtatg	taggcgggtgc	tacagagttc	10140
ttgaagtgg	ggcctaacta	cggctacact	agaagaacag	tatttggtat	ctgcgctctg	10200
ctgaagccag	ttaccttcgg	aaaaagagtt	ggtagctctt	gatccggcaa	acaaaccacc	10260
gctggtagcg	gtgggttttt	tgtttgcaag	cagcagatta	cgcgagaaa	aaaaggatct	10320
caagaagatc	ctttgatctt	ttctacgggg	tctgacgctc	agtggaacga	aaactcacgt	10380
taagggattt	tggtcatgag	attatcaaaa	aggatcttca	cctagatcct	tttaaattaa	10440
aaatgaagtt	ttaaataaat	ctaaagtata	tatgagtaaa	cttgggtctga	cagttaccaa	10500
tgcttaatca	gtgaggcacc	tatctcagcg	atctgtctat	ttcgttcatc	catagttgcc	10560
tgactccccg	tcgtgtagat	aactacgata	cgggagggct	taccatctgg	ccccagtgtc	10620
gcaatgatac	cgcgagaccc	acgctcaccg	gctccagatt	tatcagcaat	aaaccagcca	10680
gccggaaggg	ccgagcgcag	aagtggtcct	gcaactttat	ccgcctccat	ccagtctatt	10740
aattgttgcc	gggaagctag	agtaagtagt	tcgccagtta	atagtttgcg	caacgttggt	10800
gccattgcta	caggcatcgt	ggtgtcacgc	tcgtcgtttg	gtatggcttc	attcagctcc	10860
ggttcccaac	gatcaaggcg	agttacatga	tcccccatgt	tgtgcaaaaa	agcggttagc	10920
tccttcggtc	ctccgatcgt	tgtcagaagt	aagttggccg	cagtgttatc	actcatggtt	10980
atggcagcac	tgcataattc	tcttactgtc	atgccatccg	taagatgctt	ttctgtgact	11040
ggtgagtact	caaccaagtc	attctgagaa	tagtgtatgc	ggcgaccgag	ttgctcttgc	11100

20110907_CL5178USNA_SeqList.txt

ccggcgtaa	tacgggataa	taccgcgcca	catagcagaa	ctttaaaagt	gctcatcatt	11160
ggaaaacgtt	cttcggggcg	aaaactctca	aggatcttac	cgctgttgag	atccagttcg	11220
atgtaacca	ctcgtgcacc	caactgatct	tcagcatctt	ttactttcac	cagcgtttct	11280
gggtgagcaa	aaacaggaag	gcaaaatgcc	gcaaaaaagg	gaataagggc	gacacggaaa	11340
tgttgaatac	tcatactctt	cctttttcaa	tattattgaa	gcatttatca	gggttattgt	11400
ctcatgagcg	gatacatatt	tgaatgtatt	tagaaaaata	aacaaatagg	ggttccgcgc	11460
acatttcccc	gaaaagtgcc	acctgaacga	agcatctgtg	cttcattttg	tagaacaaaa	11520
atgcaacgcg	agagcgctaa	tttttcaaac	aaagaatctg	agctgcattt	ttacagaaca	11580
gaaatgcaac	gcgaaagcgc	tattttacca	acgaagaatc	tgtgcttcat	ttttgtaaaa	11640
caaaaatgca	acgcgagagc	gctaattttt	caaacaaaga	atctgagctg	cattttttaca	11700
gaacagaaat	gcaacgcgag	agcgctattt	taccaacaaa	gaatctatac	ttcttttttg	11760
ttctacaaaa	atgcatcccg	agagcgctat	ttttctaaca	aagcatctta	gattactttt	11820
tttctccttt	gtgcgctcta	taatgcagtc	tcttgataac	tttttgact	gtagggtccgt	11880
taaggtaga	agaaggctac	tttggtgtct	attttctctt	ccataaaaaa	agcctgactc	11940
cacttcccgc	gtttactgat	tactagcgaa	gctgcgggtg	cattttttca	agataaaggc	12000
atccccgatt	atattctata	ccgatgtgga	ttgcgcatac	tttgtgaaca	gaaagtgata	12060
gcgttgatga	ttcttcattg	gtcagaaaaat	tatgaacggt	ttcttctatt	ttgtctctat	12120
atactacgta	taggaaatgt	ttacattttc	gtattgtttt	cgattcactc	tatgaatagt	12180
tcttactaca	atttttttgt	ctaaagagta	atactagaga	taaacataaa	aaatgtagag	12240
gtcgagttaa	gatgcaagtt	caaggagcga	aagggtgatg	ggtaggttat	atagggatat	12300
agcacagaga	tatatagcaa	agagatactt	ttgagcaatg	tttgtggaag	cggtattcgc	12360
aatattttag	tagctcgtaa	cagtcgggtg	cgtttttggg	tttttgaaag	tgcgtcttca	12420
gagcgctttt	ggttttcaaa	agcgctctga	agttcctata	ctttctagag	aataggaact	12480
tcggaatagg	aacttcaaag	cgtttccgaa	aacgagcgct	tccgaaaatg	caacgcgagc	12540
tgcgcacata	cagctcactg	ttcacgtcgc	acctatatct	gcgtgttgcc	tgtatatata	12600
tatacatgag	aagaacggca	tagtgcggtg	ttatgcttaa	atgcgtactt	atatgcgtct	12660
atttatgtag	gatgaaagg	agtctagtac	ctcctgtgat	attatcccat	tccatgcggg	12720
gtatcgtag	cttccttcag	cactaccctt	tagctgttct	atatgctgcc	actcctcaat	12780
tggattagtc	tcaccttca	atgctatcat	ttcctttgat	attggatcat	actaagaaac	12840
cattattatc	atgacattaa	cctataaaaa	taggcgtatc	acgaggccct	ttcgtc	12896

<210> 134
 <211> 11856
 <212> DNA
 <213> Artificial sequence

<220>

<223> Plasmid

<400> 134

tcccattacc	gacatttggg	cgctatacgt	gcatatgttc	atgtatgtat	ctgtatttta	60
aacacttttg	tattatTTTT	cctcatatat	gtgtataggt	ttatacggat	gatttaatta	120
ttacttcacc	accctttatt	tcaggctgat	atcttagcct	tgttactagt	tagaaaaaga	180
catttttgct	gtcagtcact	gtcaagagat	tcttttgctg	gcatttcttc	tagaagcaaa	240
aagagcgatg	cgtcttttcc	gctgaaccgt	tccagcaaaa	aagactacca	acgcaatatg	300
gattgtcaga	atcatataaa	agagaagcaa	ataactcctt	gtcttgtatc	aattgcatta	360
taatatcttc	ttgttagtgc	aatatcatat	agaagtcatc	gaaatagata	ttaagaaaaa	420
caaaactgtac	aatcaatcaa	tcaatcatcg	ctgaggatgt	tgacaaaagc	aacaaaagaa	480
caaaaatccc	ttgtgaaaaa	cagaggggcg	gagcttgttg	ttgattgctt	agtggagcaa	540
gggtgcacac	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
gtgagctttc	cgcaagatgt	tgtgaatgaa	gtcacaaata	cgaaaaacgt	gcgtgctggt	1020
gcagcgccaa	aactcgggtc	tgcagcagat	gatgcaatca	gtgcggccat	agcaaaaatc	1080
caaacagcaa	aacttcctgt	cgttttggtc	ggcatgaaag	gcggaagacc	ggaagcaatt	1140
aaagcggttc	gcaagctttt	gaaaaagggt	cagcttccat	ttgttgaaac	atatcaagct	1200
gccggtaccc	tttctagaga	tttagaggat	caatatTTTT	gccgtatcgg	tttgttccgc	1260
aaccagcctg	gcgatttact	gctagagcag	gcagatgttg	ttctgacgat	cggctatgac	1320
ccgattgaat	atgatccgaa	attctggaat	atcaatggag	accggacaat	tatccattta	1380
gacgagatta	tcgctgacat	tgatcatgct	taccagcctg	atcttgaatt	gatcggtgac	1440
attccgtcca	cgatcaatca	tatcgaacac	gatgctgtga	aagtggaatt	tgcagagcgt	1500
gagcagaaaa	tcctttctga	tttaaaacaa	tatatgcatg	aagggtgagca	ggtgcctgca	1560
gattggaaat	cagacagagc	gcaccctctt	gaaatcgtta	aagagttgcg	taatgcagtc	1620
gatgatcatg	ttacagtaac	ttgcgatatc	ggttcgcacg	ccatttggat	gtcacgttat	1680
ttccgcagct	acgagccggt	aacattaatg	atcagtaacg	gtatgcaaac	actcggcggt	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

20110907_CL5178USNA_SeqList.txt

gcggaaagct	tcggagcaac	tggcttgccg	gtagaatcac	cagaccagct	ggcagatggt	2040
ctgcgtcaag	gcatgaacgc	tgaaggtcct	gtcatcatcg	atgtcccggg	tgactacagt	2100
gataacatta	atttagcaag	tgacaagctt	ccgaaagaat	tcgggggaact	catgaaaacg	2160
aaagctctct	agttaattaa	tcatgtaatt	agttatgtca	cgcttacatt	cacgccctcc	2220
ccccacatcc	gctctaaccg	aaaaggaagg	agttagacaa	cctgaagtct	aggtccttat	2280
ttatTTTTTT	atagttatgt	tagtattaag	aacgttatTT	atatttcaaa	TTTTTctttt	2340
TTTTctgtac	agacgcgtgt	acgcatgtaa	cattatactg	aaaaccttgc	ttgagaagggt	2400
TTTgggacgc	tcgaaggctt	taatttgcgg	gcggccgcac	ctggtaaaac	ctctagtggg	2460
gtagtagatg	taatcaatga	agcgggaagg	aaaagaccag	agtagaggcc	tatagaagaa	2520
actgcgatac	cttttgtgat	ggctaaacaa	acagacatct	TTTTatatgt	TTTTacttct	2580
gtatatcgtg	aagtagtaag	tgataagcga	atttggctaa	gaacgttgta	agtgaacaag	2640
ggacctcttt	tgcttttcaa	aaaaggatta	aatggagtta	atcattgaga	tttagttttc	2700
gtagatttct	gtatccctaa	ataactccct	taccgcacgg	gaaggcacia	aagacttgaa	2760
taatagcaaa	cggccagtag	ccaagaccaa	ataatactag	agttaactga	tggtcttaaa	2820
caggcattac	gtggtgaact	ccaagaccaa	tatacaaaat	atcgataagt	tattcttgcc	2880
caccaattta	aggagcctac	atcaggacag	tagtaccatt	cctcagagaa	gaggatataca	2940
taacaagaaa	atcgcgtgaa	caccttatat	aacttagccc	gttattgagc	taaaaaacct	3000
tgcaaaattt	cctatgaata	agaatacttc	agacgtgata	aaaatttact	ttctaactct	3060
tctcacgctg	cccctatctg	ttcttccgct	ctaccgtgag	aaataaagca	tcgagtacgg	3120
cagttcgctg	tactgaact	aaaacaataa	ggctagttcg	aatgatgaac	ttgcttgctg	3180
tcaaacttct	gagttgccgc	tgatgtgaca	ctgtgacaat	aaattcaaac	cggttatagc	3240
ggtctcctcc	ggtaccggtt	ctgccacctc	caatagagct	cagtaggagt	cagaacctct	3300
gcggtggctg	tcagtgaact	atccgcgttt	cgtaagttgt	gcgcgtgcac	atttcgcccc	3360
ttccccgtca	tcttgagca	ggcggaaatt	ttcatcacgc	tgtaggacgc	aaaaaaaaaa	3420
taattaatcg	tacaagaatc	ttggaaaaaa	aattgaaaaa	TTTTgtataa	aagggatgac	3480
ctaacttgac	tcaatggctt	ttacaccag	tattttccct	ttccttggtt	gttacaatta	3540
tagaagcaag	acaaaaacat	atagacaacc	tatttcctagg	agttatatTT	TTTTacccta	3600
ccagcaatat	aagtaaaaaa	ctgttttaac	agtatggcag	ttacaatgta	ttatgaagat	3660
gatgtagaag	tatcagcact	tgctggaaag	caaattgcag	taatcggtta	tggttcacaa	3720
ggacatgctc	acgcacagaa	tttgcgtgat	tctggtcaca	acgttatcat	tggtgtgcgc	3780
cacggaaaat	cttttgataa	agcaaaagaa	gatggctttg	aaacatttga	agtaggagaa	3840
gcagtagcta	aagctgatgt	tattatgggt	ttggcaccag	atgaacttca	acaatccatt	3900
tatgaagagg	acatcaaacc	aaacttgaaa	gcaggttcag	cacttggttt	tgctcacgga	3960
tttaatatcc	attttggcta	tattaaagta	ccagaagacg	ttgacgtctt	tatggttgcg	4020

20110907_CL5178USNA_SeqList.txt

cctaaggctc	caggtcacct	tgtccgtcgg	acttatactg	aagggttttg	tacaccagct	4080
ttgtttgttt	cacaccaaaa	tgcaagtgg	catgcgcgtg	aaatcgcaat	ggattggggc	4140
aaaggaattg	gttggtctcg	agtgggaatt	attgaaacaa	cttttaaaga	agaaacagaa	4200
gaagatttgt	ttggagaaca	agctgttcta	tgtggagggt	tgacagcact	tgttgaagcc	4260
ggttttgaaa	cactgacaga	agctggatac	gctggcgaat	tggcttactt	tgaagttttg	4320
cacgaaatga	aattgattgt	tgacctcatg	tatgaagggt	gttttactaa	aatgcgtcaa	4380
tccatctcaa	atactgctga	gtttggcgat	tatgtgactg	gtccacggat	tattactgac	4440
gaagttaaaa	agaatatgaa	gcttgttttg	gctgatattc	aatctggaaa	atttgctcaa	4500
gatttcgttg	atgacttcaa	agcggggcgt	ccaaaattaa	tagcctatcg	cgaagctgca	4560
aaaaatcttg	aaattgaaaa	aattggggca	gagctacgtc	aagcaatgcc	attcacacaa	4620
tctggtgatg	acgatgcctt	taaaatctat	cagtaaggcc	ctgcaggcca	gaggaaaata	4680
atatcaagtg	ctggaaactt	tttctcttgg	aatttttgca	acatcaagtc	atagtcaatt	4740
gaattgaccc	aatttcacat	ttaagatttt	ttttttttca	tccgacatac	atctgtacac	4800
taggaagccc	tgtttttctg	aagcagcttc	aaatatatat	attttttaca	tatttattat	4860
gattcaatga	acaatctaata	taaatcga	acaagaaccg	aaacgcgaat	aaataattta	4920
tttagatgg	gacaagtgt	taagtcctca	tcgggacagc	tacgatttct	ctttcggttt	4980
tggtgagct	actggttgc	gtgacgcagc	ggcattagcg	cggcgttatg	agctaccctc	5040
gtggcctgaa	agatggcggg	aataaagcgg	aactaaaaat	tactgactga	gccatattga	5100
ggtcaatttg	tcaactcgtc	aagtcacgtt	tggtggacgg	cccctttcca	acgaatcgta	5160
tatactaaca	tgcgcgcgct	tcctatatac	acatacatat	atatatatat	atatatatgt	5220
gtgcgtgtat	gtgtacacct	gtatttaatt	tccttactcg	cgggtttttc	ttttttctca	5280
attcttggct	tcctctttct	cgagcggacc	ggatcctccg	cggtgccggc	agatctattt	5340
aaatggcgcg	ccgacgtcag	gtggcacttt	tcggggaaat	gtgcgcggaa	cccctatttg	5400
tttatttttc	taaatacatt	caaatatgta	tccgctcatg	agacaataac	cctgataaat	5460
gcttcaataa	tattgaaaaa	ggaagagtat	gagtattcaa	catttccgtg	tcgcccttat	5520
tccttttttt	gcggcatttt	gccttcctgt	ttttgtctac	ccagaaacgc	tggtgaaagt	5580
aaaagatgct	gaagatcagt	tggtgacag	agtgggttac	atcgaactgg	atctcaacag	5640
cggtaagatc	cttgagagtt	ttcgccccga	agaacgtttt	ccaatgatga	gcacttttaa	5700
agttctgcta	tgtggcgcg	tattatcccg	tattgacgcc	gggcaagagc	aactcggtcg	5760
ccgcatacac	tattctcaga	atgacttgg	tgagtactca	ccagtcacag	aaaagcatct	5820
tacggatggc	atgacagtaa	gagaattatg	cagtgtctgc	ataacctga	gtgataaac	5880
tgcgccaac	ttacttctga	caacgatcgg	aggaccgaag	gagctaaccg	cttttttgca	5940
caacatgggg	gatcatgtaa	ctcgccttga	tcgttgggaa	ccggagctga	atgaagccat	6000
accaaacgac	gagcgtgaca	ccacgatgcc	tgtagcaatg	gcaacaacgt	tgcgcaaact	6060

20110907_CL5178USNA_SeqList.txt

attaactggc	gaactactta	ctctagcttc	ccggcaacaa	ttaatagact	ggatggaggc	6120
ggataaagtt	gcaggaccac	ttctgcgctc	ggcccttccg	gctggctggg	ttattgctga	6180
taaatctgga	gccggtgagc	gtgggtctcg	cggtatcatt	gcagcactgg	ggccagatgg	6240
taagccctcc	cgtatcgtag	ttatctacac	gacggggagt	caggcaacta	tggatgaacg	6300
aaatagacag	atcgtgaga	taggtgcctc	actgattaag	cattggtaac	tgtcagacca	6360
agtttactca	tatatacttt	agattgattt	aaaacttcat	ttttaattta	aaaggatcta	6420
ggatgaagatc	ctttttgata	atctcatgac	caaaatccct	taacgtgagt	tttcgttcca	6480
ctgagcgtca	gaccccgtag	aaaagatcaa	aggatcttct	tgagatcctt	tttttctgcg	6540
cgtaatctgc	tgcttgcaaa	caaaaaaacc	accgctacca	gcggtgggtt	gtttgccgga	6600
tcaagagcta	ccaactcttt	ttccgaaggt	aactggcttc	agcagagcgc	agataccaaa	6660
tactgttctt	ctagtgtagc	cgtagttagg	ccaccacttc	agaactctg	tagcaccgcc	6720
tacatacctc	gctctgctaa	tcctgttacc	agtggctgct	gccagtggcg	ataagtcgtg	6780
tcttaccggg	ttggactcaa	gacgatagtt	accggataag	gcgagcggt	cgggctgaac	6840
gggggggttcg	tgcacacagc	ccagcttgga	gcgaacgacc	tacaccgaac	tgagatacct	6900
acagcgtgag	ctatgagaaa	gcgccacgct	tcccgaagg	agaaaggcgg	acaggtatcc	6960
ggtaagcggc	agggtcggaa	caggagagcg	cacgaggag	cttccagggg	gaaacgcctg	7020
gtatctttat	agtcctgtcg	ggtttcgcca	cctctgactt	gagcgtcgat	ttttgtgatg	7080
ctcgtcagg	gggcggagcc	tatggaaaaa	cgccagcaac	gcggcctttt	tacggttcct	7140
ggccttttgc	tggccttttg	ctcacatgtt	ctttcctgcg	ttatcccctg	attctgtgga	7200
taaccgtatt	accgcctttg	agtgagctga	taccgctcgc	cgcagccgaa	cgaccgagcg	7260
cagcgagtca	gtgagcgagg	aagcggaaga	gcgccaata	cgcaaaccgc	ctctccccgc	7320
gcgttgggcg	attcattaat	gcagctggca	cgacaggttt	cccgactgga	aagcgggcag	7380
tgagcgcaac	gcaattaatg	tgagttagct	cactcattag	gcaccccagg	ctttacactt	7440
tatgcttccg	gctcgtatgt	tgtgtggaat	tgtgagcgga	taacaatttc	acacaggaaa	7500
cagctatgac	catgattacg	ccaagctttt	tctttccaat	tttttttttt	tcgtcattat	7560
aaaaatcatt	acgaccgaga	ttcccgggta	ataactgata	taattaaatt	gaagctctaa	7620
tttgtgagtt	tagtatacat	gcatttactt	ataatacagt	tttttagttt	tgctggccgc	7680
atcttctcaa	atatgcttcc	cagcctgctt	ttctgtaacg	ttcacctctt	accttagcat	7740
cccttccctt	tgcaaatagt	cctcttccaa	caataataat	gtcagatcct	gtagagacca	7800
catcatccac	ggttctatac	tgttgacca	atgcgtctcc	cttgtcatct	aaaccacac	7860
cgggtgtcat	aatcaaccaa	tcgtaacctt	catctcttcc	acccatgtct	ctttgagcaa	7920
taaagccgat	aacaaaatct	ttgtcgctct	tcgcaatgtc	aacagtaccc	ttagtatatt	7980
ctccagtaga	tagggagccc	ttgcatgaca	attctgctaa	catcaaaagg	cctctaggtt	8040
cctttgttac	ttcttctgcc	gcctgcttca	aaccgctaac	aatacctggg	cccaccacac	8100

20110907_CL5178USNA_SeqList.txt

cgtgtgcatt	cgtaatgtct	gccattcttg	ctattctgta	tacacccgca	gagtactgca	8160
atttgactgt	attaccaatg	tcagcaaatt	ttctgtcttc	gaagagtaaa	aaattgtact	8220
tggcggataa	tgccttttagc	ggcttaactg	tgccctccat	ggaaaaatca	gtcaagatat	8280
ccacatgtgt	ttttagtaaa	caaatttttg	gacctaattgc	ttcaactaac	tccagtaatt	8340
ccttggtggg	acgaacatcc	aatgaagcac	acaagtttgt	ttgcttttcg	tgcattgatat	8400
taaatagctt	ggcagcaaca	ggactaggat	gagtagcagc	acgttcctta	tatgtagctt	8460
tcgacatgat	ttatcttcgt	ttcctgcagg	tttttgttct	gtgcagttgg	gttaagaata	8520
ctgggcaatt	tcatgtttct	tcaacactac	atatgcgtat	atataccaat	ctaagtctgt	8580
gctccttcct	tcgttcttcc	ttctgttcgg	agattaccga	atcaaaaaaa	tttcaaggaa	8640
accgaaatca	aaaaaaagaa	taaaaaaaaa	atgatgaatt	gaaaagcttg	catgcctgca	8700
ggctgactct	agtatactcc	gtctactgta	cgatacactt	ccgctcaggt	ccttgtcctt	8760
taacgaggcc	ttaccactct	tttgttactc	tattgatcca	gctcagcaaa	ggcagtggtga	8820
tctaagattc	tatcttcgcg	atgtagtaaa	actagctaga	ccgagaaaga	gactagaaat	8880
gcaaaaggca	cttctacaat	ggctgccatc	attattatcc	gatgtgacgc	tgcatttttt	8940
tttttttttt	tttttttttt	tttttttttt	tttttttttt	ttttttgtac	aaatatcata	9000
aaaaaagaga	atctttttta	gcaaggattt	tcttaacttc	ttcggcgaca	gcatcaccga	9060
cttcggtggg	actgttgtaa	ccacctaaat	caccagttct	gatacctgca	tccaaaacct	9120
ttttaactgc	atcttcaatg	gctttacctt	cttcaggcaa	gttcaatgac	aatttcaaca	9180
tcattgcagc	agacaagata	gtggcgatag	ggttgacctt	attctttggc	aaatctggag	9240
cggaaccatg	gcatggttcg	tacaaaccaa	atgcggtggt	cttgtctggc	aaagaggcca	9300
aggacgcaga	tggcaacaaa	ccaaggagc	ctgggataac	ggaggcttca	tcggagatga	9360
tatcaccaaa	catgttgctg	gtgattataa	taccatttag	gtgggttggg	ttcttaacta	9420
ggatcatggc	ggcagaatca	atcaattgat	gttgaacttt	caatgtaggg	aattcgttct	9480
tgatggtttc	ctccacagtt	tttctccata	atcttgaaga	ggccaaaaca	ttagctttat	9540
ccaaggacca	aataggcaat	ggtggctcat	gttgtagggc	catgaaagcg	gccattcttg	9600
tgattctttg	cacttctgga	acggtgtatt	gttcactatc	ccaagcgaca	ccatcaccat	9660
cgtcttcctt	tctcttacca	aagtaaatac	ctcccactaa	ttctctaaca	acaacgaagt	9720
cagtaccttt	agcaaattgt	ggcttgattg	gagataagtc	taaaagagag	tcggatgcaa	9780
agttacatgg	tcttaagttg	gcgtacaatt	gaagttcttt	acggattttt	agtaaacctt	9840
gttcaggctt	aacactaccg	gtaccccat	taggaccacc	cacagcacct	aacaaaacgg	9900
catcagcctt	cttggaggct	tccagcgcct	catctggaag	tggaacacct	gtagcatcga	9960
tagcagcacc	accaattaaa	tgattttcga	aatcgaactt	gacattggaa	cgaacatcag	10020
aaatagcttt	aagaacctta	atggcttcgg	ctgtgatttc	ttgaccaacg	tggtcacctg	10080
gcaaaacgac	gatcttctta	ggggcgagaca	ttacaatggt	atataccttga	aatatatata	10140

20110907_CL5178USNA_SeqList.txt

aaaaaaaaa	aaaaaaaaa	aaaaaaaaa	gcagcttctc	aatgatattc	gaatacgctt	10200
tgaggagata	cagcctaata	tccgacaaac	tgttttacag	atttacgatac	gtacttggtta	10260
cccatcattg	aattttgaac	atccgaacct	gggagttttc	cctgaaacag	atagtatat	10320
tgaacctgta	taataatata	tagtctagcg	ctttacggaa	gacaatgtat	gtatttcggt	10380
tcctggagaa	actattgcat	ctattgcata	ggtaatcttg	cacgtcgcat	ccccggttca	10440
ttttctgctg	ttccatcttg	cacttcaata	gcatactttt	gttaacgaag	catctgtgct	10500
tcattttgta	gaacaaaaat	gcaacgcgag	agcgctaatt	tttcaaaca	agaatctgag	10560
ctgcattttt	acagaacaga	aatgcaacgc	gaaagcgcta	ttttaccaac	gaagaatctg	10620
tgcttcattt	ttgtaaaaca	aaaatgcaac	gcgagagcgc	taatttttca	aacaaagaat	10680
ctgagctgca	tttttacaga	acagaaatgc	aacgcgagag	cgctatttta	ccaacaaaga	10740
atctatactt	cttttttgtt	ctacaaaaat	gcatcccgag	agcgctattt	ttctaacaaa	10800
gcactcttaga	ttactttttt	tctcctttgt	gcgctctata	atgcagtctc	ttgataactt	10860
tttgcactgt	aggtccggtta	aggttagaag	aaggctactt	tgggtgtctat	tttctcttcc	10920
ataaaaaaag	cctgactcca	cttcccgctg	ttactgatta	ctagcgaagc	tgcggggtgca	10980
ttttttcaag	ataaaggcat	ccccgattat	attctatacc	gatgtggatt	gcgcatactt	11040
tgtgaacaga	aagtgatagc	gttgatgatt	cttcattggt	cagaaaatta	tgaacggttt	11100
cttctatttt	gtctctatat	actacgtata	ggaaatgttt	acatttttcgt	attgttttcg	11160
attcactcta	tgaatagttc	ttactacaat	ttttttgtct	aaagagtaat	actagagata	11220
aacataaaaa	atgtagaggt	cgagtttaga	tgcaagttca	aggagcgaaa	ggtggatggg	11280
taggttatat	agggatatag	cacagagata	tatagcaaag	agatactttt	gagcaatggt	11340
tgtggaagcg	gtattcgcaa	tatttttagta	gctcgtttaca	gtccggtgcg	tttttggttt	11400
tttgaaagtg	cgtcttcaga	gcgcttttgg	ttttcaaaag	cgctctgaag	ttcctatact	11460
ttctagagaa	taggaacttc	ggaataggaa	cttcaaagcg	tttccgaaaa	cgagcgcttc	11520
cgaaaatgca	acgcgagctg	cgcacataca	gctcactggt	cacgtcgcac	ctatatctgc	11580
gtgttgccctg	tatatatata	tacatgagaa	gaacggcata	gtgcgtgttt	atgcttaaat	11640
gcgtacttat	atgcgtctat	ttatgtagga	tgaaaggtag	tctagtacct	cctgtgatat	11700
tatcccattc	catgcggggg	atcgtagtct	tccttcagca	ctacccttta	gctgttctat	11760
atgctgccac	tcctcaattg	gattagtctc	atccttcaat	gctatcattt	cctttgatat	11820
tggatcatat	gcatagtacc	gagaaactag	aggatc			11856

<210> 135
 <211> 15539
 <212> DNA
 <213> Artificial sequence

<220>
 <223> Plasmid

20110907_CL5178USNA_SeqList.txt

<400> 135
tcgcgcgttt cggtgatgac ggtgaaaacc tctgacacat gcagctcccg gagacggtca 60
cagcttgtct gtaagcggat gccgggagca gacaagcccg tcagggcgcg tcagcgggtg 120
ttggcgggtg tcggggctgg cttactatg cggcatcaga gcagattgta ctgagagtgc 180
accataaatt cccgttttaa gagcttggtg agcgctagga gtcactgcca ggtatcgttt 240
gaacacggca ttagtcaggg aagtcataac acagtccttt cccgcaattt tctttttcta 300
ttactcttgg cctcctctag tacactctat atttttttat gcctcggtaa tgattttcat 360
tttttttttt ccacctagcg gatgactctt tttttttctt agcgattggc attatcacat 420
aatgaattat acattatata aagtaatgtg atttcttcga agaataact aaaaaatgag 480
caggcaagat aaacgaaggc aaagatgaca gagcagaaag ccctagtaaa gcgtattaca 540
aatgaaacca agattcagat tgcgatctct ttaaagggtg gtcccctagc gatagagcac 600
tcgatcttcc cagaaaaaga ggcagaagca gtagcagaac aggccacaca atcgcaagtg 660
attaacgtcc acacaggtat agggtttctg gaccatatga tacatgctct ggccaagcat 720
tccggctggt cgctaactgt tgagtgcatt ggtgacttac acatagacga ccatcacacc 780
actgaagact gcgggattgc tctcgggtcaa gcttttaaag aggccctagg ggccgtgcgt 840
ggagtaaaaa ggtttggatc aggatttgcg cctttggatg aggcactttc cagagcgggtg 900
gtagatcttt cgaacaggcc gtacgcagtt gtcgaacttg gtttgcaaag ggagaaagta 960
ggagatctct ctgagcagat gatcccgcat tttcttgaaa gctttgcaga ggctagcaga 1020
attaccctcc acgttgattg tctgagaggc aagaatgac atcacgtag tgagagtgcg 1080
ttcaaggctc ttgcggttgc cataagagaa gccacctcgc ccaatggtac caacgatgtt 1140
ccctccacca aaggtgttct tatgtagtga caccgattat ttaaagctgc agcatagat 1200
atatatacat gtgtatatat gtatacctat gaatgtcagt aagtatgtat acgaacagta 1260
tgatactgaa gatgacaagg taatgcatca ttctatacgt gtcattctga acgaggcgcg 1320
ctttcctttt ttctttttgc tttttctttt tttttctctt gaactcgacg gatctatgcg 1380
gtgtgaaata ccgcacagat gcgtaaggag aaaataccgc atcaggaaat tgtaagcggtt 1440
aatattttgt taaaattcgc gttaaatttt tgttaaataca gctcattttt taaccaatag 1500
gccgaaatcg gcaaaatccc ttataaatca aaagaataga ccgagatagg gttgagtgtt 1560
gttccagttt ggaacaagag tccactatta aagaacgtgg actccaacgt caaagggcga 1620
aaaaccgtct atcagggcga tggcccacta cgtgaaccat caccctaatac aagttttttg 1680
gggtcgaggt gccgtaaagc actaaatcgg aaccctaaag ggagcccccg atttagagct 1740
tgacggggaa agccggcgaa cgtggcgaga aaggaagga agaaagcgaa aggagcgggc 1800
gctagggcgcg tggcaagtgt agcgggtcacg ctgcgcgtaa ccaccacacc cgccgcgctt 1860
aatgcgccgc tacagggcgcg gtccattcgc cattcaggct gcgcaactgt tgggaagggc 1920
gcggtgcggg cctcttcgct attacgccag ctggcgaaag ggggatgtgc tgcaaggcga 1980
ttaagttggg taacgccagg gttttcccag tcacgacgtt gtaaaacgac ggccagttag 2040

20110907_CL5178USNA_SeqList.txt

cgcgcgtaat	acgactcact	atagggcgaa	ttgggtaccg	ggccccccct	cgaggctcgac	2100
ggcgcgccac	tggtagagag	cgactttgta	tgccccaatt	gcgaaaccg	cgatatcctt	2160
ctcgattctt	tagtaccga	ccaggacaag	gaaaaggagg	tcgaaacgtt	tttgaagaaa	2220
caagaggaa	tacacggaag	ctctaaagat	ggcaaccagc	cagaaactaa	gaaaatgaag	2280
ttgatggatc	caactggcac	cgctggcttg	aacaacaata	ccagccttcc	aacttctgta	2340
aataacggcg	gtacgccagt	gccaccagta	ccgttacctt	tcggtatacc	tcctttcccc	2400
atgtttccaa	tgcccttcat	gcctccaacg	gctactatca	caaatcctca	tcaagctgac	2460
gcaagcccta	agaaatgaat	aacaatactg	acagtactaa	ataattgcct	acttggcttc	2520
acatacgttg	catacgtcga	tatagataat	aatgataatg	acagcaggat	tatcgtaata	2580
cgtaatagct	gaaaatctca	aaaatgtgtg	ggtcattacg	taaataatga	taggaatggg	2640
attcttctat	ttttcctttt	tccattctag	cagccgtcgg	gaaaacgtgg	catcctctct	2700
ttcgggctca	attggagtca	cgctgccgtg	agcatcctct	ctttccatat	ctaacaactg	2760
agcacgtaac	caatggaaaa	gcatgagctt	agcgttgctc	caaaaaagta	ttggatgggt	2820
aataccattt	gtctgttctc	ttctgacttt	gactcctcaa	aaaaaaaaat	ctacaatcaa	2880
cagatcgctt	caattacgcc	ctcacaaaaa	cttttttctt	tcttcttcgc	ccacgttaaa	2940
ttttatccct	catgttgtct	aacggatttc	tgcacttgat	ttattataaa	aagacaaaga	3000
cataatactt	ctctatcaat	ttcagttatt	gttcttcctt	gcgttattct	tctgttcttc	3060
tttttctttt	gtcatatata	accataacca	agtaatacat	attcaaacta	gtatgactga	3120
caaaaaaact	cttaaagact	taagaaatcg	tagttctggt	tacgattcaa	tggttaaatac	3180
acctaatacgt	gctatgttgc	gtgcaactgg	tatgcaagat	gaagactttg	aaaaacctat	3240
cgtcggtgtc	atttcaactt	gggctgaaaa	cacaccttgt	aatatccact	tacatgactt	3300
tggtaaaacta	gccaaagtcg	gtgttaagga	agctgggtgt	tggccagttc	agttcggaac	3360
aatcacgggt	tctgatggaa	tcgccatggg	aaccaagga	atgcgtttct	ccttgacatc	3420
tcgtgatatt	attgcagatt	ctattgaagc	agccatggga	ggtcataatg	cggatgcttt	3480
tgtagccatt	ggcggttgtg	ataaaaacat	gcccggttct	gttatcgcta	tggctaacat	3540
ggatatccca	gccatttttg	cttacggcgg	aacaattgca	cctggtaatt	tagacggcaa	3600
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	aatgggctt	aaaaccttct	gacattttta	cgcggtgaagc	3900
ttttgaagat	gctattactg	taactatggc	tctgggaggt	tcaaccaact	caacccttca	3960
cctcttagct	attgcccatg	ctgctaattg	ggaattgaca	cttgatgatt	tcaatacttt	4020
ccaagaaaaa	gttcctcatt	tggctgattt	gaaaccttct	ggtcaatatg	tattccaaga	4080

20110907_CL5178USNA_SeqList.txt

cctttacaag	gtcggagggg	taccagcagt	tatgaaatat	ctccttaaaa	atggcttcct	4140
tcattggtgac	cgtatcactt	gtactggcaa	aacagtcgct	gaaaatttga	aggcttttga	4200
tgatttaaca	cctgggtcaaa	aggttattat	gccgcttgaa	aatcctaaac	gtgaagatgg	4260
tccgctcatt	attctccatg	gtaacttggc	tccagacggg	gccgttgcca	aagtttctgg	4320
tgtaaaagt	cgctgcatg	tcgggtcctgc	taagggtcttt	aattctgaag	aagaagccat	4380
tgaagctgtc	ttgaatgatg	atattgttga	tggtgatggt	gttgctgtac	gtttttagg	4440
accaaagggc	ggccttggtg	tgccctgaaat	gctttccctt	tcatcaatga	ttgttggtaa	4500
agggaaggt	gaaaaagttg	cccttctgac	agatggccgc	ttctcagggtg	gtacttatgg	4560
tcttgctgtg	ggcatatcg	ctcctgaagc	acaagatggc	ggccaatcg	cctacctgca	4620
aacaggagac	atagtcacta	ttgaccaaga	cactaaggaa	ttacactttg	atatctccga	4680
tgaagagtta	aaacatcgct	aagagaccat	tgaattgcca	ccgctctatt	cacgcggtat	4740
ccttggtaaa	tatgctcaca	tcgtttcgtc	tgcttctagg	ggagccgtaa	cagacttttg	4800
gaagcctgaa	gaaactggca	aaaaatgttg	tcctggttgc	tgtggttaag	cggccgcgtt	4860
aattcaaatt	aattgatata	gttttttaat	gagtattgaa	tctgtttaga	aataatggaa	4920
tattattttt	atttatttat	ttatattatt	ggtcggctct	tttcttctga	aggatcaatga	4980
caaaatgata	tgaaggaaat	aatgatttct	aaaattttac	aacgtaagat	atttttacaa	5040
aagcctagct	catcttttgt	catgcactat	tttactcacg	cttgaaatta	acggccagtc	5100
cactgcggag	tcatttcaaa	gtcatcctaa	tcgatctatc	gtttttgata	gctcattttg	5160
gagttcgca	ttgtcttctg	ttattcacaa	ctgttttaat	ttttatttca	ttctggaact	5220
cttcgagttc	tttgtaaagt	ctttcatagt	agcttacttt	atcctccaac	atattttaact	5280
tcatgtcaat	ttcgggtctt	aaattttcca	catcatcaag	ttcaacatca	tctttttaact	5340
tgaatttatt	ctctagctct	tccaaccaag	cctcattgct	ccttgattta	ctggtgaaaa	5400
gtgatacact	ttgcgcgcaa	tccagggtcaa	aactttcctg	caaagaattc	accaatttct	5460
cgacatcata	gtacaatttg	ttttgttctc	ccatcacaa	ttaatatacc	tgatggattc	5520
ttatgaagcg	ctgggtaatg	gacgtgtcac	tctacttcgc	ctttttccct	actcctttta	5580
gtacggaaga	caatgcta	aaataagagg	gtaataataa	tattattaat	cggcaaaaaa	5640
gattaaacgc	caagcgttta	attatcagaa	agcaaacgct	gtaccaatcc	ttgaatgctt	5700
cccaattgta	tattaagagt	catcacagca	acatattctt	gttattaaat	taattattat	5760
tgatttttga	tattgtataa	aaaaaccaa	tatgtataaa	aaaagtgaat	aaaaaatacc	5820
aagtatggag	aaatatatta	gaagtctata	cgttaaacca	cccgggcccc	ccctcgaggt	5880
cgacgggtatc	gataagcttg	atctgaatt	cctgcagccc	gggggatcca	ctagttctag	5940
agcggccgct	ctagaactag	taccacaggt	gttgctctct	gaggacataa	aatacacacc	6000
gagattcatc	aactcattgc	tggagttagc	atatctacaa	ttgggtgaaa	tggggagcga	6060
tttgaggga	tttgctcggc	atgccggtag	agggtgtggtc	aataagagcg	acctcatgct	6120

