

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

(19) World Intellectual Property
Organization
International Bureau



(43) International Publication Date
26 October 2012 (26.10.2012)

(10) International Publication Number
WO 2012/143513 A3

(51) International Patent Classification:
C07K 14/395 (2006.01) C12R 1/865 (2006.01)
C12N 1/36 (2006.01) C12P 7/10 (2006.01)

(21) International Application Number:
PCT/EP2012/057273

(22) International Filing Date:
20 April 2012 (20.04.2012)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/478,142 22 April 2011 (22.04.2011) US
11163579.3 22 April 2011 (22.04.2011) EP

(71) Applicant (for all designated States except US): DSM IP ASSETS B.V. [NL/NL]; Het Overloon 1, NL-6411 TE Heerlen (NL).

(72) Inventors; and

(75) Inventors/Applicants (for US only): KLAASSEN, Paul [NL/NL]; P.O.Box 130, NL-6100 AC Echt (NL). GIELESEN, Bianca, Elisabeth, Maria [NL/NL]; P.O.Box 130, NL-6100 AC Echt (NL). VAN SUYLEKOM, Gijsberdina, Pieternella [NL/NL]; P.O.Box 130, NL-6100 AC Echt (NL). SARANTINO-POULOS, Panagiotis [GR/NL]; P.O.Box 130, NL-6100 AC Echt (NL). HEIJNE, Wilbert, Herman, Marie [NL/NL]; PO Box 130, NL-6100 AC Echt (NL). GREEVE, Aldo [NL/NL]; P.O.Box 130, NL-6100 AC Echt (NL).

(74) Agents: KLEIBORN, Paul, Erik et al.; DSM Intellectual Property, P.O. Box 130, NL-6100 AC Echt (NL).

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

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

Published:

- with international search report (Art. 21(3))
- 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:
27 December 2012

(54) Title: YEAST CELL CAPABLE OF CONVERTING SUGARS INCLUDING ARABINOSE AND XYLOSE

(57) Abstract: Yeast cell belonging to the genus *Saccharomyces* having introduced into its genome at least one xylA gene and at least one of each of araA, araB and araD genes and that is capable of consuming a mixed sugar mixture comprising glucose, xylose and arabinose, wherein the cell co-consumes glucose and arabinose, has genetic variations obtained during adaptive evolution and has a specific xylose consumption rate in the presence of glucose that is 0.25 g xylose/h, g DM or more.



WO 2012/143513 A3

5

**YEAST CELL CAPABLE OF CONVERTING SUGARS INCLUDING ARABINOSE
AND XYLOSE**

Field of the invention

10 The invention relates to yeast cells which are capable of converting sugars including arabinose and xylose. The invention further relates to a process in which such cells are used for the production of a fermentation product, such as ethanol.

15

Background of the invention

 Large-scale consumption of traditional, fossil fuels (petroleum-based fuels) in recent decades has contributed to high levels of pollution. This, along with the realisation that the world stock of fossil fuels is limited and a growing environmental awareness, has stimulated new initiatives to investigate the feasibility of alternative fuels such as ethanol, which is a particulate-free burning fuel source that releases less CO₂ than unleaded gasoline on a per litre basis. Although biomass-derived ethanol may be produced by the fermentation of hexose sugars obtained from many different sources, the substrates typically used for commercial scale production of fuel alcohol, such as cane sugar and corn starch, are expensive. Increases in the production of fuel ethanol will therefore require the use of lower-cost feedstocks. Currently, only lignocellulosic feedstock derived from plant biomass is available in sufficient quantities to substitute the crops currently used for ethanol production. In most lignocellulosic material, next to C₆ sugar also contain considerable amounts of C₅ sugars, including arabinose and xylose. Thus, for an economically feasible fuel production process, both hexose and pentose sugars must be fermented to form ethanol. The yeast *Saccharomyces cerevisiae* is robust and well adapted for ethanol production, but it is unable to convert arabinose and xylose. Also, no naturally-occurring organisms are known which can ferment xylose or arabinose to ethanol

20

25

30

with both a high ethanol yield and high ethanol productivity. There is therefore a need for an organism possessing these properties so as to enable the commercially-viable production of ethanol from lignocellulosic feedstocks. In co-pending patent application that has not been published at the date of filing of this application (EP10160647.3 and the PCT application claiming its priority), strain BIE252 is described. This strain is able to ferment a mixed sugar composition that includes glucose, xylose, arabinose, galactose, and mannose and to produce fermentation product. Strain BIE252 is able to convert all these sugars, but for most part, glucose is consumed first, and the other sugars thereafter, it would be desirable to have co-consumption of glucose and C5-sugars including arabinose and xylose, since it maybe expected that shorter fermentation times are then possible.

Summary of the invention

An object of the invention is to provide a cell, in particular a yeast cell that is capable of converting a mixed sugar composition that comprises glucose, xylose and arabinose. Another object is to provide such cell that converts a mixed sugar composition that comprises glucose, xylose and arabinose in high yield. Another object is to provide such cell that is able to co-consume C5 and C6 sugar. Another object is to provide such cell, that has a high productivity. A further object is to provide such cell that is genetically stable. One or more of these objects is attained according to the invention that provides A yeast cell belonging to the genus *Saccharomyces* having introduced into its genome at least one xylA gene and at least one of each of araA, araB and araD genes and that is capable of consuming a mixed sugar mixture comprising glucose, xylose and arabinose, wherein the cell co-consumes glucose and arabinose, has genetic variations obtained during adaptive evolution and has a specific xylose consumption rate in the presence of glucose that is 0.25 g xylose/h, g DM or more.

The yeast cell of the invention is capable of converting a mixed sugar composition that comprises glucose, xylose and arabinose in high yield. Further the yeast cell has a high productivity as defined hereinafter. This allows a reduction of fermentation time. Additionally the yeast cell is genetically stable. The latter is advantageous when the yeast is used in an industrial process.

In an embodiment the yeast cell is *Saccharomyces cerevisiae*.

Brief description of the drawings

5 Figure 1 sets out the growth rates on arabinose and xylose after each cycle in the SBR cultivation system of the improved cultures of strain *S. cerevisiae* BIE252.

 Figure 2 sets out the sugar conversion and product formation of strain BIE252 on synthetic medium, in the BAM system. CO₂ production was measured constantly. Growth was monitored by following optical density of the culture. Preculture was grown on 2%
10 glucose.

 Figure 3 sets out the sugar conversion and product formation of strain BIE272 on synthetic medium, in the BAM system. CO₂ production was measured constantly. Growth was monitored by following optical density of the culture. Preculture was grown on 2%
glucose.

15 Figure 4 sets out the xylose consumption by strains BIE252 and BIE272 in the AFM fermentations in real hydrolysates at 10 and 20% dry matter pCS.

 Figure 5 sets out the arabinose consumption by strains BIE252 and BIE272 in the AFM fermentations in real hydrolysates at 10 and 20% dry matter pCS.

 Figure 6 sets out the ethanol that was produced by strains BIE252 and BIE272 in
20 the AFM fermentations in real hydrolysates at 10 and 20% dry matter pCS.

 Figure 7 sets out the CO₂ that was produced by strains BIE252 and BIE272 in the AFM fermentations in real hydrolysates at 10 and 20% dry matter pCS.

 Figure 8 sets out the performance of strain BIE272 in pretreated, hydrolyzed corn stover at 20% dry matter. Ethanol production and sugar conversion are shown.

25 Figure 9 sets out the performance stability of strain BIE272. Two colonies isolated directly from the glycerol stock of strain BIE272 and six colonies after cultivation in YEP 2% glucose for 10, 19, 28, 37 and 46 generations were tested for their ability to grow on Verduyn medium supplemented with 2% xylose. The grey parts of the bars represent the number of colonies exhibiting xylose growth better than or equal to the reference strain, BIE272. The black parts of the bar indicate the number of colonies that are lagging behind.
30 The experiment was performed in duplicate. The left panel represents the results of shake flask 1, the right panel the results of shake flask 2.

Figure 10 sets out a CHEF gel, stained with ethidium bromide. Chromosomes were separated on their size using the CHEF technique. Strains analyzed are BIE104; BIE104A2P1a, (synonym of BIE104A2P1); BIE104A2P1c; strain BIE201; BIE201X9; BIE252 and BIE272. Shifts in chromosomes are observed (see text). Strain YNN295 is a marker strain (Bio-Rad), used as a reference for the size of the chromosomes.

Figure 11 sets out an autoradiogram of a CHEF gel, blotted to a membrane and hybridized with a PNC1 probe. Strains analyzed are BIE104; BIE104A2P1a, (synonym of BIE104A2P1); BIE104A2P1c; strain BIE201; BIE201X9; BIE252 and BIE272. Shifts in chromosomes are observed (see text).

Figure 12 sets out an autoradiogram of a CHEF gel, blotted to a membrane and hybridized with a ACT1 probe (left panel, a) and the xylA probe (right panel, b). Strains analyzed are BIE104; BIE104A2P1a, (synonym of BIE104A2P1); BIE104A2P1c; strain BIE201; BIE201X9; BIE252 and BIE272. Shifts in chromosomes are observed (see text).

Figure 13 sets out the CO₂ production rate (in ml CO₂ per minute) of strains BIE104, BIE201, BIE252 and BIE272.

Figure 14 sets out the CO₂ production rate (in ml CO₂ per minute) of strains BIE104 and BIE201.

Figure 15 sets out the CO₂ production rate (in ml CO₂ per minute) of strains BIE201 and BIE252.

Figure 16 sets out the CO₂ production rate (in ml CO₂ per minute) of strains BIE252 and BIE272.

Figure 17 sets out the sugar conversion and product formation of strain BIE104 on synthetic medium, in the BAM system. CO₂ production was measured constantly. Growth was monitored by following optical density of the culture.

Figure 18 sets out the sugar conversion and product formation of strain BIE201 on synthetic medium, in the BAM system. CO₂ production was measured constantly. Growth was monitored by following optical density of the culture.

Figure 19 sets out the sugar conversion and product formation of strain BIE252 on synthetic medium, in the BAM system. CO₂ production was measured constantly. Growth was monitored by following optical density of the culture.

Figure 20 sets out the sugar conversion and product formation of strain BIE272 on synthetic medium, in the BAM system. CO₂ production was measured constantly. Growth was monitored by following optical density of the culture.

Figure 21 sets out the normalized read depth (or coverage) of the *PMA1*-gene.

Figure 22 sets out the the normalized read depth (or coverage) of the *xyIA* gene.

5

Brief description of the sequence listing

SEQ ID NO: 1: synthetic DNA, forward primer *xyIA*,
CACCGTTAGCCTTGGCGTAAGC

10

SEQ ID NO: 2 synthetic DNA, reverse primer *xyIA*,
CACTTTCGAACACGAATTGGC

SEQ ID NO: 3 synthetic DNA, forward primer *ACT1*,
GTTACGTCGCCTTGGACTTCG

15

SEQ ID NO: 4 synthetic DNA, reverse primer *ACT1*,
CGGCAATACCTGGGAACATGG

20

SEQ ID NO: 5 synthetic DNA, forward primer *PNC1*,
GATAGAGACTGGCACAGGATTG

SEQ ID NO: 6 synthetic DNA, reverse primer *PNC1*,
ACAATACTCCAAAGCTACACC

25

SEQ ID NO: 7 wild-type *PMR1* protein sequence

SEQ ID NO: 8 *PMR1* protein sequence of yeast strain BIE272

30

Detailed description of the invention

Throughout the present specification and the accompanying claims, the words "comprise" and "include" and variations such as "comprises", "comprising", "includes" and "including" are to be interpreted inclusively. That is, these words are intended to convey the possible inclusion of other elements or integers not specifically recited, where the context
5 allows.

The articles "a" and "an" are used herein to refer to one or to more than one (i.e. to one or at least one) of the grammatical object of the article. By way of example, "an element" may mean one element or more than one element.

Yeast cell or yeast cells may herein also called yeast strain.

10 The various embodiments of the invention described herein may be cross-combined.

The invention relates to a yeast cell belonging to the genus *Saccharomyces* having introduced into its genome at least one *xylA* gene and at least one of each of *araA*, *araB* and *araD* genes and that is capable of consuming a mixed sugar mixture comprising
15 glucose, xylose and arabinose, wherein the cell co-consumes glucose and arabinose, and the specific xylose consumption rate in the presence of glucose is 0.25 g xylose/h, g DM or more. Herein DM is dry yeast biomass.

In an embodiment, the specific xylose consumption rate in the presence of glucose is 0.25 or more, 0.30 or more, 0.35 or more, 0.40 or more, or about 0.41g xylose/h, g DM.
20 In an embodiment the the specific xylose consumption rate in the presence of glucose is 0.25 to 0.60 g arabinose/h/g DM. In an embodiment, the yeast cell wherein the copy number of the *araA*, *araB* and *araD* genes is three or four each. In another embodiment the yeast cell has a copy number of *xylA* of about 9 or 10.

In another embodiment, the yeast cell has one or more of the single nucleotide polymorphism chosen from the group consisting of mutations G1363T in the *SSY1* gene,
25 A512T in *YJR154w* gene, A1186G in *CEP3* gene, A436C in *GAL80* gene and A113G in *PMR1* gene.

In an embodiment the yeast cell has a single polymorphism A436C in *GAL80* gene. Optionally it also has a single nucleotide polymorphism A1186G in *CEP3* gene or a single
30 nucleotide polymorphism A113G in *PMR1* gene.

In an embodiment, the yeast cell has a yield of 0.40 g ethanol/g sugar or more or about 0.42. In another embodiment, the yeast cell has a productivity of 1.20 or more g EtOH/l, h. In an embodimet the yeast cell has a productivity of 1.25 or more, 1.30 or more,

1.35 or more, 1.40 or more, 1.45 or more, 1.50 or more, 1.55 or more, 1.60 or more or 1.65 or more g EtOH/l, h. In an embodiment the yeast cell has a productivity of about 1.69 g EtOH/l, h. The productivity is herein measured in the interval of 0-24h after start of the fermentation. Also in other time intervals the productivity of the yeast cells is high. See
5 table 11.

The invention further relates to a polypeptide having aminoacid sequence of SEQ ID NO: 7 having a substitution Tyr38Cys in *PMR1*, resulting in SEQ ID NO: 8; and variant polypeptides thereof wherein one or more of the other positions may have mutation of the aminoacid with an aminoacid that is an existing conserved aminoacid in the SPCA
10 (Secretory Pathway calcium ATP-ase) family. The invention further relates to a process for the production of one or more fermentation products from a sugar composition comprising glucose, xylose, arabinose galactose and mannose and wherein the sugar composition is fermented with a yeast cell according to the invention. In an embodiment of the process, the sugar composition is produced from lignocellulosic material by: pretreatment of one or
15 more lignocellulosic material to produce pretreated lignocellulosic material; enzymatic treatment of the pretreated lignocellulosic material to produce the sugar composition.

In another embodiment, in the process, the fermentation is conducted anaerobically. The fermentation product may be selected from the group consisting of ethanol, n-butanol, isobutanol, lactic acid, 3-hydroxy-propionic acid, acrylic acid, acetic
20 acid, succinic acid, fumaric acid, malic acid, itaconic acid, maleic acid, citric acid, adipic acid, an amino acid, such as lysine, methionine, tryptophan, threonine, and aspartic acid, 1,3-propane-diol, ethylene, glycerol, a β -lactam antibiotic and a cephalosporin, vitamins, pharmaceuticals, animal feed supplements, specialty chemicals, chemical feedstocks, plastics, solvents, fuels, including biofuels and biogas or organic polymers, and an
25 industrial enzyme, such as a protease, a cellulase, an amylase, a glucanase, a lactase, a lipase, a lyase, an oxidoreductases, a transferase or a xylanase.

In an embodiment, the yeast cell has a chromosome that is amplified compared to the host strain, wherein the amplified chromosome has the same number as the chromosome in which the *araA*, *araB* and *araD* genes were introduced in the host strain. In
30 an embodiment the amplified chromosome is chromosome VII. In an embodiment, in the yeast cell parts of chromosome VII, surrounding the centromere, are amplified (as compared to the host strain). In an embodiment, part of the right arm of chromosome VII was amplified twice, and an adjacent part was amplified three times.

The part on the right arm of chromosome VII that was amplified three times contains the arabinose expression cassette, i.e. the genes *araA*, *araB* and *araD* under control of strong constitutive promoters.

The invention further relates to a yeast cell having *araA*, *araB* and *araD* genes wherein chromosome VII has a size of from 1300 to 1400 kb, or 1375 kb as determined by electrophoresis as measured as described hereinafter.

In an embodiment, in the yeast cell, the copy number of the *araA*, *araB* and *araD* genes is two to ten, in an embodiment two to eight or three to five each. The copy number of the *araA*, *araB* and *araD* genes may be 2, 3, 4, 5, 6, 7, 8, 9, or 10. The copy number may be determined with methods known to the skilled person, Suitable methods are illustrated in the examples, and results are e.g. shown in figure 5.

In an embodiment, the yeast cell one or more of the single nucleotide polymorphism chosen from the group consisting of mutations G1363T in the *SSY1* gene, A512T in *YJR154w* gene, A1186G in *CEP3* gene, A436C in *GAL80* gene and A113G in *PMR1*. In an embodiment, the yeast cell has a single polymorphism A436C in *GAL80* gene. In an embodiment, the yeast cell has a single polymorphism A1186G in *CEP3* gene. In an embodiment, the yeast cell has a single polymorphism A113G in *PMR1*.

Adaptation

Adaptation is the evolutionary process whereby a population becomes better suited (adapted) to its habitat or habitats. This process takes place over several to many generations, and is one of the basic phenomena of biology.

The term adaptation may also refer to a feature which is especially important for an organism's survival. Such adaptations are produced in a variable population by the better suited forms reproducing more successfully, by natural selection.

Changes in environmental conditions alter the outcome of natural selection, affecting the selective benefits of subsequent adaptations that improve an organism's fitness under the new conditions. In the case of an extreme environmental change, the appearance and fixation of beneficial adaptations can be essential for survival. A large number of different factors, such as e.g. nutrient availability, temperature, the availability of oxygen, etcetera, can drive adaptive evolution.

For example, a haploid yeast strain, transformed with genes necessary for or enhancing the ability to ferment arabinose (designated all together as *ARA*) was enhanced

by a process called adaptive evolution. During the adaptive evolution process, three mutations have been introduced into the genome, designated *mut1*, *mut2* and *mut3*. The genotype of such a yeast strain could be written as *mut1 mut2 mut3 ARA*.

5 **Fitness**

There is a clear relationship between adaptedness (the degree to which an organism is able to live and reproduce in a given set of habitats) and fitness. Fitness is an estimate and a predictor of the rate of natural selection. By the application of natural selection, the relative frequencies of alternative phenotypes will vary in time, if they are
10 heritable.

Genetic changes/variations

When natural selection acts on the genetic variability of the population, genetic changes are the underlying mechanism. By this means, the population adapts genetically
15 to its circumstances. Genetic changes may result in visible structures, or may adjust the physiological activity of the organism in a way that suits the changed habitat.

It may occur that habitats frequently change. Therefore, it follows that the process of adaptation is never finally complete. In time, it may happen that the environment changes gradually, and the species comes to fit its surroundings better and better. On the
20 other hand, it may happen that changes in the environment occur relatively rapidly, and then the species becomes less and less well adapted. Adaptation is a genetic process, which goes on all the time to some extent, also when the population does not change the habitat or environment.

Single nucleotides in a DNA sequence may be changed (substitution), removed
25 (deletions) or added (insertion). Insertion or deletion SNPs (InDels) may shift the translational frame.

Single nucleotide polymorphisms may fall within coding sequences of genes (Open Reading Frames or ORFs), non-coding regions of genes (like promoter sequences, terminator sequences and the like), or in the intergenic regions between genes. SNPs
30 within a coding sequence will not necessarily change the amino acid sequence of the corresponding protein that is produced after transcription and translation, due to degeneracy of the genetic code. A SNP in which both forms lead to the same polypeptide sequence is termed synonymous (a silent mutation). If a different polypeptide sequence is

produced they are nonsynonymous. A nonsynonymous change may either be missense or nonsense. A missense change results in a different amino acid in the corresponding polypeptide, while a nonsense change results in a premature stop codon, sometimes leading to the formation of a truncated protein.

5 SNPs that are not in protein-coding regions may still have consequences for gene expression, for instance by a changed transcription factor binding or stability of the corresponding mRNA.

The changes that may occur in the DNA are not necessarily limited to the change (substitution, deletion or insertion) of a single nucleotide, but may also comprise a change
10 of two or more nucleotides (Small Nuclear Variations).

In addition, chromosomal translocations may occur. A chromosome translocation is a chromosome abnormality caused by rearrangement of parts between nonhomologous chromosomes.

In particular, in the cell according to the invention SNP are created in the following
15 reading frames: *SSY1*, *CEP3*, *GAL80* and *PMR1*.

SSY1 is herein a component of the SPS plasma membrane amino acid sensor system (Ssy1p-Ptr3p-Ssy5p), which senses external amino acid concentration and transmits intracellular signals that result in regulation of expression of amino acid permease genes.

20 *CEP3* is herein an essential kinetochore protein, component of the CBF3 complex that binds the CDEIII region of the centromere; contains an N-terminal Zn2Cys6 type zinc finger domain, a C-terminal acidic domain, and a putative coiled coil dimerization domain.

GAL80 is herein a transcriptional regulator involved in the repression of GAL genes in the absence of galactose. Typically it inhibits transcriptional activation by Gal4p and
25 inhibition is relieved by Gal3p or Gal1p binding.

PMR1 (systematic name YGL167c) is herein a High affinity Ca²⁺/Mn²⁺ P-type ATPase required for Ca²⁺ and Mn²⁺ transport into Golgi; involved in Ca²⁺ dependent protein sorting and processing. Pmr1p is the prototype of a family of transporters known as SPCA (Secretory Pathway Ca²⁺-ATPases) with members found in fungi, *C. elegans*, *D.*
30 *melanogaster*, and mammals.

According to the invention, SNP's in the genes *SSY1*, *CEP3*, *GAL80* and *PMR1* have been shown to be important for the cell to be able to ferment a mixed sugar composition.

BLAST searches were conducted for the SNP's found in these genes.

5

10

An overview of the SNP that were identified is given in table 1:

Table 1: Overview of SNP's

Gene	Nucleotide mutation position in ORF*	Amino acid mutation position in protein
<i>SSY1</i>	G1363T	E455stop
<i>YJR154w</i>	A512G	D171G
<i>CEP3</i>	A1186G	S396G
<i>GAL80</i>	A436C	T146P
<i>PMR1</i>	A113G	Y38C

* the A of the start codon ATG is the first nucleotide position

15

A blast of the genes containing the SNP resulted in the following data:

Ssy1p (member of the AA trans superfamily)

20

Component of the SPS plasma membrane amino acid sensor system (Ssy1p-Ptr3p-Ssy5p), which senses external amino acid concentration and transmits intracellular signals that result in regulation of expression of amino acid permease genes [Saccharomyces cerevisiae]

25

Ssy1p
identity

S. cerevisiae JAY291

852 aa

99%

	Ssy1p	S. cerevisiae YJM789	852 aa	99% identity
	YDR160w-like protein	S. cerevisiae AWRI1631	791 aa	99% identity
	ZYRO0F13838p	Z. rouxii CBS 732	836 aa	56% identity
	hypothetical protein	C. glabrata CBS 138	853 aa	53% identity
5	KLTH0G11726p	Lachancea thermotolerans	824 aa	46% identity

Shorter protein found in S. cerevisiae BIE201 is a unique feature.

10

YJR154w (member of the PhyH superfamily)

15 Putative protein of unknown function; green fluorescent protein (GFP)-fusion protein localizes to the cytoplasm [Saccharomyces cerevisiae]

	YJR154w	S. cerevisiae JAY291	346 aa	100% identity
	conserved protein	S. cerevisiae YJM789	346 aa	99% identity
20	putative pimeloyl-CoA synth.	S. cerevisiae	346 aa	71% identity
	YJR154Wp-like protein	S. cerevisiae AWRI1631	227 aa	99% identity
	KLTH0E09900p	Lachancea thermotolerans	340 aa	48% identity

25 In all these proteins, the D-residue at position 171 (or equivalent position based on the BLAST results) is conserved.

CEP3 (GAL4-like Zn2Cys6 binuclear cluster DNA-binding domain; found in transcription regulators like GAL4)

30 Centromere DNA-binding protein complex CBF3 subunit B

	CEP3	S. cerevisiae JAY291	608 aa	100% identity
	ZYRO0A07260p	Z. rouxii CBS 732	596 aa	46% identity

unnamed protein product	<i>Candida glabrata</i> CBS138	611 aa	44% identity
AFL200Wp	<i>A. gossypii</i> ATCC 10895	596 aa	41% identity

In all these proteins, the S-residue at position 396 (or equivalent position based on
5 the BLAST results) is conserved.

GAL80 (member of the NADB Rossmann superfamily)

Galactose/lactose metabolism regulatory protein GAL80

10

transcriptional regulator	<i>S. cerevisiae</i> YJM789	435 aa	100% identity
GAL80p	<i>S. kudriavzevii</i>	435 aa	89% identity
protein Kpol_1059p5	<i>V. polyspora</i> DSM 70294	429 aa	73% identity
ZYRO0G04664p	<i>Z. rouxii</i> CBS 732	437 aa	67% identity
15 KLTH0C02838p	<i>L. thermotolerans</i>	424 aa	64% identity
KIGAL80 protein	<i>Kluyveromyces lactis</i>	457 aa	58% identity
NECHADRAFT_86878	<i>N. haematococca</i> mpVI 77-13-4 367aa		30% identity

In all these proteins, the T-residue at position 146 (or equivalent position based on
20 the BLAST results) is conserved.

PMR1 from strain BIE272 (member of the family SPCA (Secretory Pathway Ca²⁺-ATPases))

25 Pmr1p	<i>S. cerevisiae</i> S288c	950 aa	99% identity
Pmr1p	<i>S. cerevisiae</i> JAY291	950 aa	99% identity
hypothetical protein	<i>C. glabrata</i> CBS 138	951 aa	79% identity
ZYRO0E02860p	<i>Z. rouxii</i>	943 aa	75% identity
AEL301Wp	<i>A. gossypii</i> ATCC 10895	957 aa	71% identity
30 Ca ²⁺ -pump ATPase	<i>C. tropicalis</i> MYA-3404	919 aa	61% identity

Structural variations

Structural variation (also genomic structural variation) consists of many kinds of variation in the genome of one species, and usually includes microscopic and submicroscopic types, such as deletions, duplications, copy number variations, insertions, inversions and translocations.

5

Read depth

The read depth (or coverage) represents the (average) number of nucleotides contributing to a portion of a Next Generation Sequencing assembly. The read depth expresses the number of times each base has been read. The read depth varies
10 depending on the genomic region. The average read depth may also vary depending on the mapping criteria, such as stringency and read quality.

The average sequencing depth of genomic regions is compared between sequences. This allows for the detection of regions that are over- or underrepresented.

15

Copy number variation

Copy-number variation (CNV) is a large category of structural variations, which includes insertions, deletions and duplications.

Single nucleotide polymorphism

20 A single nucleotide polymorphism (SNP) is a DNA sequence variation occurring when a single nucleotide — A, T, C, or G — in the genome (or other shared sequence) differs between members of a biological species or paired chromosomes in an individual cell.

Single nucleotide polymorphisms may fall within coding sequences of genes, non-
25 coding regions of genes, or in the intergenic regions between genes. SNPs within a coding sequence do not necessarily change the amino acid sequence of the protein that is produced, due to degeneracy of the genetic code.

Indel or DIP

30 In evolutionary studies, indel is used to mean an insertion or a deletion. Indels refer to the mutation class that includes both insertions, deletions, and the combination thereof.

Pulsed Field Gel Electrophoresis (PFGE)

PFGE is a technique used for the separation of large deoxyribonucleic acid (DNA) molecules by applying an electric field that periodically changes direction to a gel matrix. PFGE can be performed using a variety of alternative systems, such as transverse Alternating Pulsed-Field Electrophoresis (TAFE), Orthogonal Field Alternation Gel Electrophoresis (OFAGE), Field Inversion Gel Electrophoresis (FIGE) and Contour-clamped Homogeneous Electric Field (CHEF) gel electrophoresis. Each method has its advantages and disadvantages with respect to ease of use, time required to perform the electrophoresis and the resolution of the chromosomes, as reviewed by Basim and Basim (Turk J Biol 25 (2001) 405-418) and references therein.