20110907_CL5178USNA_SeqList.txt

atacctgaga	aagcaacctg	acctacagga	aagagttact	caagaataag	aatttttcggt	6180
ttaaaacctta	agagtcactt	taaaatttgt	atacacttat	tttttttata	acttatattaa	6240
taataaaaaat	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	tcacgcacaa	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	tccagtgggt	aagctgcac	gatttttagcg	acagatatct	cgtccaccac	6960
tgtgtattgg	gaaaatgtag	aagtaccaag	gaaatgggtg	ataggtttcc	ctctgcatgt	7020
aaatctgctt	gtaccatcct	gcatagtacc	tctaggcata	gacaaatcat	ttttaaggca	7080
gaaattaccc	tcaggatggt	tgcagactct	acacttacca	cattgaggag	tgaacagtgg	7140
gatcacttta	tcaccaggac	gaacagtgg	aacaccttca	cctatggatt	caacgattcc	7200
ggcagcctcg	tgtcccgcga	ttactggcaa	aggagtaact	agagtgccac	tcaccacatg	7260
gtcgtcggat	ctacagattc	cgggtggcaac	catcttgatt	ctaacctcgt	gtgcttttgg	7320
tggcgtact	tctacttctt	ctatgctaaa	cggctttttc	tcttcccaca	aaactgccgc	7380
tttacactta	ataactttac	cggctgttga	catcctcagc	tagctattgt	aatatgtgtg	7440
tttgtttga	ttattaagaa	gaataattac	aaaaaaaaatt	acaaaggaag	gtaattacaa	7500
cagaattaag	aaaggacaag	aaggaggaag	agaatcagtt	cattatttct	tctttgttat	7560
ataacaaacc	caagtagcga	tttgccata	cattaaaagt	tgagaaccac	cctccctggc	7620
aacagccaca	actcgttacc	attgttcac	acgatcatga	aactcgctgt	cagctgaaat	7680
ttcacctcag	tggatctctc	tttttattct	tcacgttcc	actaaccttt	ttccatcagc	7740
tggcagggaa	cggaaagtgg	aatcccattt	agcgagcttc	ctcttttctt	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	tcttctgtct	cacaaatctt	aaagtcatac	8160

20110907_CL5178USNA_SeqList.txt

attgcacgac	taaatgcaag	catgcggatc	ccccgggctg	caggaattcg	atatcaagct	8220
tatcgatacc	gtcgactggc	cattaatctt	tcccatatta	gatttcgcca	agccatgaaa	8280
gttcaagaaa	ggtcttttaga	cgaattaccc	ttcattttctc	aaactggcgt	caagggatcc	8340
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	aaacttttaa	aaaaaaagcc	aatatcccca	aaattattaa	9000
gagcgcctcc	attattaact	aaaatttcac	tcagcatcca	caatgtatca	ggtatctact	9060
acagatatta	catgtggcga	aaaagacaag	aacaatgcaa	tagcgcacat	agaaaaaaca	9120
caaagctttc	aatcaatgaa	tcgaaaatgt	cattaaaata	gtatataaat	tgaaactaag	9180
tcataaagct	ataaaaagaa	aatttattta	aatgcaagat	ttaaagtaaa	ttcacggccc	9240
tgcaggcctc	agctcttggt	ttgttctgca	aataacttac	ccatcttttt	caaaacttta	9300
ggtgcaccct	cctttgctag	aataagttct	atccaataca	tcctatttgg	atctgcttga	9360
gcttctttca	tcacggatac	gaattcattt	tctgtttctca	caattttgga	cacaactctg	9420
tcttccgttg	ccccgaaact	ttctggcagt	tttgagtaat	tccacatagg	aatgtcatta	9480
taactctggt	tcggaccatg	aatttccttc	tcaaccgtgt	aaccatcggt	attaatgata	9540
aagcagattg	ggtttatctt	ctctctaagt	gctagtccta	attcttggac	agtcagttgc	9600
aatgatccat	ctccgataaa	caataaatgt	ctagattctt	tatctgcaat	ttggctgcct	9660
agagctgcgg	ggaaagtgtg	tcctatagat	ccccacaagg	gttgaccaat	aaaatgtgat	9720
ttcgatttca	gaaatataga	tgaggcaccg	agaaaagaag	tgctttgttc	agccacgac	9780
gtctcattac	tttgggtcaa	attttcgaca	gcttgccaca	gtctatcttg	tgacaacagc	9840
gcgttagaag	gtacaaaatc	ttcttgcttt	ttatctatgt	acttgccttt	atattcaatt	9900
tcggacaagt	caagaagaga	tgatatcagg	gattcgaagt	cgaaattttg	gattctttcg	9960
ttgaaaattt	taccttcac	gatattcaag	gaaatcattt	tattttcatt	aagatggtga	10020
gtaaatgcac	ccgtactaga	atcggtgaag	tttacaccca	acataagaat	aaaatcagca	10080
gattccacaa	attccttcaa	gtttggctct	gacagagtac	cgttgtaaat	ccccaaaaat	10140
gagggcaatg	cttcatcaac	agatgattta	caaagtttca	aagtagtaat	aggtaactta	10200

20110907_CL5178USNA_SeqList.txt

gtctttgaaa	taaactgagt	aacagtcttc	tctaggccga	acgatataat	ttcatggcct	10260
gtgattacaa	ttggtttctt	ggcattcttc	agactttcct	gtattttggt	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	gcccgcctaag	ccattaactg	cggataattc	gccaacacca	10680
aatgtagtca	agaatgccgc	agcctttttc	gttcttgctg	acccgtcggc	catataggag	10740
gcatttaact	cattagcatt	tcccacccat	ttcatatctt	tgtgtgaaat	aatttgatct	10800
agaaaattgca	aattgtagtc	acctggtact	ccgaatatatt	cttctataacc	taattcgtgt	10860
aatctgtcca	acagatagtc	acctactgta	tacattttgt	ttactagttt	atgtgtgttt	10920
attcgaaact	aagtctcttg	tgttttaaaa	ctaaaaaaaa	gactaactat	aaaagtagaa	10980
tttaagaagt	ttaagaaata	gatttacaga	attacaatca	atacctaccg	tctttatata	11040
cttattagtc	aagtagggga	ataatttcag	ggaactgggt	tcaacctttt	ttttcagctt	11100
tttccaaatc	agagagagca	gaaggtaata	gaagggtgaa	gaaaatgaga	tagatacatg	11160
cgtgggtcaa	ttgccttggt	tcatcattta	ctccaggcag	gttgcatcac	tccattgagg	11220
ttgtgcccgt	tttttgccgt	tttgtgcccc	tgttctctgt	agttgcgcta	agagaatgga	11280
cctatgaact	gatggttggt	gaagaaaaca	atattttggt	gctgggattc	tttttttttc	11340
tggatgccag	cttaaaaagc	gggctccatt	atattttagt	gatgccagga	ataaactggt	11400
caccagaca	cctacgatgt	tatatattct	gtgtaaccgc	ccccctattt	tgggcatgta	11460
cgggttacag	cagaattaaa	aggctaattt	tttgactaaa	taaagttagg	aaaatcacta	11520
ctattaatta	ttacgtatt	ctttgaaatg	gcagtattga	taatgataaa	ctcgaactga	11580
aaaagcgtgt	tttttattca	aatgattct	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	ttaaactctg	gctctcttgg	gctcaagggtg	11880
acaaggctct	cgaaaatagg	gcgcgcccc	ccgcggtgga	gctccagctt	ttgttccctt	11940
tagtgagggt	taattgcgcg	cttggcgtaa	tcatggtcac	agctgtttcc	tgtgtgaaat	12000
tggtatccgc	tcacaattcc	acacaacata	cgagccggaa	gcataaagtg	taaagcctgg	12060
ggtgcctaata	gagtgaagta	actcacatta	attgcgttgc	gctcactgcc	cgctttccag	12120
tcgggaaacc	tgtcgtgcca	gctgcattaa	tgaatcggcc	aacgcgcggg	gagaggcggt	12180
ttgcgtattg	ggcgcctctc	cgcttcctcg	ctcactgact	cgctgcgctc	ggtcgttcgg	12240

20110907_CL5178USNA_SeqList.txt

ctgcggcgag	cggtatcagc	tcactcaaag	gcggtaat	ggttatccac	agaatcaggg	12300
gataacgcag	gaaagaacat	gtgagcaaaa	ggccagcaaa	aggccaggaa	ccgtaaaaag	12360
gccgcgttgc	tggcggtttt	ccataggctc	cgccccctg	acgagcatca	caaaaatcga	12420
cgctcaagtc	agaggtggcg	aaacccgaca	ggactataaa	gataccaggc	gtttccccct	12480
ggaagctccc	tcgtgcgctc	tcctgttccg	accctgccgc	ttaccggata	cctgtccgcc	12540
tttctccctt	cgggaagcgt	ggcgctttct	catagctcac	gctgtaggta	tctcagttcg	12600
gtgtaggtcg	ttcgctccaa	gctgggctgt	gtgcacgaac	cccccgttca	gcccgaaccgc	12660
tgcgccttat	ccggtaaact	tcgtcttgag	tccaacccgg	taagacacga	cttatcgcca	12720
ctggcagcag	ccactggtaa	caggattagc	agagcgaggt	atgtaggcgg	tgctacagag	12780
ttcttgaagt	ggtggcctaa	ctacggctac	actagaagaa	cagtatttgg	tatctgcgct	12840
ctgctgaagc	cagttacctt	cggaaaaaga	gttggtagct	cttgatccgg	caaacaaacc	12900
accgctggta	gcggtggttt	ttttgtttgc	aagcagcaga	ttacgcgcag	aaaaaaagga	12960
tctcaagaag	atcctttgat	cttttctacg	gggtctgacg	ctcagtggaa	cgaaaactca	13020
cgtaagggga	ttttggtcat	gagattatca	aaaaggatct	tcacctagat	ccttttaaat	13080
taaaaatgaa	gttttaaatc	aatctaaagt	atatatgagt	aaacttggtc	tgacagttac	13140
caatgcttaa	tcagtgaggc	acctatctca	gcgatctgtc	tatttcgttc	atccatagtt	13200
gcctgactcc	ccgtcgtgta	gataactacg	atacgggagg	gcttaccatc	tggccccagt	13260
gctgcaatga	taccgcgaga	cccacgtcga	ccggctccag	atttatcagc	aataaaccag	13320
ccagccggaa	gggccgagcg	cagaagtggg	cctgcaactt	tatccgcctc	catccagtct	13380
attaattggt	gccgggaagc	tagagtaagt	agttcgccag	ttaatagttt	gcgcaacggt	13440
gttgccattg	ctacaggcat	cgtggtgtca	cgctcgctcg	ttggtatggc	ttcattcagc	13500
tccggttccc	aacgatcaag	gcgagttaca	tgatccccc	tgttgtgcaa	aaaagcggtt	13560
agtccttctg	gtcctccgat	cgttgtcaga	agtaagttgg	ccgcagtgtt	atcactcatg	13620
gttatggcag	cactgcataa	ttctcttact	gtcatgccat	ccgtaagatg	cttttctgtg	13680
actggtgagt	actcaaccaa	gtcattctga	gaatagtgtg	tgcggcgacc	gagttgctct	13740
tgcccggcgt	caatacggga	taataccgcg	ccacatagca	gaactttaaa	agtgctcatc	13800
attggaaaac	gttcttcggg	gcgaaaactc	tcaaggatct	taccgctgtt	gagatccagt	13860
tcgatgtaac	ccactcgtgc	acccaactga	tcttcagcat	cttttacttt	caccagcggt	13920
tctgggtgag	caaaaacagg	aaggcaaaat	gccgcaaaaa	agggaataag	ggcgacacgg	13980
aaatgttgaa	tactcatact	cttccttttt	caatattatt	gaagcattta	tcagggttat	14040
tgtctcatga	gcggatacat	atttgaatgt	atttagaaaa	ataaacaat	aggggttccg	14100
cgcacatttc	cccgaagagt	gccacctgaa	cgaagcatct	gtgcttcatt	ttgtagaaca	14160
aaaatgcaac	gcgagagcgc	taatttttca	aacaaagaat	ctgagctgca	tttttacaga	14220
acagaaatgc	aacgcgaaag	cgctatttta	ccaacgaaga	atctgtgctt	catttttgta	14280

20110907_CL5178USNA_SeqList.txt

aaacaaaaat	gcaacgcgag	agcgctaatt	tttcaaaca	agaatctgag	ctgcattttt	14340
acagaacaga	aatgcaacgc	gagagcgcta	ttttaccaac	aaagaatcta	tactttctttt	14400
ttgttctaca	aaaatgcatc	ccgagagcgc	tatTTTTtcta	acaaagcatc	ttagattact	14460
ttttttctcc	tttgtgcgct	ctataatgca	gtctcttgat	aactTTTTtg	actgtaggtc	14520
cgtaaggtt	agaagaaggc	tactttggtg	tctatTTTTt	cttcataaaa	aaaagcctga	14580
ctccacttcc	cgcgtttact	gattactagc	gaagctgcgg	gtgcattttt	tcaagataaa	14640
ggcatccccg	attatattct	ataccgatgt	ggattgcgca	tactttgtga	acagaaagtg	14700
atagcgttga	tgattcttca	ttggtcagaa	aattatgaac	ggtttcttct	atTTTgtctc	14760
tatatactac	gtataggaaa	tgtttacatt	ttcgtattgt	tttcgattca	ctctatgaat	14820
agttcttact	acaatttttt	tgtctaaaga	gtaatactag	agataaacat	aaaaaatgta	14880
gaggtcgagt	ttagatgcaa	gttcaaggag	cgaaagggtg	atgggtaggt	tatatagga	14940
tatagcacag	agatatatag	caaagagata	cttttgagca	atgtttgtgg	aagcgggtatt	15000
cgcaatattt	tagtagctcg	ttacagtccg	gtgcgttttt	ggTTTTttga	aagtgcgtct	15060
tcagagcgct	tttggttttc	aaaagcgctc	tgaagttcct	atactttcta	gagaatagga	15120
acttcggaat	aggaacttca	aagcgtttcc	gaaaacgagc	gcttccgaaa	atgcaacgcg	15180
agctgcgcac	atacagctca	ctgttcacgt	cgcacctata	tctgcgtggt	gcctgtatat	15240
atatatacat	gagaagaacg	gcatagtgcg	tgTTTTatgct	taaatgcgta	cttatatgcg	15300
tctatttatg	taggatgaaa	ggtagtctag	tacctcctgt	gatattatcc	cattccatgc	15360
ggggtatcgt	atgcttcctt	cagcactacc	ctttagctgt	tctatatgct	gccactcctc	15420
aattggatta	gtctcatcct	tcaatgctat	catttccttt	gatattggat	catactaaga	15480
aaccattatt	atcatgacat	taacctataa	aaataggcgt	atcacgaggc	cctttcgtc	15539

<210> 136
 <211> 4586
 <212> DNA
 <213> Artificial sequence

<220>
 <223> Template

<400> 136						
gggtaccgag	ctcgaattca	ctggccgtcg	ttttacaacg	tcgtgactgg	gaaaaccctg	60
gcgttaccca	acttaatcgc	cttgacgac	atcccccttt	cgccagctgg	cgtaatagcg	120
aagaggcccg	caccgatcgc	ccttcccaac	agttgcgcag	cctgaatggc	gaatggcgcc	180
tgatgcggta	ttttctcctt	acgcatctgt	gcggtatttc	acaccgcata	tggtgcactc	240
tcagtacaat	ctgctctgat	gccgcatagt	taagccagcc	ccgacacccg	ccaacacccg	300
ctgacgcgcc	ctgacgggct	tgtctgctcc	cggcatccgc	ttacagacaa	gctgtgaccg	360
tctccgggag	ctgcatgtgt	cagaggTTTT	caccgtcatc	accgaaacgc	gcgagacgaa	420
agggcctcgt	gatacgccta	tttttatagg	ttaatgtcat	gataataatg	gtttcttaga	480

20110907_CL5178USNA_SeqList.txt

cgtcaggtgg	cacttttcgg	ggaaatgtgc	gcggaacccc	tatttgttta	tttttctaaa	540
tacattcaaa	tatgtatccg	ctcatgagac	aataaccctg	ataaatgctt	caataatatt	600
gaaaaaggaa	gagtatgagt	attcaacatt	tccgtgtcgc	ccttattccc	ttttttgcgg	660
cattttgcct	tcctgttttt	gctcacccag	aaacgctggg	gaaagtaaaa	gatgctgaag	720
atcagttggg	tgcacgagtg	ggttacatcg	aactggatct	caacagcggg	aagatccttg	780
agagttttcg	ccccgaagaa	cgttttccaa	tgatgagcac	ttttaaagtt	ctgctatgtg	840
gcgcggtatt	atcccgtatt	gacgccgggc	aagagcaact	cggtcgccgc	atacactatt	900
ctcagaatga	cttggttgag	tactcaccag	tcacagaaaa	gcattcttacg	gatggcatga	960
cagtaagaga	attatgcagt	gctgccataa	ccatgagtga	taacactgcg	gccaaacttac	1020
ttctgacaac	gatcggagga	ccgaaggagc	taaccgcttt	tttgcacaa	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	cttcgggctg	gctggtttat	tgctgataaa	tctggagccg	1320
gtgagcgtgg	gtctcgcggt	atcattgcag	cactggggcc	agatggtaag	ccctcccgtg	1380
tcgtagttat	ctacacgacg	gggagtcagg	caactatgga	tgaacgaaat	agacagatcg	1440
ctgagatagg	tgcctcactg	attaagcatt	ggtaactgtc	agaccaagtt	tactcatata	1500
tactttagat	tgatttaaaa	cttcattttt	aatttaaaag	gatctaggtg	aagatccttt	1560
ttgataatct	catgacaaa	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
tgctaatacct	gttaccagtg	gctgctgcca	gtggcgataa	gtcgtgtctt	accgggttgg	1920
actcaagacg	atagttaccg	gataaggcgc	agcggtcggg	ctgaacgggg	ggttcgtgca	1980
cacagcccag	cttgagcgga	acgacctaca	ccgaactgag	atacctacag	cgtgagctat	2040
gagaaagcgc	cacgcttccc	gaaggagaaa	aggcggacag	gtatccggta	agcggcaggg	2100
tcggaacagg	agagcgcacg	agggagcttc	cagggggaaa	cgcttggtat	ctttatagtc	2160
ctgtcgggtt	tcgccacctc	tgacttgagc	gtcgattttt	gtgatgctcg	tcaggggggc	2220
ggagcctatg	gaaaaacgcc	agcaacgcgg	ccttttttacg	gttcctggcc	ttttgctggc	2280
cttttgctca	catgtttctt	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

20110907_CL5178USNA_SeqList.txt

ttaatgtgag	ttagctcact	cattaggcac	cccaggcttt	acactttatg	cttccggctc	2580
gtatgtttgtg	tggaattgtg	agcggataac	aatttcacac	aggaaacagc	tatgaccatg	2640
attacgccaa	gcttgcacgc	ctgcaggctc	actctagagg	atccccgcat	tgcggattac	2700
gtattctaata	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	ctcccgcctc	2940
ctcgcataaa	caatgaacac	tgggtcaatc	atagcctaca	caggtgaaca	gagtagcggt	3000
tatacagggg	ttatacgggt	attcctacgg	caaaaatttt	tcattttctaa	aaaaaaaaaag	3060
aaaaattttt	ctttccaacg	ctagaaggaa	aagaaaaatc	taattaaatt	gatttggtga	3120
ttttctgaga	gttccctttt	tcatatatcg	aattttgaat	ataaaaggag	atcgaaaaaa	3180
tttttctatt	caatctgttt	tctggtttta	tttgatagtt	tttttgtgta	ttattattat	3240
ggattagtag	tggtttatat	gggtttttct	gtataacttc	tttttatttt	agtttgttta	3300
atcttatttt	gagttacatt	atagttccct	aactgcaaga	gaagtaacat	taaaactcga	3360
gatgggtaag	gaaaagactc	acgttttcgag	gccgcgatta	aattccaaca	tggatgctga	3420
tttatatggg	tataaatggg	ctcgcgataa	tgctgggcaa	tcaggtgcga	caatctatcg	3480
attgtatggg	aagcccgatg	cgccagagtt	gtttctgaaa	catggcaaag	gtagcggtgc	3540
caatgatgtt	acagatgaga	tggtcagact	aaactggctg	acggaattta	tgcctcttcc	3600
gaccatcaag	cattttatcc	gtactcctga	tgatgcatgg	ttactcacca	ctgcgatccc	3660
cggcaaaaaca	gcattccagg	tattagaaga	atctcctgat	tcaggtgaaa	atattgttga	3720
tgcgctggca	gtgttcctgc	gccggttgca	ttcgattcct	gtttgtaatt	gtccttttaa	3780
cagcgatcgc	gtatttcgtc	tcgctcaggc	gcaatcacga	atgaataacg	gtttggttga	3840
tgcgagtgat	tttgatgacg	agcgtaatgg	ctggcctggt	gaacaagtct	ggaaagaaat	3900
gcataagctt	ttgccattct	caccggattc	agtcgtcact	catggtgatt	tctcacttga	3960
taaccttatt	tttgacgagg	ggaaattaat	aggttgtatt	gatgttggac	gagtcggaat	4020
cgcagaccga	taccaggatc	ttgccatcct	atggaactgc	ctcggtgagt	tttctccttc	4080
attacagaaa	cggctttttc	aaaaatatgg	tattgataat	cctgatatga	ataaattgca	4140
gtttcatttg	atgctcgatg	agtttttcta	agttaaactt	gatactacta	gattttttct	4200
cttcatttat	aaaatttttg	gttataattg	aagctttaga	agtatgaaaa	aatccttttt	4260
tttcattctt	tgcaaccaa	ataagaagct	tcttttattc	attgaaatga	tgaatataaa	4320
cctaacaaaa	gaaaaagact	cgaatatcaa	acattaaaaa	aaaataaaaag	aggttatctg	4380
ttttccatt	tagttggagt	ttgcattttc	taatagatag	aactctcaat	taatgtggat	4440
ttagtttctc	tgttcgtttt	tttttgtttt	gttctcactg	tatttacatt	tctatttagt	4500
atttagttat	tcatataatc	tataacttcg	tatagcatac	attatacgaa	gttatccagt	4560

gatgatacaa cgagttagcc aaggtg 4586

<210> 137
 <211> 80
 <212> DNA
 <213> Artificial sequence

<220>
 <223> Primer

<400> 137
 ttccggtttc tttgaaattt ttttgattcg gtaatctccg agcagaagga gcattgcgga 60
 ttacgtattc taatgttcag 80

<210> 138
 <211> 81
 <212> DNA
 <213> Artificial sequence

<220>
 <223> Primer

<400> 138
 gggtaataac tgatataatt aaattgaagc tctaatttgt gagtttagta caccttggct 60
 aactcgttgt atcatcactg g 81

<210> 139
 <211> 38
 <212> DNA
 <213> Artificial sequence

<220>
 <223> Primer

<400> 139
 gcctcgagtt ttaatgttac ttctcttgca gttaggga 38

<210> 140
 <211> 31
 <212> DNA
 <213> Artificial sequence

<220>
 <223> Primer

<400> 140
 gctaaattcg agtgaaacac aggaagacca g 31

<210> 141
 <211> 24
 <212> DNA
 <213> Artificial sequence

<220>
 <223> Primer

<400> 141
 ttctcgacgt gggccttttt cttg 24

<210> 142
 <211> 49
 <212> DNA
 <213> Artificial sequence

 <220>
 <223> Primer

 <400> 142
 tgcagcttta aataatcggg gtcactactt tgccttcggt tatcttgcc 49

 <210> 143
 <211> 49
 <212> DNA
 <213> Artificial sequence

 <220>
 <223> Primer

 <400> 143
 gagcaggcaa gataaacgaa ggcaaagtag tgacaccgat tatttaaag 49

 <210> 144
 <211> 49
 <212> DNA
 <213> Artificial sequence

 <220>
 <223> Primer

 <400> 144
 tatggaccct gaaaccacag ccacattgta accaccacga cggttgttg 49

 <210> 145
 <211> 49
 <212> DNA
 <213> Artificial sequence

 <220>
 <223> Primer

 <400> 145
 tttagcaaca accgtcgtgg tggttacaat gtggctgtgg tttcagggt 49

 <210> 146
 <211> 49
 <212> DNA
 <213> Artificial sequence

 <220>
 <223> Primer

 <400> 146
 ccagaaacc tatactgtg tggacgtaag gccatgaagc tttttcttt 49

 <210> 147
 <211> 49
 <212> DNA
 <213> Artificial sequence

 <220>
 <223> Primer

<400> 147
attggaaaga aaaagcttca tggccttacg tccacacagg tatagggtt 49

<210> 148
<211> 22
<212> DNA
<213> Artificial sequence

<220>
<223> Primer

<400> 148
cataagaaca cctttggtgg ag 22

<210> 149
<211> 22
<212> DNA
<213> Artificial sequence

<220>
<223> Primer

<400> 149
aggattatca ttcataagtt tc 22

<210> 150
<211> 23
<212> DNA
<213> Artificial sequence

<220>
<223> Primer

<400> 150
ttcttgagc tgggacatgt ttg 23

<210> 151
<211> 22
<212> DNA
<213> Artificial sequence

<220>
<223> Primer

<400> 151
tgatgatatt tcataaataa tg 22

<210> 152
<211> 23
<212> DNA
<213> Artificial sequence

<220>
<223> Primer

<400> 152
atgcgtccat ctttacagtc ctg 23

<210> 153
<211> 22
<212> DNA
<213> Artificial sequence

```

<220>
<223> Primer

<400> 153
tacgtacgga ccaatcgaag tg
22

<210> 154
<211> 49
<212> DNA
<213> Artificial sequence

<220>
<223> Primer

<400> 154
aatcgtttg agtacactac taatggcttt gttggcaata tgtttttgc
49

<210> 155
<211> 49
<212> DNA
<213> Artificial sequence

<220>
<223> Primer

<400> 155
atatagcaaa aacatattgc caacaaagcc attagtagtg tactcaaac
49

<210> 156
<211> 49
<212> DNA
<213> Artificial sequence

<220>
<223> Primer

<400> 156
tatggaccct gaaaccacag ccacattctt gttatttata aaaagacac
49

<210> 157
<211> 49
<212> DNA
<213> Artificial sequence

<220>
<223> Primer

<400> 157
ctcccgtgtc tttttataaa taacaagaat gtggctgtgg tttcaggg
49

<210> 158
<211> 49
<212> DNA
<213> Artificial sequence

<220>
<223> Primer

<400> 158
taccgtaggc gtccttagga aagatagaag gccatgaagc tttttcttt
49

```


<210> 159
 <211> 49
 <212> DNA
 <213> Artificial sequence

 <220>
 <223> Primer

 <400> 159
 attggaaaga aaaagcttca tggccttcta tctttcctaa ggacgccta 49

 <210> 160
 <211> 21
 <212> DNA
 <213> Artificial sequence

 <220>
 <223> Primer

 <400> 160
 ttattgtttg gcatttgtag c 21

 <210> 161
 <211> 22
 <212> DNA
 <213> Artificial sequence

 <220>
 <223> Primer

 <400> 161
 ccaagcatct cataaaccta tg 22

 <210> 162
 <211> 22
 <212> DNA
 <213> Artificial sequence

 <220>
 <223> Primer

 <400> 162
 tgtgcagatg cagatgtgag ac 22

 <210> 163
 <211> 17
 <212> DNA
 <213> Artificial sequence

 <220>
 <223> Primer

 <400> 163
 agttattgat accgtac 17

 <210> 164
 <211> 19
 <212> DNA
 <213> Artificial sequence

 <220>
 <223> Primer

<400> 164
cgagataccg taggcgtcc

19

<210> 165
<211> 2237
<212> DNA
<213> Artificial sequence

<220>
<223> Fragment

<400> 165
ttatgtatgc tcttctgact tttcgtgtga tgaggctcgt ggaaaaaatg aataatttat 60
gaatttgaga acaattttgt gttgttacgg tattttacta tggaataatc aatcaattga 120
ggattttatg caaatatcgt ttgaatattt ttccgaccct ttgagtactt ttcttcataa 180
ttgcataata ttgtccgctg cccctttttc tgtagacgg tgtcttgatc tacttgctat 240
cgttcaacac caccttattt tctaactatt ttttttttag ctcatgtgaa tcagcttatg 300
gtgatggcac atttttgcat aaacctagct gtcctcgttg aacataggaa aaaaaatat 360
ataaacaagg ctctttcact ctccctgcaa tcagatttgg gtttgttccc tttattttca 420
tatttcttgt catattcctt tctcaattat tattttctac tcataacctc acgcaaaata 480
acacagtcaa atcaatcaaa atgactgaca aaaaaactct taaagactta agaaatcgta 540
gttctgttta cgattcaatg gttaaatcac ctaatcgtgc tatgttgctg gcaactggta 600
tgcaagatga agactttgaa aaacctatcg tcggtgtcat ttcaacttgg gctgaaaaca 660
caccttgtaa tatccactta catgactttg gtaaacctagc caaagtcggt gttaaggaag 720
ctggtgcttg gccagttcag ttcggaacaa tcacggtttc tgatggaatc gccatgggaa 780
cccaaggaat gcgtttctcc ttgacatctc gtgatattat tgcagattct attgaagcag 840
ccatgggagg tcataatgcg gatgcttttg tagccattgg cggttgtgat aaaaacatgc 900
ccggttctgt tatcgctatg gctaacatgg atatcccagc cttttttgct tacggcggaa 960
caattgcacc tggtaattta gacggcaaag atatcgattt agtctctgtc tttgaagggtg 1020
tcggccattg gaaccacggc gatatgacca aagaagaagt taaagctttg gaatgtaatg 1080
cttggtcccgg tcctggaggc tgcggtggta tgtatactgc taacacaatg gcgacagcta 1140
ttgaagtttt gggacttagc cttccgggtt catcttctca cccggctgaa tccgcagaaa 1200
agaaagcaga tattgaagaa gctggtcgcg ctgttgctaa aatgctcgaa atgggcttaa 1260
aaccttctga cattttaacg cgtgaagctt ttgaagatgc tattactgta actatggctc 1320
tgggaggttc aaccaactca acccttcacc tcttagctat tgcccatgct gctaattgtg 1380
aattgacact tgatgatttc aatactttcc aagaaaaagt tcctcatttg gctgatttga 1440
aaccttctgg tcaatatgta ttccaagacc tttacaaggc cggaggggta ccagcagtta 1500
tgaaatatct ccttaaaaat ggcttccttc atgggtgaccg tatcacttgt actggcaaaa 1560
cagtcgctga aaatttgaag gcttttgatg atttaacacc tgggtcaaaag gttattatgc 1620
cgcttgaaaa tcctaaacgt gaagatggtc cgctcattat tctccatggg aacttggctc 1680

20110907_CL5178USNA_SeqList.txt

cagacggtgc cgttgccaaa gtttctggtg taaaagtgcg tcgtcatgtc ggtcctgcta	1740
aggtctttaa ttctgaagaa gaagccattg aagctgtctt gaatgatgat attgttgatg	1800
gtgatgttgt tgtcgtacgt tttgtaggac caaagggcgg tcctggtatg cctgaaatgc	1860
tttccctttc atcaatgatt gttggtaaag ggcaaggatga aaaagttgcc cttctgacag	1920
atggccgctt ctcagggtgt acttatggtc ttgtcgtggg tcatatcgct cctgaagcac	1980
aagatggcgg tccaatcgcc tacctgcaaa caggagacat agtcactatt gaccaagaca	2040
ctaaggaatt acactttgat atctccgatg aagagttaaa acatcgtcaa gagaccattg	2100
aattgccacc gctctattca cgcggtatcc ttggtaaata tgctcacatc gtttcgtctg	2160
cttctagggg agccgtaaca gacttttgga agcctgaaga aactggcaaa aaatgagcga	2220
tttaatctct aattatt	2237

<210> 166
 <211> 24
 <212> DNA
 <213> Artificial sequence

<220>
 <223> Primer

<400> 166 ttatgtatgc tcttctgact tttc	24

<210> 167
 <211> 49
 <212> DNA
 <213> Artificial sequence

<220>
 <223> Primer

<400> 167 aataattaga gattaaatcg ctcatttttt gccagtttct tcaggcttc	49

<210> 168
 <211> 49
 <212> DNA
 <213> Artificial sequence

<220>
 <223> Primer

<400> 168 agcctgaaga aactggcaaa aaatgagcga tttaatctct aattattag	49

<210> 169
 <211> 49
 <212> DNA
 <213> Artificial sequence

<220>
 <223> Primer

<400> 169 tatggaccct gaaaccacag ccacattttt caatcattgg agcaatcat	49

```

<210> 170
<211> 49
<212> DNA
<213> Artificial sequence

<220>
<223> Primer

<400> 170
taaaatgatt gctccaatga ttgaaaaatg tggctgtggt ttcagggtc          49

<210> 171
<211> 49
<212> DNA
<213> Artificial sequence

<220>
<223> Primer

<400> 171
accgtaggtg ttgtttggga aagtggaagg ccatgaagct ttttctttc          49

<210> 172
<211> 49
<212> DNA
<213> Artificial sequence

<220>
<223> Primer

<400> 172
ttggaagaa aaagcttcat ggccttccac tttcccaaac aacacctac          49

<210> 173
<211> 23
<212> DNA
<213> Artificial sequence

<220>
<223> Primer

<400> 173
ttattgctta gcgttggtag cag          23

<210> 174
<211> 4420
<212> DNA
<213> Artificial sequence

<220>
<223> Cassette

<400> 174
ttatgtatgc tcttctgact tttcgtgtga tgaggctcgt ggaaaaaatg aataatttat          60
gaatttgaga acaattttgt gttgttacgg tattttacta tggaataatc aatcaattga          120
ggattttatg caaatatcgt ttgaatattt ttccgaccct ttgagtactt ttcttcataa          180
ttgcataata ttgtccgctg cccctttttc tgttagacgg tgtcttgatc tacttgctat          240
cgttcaacac caccttattt tctaactatt ttttttttag ctcatttgaa tcagcttatg          300