CHEF gel electrophoresis can produce substantial chromosomal separation of chromosomes from 100 to 2500 kb of *Saccharomyces* yeast strains on one gel, although not all the larger sized and similarly sized chromosomes resolve (Sheehan et al (1991) J. Inst. Brew., Vol. 97, 163-167).

15

The mixed sugar composition

The sugar composition according to the invention comprises glucose, arabinose and xylose. Any sugar composition may be used in the invention that suffices those criteria. Optional sugars in the sugar composition are galactose and mannose, and rhamnose. In a preferred embodiment, the sugar composition is a hydrolysate of one or more lignocellulosic material. Lignocellulose herein includes hemicellulose and hemicellulose parts of biomass. Also lignocellulose includes lignocellulosic fractions of biomass. Suitable lignocellulosic materials may be found in the following list: orchard primings, chaparral, mill waste, urban wood waste, municipal waste, logging waste, forest thinnings, short- rotation woody crops, industrial waste, wheat straw, oat straw, rice straw, barley straw, rye straw, flax straw, soy hulls, rice hulls, corn gluten feed, oat hulls, sugar cane, corn stover, corn stalks, corn cobs, corn husks, switch grass, miscanthus, sweet sorghum, canola stems, soybean stems, prairie grass, gamagrass, foxtail; sugar beet pulp, citrus fruit pulp, seed hulls, cellulosic animal wastes, lawn clippings, cotton, seaweed, trees, softwood, hardwood, poplar, pine, shrubs, grasses, wheat, wheat straw, sugar cane bagasse, corn,, corn kernel, fiber from kernels, products and by-products from wet or dry milling of grains, municipal solid waste, waste paper, yard waste, herbaceous material, agricultural residues, forestry residues, pulp, paper mill residues, branches, bushes,

canes, an energy crop, forest, a fruit, a flower, a grain, a grass, a herbaceous crop, a leaf, bark, a needle, a log, a root, a sapling, a shrub, switch grass, a tree, a vegetable, fruit peel, a vine, sugar beet pulp, wheat midlings, organic waste material generated from an agricultural process, forestry wood waste, or a combination of any two or more thereof. In
5 an embodiment, the lignocellulosic material is from wheat, corn, sugar cane, rice or grass, e.g. corn stover, corn fiber, corn cobs, wheat straw, rice hulls, sugar cane bagasse or types of grass or other energy crops.

An overview of some suitable sugar compositions derived from lignocellulose and the sugar composition of their hydrolysates is given in table 2. The listed lignocelluloses
10 include: corn cobs, corn fiber, rice hulls, melon shells, sugar beet pulp, wheat straw, sugar cane bagasse, wood, grass and olive pressings.

Table 2: Overview of sugar compositions from lignocellulosic materials. Gal=galactose, Xyl=xylose, Ara=arabinose, Man=mannose, Glu=glutamate, Rham=rhamnose. The percentage galactose (% Gal) and literature source is given.

Lignocellulosic material	Gal	Xyl	Ara	Man	Glu	Rham	Sum	% Gal.	Lit.
Corn cob a	10	286	36		227	11	570	1,7	(1)
Corn cob b	131	228	160		144		663	19,8	(1)
Rice hulls a	9	122	24	18	234	10	417	2,2	(1)
Rice hulls b	8	120	28		209	12	378	2,2	(1)
Melon Shells	6	120	11		208	16	361	1,7	(1)
Sugar beet pulp	51	17	209	11	211	24	523	9,8	(2)
Wheat straw Idaho	15	249	36		396		696	2,2	(3)
Corn fiber	36	176	113		372		697	5,2	(4)
Cane Bagasse	14	180	24	5	391		614	2,3	(5)
Corn stover	19	209	29		370		626		(6)
Athel (wood)	5	118	7	3	493		625	0,7	(7)
Eucalyptus (wood)	22	105	8	3	445		583	3,8	(7)
CWR (grass)	8	165	33		340		546	1,4	(7)
JTW (grass)	7	169	28		311		515	1,3	(7)
MSW	4	24	5	20	440		493	0,9	(7)
Reed Canary Grass Veg	16	117	30	6	209	1	379	4,2	(8)
Reed Canary Grass Seed	13	163	28	6	265	1	476	2,7	(9)
Olive pressing residue	15	111	24	8	329		487	3,1	(9)

5 It is clear from table 2 that in these lignocelluloses a high amount of sugar is presence in de form of glucose, xylose, arabinose and galactose. The conversion of glucose, xylose, arabinose and galactose to fermentation product is thus of great economic importance. Also mannose and rhamnose is present in some lignocellulose materials be it in lower amounts than the previously mentioned sugars. Advantageously
10 therefore also mannose and rhamnose is converted by the mixed sugar cell.

Pretreatment and enzymatic hydrolysis

Pretreatment and enzymatic hydrolysis may be needed to release sugars that may be fermented according to the invention from the lignocellulosic (including hemicellulosic) material. These steps may be executed with conventional methods.

The mixed sugar cell

The mixed sugar cell comprising the genes *araA*, *araB* and *araD* integrated into the mixed sugar cell genome as defined hereafter. It is able to ferment glucose, arabinose, xylose, galactose and mannose. In one embodiment of the invention the mixed sugar cell is able to ferment one or more additional sugar, preferably C5 and/or C6 sugar. In an embodiment of the invention the mixed sugar cell comprises one or more of: a *xylA*-gene and/or *XKS1*-gene, to allow the mixed sugar cell to ferment xylose; deletion of the aldose reductase (*GRE3*) gene; overexpression of PPP-genes *TAL1*, *TKL1*, *RPE1* and *RKI1* to allow the increase of the flux through the pentose phosphate pass-way in the cell.

Construction of the mixed sugar strain

The genes may be introduced in the mixed sugar cell by introduction into a host cell:

- a) a cluster consisting of PPP-genes *TAL1*, *TKL1*, *RPE1* and *RKI1*, under control of strong promoters;
 - b) a cluster consisting of a *xylA*-gene and a *XKS1*-gene both under control of constitutive promoters,
 - c) a cluster consisting of the genes *araA*, *araB* and *araD* and/or a cluster of *xylA*-gene and/or the *XKS1*-gene;
- and
- d) deletion of an aldose reductase gene
- and adaptive evolution to produce the mixed sugar cell. The above cell may be constructed using recombinant expression techniques.
- e) sampling a single colony isolate
 - f) subjecting the single colony isolate to adaptive evolution in sequential batch reactors
 - g) sampling single colony isolates

- h) characterizing the single colony isolates for their sugar consumption properties

These steps will be described in details below.

5 Recombinant expression

The cell of the invention is a recombinant cell. That is to say, a cell of the invention comprises, or is transformed with or is genetically modified with a nucleotide sequence that does not naturally occur in the cell in question.

10 Techniques for the recombinant expression of enzymes in a cell, as well as for the additional genetic modifications of a cell of the invention are well known to those skilled in the art. Typically such techniques involve transformation of a cell with nucleic acid construct comprising the relevant sequence. Such methods are, for example, known from standard handbooks, such as Sambrook and Russel (2001) "Molecular Cloning: A Laboratory Manual (3rd edition), Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, or F. Ausubel *et al.*, eds., "Current protocols in molecular biology",
15 Green Publishing and Wiley Interscience, New York (1987). Methods for transformation and genetic modification of fungal host cells are known from e.g. EP-A- 0635 574, WO 98/46772, WO 99/60102, WO 00/37671, WO90/14423, EP-A-0481008, EP-A-0635574 and US 6,265,186.

20 Typically, the nucleic acid construct may be a plasmid, for instance a low copy plasmid or a high copy plasmid. The cell according to the present invention may comprise a single or multiple copies of the nucleotide sequence encoding a enzyme, for instance by multiple copies of a nucleotide construct or by use of construct which has multiple copies of the enzyme sequence.

25 The nucleic acid construct may be maintained episomally and thus comprise a sequence for autonomous replication, such as an autosomal replication sequence sequence. A suitable episomal nucleic acid construct may e.g. be based on the yeast 2 μ or pKD1 plasmids (Gleer *et al.*, 1991, Biotechnology 9: 968-975), or the AMA plasmids (Fierro *et al.*, 1995, Curr Genet. 29:482-489). Alternatively, each nucleic acid construct
30 may be integrated in one or more copies into the genome of the cell. Integration into the cell's genome may occur at random by non-homologous recombination but preferably, the nucleic acid construct may be integrated into the cell's genome by homologous

recombination as is well known in the art (see e.g. WO90/14423, EP-A-0481008, EP-A-0635 574 and US 6,265,186).

Most episomal or 2μ plasmids are relatively unstable, being lost in approximately 10^{-2} or more cells after each generation. Even under conditions of selective growth, only 60% to 95% of the cells retain the episomal plasmid. The copy number of most episomal
5 plasmids ranges from 10-40 per cell of cir^+ hosts. However, the plasmids are not equally distributed among the cells, and there is a high variance in the copy number per cell in populations. Strains transformed with integrative plasmids are extremely stable, even in the absence of selective pressure. However, plasmid loss can occur at approximately 10^{-3}
10 to 10^{-4} frequencies by homologous recombination between tandemly repeated DNA, leading to looping out of the vector sequence. Preferably, the vector design in the case of stable integration is thus, that upon loss of the selection marker genes (which also occurs by intramolecular, homologous recombination) that looping out of the integrated construct is no longer possible. Preferably the genes are thus stably integrated. Stable integration is
15 herein defined as integration into the genome, wherein looping out of the integrated construct is no longer possible. Preferably selection markers are absent. Typically, the enzyme encoding sequence will be operably linked to one or more nucleic acid sequences, capable of providing for or aiding the transcription and/or translation of the enzyme sequence.

20 The term "operably linked" refers to a juxtaposition wherein the components described are in a relationship permitting them to function in their intended manner. For instance, a promoter or enhancer is operably linked to a coding sequence the said promoter or enhancer affects the transcription of the coding sequence.

As used herein, the term "promoter" refers to a nucleic acid fragment that functions
25 to control the transcription of one or more genes, located upstream with respect to the direction of transcription of the transcription initiation site of the gene, and is structurally identified by the presence of a binding site for DNA-dependent RNA polymerase, transcription initiation sites and any other DNA sequences known to one of skilled in the art. A "constitutive" promoter is a promoter that is active under most environmental and
30 developmental conditions. An "inducible" promoter is a promoter that is active under environmental or developmental regulation.

The promoter that could be used to achieve the expression of a nucleotide sequence coding for an enzyme according to the present invention, may be not native to

the nucleotide sequence coding for the enzyme to be expressed, i.e. a promoter that is heterologous to the nucleotide sequence (coding sequence) to which it is operably linked. The promoter may, however, be homologous, i.e. endogenous, to the host cell.

Promoters are widely available and known to the skilled person. Suitable examples of such promoters include e.g. promoters from glycolytic genes, such as the phosphofructokinase (PFK), triose phosphate isomerase (TPI), glyceraldehyde-3 - phosphate dehydrogenase (GPD, TDH3 or GAPDH), pyruvate kinase (PYK), phosphoglycerate kinase (PGK) promoters from yeasts or filamentous fungi; more details about such promoters from yeast may be found in (WO 93/03159). Other useful promoters are ribosomal protein encoding gene promoters, the lactase gene promoter (LAC4), alcohol dehydrogenase promoters (ADH1, ADH4, and the like), and the enolase promoter (ENO). Other promoters, both constitutive and inducible, and enhancers or upstream activating sequences will be known to those of skill in the art. The promoters used in the host cells of the invention may be modified, if desired, to affect their control characteristics. Suitable promoters in this context include both constitutive and inducible natural promoters as well as engineered promoters, which are well known to the person skilled in the art. Suitable promoters in eukaryotic host cells may be *GAL7*, *GAL10*, or *GAL1*, *CYC1*, *HIS3*, *ADH1*, *PGL*, *PH05*, *GAPDH*, *ADC1*, *TRP1*, *URA3*, *LEU2*, *ENO1*, *TPI1*, and *AOX1*. Other suitable promoters include *PDC1*, *GPD1*, *PGK1*, *TEF1*, and *TDH3*.

In a cell of the invention, the 3'-end of the nucleotide acid sequence encoding enzyme preferably is operably linked to a transcription terminator sequence. Preferably the terminator sequence is operable in a host cell of choice, such as e.g. the yeast species of choice. In any case the choice of the terminator is not critical; it may e.g. be from any yeast gene, although terminators may sometimes work if from a non-yeast, eukaryotic, gene. Usually a nucleotide sequence encoding the enzyme comprises a terminator. Preferably, such terminators are combined with mutations that prevent nonsense mediated mRNA decay in the host cell of the invention (see for example: Shirley et al., 2002, Genetics 161:1465-1482).

The transcription termination sequence further preferably comprises a polyadenylation signal.

Optionally, a selectable marker may be present in a nucleic acid construct suitable for use in the invention. As used herein, the term "marker" refers to a gene encoding a trait or a phenotype which permits the selection of, or the screening for, a host cell containing

the marker. The marker gene may be an antibiotic resistance gene whereby the appropriate antibiotic can be used to select for transformed cells from among cells that are not transformed. Examples of suitable antibiotic resistance markers include e.g. dihydrofolate reductase, hygromycin-B-phosphotransferase, 3'-O-phosphotransferase II (kanamycin, neomycin and G418 resistance). Antibiotic resistance markers may be most convenient for the transformation of polyploid host cells, Also non- antibiotic resistance markers may be used, such as auxotrophic markers (*URA3*, *TRPI*, *LEU2*) or the *S. pombe* *TPI* gene (described by Russell P R, 1985, Gene 40: 125-130). In a preferred embodiment the host cells transformed with the nucleic acid constructs are marker gene free. Methods for constructing recombinant marker gene free microbial host cells are disclosed in EP-A-O 635 574 and are based on the use of bidirectional markers such as the *A. nidulans* *amdS* (acetamidase) gene or the yeast *URA3* and *LYS2* genes. Alternatively, a screenable marker such as Green Fluorescent Protein, *lacL*, luciferase, chloramphenicol acetyltransferase, beta-glucuronidase may be incorporated into the nucleic acid constructs of the invention allowing to screen for transformed cells.

Optional further elements that may be present in the nucleic acid constructs suitable for use in the invention include, but are not limited to, one or more leader sequences, enhancers, integration factors, and/or reporter genes, intron sequences, centromeres, telomeres and/or matrix attachment (MAR) sequences. The nucleic acid constructs of the invention may further comprise a sequence for autonomous replication, such as an ARS sequence.

The recombination process may thus be executed with known recombination techniques. Various means are known to those skilled in the art for expression and overexpression of enzymes in a cell of the invention. In particular, an enzyme may be overexpressed by increasing the copy number of the gene coding for the enzyme in the host cell, e.g. by integrating additional copies of the gene in the host cell's genome, by expressing the gene from an episomal multicopy expression vector or by introducing a episomal expression vector that comprises multiple copies of the gene.

Alternatively, overexpression of enzymes in the host cells of the invention may be achieved by using a promoter that is not native to the sequence coding for the enzyme to be overexpressed, i.e. a promoter that is heterologous to the coding sequence to which it is operably linked. Although the promoter preferably is heterologous to the coding sequence to which it is operably linked, it is also preferred that the promoter is

homologous, i.e. endogenous to the host cell. Preferably the heterologous promoter is capable of producing a higher steady state level of the transcript comprising the coding sequence (or is capable of producing more transcript molecules, i.e. mRNA molecules, per unit of time) than is the promoter that is native to the coding sequence. Suitable promoters in this context include both constitutive and inducible natural promoters as well as engineered promoters.

The coding sequence used for overexpression of the enzymes mentioned above may preferably be homologous to the host cell of the invention. However, coding sequences that are heterologous to the host cell of the invention may be used.

Overexpression of an enzyme, when referring to the production of the enzyme in a genetically modified cell, means that the enzyme is produced at a higher level of specific enzymatic activity as compared to the unmodified host cell under identical conditions. Usually this means that the enzymatically active protein (or proteins in case of multi-subunit enzymes) is produced in greater amounts, or rather at a higher steady state level as compared to the unmodified host cell under identical conditions. Similarly this usually means that the mRNA coding for the enzymatically active protein is produced in greater amounts, or again rather at a higher steady state level as compared to the unmodified host cell under identical conditions. Preferably in a host cell of the invention, an enzyme to be overexpressed is overexpressed by at least a factor of about 1.1, about 1.2, about 1.5, about 2, about 5, about 10 or about 20 as compared to a strain which is genetically identical except for the genetic modification causing the overexpression. It is to be understood that these levels of overexpression may apply to the steady state level of the enzyme's activity, the steady state level of the enzyme's protein as well as to the steady state level of the transcript coding for the enzyme.

25

The adaptive evolution

The mixed sugar cells are in their preparation subjected to adaptive evolution. A cell of the invention may be adapted to sugar utilisation by selection of mutants, either spontaneous or induced (e.g. by radiation or chemicals), for growth on the desired sugar, preferably as sole carbon source, and more preferably under anaerobic conditions. Selection of mutants may be performed by techniques including serial transfer of cultures as e.g. described by Kuyper *et al.* (2004, FEMS Yeast Res. 4: 655-664) or by cultivation under selective pressure in a chemostat culture. E.g. in a preferred host cell of the

30

invention at least one of the genetic modifications described above, including modifications obtained by selection of mutants, confer to the host cell the ability to grow on the xylose as carbon source, preferably as sole carbon source, and preferably under anaerobic conditions. Preferably the cell produce essentially no xylitol, e.g. the xylitol produced is below the detection limit or e.g. less than about 5, about 2, about 1, about 0.5, or about 0.3 % of the carbon consumed on a molar basis.

Adaptive evolution is also described e.g. in Wisselink H.W. et al, Applied and Environmental Microbiology Aug. 2007, p. 4881–4891

In one embodiment of adaptive evolution a regimen consisting of repeated batch cultivation with repeated cycles of consecutive growth in different media is applied, e.g. three media with different compositions (glucose, xylose, and arabinose; xylose and arabinose. See Wisselink et al. (2009) Applied and Environmental Microbiology, Feb. 2009, p. 907–914.

In one embodiment, the yeast cell BIE252 was adapted in a SBR set-up. The following media were used: (1) mixed sugars medium: 10 g/l glucose, 10 g/l xylose, 7 g/l arabinose, 2 g/l galactose and 1 g/l mannose; (2) arabinose medium: 27 g/l arabinose and 3 g/l xylose and (3) xylose medium: 27 g/l xylose and 3 g/l arabinose. After completion of the batch growth in medium (1), media (2) and (3) were alternated and that sequence of cultivation in media (2) and (3) was repeated for six cycles. Growth in medium (1) was repeated after cycle three to confirm that the culture was still able to utilize the C6-sugars as fast as in the beginning of the SBR cultivation. For each run, the maximum specific growth rate (μ_{\max}) was estimated from the CO₂ profile in the exponential growth phase.

The host cell

The host cell may be any host cell suitable for production of a useful product. A cell of the invention may be any suitable cell, such as a prokaryotic cell, such as a bacterium, or a eukaryotic cell. Typically, the cell will be a eukaryotic cell, for example a yeast or a filamentous fungus.

Yeasts are herein defined as eukaryotic microorganisms and include all species of the subdivision Eumycotina (Alexopoulos, C. J., 1962, In : Introductory Mycology, John Wiley & Sons, Inc. , New York) that predominantly grow in unicellular form.

Yeasts may either grow by budding of a unicellular thallus or may grow by fission of the organism. A preferred yeast as a cell of the invention may belong to the genera

Saccharomyces, *Kluyveromyces*, *Candida*, *Pichia*, *Schizosaccharomyces*, *Hansenula*, *Kloeckera*, *Schwanniomyces* or *Yarrowia*. Preferably the yeast is one capable of anaerobic fermentation, more preferably one capable of anaerobic alcoholic fermentation.

Filamentous fungi are herein defined as eukaryotic microorganisms that include all
5 filamentous forms of the subdivision Eumycotina. These fungi are characterized by a vegetative mycelium composed of chitin, cellulose, and other complex polysaccharides. The filamentous fungi of the suitable for use as a cell of the present invention are morphologically, physiologically, and genetically distinct from yeasts. Filamentous fungal cells may be advantageously used since most fungi do not require sterile conditions for
10 propagation and are insensitive to bacteriophage infections. Vegetative growth by filamentous fungi is by hyphal elongation and carbon catabolism of most filamentous fungi is obligately aerobic. Preferred filamentous fungi as a host cell of the invention may belong to the genus *Aspergillus*, *Trichoderma*, *Humicola*, *Acremonium*, *Fusarium* or *Penicillium*. More preferably, the filamentous fungal cell may be a *Aspergillus niger*, *Aspergillus oryzae*,
15 a *Penicillium chrysogenum*, or *Rhizopus oryzae* cell.

In one embodiment the host cell may be yeast.

Preferably the host is an industrial host, more preferably an industrial yeast. An industrial host and industrial yeast cell may be defined as follows. The living environments
20 of yeast cells in industrial processes are significantly different from that in the laboratory. Industrial yeast cells must be able to perform well under multiple environmental conditions which may vary during the process. Such variations include change in nutrient sources, pH, ethanol concentration, temperature, oxygen concentration, etc., which together have potential impact on the cellular growth and ethanol production of *Saccharomyces cerevisiae*. Under adverse industrial conditions, the environmental tolerant strains should
25 allow robust growth and production. Industrial yeast strains are generally more robust towards these changes in environmental conditions which may occur in the applications they are used, such as in the baking industry, brewing industry, wine making and the ethanol industry. Examples of industrial yeast (*S. cerevisiae*) are Ethanol Red® (Fermentis) Fermiol® (DSM) and Thermosacc® (Lallemand).

30 In an embodiment the host is inhibitor tolerant. Inhibitor tolerant host cells may be selected by screening strains for growth on inhibitors containing materials, such as illustrated in Kadar et al, Appl. Biochem. Biotechnol. (2007), Vol. 136-140, 847-858, wherein an inhibitor tolerant *S. cerevisiae* strain ATCC 26602 was selected.

Preferably the host cell is industrial and inhibitor tolerant.

araA, araB and araD genes

A cell of the invention is capable of using arabinose. A cell of the invention is therefore, be capable of converting L-arabinose into L-ribulose and/or xylulose 5-phosphate and/or into a desired fermentation product, for example one of those mentioned herein.

Organisms, for example *S. cerevisiae* strains, able to produce ethanol from L-arabinose may be produced by modifying a cell introducing the *araA* (L-arabinose isomerase), *araB* (L-ribulokinase) and *araD* (L-ribulose-5-P4-epimerase) genes from a suitable source. Such genes may be introduced into a cell of the invention in order that it is capable of using arabinose. Such an approach is given is described in WO2003/095627. *araA*, *araB* and *araD* genes from *Lactobacillus plantanum* may be used and are disclosed in WO2008/041840. The *araA* gene from *Bacillus subtilis* and the *araB* and *araD* genes from *Escherichia coli* may be used and are disclosed in EP1499708.

PPP-genes

A cell of the invention may comprise one or more genetic modifications that increases the flux of the pentose phosphate pathway. In particular, the genetic modification(s) may lead to an increased flux through the non-oxidative part pentose phosphate pathway. A genetic modification that causes an increased flux of the non-oxidative part of the pentose phosphate pathway is herein understood to mean a modification that increases the flux by at least a factor of about 1.1, about 1.2, about 1.5, about 2, about 5, about 10 or about 20 as compared to the flux in a strain which is genetically identical except for the genetic modification causing the increased flux. The flux of the non-oxidative part of the pentose phosphate pathway may be measured by growing the modified host on xylose as sole carbon source, determining the specific xylose consumption rate and subtracting the specific xylitol production rate from the specific xylose consumption rate, if any xylitol is produced. However, the flux of the non-oxidative part of the pentose phosphate pathway is proportional with the growth rate on xylose as sole carbon source, preferably with the anaerobic growth rate on xylose as sole carbon source. There is a linear relation between the growth rate on xylose as sole carbon source (μ_{\max}) and the flux of the non-oxidative part of the pentose phosphate pathway. The

specific xylose consumption rate (Q_s) is equal to the growth rate (μ) divided by the yield of biomass on sugar (Y_{xs}) because the yield of biomass on sugar is constant (under a given set of conditions: anaerobic, growth medium, pH, genetic background of the strain, etc.; i.e. $Q_s = \mu / Y_{xs}$). Therefore the increased flux of the non-oxidative part of the pentose phosphate pathway may be deduced from the increase in maximum growth rate under
5 these conditions unless transport (uptake is limiting).

One or more genetic modifications that increase the flux of the pentose phosphate pathway may be introduced in the host cell in various ways. These including e.g. achieving higher steady state activity levels of xylulose kinase and/or one or more of the enzymes of
10 the non-oxidative part pentose phosphate pathway and/or a reduced steady state level of unspecific aldose reductase activity. These changes in steady state activity levels may be effected by selection of mutants (spontaneous or induced by chemicals or radiation) and/or by recombinant DNA technology e.g. by overexpression or inactivation, respectively, of genes encoding the enzymes or factors regulating these genes.

In a preferred host cell, the genetic modification comprises overexpression of at
15 least one enzyme of the (non-oxidative part) pentose phosphate pathway. Preferably the enzyme is selected from the group consisting of the enzymes encoding for ribulose-5-phosphate isomerase, ribulose-5-phosphate epimerase, transketolase and transaldolase. Various combinations of enzymes of the (non-oxidative part) pentose phosphate pathway
20 may be overexpressed. E.g. the enzymes that are overexpressed may be at least the enzymes ribulose-5-phosphate isomerase and ribulose-5-phosphate epimerase; or at least the enzymes ribulose-5-phosphate isomerase and transketolase; or at least the enzymes ribulose-5-phosphate isomerase and transaldolase; or at least the enzymes ribulose-5-phosphate epimerase and transketolase; or at least the enzymes ribulose-5-phosphate
25 epimerase and transaldolase; or at least the enzymes transketolase and transaldolase; or at least the enzymes ribulose-5-phosphate epimerase, transketolase and transaldolase; or at least the enzymes ribulose-5-phosphate isomerase, transketolase and transaldolase; or at least the enzymes ribulose-5-phosphate isomerase, ribulose-5-phosphate epimerase,
30 and transaldolase; or at least the enzymes ribulose-5-phosphate isomerase, ribulose-5-phosphate epimerase, and transketolase. In one embodiment of the invention each of the enzymes ribulose-5-phosphate isomerase, ribulose-5-phosphate epimerase, transketolase and transaldolase are overexpressed in the host cell. More preferred is a host cell in which the genetic modification comprises at least overexpression of both the enzymes

transketolase and transaldolase as such a host cell is already capable of anaerobic growth on xylose. In fact, under some conditions host cells overexpressing only the transketolase and the transaldolase already have the same anaerobic growth rate on xylose as do host cells that overexpress all four of the enzymes, i.e. the ribulose-5-phosphate isomerase, 5 ribulose-5-phosphate epimerase, transketolase and transaldolase. Moreover, host cells overexpressing both of the enzymes ribulose-5-phosphate isomerase and ribulose-5-phosphate epimerase are preferred over host cells overexpressing only the isomerase or only the epimerase as overexpression of only one of these enzymes may produce metabolic imbalances.

10 The enzyme "ribulose 5-phosphate epimerase" (EC 5.1.3.1) is herein defined as an enzyme that catalyses the epimerisation of D-xylulose 5-phosphate into D-ribulose 5-phosphate and vice versa. The enzyme is also known as phosphoribulose epimerase; erythrose-4-phosphate isomerase; phosphoketopentose 3-epimerase; xylulose phosphate 3-epimerase; phosphoketopentose epimerase; ribulose 5-phosphate 3- epimerase; D- 15 ribulose phosphate-3-epimerase; D-ribulose 5-phosphate epimerase; D- ribulose-5-P 3-epimerase; D-xylulose-5-phosphate 3-epimerase; pentose-5-phosphate 3-epimerase; or D-ribulose-5-phosphate 3-epimerase. A ribulose 5-phosphate epimerase may be further defined by its amino acid sequence. Likewise a ribulose 5-phosphate epimerase may be defined by a nucleotide sequence encoding the enzyme as well as by a nucleotide 20 sequence hybridising to a reference nucleotide sequence encoding a ribulose 5-phosphate epimerase. The nucleotide sequence encoding for ribulose 5-phosphate epimerase is herein designated *RPE1*.