```

20110907_CL5178USNA_SeqList.txt

gtgatggcac	atTTTTgcat	aaacctagct	gtcctcgttg	aacataggaa	aaaaaaatat	360
ataaacaagg	ctcttttact	ctccttgcaa	tcagatttgg	gtttgttccc	tttatttttca	420
tatttcttgt	catattcctt	tctcaattat	tattttctac	tcataacctc	acgcaaaata	480
acacagtcaa	atcaatcaaa	atgactgaca	aaaaaactct	taaagactta	agaaatcgta	540
gttctgttta	cgattcaatg	gttaaatacac	ctaatacgtc	tatgtttgct	gcaactggta	600
tgcaagatga	agactttgaa	aaacctatcg	tcggtgtcat	ttcaacttgg	gctgaaaaca	660
caccttgtaa	tatccactta	catgactttg	gtaaactagc	caaagtcggt	gttaaggaag	720
ctggtgcttg	gccagttcag	ttcggaacaa	tcacggtttc	tgatggaatc	gccatgggaa	780
cccaaggaat	gcgttttctc	ttgacatctc	gtgatattat	tgacagattct	attgaagcag	840
ccatgggagg	tcataatgcg	gatgcttttg	tagccatttg	cggttgtgat	aaaaacatgc	900
ccggttctgt	tatcgctatg	gctaacatgg	atatcccagc	cattttttgct	tacggcggaa	960
caattgcacc	tggtaattta	gacggcaaag	atatcgattt	agtctctgtc	tttgaagggtg	1020
tcggccattg	gaaccacggc	gatatgacca	aagaagaagt	taaagctttg	gaatgtaatg	1080
cttgtcccgg	tcctggaggc	tgcggtggta	tgtatactgc	taacacaatg	gcgacagcta	1140
ttgaagtttt	gggacttagc	cttccgggtt	catcttctca	cccggctgaa	tccgcagaaa	1200
agaaagcaga	tattgaagaa	gctggtcgcg	ctgttgtcaa	aatgctcgaa	atgggcttaa	1260
aaccttctga	cattttaacg	cgtgaagctt	ttgaagatgc	tattactgta	actatggctc	1320
tgggagggtc	aaccaactca	acctttcacc	tcttagctat	tgcccatgct	gctaattgtg	1380
aattgacact	tgatgatttc	aatactttcc	aagaaaaagt	tcctcatttg	gctgatttga	1440
aaccttctgg	tcaatatgta	ttccaagacc	tttacaaggt	cggaggggta	ccagcagtta	1500
tgaaatatct	ccttaaaaaat	ggcttccttc	atggtgaccg	tatcacttgt	actggcaaaa	1560
cagtcgctga	aaatttgaag	gctttttagt	atttaacacc	tggtcaaaaag	gttattatgc	1620
cgcttgaaaa	tcctaaacgt	gaagatggtc	cgctcattat	tctccatggt	aacttggctc	1680
cagacgggtc	cgttgccaaa	gtttctgggt	taaaagtgcg	tcgtcatgtc	ggctctgcta	1740
aggtctttta	ttctgaagaa	gaagccattg	aagctgtctt	gaatgatgat	attgtttgatg	1800
gtgatgttgt	tgctgtacgt	ttttagtagc	caaagggcgg	tcctgggtatg	cctgaaatgc	1860
tttccctttc	atcaatgatt	gttggtaaaag	ggcaaggtga	aaaagttgcc	cttctgacag	1920
atggccgctt	ctcaggtggg	acttatggtc	ttgtcgtggg	tcatatcgct	cctgaagcac	1980
aagatggcgg	tccaatcgcc	tacctgcaaa	caggagacat	agtcactatt	gaccaagaca	2040
ctaaggaatt	acactttgat	atctccgatg	aagagttaaa	acatcgtcaa	gagaccattg	2100
aattgccacc	gctctattca	cgcggtatcc	ttggtaaata	tgctcacatc	gtttcgtctg	2160
cttctagggg	agccgtaaca	gacttttggg	agcctgaaga	aactggcaaa	aaatgagcga	2220
tttaatctct	aattattagt	taaagtttta	taagcatttt	tatgtaacga	aaaataaatt	2280
ggttcatatt	attactgcac	tgtcacttac	catggaaaga	ccagacaaga	agttgccgac	2340

20110907_CL5178USNA_SeqList.txt

agtctgttga attggcctgg ttaggcttaa gtctgggtcc gcttctttac aaatttggag	2400
aatttctctt aaacgatatg tatattcttt tcgttggaaa agatgtcttc caaaaaaaaa	2460
accgatgaat tagtgggaacc aaggaaaaaa aaagaggtat ccttgattaa ggaacactgt	2520
ttaaacagtg tggtttccaa aaccctgaaa ctgcattagt gtaatagaag actagacacc	2580
tcgatacaaa taatggttac tcaattcaaa actgccagcg aattcgactc tgcaattgct	2640
caagacaagc tagttgtcgt agatttctac gccacttggg gcggtccatg taaaatgatt	2700
gctccaatga ttgaaaaatg tggctgtggg ttcaggggcc ataaagcttt tcaattcatc	2760
tttttttttt ttgttctttt ttttgattcc ggtttctttg aaattttttt gattcggtaa	2820
tctccgagca gaaggaagaa cgaaggaagg agcacagact tagattggta tatatacgca	2880
tatgtggtgt tgaagaaaca tgaaattgcc cagtattctt aaccaactg cacagaacaa	2940
aaacctgcag gaaacgaaga taaatcatgt cgaaagctac atataaggaa cgtgctgcta	3000
ctcatcctag tcctgttgct gccaagctat ttaatatcat gcacgaaaag caaacaact	3060
tgtgtgcttc attggatgtt cgtaccacca aggaattact ggagttagtt gaagcattag	3120
gtcccaaaat ttgtttacta aaaacacatg tggatatctt gactgatttt tccatggagg	3180
gcacagttaa gccgctaaag gcattatccg ccaagtacaa ttttttactc ttcgaagaca	3240
gaaaatttgc tgacattggg aatacagtca aattgcagta ctctgcgggt gtatacagaa	3300
tagcagaatg ggcagacatt acgaatgcac acggtgtggg gggcccagggt attgttagcg	3360
gtttgaagca ggcggcggaa gaagtaacaa aggaacctag aggccttttg atgttagcag	3420
aattgtcatg caagggctcc ctagctactg gagaatatac taaggggtact gttgacattg	3480
cgaagagcga caaagatttt gttatcggct ttattgtctca aagagacatg ggtggaagag	3540
atgaaggtta cgattgggtg attatgacac ccggtgtggg tttagatgac aaggagacg	3600
cattgggtca acagtataga accgtggatg atgtggtctc tacaggatct gacattatta	3660
ttgttggaag aggactatth gcaaagggaa gggatgctaa ggtagagggt gaacgttaca	3720
gaaaagcagg ctgggaagca tatttgagaa gatgcggcca gcaaaactaa aaaactgtat	3780
tataagtaaa tgcattgata ctaaactcac aaattagagc ttcaatttaa ttatatcagt	3840
tattacccgg gaatctcggg cgtaattgatt tctataatga cgaaaaaaaa aaaattggaa	3900
agaaaaagct tcatggcctt ccactttccc aaacaacacc tacggtatct ctcaagtctt	3960
atgggggttcc attggtttca ccactgggtg taccttgggt gctgctttcg ctgctgaaga	4020
aattgatcca aagaagagag ttatcttatt cattggtgac gggtctttgc aattgactgt	4080
tcaagaaatc tccaccatga tcagatgggg cttgaagcca tacttgttcg tcttgaacaa	4140
cgatgggttac accattgaaa agttgattca cgggtccaaag gctcaatata acgaaattca	4200
agggtgggac cacctatcct tgttgccaac tttcgggtgct aaggactatg aaaccacag	4260
agtcgctacc accggtgaat gggacaagtt gaccaagac aagtctttca acgacaactc	4320
taagatcaga atgattgaaa tcatgttgcc agtcttcgat gctccacaaa acttggttga	4380

acaagctaag ttgactgctg ctaccaacgc taagcaataa 4420

<210> 175
 <211> 21
 <212> DNA
 <213> Artificial sequence

<220>
 <223> Primer

<400> 175
 tttttggtgg ttccggcttc c 21

<210> 176
 <211> 22
 <212> DNA
 <213> Artificial sequence

<220>
 <223> Primer

<400> 176
 aaagttggca tagcggaaac tt 22

<210> 177
 <211> 16
 <212> DNA
 <213> Artificial sequence

<220>
 <223> Primer

<400> 177
 gtcattgaca ccatct 16

<210> 178
 <211> 19
 <212> DNA
 <213> Artificial sequence

<220>
 <223> Primer

<400> 178
 agagataccg taggtgttg 19

<210> 179
 <211> 57
 <212> DNA
 <213> Artificial sequence

<220>
 <223> Primer

<400> 179
 aattggatcc ggcgcgccgt ttaaacggcc ggccaatgtg gctgtggttt cagggtc 57

<210> 180
 <211> 49
 <212> DNA
 <213> Artificial sequence

<220>

<223> Primer

<400> 180

aatttctaga ttaattaagc ggccgcaagg ccatgaagct ttttctttc

49

<210> 181

<211> 2686

<212> DNA

<213> Artificial sequence

<220>

<223> Plasmid

<400> 181

tcgcgcgttt cggatgatgac ggtgaaaacc tctgacacat gcagctcccg gagacgggtca 60

cagcttgtct gtaagcggat gccgggagca gacaagcccg tcagggcgcg tcagcgggtg 120

ttggcgggtg tcggggctgg cttaactatg cggcatcaga gcagattgta ctgagagtgc 180

accatatgcg gtgtgaaata ccgcacagat gcgtaaggag aaaataccgc atcaggcgcc 240

attcgccatt caggctgcgc aactgttggg aagggcgatc ggtgcggggc tcttcgctat 300

tacgccagct ggcgaaaggg ggatgtgctg caaggcgatt aagttgggta acgccagggc 360

tttcccagtc acgacgttgt aaaacgacgg ccagtgaatt cgagctcggc acccggggat 420

cctctagagt cgacctgcag gcatgcaagc ttggcgtaat catggtcata gctgtttcct 480

gtgtgaaatt gttatccgct cacaattcca cacaacatac gagccggaag cataaagtgt 540

aaagcctggg gtgcctaata agtgagctaa ctcacattaa ttgcgttgcg ctcactgccc 600

gctttccagt cgggaaacct gtcgtgccag ctgcattaat gaatcggcca acgcgcgggg 660

agaggcggtt tgcgtatttg gcgctcttcc gcttcctcgc tctactgactc gctgcgctcg 720

gtcgttcggc tgcggcgagc ggtatcagct cactcaaagg cggtaatacg gttatccaca 780

gaatcagggg ataacgcagg aaagaacatg tgagcaaaag gccagcaaaa ggccaggaac 840

cgtaaaaagg ccgcgttgct ggcgtttttc cataggctcc gccccctga cgagcatcac 900

aaaaatcgac gctcaagtca gaggtggcga aaccgcagag gactataaag ataccaggcg 960

tttccccctg gaagctccct cgtgcgctct cctgtttccga ccctgccgct taccggatac 1020

ctgtccgcct ttctcccttc gggaagcgtg gcgctttctc atagctcacg ctgtaggtat 1080

ctcagttcgg ttaggtcgt tcgctccaag ctgggctgtg tgcacgaacc ccccgttcag 1140

cccgaccgct gcgccttatc cggtaactat cgtcttgagt ccaaccggg aagacacgac 1200

ttatcgccac tggcagcagc cactggtaac aggattagca gagcgaggta ttaggcgggt 1260

gctacagagt tcttgaagt gtggcctaac tacggctaca ctagaaggac agtatttggc 1320

atctgcgctc tgctgaagcc agttaccttc ggaaaaagag ttggtagctc ttgatccggc 1380

aaacaaacca ccgctggtag cgggtggtttt tttgtttgca agcagcagat tacgcgcaga 1440

aaaaaaggat ctcaagaaga tcctttgatc ttttctacgg ggtctgacgc tcagtgaac 1500

gaaaactcac gttaagggtat tttggatcatg agattatcaa aaaggatctt cacctagatc 1560

20110907_CL5178USNA_SeqList.txt

cttttaaatt aaaaatgaag ttttaaata atctaaagta tatatgagta aacttggtct	1620
gacagttacc aatgcttaat cagtgaggca cctatctcag cgatctgtct atttcgttca	1680
tccatagttg cctgactccc cgtcgtgtag ataactacga tacgggaggg cttaccatct	1740
ggccccagtg ctgcaatgat accgcgagac ccacgctcac cggctccaga tttatcagca	1800
ataaaccagc cagccggaag ggccgagcgc agaagtggtc ctgcaacttt atccgcctcc	1860
atccagtcta ttaattgttg ccgggaagct agagtaagta gttcgccagt taatagtttg	1920
cgcaacgttg ttgccattgc tacaggcatc gtggtgtcac gctcgtcgtt tggatatggct	1980
tcattcagct ccggttccca acgatcaagg cgagttacat gatcccccatt gttgtgcaaa	2040
aaagcggtta gtccttcgg tcctccgatc gttgtcagaa gtaagttggc cgcagtgtta	2100
tcactcatgg ttatggcagc actgcataat tctcttactg tcatgccatc cgtaagatgc	2160
ttttctgtga ctggtgagta ctcaaccaag tcattctgag aatagtgtat gcggcgaccg	2220
agttgctctt gcccggcgtc aatacgggat aataccgcgc cacatagcag aactttaaaa	2280
gtgctcatca ttgaaaacg ttcttcgggg cgaaaactct caaggatctt accgctgttg	2340
agatccagtt cgatgtaacc cactcgtgca cccaactgat cttcagcatc ttttactttc	2400
accagcgttt ctgggtgagc aaaaacagga aggcaaatg ccgcaaaaaa gggaataagg	2460
gcgacacgga aatgttgaat actcatactc ttcctttttc aatattattg aagcatttat	2520
cagggttatt gtctcatgag cggatacata tttgaatgta tttagaaaaa taaacaaata	2580
ggggttccgc gcacatttcc ccgaaaagtg ccacctgacg tctaagaaac cattattatc	2640
atgacattaa cctataaaaa taggcgtatc acgaggccct ttcgtc	2686

<210> 182
 <211> 21
 <212> DNA
 <213> Artificial sequence

<220>
 <223> Primer

<400> 182	
tcggtgcggg cctcttcgct a	21

<210> 183
 <211> 21
 <212> DNA
 <213> Artificial sequence

<220>
 <223> Primer

<400> 183	
aatgtgagtt agtcactca t	21

<210> 184
 <211> 15456
 <212> DNA
 <213> Artificial sequence

20110907_CL5178USNA_SeqList.txt

<220>

<223> Template

<400> 184

aaagagtaat	actagagata	aacataaaaa	atgtagaggt	cgagtttaga	tgcaagttca	60
aggagcgaag	ggtggatggg	taggttatat	agggatatag	cacagagata	tatagcaaag	120
agatactttt	gagcaatggt	tgtggaagcg	gtattcgcaa	tatttttagta	gctcgttaca	180
gtccggtgcg	tttttggttt	tttgaaagtg	cgtcttcaga	gcgcttttgg	ttttcaaaag	240
cgctctgaag	ttcctatact	ttctagagaa	taggaacttc	ggaataggaa	cttcaaagcg	300
tttccgaaaa	cgagcgcttc	cgaaaatgca	acgcgagctg	cgcacatata	gctcactggt	360
cacgtcgcac	ctatatctgc	gtgttgccctg	tatatatata	tacatgagaa	gaacggcata	420
gtgctgtgtt	atgcttaaat	gcgtacttat	atgcgtctat	ttatgtagga	tgaaaggtag	480
tctagtacct	cctgtgatat	tatcccattc	catgcggggg	atcgtatgct	tccttcagca	540
ctacccttta	gctgttctat	atgctgccac	tcctcaattg	gattagtctc	atccttcaat	600
gctatcattt	cctttgatat	tggatcatac	taagaaacca	ttattatcat	gacattaacc	660
tataaaaaata	ggcgtatcac	gaggcccttt	cgtctcgcg	gtttcggtga	tgacggtgaa	720
aacctctgac	acatgcagct	cccggagacg	gtcacagctt	gtctgtaagc	ggatgccggg	780
agcagacaag	cccgtcaggg	cgcgtcagcg	ggtgttggcg	ggtgtcgggg	ctggcttaac	840
tatgcggcat	cagagcagat	tgtactgaga	gtgcaccata	aattcccgtt	ttaagagctt	900
ggtgagcgct	aggagtcact	gccaggatc	gtttgaacac	ggcattagtc	agggaagtca	960
taacacagtc	ctttcccga	attttctttt	tctattactc	ttggcctcct	ctagtacact	1020
ctatatTTTT	ttatgcctcg	gtaatgattt	tcattttttt	ttttccacct	agcggatgac	1080
tctttttttt	tcttagcgat	tggcattatc	acataatgaa	ttatacatta	tataaagtaa	1140
tgtgatttct	tcgaagaata	tactaaaaaa	tgagcaggca	agataaacga	aggcaaagat	1200
gacagagcag	aaagccctag	taaagcgtat	tacaaatgaa	accaagattc	agattgcgat	1260
ctcttttaaag	ggtggtcccc	tagcgataga	gcactcgatc	ttcccagaaa	aagaggcaga	1320
agcagtagca	gaacaggcca	cacaatcgca	agtgattaac	gtccacacag	gtatagggtt	1380
tctggaccat	atgatacatg	ctctggccaa	gcattccggc	tggtcgctaa	tcgttgagt	1440
cattggtgac	ttacacatag	acgaccatca	caccactgaa	gactgcggga	ttgctctcgg	1500
tcaagctttt	aaagaggccc	tagggggcgt	gcgtggagta	aaaaggtttg	gatcaggatt	1560
tgcgcttttg	gatgaggcac	tttcagagc	ggtggtagat	ctttcgaaca	ggccgtacgc	1620
agttgtcgaa	cttggtttgc	aaaggagaa	agtaggagat	ctctcttgcg	agatgatccc	1680
gcattttctt	gaaagctttg	cagaggctag	cagaattacc	ctccacgttg	attgtctg	1740
aggcaagaat	gatcatcacc	gtagtgagag	tgcgttcaag	gctcttg	ttgccataag	1800
agaagccacc	tcgccaatg	gtaccaacga	tgttccctcc	accaaagg	ttcttatgta	1860
gtgacaccga	ttatttaaag	ctgcagcata	cgatatatat	acatgtgtat	atatgtatac	1920

20110907_CL5178USNA_SeqList.txt

ctatgaatgt	cagtaagtat	gtatacgaac	agtatgatac	tgaagatgac	aaggtaatgc	1980
atcatttctat	acgtgtcatt	ctgaacgagg	cgcgctttcc	ttttttcttt	ttgctttttc	2040
tttttttttc	tcttgaactc	gacggatcta	tgcggtgtga	aataccgcac	agatgcgtaa	2100
ggagaaaata	ccgcatcagg	aaattgtaag	cgттаататт	ttgttaaaat	tcgcgttaaa	2160
tttttgttaa	atcagctcat	tttttaacca	ataggccgaa	atcggcacaaa	tcctttataa	2220
atcaaaaagaa	tagaccgaga	tagggttgag	tgttgttcca	gtttggaaca	agagtccact	2280
attaaagaac	gtggactcca	acgtcaaagg	gcgaaaaacc	gtctatcagg	gcgatggccc	2340
actacgtgaa	ccatcacctt	aatcaagttt	tttggggtcg	aggtgccgta	aagcactaaa	2400
tcggaaccct	aaagggagcc	cccgatttag	agcttgacgg	ggaaagccgg	cgaacgtggc	2460
gagaaaggaa	gggaagaaag	cgaaaggagc	gggcgctagg	gcgctggcaa	gtgtagcggt	2520
cacgctgcgc	gtaaccacca	caccgcgcgc	gcttaatgcg	ccgctacagg	gcgctccat	2580
tcgccattca	ggctgcgcaa	ctgttgggaa	gggcgcgggtg	cgggcctctt	cgctattacg	2640
ccagctggcg	aaagggggat	gtgctgcaag	gcgattaagt	tgggtaacgc	cagggttttc	2700
ccagtcacga	cgttgtaaaa	cgacggccag	tgagcgcgcg	taatacgact	cactataggg	2760
cgaattgggt	accgggcccc	ccctcgaggt	cgacggcgcg	ccactggtag	agagcgactt	2820
tgtatgcccc	aattgcgaaa	cccgcgatat	ccttctcgat	tcttttagtac	ccgaccagga	2880
caaggaaaag	gaggtcgaaa	cgtttttgaa	gaaacaagag	gaactacacg	gaagctctaa	2940
agatggcaac	cagccagaaa	ctaagaaaat	gaagttgatg	gatccaactg	gcaccgctgg	3000
cttgaacaac	aataccagcc	ttccaacttc	tgtaaataac	ggcggtagcg	cagtgccacc	3060
agtaccgtta	cctttcggtta	tacctccttt	ccccatgttt	ccaatgccct	tcatgcctcc	3120
aacggctact	atcacaatc	ctcatcaagc	tgacgcaagc	cctaagaaat	gaataacaat	3180
actgacagta	ctaaataatt	gcctacttgg	cttcacatac	gttgcatagc	tcgatataga	3240
taataatgat	aatgacagca	ggattatcgt	aatacgtaat	agctgaaaat	ctcaaaaatg	3300
tgtgggtcat	tacgtaaata	atgataggaa	tgggattctt	ctatttttcc	tttttccatt	3360
ctagcagccg	tcgggaaaac	gtggcatcct	ctctttcggg	ctcaattgga	gtcacgctgc	3420
cgtgagcatc	ctctctttcc	atatctaaca	actgagcacg	taaccaatgg	aaaagcatga	3480
gcttagcggt	gtcctcaaaa	agtattggat	ggttaatacc	atttgtctgt	tctcttctga	3540
ctttgactcc	tcaaaaaaaaa	aatctacaa	tcaacagatc	gcttcaatta	cgccctcaca	3600
aaaacttttt	tccttcttct	tcgcccacgt	taaattttat	ccctcatgtt	gtctaacgga	3660
tttctgcact	tgatttatta	taaaaagaca	aagacataat	acttctctat	caatttcagt	3720
tattgttctt	ccttgcggtta	ttcttctgtt	cttctttttc	ttttgtcata	tataaccata	3780
accaagtaat	acatattcaa	actagtatga	ctgacaaaaa	aactcttaaa	gacttaagaa	3840
atcgtagttc	tgtttacgat	tcaatgggtta	aatcacctaa	tcgtgctatg	ttgcgtgcaa	3900
ctgggtatgca	agatgaagac	tttgaaaaac	ctatcgctcg	tgtcatttca	acttgggctg	3960

20110907_CL5178USNA_SeqList.txt

aaaacacacc	ttgtaatatc	cacttacatg	acttttgtaa	actagccaaa	gtcgggtgta	4020
aggaagctgg	tgcttggcca	gttcagttcg	gaacaatcac	ggtttctgat	ggaatcgcca	4080
tgggaaccca	aggaatgcgt	ttctccttga	catctcgtga	tattattgca	gattctattg	4140
aagcagccat	gggaggtcat	aatgcggatg	cttttgtagc	cattggcggt	tgtgataaaa	4200
acatgcccgg	ttctgttatc	gctatggcta	acatggatat	cccagccatt	tttgcttacg	4260
gcggaacaat	tgcacctggt	aatttagacg	gcaaagatat	cgatttagtc	tctgtctttg	4320
aagggtgctg	ccattggaac	cacggcgata	tgaccaaaga	agaagttaaa	gctttggaat	4380
gtaatgcttg	ttccggctct	ggaggctg	gtggtatgta	tactgctaac	acaatggcga	4440
cagctattga	agttttggga	cttagccttc	cgggttcac	ttctcaccg	gctgaatccg	4500
cagaaaagaa	agcagatatt	gaagaagctg	gtcgcgctgt	tgtcaaaatg	ctcgaaatgg	4560
gcttaaaacc	ttctgacatt	ttaacgcgtg	aagcttttga	agatgctatt	actgtaacta	4620
tggctctggg	aggttcaacc	aactcaacc	ttcacctctt	agctattgcc	catgctgcta	4680
atgtggaatt	gacacttgat	gatttcaata	ctttccaaga	aaaagttcct	catttggtg	4740
atttgaaacc	ttctgggtcaa	tatgtattcc	aagaccttta	caaggctcga	ggggtaccag	4800
cagttatgaa	atatctcctt	aaaaatggct	tccttcatgg	tgaccgtatc	acttgtagt	4860
gcaaaacagt	cgctgaaaat	ttgaaggctt	ttgatgattt	aacacctggt	caaaaggtta	4920
ttatgccgct	tgaaaatcct	aaacgtgaag	atgggtccgct	cattattctc	catggtaact	4980
tggctccaga	cgggtgccgtt	gccaaagttt	ctgggtgtaaa	agtgcgtcgt	catgtcggtc	5040
ctgctaagg	ctttaattct	gaagaagaag	ccattgaagc	tgtcttgaat	gatgatattg	5100
ttgatgggtga	tggtgtgtgc	gtacgttttg	taggacccaa	gggcggctc	ggtatgcctg	5160
aaatgctttc	cctttcatca	atgattgttg	gtaaagggca	aggtgaaaaa	gttgcccttc	5220
tgacagatgg	ccgcttctca	ggtgggtactt	atgggtcttgt	cgtgggtcat	atcgctcctg	5280
aagcacaaga	tggcgggtcca	atcgccctacc	tgcaaacagg	agacatagtc	actattgacc	5340
aagacactaa	ggaattacac	tttgatatct	ccgatgaaga	gttaaaacat	cgtcaagaga	5400
ccattgaatt	gccaccgctc	tattcacg	gtatccttgg	taaatatgct	cacatcgttt	5460
cgtctgcttc	taggggagcc	gtaacagact	tttggaagcc	tgaagaaact	ggcaaaaaat	5520
gttgcctg	ttgctgtggt	taagcggccg	cgtaatttca	aattaatgga	tatagttttt	5580
taatgagtat	tgaatctggt	tagaaataat	ggaatattat	ttttatttat	ttatttatat	5640
tattggtcgg	ctcttttctt	ctgaagggtca	atgacaaaat	gatatgaagg	aaataatgat	5700
ttctaaaatt	ttacaacgta	agatattttt	acaaaagcct	agctcatctt	ttgtcatgca	5760
ctattttact	cacgcttgaa	attaacggcc	agtccactgc	ggagtcattt	caaagtcac	5820
ctaactcgatc	tatcgttttt	gatagctcat	tttgaggttc	gcgattgtct	tctgttattc	5880
acaactgttt	taatttttat	ttcattctgg	aactcttcga	gttctttgta	aagtctttca	5940
tagtagctta	ctttatcctc	caacatattt	aacttcatgt	caatttcggc	tcttaaat	6000

20110907_CL5178USNA_SeqList.txt

tccacatcat	caagttcaac	atcatctttt	aacttgaatt	tattctctag	ctcttccaac	6060
caagcctcat	tgctccttga	tttactggtg	aaaagtgata	cactttgcgc	gcaatccagg	6120
tcaaaacttt	cctgcaaaga	attcaccaat	ttctcgacat	catagtacaa	tttgttttgt	6180
tctcccatca	caatttaata	tacctgatgg	attcttatga	agcgctgggt	aatggacgtg	6240
tcactctact	tcgccttttt	ccctactcct	tttagtacgg	aagacaatgc	taataaataa	6300
gagggttaata	ataatattat	taatcgga	aaaagattaa	acgccaagcg	tttaattatc	6360
agaaagcaaa	cgtcgtacca	atccttgaat	gcttcccaat	tgtatattaa	gagtcatcac	6420
agcaacatat	tcttgttatt	aaattaatta	ttattgattt	ttgatattgt	ataaaaaaac	6480
caaatatgta	taaaaaaagt	gaataaaaaa	taccaagtat	ggagaaatat	attagaagtc	6540
tatacgtaa	accacccggg	ccccccctcg	aggctgcacg	tatcgataag	cttgatatcg	6600
aattcctgca	gcccggggga	tccactagtt	ctagagcggc	cgctctagaa	ctagtaccac	6660
agggtgtgtc	ctctgaggac	ataaaataca	caccgagatt	catcaactca	ttgctggagt	6720
tagcatatct	acaattgggt	gaaatgggga	gcgatttgca	ggcatttgct	cggcatgccg	6780
gtagagggtg	ggtcaataag	agcgacctca	tgctatacct	gagaaagcaa	cctgacctac	6840
aggaaagagt	tactcaagaa	taagaatttt	cgttttaaaa	cctaagagtc	actttaaaat	6900
ttgtatacac	ttatTTTTTT	tataacttat	ttaataataa	aatcataaa	tcataagaaa	6960
ttcgcttact	cttaattaat	caggcagcgc	ctgcgttcga	gaggatgatc	ttcatcgcct	7020
tctccttggc	gccattgagg	aatacctgat	aggcgtgctc	gatctcggcc	agctcgaagc	7080
gatgggtaat	catcttcttc	aacggaagct	tgctcggctga	ggcgaccttc	atcagcatgg	7140
gcgtcgtgtt	cgtgttcacc	agtcccgtgg	tgatcgtcag	gttcttgatc	cagagcttct	7200
gaatctcgaa	gtcaaccttg	acgccatgca	cgccgacggt	ggcgatgtgc	gcgccgggct	7260
tgacgatctc	ctggcagatg	tccaagtcg	ccggtatgcc	caccgcctcg	atcgcaacat	7320
cgactccctc	tgccgcaatc	ctatgcacgg	cttcgacaac	gttctccgtg	ccggagttga	7380
tggtgtgcgt	tgccccgagc	tccttggcga	gctggaggcg	attctcgtcc	atgtcgatca	7440
cgatgatggt	cgagggggag	tagaactggg	cgggtcaacag	tacggacatg	ccgacggggc	7500
ccgcgccgac	aatagccacc	gcatcgcccc	gctggacatt	cccatactgg	acgccgattt	7560
cgtggccggt	gggcaggatg	tcgctcagca	ggacggcgat	ttcgtcgtca	attgtctggg	7620
ggatcttgta	gaggctgttg	tcggcatgcg	ggatgcggac	gtattcggcc	tgcacgccat	7680
cgatcatgta	accaggatc	cacccgccgt	cgcggaatg	ggagtaaagc	tgcttcttgc	7740
agtagtcgca	cgagccgcaa	gaagtgcgc	aggaaatcag	gaccttgctg	cctttcttga	7800
actgcgtgac	actctcgccc	acttctcga	tgacgcctac	cccttcatgg	cccaggatgc	7860
gcccgtcggc	gacctctgga	ttcttgccct	tgtagatgcc	gagatccgtg	ccgcagatcg	7920
tggtcttcaa	aaccctgact	actacatccg	tgggcttttg	aagggtgggc	ttgggcttgt	7980
cttcaagcga	gatcttgtgg	tcaccgtgat	aaaccagagc	tttcatcctc	agctattgta	8040

20110907_CL5178USNA_SeqList.txt

atatgtgtgt	ttgtttggat	tattaagaag	aataattaca	aaaaaaatta	caaaggaagg	8100
taattacaac	agaattaaga	aaggacaaga	aggaggaaga	gaatcagttc	attattttctt	8160
ctttgttata	taacaaaccc	aagtagcgat	ttggccatac	attaaaagtt	gagaaccacc	8220
ctccctggca	acagccacaa	ctcgttacca	ttgttcatca	cgatcatgaa	actcgctgtc	8280
agctgaaatt	tcacctcagt	ggatctctct	ttttattctt	catcgttcca	ctaacctttt	8340
tccatcagct	ggcaggaac	ggaaagtgga	atcccattta	gcgagcttcc	tcttttcttc	8400
aagaaaagac	gaagcttgtg	tgtgggtgcg	cgcgctagta	tctttccaca	ttaagaaata	8460
taccataaag	gttacttaga	catcactatg	gctatatata	tatatatata	tatatatgta	8520
acttagcacc	atcgcgctg	catcactgca	tgtgttaacc	gaaaagtttg	gcgaacactt	8580
caccgacacg	gtcatttaga	tctgtcgtct	gcattgcacg	tcccttagcc	ttaaatccta	8640
ggcgggagca	ttctcgtgta	attgtgcagc	ctgcgtagca	actcaacata	gcgtagtcta	8700
cccagttttt	caagggttta	tcgttagaag	attctccctt	ttcttcctgc	tcacaaatct	8760
taaagtcata	cattgcacga	ctaaatgcaa	gcatgcggat	ccccgggct	gcaggaattc	8820
gatatcaagc	ttatcgatac	cgtcgactgg	ccattaatct	ttcccatatt	agatttcgcc	8880
aagccatgaa	agttcaagaa	aggtcttttag	acgaattacc	cttcatttct	caaactggcg	8940
tcaagggatc	ctggtatggt	tttatcgttt	tatttctggt	tcttatagca	tcgttttgga	9000
cttctctgtt	cccattaggc	ggttcaggag	ccagcgcaga	atcattcttt	gaaggatact	9060
tatcctttcc	aattttgatt	gtctgttacg	ttggacataa	actgtatact	agaaattgga	9120
ctttgatggt	gaaactagaa	gatatggatc	ttgataccgg	cagaaaacaa	gtagatttga	9180
ctcttcgtag	ggaagaaatg	aggattgagc	gagaaacatt	agcaaaaaga	tccttcgtaa	9240
caagattttt	acatttctgg	tgttgaaggg	aaagatatga	gctatacagc	ggaatttcca	9300
tatcactcag	attttgttat	ctaatttttt	ccttcccacg	tccgcgggaa	tctgtgtata	9360
ttactgcatc	tagatatatg	ttatcttatc	ttggcgcgta	catttaattt	tcaacgtatt	9420
ctataagaaa	ttgcgggagt	ttttttcatg	tagatgatac	tgactgcacg	caaatatagg	9480
catgatttat	aggcatgatt	tgatggctgt	accgatagga	acgctaagag	taacttcaga	9540
atcgttatcc	tggcggaaaa	aattcatttg	taaactttaa	aaaaaaaagc	caatatcccc	9600
aaaattatta	agagcgcctc	cattattaac	taaaatttca	ctcagcatcc	acaatgtatc	9660
aggtatctac	tacagatatt	acatgtggcg	aaaaagacaa	gaacaatgca	atagcgcatc	9720
aagaaaaaac	acaaagcttt	caatcaatga	atcgaaaatg	tcattaaaat	agtatataaa	9780
ttgaaactaa	gtcataaagc	tataaaaaga	aaattttatt	aatgcaaga	tttaaagtaa	9840
attcacggcc	ctgcaggcct	cagctcttgt	tttgttctgc	aaataactta	cccatctttt	9900
tcaaaacttt	aggtgcaccc	tcctttgcta	gaataagttc	tatccaatac	atcctatttg	9960
gatctgcttg	agcttctttc	atcacggata	cgaattcatt	ttctgttctc	acaatttttg	10020
acacaactct	gtcttccggt	gccccgaaac	tttctggcag	ttttgagtaa	ttccacatag	10080

20110907_CL5178USNA_SeqList.txt

gaatgtcatt	ataactctgg	ttcggaccat	gaatttcctt	ctcaaccgtg	taaccatcgt	10140
tattaatgat	aaagcagatt	gggtttatct	tctctcta	ggctagtcct	aattcttgga	10200
cagtcagttg	caatgatcca	tctccgataa	acaataaatg	tctagattct	ttatctgcaa	10260
tttggctgcc	tagagctgcg	gggaaagtgt	atcctataga	tccccacaag	ggttgaccaa	10320
taaaatgtga	tttcgatttc	agaaatatag	atgaggcacc	gaagaaagaa	gtgccttggt	10380
cagccacgat	cgtctcatta	ctttgggtca	aattttcgac	agcttgccac	agtctatctt	10440
gtgacaacag	cgcgttagaa	ggtacaaaat	cttcttgctt	tttatctatg	tacttgccctt	10500
tatattcaat	ttcggacaag	tcaagaagag	atgatatcag	ggattcgaag	tcgaaatctt	10560
ggattctttc	gttgaaaatt	ttaccttcat	cgatattcaa	ggaaatcatt	ttatcttcat	10620
taagatggtg	agtaaagtca	cccgtactag	aatcggttaag	ctttacaccc	aacataagaa	10680
taaaatcagc	agattccaca	aattccttca	agtttggtct	tgacagagta	ccgttgtaaa	10740
tccccaaaaa	tgagggcaat	gcttcatcaa	cagatgattt	accaaagttc	aaagtagtaa	10800
taggtaactt	agtctttgaa	ataaactgag	taacagtctt	ctctaggccg	aacgatataa	10860
tttcatggcc	tgtgattaca	attggtttct	tggcattctt	cagactttcc	tgtattttgt	10920
tcagaatctc	ttgatcagat	gtattcgacg	tggaaatctt	cttcttaaga	ggcaaggatg	10980
gtttttcagc	cttagcggca	gctacatcta	caggtaaatt	gatgtaaacc	ggctttcttt	11040
ccttttagtaa	ggcagacaac	actctatcaa	tttcaacagt	tgcattctcg	gctgtcaata	11100
aagtcctggc	agcagtaacc	ggttcgtgca	tcttcataaa	gtgcttgaaa	tcaccatcag	11160
ccaacgtatg	gtgaacaaac	ttaccttcgt	tctgcacttt	cgaggtagga	gatcccacga	11220
tctcaacaac	aggcagggtc	tcagcatagg	agcccgttaa	gccattaact	gcgataaatt	11280
cgccaacacc	aaatgtagtc	aagaatgccg	cagccttttt	cgttcttgcg	taccgcgcgg	11340
ccatatagga	ggcatttaac	tcattagcat	ttcccaccca	tttcatatct	ttgtgtgaaa	11400
taatttgatc	tagaaattgc	aaattgtagt	cacctggtac	tccgaatatt	tcttctatac	11460
ctaattcgtg	taatctgtcc	aacagatagt	cacctactgt	atacatcttg	tttactagtt	11520
tatgtgtgtt	tattcgaaac	taagttcttg	gtgttttaaa	actaaaaaaaa	agactaacta	11580
taaaagtaga	atttaagaag	tttaagaaat	agattttacag	aattacaatc	aatacctacc	11640
gtctttatat	acttattagt	caagtagggg	aataatttca	gggaactggt	ttcaaccttt	11700
tttttcagct	ttttccaaat	cagagagagc	agaaggtaat	agaagggtga	agaaaatgag	11760
atagatacat	gcgtgggtca	attgccttgt	gtcatcattt	actccaggca	ggttgcatca	11820
ctccattgag	gttgtgccc	ttttttgcct	gtttgtgccc	ctgttctctg	tagttgcgct	11880
aagagaatgg	acctatgaac	tgatgggttg	tgaagaaaac	aatatcttgg	tgctgggatt	11940
cttttttttt	ctggatgcca	gcttaaaaag	cgggctccat	tatatcttagt	ggatgccagg	12000
aataaaactgt	tcaccagac	acctacgatg	ttatatattc	tgtgtaaccc	gccccctatt	12060
ttgggcatgt	acgggttaca	gcagaattaa	aaggctaatt	ttttgactaa	ataaagttag	12120

20110907_CL5178USNA_SeqList.txt

gaaaatcact	actattaatt	atttacgtat	tctttgaaat	ggcagtattg	ataatgataa	12180
actcgaactg	aaaaagcgtg	ttttttattc	aaaatgattc	taactccctt	acgtaatcaa	12240
ggaatctttt	tgccttggcc	tccgcgtcat	taaacttctt	gttgttgacg	ctaacattca	12300
acgctagtat	atattcgttt	ttttcaggta	agttcttttc	aacgggtctt	actgatgagg	12360
cagtcgcgtc	tgaacctgtt	aagaggtcaa	atatgtcttc	ttgaccgtac	gtgtcttgca	12420
tgttattagc	tttggaatt	tgcatacaag	cataggaaaa	tttaaactct	ggctctcttg	12480
ggctcaagg	gacaaggctc	tcgaaaatag	ggcgcgcccc	accgcggtgg	agctccagct	12540
tttgttccct	ttagtgagg	ttaattgcgc	gcttggcgta	atcatggtca	tagctgtttc	12600
ctgtgtgaaa	ttgttatccg	ctcacaattc	cacacaacat	acgagccgga	agcataaagt	12660
gtaaagcctg	gggtgcctaa	tgagttagct	aactcacatt	aattgcgttg	cgctcactgc	12720
ccgctttcca	gtcgggaaac	ctgtcgtgcc	agctgcatta	atgaatcggc	caacgcgcgg	12780
ggagaggcgg	tttgcgtatt	gggcgctctt	ccgcttcctc	gctcactgac	tcgctgcgct	12840
cggctcgttc	gctgcggcga	gcggtatcag	ctcactcaaa	ggcggtaata	cggttatcca	12900
cagaatcagg	ggataacgca	ggaaagaaca	tgtgagcaaa	aggccagcaa	aaggccagga	12960
accgtaaaaa	ggccgcgttg	ctggcgtttt	tccataggct	ccgccccctc	gacgagcatc	13020
acaaaaatcg	acgtcaagt	cagaggtggc	gaaacccgac	aggactataa	agataccagg	13080
cgtttcccc	tggaagctcc	ctcgtgcgct	ctcctgttcc	gaccctgccg	cttaccggat	13140
acctgtccgc	ctttctccct	tcgggaagcg	tggcgctttc	tcatagctca	cgctgtaggt	13200
atctcagttc	ggtgtaggtc	gttcgctcca	agctgggctg	tgtgcacgaa	cccccgttc	13260
agcccgaccg	ctgcgcctta	tccggttaact	atcgtcttga	gtccaaccgc	gtaagacacg	13320
acttatcgcc	actggcagca	gccactggta	acaggattag	cagagcgagg	tatgtaggcg	13380
gtgctacaga	gttcttgaag	tgggtggccta	actacggcta	cactagaaga	acagtatttg	13440
gtatctgcgc	tctgctgaag	ccagttacct	tcggaaaaag	agttggtagc	tcttgatccg	13500
gcaaacaaac	caccgctggg	agcgggtggt	tttttgtttg	caagcagcag	attacgcgca	13560
gaaaaaaaag	atctcaagaa	gatcctttga	tcttttctac	ggggtctgac	gctcagtggg	13620
acgaaaactc	acgttaagg	attttggtca	tgagattatc	aaaaaggatc	ttcacctaga	13680
tccttttaaa	ttaaaaatga	agttttaaat	caatctaaag	tatatatgag	taaacttggt	13740
ctgacagtta	ccaatgctta	atcagttagg	cacctatctc	agcgatctgt	ctatttcggt	13800
catccatagt	tgcctgactc	cccgtcgtgt	agataactac	gatacgggag	ggcttaccat	13860
ctggccccag	tgctgcaatg	ataccgcgag	accacgctc	accggctcca	gatttatcag	13920
caataaacca	gccagccgga	agggccgagc	gcagaagtgg	tcctgcaact	ttatccgcct	13980
ccatccagtc	tattaattgt	tgccgggaag	ctagagtaag	tagttcgcca	gttaatatgt	14040
tgcgcaacgt	tgttgccatt	gctacaggca	tcgtggtgtc	acgctcgtcg	tttggtatgg	14100
cttcattcag	ctccggttcc	caacgatcaa	ggcgagttac	atgatcccc	atgttgtgca	14160

20110907_CL5178USNA_SeqList.txt

aaaaagcggg	tagctccttc	ggtcctccga	tcgttgctcag	aagtaagttg	gccgcagtg	14220
tatcactcat	ggttatggca	gcactgcata	attctcttac	tgtcatgcc	tccgtaagat	14280
gcttttctgt	gactgggtgag	tactcaacca	agtcattctg	agaatagtgt	atgcggcgac	14340
cgagttgctc	ttgcccggcg	tcaatacggg	ataataccgc	gccacatagc	agaactttta	14400
aagtgtcat	cattggaaaa	cgttcttcgg	ggcgaaaact	ctcaaggatc	ttaccgctgt	14460
tgagatccag	ttcgatgtaa	cccactcgtg	cacccaactg	atcttcagca	tcttttactt	14520
tcaccagcgt	ttctgggtga	gcaaaaacag	gaaggcaaaa	tgccgcaaaa	aagggaataa	14580
gggcgacacg	gaaatgttga	atactcatac	tcttcctttt	tcaatattat	tgaagcattt	14640
atcaggggta	ttgtctcatg	agcggataca	tatttgaatg	tatttagaaa	aataaacaaa	14700
taggggttcc	gcgcacattt	ccccgaaaag	tgccacctga	acgaagcatc	tgtgcttcat	14760
ttttagaac	aaaaatgcaa	cgcgagagcg	ctaatttttc	aaacaaagaa	tctgagctgc	14820
atttttacag	aacagaaatg	caacgcgaaa	gcgctatttt	accaacgaag	aatctgtgct	14880
tcatttttgt	aaaacaaaaa	tgcaacgcga	gagcgcta	ttttcaaaca	aagaatctga	14940
gctgcatttt	tacagaacag	aaatgcaacg	cgagagcgct	attttaccaa	caaagaatct	15000
atacttcttt	tttgttctac	aaaaatgcat	cccgagagcg	ctatttttct	aacaaagcat	15060
cttagattac	tttttttctc	ctttgtgctc	tctataatgc	agtctcttga	taactttttg	15120
cactgtaggt	ccgttaaggt	tagaagaagg	ctacttttgt	gtctattttc	tcttcataa	15180
aaaaagcctg	actccacttc	ccgcgtttac	tgattactag	cgaagctgcg	ggtgcatttt	15240
ttcaagataa	aggcatcccc	gattatattc	tataccgatg	tggattgctc	atactttgtg	15300
aacagaaagt	gatagcgttg	atgattcttc	attggtcaga	aaattatgaa	cggtttcttc	15360
tattttgtct	ctatatacta	cgtataggaa	atgtttacat	tttcgtattg	ttttcgattc	15420
actctatgaa	tagttcttac	tacaattttt	ttgtct			15456

<210> 185
 <211> 33
 <212> DNA
 <213> Artificial sequence

<220>
 <223> Primer

<400>	185	
aattggcgcg	ccatgaaagc	tctggtttat cac
		33

<210> 186
 <211> 49
 <212> DNA
 <213> Artificial sequence

<220>
 <223> Primer

<400>	186	
tgaatcatga	gttttatggt	aattagctca ggcagcgct gcgttcgag
		49

<210> 187
 <211> 49
 <212> DNA
 <213> Artificial sequence

 <220>
 <223> Primer

 <400> 187
 atcctctcga acgcaggcgc tgcctgagct aattaacata aaactcatg 49

 <210> 188
 <211> 34
 <212> DNA
 <213> Artificial sequence

 <220>
 <223> Primer

 <400> 188
 aattgttttaa acaagtaaataa aaattaatca gcat 34

 <210> 189
 <211> 49
 <212> DNA
 <213> Artificial sequence

 <220>
 <223> Primer

 <400> 189
 acacaatata ataacaagaa gaacaaaatg aaagctctgg tttatcacg 49

 <210> 190
 <211> 49
 <212> DNA
 <213> Artificial sequence

 <220>
 <223> Primer

 <400> 190
 agcgtatata tctgttggga aagtagaagg ccatgaagct ttttctttc 49

 <210> 191
 <211> 49
 <212> DNA
 <213> Artificial sequence

 <220>
 <223> Primer

 <400> 191
 ttggaaagaa aaagcttcat ggccttctac tttccaaca gatgtatac 49

 <210> 192
 <211> 22
 <212> DNA
 <213> Artificial sequence

 <220>

<223> Primer

<400> 192

ttattgttta gcgtagtag cg

22

<210> 193

<211> 3521

<212> DNA

<213> Artificial sequence

<220>

<223> Cassette

<400> 193

aaggaaataa agcaaataac aataacacca ttattttaat tttttttcta ttactgtcgc 60

taacacctgt atggttgcaa ccaggtgaga atccttctga tgcatacttt atgcgtttat 120

gcgttttgcg ccccttgaa aaaaattgat tctcatcgta aatgcatact acatgcgttt 180

atgggaaaag cctccatc caaaggtcgc gtttctttta gaaaaactaa tacgtaaacc 240

tgcattaagg taagattata tcagaaaatg tgttgcaaga aatgcattat gcaatTTTTT 300

gattatgaca atctctcgaa agaaatttca tatgatgaga cttgaataat gcagcggcgc 360

ttgctaaaag aacttgata taagagctgc cattctcgat caatatactg tagtaagtcc 420

tttctctctt ttcttattac acttatttca cataatcaat ctcaaagaga acaacacaat 480

acaataacaa gaagaacaaa atgaaagctc tggtttatca cggtgaccac aagatctcgc 540

ttgaagacaa gcccaagccc acccttcaaa agcccacgga tgtagtagta cgggttttga 600

agaccacgat ctgcggcacg gatctcggca tctacaaagg caagaatcca gaggtcgccg 660

acgggcgcat cctgggcat gaagggtag gcgtcatcga ggaagtgggc gagagtgtca 720

cgcagttcaa gaaaggcgac aaggtcctga tttcctgcgt cacttcttgc ggctcgtgcg 780

actactgcaa gaagcagctt tactccatt gccgcgacgg cgggtggatc ctgggttaca 840

tgatcgatgg cgtgcaggcc gaatacgtcc gcatcccga tgccgacaac agcctctaca 900

agatcccca gacaattgac gacgaaatcg ccgtcctgct gagcgacatc ctgcccaccg 960

gccacgaaat cggcgtccag tatgggaatg tccagccggg cgatgcggtg gctattgtcg 1020

gcgcgggccc cgtcggcatg tccgtactgt tgaccgccc gttctactcc ccctcgacca 1080

tcatcgtgat cgacatggac gagaatcgcc tccagctcgc caaggagctc ggggcaacgc 1140

acaccatcaa ctccggcacg gagaacgttg tcgaagccgt gcataggatt gcggcagagg 1200

gagtcgatgt tgcgatcgag gcggtgggca taccggcgac ttgggacatc tgccaggaga 1260

tcgtcaagcc cggcgcgcac atcgccaacg tcggcgtgca tggcgtcaag gttgacttcg 1320

agattcagaa gctctggatc aagaacctga cgatcaccac gggactggtg aacacgaaca 1380

cgacgcccac gctgatgaag gtcgcctcga ccgacaagct tccgttgaag aagatgatta 1440

cccatcgctt cgagctggcc gagatcgagc acgcctatca ggtattcctc aatggcgcca 1500

aggagaaggc gatgaagatc atcctctcga acgcaggcgc tgcctgagct aattaacata 1560

aaactcatga ttcaacgttt gtgtattttt ttacttttga aggttataga tgtttaggta 1620

20110907_CL5178USNA_SeqList.txt

aataattggc	atagatatag	ttttagtata	ataaatttct	gatttggttt	aaaatatcaa	1680
ctatTTTTTT	tcacatatgt	tcttgtaatt	acttttctgt	cctgtcttcc	aggTTaaaga	1740
ttagcttcta	atatttttagg	tggtttatta	tttaatttta	tgctgattaa	tttatttact	1800
tgTTtaaacg	gccggccaat	gtggctgtgg	tttcagggtc	cataaagctt	ttcaattcat	1860
ctTTTTTTTT	tttgttcttt	tttttgattc	cggtttcttt	gaaatttttt	tgattcggta	1920
atctccgagc	agaaggaaga	acgaaggaag	gagcacagac	ttagattggt	atatatacgc	1980
atatgtggtg	ttgaagaaac	atgaaattgc	ccagtattct	taacccaact	gcacagaaca	2040
aaaacctgca	ggaaacgaag	ataaatcatg	tcgaaagcta	catataagga	acgtgctgct	2100
actcatccta	gtcctgttgc	tgccaagcta	tttaatatca	tgcacgaaaa	gcaaacaaac	2160
ttgtgtgctt	cattggatgt	tcgtaccacc	aaggaattac	tggagttagt	tgaagcatta	2220
ggTcccaaaa	tttgtttact	aaaaacacat	gtggatatct	tgactgattt	ttccatggag	2280
ggcacagtta	agccgctaaa	ggcattatcc	gccaaagtaca	atTTTTtact	cttcgaagac	2340
agaaaatttg	ctgacattgg	taatacagtc	aaattgcagt	actctgcggg	tgtatacaga	2400
atagcagaat	gggcagacat	tacgaatgca	cacggtgtgg	tgggcccgagg	tattgttagc	2460
ggTTtgaagc	aggcggcgga	agaagtaaca	aaggaaccta	gaggcctttt	gatgttagca	2520
gaattgtcat	gcaagggctc	cctagctact	ggagaatata	ctaaggggtac	tgTTgacatt	2580
gcgaagagcg	acaaagattt	tgTTatcggc	tttattgctc	aaagagacat	gggtggaaga	2640
gatgaaggTT	acgattggTT	gattatgaca	cccggTgtgg	gTTtagatga	caagggagac	2700
gcattgggtc	aacagtatag	aaccgtggat	gatgtggtct	ctacaggatc	tgacattatt	2760
attgttgga	gaggactatt	tgcaaaggga	agggatgcta	aggtagaggg	tgaacgttac	2820
agaaaagcag	gctgggaagc	atatttgaga	agatgcggcc	agcaaaaacta	aaaaactgta	2880
ttataagtaa	atgcatgtat	actaaactca	caaattagag	cttcaattta	attatatcag	2940
ttattaccCG	ggaatctcgg	tcgtaatgat	ttctataatg	acgaaaaaaaa	aaaaattgga	3000
aagaaaaagc	ttcatggcct	tctactttcc	caacagatgt	atacgctatc	gtccaagtct	3060
tgtgggggtc	cattggtttc	acagtcggcg	ctctattggg	tgctactatg	gccgctgaag	3120
aacttgatcc	aaagaagaga	gttattttat	tcattggtga	cggttctcta	caattgactg	3180
ttcaagaaat	ctctaccatg	attagatggg	gtttgaagcc	atacattttt	gtcttgaata	3240
acaacggtta	caccattgaa	aaattgattc	acggtcctca	tgccgaatat	aatgaaattc	3300
aaggTTggga	ccacttggcc	ttattgccaa	cttttggtgc	tagaaactac	gaaaccacaca	3360
gagttgctac	cactggtgaa	tgggaaaagt	tgactcaaga	caaggacttc	caagacaact	3420
ctaagattag	aatgattgaa	gttatgttgc	cagtctttga	tgctccacaa	aacttggtta	3480
aacaagctca	attgactgcc	gctactaacg	ctaaacaata	a		3521

<210> 194

<211> 72

<212> DNA
 <213> Artificial sequence

 <220>
 <223> Primer

 <400> 194
 cataatcaat ctcaaagaga acaacacaat acaataacaa gaagaacaaa atgaaagctc 60
 tggtttatca cg 72

 <210> 195
 <211> 21
 <212> DNA
 <213> Artificial sequence

 <220>
 <223> Primer

 <400> 195
 taggcataat caccgaagaa g 21

 <210> 196
 <211> 21
 <212> DNA
 <213> Artificial sequence

 <220>
 <223> Primer

 <400> 196
 aaaatggtaa gcagctgaaa g 21

 <210> 197
 <211> 17
 <212> DNA
 <213> Artificial sequence

 <220>
 <223> Primer

 <400> 197
 agttgtaga actggtg 17

 <210> 198
 <211> 22
 <212> DNA
 <213> Artificial sequence

 <220>
 <223> Primer

 <400> 198
 cttagcctct agccatagcc at 22

 <210> 199
 <211> 1685
 <212> DNA
 <213> Artificial sequence

 <220>
 <223> Cassette

20110907_CL5178USNA_SeqList.txt

```

<400> 199
gtattttggg agattcaatt ctctttccct ttccttttcc ttcgctcccc ttccttatca      60
gcattgcgga ttacgtattc taatgttcag ataacttcgt atagcataca ttatacgaag      120
ttatgcagat tgtactgaga gtgcaccata ccacagcttt tcaattcaat tcatcatttt      180
ttttttattc ttttttttga tttcggtttc tttgaaatth ttttgattcg gtaatctccg      240
aacagaagga agaacgaagg aaggagcaca gacttagatt ggtatatata cgcataatgta      300
gtgttgaaga aacatgaaat tgcccagtat tcttaaccca actgcacaga aaaaaaacct      360
gcaggaaacg aagataaatc atgtcgaaag ctacatataa ggaacgtgct gctactcatc      420
ctagtccctgt tgctgccaa gctatttaata tcatgcacga aaagcaaaca aacttggtgtg      480
cttcattgga tgttcgtacc accaaggaat tactggagtt agttgaagca ttaggtccca      540
aaatttgtht actaaaaaca catgtggata tcttgactga tttttccatg gagggcacag      600
ttaagccgct aaaggcatta tccgccaa gtaatttttt actcttcgaa gacagaaaat      660
ttgctgacat tggtaataca gtcaaattgc agtactctgc ggggtgtatac agaatagcag      720
aatgggcaga cattacgaat gcacacggtg tgggtgggcc aggtattgth agcggthtga      780
agcaggcggc agaagaagta acaaaggaac ctagaggcct tttgatgtta gcagaattgt      840
catgcaaggg ctccctatct actggagaat atactaaggg tactgttgac attgcgaaga      900
gcgacaaaga ttttgthtct ggctttattg ctcaaagaga catgggtgga agagatgaag      960
gttacgattg gttgattatg acaccggtg tgggtthtga tgacaaggga gacgcattgg     1020
gtcaacagta tagaacgtg gatgatgtgg tctctacagg atctgacatt attattgthg     1080
gaagaggact atthgcaaag ggaagggatg ctaaggtaga ggggtgaacgt tacagaaaag     1140
caggctggga agcatatttg agaagatgag gccagcaaaa ctaaaaaact gtattataag     1200
taaathgatg tataactaac tcacaaatta gagctthcat ttaattatat cagthattac     1260
cctatgcggt gtgaaatacc gcacagatgc gtaaggagaa aataccgcat caggaaattg     1320
taaagcttaa taththtga aaathcgctg taaaththth ttaaatcagc tcatththth     1380
accaataggc cgaaatcggc aaaatccctt ataaatcaaa agaatagacc gagatagggt     1440
tgagtgtthg tccagththg aacaagagtc cactattaaa gaacgtggac tccaacgtca     1500
aagggcgaaa aaccgtctat cagggcgatg gccactacg tgaacctca ccctaatcaa     1560
gataacttcg tatagcatc attatacgaa gthtaccagt gatgatacaa cgagthtagcc     1620
aaggtgacac tctccccccc cctccccctc tgatctthtc tgttgctctt tthtccccca     1680
accaa                                             1685

```

```

<210> 200
<211> 1559
<212> DNA
<213> Artificial sequence

```

```

<220>
<223> Template

```

20110907_CL5178USNA_SeqList.txt

```

<400> 200
gcattgcgga ttacgtattc taatgttcag taccgttcgt ataatgtatg ctatacgaag      60
ttatgcagat tgtactgaga gtgcaccata ccaccttttc aattcatcat ttttttttta      120
ttcttttttt tgatttcggt ttccttgaaa tttttttgat tcggtaatct ccgaacagaa      180
ggaagaacga aggaaggagc acagacttag attggtatat atacgcatat gtagtgttga      240
agaaacatga aattgcccag tattcttaac ccaactgcac agaacaaaaa cctgcaggaa      300
acgaagataa atcatgtcga aagctacata taaggaacgt gctgctactc atcctagtcc      360
tgttgctgcc aagctattta atatcatgca cgaaaagcaa acaaacttgt gtgcttcatt      420
ggatgttcgt accaccaagg aattactgga gttagttgaa gcattagggtc ccaaaatttg      480
tttactaaaa acacatgtgg atatcttgac tgatttttcc atggaggggca cagttaagcc      540
gctaaaggca ttatccgcca agtacaattt tttactcttc gaagacagaa aatttgctga      600
cattggtaat acagtcaaat tgcagtactc tgcgggtgta tacagaatag cagaatgggc      660
agacattacg aatgcacacg gtgtggtggg cccaggtatt gttagcgggt tgaagcaggc      720
ggcagaagaa gtaacaaagg aacctagagg ctttttgatg ttagcagaat tgtcatgcaa      780
gggctcccta tctactggag aatatactaa gggtagtgtt gacattgcga agagcgacaa      840
agattttgtt atcggccttta ttgctcaaag agacatgggt ggaagagatg aaggttacga      900
ttggttgatt atgacacccg gtgtgggttt agatgacaag ggagacgcat tgggtcaaca      960
gtatagaacc gtggatgatg tggctctctac aggatctgac attattattg ttggaagagg     1020
actatttgca aagggaaggg atgctaaggt agagggtgaa cgttacagaa aagcaggctg     1080
ggaagcatat ttgagaagat gcggccagca aaactaaaaa actgtattat aagtaaatgc     1140
atgtatacta aactacaaaa ttagagcttc aatttaatta tatcagttat taccctatgc     1200
ggtgtgaaat accgcacaga tgcgtaagga gaaaataccg catcaggaaa ttgtaaacgt     1260
taatattttg ttaaaattcg cgttaaattt ttgttaaadc agctcatttt ttaaccaata     1320
ggccgaaatc ggcaaaatcc cttataaatc aaaagaatag accgagatag ggttgagtgt     1380
tgttccagtt tggaacaaga gtccactatt aaagaacgtg gactccaacg tcaaagggcg     1440
aaaaaccgtc tatcagggcg atggcccact acgtgaacca tcaccctaata caagataact     1500
tcgtataatg tatgctatac gaacggtacc agtgatgata caacgagtta gccaaagggt     1559

```

```

<210> 201
<211> 90
<212> DNA
<213> Artificial sequence

```

```

<220>
<223> Primer

```

```

<400> 201
gtattttggg agattcaatt ctctttccct ttccttttcc ttcgctcccc ttccttatca      60
gcattgcgga ttacgtattc taatgttcag                                     90

```

```

<210> 202
<211> 90
<212> DNA
<213> Artificial sequence

<220>
<223> Primer

<400> 202
ttggttgggg gaaaaagagg caacaggaaa gatcagaggg ggaggggggg ggagagtgtc      60
accttggtta actcgttgta tcatcactgg      90

<210> 203
<211> 23
<212> DNA
<213> Artificial sequence

<220>
<223> Primer

<400> 203
ttagttttgc tggccgcatc ttc      23

<210> 204
<211> 7523
<212> DNA
<213> Artificial sequence

<220>
<223> Plasmid

<400> 204
ccagcttttg ttccctttag tgagggttaa ttgcgcgctt ggcgtaatca tggtcatagc      60
tgtttcctgt gtgaaattgt tatccgctca caattccaca caacatagga gccggaagca      120
taaagtgtaa agcctggggg gcctaattgag tgaggtaact cacattaatt gcgttgcgct      180
cactgcccgc tttccagtcg ggaaacctgt cgtgccagct gcattaatga atcggccaac      240
gcgcggggag aggcggtttg cgtattgggc gctcttccgc ttcctcgctc actgactcgc      300
tgcgctcggt cgttcggctg cggcgagcgg tatcagctca ctcaaaggcg gtaatacggg      360
tatccacaga atcaggggat aacgcaggaa agaacatgtg agcaaaaggc cagcaaaagg      420
ccaggaaccg taaaaaggcc gcgttgctgg cgtttttcca taggctccgc cccctgacg      480
agcatcacia aaatcgacgc tcaagtcaga ggtggcgaaa cccgacagga ctataaagat      540
accaggcggt tccccctgga agtccctcg tgcgctctcc tgttccgacc ctgccgctta      600
ccggatacct gtccgccttt ctcccttcgg gaagcgtggc gctttctcat agctcacgct      660
gtaggtatct cagttcgggt taggtcggtt gctccaagct gggctgtgtg cacgaacccc      720
ccgttcagcc cgaccgctgc gccttatccg gtaactatcg tcttgagtcc aaccggtaaa      780
gacacgactt atcgccactg gcagcagcca ctggtaacag gattagcaga gcgaggtatg      840
taggcgggtg tacagagttc ttgaagtggg ggcctaacta cggctacact agaaggacag      900
tatttggtat ctgcgctctg ctgaagccag ttaccttcgg aaaaagagtt ggtagctctt      960
gatccggcaa acaaaccacc gctggtagcg gtggtttttt tgtttgcaag cagcagatta      1020

```


20110907_CL5178USNA_SeqList.txt

cgcgagaaa	aaaaggatct	caagaagatc	ctttgatctt	ttctacgggg	tctgacgctc	1080
agtggaaacga	aaactcacgt	taagggatctt	tggtcatgag	attatcaaaa	aggatcttca	1140
cctagatcct	tttaaattaa	aaatgaagtt	ttaaataaat	ctaaagtata	tatgagtaaa	1200
cttgggtctga	cagttaccaa	tgcttaata	gtgaggcacc	tatctcagcg	atctgtctat	1260
ttcggttcac	catagttgcc	tgactccccg	tcgtgtagat	aactacgata	cgggagggct	1320
taccatctgg	ccccagtgt	gcaatgatac	cgcgagaccc	acgctcaccg	gctccagatt	1380
tatcagcaat	aaaccagcca	gccggaaggg	ccgagcgag	aagtggctct	gcaactttat	1440
ccgcctccat	ccagtctatt	aattgttgcc	gggaagctag	agtaagtagt	tcgccagtta	1500
atagtttgcg	caacgttggt	gccattgcta	caggcatcgt	gggtgtcacgc	tcgtcgtttg	1560
gtatggcttc	attcagctcc	ggttcccaac	gatcaaggcg	agttacatga	tcccccatgt	1620
tgtgcaaaaa	agcggtagc	tccttcggtc	ctccgatcgt	tgtcagaagt	aagttggccg	1680
cagtgttatc	actcatgggt	atggcagcac	tgcataattc	tcttactgtc	atgccatccg	1740
taagatgctt	ttctgtgact	ggtgagtact	caaccaagtc	attctgagaa	tagtgtatgc	1800
ggcgaccgag	ttgtctttgc	ccggcgtaaa	tacgggataa	taccgcgcca	catagcagaa	1860
ctttaaaggt	gctcatcatt	ggaaaacgtt	cttcggggcg	aaaactctca	aggatcttac	1920
cgctgttgag	atccagttcg	atgtaacca	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	gggtccgcgc	acatttcccc	gaaaagtgcc	acctgaacga	agcatctgtg	2220
cttcattttg	tagaacaaaa	atgcaacgcg	agagcgctaa	tttttcaaac	aaagaatctg	2280
agctgcattt	ttacagaaca	gaaatgcaac	gcgaaagcg	tattttacca	acgaagaatc	2340
tgtgcttcat	ttttgtaaaa	caaaaatgca	acgcgagagc	gctaattttt	caaacaaaga	2400
atctgagctg	cattttttaca	gaacagaaat	gcaacgcgag	agcgctatct	taccaacaaa	2460
gaatctatac	ttcttttttg	ttctacaaaa	atgcatcccc	agagcgctat	ttttctaaca	2520
aagcatctta	gattactttt	tttctccttt	gtgcgctcta	taatgcagtc	tcttgataac	2580
tttttgact	gtaggtccgt	taaggttaga	agaaggctac	tttggtgtct	atcttctctt	2640
ccataaaaaa	agcctgactc	cacttcccgc	gtttactgat	tactagcgaa	gctgcgggtg	2700
cattttttca	agataaaggc	atccccgatt	atattctata	ccgatgtgga	ttgcgcatac	2760
tttgtgaaca	gaaagtgata	gcgttgatga	ttcttcattg	gtcagaaaaat	tatgaacggt	2820
ttcttctatt	ttgtctctat	atactacgta	taggaaatgt	ttacattttc	gtattgtttt	2880
cgattcactc	tatgaatagt	tcttactaca	atttttttgt	ctaaagagta	atactagaga	2940
taaacataaa	aaatgtagag	gtcgagttaa	gatgcaagtt	caaggagcga	aaggtggatg	3000
ggtaggttat	atagggatat	agcacagaga	tatatagcaa	agagatactt	ttgagcaatg	3060

20110907_CL5178USNA_SeqList.txt

tttgtggaag	cggtattcgc	aatatTTTTag	tagctcgtta	cagtccggtg	cgTTTTtggT	3120
TTTTtgaaag	tgcgtcttca	gagcgctttt	ggttttcaaa	agcgctctga	agttcctata	3180
ctttctagag	aataggaact	tcggaatagg	aacttcaaag	cgtttccgaa	aacgagcgct	3240
tccgaaaatg	caacgcgagc	tgcgcacata	cagctcactg	ttcacgtcgc	acctatatct	3300
gcgtgttgcc	tgtatatata	tatacatgag	aagaacggca	tagtgcggtg	ttatgcttaa	3360
atgcgtactt	atatgcgtct	atttatgtag	gatgaaaggt	agtctagtac	ctcctgtgat	3420
attatcccat	tccatgcggg	gtatcgtatg	cttccttcag	cactaccctt	tagctgtttct	3480
atatgctgcc	actcctcaat	tggattagtc	tcatccttca	atgctatcat	ttcctttgat	3540
attggatcat	ctaagaaacc	attattatca	tgacattaac	ctataaaaaat	aggcgatatca	3600
cgaggccctt	tcgtctcgcg	cgtttcggtg	atgacggtga	aaacctctga	cacatgcagc	3660
tcccggagac	ggcacagct	tgtctgtaag	cggatgccgg	gagcagacaa	gcccgtcagg	3720
gcgcgtcagc	gggtgttgcc	gggtgtcggg	gctggcttaa	ctatgcggca	tcagagcaga	3780
ttgtactgag	agtgacccat	aaattcccgt	tttaagagct	tggtagagcg	taggagtcac	3840
tgccaggtat	cgtttgaaca	cggcattagt	caggaagtc	ataacacagt	cctttcccg	3900
aattttcttt	ttctattact	cttggcctcc	tctagtacac	tctatatattt	tttatgcctc	3960
ggtaatgatt	ttcatttttt	tttttcccct	agcggatgac	tctttttttt	tcttagcgat	4020
tggcattatc	acataatgaa	ttatacatta	tataaagtaa	tgtgatttct	tcgaagaata	4080
tactaaaaaa	tgagcaggca	agataaacga	aggcaaagat	gacagagcag	aaagccctag	4140
taaagcgtat	tacaaatgaa	accaagattc	agattgcgat	ctcttttaaag	ggtggtcccc	4200
tagcgataga	gcactcgatc	ttcccagaaa	aagaggcaga	agcagtagca	gaacaggcca	4260
cacaatcgca	agtgattaac	gtccacacag	gtataggggt	tctggaccat	atgatacatg	4320
ctctggccaa	gcattccggc	tggtcgctaa	tcgttgagtg	cattggtgac	ttacacatag	4380
acgaccatca	caccactgaa	gactgcggga	ttgctctcgg	tcaagctttt	aaagaggccc	4440
tactggcgcg	tggagtaaaa	aggtttggat	caggatttgc	gcctttggat	gaggcacttt	4500
ccagagcggg	ggtagatctt	tcgaacaggc	cgtacgcagt	tgtcgaactt	ggtttgcaaa	4560
gggagaaagt	aggagatctc	tcttgcgaga	tgatcccgca	ttttcttgaa	agctttgcag	4620
aggctagcag	aattaccctc	cacgttgatt	gtctgcgagg	caagaatgat	catcaccgta	4680
gtgagagtgc	gttcaaggct	cttgcgggtt	ccataagaga	agccacctcg	cccaatggta	4740
ccaacgatgt	tccctccacc	aaagggtgtt	ttatgtagtg	acaccgatta	tttaaagctg	4800
cagcatacga	tatatataca	tgtgtatata	tgtataccta	tgaatgtcag	taagtatgta	4860
tacgaacagt	atgatactga	agatgacaag	gtaatgcata	attctatacg	tgtcattctg	4920
aacgaggcgc	gctttccttt	tttctttttg	ctttttcttt	ttttttctct	tgaactcgac	4980
ggatctatgc	ggtgtgaaat	accgcacaga	tcgtaagga	gaaaataccg	catcaggaaa	5040
ttgtaaacgt	taatatTTTT	ttaaaattcg	cgttaaattt	ttgttaaata	agctcatttt	5100

20110907_CL5178USNA_SeqList.txt

ttaaccaata	ggccgaaatc	ggcaaaatcc	cttataaatc	aaaagaatag	accgagatag	5160
ggttgagtgt	tgttccagtt	tggaacaaga	gtccactatt	aaagaacgtg	gactccaacg	5220
tcaaagggcg	aaaaaccgtc	tatcagggcg	atggcccact	acgtgaacca	tcaccctaata	5280
caagtttttt	ggggtcgagg	tgccgtaaag	cactaaatcg	gaaccctaaa	gggagccccc	5340
gatttagagc	ttgacgggga	aagccggcga	acgtggcgag	aaaggaaggg	aagaaagcga	5400
aaggagcggg	cgctagggcg	ctggcaagtg	tagcgggtcac	gctgcgcgta	accaccacac	5460
ccgccgcgct	taatgcgccg	ctacagggcg	cgtcgcgccca	ttcgccattc	aggctgcgca	5520
actgttggga	agggcgatcg	gtgcgggcct	cttcgctatt	acgccagctg	gcgaaagggg	5580
gatgtgctgc	aaggcgatta	agttgggtaa	cgccaggggt	ttcccagtca	cgacgttgta	5640
aaacgacggc	cagtgcgcgc	gcgtaatacg	actcactata	gggcgaattg	ggtaccgggc	5700
ccccctcga	ggtattagaa	gccgccgagc	gggcgacagc	cctccgacgg	aagactctcc	5760
tccgtgcgtc	ctcgtcttca	ccggtcgcgt	tcctgaaacg	cagatgtgcc	tcgcgccgca	5820
ctgctccgaa	caataaagat	tctacaatac	tagcttttat	ggttatgaag	aggaaaaatt	5880
ggcagtaacc	tggccccaca	aaccttcaaa	ttaacgaatc	aaattaacaa	ccataggatg	5940
ataatgcgat	tagtttttta	gccttatttc	tggggtaatt	aatcagcgaa	gcgatgattt	6000
ttgatctatt	aacagatata	taaatggaaa	agctgcataa	ccactttaac	taatactttc	6060
aacattttca	gtttgtatta	cttcttattc	aaatgtcata	aaagtatcaa	caaaaaattg	6120
ttaatatacc	tctatacttt	aacgtcaagg	agaaaaatgt	ccaatttact	gcccgtagac	6180
caaaatttgc	ctgcattacc	ggtcgaatgc	acgagtgatg	aggttcgcaa	gaacctgatg	6240
gacatgttca	gggatcgcca	ggcgttttct	gagcatacct	ggaaaatgct	tctgtccggt	6300
tgccggtcgt	gggcggcatg	gtgcaagttg	aataaccgga	aatgggtttcc	cgcagaacct	6360
gaagatgttc	gcgattatct	tctatatctt	caggcgcgcg	gtctggcagt	aaaaactatc	6420
cagcaacatt	tgggccagct	aaacatgctt	catcgtcggt	ccgggctgcc	acgaccaagt	6480
gacagcaatg	ctgtttcact	ggttatgcgg	cggatccgaa	aagaaaacgt	tgatgccggt	6540
gaacgtgcaa	aacaggctct	agcgttcgaa	cgcactgatt	tcgaccaggt	tcgttctactc	6600
atggaaaata	gcgatcgctg	ccaggatata	cgtaatctgg	catttctggg	gattgcttat	6660
aacaccctgt	tacgtatagc	cgaaattgcc	aggatcaggg	ttaaagatat	ctcacgtact	6720
gacggtggga	gaatgttaat	ccatattggc	agaacgaaaa	cgctgggttag	caccgcaggt	6780
gtagagaagg	cacttagcct	gggggtaact	aaactggtcg	agcgatggat	ttccgtctct	6840
ggtgtagctg	atgatccgaa	taactacctg	ttttgccggg	tcagaaaaaa	tgggtgtgcc	6900
gcgccatctg	ccaccagcca	gctatcaact	cgcgccctgg	aagggttttt	tgaagcaact	6960
catcgattga	tttacggcgc	taaggatgac	tctggtcaga	gataacctggc	ctggtctgga	7020
cacagtgcc	gtgtcggagc	cgcgcgagat	atggcccgcg	ctggagtttc	aataccggag	7080
atcatgcaag	ctgggtggctg	gaccaatgta	aatattgtca	tgaactatat	ccgtaacctg	7140

20110907_CL5178USNA_SeqList.txt

gatagtga	aaa	cagggg	caat	ggtg	cg	cctg	ctgga	agatg	g	cgatt	agga	gtaag	cg	aat	7200
ttcttat	gat	ttatga	tttt	tattatta	aaa	taagtt	tataa	aaaaa	ataag	tgtata	caaaa	7260			
ttttaaa	agt	actctt	aggt	tttaaaa	acga	aaattc	ttat	tcttg	agtaa	ctcttt	cctg	7320			
taggtc	aggt	tgcttt	tca	ggtat	agcat	gaggtc	gctc	ttattg	acca	cacct	ctacc	7380			
ggcatg	ccga	gcaaat	gcct	gcaaat	cgct	ccccatt	tca	ccaatt	gta	gatat	gctaa	7440			
ctccag	caat	gagttg	atga	atctcg	gtgt	gtatttt	atg	tcctc	agagg	acaac	acctg	7500			
tgggcc	gcca	ccgcgg	tgga	gct								7523			

<210> 205
 <211> 21
 <212> DNA
 <213> Artificial sequence

<220>
 <223> Primer

<400>	205	
cccat	taata	tactattgag a
		21

<210> 206
 <211> 8366
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> pRS316-UAS(PGK1)-PFBA1-GUS