The enzyme "ribulose 5-phosphate isomerase" (EC 5.3.1.6) is herein defined as an enzyme that catalyses direct isomerisation of D-ribose 5-phosphate into D-ribulose 5- 25 phosphate and vice versa. The enzyme is also known as phosphopentosisomerase; phosphoriboisomerase; ribose phosphate isomerase; 5-phosphoribose isomerase; D-ribose 5-phosphate isomerase; D-ribose-5-phosphate ketol-isomerase; or D-ribose-5-phosphate aldose-ketose-isomerase. A ribulose 5-phosphate isomerase may be further defined by its amino acid sequence. Likewise a ribulose 5-phosphate isomerase may be 30 defined by a nucleotide sequence encoding the enzyme as well as by a nucleotide sequence hybridising to a reference nucleotide sequence encoding a ribulose 5-phosphate isomerase. The nucleotide sequence encoding for ribulose 5-phosphate isomerase is herein designated *RPI1*.

The enzyme "transketolase" (EC 2.2.1.1) is herein defined as an enzyme that catalyses the reaction: D-ribose 5-phosphate + D-xylulose 5-phosphate \leftrightarrow sedoheptulose 7-phosphate + D-glyceraldehyde 3-phosphate and vice versa. The enzyme is also known as glycolaldehydetransferase or sedoheptulose-7-phosphate:D-glyceraldehyde-3-phosphate glycolaldehydetransferase. A transketolase may be further defined by its amino acid. Likewise a transketolase may be defined by a nucleotide sequence encoding the enzyme as well as by a nucleotide sequence hybridising to a reference nucleotide sequence encoding a transketolase. The nucleotide sequence encoding for transketolase is herein designated *TKL1*.

The enzyme "transaldolase" (EC 2.2.1.2) is herein defined as an enzyme that catalyses the reaction: sedoheptulose 7-phosphate + D-glyceraldehyde 3-phosphate \leftrightarrow D-erythrose 4-phosphate + D-fructose 6-phosphate and vice versa. The enzyme is also known as dihydroxyacetone transferase; dihydroxyacetone synthase; formaldehyde transketolase; or sedoheptulose-7-phosphate :D-glyceraldehyde-3-phosphate glyceronetransferase. A transaldolase may be further defined by its amino acid sequence. Likewise a transaldolase may be defined by a nucleotide sequence encoding the enzyme as well as by a nucleotide sequence hybridising to a reference nucleotide sequence encoding a transaldolase. The nucleotide sequence encoding for transketolase from is herein designated *TAL1*.

Xylose Isomerase gene

The presence of the nucleotide sequence encoding a xylose isomerase confers on the cell the ability to isomerise xylose to xylulose. According to the invention, two to fifteen copies of one or more xylose isomerase gene are introduced into the host cell.

In one embodiment, the two to fifteen copies of one or more xylose isomerase gene are introduced into the host cell.

A "xylose isomerase" (EC 5.3.1.5) is herein defined as an enzyme that catalyses the direct isomerisation of D-xylose into D-xylulose and/or *vice versa*. The enzyme is also known as a D-xylose ketoisomerase. A xylose isomerase herein may also be capable of catalysing the conversion between D-glucose and D-fructose (and accordingly may therefore be referred to as a glucose isomerase). A xylose isomerase herein may require a bivalent cation, such as magnesium, manganese or cobalt as a cofactor.

Accordingly, a cell of the invention is capable of isomerising xylose to xylulose. The ability of isomerising xylose to xylulose is conferred on the host cell by transformation of the host cell with a nucleic acid construct comprising a nucleotide sequence encoding a defined xylose isomerase. A cell of the invention isomerises xylose into xylulose by the direct isomerisation of xylose to xylulose. This is understood to mean that xylose is isomerised into xylulose in a single reaction catalysed by a xylose isomerase, as opposed to two step conversion of xylose into xylulose via a xylitol intermediate as catalysed by xylose reductase and xylitol dehydrogenase, respectively.

A unit (U) of xylose isomerase activity may herein be defined as the amount of enzyme producing 1 nmol of xylulose per minute, under conditions as described by Kuyper *et al.* (2003, FEMS Yeast Res. 4: 69-78). The Xylose isomerase gene may have various origin, such as for example *Pyromyces* sp. as disclosed in WO2006/009434. Other suitable origins are *Bacteroides*, in particular *Bacteroides uniformis* as described in PCT/EP2009/52623, *Bacillus*, in particular *Bacillus stearothermophilus* as described in PCT/EP2009/052625, *Thermotoga*, in particular *Thermotoga maritima*, as described in PCT/EP2009/052621 and *Clostridium*, in particular *Clostridium cellulolyticum* as described in PCT/EP2009/052620.

XKS1 gene

A cell of the invention may comprise one or more genetic modifications that increase the specific xylulose kinase activity. Preferably the genetic modification or modifications causes overexpression of a xylulose kinase, e.g. by overexpression of a nucleotide sequence encoding a xylulose kinase. The gene encoding the xylulose kinase may be endogenous to the host cell or may be a xylulose kinase that is heterologous to the host cell. A nucleotide sequence used for overexpression of xylulose kinase in the host cell of the invention is a nucleotide sequence encoding a polypeptide with xylulose kinase activity.

The enzyme "xylulose kinase" (EC 2.7.1.17) is herein defined as an enzyme that catalyses the reaction $\text{ATP} + \text{D-xylulose} = \text{ADP} + \text{D-xylulose 5-phosphate}$. The enzyme is also known as a phosphorylating xylulokinase, D-xylulokinase or ATP :D- xylulose 5-phosphotransferase. A xylulose kinase of the invention may be further defined by its amino acid sequence. Likewise a xylulose kinase may be defined by a nucleotide sequence

encoding the enzyme as well as by a nucleotide sequence hybridising to a reference nucleotide sequence encoding a xylulose kinase.

In a cell of the invention, a genetic modification or modifications that increase(s) the specific xylulose kinase activity may be combined with any of the modifications increasing the flux of the pentose phosphate pathway as described above. This is not, however, essential.

Thus, a host cell of the invention may comprise only a genetic modification or modifications that increase the specific xylulose kinase activity. The various means available in the art for achieving and analysing overexpression of a xylulose kinase in the host cells of the invention are the same as described above for enzymes of the pentose phosphate pathway. Preferably in the host cells of the invention, a xylulose kinase to be overexpressed is overexpressed by at least a factor of about 1.1, about 1.2, about 1.5, about 2, about 5, about 10 or about 20 as compared to a strain which is genetically identical except for the genetic modification(s) causing the overexpression. It is to be understood that these levels of overexpression may apply to the steady state level of the enzyme's activity, the steady state level of the enzyme's protein as well as to the steady state level of the transcript coding for the enzyme.

20

Aldose reductase (*GRE3*) gene deletion

A cell of the invention may comprise one or more genetic modifications that reduce unspecific aldose reductase activity in the host cell. Preferably, unspecific aldose reductase activity is reduced in the host cell by one or more genetic modifications that reduce the expression of or inactivates a gene encoding an unspecific aldose reductase. Preferably, the genetic modification(s) reduce or inactivate the expression of each endogenous copy of a gene encoding an unspecific aldose reductase in the host cell (herein called *GRE3* deletion). Host cells may comprise multiple copies of genes encoding unspecific aldose reductases as a result of di-, poly- or aneu-ploidy, and/or the host cell may contain several different (iso)enzymes with aldose reductase activity that differ in amino acid sequence and that are each encoded by a different gene. Also in such instances preferably the expression of each gene that encodes an unspecific aldose reductase is reduced or inactivated. Preferably, the gene is inactivated by deletion of at

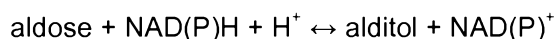
30

least part of the gene or by disruption of the gene, whereby in this context the term gene also includes any non-coding sequence up- or down-stream of the coding sequence, the (partial) deletion or inactivation of which results in a reduction of expression of unspecific aldose reductase activity in the host cell.

5 A nucleotide sequence encoding an aldose reductase whose activity is to be reduced in the host cell of the invention is a nucleotide sequence encoding a polypeptide with aldose reductase activity.

Thus, a host cell of the invention comprising only a genetic modification or modifications that reduce(s) unspecific aldose reductase activity in the host cell is
10 specifically included in the invention.

The enzyme "aldose reductase" (EC 1.1.1.21) is herein defined as any enzyme that is capable of reducing xylose or xylulose to xylitol. In the context of the present invention an aldose reductase may be any unspecific aldose reductase that is native (endogenous) to a host cell of the invention and that is capable of reducing xylose
15 or xylulose to xylitol. Unspecific aldose reductases catalyse the reaction:



The enzyme has a wide specificity and is also known as aldose reductase; polyol dehydrogenase (NADP⁺); alditol:NADP oxidoreductase; alditol:NADP⁺ 1- oxidoreductase; NADPH-aldopentose reductase; or NADPH-aldose reductase.

20 A particular example of such an unspecific aldose reductase that is endogenous to *S. cerevisiae* and that is encoded by the *GRE3* gene (Traff *et al.*, 2001, Appl. Environ. Microbiol. 67: 5668-74). Thus, an aldose reductase of the invention may be further defined by its amino acid sequence. Likewise an aldose reductase may be defined by the nucleotide sequences encoding the enzyme as well as by a nucleotide sequence
25 hybridising to a reference nucleotide sequence encoding an aldose reductase.

Bioproducts production

Over the years suggestions have been made for the introduction of various organisms for the production of bio-ethanol from crop sugars. In practice, however, all
30 major bio-ethanol production processes have continued to use the yeasts of the genus *Saccharomyces* as ethanol producer. This is due to the many attractive features of *Saccharomyces* species for industrial processes, i. e. , a high acid-, ethanol-and osmo-tolerance, capability of anaerobic growth, and of course its high alcoholic fermentative

capacity. Preferred yeast species as host cells include *S. cerevisiae*, *S. bulderi*, *S. barnetti*, *S. exiguus*, *S. uvarum*, *S. diastaticus*, *K. lactis*, *K. marxianus* or *K. fragilis*.

A cell of the invention may be able to convert plant biomass, celluloses, hemicelluloses, pectins, rhamnase, galactose, fructose, maltose, maltodextrines, ribose, ribulose, or starch, starch derivatives, sucrose, lactose and glycerol, for example into fermentable sugars. Accordingly, a cell of the invention may express one or more enzymes such as a cellulase (an endocellulase or an exocellulase), a hemicellulase (an endo- or exo-xylanase or arabinase) necessary for the conversion of cellulose into glucose monomers and hemicellulose into xylose and arabinose monomers, a pectinase able to convert pectins into glucuronic acid and galacturonic acid or an amylase to convert starch into glucose monomers.

The cell further preferably comprises those enzymatic activities required for conversion of pyruvate to a desired fermentation product, such as ethanol, butanol, lactic acid, 3 -hydroxy- propionic acid, acrylic acid, acetic acid, succinic acid, citric acid, fumaric acid, malic acid, itaconic acid, an amino acid, 1,3- propane-diol, ethylene, glycerol, a β -lactam antibiotic or a cephalosporin.

A preferred cell of the invention is a cell that is naturally capable of alcoholic fermentation, preferably, anaerobic alcoholic fermentation. A cell of the invention preferably has a high tolerance to ethanol, a high tolerance to low pH (i.e. capable of growth at a pH lower than about 5, about 4, about 3, or about 2.5) and towards organic acids like lactic acid, acetic acid or formic acid and/or sugar degradation products such as furfural and hydroxy- methylfurfural and/or a high tolerance to elevated temperatures.

Any of the above characteristics or activities of a cell of the invention may be naturally present in the cell or may be introduced or modified by genetic modification.

A cell of the invention may be a cell suitable for the production of ethanol. A cell of the invention may, however, be suitable for the production of fermentation products other than ethanol. Such non-ethanolic fermentation products include in principle any bulk or fine chemical that is producible by a eukaryotic microorganism such as a yeast or a filamentous fungus.

Such fermentation products may be, for example, butanol, lactic acid, 3 -hydroxy-propionic acid, acrylic acid, acetic acid, succinic acid, citric acid, malic acid, fumaric acid, itaconic acid, an amino acid, 1,3-propane-diol, ethylene, glycerol, a β -lactam antibiotic or a cephalosporin. A preferred cell of the invention for production of non-ethanolic

fermentation products is a host cell that contains a genetic modification that results in decreased alcohol dehydrogenase activity.

In a further aspect the invention relates to fermentation processes in which the cells of the invention are used for the fermentation of a carbon source comprising a source of xylose, such as xylose. In addition to a source of xylose the carbon source in the fermentation medium may also comprise a source of glucose. The source of xylose or glucose may be xylose or glucose as such or may be any carbohydrate oligo- or polymer comprising xylose or glucose units, such as e.g. lignocellulose, xylans, cellulose, starch and the like. For release of xylose or glucose units from such carbohydrates, appropriate carbohydrases (such as xylanases, glucanases, amylases and the like) may be added to the fermentation medium or may be produced by the cell. In the latter case the cell may be genetically engineered to produce and excrete such carbohydrases. An additional advantage of using oligo- or polymeric sources of glucose is that it enables to maintain a low(er) concentration of free glucose during the fermentation, e.g. by using rate-limiting amounts of the carbohydrases. This, in turn, will prevent repression of systems required for metabolism and transport of non-glucose sugars such as xylose.

In a preferred process the cell ferments both the xylose and glucose, preferably simultaneously in which case preferably a cell is used which is insensitive to glucose repression to prevent diauxic growth. In addition to a source of xylose (and glucose) as carbon source, the fermentation medium will further comprise the appropriate ingredient required for growth of the cell. Compositions of fermentation media for growth of microorganisms such as yeasts are well known in the art. The fermentation process is a process for the production of a fermentation product such as e.g. ethanol, butanol, lactic acid, 3-hydroxy-propionic acid, acrylic acid, acetic acid, succinic acid, citric acid, malic acid, fumaric acid, itaconic acid, an amino acid, 1,3-propane-diol, ethylene, glycerol, a β -lactam antibiotic, such as Penicillin G or Penicillin V and fermentative derivatives thereof, and a cephalosporin.

Bioproducts production

Over the years suggestions have been made for the introduction of various organisms for the production of bio-ethanol from crop sugars. In practice, however, all major bio-ethanol production processes have continued to use the yeasts of the genus *Saccharomyces* as ethanol producer. This is due to the many attractive features of

Saccharomyces species for industrial processes, i. e. , a high acid-, ethanol-and osmo-tolerance, capability of anaerobic growth, and of course its high alcoholic fermentative capacity. Preferred yeast species as host cells include *S. cerevisiae*, *S. bulderi*, *S. barnetti*, *S. exiguus*, *S. uvarum*, *S. diastaticus*, *K. lactis*, *K. marxianus* or *K fragilis*.

5 A mixed sugar cell may be a cell suitable for the production of ethanol. A mixed sugar cell may, however, be suitable for the production of fermentation products other than ethanol. Such non-ethanolic fermentation products include in principle any bulk or fine chemical that is producible by a eukaryotic microorganism such as a yeast or a filamentous fungus.

10 A mixed sugar cell may be used for production of non-ethanolic fermentation products is a host cell that contains a genetic modification that results in decreased alcohol dehydrogenase activity.

In an embodiment the mixed sugar cell may be used in a process wherein sugars originating from lignocellulose are converted into ethanol.

15

Lignocellulose

Lignocellulose, which may be considered as a potential renewable feedstock, generally comprises the polysaccharides cellulose (glucans) and hemicelluloses (xylans, heteroxylans and xyloglucans). In addition, some hemicellulose may be present as glucomannans, for example in wood-derived feedstocks. The enzymatic hydrolysis of
20 these polysaccharides to soluble sugars, including both monomers and multimers, for example glucose, cellobiose, xylose, arabinose, galactose, fructose, mannose, rhamnose, ribose, galacturonic acid, glucuronic acid and other hexoses and pentoses occurs under the action of different enzymes acting in concert.

25 In addition, pectins and other pectic substances such as arabinans may make up considerably proportion of the dry mass of typically cell walls from non-woody plant tissues (about a quarter to half of dry mass may be pectins).

Pretreatment

30 Before enzymatic treatment, the lignocellulosic material may be pretreated. The pretreatment may comprise exposing the lignocellulosic material to an acid, a base, a solvent, heat, a peroxide, ozone, mechanical shredding, grinding, milling or rapid

depressurization, or a combination of any two or more thereof. This chemical pretreatment is often combined with heat-pretreatment, e.g. between 150-220 °C for 1 to 30 minutes.

Enzymatic hydrolysis

5 The pretreated material is commonly subjected to enzymatic hydrolysis to release sugars that may be fermented according to the invention. This may be executed with conventional methods, e.g. contacting with cellulases, for instance cellobiohydrolase(s), endoglucanase(s), beta-glucosidase(s) and optionally other enzymes. The conversion with the cellulases may be executed at ambient temperatures or at higher temperatures, at a
10 reaction time to release sufficient amounts of sugar(s). The result of the enzymatic hydrolysis is hydrolysis product comprising C5/C6 sugars, herein designated as the sugar composition.

Fermentation

15 The fermentation process may be an aerobic or an anaerobic fermentation process. An anaerobic fermentation process is herein defined as a fermentation process run in the absence of oxygen or in which substantially no oxygen is consumed, preferably less than about 5, about 2.5 or about 1 mmol/L/h, more preferably 0 mmol/L/h is consumed (i.e. oxygen consumption is not detectable), and wherein organic molecules serve as both
20 electron donor and electron acceptors. In the absence of oxygen, NADH produced in glycolysis and biomass formation, cannot be oxidised by oxidative phosphorylation. To solve this problem many microorganisms use pyruvate or one of its derivatives as an electron and hydrogen acceptor thereby regenerating NAD⁺.

Thus, in a preferred anaerobic fermentation process pyruvate is used as an
25 electron (and hydrogen acceptor) and is reduced to fermentation products such as ethanol, butanol, lactic acid, 3-hydroxy-propionic acid, acrylic acid, acetic acid, succinic acid, citric acid, malic acid, fumaric acid, an amino acid, 1,3-propane-diol, ethylene, glycerol, a β -lactam antibiotic and a cephalosporin.

The fermentation process is preferably run at a temperature that is optimal for the
30 cell. Thus, for most yeasts or fungal host cells, the fermentation process is performed at a temperature which is less than about 42°C, preferably less than about 38°C. For yeast or filamentous fungal host cells, the fermentation process is preferably performed at a

temperature which is lower than about 35, about 33, about 30 or about 28°C and at a temperature which is higher than about 20, about 22, or about 25°C.

The ethanol yield on xylose and/or glucose in the process preferably is at least about 50, about 60, about 70, about 80, about 90, about 95 or about 98%. The ethanol yield is herein defined as a percentage of the theoretical maximum yield.

The invention also relates to a process for producing a fermentation product.,

The fermentation processes may be carried out in batch, fed-batch or continuous mode. A separate hydrolysis and fermentation (SHF) process or a simultaneous saccharification and fermentation (SSF) process may also be applied. A combination of these fermentation process modes may also be possible for optimal productivity.

The fermentation process according to the present invention may be run under aerobic and anaerobic conditions. Preferably, the process is carried out under micro-aerophilic or oxygen limited conditions.

An anaerobic fermentation process is herein defined as a fermentation process run in the absence of oxygen or in which substantially no oxygen is consumed, preferably less than about 5, about 2.5 or about 1 mmol/L/h, and wherein organic molecules serve as both electron donor and electron acceptors.

An oxygen-limited fermentation process is a process in which the oxygen consumption is limited by the oxygen transfer from the gas to the liquid. The degree of oxygen limitation is determined by the amount and composition of the ingoing gasflow as well as the actual mixing/mass transfer properties of the fermentation equipment used. Preferably, in a process under oxygen-limited conditions, the rate of oxygen consumption is at least about 5.5, more preferably at least about 6, such as at least 7 mmol/L/h. A process of the invention comprises recovery of the fermentation product.

Fermentation product

The fermentation product of the invention may be any useful product. In one embodiment, it is a product selected from the group consisting of ethanol, n-butanol, isobutanol, lactic acid, 3-hydroxy-propionic acid, acrylic acid, acetic acid, succinic acid, fumaric acid, malic acid, itaconic acid, maleic acid, citric acid, adipic acid, an amino acid, such as lysine, methionine, tryptophan, threonine, and aspartic acid, 1,3-propane-diol, ethylene, glycerol, a β -lactam antibiotic and a cephalosporin, vitamins, pharmaceuticals, animal feed supplements, specialty chemicals, chemical feedstocks, plastics, solvents,

fuels, including biofuels and biogas or organic polymers, and an industrial enzyme, such as a protease, a cellulase, an amylase, a glucanase, a lactase, a lipase, a lyase, an oxidoreductases, a transferase or a xylanase. For example the fermentation products may be produced by cells according to the invention, following additionally prior art cell preparation methods and fermentation processes, which examples however should herein not be construed as limiting. For example. n-butanol may be produced by cells as described in WO2008121701 or WO2008086124; lactic acid as described in US2011053231 or US2010137551; 3-hydroxy-propionic acid as described in WO2010010291; acrylic acid as described in WO2009153047. An overview of all kind of fermentation products is and how they can be prepared in yeast is given in Romanos, MA, et al, "Foreign Gene Expression in Yeast: a Review", yeast vol. 8: 423-488 (1992), see e.g. table 7. The production of glycerol, 1,3 propane diol, organic acids, and vitamin C (table 2) is described in Nevoigt, E., Microbiol. Mol. Biol. Rev. 72(3) 379-412 (2008). Giddijala, L., et al, BMC Biotechnology 8(29) (2008) describes production of beta-lactams in yeast.

Recovery of the fermentation product

For the recovery of the fermentation product existing technologies are used. For different fermentation products different recovery processes are appropriate. Existing methods of recovering ethanol from aqueous mixtures commonly use fractionation and adsorption techniques. For example, a beer still can be used to process a fermented product, which contains ethanol in an aqueous mixture, to produce an enriched ethanol-containing mixture that is then subjected to fractionation (e.g., fractional distillation or other like techniques). Next, the fractions containing the highest concentrations of ethanol can be passed through an adsorber to remove most, if not all, of the remaining water from the ethanol.

The following examples illustrate the invention:

EXAMPLES

Unless indicated otherwise, the methods described in here are standard biochemical techniques. Examples of suitable general methodology textbooks include
5 Sambrook et al., Molecular Cloning, a Laboratory Manual (1989) and Ausubel et al.,
Current
Protocols in Molecular Biology (1995), John Wiley & Sons, Inc.

Medium composition

10 Growth experiments: *Saccharomyces cerevisiae* strains are grown on medium
having the following composition: 0.67% (w/v) yeast nitrogen base or synthetic medium
(Verduyn et al., Yeast 8:501-517, 1992) and glucose, arabinose, mannose, galactose or
xylose, or a combination of these substrates, at varying concentrations (see examples for
specific details; concentrations in % weight over volume (w/v)). For agar plates the medium
15 is supplemented with 2% (w/v) bacteriological agar.

Ethanol production

Pre-cultures were prepared by inoculating 25 ml Verduyn-medium (Verduyn et al.,
Yeast 8:501-517, 1992) supplemented with 2% glucose in a 100 ml shake flask with a
20 frozen stock culture or a single colony from an agar plate. After incubation at 30°C in an
orbital shaker (280 rpm) for approximately 24 hours, the culture was harvested and used
for determination of CO₂ evolution and ethanol production experiments.

Cultivations for ethanol production were performed at 30°C in 100 ml synthetic
model medium (Verduyn-medium (Verduyn et al., Yeast 8:501-517, 1992) supplemented
25 with 5% glucose, 5% xylose, 3.5% arabinose, 1% galactose and 0.5% mannose) in the
BAM (Biological Activity Monitor, Halotec, the Netherlands). The pH of the medium was
adjusted to 4.2 with 2 M NaOH/H₂SO₄ prior to sterilisation. The synthetic medium for
anaerobic cultivation was supplemented with 0.01 g l⁻¹ ergosterol and 0.42 g l⁻¹ Tween 80
30 dissolved in ethanol (Andreasen and Stier. J. Cell Physiol. 41:23-36, 1953; and Andreasen
and Stier. J. Cell Physiol. 43:271-281, 1954). The medium was inoculated at an initial
OD₆₀₀ of approximately 2. Cultures were stirred by a magnetic stirrer. Anaerobic
conditions developed rapidly during fermentation as the culture was not aerated. CO₂
production was monitored constantly. Sugar conversion and product formation (ethanol,

glycerol) was analyzed by NMR. Growth was monitored by following optical density of the culture at 600nm on a LKB Ultrospec K spectrophotometer.

Transformation of *S. cerevisiae*

5 Transformation of *S. cerevisiae* was done as described by Gietz and Woods (2002; Transformation of the yeast by the LiAc/SS carrier DNA/PEG method. Methods in Enzymology 350: 87-96).

Colony PCR

10 A single colony isolate was picked with a plastic toothpick and resuspended in 50µl milliQ water. The sample was incubated for 10 minutes at 99°C. 5µl of the incubated sample was used as a template for the PCR reaction, using Phusion® DNA polymerase (Finnzymes) according to the instructions provided by the supplier.

15 PCR reaction conditions:

step 1 3' 98°C

step 2 10" 98°C

step 3 15" 58°C repeat step 2 to 4 for 30 cycles

step 4 30" 72°C

20 step 5 4' 72°C

step 6 30" 20°C

Chromosomal DNA isolation

25 Yeast cells were grown in YEP-medium containing 2% glucose, in a rotary shaker (overnight, at 30°C and 280 rpm). 1.5 ml of these cultures were transferred to an Eppendorf tube and centrifuged for 1 minute at maximum speed. The supernatant was decanted and the pellet was resuspended in 200 µl of YCPS (0.1% SB3-14 (Sigma Aldrich, the Netherlands) in 10 mM Tris.HCl pH 7.5; 1 mM EDTA) and 1 µl RNase (20 mg/ml RNase A from bovine pancreas, Sigma, the Netherlands). The cell suspension was
30 incubated for 10 minutes at 65°C. The suspension was centrifuged in an Eppendorf centrifuge for 1 minute at 7000 rpm. The supernatant was discarded. The pellet was carefully dissolved in 200 µl CLS (25mM EDTA, 2% SDS) and 1µl RNase A. After incubation at 65°C for 10 minutes, the suspension was cooled on ice. After addition of 70

µl PPS (10M ammonium acetate) the solutions were thoroughly mixed on a Vortex mixer. After centrifugation (5 minutes in Eppendorf centrifuge at maximum speed), the supernatant was mixed with 200 µl ice-cold isopropanol. The DNA readily precipitated and was pelleted by centrifugation (5 minutes, maximum speed). The pellet was washed with
5 400 µl ice-cold 70% ethanol. The pellet was dried at room temperature and dissolved in 50 µl TE (10 mM Tris.HCl pH7.5, 1 mM EDTA).

Yeast application test on real hydrolysates

Dilute acid pretreated samples of corn stover were enzymatically hydrolyzed by
10 using an experimental broad spectrum cellulase preparation at 60°C for 3 days (72 hours). The pH at the start of the hydrolysis was 5.0. The dry matter content at the start of the hydrolysis was 10 and 20% w/w. After hydrolysis (72 hrs), the samples were allowed to be cooled to room temperature. The pH was adjusted to 5.5 using 10% NaOH. Subsequently, 1 milliliter of a 200 gram per liter (NH₄)₂SO₄ and 1 milliliter of 100 gram per liter KH₂PO₄
15 was added. Finally, yeast samples were added corresponding to a yeast dry matter content of 1 or 2 gram yeast per kilogram hydrolysate at 10 or 20% w/w, respectively. The CO₂ evolution in time was followed using the AFM (Alcohol Fermentation Monitor; HaloteC Instruments BV, Veenendaal, the Netherlands). Experiments were performed in at least triplicate, for 72 hours at 33°C. One of these is sampled at regular intervals in order to be
20 able to analyze ethanol formation and residual sugar concentrations. These data can be used to calculate fermentation yields. The broth of the other two experiments is not sampled. Instead, at the end of the fermentation the broth is distilled using a Buchi K-355 distillation unit at 45% steam for 15 minutes. The alcohol produced is being determined using an Anton Paar DMA 5000 density meter (Anton Paar Benelux BVBA, Dongen, the
25 Netherlands).