<400>	206																
tcg	cg	gttt	cggt	gat	gac	ggt	gaaa	acc	tct	gac	acat	gcag	ctccc	gag	acg	gtca	60
cag	ctt	gtct	gta	agc	ggat	gcc	ggg	agca	gaca	agccc	tcag	ggc	gcg	tcag	cgg	gtg	120
ttg	gcg	gggtg	tcg	ggg	ctgg	ctta	act	atg	cgg	cat	caga	gcag	att	gta	ctg	agagt	180
acc	acg	cttt	tca	att	caat	tcat	catt	ttt	tttt	tattc	tttt	tttt	tga	tttc	cggt	tttc	240
ttt	gaa	at	ttt	gatt	cg	gta	at	ctccg	aac	aga	agga	aga	acga	agg	aagg	agcaca	300
gact	tag	att	ggt	atat	tata	cg	cat	atgta	gtgt	tga	aga	aacat	gaa	at	tgccc	agtat	360
tct	taac	cca	act	gc	acaga	acaaa	ac	ct	gcag	gaa	acg	aag	ata	aatc	atgt	cgaa	420
ctac	atata	aa	gac	gtg	ct	gtact	catc	ctag	tc	cctgt	tgct	gcca	ag	ctatt	taata		480
tcat	gc	acga	aa	agcaa	aca	aact	tgt	gtg	ctt	catt	gga	tgtt	cgt	acc	acca	agga	540
tact	gg	agtt	ag	ca	ttag	gtccc	aa	attt	gttt	act	aaaa	aca	cat	gtg	gata		600
tct	tga	ctga	tttt	tcc	atg	gaggg	cac	ag	tta	agcc	gct	aa	agg	catta	tcc	gcca	660
aca	atttt	ttt	act	cttc	gaa	gac	agaaa	at	ttg	ctg	acat	tg	gta	ataca	gtc	aaatt	720
agt	act	ctg	ggg	tgt	atac	aga	atag	cag	aat	ggg	caga	catt	ac	gaat	gcac	acgg	780
tg	gtggg	ccc	ag	gtatt	gtt	ag	cggtt	tga	ag	cagg	cggc	aga	aga	agta	aca	aagga	840
ctag	agg	cct	ttt	gat	gtta	gc	aga	attgt	cat	gca	agg	ctcc	ctat	ct	act	ggaga	900
ata	cta	aggg	tact	gtt	gac	att	gc	gaaga	gcg	acaa	aga	tttt	gtt	atc	gg	cttt	960

20110907_CL5178USNA_SeqList.txt

ctcaaagaga	catgggtgga	agagatgaag	gttacgattg	gttgattatg	acacccggtg	1020
tgggtttaga	tgacaagga	gacgcattgg	gtcaacagta	tagaaccgtg	gatgatgtgg	1080
tctctacagg	atctgacatt	attattgttg	gaagaggact	atttgcaaag	ggaagggatg	1140
ctaaggtaga	gggtgaacgt	tacagaaaag	caggctggga	agcatatttg	agaagatgcg	1200
gccagcaaaa	ctaaaaaact	gtattataag	taaatgcatg	tatactaaac	tcacaaatta	1260
gagcttcaat	ttaattatat	cagttattac	cctgcggtgt	gaaataccgc	acagatgcgt	1320
aaggagaaaa	taccgcatca	ggaaattgta	aacgttaata	ttttgttaaa	attcgcgtta	1380
aatttttgtt	aatcagctc	attttttaac	caataggccg	aaatcggcaa	aatcccttat	1440
aatcaaaaag	aatagaccga	gatagggttg	agtgttgttc	cagtttgga	caagagtcca	1500
ctattaaaga	acgtggactc	caacgtcaaa	gggcgaaaaa	ccgtctatca	gggcgatggc	1560
ccactacgtg	aaccatcacc	ctaatacagt	tttttggggt	cgagggtgccg	taaagcacta	1620
aatcggaacc	ctaaaggag	ccccgattt	agagcttgac	ggggaaagcc	ggcgaacgtg	1680
gcgagaaagg	aagggaagaa	agcgaaagga	gcgggcgcta	gggcgctggc	aagtgtagcg	1740
gtcacgctgc	gcgtaaccac	cacacccgcc	gcgcttaatg	cgccgctaca	gggcgcgtcg	1800
cgccattcgc	cattcaggct	gcgcaactgt	tgggaagggc	gatcgggtcg	ggcctcttcg	1860
ctattacgcc	agctggcgaa	ggggggatgt	gctgcaaggc	gattaagttg	ggtaacgcca	1920
gggttttccc	agtcacgacg	ttgtaaaacg	acggccagtg	aattgtaata	cgactcacta	1980
tagggcgaat	tggagctcca	ccgcggtggt	ttaacgtata	gacttctaata	atatttctcc	2040
atacttggtg	ttttttattc	acttttttta	tacatatttg	gtttttttat	acaatatcaa	2100
aaatcaataa	taattaattt	aataacaaga	atatgttgct	gtgatgactc	ttaatataca	2160
attggaagc	attcaaggat	tggtagcg	tttgctttct	gataattaaa	cgcttggcgt	2220
ttaatctttt	ttgccgatta	ataatattat	tattaccctc	ttatttatta	gcattgtctt	2280
ccgtactaaa	aggagtaggg	aaaaaggcga	agtagagtga	cacgtccatt	accagcgct	2340
tcataagaat	ccatcaggta	tattaaattg	tgatgggaga	acaaaacaaa	ttgtactatg	2400
atgtcgagaa	attggtgaat	tctttgcagg	aaagttttga	cctggattgc	gcgcaaagtg	2460
tatcactttt	caccagtaaa	tcaaggagca	atgaggcttg	gttggaagag	ctagagaata	2520
aattcaagtt	aaaagatgat	gttgaacttg	atgatgtgga	aaatttaaga	gccgaaattg	2580
acatgaagtt	aaatatgttg	gaggataaag	taagctacta	tgaaagactt	tacaaagaac	2640
tcgaagagtt	ccagaatgaa	ataaaaatta	aaacagttgt	gaataacaga	agacaatcgc	2700
gaactccaaa	atgagctatc	aaaaacgata	gatcgattag	gatgactttg	aaatgactcc	2760
gcagtggact	ggccgttaat	ttcaagcgtg	agtaaaatag	tgcatgacaa	aagatgagct	2820
aggcttttgt	aaaaatatct	tacgttgtaa	aatttttagaa	atcattattt	ccttcatatc	2880
attttgtcat	tgaccttcag	aagaaaagag	ccgaccaata	atataaataa	ataaataaaa	2940
ataatattcc	attatttcta	aacagattca	atactcatta	aaaaactata	tcaattaatt	3000

20110907_CL5178USNA_SeqList.txt

tgaattaacg	cggccgctct	agttaattaa	tcattgtttg	cctccctgct	gcggtttttc	3060
accgaagttc	atgccagtcc	agcgtttttg	cagcagaaaa	gccgccgact	tcggtttgcg	3120
gtcgcgagtg	aagatcccctt	tcttgttacc	gccaacgcgc	aatatgcctt	gcgaggtcgc	3180
aaaatcggcg	aaattccata	cctgttcacc	gacgacggcg	ctgacgcgat	caaagacgcg	3240
gtgatacata	tccagccatg	cacactgata	ctcttcactc	cacatgtcgg	tgtacattga	3300
gtgcagcccc	gctaacgtat	ccacgccgta	ttcggtgatg	ataatcggct	gatgcagttt	3360
ctcctgccag	gccagaagtt	ctttttccag	tacctttctt	gccgtttcca	aatcggccgt	3420
ttggacatac	catccgtaat	aacggttcag	gcacagcaca	tcaaagagat	cgctgatggt	3480
atcgggtgtg	gcgtcgcaga	acattacatt	gacgcaggtg	atcggacgcg	tcgggtcgcg	3540
tttacgcgtt	gcttccgcca	gtggcgcgaa	atattcccgt	gcaccttgcg	gacgggtatc	3600
cggttcgttg	gcaatactcc	acatcaccac	gcttggggtg	tttttgtcac	gcgctatcag	3660
ctctttaatc	gcctgtaagt	gcgcttgctg	agtttccccg	ttgactgcct	cttcgctgta	3720
cagttctttc	ggcttggtgc	ccgcttcgaa	accaatgcct	aaagagaggt	taaagccgac	3780
agcagcagtt	tcatcaatca	ccacgatgcc	atgttcatct	gcccagtcga	gcattctttc	3840
agcgttaagg	taatgcgagg	tacggtagga	gttggcccca	atccagtcca	ttaatgcgtg	3900
gtcgtgcacc	atcagcacgt	tatcgaatcc	tttgccacgc	aagtccgcat	cttcatgacg	3960
accaaagcca	gtaaagtaga	acggtttgtg	gttaatcagg	aactgttcgc	ccttcactgc	4020
cactgaccgg	atgccgacgc	gaagcgggta	gatatcacac	tctgtctggc	ttttggctgt	4080
gacgcacagt	tcatagagat	aaccttcacc	cggttgccag	aggtgcggat	tcaccacttg	4140
caaagtcccc	ctagtgcctt	gtccagttgc	aaccacctgt	tgatccgcat	cacgcagttc	4200
aacgctgaca	tcaccattgg	ccaccacctg	ccagtcaaca	gacgcgtggt	tacagtcttg	4260
cgcgacatgc	gtcaccacgg	tgatatcgtc	caccacaggtg	ttcggcgtgg	tgtagagcat	4320
tacgtctgca	tggattccgg	catagttaaa	gaaatcatgg	aagtaagact	gctttttctt	4380
gccgttttcg	tcggtaatca	ccattcccgg	cgggatagtc	tgccagttca	gttcgttggt	4440
cacacaaacg	gtgatacgta	cacttttccc	ggcaataaca	tacggcgtga	catcggttcc	4500
aaatggagta	tagccgccct	gatgtcccat	cacttcctga	ttattgaccc	acactttgcc	4560
gtaatgagtg	accgcacgca	aacgcagcac	gatacgttgg	cctgcccaac	ctttcggtat	4620
aaagacttcg	cgctgatacc	agacgttgcc	cgcataatta	cgaatatctg	catcggcgaa	4680
ctgatcgtta	aaactgcctg	gcacagcaat	tgcccggctt	tcttgtaacg	cgctttccca	4740
ccaacgctga	tcaattccac	agttttcgcg	atccagactg	aatgcccaac	ggccgtcgag	4800
ttttttgatt	tcacgggttg	gggtttctac	aggacgtacc	atactagttt	gaatatgtat	4860
tacttggtta	tggttatata	tgacaaaaga	aaaagaagaa	cagaagaata	acgcaaggaa	4920
gaacaataac	tgaaattgat	agagaagtat	tatgtctttg	tctttttata	ataaatcaag	4980
tgcagaaatc	cgttagacaa	catgagggat	aaaatttaac	gtgggcgaag	aagaaggaaa	5040

20110907_CL5178USNA_SeqList.txt

aaagtttttg	tgagggcgta	attgaagcga	tctgttgatt	gtagattttt	tttttttgag	5100
gagtcaaagt	cagaagagaa	cagacaaatg	gtattaacca	tccaataactt	ttttggagca	5160
acgctaagct	catgcttttc	cattggttac	gtgctcagtt	gtagatatg	gaaagagagg	5220
atgctcacgg	cagcgtgact	ccaattgagc	ccgaaagaga	ggatgccacg	ttttcccgac	5280
ggctgctaga	atggaaaaag	gaaaaataga	agaatcccat	tcctatcatt	atttacgtaa	5340
tgaccacac	atttttgaga	ttttcaacta	ttacgtatta	cgataatcct	gctgtcatta	5400
tcattattat	ctatatcgac	gtatgcaacg	tatgtgaagc	caagtaggcg	tcgacaggac	5460
cttgttgtgt	gacgaaattg	gaagctgcaa	tcaataggaa	gacaggaagt	cgagcgtgtc	5520
tgggtttttt	cagttttgtt	ctttttgcaa	acaaatcacg	agcgacggta	attggtaccc	5580
agcttttgtt	ccctttagtg	agggttaatt	ccgagcttgg	cgtaatcatg	gtcatagctg	5640
tttctgtgt	gaaattgtta	tccgctcaca	attccacaca	acataggagc	cggaagcata	5700
aagtgtaaag	cctgggggtgc	ctaattgagtg	aggttaactca	cattaattgc	gttgcgctca	5760
ctgcccgtt	tccagtcggg	aaacctgtcg	tgccagctgc	attaatgaat	cggccaacgc	5820
gcggggagag	gcggtttgcg	tattgggcgc	tcttcgctt	cctcgctcac	tgactcgctg	5880
cgctcggctg	ttcggctgcg	gcgagcggta	tcagctcact	caaaggcggg	aatacggtta	5940
tccacagaat	caggggataa	cgcaggaaaag	aacatgtgag	caaaaggcca	gcaaaaggcc	6000
aggaaccgta	aaaaggccgc	gttgctggcg	tttttccata	ggctcggccc	ccctgacgag	6060
catcacaaaa	atcgacgctc	aagtcagagg	tggcgaaacc	cgacaggact	ataaagatac	6120
caggcgttcc	cccctggaag	ctccctcgtg	cgctctcctg	ttccgaccct	gccgcttacc	6180
ggataacctgt	ccgcctttct	cccttcggga	agcgtggcgc	tttctcaatg	ctcacgctgt	6240
aggtatctca	gttcggtgta	ggtcgttcgc	tccaagctgg	gctgtgtgca	cgaaccccc	6300
gttcagcccg	accgctgcg	cttatccggg	aactatcgtc	ttgagtccaa	cccggtaaga	6360
cacgacttat	cgccactggc	agcagccact	ggtaacagga	ttagcagagc	gaggtatgta	6420
ggcggtgcta	cagagtctt	gaagtgggtg	cctaactacg	gtacactag	aaggacagta	6480
tttggtatct	gcgctctgct	gaagccagtt	accttcggaa	aaagagttgg	tagctcttga	6540
tccggcaaac	aaaccaccgc	tggtagcggg	ggtttttttg	tttgcaagca	gcagattacg	6600
cgcagaaaaa	aaggatctca	agaagatcct	ttgatctttt	ctacgggggtc	tgacgctcag	6660
tggaacgaaa	actcacgtta	agggattttg	gtcatgagat	tatcaaaaag	gatcttcacc	6720
tagatccttt	taaattaaaa	atgaagtttt	aaatcaatct	aaagtatata	tgagtaaact	6780
tggctgaca	gttaccaatg	cttaatcagt	gaggcaccta	tctcagcgat	ctgtctattt	6840
cgttcatcca	tagttgcctg	actgcccgtc	gtgtagataa	ctacgatacg	ggagggctta	6900
ccatctggcc	ccagtgtgct	aatgataccg	cgagaccac	gctcaccggc	tccagattta	6960
tcagcaataa	accagccagc	cggaagggcc	gagcgcagaa	gtggtcctgc	aactttatcc	7020
gcctccatcc	agtctattaa	ttgttgccgg	gaagctagag	taagtagttc	gccagttaat	7080

20110907_CL5178USNA_SeqList.txt

agtttgcgca acgttgttgc cattgctaca ggcacgtgg tgtcacgctc gtcgtttggt	7140
atgggttcat tcagctccgg ttcccaacga tcaaggcgag ttacatgatc ccccatgttg	7200
tgaaaaaaag cggttagctc cttcggtcct ccgatcgttg tcagaagtaa gttggccgca	7260
gtgttatcac tcatggttat ggcagcactg cataattctc ttactgtcat gccatccgta	7320
agatgctttt ctgtgactgg tgagtactca accaagtcac tctgagaata gtgtatgcgg	7380
cgaccgagtt gctcttgccc ggcgtcaata cgggataata ccgcgccaca tagcagaact	7440
ttaaaagtgc tcatcattgg aaaacgttct tcggggcgaa aactctcaag gatcttaccg	7500
ctgttgagat ccagttcgat gtaaccact cgtgcaccca actgatcttc agcatctttt	7560
actttcacca gcgtttctgg gtgagcaaaa acaggaaggc aaaatgccgc aaaaaagga	7620
ataagggcga cacggaaatg ttgaatactc atactcttcc tttttcaata ttattgaagc	7680
atttatcagg gttattgtct catgagcgga tacatatgtg aatgtattta gaaaaataaa	7740
caaatagggg ttccgcgcac atttccccga aaagtgccac ctgggtcctt ttcacacgt	7800
gctataaaaa taattataat ttaaattttt taatataaat atataaatta aaaatagaaa	7860
gtaaaaaaag aaattaaaga aaaaatagtt tttgttttcc gaagatgtaa aagactctag	7920
ggggatcgcc aacaaatact accttttctc ttgctcttcc tgctctcagg tattaatgcc	7980
gaattgtttc atcttgtctg tgtagaagac cacacacgaa aatcctgtga ttttacattt	8040
tacttatcgt taatcgaatg tatactctatt taatctgctt ttcttgtcta ataaatatat	8100
atgtaaagta cgctttttgt tgaaattttt taaacctttg tttatttttt tttcttcatt	8160
ccgtaactct tctaccttct ttatttactt tctaaaatcc aaatacaaaa cataaaaata	8220
aataaacaca gagtaaattc ccaaattatt ccatcattaa aagatacgag gcgcgtgtaa	8280
gttacaggca agcgatccgt cctaagaaac cattattatc atgacattaa cctataaaaa	8340
taggcgtatc acgaggccct ttcgtc	8366

<210> 207
 <211> 62
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> N1176 Primer

<400> 207	
gcatagcaat ctaatctaag ttccagctga ggatgacaac agattactca tcaccagcat	60
at	62

<210> 208
 <211> 62
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> N1177 Primer

20110907_CL5178USNA_SeqList.txt

<400> 208
atcaacacac aaacactaaa tcaaagctga ggatggattt atttgagtca ttagcacaaa 60
aa 62

<210> 209
<211> 44
<212> DNA
<213> Artificial Sequence

<220>
<223> N822 Primer

<400> 209
cgcctcagct ttgatttagt gtttgtgtgt tgataagcag ttgc 44

<210> 210
<211> 63
<212> DNA
<213> Artificial Sequence

<220>
<223> N1178 Primer

<400> 210
ggatcgcata agcttgatat cgaattcctg cgcccggggc actagtcaga tgccgcgggc 60
act 63

<210> 211
<211> 10934
<212> DNA
<213> Artificial Sequence

<220>
<223> pRS423::TEF(M4)-xpk1+EN01-eutD plasmid

<400> 211
ggatggagctc cagcttttgt tcccttttagt gaggggttaat tgcgcgcttg gcgtaatcat 60
ggcatagct gtttcctgtg tgaaattgtt atccgctcac aattccacac aacataggag 120
ccggaagcat aaagtgtaaa gcctgggggtg cctaattgagt gaggtaactc acattaattg 180
cgttgcgctc actgcccgtt ttccagtcgg gaaacctgtc gtgccagctg cattaatgaa 240
tcggccaacg cgcggggaga ggcgggttgc gtattgggag ctcttccgct tcctcgtcga 300
ctgactcgtc gcgctcggtc gttcggctgc ggcgagcggg atcagctcac tcaaaggcgg 360
taatacgggt atccacagaa tcaggggata acgcaggaaa gaacatgtga gcaaaaggcc 420
agcaaaaggc caggaaccgt aaaaaggccg cgttgctggc gtttttccat aggctccgcc 480
cccctgacga gcatcacaaa aatcgacgct caagtcagag gtggcgaaac ccgacaggac 540
tataaagata ccaggcgttt ccccttgga gctccctcgt gcgctctcct gttccgacct 600
tgccgcttac cggatacctg tccgcctttc tcccttcggg aagcgtggcg ctttctcata 660
gctcacgctg taggtatctc agttcgggtg aggtcggttc ctccaagctg ggctgtgtgc 720
acgaaccccc cgttcagccc gaccgctgcg ctttatccgg taactatcgt cttgagtcca 780
acccggtaag acacgactta tcgccactgg cagcagccac tggtaacagg attagcagag 840

20110907_CL5178USNA_SeqList.txt

cgaggatatgt	aggcggtgct	acagagttct	tgaagtgggtg	gcctaactac	ggctacacta	900
gaaggacagt	atttgggtatc	tgcgctctgc	tgaagccagt	taccttcgga	aaaagagttg	960
gtagctcttg	atccggcaaa	caaaccaccg	ctggtagcgg	tggttttttt	gtttgcaagc	1020
agcagattac	gcgagaaaa	aaaggatctc	aagaagatcc	tttgatcttt	tctacggggt	1080
ctgacgtca	gtggaacgaa	aactcacgtt	aagggatttt	ggtcacgaga	ttatcaaaaa	1140
ggatcttcac	ctagatcctt	ttaaattaaa	aatgaagttt	taaatcaatc	taaagtatat	1200
atgagtaaac	ttggctcgac	agttaccaat	gcttaatcag	tgaggcacct	atctcagcga	1260
tctgtctatt	tcgttcaccc	atagttgcct	gactccccgt	cgtgtagata	actacgatac	1320
gggagggctt	accatctggc	cccagtgctg	caatgatacc	gcgagacca	cgctcaccgg	1380
ctccagattt	atcagcaata	aaccagccag	ccggaagggc	cgagcgagca	agtggtcctg	1440
caactttatc	cgctccatc	cagtctatta	attgttgccg	ggaagctaga	gtaagtagtt	1500
cgccagttaa	tagtttgccg	aacgttggtt	ccattgctac	aggcatcgtg	gtgtcacgct	1560
cgctgtttgg	tatggcttca	ttcagctccg	gttcccaacg	atcaaggcga	gttacatgat	1620
cccccatgtt	gtgcaaaaa	gcggttagct	ccttcgggtc	tccgatcgtt	gtcagaagta	1680
agttggccgc	agtgttatca	ctcatggtta	tggcagcact	gcataattct	cttactgtca	1740
tgccatccgt	aagatgcttt	tctgtgactg	gtgagtactc	aaccaagtca	ttctgagaat	1800
agtgtatgcg	gcgaccgagt	tgctcttgcc	cggcgtcaat	acgggataat	accgcgccac	1860
atagcagaac	tttaaaagtg	ctcatcattg	gaaaacgttc	ttcggggcga	aaactctcaa	1920
ggatcttacc	gctgttgaga	tccagttcga	tgtaaccac	tcgtgcaccc	aactgatctt	1980
cagcatcttt	tactttcacc	agcgtttctg	ggtgagcaaa	aacaggaagg	caaatgccg	2040
caaaaaaggg	aataagggcg	acacggaaat	gttgaatact	catactcttc	ctttttcaat	2100
attattgaag	catttatcag	ggttattgtc	tcatgagcgg	atacatat	gaatgtat	2160
agaaaaataa	acaaataggg	gttccgcgca	catttccccg	aaaagtgcc	cctgaacgaa	2220
gcatctgtgc	ttcattttgt	agaacaaaa	tgcaacgcga	gagcgcta	ttttcaaaca	2280
aagaatctga	gctgcatttt	tacagaacag	aaatgcaacg	cgaaagcgct	attttacc	2340
cgaagaatct	gtgcttcatt	tttgtaaaac	aaaaatgcaa	cgcgagagcg	ctaatttttc	2400
aaacaaagaa	tctgagctgc	atttttacag	aacagaaatg	caacgcgaga	gcgctat	2460
accaacaaag	aatctatact	tcttttttgt	tctacaaaaa	tgcatcccga	gagcgctatt	2520
tttctaacia	agcatcttag	attacttttt	ttctcctttg	tgcgctctat	aatgcagtct	2580
cttgataact	ttttgactg	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
atgaacgggt	tcttctat	tgtctctata	tactacgtat	aggaaatgtt	tacattttcg	2880

20110907_CL5178USNA_SeqList.txt

tattgttttc	gattcactct	atgaatagtt	cttactacaa	tttttttgtc	taaagagtaa	2940
tactagagat	aaacataaaa	aatgtagagg	tcgagtttag	atgcaagttc	aaggagcgaa	3000
aggtggatgg	gtaggttata	tagggatata	gcacagagat	atatagcaaa	gagatacttt	3060
tgagcaatgt	ttgtggaagc	ggtattcgca	atatttttagt	agctcgttac	agtccggtgc	3120
gtttttgggt	ttttgaaagt	gcgtcttcag	agcgcttttg	gttttcaaaa	gcgctctgaa	3180
gttcctatac	tttctagaga	ataggaactt	cggaatagga	acttcaaagc	gtttccgaaa	3240
acgagcgctt	ccgaaaatgc	aacgcgagct	gcgcacatac	agctcactgt	tcacgtcgca	3300
cctatatctg	cgtgttgcc	gtatatatat	atacatgaga	agaacggcat	agtgcgtgtt	3360
tatgcttaaa	tgcgctactta	tatgcgtcta	tttatgtagg	atgaaaggta	gtctagtacc	3420
tcctgtgata	ttatcccatt	ccatgcgggg	tatcgtagtc	ttccttcagc	actacccttt	3480
agctgttcta	tatgctgcc	ctcctcaatt	ggattagtct	catccttcaa	tgctatcatt	3540
tcctttgata	ttggatcatc	taagaaacca	ttattatcat	gacattaacc	tataaaaata	3600
ggcgtatcac	gaggcccttt	cgtctcgcg	gtttcgggtga	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
ctttcccga	attttctttt	tctattactc	ttggcctcct	ctagtacact	ctatatTTTT	3960
ttatgcctcg	gtaatgattt	tcattttttt	ttttccccta	gcggatgact	cttttttttt	4020
cttagcgatt	ggcattatca	cataatgaat	tatacattat	ataaagtaat	gtgatttctt	4080
cgaagaatat	actaaaaaat	gagcaggcaa	gataaacgaa	ggcaaagatg	acagagcaga	4140
aagccctagt	aaagcgtatt	acaaatgaaa	ccaagattca	gattgcgatc	tcttttaaagg	4200
gtgggtccc	agcgatagag	cactcgatct	tcccagaaaa	agaggcagaa	gcagtagcag	4260
aacaggccac	acaatcgcaa	gtgattaacg	tccacacagg	tatagggttt	ctggaccata	4320
tgatacatgc	tctggccaag	cattccggct	ggtcgcta	cgttgagtgc	attggtgact	4380
tacacataga	cgaccatcac	accactgaag	actgcgggat	tgctctcggt	caagctttta	4440
aagaggccct	actggcgcg	ggagtaaaaa	ggtttggatc	aggatttgcg	cctttggatg	4500
aggcactttc	cagagcgggtg	gtagatcttt	cgaacaggcc	gtacgcagtt	gtcgaacttg	4560
gtttgcaaag	ggagaaagta	ggagatctct	cttgcgagat	gatcccgc	tttcttgaaa	4620
gctttgcaga	ggctagcaga	attaccctcc	acgttgattg	tctgagaggc	aagaatgatc	4680
atcaccgtag	tgagagtgcg	ttcaaggctc	ttgcgggttc	cataagagaa	gccacctcgc	4740
ccaatggtac	caacgatgtt	ccctccacca	aagggtgttct	tatgtagtga	caccgattat	4800
ttaaagctgc	agcatagat	atatatacat	gtgtatatat	gtatacctat	gaatgtcagt	4860
aagtatgtat	acgaacagta	tgatactgaa	gatgacaagg	taatgcatca	ttctatacgt	4920

20110907_CL5178USNA_SeqList.txt

gtcattctga	acgaggcgcg	ctttcctttt	ttctttttgc	tttttctttt	tttttctctt	4980
gaactcgacg	gatctatgcg	gtgtgaaata	ccgcacagat	gcgtaaggag	aaaataccgc	5040
atcaggaaat	tgtaaacggt	aatattttgt	taaaattcgc	gttaaatttt	tgttaaatca	5100
gctcattttt	taaccaatag	gccgaaatcg	gcaaaatccc	ttataaatca	aaagaataga	5160
ccgagatagg	gttgagtgtt	gttccagttt	ggaacaagag	tccactatta	aagaacgtgg	5220
actccaacgt	caaagggcga	aaaaccgtct	atcagggcga	tggcccacta	cgtgaaccat	5280
caccctaata	aagttttttg	gggtcgaggt	gccgtaaagc	actaaatcgg	aaccctaaag	5340
ggagcccccg	atttagagct	tgacggggaa	agccggcgaa	cgtggcgaga	aaggaagggg	5400
agaaagcgaa	aggagcgggc	gctagggcgc	tggcaagtgt	agcggtcacg	ctgcgcgtaa	5460
ccaccacacc	cgccgcgctt	aatgcgccgc	tacagggcgc	gtcgcgccat	tcgccattca	5520
ggctgcgcaa	ctgttgggaa	gggcgatcgg	tgcgggcctc	ttcgctatta	cgccagctgg	5580
cgaaaggggg	atgtgctgca	aggcgattaa	gttgggtaac	gccagggttt	tcccagtcac	5640
gacgttgtaa	aacgacggcc	agtgagcgcg	cgtaatacga	ctcactatag	ggcgaattgg	5700
gtaccggggc	ccccctcgag	gtcgacggta	tcgataagct	tgatatcgaa	ttcctgcgcc	5760
cggggcacta	gtcagatgcc	gcgggcactt	gagcacctca	tgcacagcaa	taacacaaca	5820
caatggttag	tagcaacctg	aattcgggtca	ttgatgcatg	catgtgccgt	gaagcgggac	5880
aaccagaaaa	gtcgtctata	aatgccggca	cgtgcgatca	tcgtggcggg	gttttaagag	5940
tgcatatcac	aaattgtcgc	attaccgcg	aaccgccaga	tattcattac	ttgacgcaaa	6000
agcgtttgaa	ataatgacga	aaaagaagga	agaaaaaaaa	agaaaaatac	cgcttctagg	6060
cgggttatct	actgatccga	gcttccacta	ggatagcacc	caaacacctg	catatttgga	6120
cgacctttac	ttacaccacc	aaaaaccact	ttcgccctctc	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
aattactgg	aaagatcaaa	caattgtttt	ccctgaagga	actgaacccc	gaattgtcgg	6660
tgcggcagcg	cgattagctg	cagacggctt	ggttaagccg	attgttttag	gtgcaacgga	6720
caaagttcag	gctgtggcta	acgatttgaa	tgcggtattta	acaggcgttc	aagtccttga	6780
tcctgcgaca	taccgggctg	aagataagca	agcaatgctt	gatgccctcg	ttgaacggcg	6840
gaaaggtaag	aatacgccag	aacaagcggc	taaaatgctg	gaagatgaaa	actactttgg	6900
cacgatgctc	gtttatatgg	gcaaagcggg	tgggatgggt	tcagggtgcaa	tccatccaac	6960

20110907_CL5178USNA_SeqList.txt

tggtgatacg	gtacggccag	cgttacaaat	tattaagacc	aagcccgggt	cacaccgaat	7020
ctcgggtgca	tttatcatgc	aaaaggggtga	ggaacgctac	gtctttgctg	actgtgccat	7080
caatattgat	cccgatgccg	atacgttagc	ggaaattgcc	actcagagtg	cggctactgc	7140
taaggctctc	gatattgacc	cgaaagttgc	gatgctcagc	ttctcaacta	agggttcggc	7200
taagggtgaa	atggctacta	aagtgcaga	agcaacggcc	aaggcgcaag	ctgctgaacc	7260
ggaattggct	atcgatgggt	aacttcaatt	tgacgcggcc	ttcgttgaaa	aagttggttt	7320
gcaaaaggct	cctgggtcca	aagtagctgg	tcatgccaat	gtctttgtat	ttccagagct	7380
tcagtctggg	aatattgggt	ataagattgc	gcaacgattt	ggtcattttg	aagcgggtggg	7440
tcctgtcttg	caaggcctga	acaagccggg	ctccgacttg	tcacgtggat	gcagtgaaga	7500
agacgtttat	aaggttgcca	ttattacagc	agcccaagga	ttagcttaat	taattaagag	7560
taagcgaatt	tcttatgatt	tatgattttt	attattaaat	aagttataaa	aaaaataagt	7620
gtatacaaat	tttaaagtga	ctcttaggtt	ttaaaacgaa	aattcttatt	cttgagtaac	7680
tctttcctgt	aggtcagggt	gctttctcag	gtatagcatg	aggtcgctct	tattgaccac	7740
acctctaccg	gcatgccgag	caaatgcctg	caaatcgctc	cccatttcac	ccaattgtag	7800
atatgctaac	tccagcaatg	agttgatgaa	tctcgggtgtg	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	tcagggtgtc	taactccttc	cttttcgggt	agagcggatg	8100
tgggaggagg	gcgtgaatgt	aagcgtgaca	taactaatta	catgattaat	taattatttt	8160
aaaccttcc	attgccaatc	attaacttct	ggcaagtcag	ttccggcatc	ccggatatag	8220
gcattgtgtt	tagcaagcat	attatccatg	gattgaacga	aggccgcacc	agtgttttcc	8280
attgctgggt	gcgccgcaat	tgccgactta	gctaagtcga	agcgggtccat	ctggttcatg	8340
acccgtacgt	cgaatgggtg	ggtaatatca	ccattttcac	ggtaaccgtg	gacgtataag	8400
ttatggttgt	gacgatcaaa	gaagatgtca	cgaactaagt	cttcgtaacc	gtggaaagca	8460
aagaccactg	gtttgtcctt	agtaaagtaa	tggtcaaact	cagcatctga	caagccccgc	8520
ggatcctttt	caggactacg	taacttcaag	atgtcgacca	cgttcacgaa	acgaatcttc	8580
atctctggga	aactgtcgtg	tagtaattgg	atggcagcca	acgtttcaag	cgttggttcc	8640
gtcccagcag	ctgcaaagac	aatgtctggg	tcgctacctt	ggcccgact	tgcccaatca	8700
atgataccaa	gaccattgtc	aactaattgc	ttagcttctt	caatgctgaa	ccattgttga	8760
cgtgggtgtt	ttgacgtaac	cacgtagttg	atcttttctt	ggctccggaa	aatgacgtca	8820
ccgacagcta	ataacgtgtt	ggcatcggct	ggtaaataat	cacgaatgta	ttctggtttc	8880
ttttcggcca	aatgagttaa	tgcacctgga	tcttggtggg	tataaccatt	atggctctgt	8940
tggaatacag	ttgaagccgc	gataatgtta	agtgatgggt	actttttacg	ccaatcaagt	9000

20110907_CL5178USNA_SeqList.txt

tcattggctt	tacgtaacca	cttgaagtgt	tgcgtcaaca	ttgagtccac	aacgcgtagg	9060
aaggcttcat	aactggcaaa	taacccatga	cgtccagtta	agacgtaacc	ttctaaccaa	9120
ccttcagctt	ggtgttcaga	taactgagca	tctaagaccc	ggccagctgg	tgcttcatat	9180
tggtcactat	ctggatgaat	gtcttccatc	cattgacgat	tagtggtttc	gaagacacca	9240
tataaacggg	tagacatggg	ttcatcaggt	ccgaacaacc	ggaagttatc	aggatttttc	9300
ttgatgacat	cccgc aaata	gtctgaccaa	acgatcatat	cttgcttaac	attcgcgctt	9360
tctttggacg	tatcgaccgc	ataatcacgg	aagtttggtg	agttcaaggc	tttcggatcg	9420
acccaccat	tggtgattgg	gtagcagcc	atccgactgt	ccccagtagg	aataattttt	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	tgctaattga	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	tatcgggtga	agtaccgtca	aggtaagagt	ttgaaaccat	cacttgacca	10260
ccatgacctg	gaccttcaac	gtagaacatc	ttcaaaccgt	acttgttgat	gacccggtta	10320
agatgagcat	agataaagtt	ttgaccggca	atcggtcccc	agtgaccaat	tggtatgaacc	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	gaagcttgag	10620
aaattgaaga	ccgtttatta	gcttaaatat	caatgggagg	tcatacgaaag	agaaaaaaat	10680
caagaaagaa	actctcaaga	aaaagaaacg	tgataaaaaat	ttttattgcc	tctctcgacg	10740
aagagaaaga	aacgaggcgg	tccctttttt	cttttccaaa	ccttttagtac	gggtaattag	10800
cgacacccta	gaggaagaaa	gaggggaaat	ttagtatgct	gtgcttgggt	gtcttgaagt	10860
ggtacggcga	tgcgcggagt	ccgagaaaat	ctggaagagt	aaaaaggggg	tagaagcggt	10920
ttgaagctat	ccgc					10934

<210> 212
 <211> 38
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> N821 Primer

 <400> 212
 cgcccgggcc actagtcaga tgccgcgggc acttgagc 38

 <210> 213
 <211> 30
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> N1115 Primer

 <400> 213
 tttttgtgct aatgactcaa ataaatccat 30

 <210> 214
 <211> 30
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> N1110 Primer

 <400> 214
 gcgatttaat ctctaattat tagttaaagt 30

 <210> 215
 <211> 30
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> N1114 Primer

 <400> 215
 atatgctggt gatgagtaat ctgttgatcat 30

 <210> 216
 <211> 22
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> N160SeqF5 Primer

 <400> 216
 cctgaagtct aggtccctat tt 22

 <210> 217
 <211> 38
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> BK468 Primer

<400> 217
gcctcagatt ttaatgttac ttctcttgca gttagggga 38

<210> 218
<211> 9220
<212> DNA
<213> Artificial Sequence

<220>
<223> pUC19-URA3::pdc1::TEF(M4)-xpk1::kan Plasmid

<400> 218
ccgcattgcg gattacgtat tctaattgttc agataacttc gtataatgta tgctatacga 60
agttatcgaa cagagaaact aaatccacat taattgagag ttctatctat tagaaaatgc 120
aaactccaac taaatgggaa aacagataac ctcttttatt tttttttaat gtttgatatt 180
cgagtctttt tcttttgta ggtttatatt catcatttca atgaataaaa gaagcttctt 240
attttggttg caaagaatga aaaaaaagga ttttttcata cttctaaagc ttcaattata 300
acaaaaaatt ttataaatga agagaaaaaa tctagtagta tcaagttaaa cttagaaaaa 360
ctcatcgagc atcaaatgaa actgcaattt attcatatca ggattatcaa taccatattt 420
ttgaaaaagc cgtttctgta atgaaggaga aaactcaccg aggcagttcc ataggatggc 480
aagatcctgg tatcgggtctg cgattccgac tcgtccaaca tcaatacaac ctattaattt 540
cccctcgtca aaaataaggt tatcaagtga gaaatcacca tgagtgcga ctgaatccgg 600
tgagaatggc aaaagcttat gcatttcttt ccagacttgt tcaacaggcc agccattacg 660
ctcgtcatca aaatcactcg catcaaccaa accgttattc attcgtgatt gcgcctgagc 720
gagacgaaat acgcgatcgc tgttaaaagg acaattacaa acaggaatcg aatgcaaccg 780
gcgcaggaac actgccagcg catcaacaat attttcacct gaatcaggat attcttctaa 840
tacctggaat gctgttttgc cggggatcgc agtgggtgagt aaccatgcat catcaggagt 900
acggataaaa tgcttgatgg tcggaagagg cataaatttc gtcagccagt ttagtctgac 960
catctcatct gtaacatcat tggcaacgct acctttgcc a tgtttcagaa acaactctgg 1020
cgcatcgggc ttcccataca atcgatagat tgtcgcacct gattgcccga cattatcgcg 1080
agcccattta taccatata aatcagcatc catgtttgaa tttaatcgcg gcctcgaaac 1140
gtgagtcttt tccttacc a tctcagattt taatgttact tctcttgag ttaggggaact 1200
ataatgtaac tcaaaataag attaaacaaa ctaaaataaa aagaagttat acagaaaaac 1260
ccatataaac cagtactaat ccataataat aatacacaaa aaaactatca aataaaacca 1320
gaaaacagat tgaatagaaa aattttttcg atctcctttt atattcaaaa ttcgatatat 1380
gaaaaaggg a actctcagaa aatcaccaaa tcaatttaat tagatttttc ttttccttct 1440
agcgttggaa agaaaaattt ttcttttttt ttttagaaat gaaaaatttt tgccgtagga 1500
atcaccgtat aaaccctgta taaacgctac tctgttcacc tgtgtaggct atgattgacc 1560
cagtgttcat tgttattgcg agagagcggg agaaaagaac cgatacaaga gatccatgct 1620
ggtatagttg tctgtccaac actttgatga acttgtagga cgatgatgtg tatttagacg 1680

20110907_CL5178USNA_SeqList.txt

agtacgtgtg	tgactattaa	gtagttatga	tagagagggtt	tgtacgggtgt	gttctgtgta	1740
attcgattga	gaaaatgggtt	atgaatccct	agataacttc	gtataatgta	tgctatacga	1800
agttatccag	tgatgataca	acgagtttagc	caagggtgggg	gatcctctag	agtcttaagg	1860
ccgcccga	attaaagcct	tcgagcgtcc	caaaaccttc	tcaagcaagg	ttttcagtat	1920
aatgttacat	gcgtacacgc	gtttgtacag	aaaaaaaaga	aaaatttgaa	atataaataa	1980
cgttcttaat	actaacataa	ctattaataaa	aaataaatag	ggacctagac	ttcaggttgt	2040
ctaactcctt	ccttttcggt	tagagcggat	gtgggaggag	ggcgtgaatg	taagcgtgac	2100
ataactaatt	acatgattaa	ttaattatctt	taaacccttc	cattgccaat	cattaacttc	2160
tggcaagtca	gttccggcat	cccggatata	ggcattgtgt	ttagcaagca	tattatccat	2220
ggattgaacg	aaggccgcac	cagtgttttc	cattgctggt	tgccgcaaa	ttgccgactt	2280
agctaagtcg	aagcgggtcca	tctggttcat	gaccggtacg	tcgaatggtg	tggtaatatc	2340
accattttca	cggtaaccgt	ggacgtataa	gttatgggtg	tgacgatcaa	agaagatgtc	2400
acgaactaag	tcttcgtaac	cgtggaaagc	aaagaccact	ggtttgtcct	tagtaaagta	2460
atgggtcaaac	tcagcatctg	acaagccccg	cggatccttt	tcaggactac	gtaacttcaa	2520
gatgtcgacc	acgttcacga	aacgaatctt	catctctggg	aaactgtcgt	gtagtaattg	2580
gatggcagcc	aacgtttcaa	gcgttggttc	cgtcccagca	gctgcaaaga	caatgtctgg	2640
ttcgctacct	tggtccgtac	ttgcccaatc	aatgatacca	agaccattgt	caactaattg	2700
cttagcttct	tcaatgctga	accattgttg	acgtgggtgt	tttgacgtaa	ccacgtagtt	2760
gatcttttct	tggctccgga	aaatgacgtc	accgacagct	aataacgtgt	tggcatcggc	2820
tggtaaatat	tcacgaatgt	attctgggtt	cttttcggcc	aaatgagtta	atgcacctgg	2880
atcttggtgg	gtataacat	tatggtcttg	ttggaataca	gttgaagccg	cgataatgtt	2940
aagtgatggg	tactttttac	gccaatcaag	ttcattgggt	ttacgtaacc	acttgaagtg	3000
ttgcgtcaac	attgagtcca	caacgcgtag	gaaggcttca	taactggcaa	ataacccatg	3060
acgtccagtt	aagacgtaac	cttctaacca	accttcagct	tggtgttcag	ataactgagc	3120
atctaagacc	cggccagctg	gtgcttcata	ttggtcacta	tctggatgaa	tgtcttccat	3180
ccattgacga	ttagtggttt	cgaagacacc	atataaacgg	ttagacatgg	tttcatcagg	3240
tccgaacaac	cgaagttat	caggattttt	cttgatgaca	tcccgcaa	agtctgacca	3300
aacgatcata	tcttgcttaa	cattcgcgcc	ttctttggac	gtatcgaccg	cataatcacg	3360
gaagtttgg	aagttcaagg	ctttcggatc	gacccaccca	ttggtgattg	ggtttagcagc	3420
catccgactg	tccccagtag	gaataatttc	tttaatatca	tccttcaaag	agccatcttc	3480
attgaagagt	tcttttgggt	gatatgattc	gagccaatca	actaaagcat	ccgcatgttc	3540
catgtcattt	tgatcaacag	gaatcggaat	ttgatgagca	cgaatgaac	cttcgatctt	3600
atcaccgtcc	catgacttcg	gaccagtcca	gcccttaggt	gcgcggaaga	cgatcattgg	3660
ccatactggc	aatgttgcac	cgttatcttc	gcgagcatgc	ttctggattg	ccttgatctt	3720

20110907_CL5178USNA_SeqList.txt

ttcaacggct	tcatccatgg	ccttagctaa	ggctgggtga	accttttcag	gacggtcacc	3780
ttcaacgaag	attgggtccc	aattcatgct	ttcgaagtat	tccttaattct	tagcatcaga	3840
agtccgacca	aaaatcggtg	gattagaaat	cttaaaacca	tttaagttca	agattggtaa	3900
aacagccccg	tcgttgattg	ggttaatgaa	cttcgttgat	tgccatgaag	ttgctaattg	3960
acccgtttcg	gattcccat	caccaacaac	aaccgcggcg	atttcgtcag	gattgtcaag	4020
aattgccccca	accccggtgtg	aaattgagta	accaagttcg	ccaccttcgt	ggattgaacc	4080
gggtgtttca	ggtgccgcat	gggaagcaac	cccacctggg	aatgagaatt	gcttgaagag	4140
cttttgcatc	ccttcaacat	cctgcgtaat	ttctggataa	atatcggtgt	aagtaccgtc	4200
aaggtaagag	tttgaacca	tcacttgacc	accatgacct	ggaccttcaa	cgtagaacat	4260
cttcaaaccg	tacttggtga	tgacctgggt	aagatgagca	tagataaagt	tttgaccggc	4320
aatcgcccc	cagtgacaa	ttggatgaac	cttaacgtca	ctggccttca	atggccgttg	4380
taatagtga	ttatctttta	aataaagttg	accaactgat	aagtagttgg	cagcacgcca	4440
gtacttatca	actttttgca	aatatgctgg	tgatgagtaa	tctgttggtca	tcctcagctg	4500
gaacttagat	tagattgcta	tgctttctct	ctaacgagca	agaagtaaaa	aaagttgtaa	4560
tagaacaaga	aaaatgaaac	tgaagcttga	gaaattgaag	accgtttatt	agcttaaata	4620
tcaatgggag	gtcatcgaaa	gagaaaaaaa	tcaagaaaga	aactctcaag	aaaaagaaac	4680
gtgataaaaa	tttttattgc	ctctctcgac	gaagagaaag	aaacgaggcg	gtcccttttt	4740
tcttttccaa	acctttagta	cgggtaatta	gcgacacct	agaggaagaa	agaggggaaa	4800
tttagtatgc	tgtgcttggg	tgtcttgaag	tggtacggcg	atgcgcggag	tccgagaaaa	4860
tctggaagag	taaaaagggg	gtagaagcgt	tttgaagcta	tccgcggtgg	ttaagcctaa	4920
ccaggccaat	tcaacagact	gtcggcaact	tcttgtctgg	tctttccatg	gtaagtgaca	4980
gtgcagtaat	aatatgaacc	aatttatatt	tcgttacata	aaaatgctta	taaaacttta	5040
actaataatt	agagattaaa	tcgcaaaccg	ccggccaatg	tggtgtggt	ttcaggggtcc	5100
ataaagcttt	tcaattcatc	tttttttttt	ttgttctttt	ttttgattcc	ggtttctttg	5160
aaattttttt	gattcggtaa	tctccgagca	gaaggaagaa	cgaaggaagg	agcacagact	5220
tagattggta	tatatacgca	tatgtggtgt	tgaagaaaca	tgaaattgcc	cagtattctt	5280
aacccaactg	cacagaacaa	aaacctgcag	gaaacgaaga	taaatcatgt	cgaaagctac	5340
atataaggaa	cgtgctgcta	ctcatcctag	tcctgttgct	gccaagctat	ttaatatcat	5400
gcacgaaaag	caaacaaact	tgtgtgcttc	attggatggt	cgtaccacca	aggaattact	5460
ggagttagtt	gaagcattag	gtcccaaaat	ttgtttacta	aaaacacatg	tgatattctt	5520
gactgatttt	tccatggagg	gcacagttaa	gccgctaag	gcattatccg	ccaagtacaa	5580
ttttttactc	ttcgaagaca	gaaaatttgc	tgacattggt	aatacagtca	aattgcagta	5640
ctctgcgggt	gtatacagaa	tagcagaatg	ggcagacatt	acgaatgcac	acggtgtggt	5700
gggcccagggt	attgttagcg	gtttgaagca	ggcgccggaa	gaagtaacaa	aggaacctag	5760

20110907_CL5178USNA_SeqList.txt

aggccttttg	atgtagcag	aattgtcatg	caagggctcc	ctagctactg	gagaatatac	5820
taaggggtact	gttgacattg	cgaagagcga	caaagatttt	gttatcggct	ttattgtctca	5880
aagagacatg	ggtggaagag	atgaagggtta	cgattgggtg	attatgacac	ccggtgtggg	5940
tttagatgac	aagggagacg	cattgggtca	acagtataga	accgtggatg	atgtggtctc	6000
tacaggatct	gacattatta	ttgttggaag	aggactattt	gcaaagggaa	gggatgctaa	6060
ggtagagggg	gaacgttaca	gaaaagcagg	ctgggaagca	tatttgagaa	gatgcggccca	6120
gcaaaactaa	aaaactgtat	tataagtaaa	tgcatgtata	ctaaactcac	aaattagagc	6180
ttcaatttaa	ttatatcagt	tattacccgg	gaatctcggg	cgtaatgatt	tctataatga	6240
cgaaaaaaaa	aaaattggaa	agaaaaagct	tcatggcctt	gcggccgctt	aattaatcta	6300
gagtcgacct	gcaggcatgc	aagcttggcg	taatcatggt	catagctggt	tcctgtgtga	6360
aattgttatc	cgctcacaat	tccacacaac	atacgagccg	gaagcataaa	gtgtaaagcc	6420
tgggggtgcct	aatgagtgag	ctaactcaca	ttaattgcgt	tgcgctcact	gcccgttttc	6480
cagtcgggaa	acctgtcgtg	ccagctgcat	taatgaatcg	gccaacgcgc	ggggagaggc	6540
ggtttgcgta	ttgggcgctc	ttccgcttcc	tcgctcactg	actcgctgcg	ctcggtcggt	6600
cggctgcggc	gagcgggtatc	agctcactca	aaggcggtaa	tacggttatc	cacagaatca	6660
ggggataacg	caggaaagaa	catgtgagca	aaaggccagc	aaaaggccag	gaaccgtaaa	6720
aaggccgcgt	tgctggcggt	tttccatagg	ctccgcccc	ctgacgagca	tcacaaaaat	6780
cgacgctcaa	gtcagagggtg	gcgaaacccg	acaggactat	aaagatacca	ggcgtttccc	6840
cctggaagct	ccctcgtgcg	ctctcctggt	ccgaccctgc	cgcttaccgg	atacctgtcc	6900
gcctttctcc	cttcgggaag	cgtggcgctt	tctcatagct	cacgctgtag	gtatctcagt	6960
tcggtgtagg	tcgttcgctc	caagctgggc	tgtgtgcacg	aacccccgt	tcagcccgcg	7020
cgctgcgcct	tatccggtaa	ctatcgtctt	gagtccaacc	cggtaagaca	cgacttatcg	7080
ccactggcag	cagccactgg	taacaggatt	agcagagcga	ggtatgtagg	cggtgctaca	7140
gagttcttga	agtgggtggc	taactacggc	tacactagaa	ggacagtatt	tgggtatctgc	7200
gctctgctga	agccagttac	cttcggaaaa	agagttggta	gctcttgatc	cggcaaacaa	7260
accaccgctg	gtagcggtgg	tttttttggt	tgcaagcagc	agattacgcg	cagaaaaaaaa	7320
ggatctcaag	aagatccttt	gatcttttct	acggggctctg	acgctcagtg	gaacgaaaac	7380
tcacgttaag	ggatttttgg	catgagatta	tcaaaaagga	tcttcaccta	gatcctttta	7440
aattaaaaat	gaagttttta	atcaatctaa	agtatatatg	agtaaacttg	gtctgacagt	7500
taccaatgct	taatcagtga	ggcacctatc	tcagcgatct	gtctatttcg	ttcatccata	7560
gttgccctgac	tccccgtcgt	gtagataact	acgatacggg	agggcttacc	atctggcccc	7620
agtgtgcaa	tgataccgcg	agaccacgc	tcaccggctc	cagatttatc	agcaataaac	7680
cagccagccg	gaagggccga	gcgcagaagt	ggtcctgcaa	ctttatccgc	ctccatccag	7740
tctattaatt	gttgccggga	agctagagta	agtagttcgc	cagttaatag	tttgcgcaac	7800

20110907_CL5178USNA_SeqList.txt

gttggtgcca	ttgctacagg	catcgtggtg	tcacgctcgt	cgtttggtat	ggcttcattc	7860
agctccgggt	cccaacgata	aaggcgagtt	acatgatccc	ccatgtttgtg	caaaaaagcg	7920
gtagctcct	tcggtcctcc	gatcgttgtc	agaagtaagt	tggccgcagt	gttatcactc	7980
atggttatgg	cagcactgca	taattctctt	actgtcatgc	catccgtaag	atgcttttct	8040
gtgactggtg	agtactcaac	caagtcattc	tgagaatagt	gtatgcggcg	accgagttgc	8100
tcttgcccgg	cgtcaatacg	ggataatacc	gcgccacata	gcagaacttt	aaaagtgctc	8160
atcattggaa	aacgtttctt	ggggcgaaaa	ctctcaagga	tcttaccgct	gttgagatcc	8220
agttcgatgt	aaccactcgc	tgcacccaac	tgatcttcag	catctttttac	tttcaccagc	8280
gtttctgggt	gagcaaaaac	aggaaggcaa	aatgccgcaa	aaaagggaat	aagggcgaca	8340
cggaaatgtt	gaatactcat	actcttcctt	tttcaatatt	attgaagcat	ttatcaggggt	8400
tattgtctca	tgagcggata	catatttgaa	tgtattttaga	aaaataaaca	aatagggggtt	8460
ccgcgcacat	ttccccgaaa	agtgccacct	gacgtctaag	aaaccattat	tatcatgaca	8520
ttaacctata	aaaataggcg	tatcacgagg	ccctttcgtc	tcgcgcggtt	cggtgatgac	8580
ggtgaaaacc	tctgacacat	gcagctcccg	gagacggtca	cagcttgtct	gtaagcggat	8640
gccgggagca	gacaagcccc	tcagggcgcg	tcagcgggtg	ttggcgggtg	tcggggctgg	8700
cttaactatg	cggcatcaga	gcagattgta	ctgagagtgc	accatatgcg	gtgtgaaata	8760
ccgcacagat	gcgtaaggag	aaaataccgc	atcaggcgcc	attcgccatt	caggctgcgc	8820
aactgttggg	aagggcgatc	ggtgcggggc	tcttcgctat	tacgccagct	ggcgaaaggg	8880
ggatgtgctg	caaggcgatt	aagttgggta	acgccaggggt	tttcccagtc	acgacgttgt	8940
aaaacgacgg	ccagtgaatt	cgagctcggt	acccggggat	ccggcgcgcc	gttttatttg	9000
tatcgaggtg	tctagtcttc	tattacacta	atgcagtttc	agggtttttg	aaaccacact	9060
gtttaaacag	tgttccttaa	tcaaggatac	ctcttttttt	ttccttggtt	ccactaattc	9120
atcggttttt	tttttggaag	acatcttttc	caacgaaaag	aatatacata	tcgtttaaga	9180
gaaattctcc	aaatttgtaa	agaagcggac	ccagacttaa			9220

<210> 219
 <211> 20
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> N886 Primer

<400> 219
 caatgattgt tggtaaaggg

20

<210> 220
 <211> 30
 <212> DNA
 <213> Artificial Sequence

<220>

<223> N1214 Primer

<400> 220
 aaaaaggggg tagaagcggtt ttgaagctat 30

<210> 221
 <211> 22
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> oBP512 Primer

<400> 221
 aaagttggca tagcggaaac tt 22

<210> 222
 <211> 19
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> N160SeqR5 Primer

<400> 222
 tgagcccgaa agagaggat 19

<210> 223
 <211> 500
 <212> PRT
 <213> Saccharomyces cerevisiae

<400> 223

Met Thr Lys Leu His Phe Asp Thr Ala Glu Pro Val Lys Ile Thr Leu
 1 5 10 15

Pro Asn Gly Leu Thr Tyr Glu Gln Pro Thr Gly Leu Phe Ile Asn Asn
 20 25 30

Lys Phe Met Lys Ala Gln Asp Gly Lys Thr Tyr Pro Val Glu Asp Pro
 35 40 45

Ser Thr Glu Asn Thr Val Cys Glu Val Ser Ser Ala Thr Thr Glu Asp
 50 55 60

Val Glu Tyr Ala Ile Glu Cys Ala Asp Arg Ala Phe His Asp Thr Glu
 65 70 75 80

Trp Ala Thr Gln Asp Pro Arg Glu Arg Gly Arg Leu Leu Ser Lys Leu
 85 90 95

Ala Asp Glu Leu Glu Ser Gln Ile Asp Leu Val Ser Ser Ile Glu Ala
 100 105 110

Leu Asp Asn Gly Lys Thr Leu Ala Leu Ala Arg Gly Asp Val Thr Ile
 115 120 125

20110907_CL5178USNA_SeqList.txt

Ala Ile Asn Cys Leu Arg Asp Ala Ala Ala Tyr Ala Asp Lys Val Asn
 130 135 140
 Gly Arg Thr Ile Asn Thr Gly Asp Gly Tyr Met Asn Phe Thr Thr Leu
 145 150 155 160
 Glu Pro Ile Gly Val Cys Gly Gln Ile Ile Pro Trp Asn Phe Pro Ile
 165 170 175
 Met Met Leu Ala Trp Lys Ile Ala Pro Ala Leu Ala Met Gly Asn Val
 180 185 190
 Cys Ile Leu Lys Pro Ala Ala Val Thr Pro Leu Asn Ala Leu Tyr Phe
 195 200 205
 Ala Ser Leu Cys Lys Lys Val Gly Ile Pro Ala Gly Val Val Asn Ile
 210 215 220
 Val Pro Gly Pro Gly Arg Thr Val Gly Ala Ala Leu Thr Asn Asp Pro
 225 230 235 240
 Arg Ile Arg Lys Leu Ala Phe Thr Gly Ser Thr Glu Val Gly Lys Ser
 245 250 255
 Val Ala Val Asp Ser Ser Glu Ser Asn Leu Lys Lys Ile Thr Leu Glu
 260 265 270
 Leu Gly Gly Lys Ser Ala His Leu Val Phe Asp Asp Ala Asn Ile Lys
 275 280 285
 Lys Thr Leu Pro Asn Leu Val Asn Gly Ile Phe Lys Asn Ala Gly Gln
 290 295 300
 Ile Cys Ser Ser Gly Ser Arg Ile Tyr Val Gln Glu Gly Ile Tyr Asp
 305 310 315 320
 Glu Leu Leu Ala Ala Phe Lys Ala Tyr Leu Glu Thr Glu Ile Lys Val
 325 330 335
 Gly Asn Pro Phe Asp Lys Ala Asn Phe Gln Gly Ala Ile Thr Asn Arg
 340 345 350
 Gln Gln Phe Asp Thr Ile Met Asn Tyr Ile Asp Ile Gly Lys Lys Glu
 355 360 365
 Gly Ala Lys Ile Leu Thr Gly Gly Glu Lys Val Gly Asp Lys Gly Tyr
 370 375 380
 Phe Ile Arg Pro Thr Val Phe Tyr Asp Val Asn Glu Asp Met Arg Ile
 385 390 395 400

20110907_CL5178USNA_SeqList.txt

Val Lys Glu Glu Ile Phe Gly Pro Val Val Thr Val Ala Lys Phe Lys
405 410 415

Thr Leu Glu Glu Gly Val Glu Met Ala Asn Ser Ser Glu Phe Gly Leu
420 425 430

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

Lys Met Leu Lys Ala Gly Thr Val Trp Ile Asn Thr Tyr Asn Asp Phe
450 455 460

Asp Ser Arg Val Pro Phe Gly Gly Val Lys Gln Ser Gly Tyr Gly Arg
465 470 475 480

Glu Met Gly Glu Glu Val Tyr His Ala Tyr Thr Glu Val Lys Ala Val
485 490 495

Arg Ile Lys Leu
500

<210> 224
<211> 343
<212> PRT
<213> Artificial sequence

<220>
<223> Anaerostipes caccae KARI variant K9D3

<400> 224

Met Glu Glu Cys Lys Met Ala Lys Ile Tyr Tyr Gln Glu Asp Cys Asn
1 5 10 15

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

Ile Ile Gly Leu Tyr Glu Gly Ala Lys Asp Trp Lys Arg Ala Glu Glu
50 55 60

Gln Gly Phe Glu Val Tyr Thr Ala Ala Glu Ala Ala Lys Lys Ala Asp
65 70 75 80

Ile Ile Met Ile Leu Ile Asn Asp Glu Lys Gln Ala Thr Met Tyr Lys
85 90 95

Asn Asp Ile Glu Pro Asn Leu Glu Ala Gly Asn Met Leu Met Phe Ala
100 105 110

His Gly Phe Asn Ile His Phe Gly Cys Ile Val Pro Pro Lys Asp Val
Page 239

115

120

125

Asp Val Thr Met Ile Ala Pro Lys Gly Pro Gly His Thr Val Arg Ser
 130 135 140

Glu Tyr Glu Glu Gly Lys Gly Val Pro Cys Leu Val Ala Val Glu Gln
 145 150 155 160

Asp Ala Thr Gly Lys Ala Leu Asp Met Ala Leu Ala Tyr Ala Leu Ala
 165 170 175

Ile Gly Gly Ala Arg Ala Gly Val Leu Glu Thr Thr Phe Arg Thr Glu
 180 185 190

Thr Glu Thr Asp Leu Phe Gly Glu Gln Ala Val Leu Cys Gly Gly Val
 195 200 205

Cys Ala Leu Met Gln Ala Gly Phe Glu Thr Leu Val Glu Ala Gly Tyr
 210 215 220

Asp Pro Arg Asn Ala Tyr Phe Glu Cys Ile His Glu Met Lys Leu Ile
 225 230 235 240

Val Asp Leu Ile Tyr Gln Ser Gly Phe Ser Gly Met Arg Tyr Ser Ile
 245 250 255

Ser Asn Thr Ala Glu Tyr Gly Asp Tyr Ile Thr Gly Pro Lys Ile Ile
 260 265 270

Thr Glu Asp Thr Lys Lys Ala Met Lys Lys Ile Leu Ser Asp Ile Gln
 275 280 285

Asp Gly Thr Phe Ala Lys Asp Phe Leu Val Asp Met Ser Asp Ala Gly
 290 295 300

Ser Gln Val His Phe Lys Ala Met Arg Lys Leu Ala Ser Glu His Pro
 305 310 315 320

Ala Glu Val Val Gly Glu Glu Ile Arg Ser Leu Tyr Ser Trp Ser Asp
 325 330 335

Glu Asp Lys Leu Ile Asn Asn
 340

<210> 225
 <211> 343
 <212> PRT
 <213> Artificial sequence

<220>
 <223> Anaerostipes caccae KARI variant K9G9

<400> 225

20110907_CL5178USNA_SeqList.txt

Met Glu Glu Cys Lys Met Ala Lys Ile Tyr Tyr Gln Glu Asp Cys Asn
1 5 10 15

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

Ile Ile Gly Leu Tyr Glu Gly Ala Lys Glu Trp Lys Arg Ala Glu Glu
50 55 60

Gln Gly Phe Glu Val Tyr Thr Ala Ala Glu Ala Ala Lys Lys Ala Asp
65 70 75 80

Ile Ile Met Ile Leu Ile Asn Asp Glu Lys Gln Ala Thr Met Tyr Lys
85 90 95

Asn Asp Ile Glu Pro Asn Leu Glu Ala Gly Asn Met Leu Met Phe Ala
100 105 110

His Gly Phe Asn Ile His Phe Gly Cys Ile Val Pro Pro Lys Asp Val
115 120 125

Asp Val Thr Met Ile Ala Pro Lys Gly Pro Gly His Thr Val Arg Ser
130 135 140

Glu Tyr Glu Glu Gly Lys Gly Val Pro Cys Leu Val Ala Val Glu Gln
145 150 155 160

Asp Ala Thr Gly Lys Ala Leu Asp Met Ala Leu Ala Tyr Ala Leu Ala
165 170 175

Ile Gly Gly Ala Arg Ala Gly Val Leu Glu Thr Thr Phe Arg Thr Glu
180 185 190

Thr Glu Thr Asp Leu Phe Gly Glu Gln Ala Val Leu Cys Gly Gly Val
195 200 205

Cys Ala Leu Met Gln Ala Gly Phe Glu Thr Leu Val Glu Ala Gly Tyr
210 215 220

Asp Pro Arg Asn Ala Tyr Phe Glu Cys Ile His Glu Met Lys Leu Ile
225 230 235 240

Val Asp Leu Ile Tyr Gln Ser Gly Phe Ser Gly Met Arg Tyr Ser Ile
245 250 255

Ser Asn Thr Ala Glu Tyr Gly Asp Tyr Ile Thr Gly Pro Lys Ile Ile
260 265 270

20110907_CL5178USNA_SeqList.txt

Thr Glu Asp Thr Lys Lys Ala Met Lys Lys Ile Leu Ser Asp Ile Gln
275 280 285

Asp Gly Thr Phe Ala Lys Asp Phe Leu Val Asp Met Ser Asp Ala Gly
290 295 300

Ser Gln Val His Phe Lys Ala Met Arg Lys Leu Ala Ser Glu His Pro
305 310 315 320

Ala Glu Val Val Gly Glu Glu Ile Arg Ser Leu Tyr Ser Trp Ser Asp
325 330 335

Glu Asp Lys Leu Ile Asn Asn
340

<210> 226
<211> 267
<212> PRT
<213> Saccharomyces cerevisiae

<400> 226

Met Ser Gln Gly Arg Lys Ala Ala Glu Arg Leu Ala Lys Lys Thr Val
1 5 10 15

Leu Ile Thr Gly Ala Ser Ala Gly Ile Gly Lys Ala Thr Ala Leu Glu
20 25 30

Tyr Leu Glu Ala Ser Asn Gly Asp Met Lys Leu Ile Leu Ala Ala Arg
35 40 45

Arg Leu Glu Lys Leu Glu Glu Leu Lys Lys Thr Ile Asp Gln Glu Phe
50 55 60

Pro Asn Ala Lys Val His Val Ala Gln Leu Asp Ile Thr Gln Ala Glu
65 70 75 80

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

Asp Ile Leu Val Asn Asn Ala Gly Lys Ala Leu Gly Ser Asp Arg Val
100 105 110

Gly Gln Ile Ala Thr Glu Asp Ile Gln Asp Val Phe Asp Thr Asn Val
115 120 125

Thr Ala Leu Ile Asn Ile Thr Gln Ala Val Leu Pro Ile Phe Gln Ala
130 135 140

Lys Asn Ser Gly Asp Ile Val Asn Leu Gly Ser Ile Ala Gly Arg Asp
145 150 155 160

20110907_CL5178USNA_SeqList.txt

Ala Tyr Pro Thr Gly Ser Ile Tyr Cys Ala Ser Lys Phe Ala Val Gly
165 170 175

Ala Phe Thr Asp Ser Leu Arg Lys Glu Leu Ile Asn Thr Lys Ile Arg
180 185 190

Val Ile Leu Ile Ala Pro Gly Leu Val Glu Thr Glu Phe Ser Leu Val
195 200 205

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

Thr Pro Leu Met Ala Asp Asp Val Ala Asp Leu Ile Val Tyr Ala Thr
225 230 235 240

Ser Arg Lys Gln Asn Thr Val Ile Ala Asp Thr Leu Ile Phe Pro Thr
245 250 255

Asn Gln Ala Ser Pro His His Ile Phe Arg Gly
260 265

<210> 227
<211> 2073
<212> DNA
<213> Saccharomyces cerevisiae

<400> 227
atggaaggct tcaatccggc tgacatagaa catgcgtcac cgattaattc atctgacagc 60
cattcatcct cctttgtata tgctctaccc aaaagtgcta gtgaatatgt agtcaaccat 120
aatgaggggtc gtgcaagtgc aagtggaaat ccagccgcag tgccgtctcc cataatgaca 180
ctgaatctca aaagcacaca ttccctcaat attgatcagc atgttcatac ctcaacatcg 240
ccgacggaaa ctattgggca tattcatcat gtggaaaagc tgaatcaaaa caatttgatt 300
catctggatc cagtacccaa ctttgaagat aagtccgata ttaagccttg gttgcaaaaag 360
atTTTTtATC ctcaaggaat agaacttgTG atagaaaggT cggacgcatt taaagttgtc 420
ttcaagtgtA aagctgctaa aaggggaagg aacgcgagaa ggaaaagaaa agataagccc 480
aaaggacagg accacgaaga cgagaaatcc aagatcaatg atgacgaatt agaatatgcg 540
agtccttcta atgccacagt aaccaatggg cctcaaacat cgcccgatca aacatcctcc 600
ataaagccaa agaaaaaag atgtgtatcg aggtttaata actgtccgtt tagagtacga 660
gctacttatt cgtaaagag gaaaagatgg agcattgttg taatggacaa taaccattca 720
catcagctaa agtttaaccc tgattccgaa gagtacaaaa aattcaaaga aaaattaaga 780
aaggataatg acgtagatgc aatcaagaaa ttcgacgaat tggaatacag aactttggcc 840
aatttgccca ttccaacagc tacaatcccc tgtgattgtg gtttaacaaa tgaatacaaa 900
agtttcaatg tcgtattgcc cactaacagt aatgttactt catcagcatc ctcttcaact 960
gtatcgtcca tatcccttga ttcacgaat gcatctaaaa ggccatgctt accctctgta 1020

20110907_CL5178USNA_SeqList.txt

```

aataacaccg gtagtatcaa taccaataac gtaaggaaac cgaaaagcca gtgtaagaat 1080
aaagacacac tcttaaaaag aaccaccatg cagaactttc tcacaactaa atcaaggctg 1140
cgtaagaccg gtacgccaac atcttcgcaa cactcatcta cagcattttc aggatatatt 1200
gatgatcctt tcaatttgaa tgaaatcttg ccactgccgg catccgattt caagctaaac 1260
actgtaacaa atttgaacga aattgacttt acgaacattt ttaccaaadc gccgcatcca 1320
catagcgggt ctacccatcc aagacaagtc ttcgaccaat tggacgattg ttcctctata 1380
ctcttctctc cattaactac aaacacgaat aatgaatttg aaggagagtc agatgatttt 1440
gttcattctc catatttgaa ctcagaggca gatttcagcc aaattcttag tagtgctccc 1500
ccagtccatc atgacccaaa tgaaacacat caggaaaacc aggatattat tgatagattt 1560
gctaatagtt cccaagaaca taatgagtat attctacaat atttgacgca ctccgatgct 1620
gctaaccaca ataacatcgg cgttccaaac aacaattcac attcgctaaa tactcagcat 1680
aacgtttctg atctgggcaa ctcactttta agacaagaag ctttagttgg cagctcttca 1740
acaaaaatct tcgacgaatt gaaatttgta caaaatggcc cacacggttc tcaacatcct 1800
atagattttc aacatgttga ccatcgtcac ctcagctcta atgaacctca agtacgatca 1860
catcaatatg gtccgcaaca gcagccaccg cagcaattgc aatatcacca aaatcagccc 1920
cacgacggcc ataaccacga acagcaccaa acagtacaaa aggatatgca aacgcatgaa 1980
tcgctagaaa taatgggaaa cacattattg gaagagttca aagacattaa aatggtgaac 2040
ggcgagtga agtatgtgaa gccagaagat tag 2073

```

```

<210> 228
<211> 690
<212> PRT
<213> Saccharomyces cerevisiae

```

```
<400> 228
```

```
Met Glu Gly Phe Asn Pro Ala Asp Ile Glu His Ala Ser Pro Ile Asn
1           5           10          15
```

```
Ser Ser Asp Ser His Ser Ser Ser Phe Val Tyr Ala Leu Pro Lys Ser
20          25          30
```

```
Ala Ser Glu Tyr Val Val Asn His Asn Glu Gly Arg Ala Ser Ala Ser
35          40          45
```

```
Gly Asn Pro Ala Ala Val Pro Ser Pro Ile Met Thr Leu Asn Leu Lys
50          55          60
```

```
Ser Thr His Ser Leu Asn Ile Asp Gln His Val His Thr Ser Thr Ser
65          70          75          80
```

```
Pro Thr Glu Thr Ile Gly His Ile His His Val Glu Lys Leu Asn Gln
85          90          95
```

20110907_CL5178USNA_SeqList.txt

Asn Asn Leu Ile His Leu Asp Pro Val Pro Asn Phe Glu Asp Lys Ser
100 105 110

Asp Ile Lys Pro Trp Leu Gln Lys Ile Phe Tyr Pro Gln Gly Ile Glu
115 120 125

Leu Val Ile Glu Arg Ser Asp Ala Phe Lys Val Val Phe Lys Cys Lys
130 135 140

Ala Ala Lys Arg Gly Arg Asn Ala Arg Arg Lys Arg Lys Asp Lys Pro
145 150 155 160

Lys Gly Gln Asp His Glu Asp Glu Lys Ser Lys Ile Asn Asp Asp Glu
165 170 175

Leu Glu Tyr Ala Ser Pro Ser Asn Ala Thr Val Thr Asn Gly Pro Gln
180 185 190

Thr Ser Pro Asp Gln Thr Ser Ser Ile Lys Pro Lys Lys Lys Arg Cys
195 200 205

Val Ser Arg Phe Asn Asn Cys Pro Phe Arg Val Arg Ala Thr Tyr Ser
210 215 220

Leu Lys Arg Lys Arg Trp Ser Ile Val Val Met Asp Asn Asn His Ser
225 230 235 240

His Gln Leu Lys Phe Asn Pro Asp Ser Glu Glu Tyr Lys Lys Phe Lys
245 250 255

Glu Lys Leu Arg Lys Asp Asn Asp Val Asp Ala Ile Lys Lys Phe Asp
260 265 270

Glu Leu Glu Tyr Arg Thr Leu Ala Asn Leu Pro Ile Pro Thr Ala Thr
275 280 285

Ile Pro Cys Asp Cys Gly Leu Thr Asn Glu Ile Gln Ser Phe Asn Val
290 295 300

Val Leu Pro Thr Asn Ser Asn Val Thr Ser Ser Ala Ser Ser Ser Thr
305 310 315 320

Val Ser Ser Ile Ser Leu Asp Ser Ser Asn Ala Ser Lys Arg Pro Cys
325 330 335

Leu Pro Ser Val Asn Asn Thr Gly Ser Ile Asn Thr Asn Asn Val Arg
340 345 350

Lys Pro Lys Ser Gln Cys Lys Asn Lys Asp Thr Leu Leu Lys Arg Thr
355 360 365

20110907_CL5178USNA_SeqList.txt

Thr	Met	Gln	Asn	Phe	Leu	Thr	Thr	Lys	Ser	Arg	Leu	Arg	Lys	Thr	Gly
370						375					380				
Thr	Pro	Thr	Ser	Ser	Gln	His	Ser	Ser	Thr	Ala	Phe	Ser	Gly	Tyr	Ile
385					390					395					400
Asp	Asp	Pro	Phe	Asn	Leu	Asn	Glu	Ile	Leu	Pro	Leu	Pro	Ala	Ser	Asp
				405					410					415	
Phe	Lys	Leu	Asn	Thr	Val	Thr	Asn	Leu	Asn	Glu	Ile	Asp	Phe	Thr	Asn
			420					425					430		
Ile	Phe	Thr	Lys	Ser	Pro	His	Pro	His	Ser	Gly	Ser	Thr	His	Pro	Arg
		435					440					445			
Gln	Val	Phe	Asp	Gln	Leu	Asp	Asp	Cys	Ser	Ser	Ile	Leu	Phe	Ser	Pro
	450					455					460				
Leu	Thr	Thr	Asn	Thr	Asn	Asn	Glu	Phe	Glu	Gly	Glu	Ser	Asp	Asp	Phe
465					470					475					480
Val	His	Ser	Pro	Tyr	Leu	Asn	Ser	Glu	Ala	Asp	Phe	Ser	Gln	Ile	Leu
				485					490					495	
Ser	Ser	Ala	Pro	Pro	Val	His	His	Asp	Pro	Asn	Glu	Thr	His	Gln	Glu
			500					505					510		
Asn	Gln	Asp	Ile	Ile	Asp	Arg	Phe	Ala	Asn	Ser	Ser	Gln	Glu	His	Asn
		515					520					525			
Glu	Tyr	Ile	Leu	Gln	Tyr	Leu	Thr	His	Ser	Asp	Ala	Ala	Asn	His	Asn
	530					535					540				
Asn	Ile	Gly	Val	Pro	Asn	Asn	Asn	Ser	His	Ser	Leu	Asn	Thr	Gln	His
545					550					555					560
Asn	Val	Ser	Asp	Leu	Gly	Asn	Ser	Leu	Leu	Arg	Gln	Glu	Ala	Leu	Val
				565					570					575	
Gly	Ser	Ser	Ser	Thr	Lys	Ile	Phe	Asp	Glu	Leu	Lys	Phe	Val	Gln	Asn
			580					585					590		
Gly	Pro	His	Gly	Ser	Gln	His	Pro	Ile	Asp	Phe	Gln	His	Val	Asp	His
		595					600					605			
Arg	His	Leu	Ser	Ser	Asn	Glu	Pro	Gln	Val	Arg	Ser	His	Gln	Tyr	Gly
	610					615					620				
Pro	Gln	Gln	Gln	Pro	Pro	Gln	Gln	Leu	Gln	Tyr	His	Gln	Asn	Gln	Pro
625					630					635					640