Example 1

Growth rate improvements on xylose and arabinose in Sequential Batch Reactor (SBR) cultivations

30

Strain *S. cerevisiae* BIE252 was grown in a Sequential Batch Reactor cultivation system according to a modified protocol described in WO 2009/112472 in order to improve

the growth rates on C5-sugars. Anaerobic cultivation was carried out at 32°C in 5-L laboratory fermentors with a working volume of 2-L. The pH was maintained at 4.0 by automatic addition of 2 M KOH. The cultures were stirred at 100 rpm and sparged with 0.01 vvm air, while 2 nL/min N₂ was used in the headspace as carrier gas for the MS off-gas measurements. The cultures were performed in media containing different C6- and C5-sugars composition. When the C-source was fully depleted, as indicated by the CO₂ level, a new cycle of batch cultivation was initiated by replacement of ~95-99% of the culture with fresh synthetic medium containing the appropriate C-source. The following media were used: (1) mixed sugars medium: 10 g/l glucose, 10 g/l xylose, 7 g/l arabinose, 2 g/l galactose and 1 g/l mannose; (2) arabinose medium: 27 g/l arabinose and 3 g/l xylose and (3) xylose medium: 27 g/l xylose and 3 g/l arabinose. After completion of the batch growth in medium (1), media (2) and (3) were alternated and that sequence of cultivation in media (2) and (3) was repeated for six cycles. Growth in medium (1) was repeated after cycle three to confirm that the culture was still able to utilize the C6-sugars as fast as in the beginning of the SBR cultivation. For each run, the maximum specific growth rate (μ_{\max}) was estimated from the CO₂ profile in the exponential growth phase. After six cycles (90-100 generations) of SBR cultivation both the growth rates of strain *S. cerevisiae* BIE252 on xylose and arabinose were almost doubled, as depicted in figure 1. The growth rate on xylose was increased from 0.1 h⁻¹ to 0.19 h⁻¹, while the growth rate on arabinose from 0.066 h⁻¹ to 0.12 h⁻¹. After the last cycle of evolutionary engineering in the SBR system, a sample was taken for single colony isolation.

Example 2

Single colony isolation

The following approach was executed in order to select a strain which has gained improved growth on xylose and arabinose as a sole carbon source without losing its ability to utilize C6-sugars (glucose and galactose). Initially, a sample was taken from the fermentor after the last cycle of evolutionary engineering in the SBR cultivation system. The broth sample (SBR culture) was streaked on YEPD-agar and incubated for 48 hours at 30°C. Nine single colony isolates of the SBR culture were re-streaked on YEPD-agar and incubated for 48 hours at 30°C. Subsequently, a pre-culture was done for each single

colony on YEP liquid medium supplemented with 2% glucose. The nine cultures were incubated overnight at 30°C and 280 rpm. All nine cultures were tested for their performance in the BAM (Biological Activity Monitor) system (Halotec, Veenendaal, the Netherlands).

5

Example 3

Performance test in BAM

10

In order to test the performance of the nine single colony isolates, the strains were inoculated in Verduyn medium, supplemented with 2% glucose. As control, strain *S. cerevisiae* BIE252, the original strain before adaptive evolution in the SBR cultivation system, was included. After overnight incubation at 30°C and 280 rpm in a rotary shaker, cells were harvested by centrifugation and cultivations for CO₂ production were performed at 33°C and pH 4.2 in 100 ml Verduyn medium supplemented with 50 g/l glucose, 50 g/l xylose, 35 g/l arabinose, 10 g/l galactose and 5 g/l mannose in the BAM. The CO₂ production was constantly monitored at intervals, and samples were taken for analysis (optical density at 600 nm using a spectrophotometer; ethanol, glycerol and residual sugars by NMR). Among the nine single colony isolates, colony number five, designated as strain BIE272, performed significantly better than strain BIE252. The results of the BAM experiment of strains BIE252 and BIE272 are shown in figure 2 and 3, respectively. The performance of strain BIE252 shows that glucose was consumed readily after the start of the fermentation experiment (figure 2). Subsequently, galactose, mannose, arabinose and xylose were being co-fermented. After approximately 30 hours, galactose and mannose were consumed, while after approximately 72 hours, arabinose and xylose were also fully consumed. Both arabinose and xylose contributed significantly to the formation of CO₂ and ethanol. However, arabinose and xylose consumption, and thus ethanol and CO₂ production slowed down after the exhaustion of galactose and mannose. Strain BIE272 performed generally much better (figure 3). Immediately after glucose exhaustion within 10 hours, galactose, mannose, arabinose and xylose were co-fermented rapidly and efficiently. The C5-sugars utilization rates were higher than BIE252. Even after galactose and mannose exhaustion, both pentoses (arabinose and xylose) were co-fermented

20
25
30

rapidly, both contributing to CO₂-production and ethanol formation. The fermentation of all sugars was completed after approximately 48 hours, in case of strain BIE272. This resulted in a higher cumulative CO₂ and ethanol production and a higher productivity (g ethanol/L/h) of strain BIE272 compared to strain BIE252.

5

Example 4

10

Performance test in real hydrolysates

The performance test in real hydrolysates was performed using strains BIE252 and BIE272, which were cultured overnight in shake flasks containing YEP medium supplemented with 2% glucose. The cells were harvested by centrifugation and re-suspended at a concentration of 50 grams dry matter per liter. Pretreated corn stover (pCS) at 10 and 20% dry matter were used as feedstocks. The hydrolysis and fermentation were performed as described in the materials and methods section. The results are presented in figures 4 (xylose consumption of BIE252 and BIE272 at 10 and 20% dry matter pCS), 5 (arabinose consumption of BIE252 and BIE272 at 10 and 20% dry matter pCS), 6 (ethanol production of BIE252 and BIE272 at 10 and 20% dry matter pCS) and 7 (CO₂ production of BIE252 and BIE272 at 10 and 20% dry matter pCS). For both strains glucose was readily consumed after the start of the fermentation and the consumption profiles were similar. Glucose was fully consumed after approximately 12 hours at 10% dry matter pCS and after 24 hours at 20% dry matter pCS (data not shown). Strain BIE252 consumed xylose fully in 96 hours at 10% dry matter pCS, while there was still approximately 9 g/l xylose left after 168 hours at 20% dry matter pCS. In case of strain BIE272, xylose was fully consumed in 72 hours at 10% dry matter pCS, while there was less xylose left (5 g/l) compared to strain BIE252 after 168 hours at 20% dry matter pCS (figure 4). Arabinose was not fully consumed in case of both strains at both 10% and 20% dry matter. However, in case of strain BIE272, less residual arabinose was measured after 168 h, compared to BIE252 (figure 5). Ethanol titer and cumulative CO₂ production were higher in case of strain BIE272 at both 10 % and 20% dry matter hydrolysates tested (figures 6 and 7). The overall performance of strain BIE272 at 20% dry matter pCS is

presented in figure 8. In the table below the yields of the fermentation are calculated, on basis of the sugars liberated at the end of the hydrolysis and the amount of ethanol that was produced at the end of the fermentation.

- 5 Table 3 *Total sugar released (g/l), ethanol produced (g/l) and ethanol yield (g ethanol/g sugar) of AFM fermentation of BIE252 and BIE272 at 10% and 20% dry matter pretreated corn stover (pCS)*

	Total sugar* (g/l)		Produced EtOH (g/l)		EtOH yield (g ethanol/g sugar)	
	BIE252	BIE272	BIE252	BIE272	BIE252	BIE272
Lignocellulosic feedstock						
10% dry matter pCS	55.9	58.4	27.3	27.7	0.49	0.47
20% dry matter pCS	108.6	113.2	45.9	48.5	0.42	0.43

*(released, monomeric sugar at start fermentation)

10

Example 5

Stability test of strain BIE272

15

5.1 Inoculation procedure

In order to test the stability of the strain BIE272, 25 µl of a glycerol stock of strain BIE272 was used to inoculated in duplicate flasks in 25 ml of YEP 2% glucose. The optical density of the culture was measured at 600 nm. The cultures were incubated overnight at
20 30°C and 280 rpm in a rotary shaker.

After the overnight incubation, the optical densities were determined. Based on the OD⁶⁰⁰ values before and after incubation, the number of generations made during the incubations was calculated. 25 µl of the overnight cultures were used to inoculate a flask containing 25 ml of fresh medium.

5 The cultures were incubated again under the same conditions as described above. This procedure was repeated until a culture was obtained in which the cells were grown for about 50 generations.

YEP medium supplemented with 2% glucose was chosen, because under these conditions no selection pressure is applied for maintaining the introduced genes and structural variations in strain BIE272, needed for the conversion of arabinose and xylose.

5.2 Isolation of single colonies after every 9 – 10 generations of growth

In addition to the inoculation procedure described above, the following was applied.

From a glycerol stock of strain BIE272, a loopful of cells was streaked on a YEPD agar plate. The plate was incubated at 30°C for two days. After two days of incubation, single colonies were visible.

From every overnight culture of section 5.1 (i.e. after 10, 19, 28, 37 or 46 generations), a loopful of cells was streaked on YEPD agar plates in order to generate single colony isolates. The performance of six such colonies was assessed in a growth experiment. To this end the colonies were inoculated in Verduyn medium containing 2% xylose, in 24 well microplates. The plates were incubated for 6 days at 30°C and 550 rpm in an Infors microplate shaker. The optical density was read at intervals and compared to the initial single colony isolates, directly streaked from the glycerol stock.

After 48 hours of growth, there was in some cases a difference in growth on xylose between the reference strain BIE272 and the single colony isolates of BIE272 generated after 10, 19, 28, 37 or 46 generations. If the growth of the latter colonies was lagging behind, as indicated by an optical density of 75% or lower as compared to the reference strains, growth was marked as such. The results of the growth at 48 hours of incubation are presented in the figure 9.

30

The results indicate that after about 46 generations, all twelve colonies (six colonies in duplicate) show the same growth phenotype as the colonies of reference strain BIE272 which were isolated directly from the glycerol stock.

After prolonged growth (i.e. beyond 48 hours, up to 6 days, vide supra) all colonies grew confluent, i.e. a maximal optical density was obtained.

5 The number of generations on the YEPD agar plate was not taken into account. The growth of the single cell, after shake flask incubation and streaked on an agar plate, will grow into a colony on the YEPD agar plate within two days at 30°C. A yeast colony typically has about $3 \cdot 10^5$ - 10^6 cells (Runge, K.W. (2006) Telomeres and Aging in the Yeast Model System. Pages 191-206. In: Handbook of models for human ageing. Edited
10 by P. Michael Conn. ISBN 13: 978-0-12-369391-4). So, starting from one cell to a fully grown colony takes about 18 to 20 divisions, and thus 18 – 20 additional generations after the culturing in the shake flasks had taken place.

From the results it may be concluded that after more than 50 generations of
15 culturing all colonies exhibited the same growth phenotype as the initial glycerol stock of strain BIE272. Hence, the strain is phenotypically stable.

5.3 Q-PCR analysis of xylose isomerase genes

In addition to the phenotypic analysis described in section 5.2, a quantitative PCR
20 experiment (Q-PCR) was done in order to assess the copy number of the xylose isomerase genes present in strain BIE272 prior to the cultivation experiment, as well as from the cultures after overnight growth.

After overnight growth, small aliquots of the cultures were used to inoculate fresh
25 medium or to isolate single colonies (see section 5.1 and 5.2). The remainder of the culture was used for the isolation of chromosomal DNA, for Q-PCR purposes.

The Q-PCR analysis was performed using the Bio-Rad iCycler iQ system from Bio-Rad (Bio-Rad Laboratories, Hercules, CA, USA). The iQ SYBR Green Supermix (Bio-Rad)
30 was used. Experiments were set up as suggested in the manual of the provider.

The stability of the strain BIE272 was assessed by determining the copy number of the *xyIA* gene encoding xylose isomerase. As a reference single copy gene, the *ACT1* gene was chosen.

The primers for the detection of the genes *xyIA* and *ACT1* are summarized in the table below.

5 Table 4 *Primers used for amplification in the Q-PCR experiment*

Gene of interest	Forward primer	Reverse primer
<i>xyIA</i>	SEQ ID NO 1	SEQ ID NO 2
<i>ACT1</i>	SEQ ID NO 3	SEQ ID NO 4

The Q-PCR conditions were as follows:

- 10 1) 95°C for 3 min
- for 40 cycli, steps 2 - 4
- 2) 95°C for 10 sec
- 3) 58°C for 45 sec
- 15 4) 72°C for 45 sec
- 5) 65°C for 10 sec
- 6) Increase of temperature with 0,5°C per 10 sec to 95°C

20 The melting curve is being determined by starting to measure fluorescence at 65°C for 10 seconds. The temperature is increased every 10 seconds with 0.5°C, until a temperature of 95°C is reached. From the reads, the copy number of the gene of interest was calculated and/or estimated. The results are presented in the table below.

productivity. Moreover, this strain showed an excellent conversion of mixtures of hexose and pentose sugars, either added to mineral media or present in lignocellulosic hydrolysates, exceeding the performance of pentose fermenting strains known to date.

5 In addition, it was shown that the selected strains were genetically and phenotypically stable (Example 5). After cultivation on a non-selective medium for about 50 generations, single colony isolates obtained prior and after the cultivations exhibit the same relevant phenotype and genotype.

10 In order to investigate which genetic variations, such as SNPs (single nucleotide polymorphisms), DIPs (deletion/insertion polymorphisms), deletions, amplifications and rearrangements in the genome of strains BIE252 and BIE272 have contributed to the observed phenotypes (improved ethanol yield and productivity in mixed sugar substrates), we resequenced the genomic DNA of the transformants, using the art known as Solexa®
15 technology, using the Illumina® Genome Analyzer.