20110907_CL5178USNA_SeqList.txt

His Asp Gly His Asn His Glu Gln His Gln Thr Val Gln Lys Asp Met
645 650 655

Gln Thr His Glu Ser Leu Glu Ile Met Gly Asn Thr Leu Leu Glu Glu
660 665 670

Phe Lys Asp Ile Lys Met Val Asn Gly Glu Leu Lys Tyr Val Lys Pro
675 680 685

Glu Asp
690

<210> 229
<211> 1251
<212> DNA
<213> Saccharomyces cerevisiae

<400> 229
atgaaagcaa agtcgatgaa gagcataata tctgtgccaa tatcagtgtc taaaacaggg 60
aaaatgaagt taacagcatc tcccgataat ctagcatcta tgatgtctaa agacccaaat 120
aaacttattc atttagacc cgttccgtct tttgaggaca ggcacgagat taaaccgtgg 180
ctacagaaaa tattctatcc tcaaggaata gatatagtca tcgaaagatc agatagtagc 240
aaagttactt ttaagtgtag gtcagtacga tcgaaagtag ggttgaaccc taaatccaaa 300
ggttcctctt caagatccca tgcattgtcca tttagaatta gagcagccta ctctgtgcgg 360
ttacagaagt ggaacgtggg tggtatgaat aacatacatt cacatgagct gagatttgat 420
ttgattacta aaacggatga ctataaaaag ttttaaggaaa atcttcgtca gaagaacgac 480
gaaaaagcca tcaaaacatt tgatgaatta gaataataag caagtctaaa cttaccactc 540
gtcacaccaa ttatttcttg tgactgtggg ttaaccaagg aaattgaagc tttcaataac 600
atttttctcc cactatcaaa cccccctttg acgtctaaaa aaaatttgct caaaacaaat 660
aaaaactccg tttctaaaat aaaatccaga cagatggata attcgaaacc taggccaagg 720
ttgaaaacga agttggatgc ggatctacat gataccggat ttttagacaa tttcaaaacg 780
cggaattcct gtgtaaagat agagaaagaa gattctttta cgaacttaaa cgaaattgat 840
tttacgaaca tgttttgcaa cgacaatttc atccaaaatt acaaccaagg cctgatggaa 900
cttttgacag aacctacacc aggtccttcg tcctcttctt gtatacttcc ttcaacacca 960
acgaggccct tgcacaaaag caaaatggat atagccctgt cagaatcaac aacatcttca 1020
ccaaatttta tggaacaga cgctccttat ggagatgaga tcattaaagt ttccaaagat 1080
acaaaaagca acgcaccaac tgctgataca gatattgcta cgaacttagg caaagagaga 1140
aacgaaaatt ttggcatgtt aaactacaac tacgaagcgt tgcttcattt caacgatgag 1200
cattttaatg aattaaattc tattgacca gccttaatat caaaatatta a 1251

<210> 230
<211> 416
<212> PRT

<213> Saccharomyces cerevisiae

<400> 230

Met Lys Ala Lys Ser Met Lys Ser Ile Ile Ser Val Pro Ile Ser Val
1 5 10 15

Ser Lys Thr Gly Lys Met Lys Leu Thr Ala Ser Pro Asp Asn Leu Ala
20 25 30

Ser Met Met Ser Lys Asp Gln Asn Lys Leu Ile His Leu Asp Pro Val
35 40 45

Pro Ser Phe Glu Asp Arg His Glu Ile Lys Pro Trp Leu Gln Lys Ile
50 55 60

Phe Tyr Pro Gln Gly Ile Asp Ile Val Ile Glu Arg Ser Asp Ser Ser
65 70 75 80

Lys Val Thr Phe Lys Cys Arg Ser Val Arg Ser Lys Val Gly Leu Asn
85 90 95

Pro Lys Ser Lys Gly Ser Ser Ser Arg Ser His Ala Cys Pro Phe Arg
100 105 110

Ile Arg Ala Ala Tyr Ser Val Arg Leu Gln Lys Trp Asn Val Val Val
115 120 125

Met Asn Asn Ile His Ser His Glu Leu Arg Phe Asp Leu Ile Thr Lys
130 135 140

Thr Asp Asp Tyr Lys Lys Phe Lys Glu Asn Leu Arg Gln Lys Asn Asp
145 150 155 160

Glu Lys Ala Ile Lys Thr Phe Asp Glu Leu Glu Tyr Lys Ala Ser Leu
165 170 175

Asn Leu Pro Leu Val Thr Pro Ile Ile Ser Cys Asp Cys Gly Leu Thr
180 185 190

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

Pro Leu Thr Ser Lys Lys Asn Leu Leu Lys Thr Asn Lys Asn Ser Val
210 215 220

Ser Lys Ile Lys Ser Arg Gln Met Asp Asn Ser Lys Pro Arg Pro Arg
225 230 235 240

Leu Lys Thr Lys Leu Asp Ala Asp Leu His Asp Thr Gly Phe Leu Asp
245 250 255

20110907_CL5178USNA_SeqList.txt

Asn Phe Lys Thr Arg Asn Ser Cys Val Lys Ile Glu Lys Glu Asp Ser
260 265 270

Leu Thr Asn Leu Asn Glu Ile Asp Phe Thr Asn Met Phe Cys Asn Asp
275 280 285

Asn Phe Ile Gln Asn Tyr Asn Gln Gly Leu Met Glu Leu Leu Thr Glu
290 295 300

Pro Thr Pro Gly Pro Ser Ser Ser Ser Cys Ile Leu Pro Ser Thr Pro
305 310 315 320

Thr Arg Pro Leu Ser Gln Ser Lys Met Asp Ile Ala Leu Ser Glu Ser
325 330 335

Thr Thr Ser Ser Pro Asn Phe Met Glu Thr Asp Ala Pro Tyr Gly Asp
340 345 350

Glu Ile Ile Lys Val Ser Lys Asp Thr Lys Ser Asn Ala Pro Thr Ala
355 360 365

Asp Thr Asp Ile Ala Thr Asn Leu Gly Lys Glu Arg Asn Glu Asn Phe
370 375 380

Gly Met Leu Asn Tyr Asn Tyr Glu Ala Leu Leu His Phe Asn Asp Glu
385 390 395 400

His Phe Asn Glu Leu Asn Ser Ile Asp Pro Ala Leu Ile Ser Lys Tyr
405 410 415

<210> 231
<211> 363
<212> DNA
<213> Saccharomyces cerevisiae

<400> 231
atgacaggtg aaagaattga aaaggtgaaa ataaatgacg aatttgcaaa atcacatttc 60
ctaacaactc agtggagaga acaaagcgc caacgacact ataagatgcc cgttactgaa 120
cagggactcc gcgagaggat agaatcagcc ataccgcaag ttaccatat cattgtgacg 180
gatctttcgt acggttgtgg tcagtcgttt gatattgtgg tggtcagcga cttctttcaa 240
ggtaaaagca aactaatgag gagtcgtgca gtgaacaagg ctgtgaaaga agagctgcag 300
gagattcatg cctttagctg caagtgctac actgaggagg aatggtctaa gattgtggta 360
tga 363

<210> 232
<211> 120
<212> PRT
<213> Saccharomyces cerevisiae

<400> 232

20110907_CL5178USNA_SeqList.txt

Met Thr Gly Glu Arg Ile Glu Lys Val Lys Ile Asn Asp Glu Phe Ala
1 5 10 15

Lys Ser His Phe Leu Thr Thr Gln Trp Arg Glu Thr Lys Arg Gln Arg
20 25 30

His Tyr Lys Met Pro Val Thr Glu Gln Gly Leu Arg Glu Arg Ile Glu
35 40 45

Ser Ala Ile Pro Gln Val Tyr His Ile Ile Val Thr Asp Leu Ser Tyr
50 55 60

Gly Cys Gly Gln Ser Phe Asp Ile Val Val Val Ser Asp Phe Phe Gln
65 70 75 80

Gly Lys Ser Lys Leu Met Arg Ser Arg Ala Val Asn Lys Ala Val Lys
85 90 95

Glu Glu Leu Gln Glu Ile His Ala Phe Ser Cys Lys Cys Tyr Thr Glu
100 105 110

Glu Glu Trp Ser Lys Ile Val Val
115 120

<210> 233
<211> 858
<212> DNA
<213> Saccharomyces cerevisiae

<400> 233
atgtgttctt ttcaggttcc atctgcattt tcttttaact acacctcgta ctgtttataaa 60
cgccaccaag caagatatta cacagcagca aaactttttc aggaaatgcc tgttattgaa 120
attaacgatc aagagcaatt tacttaccta actaccactg cggccggcga caagttaatc 180
gtgctttatt tccataccag ttgggcagaa ccatgcaaag cattaaagca ggtttttgag 240
gccattagta atgagccttc caattccaac gtctctttct tatccattga tgcggacgaa 300
aactcggaaa tttcagaact ttttgaaatc tcagctgttc catattttat cataattcac 360
aaagggacaa tcttaaaaga attatccggc gcggatccaa aggagtatgt gtctttatta 420
gaagactgca agaactcagt caattccgga tcatcacaaa ctcatactat ggaaaatgca 480
aacgtaaatg aggggagtca taatgatgaa gacgatgacg acgaagaaga ggaagaagaa 540
actgaggagc aaataaacgc tagattgact aaattgggtca atgccgcgcc ggtaatgtta 600
tttatgaagg ggagcccctc tgaacctaaa tgcgggtttt cgagacaact tgtgggtatc 660
ttgagagaac atcaagtaag atttggcttc tttgatatat taagagacga atctgttaga 720
caaaacttga aaaagttttc tgaatggcca actttccctc aactttatat aaatggggag 780
tttcaaggcg gtttagacat tatcaaggaa tccttggagg aagaccctga ttttttgcag 840
catgctctcc aatcttaa 858

20110907_CL5178USNA_SeqList.txt

<210> 234
 <211> 285
 <212> PRT
 <213> *Saccharomyces cerevisiae*
 <400> 234
 Met Cys Ser Phe Gln Val Pro Ser Ala Phe Ser Phe Asn Tyr Thr Ser
 1 5 10 15
 Tyr Cys Tyr Lys Arg His Gln Ala Arg Tyr Tyr Thr Ala Ala Lys Leu
 20 25 30
 Phe Gln Glu Met Pro Val Ile Glu Ile Asn Asp Gln Glu Gln Phe Thr
 35 40 45
 Tyr Leu Thr Thr Thr Ala Ala Gly Asp Lys Leu Ile Val Leu Tyr Phe
 50 55 60
 His Thr Ser Trp Ala Glu Pro Cys Lys Ala Leu Lys Gln Val Phe Glu
 65 70 75 80
 Ala Ile Ser Asn Glu Pro Ser Asn Ser Asn Val Ser Phe Leu Ser Ile
 85 90 95
 Asp Ala Asp Glu Asn Ser Glu Ile Ser Glu Leu Phe Glu Ile Ser Ala
 100 105 110
 Val Pro Tyr Phe Ile Ile Ile His Lys Gly Thr Ile Leu Lys Glu Leu
 115 120 125
 Ser Gly Ala Asp Pro Lys Glu Tyr Val Ser Leu Leu Glu Asp Cys Lys
 130 135 140
 Asn Ser Val Asn Ser Gly Ser Ser Gln Thr His Thr Met Glu Asn Ala
 145 150 155 160
 Asn Val Asn Glu Gly Ser His Asn Asp Glu Asp Asp Asp Asp Glu Glu
 165 170 175
 Glu Glu Glu Glu Thr Glu Glu Gln Ile Asn Ala Arg Leu Thr Lys Leu
 180 185 190
 Val Asn Ala Ala Pro Val Met Leu Phe Met Lys Gly Ser Pro Ser Glu
 195 200 205
 Pro Lys Cys Gly Phe Ser Arg Gln Leu Val Gly Ile Leu Arg Glu His
 210 215 220
 Gln Val Arg Phe Gly Phe Phe Asp Ile Leu Arg Asp Glu Ser Val Arg
 225 230 235 240

20110907_CL5178USNA_SeqList.txt

Gln Asn Leu Lys Lys Phe Ser Glu Trp Pro Thr Phe Pro Gln Leu Tyr
245 250 255

Ile Asn Gly Glu Phe Gln Gly Gly Leu Asp Ile Ile Lys Glu Ser Leu
260 265 270

Glu Glu Asp Pro Asp Phe Leu Gln His Ala Leu Gln Ser
275 280 285

<210> 235
<211> 969
<212> DNA
<213> *Saccharomyces cerevisiae*

<400> 235
atgtccattg tagcactaaa gaacgcagtg gtgaccctta tacagaaagc gaaaggtagt 60
ggtggaacct cagagttggg ggggtctgaa tcaactccct tgttgagggg tagtaatagc 120
aatagttcaa ggcatgataa cttatcctca tctagctcgg atattatcta tggtagaaat 180
tcagcgcagg atctagaaaa ctcaccgatg tcagtaggga aagataatag gaatggcgat 240
aacggttcgg ataacgaaaa ggcgaaccta gggttcttcc aatcagtaga tcctcgcgta 300
attagcgatt tgattatcgg gctaagcgac ggtttaaccg tcccctttgc tctaacagct 360
ggtctatctt cacttggtga tgcgaaattg gtcattaccg gtggctttgc tgagttgatc 420
tcaggtgcga tttctatggg ctttggtggc tatttgaggag cgaagagtga atctgattat 480
tatcatgctg aagtaaagaa ggagaaaagg aaattttacg ataactccaa cctaattaac 540
agagaaattg aagacattct gttagagatt aaccctaatt tctcggacga aacaatcggt 600
tcgttcatca aagatttaca aagaacacct gagctaattg ttgacttcat tatcagggtac 660
ggtaggggtc tagacgaacc tgccgagaat agagaactaa tcagtgcagt cactatcggg 720
ggcggttatc tacttggtgg gctagtacca ctagtgccat atttttttgt ctcagatgtg 780
ggcacaggtc tcatattatc tataatagtc atggttgtaa cattattctg gtttggttac 840
gtaaagacaa aactatccat gggtagtggc agttcgactt caaagaaagt tacggaaggt 900
gttgaaatgg ttgttggtgg tggtgtagca gcagggtcgg cttgggttctt tgттаagtta 960
ctgggttaa 969

<210> 236
<211> 322
<212> PRT
<213> *Saccharomyces cerevisiae*

<400> 236

Met Ser Ile Val Ala Leu Lys Asn Ala Val Val Thr Leu Ile Gln Lys
1 5 10 15

Ala Lys Gly Ser Gly Gly Thr Ser Glu Leu Gly Gly Ser Glu Ser Thr
20 25 30

20110907_CL5178USNA_SeqList.txt

Pro Leu Leu Arg Gly Ser Asn Ser Asn Ser Ser Arg His Asp Asn Leu
35 40 45

Ser Ser Ser Ser Ser Asp Ile Ile Tyr Gly Arg Asn Ser Ala Gln Asp
50 55 60

Leu Glu Asn Ser Pro Met Ser Val Gly Lys Asp Asn Arg Asn Gly Asp
65 70 75 80

Asn Gly Ser Asp Asn Glu Lys Ala Asn Leu Gly Phe Phe Gln Ser Val
85 90 95

Asp Pro Arg Val Ile Ser Asp Leu Ile Ile Gly Leu Ser Asp Gly Leu
100 105 110

Thr Val Pro Phe Ala Leu Thr Ala Gly Leu Ser Ser Leu Gly Asp Ala
115 120 125

Lys Leu Val Ile Thr Gly Gly Phe Ala Glu Leu Ile Ser Gly Ala Ile
130 135 140

Ser Met Gly Leu Gly Gly Tyr Leu Gly Ala Lys Ser Glu Ser Asp Tyr
145 150 155 160

Tyr His Ala Glu Val Lys Lys Glu Lys Arg Lys Phe Tyr Asp Asn Ser
165 170 175

Asn Leu Ile Asn Arg Glu Ile Glu Asp Ile Leu Leu Glu Ile Asn Pro
180 185 190

Asn Phe Ser Asp Glu Thr Ile Val Ser Phe Ile Lys Asp Leu Gln Arg
195 200 205

Thr Pro Glu Leu Met Val Asp Phe Ile Ile Arg Tyr Gly Arg Gly Leu
210 215 220

Asp Glu Pro Ala Glu Asn Arg Glu Leu Ile Ser Ala Val Thr Ile Gly
225 230 235 240

Gly Gly Tyr Leu Leu Gly Gly Leu Val Pro Leu Val Pro Tyr Phe Phe
245 250 255

Val Ser Asp Val Gly Thr Gly Leu Ile Tyr Ser Ile Ile Val Met Val
260 265 270

Val Thr Leu Phe Trp Phe Gly Tyr Val Lys Thr Lys Leu Ser Met Gly
275 280 285

Ser Gly Ser Ser Thr Ser Lys Lys Val Thr Glu Gly Val Glu Met Val
290 295 300

Val Val Gly Gly Val Ala Ala Gly Ala Ala Trp Phe Phe Val Lys Leu
 305 310 315 320

Leu Gly

<210> 237
 <211> 347
 <212> PRT
 <213> Beijerinkia indica

<400> 237

Met Lys Ala Leu Val Tyr Arg Gly Pro Gly Gln Lys Leu Val Glu Glu
 1 5 10 15

Arg Gln Lys Pro Glu Leu Lys Glu Pro Gly Asp Ala Ile Val Lys Val
 20 25 30

Thr Lys Thr Thr Ile Cys Gly Thr Asp Leu His Ile Leu Lys Gly Asp
 35 40 45

Val Ala Thr Cys Lys Pro Gly Arg Val Leu Gly His Glu Gly Val Gly
 50 55 60

Val Ile Glu Ser Val Gly Ser Gly Val Thr Ala Phe Gln Pro Gly Asp
 65 70 75 80

Arg Val Leu Ile Ser Cys Ile Ser Ser Cys Gly Lys Cys Ser Phe Cys
 85 90 95

Arg Arg Gly Met Phe Ser His Cys Thr Thr Gly Gly Trp Ile Leu Gly
 100 105 110

Asn Glu Ile Asp Gly Thr Gln Ala Glu Tyr Val Arg Val Pro His Ala
 115 120 125

Asp Thr Ser Leu Tyr Arg Ile Pro Ala Gly Ala Asp Glu Glu Ala Leu
 130 135 140

Val Met Leu Ser Asp Ile Leu Pro Thr Gly Phe Glu Cys Gly Val Leu
 145 150 155 160

Asn Gly Lys Val Ala Pro Gly Ser Ser Val Ala Ile Val Gly Ala Gly
 165 170 175

Pro Val Gly Leu Ala Ala Leu Leu Thr Ala Gln Phe Tyr Ser Pro Ala
 180 185 190

Glu Ile Ile Met Ile Asp Leu Asp Asp Asn Arg Leu Gly Leu Ala Lys
 195 200 205

Gln Phe Gly Ala Thr Arg Thr Val Asn Ser Thr Gly Gly Asn Ala Ala
 Page 254

210

215

220

Ala Glu Val Lys Ala Leu Thr Glu Gly Leu Gly Val Asp Thr Ala Ile
225 230 235 240

Glu Ala Val Gly Ile Pro Ala Thr Phe Glu Leu Cys Gln Asn Ile Val
245 250 255

Ala Pro Gly Gly Thr Ile Ala Asn Val Gly Val His Gly Ser Lys Val
260 265 270

Asp Leu His Leu Glu Ser Leu Trp Ser His Asn Val Thr Ile Thr Thr
275 280 285

Arg Leu Val Asp Thr Ala Thr Thr Pro Met Leu Leu Lys Thr Val Gln
290 295 300

Ser His Lys Leu Asp Pro Ser Arg Leu Ile Thr His Arg Phe Ser Leu
305 310 315 320

Asp Gln Ile Leu Asp Ala Tyr Glu Thr Phe Gly Gln Ala Ala Ser Thr
325 330 335

Gln Ala Leu Lys Val Ile Ile Ser Met Glu Ala
340 345

<210> 238

<211> 719

<212> DNA

<213> Artificial sequence

<220>

<223> UAS(PGK1)-FBA1

<400> 238

aattaccgtc gctcgtgatt tgtttgcaaa aagaacaaaa ctgaaaaaac ccagacacgc 60
tcgacttcct gtcttcctat tgattgcagc ttccaatttc gtcacacaac aaggtcctgc 120
ctacttggct tcacatacgt tgcatacgtc gatatagata ataatgataa tgacagcagg 180
attatcgtaa tacgtaatag ttgaaaatct caaaaatgtg tgggtcatta cgtaaataat 240
gataggaatg ggattcttct atttttcctt tttccattct agcagccgtc gggaaaacgt 300
ggcatcctct ctttcgggct caattggagt cacgctgccg tgagcatcct ctctttccat 360
atctaacaac tgagcacgta accaatggaa aagcatgagc ttagcgttgc tccaaaaaag 420
tattggatgg ttaataccat ttgtctgttc tcttctgact ttgactcctc aaaaaaaaaa 480
aatctacaat caacagatcg cttcaattac gccctcacia aaactttttt ccttcttctt 540
cgcccacgtt aaattttatc cctcatgttg tctaacggat ttctgcactt gatttattat 600
aaaaagacaa agacataata cttctctatc aatttcagtt attgttcttc cttgcgttat 660
tcttctgttc ttctttttct tttgtcatat ataaccataa ccaagtaata catattcaa 719

20110907_CL5178USNA_SeqList.txt

<210> 239
<211> 636
<212> DNA
<213> Artificial sequence

<220>
<223> Promoter UAS(PGK1)-EN02

<400> 239
aattaccgtc gctcgtgatt tgtttgcaaa aagaacaaaa ctgaaaaaac ccagacacgc 60
tcgacttcct gtcttcctat tgattgcagc ttccaatttc gtcacacaac aaggctcctgc 120
ggcgttatgt cactaacgac gtgcaccaac ttgcggaaag tggaatcccg ttccaaaact 180
ggcatccact aattgataca tctacacacc gcacgccttt tttctgaagc ccactttcgt 240
ggactttgcc atatgcaaaa ttcattgaagt gtgataccaa gtcagcatac acctcactag 300
ggtagtttct ttggttgtat tgatcatttg gttcatcgtg gttcattaat tttttttctc 360
cattgctttc tggctttgat cttactatca tttggatttt tgtcgaaggt tgtagaattg 420
tatgtgacaa gtggcaccaa gcatatataa aaaaaaaaaag cattatcttc ctaccagagt 480
tgattgttaa aaacgtattt atagcaaacg caattgtaat taattcttat tttgtatctt 540
ttcttccctt gtctcaatct tttattttta tttatttttt cttttcttag tttctttcat 600
aacaccaagc aactaatact ataacataca ataata 636

<210> 240
<211> 979
<212> DNA
<213> Artificial sequence

<220>
<223> Promoter UAS(FBA1)-PDC1

<400> 240
tgacagcagg attatcgtaa tacgtaatag ttgaaaatct caaaaatgtg tgggtcatta 60
cgtaaataat gataggaatg ggattcttct atttttcctt tttccattct cgcacgccga 120
aatgcatgca agtaacctat tcaaagtaat atctcataca tgtttcatga gggtaacaac 180
atgcgactgg gtgagcatat gttccgctga tgtgatgtgc aagataaaca agcaaggcag 240
aaactaactt cttcttcatg taataaacac accccgcgtt tatttaccta tctctaaact 300
tcaacacctt atatcataac taatatttct tgagataagc aactgcacc cataccttcc 360
ttaaaaacgt agcttccagt ttttggtggt tccggcttcc ttcccgattc cgcccgttaa 420
acgcatatth ttgttgcttg gtggcatttg caaaatgcat aacctatgca tttaaaagat 480
tatgtatgct cttctgactt ttcgtgtgat gaggctcgtg gaaaaaatga ataatttatg 540
aatgtgagaa caattttgtg ttgttacggg attttactat ggaataatca atcaattgag 600
gattttatgc aaatatcggt tgaatattht tccgaccctt tgagtacttt tcttcataat 660
tgcataatat tgtccgctgc ccttttttct gttagacggg gtcttgatct acttgctatc 720
gttcaacacc acctattht ctaactatth tttttttagc tcatttgaat cagcttatgg 780

tgatggcaca	tttttgcata	aacctagctg	tcctcgttga	acataggaaa	aaaaaatata	840
taaacaaggc	tctttcactc	tccttgcaat	cagatttggg	tttgttccct	ttattttcat	900
atttcttgtc	atattccttt	ctcaattatt	attttctact	cataacctca	cgcaaaataa	960
cacagtcaaa	tcaatcaaa					979

<210> 241
 <211> 994
 <212> DNA
 <213> Artificial sequence

<220>
 <223> UAS(PGK1)-PDC1

<400> 241						
aattaccgtc	gctcgtgatt	tgtttgcaaa	aagaacaaaa	ctgaaaaaac	ccagacacgc	60
tcgacttcct	gtcttcctat	tgattgcagc	ttccaatttc	gtcacacaac	aaggctcctgt	120
cgaccgcac	gccgaaatgc	atgcaagtaa	cctattcaaa	gtaatatctc	atacatgttt	180
catgagggtg	acaacatgcg	actgggtgag	catatgttcc	gctgatgtga	tgtgcaagat	240
aaacaagcaa	ggcagaaact	aacttcttct	tcatgtaata	aacacacccc	gcgtttattt	300
acctatctct	aaacttcaac	accttatatc	ataactaata	tttcttgaga	taagcacact	360
gcaccatac	cttccttaaa	aacgtagctt	ccagtttttg	gtggttccgg	cttccttccc	420
gattccgccc	gctaaacgca	tatttttggt	gcctgggtggc	atttgcaaaa	tgcataacct	480
atgcatttaa	aagattatgt	atgctcttct	gacttttcgt	gtgatgaggc	tcgtggaaaa	540
aatgaataat	ttatgaattt	gagaacaatt	ttgtgttggt	acggtatttt	actatggaat	600
aatcaatcaa	ttgaggattt	tatgcaata	tcgtttgaat	atttttccga	ccctttgagt	660
acttttcttc	ataattgcat	aatattgtcc	gctgcccctt	tttctgtag	acggtgtcct	720
gatctacttg	ctatcgttca	acaccacctt	attttctaac	tatttttttt	ttagctcatt	780
tgaatcagct	tatgggtgatg	gcacattttt	gcataaacct	agctgtcctc	gttgaaacata	840
ggaaaaaaaa	atatataaac	aaggctcttt	cactctcctt	gcaatcagat	ttgggtttgt	900
tccctttatt	ttcatatttc	ttgtcatatt	cctttctcaa	ttattatttt	ctactcataa	960
cctcacgcaa	aataacacag	tcaaatcaat	caaa			994

<210> 242
 <211> 731
 <212> DNA
 <213> Artificial sequence

<220>
 <223> UAS(PGK1)-OLE1

<400> 242						
aattaccgtc	gctcgtgatt	tgtttgcaaa	aagaacaaaa	ctgaaaaaac	ccagacacgc	60
tcgacttcct	gtcttcctat	tgattgcagc	ttccaatttc	gtcacacaac	aaggctcctgt	120
cgaccggggc	gcccgtgcaa	aaactaactc	cgagcccggg	catgtcccgg	gttagcgggc	180

20110907_CL5178USNA_SeqList.txt

ccaacaaagg cgcttatctg gtgggcttcc gtagaagaaa aaaagctggt gagcgagcta	240
tttcgggtat cccagccttc tctgcagacc gccccagttg gcttggtctt ggtgctgttc	300
gtagcatca catcgctgt gacaggcaga ggtaataacg gcttaagggt ctcttcgcat	360
agtcggcagc tttctttcgg acgttgaaca ctcaacaaac cttatctagt gccaaccag	420
gtgtgcttct acgagtcttg ctactcaga cacacctatc cctattgtta cggctatggg	480
gatggcacac aaagggtggaa ataatagtag ttaacaatat atgcagcaaa tcatcggctc	540
ctggctcatc gagtcttgca aatcagcata tacatatata tatgggggca gatcttgatt	600
catttattgt tctatttcca tctttcctac ttctgtttcc gtttatattt tgtattacgt	660
agaatagaac atcatagtaa tagatagttg tggatgatcat attataaaca gcactaaaac	720
attacaacaa a	731

<210> 243
 <211> 4306
 <212> DNA
 <213> Artificial sequence

<220>
 <223> Synthetic construct-pUC19-URA3::ilvD-TRX1

<400> 243	
tatttgatc gaggtgtcta gtcttctatt acactaatgc agtttcaggg ttttgaaac	60
cacactgttt aaacagtgtt ccttaatcaa ggatacctct ttttttttcc ttggttccac	120
taattcatcg gttttttttt tggaagacat cttttccaac gaaaagaata tacatatcgt	180
ttaagagaaa ttctccaaat ttgtaaagaa gcggaccag acttaagcct aaccaggcca	240
attcaacaga ctgtcggcaa cttcttgtct ggtctttcca tggtaagtga cagtgcagta	300
ataatatgaa ccaatttatt tttcgttaca taaaaatgct tataaaactt taactaataa	360
ttagagatta aatcgcaaac ggccggccaa tgtggctgtg gtttcagggt ccataaagct	420
tttcaattca tctttttttt ttttgttctt ttttttgatt ccggtttctt tgaaattttt	480
ttgattcggg aatctccgag cagaaggaag aacgaaggaa ggagcacaga cttagattgg	540
tatatatacg catatgtggt gttgaagaaa catgaaattg ccagtatctc ttaacccaac	600
tgcacagaac aaaaacctgc aggaacgaa gataaatcat gtcgaaagct acatataagg	660
aacgtgctgc tactcatcct agtcctgttg ctgccaagct atttaatatc atgcacgaaa	720
agcaaacaaa cttgtgtgct tcattggatg ttcgtaccac caaggaatta ctggagttag	780
ttgaagcatt aggtcccaaa atttgtttac taaaaacaca tgtggatatc ttgactgatt	840
tttccatgga gggcacagtt aagccgctaa aggcattatc cgccaagtac aattttttac	900
tcttcgaaga cagaaaattt gctgacattg gtaatacagt caaattgcag tactctgcgg	960
gtgtatacag aatagcagaa tgggcagaca ttacgaatgc acacggtgtg gtggggccag	1020
gtattgttag cggtttgaag caggcggcgg aagaagtaac aaaggaacct agaggccttt	1080
tgatgttagc agaattgtca tgcaagggtc ccctagctac tggagaatat actaagggta	1140

20110907_CL5178USNA_SeqList.txt

ctgttgacat	tgccaagagc	gacaaagatt	ttgttatcgg	ctttattgct	caaagagaca	1200
tgggtggaag	agatgaaggt	tacgattggg	tgattatgac	acccgggtgtg	ggtttagatg	1260
acaagggaga	cgcattgggt	caacagtata	gaaccgtgga	tgatgtgggc	tctacaggat	1320
ctgacattat	tattgttggg	agaggactat	ttgcaaaggg	aagggatgct	aaggtagagg	1380
gtgaacgtta	cagaaaagca	ggctgggaag	catatttgag	aagatgcggc	cagcaaaact	1440
aaaaaactgt	attataagta	aatgcatgta	tactaaactc	acaaattaga	gcttcaattt	1500
aattatatca	gttattaccc	gggaatctcg	gtcgtaatga	tttctataat	gacgaaaaaa	1560
aaaaaattgg	aaagaaaaag	cttcatggcc	ttgcggccgc	ttaattaatc	tagagtgcac	1620
ctgcaggcat	gcaagcttgg	cgtaatcatg	gtcatagctg	tttcctgtgt	gaaattgtta	1680
tccgctcaca	attccacaca	acatacgagc	cggaagcata	aagtgtaaag	cctgggggtgc	1740
ctaattgagt	agctaactca	cattaattgc	gttgcgctca	ctgcccgtt	tccagtcggg	1800
aaacctgtcg	tgccagctgc	attaatgaat	cggccaacgc	gcggggagag	gcggtttgcg	1860
tattgggcgc	tcttccgctt	cctcgctcac	tgactcgctg	cgctcggtcg	ttcggctgcg	1920
gcgagcggta	tcagctcact	caaaggcggg	aatacggtta	tccacagaat	caggggataa	1980
cgcaggaaaag	aacatgtgag	caaaaggcca	gcaaaaggcc	aggaaccgta	aaaaggccgc	2040
gttgctggcg	ttttccata	ggctccgccc	ccctgacgag	catcacaaaa	atcgacgctc	2100
aagtcagagg	tggcgaaacc	cgacaggact	ataaagatac	caggcgtttc	cccctggaag	2160
ctccctcgtg	cgctctcctg	ttccgaccct	gccgcttacc	ggataacctgt	ccgcctttct	2220
cccttcggga	agcgtggcgc	tttctcatag	ctcacgctgt	aggtatctca	gttcgggtgta	2280
ggctgttcgc	tccaagctgg	gctgtgtgca	cgaaccccc	gttcagcccg	accgctgcgc	2340
cttatccggg	aactatcgtc	ttgagtccaa	cccggtaaga	cacgacttat	cgccactggc	2400
agcagccact	ggtaacagga	ttagcagagc	gaggtatgta	ggcgggtgcta	cagagttctt	2460
gaagtgggtg	cctaactacg	gctacactag	aaggacagta	tttggtatct	gcgctctgct	2520
gaagccagtt	accttcggaa	aaagagttgg	tagctcttga	tccggcaaac	aaaccaccgc	2580
tggtagcggg	ggtttttttg	tttgcaagca	gcagattacg	cgcagaaaaa	aaggatctca	2640
agaagatcct	ttgatctttt	ctacgggggc	tgacgctcag	tggaacgaaa	actcacgtta	2700
agggattttg	gtcatgagat	tatcaaaaag	gatcttcacc	tagatccttt	taaattaaaa	2760
atgaagtttt	aaatcaatct	aaagtatata	tgagtaaact	tggctctgaca	gttaccaatg	2820
cttaatcagt	gaggcaccta	tctcagcgat	ctgtctattt	cgttcatcca	tagttgcctg	2880
actccccgtc	gtgtagataa	ctacgatacg	ggagggttta	ccatctggcc	ccagtgtgc	2940
aatgataccg	cgagaccac	gctcaccggc	tccagattta	tcagcaataa	accagccagc	3000
cgggaagggc	gagcgagaa	gtggctcctg	aactttatcc	gcctccatcc	agtcatttaa	3060
ttgttgccgg	gaagctagag	taagtagttc	gccagttaat	agtttgcgca	acgttgttgc	3120
cattgctaca	ggcatcgtgg	tgtcacgctc	gtcgtttggg	atggcttcat	tcagctccgg	3180

20110907_CL5178USNA_SeqList.txt

```

ttcccaacga tcaaggcgag ttacatgata ccccatgttg tgcaaaaaag cggttagctc 3240
cttcggtcct ccgatcggtg tcagaagtaa gttggccgca gtgttatcac tcatggttat 3300
ggcagcactg cataattctc ttactgtcat gccatccgta agatgctttt ctgtgactgg 3360
tgagtactca accaagtcac tctgagaata gtgtatgcgg cgaccgagtt gctcttgccc 3420
ggcgtcaata cgggataata ccgcgccaca tagcagaact ttaaaagtgc tcatcattgg 3480
aaaacgttct tcggggcgaa aactctcaag gatcttaccg ctgttgagat ccagttcgat 3540
gtaaccact cgtgcacca actgatcttc agcatctttt actttcacca gcgtttctgg 3600
gtgagcaaaa acaggaaggc aaaatgccgc aaaaaagga ataagggcga cacggaaatg 3660
ttgaatactc atactcttcc tttttcaata ttattgaagc atttatcagg gttattgtct 3720
catgagcggg tacatatttg aatgtattta gaaaaataaa caaatagggg ttccgcgcac 3780
atttccccga aaagtgccac ctgacgtcta agaaaccatt attatcatga cattaaccta 3840
taaaaatagg cgtatcacga ggccctttcg tctcgcgcgt ttcggtgatg acggtgaaaa 3900
cctctgacac atgcagctcc cggagacggt cacagcttgt ctgtaagcgg atgccgggag 3960
cagacaagcc cgtcagggcg cgtcagcggg tgttggcggg tgtcggggct ggcttaacta 4020
tgcggcatca gagcagattg tactgagagt gcacatatg cgggtgtgaaa taccgcacag 4080
atgcgtaagg agaaaatacc gcatcaggcg ccattcgcca ttcaggctgc gcaactgttg 4140
ggaagggcga tcggtgcggg cctcttcgct attacgccag ctggcgaaag ggggatgtgc 4200
tgcaaggcga ttaagttggg taacgccagg gttttccag tcacgacgtt gtaaaacgac 4260
ggccagtga ttcgagctcg gtaccgggg atccggcgcg ccgttt 4306

```

<210> 244
<211> 1772
<212> DNA
<213> Artificial sequence

<220>
<223> chimeric construct

<220>
<221> misc_feature
<222> (613)..(616)
<223> Start codon

<220>
<221> misc_feature
<222> (1420)..(1422)
<223> Stop codon

```

<400> 244
cctagggatt cataaccatt ttctcaatcg aattacacag aacacaccgt acaaacctct 60
ctatcataac tacttaatag tcacacacgt actcgtctaa atacacatca tcgtcctaca 120
agttcatcaa agtgttggac agacaactat accagcatgg atctcttgta tcggttcttt 180
tctcccgctc tctcgcaata acaatgaaca ctgggtcaat catagcctac acaggtgaac 240
agagtagcgt ttatacaggg tttatacggg gattcctacg gcaaaaattt ttcatcttcta 300

```

20110907_CL5178USNA_SeqList.txt

aaaaaaaaa	gaaaaatttt	tctttccaac	gctagaagga	aaagaaaaat	ctaattaaat	360
tgatttggtg	atcttctgag	agttcccttt	ttcatatata	gaattttgaa	tataaaagga	420
gatcgaaaaa	atcttctat	tcaatctgtt	ttctggtttt	atctgatagt	ttttttgtgt	480
attattatta	tggattagta	ctggtttata	tgggtttttc	tgtataactt	ctttttatct	540
tagtttggtt	aatcttattt	tgagttacat	tatagttccc	taactgcaag	agaagtaaca	600
ttaaaactcg	agatgggtaa	ggaaaagact	cacgtttcga	ggccgcgatt	aaattccaac	660
atggatgctg	atctatatgg	gtataaatgg	gctcgcgata	atgtcgggca	atcaggtgcg	720
acaatctatc	gattgtatgg	gaagcccgat	gcgccagagt	tgtttctgaa	acatggcaaa	780
ggtagcgttg	ccaatgatgt	tacagatgag	atggtcagac	taaactggct	gacggaatct	840
atgcctcttc	cgaccatcaa	gcattttatc	cgtactcctg	atgatgcatg	gttactcacc	900
actgcgatcc	ccggcaaaac	agcattccag	gtattagaag	aatatcctga	ttcaggtgaa	960
aatattgttg	atgcgctggc	agtgttcctg	cgccggttgc	attcgattcc	tgtttgtaat	1020
tgtcctttta	acagcgatcg	cgtatttcgt	ctcgtcagg	cgcaatcacg	aatgaataac	1080
ggtttggttg	atgcgagtga	ttttgatgac	gagcgtaatg	gctggcctgt	tgaacaagtc	1140
tggaaagaaa	tgcataagct	tttgccattc	tcaccggatt	cagtcgtcac	tcattggtgat	1200
ttctcacttg	ataaccttat	ttttgacgag	gggaaattaa	taggttggtat	tgatgttgga	1260
cgagtcggaa	tcgcgaccg	ataccaggat	cttgccatcc	tatggaactg	cctcggtgag	1320
ttttctcctt	cattacagaa	acggcttttt	caaaaatatg	gtattgataa	tcctgatatg	1380
aataaattgc	agtttcattt	gatgctcgat	gagtttttct	aagtttaact	tgatactact	1440
agattttttc	tcttcattta	taaaattttt	ggttataatt	gaagcttttag	aagtatgaaa	1500
aaatcctttt	ttttcattct	ttgcaaccaa	aataagaagc	ttctttttatt	cattgaaatg	1560
atgaatataa	acctaacaaa	agaaaaagac	tcgaatatca	aacattaaaa	aaaaataaaa	1620
gaggttatct	gttttcccat	ttagttggag	tttgcatctt	ctaatagata	gaactctcaa	1680
ttaatgtgga	tttagtttct	ctgttcgttt	ttttttgttt	tgtttctact	gtattttacat	1740
ttctatttag	tatttagtta	ttcatataat	ct			1772

<210> 245

<211> 1854

<212> DNA

<213> Artificial sequence

<220>

<223> Synthetic construct-HincII-KpnI fragment

<400> 245

gactctagag	gatccccac	cttggttaac	tcgttgatct	atcactggat	aacttcgtat	60
agcatacatt	atacgaagtt	atctagggat	tcataacat	tttctcaatc	gaattacaca	120
gaacacaccg	tacaaacctc	tctatcataa	ctacttaata	gtcacacacg	tactcgtcta	180
aatacacatc	atcgtcctac	aagttcatca	aagtgttgga	cagacaacta	taccagcatg	240

20110907_CL5178USNA_SeqList.txt

gatctcttgt atcggttctt ttctcccgt ctctcgcaat aacaatgaac actgggtcaa	300
tcatagccta cacaggtgaa cagagtagcg tttatacagg gtttatacgg tgattcctac	360
ggcaaaaatt tttcatttct aaaaaaaaaa agaaaaatth ttctttccaa cgctagaagg	420
aaaagaaaaa tctaattaaa ttgatttggt gattttctga gagttccctt tttcatatat	480
cgaattttga atataaaagg agatcgaaaa aatttttcta ttcaatctgt tttctggttt	540
tatttgatag tttttttgtg tattattatt atggattagt actggtttat atgggttttt	600
ctgtataact tctttttatt ttagtttggt taatcttatt ttgagttaca ttatagttcc	660
ctaactgcaa gagaagtaac attaaaactc gagatgggta aggaaaagac tcacgtttcg	720
aggccgcgat taaattccaa catggatgct gatttatatg ggtataaatg ggctcgcgat	780
aatgtcgggc aatcagggtgc gacaatctat cgattgtatg ggaagcccga tgcgccagag	840
ttgtttctga aacatggcaa aggtagcggt gccaatgatg ttacagatga gatggtcaga	900
ctaaactggc tgacggaatt tatgcctctt ccgaccatca agcattttat ccgtactcct	960
gatgatgcat ggttactcac cactgcgatc cccggcaaaa cagcattcca ggtattagaa	1020
gaatatcctg attcagggtga aaatattggt gatgcgctgg cagtgttcct gcgccggttg	1080
cattcgattc ctgtttgtaa ttgtcctttt aacagcgatc gcgtatttcg tctcgctcag	1140
gcgcaatcac gaatgaataa cggtttggtt gatgcgagtg attttgatga cgagcgtaat	1200
ggctggcctg ttgaacaagt ctggaaagaa atgcataagc ttttgccatt ctcaccggat	1260
tcagtcgtca ctcatggtga tttctcactt gataacctta tttttgacga ggggaaatta	1320
ataggttgta ttgatgttg acgagtcgga atcgcagacc gataccagga tcttgccatc	1380
ctatggaact gcctcgggtga gttttctcct tcattacaga aacggctttt tcaaaaatat	1440
ggtattgata atcctgatat gaataaattg cagtttcatt tgatgctcga tgagtttttc	1500
taagtttaac ttgatactac tagatttttt ctcttcattt ataaaattht tggttataat	1560
tgaagcttta gaagtatgaa aaaatccttt tttttcattc tttgcaacca aaataagaag	1620
cttcttttat tcattgaaat gatgaatata aacctaacaa aagaaaaaga ctcgaatatc	1680
aaacattaaa aaaaaataaa agaggttatc tgttttccca tttagttgga gtttgcatth	1740
tctaatagat agaactctca attaatgtgg atttagtttc tctgttcgat aacttcgtat	1800
agcatacatt atacgaagtt atctgaacat tagaatacgt aatccgcaat gcgg	1854

<210> 246
 <211> 11017
 <212> DNA
 <213> Artificial sequence

<220>
 <223> Synthetic construct-pRS426::GPD-xpk1+ADH-eutD

<400> 246	
gatccccgg gctgcaggaa ttcgatatca agcttatcga taccgtcgac ctcgaggggg	60
ggccccgtac ccaattcgcc ctatagttag tcgtattacg cgcgctcact ggccgctcgtt	120

20110907_CL5178USNA_SeqList.txt

ttacaacgtc	gtgactggga	aaaccctggc	gttaccacaac	ttaatcgccct	tgcagcacat	180
ccccctttcg	ccagctggcg	taatagcgaa	gaggcccgcga	ccgatcgccc	ttcccaacag	240
ttgcgcagcc	tgaatggcga	atggcgcgac	gcgccctgta	gcggcgccatt	aagcgcggcg	300
ggtgtggtgg	ttacgcgcag	cgtgaccgct	acacttgcca	gcgccctagc	gcccgtcctt	360
ttcgctttct	tcccttcctt	tctcgccacg	ttcgccgggt	ttccccgtca	agctctaaat	420
cgggggctcc	ctttaggggt	ccgatttagt	gctttacggc	acctcgaccc	caaaaaactt	480
gattaggggtg	atgggttcacg	tagtgggcca	tcgccctgat	agacggtttt	tcgccctttg	540
acgttgaggat	ccacgttctt	taatagtggga	ctcttgttcc	aaactggaac	aacactcaac	600
cctatctcgg	tctattcttt	tgatttataa	gggattttgc	cgatttcggc	ctattggtta	660
aaaaatgagc	tgatttaaca	aaaatttaac	gcgaatttta	acaaaatatt	aacgtttaca	720
atttcctgat	gcgggtat	ctccttacgc	atctgtgcgg	tatttcacac	cgcataggggt	780
aataactgat	ataattaaat	tgaagctcta	atgtgtgagt	ttagtataca	tgcatttact	840
tataatacag	ttttttagtt	ttgctggccg	catcttctca	aatatgcttc	ccagcctgct	900
tttctgtaac	gttcaccctc	taccttagca	tcccttcctt	ttgcaaatag	tcctcttcca	960
acaataataa	tgtcagatcc	tgtagagacc	acatcatcca	cggttctata	ctgttgaccc	1020
aatgcgtctc	ccttgtcatc	taaaccaca	ccgggtgtca	taatcaacca	atcgtaacct	1080
tcctctcttc	cacctatgtc	tctttgagca	ataaagccga	taacaaaatc	tttgtcgctc	1140
ttcgcaatgt	caacagtacc	cttagtata	tctccagtag	atagggagcc	cttgcatgac	1200
aattctgcta	acatcaaaag	gcctctaggt	tcctttgtta	cttcttctgc	cgctgtcttc	1260
aaaccgctaa	caatacctgg	gccaccaca	ccgtgtgcat	tcgtaatgtc	tgccattctt	1320
gctattctgt	atacaccgcg	agagtactgc	aatttgactg	tattaccaat	gtcagcaaat	1380
tttctgtctt	cgaagagtaa	aaaattgtac	ttggcggata	atgccttttag	cggcttaact	1440
gtgccctcca	tggaaaaatc	agtcaagata	tccacatgtg	tttttagtaa	acaaattttg	1500
ggacctaatg	cttcaactaa	ctccagtaat	tccttggtgg	tacgaacatc	caatgaagca	1560
cacaagtttg	tttgcttttc	gtgcatgata	ttaaatagct	tggcagcaac	aggactagga	1620
tgagtagcag	cacgttcctt	atatgtagct	ttcgacatga	tttatcttcg	tttcctgcag	1680
gtttttgttc	tgtgcagttg	ggttaagaat	actgggcaat	ttcatgtttc	ttcaacacta	1740
catatgcgta	tatatacca	tctaagtctg	tgctccttcc	ttcgttcttc	cttctgttcg	1800
gagattaccg	aatcaaaaaa	atttcaaaga	aaccgaaatc	aaaaaaaaga	ataaaaaaaa	1860
aatgatgaat	tgaattgaaa	agctgtggta	tggtgcactc	tcagtacaat	ctgctctgat	1920
gccgcatagt	taagccagcc	ccgacaccgc	ccaacaccgc	ctgacgcgcc	ctgacgggct	1980
tgtctgctcc	cggcatccgc	ttacagacaa	gctgtgaccg	tctccgggag	ctgcatgtgt	2040
cagaggtttt	caccgtcatc	accgaaacgc	gcgagacgaa	agggcctcgt	gatacgccta	2100
tttttatagg	ttaatgtcat	gataataatg	gtttcttagt	atgatccaat	atcaaaggaa	2160

20110907_CL5178USNA_SeqList.txt

atgatagcat	tgaaggatga	gactaatcca	attgaggagt	ggcagcatat	agaacagcta	2220
aagggtagtg	ctgaaggaag	catacgatac	cccgcattga	atgggataat	atcacaggag	2280
gtactagact	acctttcatc	ctacataaat	agacgcatat	aagtacgcat	ttaagcataa	2340
acacgcacta	tgccgttctt	ctcatgtata	tatatataca	ggcaacacgc	agatataggt	2400
gcgacgtgaa	cagttagctg	tatgtgcgca	gctcggttg	cattttcgga	agcgctcggt	2460
ttcggaaacg	ctttgaagtt	cctattccga	agttcctatt	ctctagaaag	tataggaact	2520
tcagagcgct	tttgaaaacc	aaaagcgctc	tgaagacgca	ctttcaaaaa	acaaaaaacg	2580
caccggactg	taacgagcta	ctaaaatatt	gcgaataccg	cttcacaaaa	cattgctcaa	2640
aagtatctct	ttgctatata	tctctgtgct	atatccctat	ataacctacc	catccacctt	2700
tcgctccttg	aacttgcac	taaactcgac	ctctacattt	tttatgttta	tctctagtat	2760
tactctttag	acaaaaaaat	tgtagtaaga	actattcata	gagtgaatcg	aaaacaatac	2820
gaaaatgtaa	acatttccta	tacgtagtat	atagagacaa	aatagaagaa	accgttcata	2880
attttctgac	caatgaagaa	tcatcaacgc	tatcactttc	tgttcacaaa	gtatgcgcaa	2940
tccacatcgg	tatagaatat	aatcggggat	gcctttatct	tgaaaaaatg	caccgcgacg	3000
ttcgctagta	atcagtaaac	gcgggaagtg	gagtcaggct	ttttttatgg	aagagaaaat	3060
agacaccaaa	gtagccttct	tctaacctta	acggacctac	agtgcaaaaa	gttatcaaga	3120
gactgcatta	tagagcgac	aaaggagaaa	aaaagtaatc	taagatgctt	tgttagaaaa	3180
atagcgctct	cgggatgcat	ttttgtagaa	caaaaaagaa	gtatagattc	tttgttggta	3240
aaatagcgct	ctcgcggttg	atctctgttc	tgtaaaaatg	cagctcagat	tctttgtttg	3300
aaaaattagc	gctctcgctg	tgcatttttg	ttttacaaaa	atgaagcaca	gattcttcgt	3360
tggtaaaata	gcgctttcgc	gttgcatctt	tgttctgtaa	aatgcagct	cagattcttt	3420
gtttgaaaaa	ttagcgctct	cgcgttgcat	ttttgttcta	caaaatgaag	cacagatgct	3480
tcgttcaggt	ggcacttttc	ggggaaatgt	gcgcggaacc	cctatttggt	tatttttcta	3540
aatacattca	aatatgtatc	cgctcatgag	acaataaccc	tgataaatgc	ttcaataata	3600
ttgaaaaagg	aagagtatga	gtattcaaca	tttccgtgtc	gcccttattc	ccttttttgc	3660
ggcattttgc	cttcctgttt	ttgctcacc	agaaacgctg	gtgaaagtaa	aagatgctga	3720
agatcagttg	ggtgcacgag	tgggttacat	cgaactggat	ctcaacagcg	gtaagatcct	3780
tgagagtttt	cgccccgaag	aacgttttcc	aatgatgagc	acttttaaag	ttctgctatg	3840
tggcgcggtg	ttatcccgtg	ttgacgccgg	gcaagagcaa	ctcggtcgcc	gcatacacta	3900
ttctcagaat	gacttggttg	agtactcacc	agtcacagaa	aagcatctta	cggatggcat	3960
gacagtaaga	gaattatgca	gtgctgccat	aaccatgagt	gataaactg	cggccaactt	4020
acttctgaca	acgatcggag	gaccgaagga	gctaaccgct	tttttgca	acatggggga	4080
tcatgtaact	cgccttgatc	gttggaacc	ggagctgaat	gaagccatac	caaacgacga	4140
gcgtgacacc	acgatgcctg	tagcaatggc	aacaacgttg	cgaaaactat	taactggcga	4200

20110907_CL5178USNA_SeqList.txt

actacttact	ctagcttccc	ggcaacaatt	aatagactgg	atggaggcgg	ataaagttgc	4260
aggaccactt	ctgcgctcgg	cccttcgggc	tggctggttt	attgctgata	aatctggagc	4320
cggtgagcgt	gggtctcgcg	gtatcattgc	agcactgggg	ccagatggta	agccctcccg	4380
tatcgtagtt	atctacacga	cggggagtca	ggcaactatg	gatgaacgaa	atagacagat	4440
cgctgagata	ggtgcctcac	tgattaagca	ttggtaactg	tcagaccaag	tttactcata	4500
tatactttag	attgatttaa	aacttcattt	ttaatttaaa	aggatctagg	tgaagatcct	4560
ttttgataat	ctcatgacca	aaatccctta	acgtgagttt	tcgttccact	gagcgtcaga	4620
ccccgtagaa	aagatcaaag	gatcttcttg	agatcctttt	tttctgcgcg	taatctgctg	4680
cttgcaaaca	aaaaaaccac	cgctaccagc	ggtggtttgt	ttgccggatc	aagagctacc	4740
aactcttttt	ccgaaggtaa	ctggcttcag	cagagcgcag	ataccaaata	ctgtccttct	4800
agtgtagccg	tagttaggcc	accacttcaa	gaactctgta	gcaccgccta	catacctcgc	4860
tctgctaadc	ctgttaccag	tggctgctgc	cagtggcgat	aagtcgtgtc	ttaccggggt	4920
ggactcaaga	cgatagttac	cggataaggc	gcagcggtcg	ggctgaacgg	ggggttcgtg	4980
cacacagccc	agcttggagc	gaacgacct	caccgaactg	agatacctac	agcgtgagct	5040
atgagaaagc	gccacgcttc	ccgaaggag	aaaggcggac	aggtatccgg	taagcggcag	5100
ggtcggaaac	ggagagcgca	cgaggagct	tccaggggga	aacgcctggt	atctttatag	5160
tcctgtcggg	tttcgccacc	tctgacttga	gcgtcgattt	ttgtgatgct	cgtcaggggg	5220
gcggagccta	tggaaaaacg	ccagcaacgc	ggccttttta	cggttcctgg	ccttttgctg	5280
gccttttgct	cacatgttct	ttcctgcgtt	atcccctgat	tctgtggata	accgtattac	5340
cgcctttgag	tgagctgata	ccgctcgccg	cagccgaacg	accgagcgca	gcgagtcagt	5400
gagcgaggaa	gcggaagagc	gccaatacgc	caaaccgcct	ctccccgcgc	gttggccgat	5460
tcattaatgc	agctggcacg	acaggtttcc	cgactggaaa	gcgggcagtg	agcgcaacgc	5520
aattaatgtg	agttacctca	ctcattaggc	accccaggct	ttacacttta	tgcttccggc	5580
tcctatgttg	tgtggaattg	tgagcggata	acaatttcac	acaggaaaca	gctatgacca	5640
tgattacgcc	aagcgcgcaa	ttaaccctca	ctaaagggaa	caaaagctgg	agctccaccg	5700
cggatagatc	tagttcgagt	ttatcattat	caatactgcc	atttcaaaga	atacgtaaata	5760
aattaatagt	agtgattttc	ctaactttat	ttagtcaaaa	aattagcctt	ttaattctgc	5820
tgaacccgt	acatgcccaa	aatagggggc	gggttacaca	gaatatataa	catcgtaggt	5880
gtctgggtga	acagtttatt	cctggcatcc	actaaatata	atggagcccg	ctttttaagc	5940
tggcatccag	aaaaaaaaag	aatcccagca	ccaaaatatt	gttttcttca	ccaaccatca	6000
gttcataaggt	ccattctctt	agcgcaacta	cagagaacag	gggcacaaac	aggcaaaaaa	6060
cgggcacaac	ctcaatggag	tgatgcaacc	tgccctggagt	aaatgatgac	acaaggcaat	6120
tgaccacgcg	atgtatctat	ctcattttct	tacaccttct	attaccttct	gctctctctg	6180
atttggaaaa	agctgaaaaa	aaagggtgaa	accagttccc	tgaaattatt	cccctacttg	6240

20110907_CL5178USNA_SeqList.txt

actaataagt	atataaagac	ggtaggtatt	gatttgtaatt	ctgtaaatct	atttcttaaa	6300
cttcttaaat	tctactttta	tagttagtct	tttttttagt	tttaaaacac	caagaactta	6360
gtttcgaata	aacacacata	aacgctgagg	atgacaacag	attactcatc	accagcatat	6420
ttgcaaaaag	ttgataagta	ctggcgtgct	gccaaactact	tatcagttgg	tcaactttat	6480
ttaaaagata	atccactatt	acaacggcca	ttgaaggcca	gtgacgttaa	ggttcatcca	6540
attggtcact	gggggacgat	tgccggtcaa	aactttatct	atgctcatct	taaccggggtc	6600
atcaacaagt	acggtttgaa	gatgttctac	gttgaagggtc	cagggtcatgg	tggtcaagtg	6660
atggtttcaa	actcttacct	tgacggtact	tacaccgata	tttatccaga	aattacgcag	6720
gatgttgaag	ggatgcaaaa	gctcttcaag	caattctcat	tcccagggtgg	ggttgcttcc	6780
catgcggcac	ctgaaacacc	cggttcaatc	cacgaagggtg	gcgaacttgg	ttactcaatt	6840
tcacacgggg	ttggggcaat	tcttgacaat	cctgacgaaa	tcgccgcggt	tgttgttggt	6900
gatggggaat	ccgaaacggg	tccattagca	acttcatggc	aatcaacgaa	gttcattaac	6960
ccaatcaacg	acggggctgt	tttaccaatc	ttgaacttaa	atggttttta	gatttctaata	7020
ccaacgattt	ttggctcgac	ttctgatgct	aagattaagg	aatacttcga	aagcatgaat	7080
tggaaccaa	tcttcgttga	agggtgacgat	cctgaaaagg	ttcaccagc	cttagctaag	7140
gcatgggatg	aagccgttga	aaagatcaag	gcaatccaga	agcatgctcg	cgaaaataac	7200
gatgcaacat	tgccagtatg	gccaatgatc	gtcttccgcg	cacctaaggg	ctggactggt	7260
ccgaagtcat	gggacggtga	taagatcgaa	ggttcattcc	gtgctcatca	aattccgatt	7320
cctgttgatc	aaaatgacat	ggaacatgcg	gatgcttttag	ttgattggct	cgaatcatat	7380
caacaaaaag	aactcttcaa	tgaagatggc	tctttgaagg	atgatattaa	agaaattatt	7440
cctactgggg	acagtcggat	ggctgctaac	ccaatcacca	atgggtgggg	cgatccgaaa	7500
gccttgaact	taccaaactt	ccgtgattat	gcggctgata	cgtccaaaga	aggcgcgaat	7560
gttaagcaag	atatgatcgt	ttggtcagac	tatttgcggg	atgtcatcaa	gaaaaatcct	7620
gataacttcc	ggttgttcgg	acctgatgaa	accatgtcta	accgtttata	tggtgtcttc	7680
gaaaccacta	atcgtcaatg	gatggaagac	attcatccag	atagtgacca	atatgaagca	7740
ccagctggcc	gggtcttaga	tgctcagtta	tctgaacacc	aagctgaagg	ttggttagaa	7800
ggttacgtct	taactggacg	tcatgggtta	tttgccagtt	atgaagcctt	cctacgcgtt	7860
gtggactcaa	tgttgacgca	acacttcaag	tggttacgta	aagccaatga	acttgattgg	7920
cgtaaaaagt	accatcact	taacattatc	gcggcttcaa	ctgtattcca	acaagaccat	7980
aatgggtata	cccaccaaga	tccagggtgca	ttactcatt	tggccgaaaa	gaaaccagaa	8040
tacattcgtg	aatatttacc	agccgatgcc	aacacgttat	tagctgtcgg	tgacgtcatt	8100
ttccggagcc	aagaaaagat	caactacgtg	gttacgtcaa	aacaccacg	tcaacaatgg	8160
ttcagcattg	aagaagctaa	gcaattagtt	gacaatgggtc	ttggtatcat	tgattgggca	8220
agtacggacc	aaggtagcga	accagacatt	gtctttgcag	ctgctgggac	ggaaccaacg	8280

20110907_CL5178USNA_SeqList.txt

cttgaaacgt	tggctgccat	ccaattacta	cacgacagtt	tcccagagat	gaagattcgt	8340
ttcgtgaacg	tggctgcacat	cttgaagtta	cgtagtcctg	aaaaggatcc	gcggggccttg	8400
tcagatgctg	agtttgacca	ttactttact	aaggacaaac	cagtgggtctt	tgctttccac	8460
ggttacgaag	acttagttcg	tgacatcttc	tttgatcgtc	acaaccataa	cttatacgtc	8520
cacggttacc	gtgaaaatgg	tgatattacc	acaccattcg	acgtacgggt	catgaaccag	8580
atggaccgct	tcgacttagc	taagtcggca	attgcgggcg	aaccagcaat	ggaaaacact	8640
ggtgcggcct	tcgttcaatc	catggataat	atgcttgcta	aacacaatgc	ctatatccgg	8700
gatgccggaa	ctgacttgcc	agaagttaat	gattggcaat	ggaaggggtt	aaaataatta	8760
attaatcatg	taattagtta	tgtcacgctt	acattcacgc	cctcctccca	catccgctct	8820
aaccgaaaag	gaaggagtta	gacaacctga	agtctaggtc	cctatattt	ttttttaata	8880
gttatgttag	tattaagaac	gttatattata	tttcaaattt	ttcttttttt	tctgtacaaa	8940
cgcgtgtacg	catgtaacat	tatactgaaa	accttgcttg	agaagggttt	gggacgctcg	9000
aaggcttta	tttgcggcg	gccgctctag	aactagtacc	acaggtgttg	tcctctgagg	9060
acataaaata	cacaccgaga	ttcatcaact	cattgctgga	gttagcatat	ctacaattgg	9120
gtgaaatggg	gagcgatttg	caggcatttg	ctcggcatgc	cggtagaggt	gtggtcaata	9180
agagcgacct	catgctatac	ctgagaaagc	aacctgacct	acaggaaaga	gttactcaag	9240
aataagaatt	ttcgttttaa	aacctaaagag	tcacttttaa	atttgtatac	acttatTTTT	9300
tttataactt	atttaataat	aaaaatcata	aatcataaga	aattcgctta	ctcttaatta	9360
attaagctaa	tccttgggct	gctgtaataa	tcgcaacctt	ataaacgtct	tcttactgc	9420
atccacgtga	caagtcggag	accggcttgt	tcaggccttg	caagacagga	cccaccgctt	9480
caaaatgacc	aaatcgttgc	gcaatcttat	agccaatatt	accagactga	agctctggaa	9540
atacaaagac	attggcatga	ccagctactt	tggaaccagg	agccttttgc	aaaccaactt	9600
tttcaacgaa	ggccgcgtca	aattgaagtt	caccatcgat	agccaattcc	ggttcagcag	9660
cttgcgctt	ggccgttgct	tcttgactt	tagtgaccat	ttcaccctta	gccgaacctt	9720
tagttgagaa	gctgagcatc	gcaactttcg	ggtcaatatc	gaagacctta	gcagtagccg	9780
cactctgagt	ggcaatttcc	gctaacgtat	cggcatcggg	atcaatattg	atggcacagt	9840
cagcaaagac	gtagcgttcc	tcaccctttt	gcatgataaa	tgcacccgag	attcgggtgtg	9900
aaccgggctt	ggtcttaata	atttgtaacg	ctggccgtac	cgtatcacca	gttggtatgga	9960
ttgcacctga	aaccatccca	tccgctttgc	ccatataaac	gagcatcgtg	caaagtagt	10020
tttcatcttc	cagcatttta	gccgcttggt	ctggcgattt	cttacctttc	cgccgttcaa	10080
cgagggcatc	aagcattgct	tgcttatctt	cagccgggta	tgtcgcagga	tcaaggactt	10140
gaacgcctgt	taaatccgca	ttcaaatcgt	tagccacagc	ctgaactttg	tccgttgcac	10200
ctaaaacaat	cggcttaacc	aagccgtctg	cagctaatac	cgctgccgca	ccgacaattc	10260
ggggttcagt	tccttcaggg	aaaacaattg	tttgatcttt	accagtaatt	ttttgtgcta	10320

20110907_CL5178USNA_SeqList.txt

```

atgactcaaa taaatccatc ctcagcgaga tagttgattg tatgcttggt atagcttgaa 10380
atattgtgca gaaaaagaaa caaggaagaa aggggaacgag aacaatgacg aggaaacaaa 10440
agattaataa ttgcaggtct atttatactt gatagcaaga cagcaaactt ttttttattt 10500
caaattcaag taactggaag gaaggccgta taccgttgct cattagagag tagtgtgcgt 10560
gaatgaagga agggaaaaagt ttcgtgtgct tcgagatacc cctcatcagc tctggaacaa 10620
cgacatctgt tggtgctgtc tttgtcgta attttttcct ttagtgtctt ccatcatttt 10680
tttgtcattg cggatatggt gagacaacaa cgggggagag agaaaagaaa aaaaaagaaa 10740
agaagttgca tgcgcctatt attacttcaa tagatggcaa atggaaaaag ggtagtgaaa 10800
cttcgatatg atgatggcta tcaagtctag ggctacagta ttagttcggt atgtaccacc 10860
atcaatgagg cagtgttaatt ggtgtagtct tgtttagccc attatgtctt gtctggtatc 10920
tgttctattg tatatctccc ctccgccacc tacatgttag ggagaccaac gaaggtatta 10980
taggaatccc gatgtatggg tttggttgcc agaaaag 11017

```

<210> 247
 <211> 548
 <212> PRT
 <213> *Listeria grayi*

<400> 247

Met Tyr Thr Val Gly Gln Tyr Leu Val Asp Arg Leu Glu Glu Ile Gly
1 5 10 15

Ile Asp Lys Val Phe Gly Val Pro Gly Asp Tyr Asn Leu Thr Phe Leu
20 25 30

Asp Tyr Ile Gln Asn His Glu Gly Leu Ser Trp Gln Gly Asn Thr Asn
35 40 45

Glu Leu Asn Ala Ala Tyr Ala Ala Asp Gly Tyr Ala Arg Glu Arg Gly
50 55 60

Val Ser Ala Leu Val Thr Thr Phe Gly Val Gly Glu Leu Ser Ala Ile
65 70 75 80

Asn Gly Thr Ala Gly Ser Phe Ala Glu Gln Val Pro Val Ile His Ile
85 90 95

Val Gly Ser Pro Thr Met Asn Val Gln Ser Asn Lys Lys Leu Val His
100 105 110

His Ser Leu Gly Met Gly Asn Phe His Asn Phe Ser Glu Met Ala Lys
115 120 125

Glu Val Thr Ala Ala Thr Thr Met Leu Thr Glu Glu Asn Ala Ala Ser
130 135 140

20110907_CL5178USNA_SeqList.txt

Glu Ile Asp Arg Val Leu Glu Thr Ala Leu Leu Glu Lys Arg Pro Val
 145 150 155 160
 Tyr Ile Asn Leu Pro Ile Asp Ile Ala His Lys Ala Ile Val Lys Pro
 165 170 175
 Ala Lys Ala Leu Gln Thr Glu Lys Ser Ser Gly Glu Arg Glu Ala Gln
 180 185 190
 Leu Ala Glu Ile Ile Leu Ser His Leu Glu Lys Ala Ala Gln Pro Ile
 195 200 205
 Val Ile Ala Gly His Glu Ile Ala Arg Phe Gln Ile Arg Glu Arg Phe
 210 215 220
 Glu Asn Trp Ile Asn Gln Thr Lys Leu Pro Val Thr Asn Leu Ala Tyr
 225 230 235 240
 Gly Lys Gly Ser Phe Asn Glu Glu Asn Glu His Phe Ile Gly Thr Tyr
 245 250 255
 Tyr Pro Ala Phe Ser Asp Lys Asn Val Leu Asp Tyr Val Asp Asn Ser
 260 265 270
 Asp Phe Val Leu His Phe Gly Gly Lys Ile Ile Asp Asn Ser Thr Ser
 275 280 285
 Ser Phe Ser Gln Gly Phe Lys Thr Glu Asn Thr Leu Thr Ala Ala Asn
 290 295 300
 Asp Ile Ile Met Leu Pro Asp Gly Ser Thr Tyr Ser Gly Ile Ser Leu
 305 310 315 320
 Asn Gly Leu Leu Ala Glu Leu Glu Lys Leu Asn Phe Thr Phe Ala Asp
 325 330 335
 Thr Ala Ala Lys Gln Ala Glu Leu Ala Val Phe Glu Pro Gln Ala Glu
 340 345 350
 Thr Pro Leu Lys Gln Asp Arg Phe His Gln Ala Val Met Asn Phe Leu
 355 360 365
 Gln Ala Asp Asp Val Leu Val Thr Glu Gln Gly Thr Ser Ser Phe Gly
 370 375 380
 Leu Met Leu Ala Pro Leu Lys Lys Gly Met Asn Leu Ile Ser Gln Thr
 385 390 395 400
 Leu Trp Gly Ser Ile Gly Tyr Thr Leu Pro Ala Met Ile Gly Ser Gln
 405 410 415

20110907_CL5178USNA_SeqList.txt

Ile Ala Ala Pro Glu Arg Arg His Ile Leu Ser Ile Gly Asp Gly Ser
420 425 430

Phe Gln Leu Thr Ala Gln Glu Met Ser Thr Ile Phe Arg Glu Lys Leu
435 440 445

Thr Pro Val Ile Phe Ile Ile Asn Asn Asp Gly Tyr Thr Val Glu Arg
450 455 460

Ala Ile His Gly Glu Asp Glu Ser Tyr Asn Asp Ile Pro Thr Trp Asn
465 470 475 480

Leu Gln Leu Val Ala Glu Thr Phe Gly Gly Asp Ala Glu Thr Val Asp
485 490 495

Thr His Asn Val Phe Thr Glu Thr Asp Phe Ala Asn Thr Leu Ala Ala
500 505 510

Ile Asp Ala Thr Pro Gln Lys Ala His Val Val Glu Val His Met Glu
515 520 525

Gln Met Asp Met Pro Glu Ser Leu Arg Gln Ile Gly Leu Ala Leu Ser
530 535 540

Lys Gln Asn Ser
545

<210> 248
<211> 546
<212> PRT
<213> *Macrococcus caseolyticus*

<400> 248

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

20110907_CL5178USNA_SeqList.txt

Ile Ser Gly Gly Pro Ser Ser Val Ala Gln Gln Glu Gly Arg Tyr Val
 100 105 110
 His His Ser Leu Gly Glu Gly Ile Phe Asp Ser Tyr Ser Lys Met Tyr
 115 120 125
 Ala His Ile Thr Ala Thr Thr Thr Ile Leu Ser Val Asp Asn Ala Val
 130 135 140
 Asp Glu Ile Asp Arg Val Ile His Cys Ala Leu Lys Glu Lys Arg Pro
 145 150 155 160
 Val His Ile His Leu Pro Ile Asp Val Ala Leu Thr Glu Ile Glu Ile
 165 170 175
 Pro His Ala Pro Lys Val Tyr Thr His Glu Ser Gln Asn Val Asp Ala
 180 185 190
 Tyr Ile Gln Ala Val Glu Lys Lys Leu Met Ser Ala Lys Gln Pro Val
 195 200 205
 Ile Ile Ala Gly His Glu Ile Asn Ser Phe Lys Leu His Glu Gln Leu
 210 215 220
 Glu Gln Phe Val Asn Gln Thr Asn Ile Pro Val Ala Gln Leu Ser Leu
 225 230 235 240
 Gly Lys Ser Ala Phe Asn Glu Glu Asn Glu His Tyr Leu Gly Ile Tyr
 245 250 255
 Asp Gly Lys Ile Ala Lys Glu Asn Val Arg Glu Tyr Val Asp Asn Ala
 260 265 270
 Asp Val Ile Leu Asn Ile Gly Ala Lys Leu Thr Asp Ser Ala Thr Ala
 275 280 285
 Gly Phe Ser Tyr Lys Phe Asp Thr Asn Asn Ile Ile Tyr Ile Asn His
 290 295 300
 Asn Asp Phe Lys Ala Glu Asp Val Ile Ser Asp Asn Val Ser Leu Ile
 305 310 315 320
 Asp Leu Val Asn Gly Leu Asn Ser Ile Asp Tyr Arg Asn Glu Thr His
 325 330 335
 Tyr Pro Ser Tyr Gln Arg Ser Asp Met Lys Tyr Glu Leu Asn Asp Ala
 340 345 350
 Pro Leu Thr Gln Ser Asn Tyr Phe Lys Met Met Asn Ala Phe Leu Glu
 355 360 365

20110907_CL5178USNA_SeqList.txt

Lys Asp Asp Ile Leu Leu Ala Glu Gln Gly Thr Ser Phe Phe Gly Ala
370 375 380

Tyr Asp Leu Ser Leu Tyr Lys Gly Asn Gln Phe Ile Gly Gln Pro Leu
385 390 395 400

Trp Gly Ser Ile Gly Tyr Thr Phe Pro Ser Leu Leu Gly Ser Gln Leu
405 410 415

Ala Asp Met His Arg Arg Asn Ile Leu Leu Ile Gly Asp Gly Ser Leu
420 425 430

Gln Leu Thr Val Gln Ala Leu Ser Thr Met Ile Arg Lys Asp Ile Lys
435 440 445

Pro Ile Ile Phe Val Ile Asn Asn Asp Gly Tyr Thr Val Glu Arg Leu
450 455 460

Ile His Gly Met Glu Glu Pro Tyr Asn Asp Ile Gln Met Trp Asn Tyr
465 470 475 480

Lys Gln Leu Pro Glu Val Phe Gly Gly Lys Asp Thr Val Lys Val His
485 490 495

Asp Ala Lys Thr Ser Asn Glu Leu Lys Thr Val Met Asp Ser Val Lys
500 505 510

Ala Asp Lys Asp His Met His Phe Ile Glu Val His Met Ala Val Glu
515 520 525

Asp Ala Pro Lys Lys Leu Ile Asp Ile Ala Lys Ala Phe Ser Asp Ala
530 535 540

Asn Lys
545