To this end, chromosomal DNA was isolated from the strains BIE252 and BIE272 from YEP 2% glucose cultures, previously grown overnight at 280 rpm and 30°C. The DNA was sent to ServiceXS (Leiden, the Netherlands) in case of BIE252 and to BaseClear
20 (Leiden, the Netherlands) in case of BIE272, for resequencing using the Illumina® Genome Analyzer (50 and 75 bp reads respectively, paired end sequencing in both cases).

Different sequence yields (i.e. the number of reads) were obtained in the sequencing
25 analysis, mostly related to the state of development in the technology. In all strains, millions of sequence reads were obtained, for example for BIE272, 25 million reads of 75 nucleotides length were obtained, being 1.8 billion nucleotides.

Sequence reads were obtained from the Illumina GAI machine and a quality
30 filtering was applied based on (Phred) quality scores. In addition, low quality and ambiguous nucleotides were trimmed off from the remaining reads.

Using software such as NextGene (SoftGenetics LLC, State College, PA 16803, USA) and CLC Genomics workbench v4.5 (CLCbio, Aarhus, Denmark), the sequencing

reads were aligned using the S288c as a template in case of strain BIE252. In case of BIE272, alignment was done using CLC Genomics workbench v4.5 to the template of S288c and in a second analysis, an assembly of previously obtained sequencing data of strain BIE104 was used as a reference template.

- 5 The average read depth per base that was obtained after mapping the reads to the template:

BIE252: read depth of 121

BIE272: read depth of 135

10

Mutations (single nucleotide polymorphisms and insertion/deletions up to 30 bp) were detected and summarised in a mutation report. The mutations called in the different strains were compared to each other to identify the unique variations between the strains.

- 15 Every entry of the mutation report was checked manually, in order to rule out the possibility of misalignment of the sequence reads, or mutation calls due to sequencing errors or calls for which sequence coverage was very low. False positive mutations were removed from the mutation report.

Table 7 presents an overview of the SNPs that were observed.

Table 7 *Single nucleotide polymorphisms (SNPs) in strains BIE252 and BIE272, relative to strain BIE104. For reference, the structural variations in strain BIE201 have been listed as well.*

ORF	Gene	Chromosome	SNP	Amino acid change	BIE 104	BIE 201	BIE 252	BIE 272
YDR160w	SSY1	IV	G>T	Glu455STOP (>stop codon)	G	T	T	T
YJR154w		X	A>G	Asp171Gly	A	G	G	G
YMR168c	CEP3	XIII	A>G	Ser396Gly	A	G	G	G
YML051w	GAL80	XIII	A>C	Thr146Pro	A	C	C	C
YGL167C	PMR1	VII	A>G	Tyr38Cys	A	A	A	G

5

In addition, using the coverage plots indicating the read depth at every single nucleotide position of the genome, searches were done for areas in the genome that are over- or underrepresented. Such structural variations comprise deletions, duplications, copy number variations, insertions, inversions and translocations.

10

Figure 21 sets out an example of an increased coverage of the *PMA1* terminator region. This terminator has been used in several constructs for the overexpression of the genes *araA*, *araB* and *araD* (in plasmid pPWT018 (see PCT/EP2011/056242) as well as the *xyIA*-gene (see EP10160647.3)). As a consequence, multiple copies of the *PMA1* terminator are present in the genome of strains BIE252 and BIE272, resulting in an increased read depth as compared to the surrounding genomic regions of the *PMA1* terminator.

15

Figure 22 sets out another example of a coverage analysis, in this case of the *xyIA* gene encoding xylose isomerase. The normalized read depth of the region consisting of the *xyIA* gene corresponds with a value of 9 to 10, which is in line with the copy number as determined by Q-PCR (see Example 5), while the read depth of the surrounding genomic regions is around 1.

20

Using this method, regions that were over- and underrepresented in strains BIE252 and BIE272 were identified. It was found that the amplification previously observed in strain BIE201 (see PCT/EP2011/056242), based on the information now available, is located on the left arm of chromosome VII, and that left arm is no longer amplified in strains BIE252 and BIE272. The amplification observed on the right arm of chromosome VII in BIE201, comprising the genes *araA*, *araB* and *araD*, are conserved in strains BIE252 and BIE272. The copy number of the arabinose genes in BIE272 was determined as three copies, based on the normalized read depth.

Example 7

Analysis of the chromosome structure of transformants

From resequencing data (see Example 6) it could be inferred that after adaptive evolution, genome variations in the genome have taken place. These genome variations include single nuclear polymorphisms (SNPs), deletion-insertion polymorphisms (DIPs) and larger variations in the structure of the chromosomes, due to events like amplification and translocation. In order to substantiate the latter, CHEF gel electrophoresis was employed.

Contour-clamped homogeneous electric field (CHEF) gel electrophoresis has been used to study the karyotypes of a range of *Saccharomyces cerevisiae* yeast strains, from the untransformed strain BIE104 up to strain BIE272, a strain that is able to ferment pentoses and hexoses rapidly in sugar mixtures.

7.1 CHEF electrophoresis

In order to determine whether the number and the size of the chromosomes were changed, or the composition with respect to certain key genes, CHEF electrophoresis (Clamped Homogeneous Electric Fields electrophoresis; CHEF-DR® III Variable Angle System; Bio-Rad, Hercules, CA 94547, USA) was applied. Agarose plugs of yeast strains (see below) were prepared using the CHEF Yeast Genomic DNA Plug Kit (BioRad) according to the instructions of the supplier. 1% Agarose gels (Pulse Field Agarose, Bio-

Rad) were prepared in 0.5x TBE (Tris-Borate-EDTA) according to the suppliers instructions. Gels were run according to the following settings:

5	Block 1	initial time 60 sec final time 80 sec ratio 1 run time 15 hours
10	Block 2	initial time 90 sec final time 120 sec ratio 1 run time 9 hours

As a marker for size determination of the chromosomes, agarose plugs of strain YNN295 (Bio-Rad) were included in the experiment.

After electrophoresis, gels were stained using ethidiumbromide at a final concentration of 70 µg per litre, for 30 minutes. In figure 10, an example of a stained gel is shown.

In strains BIE104A2P1c, BIE201 and BIE201X9, the size of chromosome VII was increased. Its size was increased to a size close to that of chromosome IV, of about 1500 - 1550 kb.

In strain BIE252 however, the large size of chromosome VII was decreased as compared to BIE201, but still larger than the original size of chromosome VII (as it is in strain BIE104, the untransformed yeast strain). Two chromosomes appeared, with a size of approximately 1375 and 1450 kb respectively. This result corroborates the observation from the resequencing data (Example 6), that the left arm of chromosome VII, which was amplified in strain BIE201, is no longer amplified in strains BIE252 and BIE272. Since the amplifications on the right arm of chromosome VII are still present in BIE252 and BIE272, as in BIE201, the size of chromosome VII is still larger than the original size of chromosome VII (as it appears in BIE104).

In strain BIE272, the situation seems even more complex. The 1600 kb band, representing chromosome IV, is no longer visible. Since on chromosome IV, like on any other chromosome, several indispensable genes are present, the chromosome size presumably has been changed by either fragmentation (i.e. split in two smaller parts) or
5 increase in size (e.g. by amplification).

In addition, the chromosome with a size of approximately 1450 kb has disappeared. The chromosome with a size of approximately 1375 kb is also present in BIE272.

10 One way of identifying how the chromosomes have been rearranged, either by amplification of parts, translocation and/or fragmentation, is to transfer the DNA of the gel by blotting followed by hybridisation with specific probes, which are representative for certain chromosomes.

15 7.2 Hybridisation with specific probes

After staining, gels were blotted onto Amersham Hybond N+ membranes (GE Healthcare Life Sciences, Diegem, Belgium).

In order to be able to identify the nature of the changes in the size of the
20 chromosomes, probes were made for hybridization with the blotted membranes. Probes (see table below) were prepared using the PCR DIG Probe Synthesis Kit (Roche, Almere, the Netherlands) according to the instructions of the supplier.

The following probes were prepared.

Table 8 Primers for amplification of the indicated probes

Probe	Systematic name gene	Forward primer	Reverse primer	Size PCR product (bp)	Chromosome
<i>xyIA</i>		SEQ ID NO 1	SEQ ID NO 2	419	V + other chromosome(s)
<i>ACT1</i>	YFL039c	SEQ ID NO 3	SEQ ID NO 4	392	VI
<i>PNC1</i>	YGL037c	SEQ ID NO 5	SEQ ID NO 6	384	VII

5 Membranes were prehybridized in DIG Easy Hyb Buffer (Roche) according to the instructions of the supplier. The probes were denatured at 99°C for 5 minutes, chilled on ice for 5 minutes, and added to the prehybridized membranes. Hybridization was done overnight at 42°C.

10 Washing of the membranes and blocking of the membranes prior to detection of the hybridized probes were done using the DIG Wash and Block Buffer Set (Roche) according to the instructions of the supplier. The detection was done by incubation with anti-dioxygenin-AP Fab fragments (Roche) followed by the addition of detection reagents using the CDP-Star ready-to-use kit (Roche). Detection of the chemiluminescent signals were performed using the Bio-Rad Chemidoc XRS+ System, using the appropriate settings provided by the Chemidoc apparatus.

15

The results are shown in figures 11 (*PNC1*), 12a (*ACT1*) and 12b (*xyIA*).

20 *PNC1* is located on the left arm of chromosome VII, and thus considered to be a specific probe for this chromosome. Hybridization resulted in a band of the expected size in case of strain BIE104, the untransformed strain. In strain BIE104A2P1 (designated in figure 11 as BIE104A2P1a), the same band is observed. In addition, a second more faint and smaller band is observed. The corresponding band is absent in the ethidiumbromide stained gel (figure 10). Hence, this signal is probably an electrophoresis (trapping) and or a hybridization artefact.

In strains BIE104A2P1c, BIE201 and BIE201X9, an increase in the size of chromosome VII is observed, as was apparent from the ethidiumbromide stained gel (figure 10). Its size was increased to a size close to that of chromosome IV, of about 1500 - 1550 kb.

5 In strains BIE252 and BIE272, a band of smaller size hybridized. The size is approximately 1375 kb. In BIE252, a second, larger but less intense band is observed, which is absent in BIE272. This band may be the result of an electrophoresis (trapping) and/or a hybridization artefact. Alternatively, it is a larger form of the same chromosome. Since the agarose plugs were prepared from a purified single colony isolate, this is not
10 very likely.

Based on intensity comparisons, these results corroborate the observations described above that the left arm of chromome VII is no longer amplified in strains BIE252 and BIE272. From the intensity of the bands, it can be deduced that the copy number of the *PNC1* gene is increased in strains BIE104A2P1c, BIE201 and BIE201X9 relative to
15 BIE104 and BIE104A2P1(a), but decreased again in strains BIE252 and BIE272.

The *ACT1*-gene is located on chromosome VI and not expected to be amplified. Hence, this probe serves as a control. Indeed, a single band was observed after hybridisation (see figure 12, panel a) in all strains tested.
20

The *xyIA*-gene was integrated as a single copy gene on chromosome V in strain BIE201X9. In strain BIE252, extra copies were introduced in the Ty1 loci of strain BIE201X9, followed by adaptive evolution, finally yielding strain BIE272.

As expected, in strain BIE201X9, one single chromosome hybridizes with the *xyIA*-
25 probe. The band observed on the autoradiogram has a size of approximately 600 kb, which is the right size. Please note that in this genetic background the resolution between chromosomes V and VIII is less pronounced as is the case in the marker strain, YNN295 (see figure 10).

The same band is observed in the strains BIE252 and BIE272.

30 In strain BIE252, at least one extra band is observed which has a high molecular weight. The band is around 2Mb, which suggests that the integration of the extra copies of the *xyIA*-gene have taken place on chromosome XII, which is the largest chromosome. The intensity of the band is high in comparison to the intensity of the band corresponding

to the integrated *xyIA*-gene on chromosome V. From the ratio of the intensities of both bands, the copy number may be inferred. It may be concluded that multiple copies of the *xyIA*-gene have been integrated in chromosome XII. The exact determination of the copy number requires more elaborate work, such as the application of different concentrations of DNA, and the application of densitometry (to quantify the DNA by measuring the density of silver grains on the photograph) of autoradiograms with several exposure times, in order to assure that the readings obtained are within the linear range of the film. In addition, although an increase in the signal intensity may suggest an increase of the copy number of a certain gene, other factors may also influence the signal strength, like the amount of DNA applied on the gel, blotting efficiency, detection saturation, and the like.

Two extra, fainter bands were observed in strain BIE252, one being slightly higher and one being slightly smaller than the most intense band. Presumably, these bands are electrophoresis artefacts caused by trapping of DNA.

In strain BIE272, the size of the strongest hybridizing band has decreased in size as compared to strain BIE252, suggesting a structural variation of chromosome XII (figure 12b). This is also observed in the ethidiumbromide stained gel (figure 10). Also in case of strain BIE272, "shadowbands" occur which are most likely due to trapping of chromosomes during electrophoresis. The intensity of the band corresponding to chromosome XII is several times higher than the intensity of the band corresponding to chromosome V, suggesting that multiple copies of the *xyIA*-gene are still present in strain BIE272, as was observed for strain BIE252. From the Q-PCR experiments it was concluded that around 9 copies of the *xyIA*-gene are present in strain BIE272 (Example 5, section 5.3).

In the stained gel (figure 10), the band corresponding to chromosome IV is no longer visible in strain BIE272, suggesting a recombinational event in which chromosome IV is involved.

In conclusion, the results of Example 7 clearly indicate that structural variations leading to shifts in chromosome sizes have taken place. More elaborate studies will be needed in order to be able to conclude which (parts of) chromosomes were involved in these processes.

Example 8

Performance test of the strains BIE104, BIE201, BIE252 and BIE272

5 The strains BIE104, BIE201, BIE252 and BIE272 have different characteristics with respect to their genetic constitution and their performance in sugar hydrolysates. The table below illustrates how the strains relate to each other.

Table 9 Relevant strains in the strain lineage of BIE272

10

Strain	Parent strain	Genes introduced	Adaptive evolution
BIE104	Wild-type strain		
BIE201	BIE104	<i>TAL1, TKL1, RPE1, RKI1, araA, araB, araD</i>	In shake flasks
BIE252	BIE201	<i>xyIA, XKS1</i>	In shake flasks
BIE272	BIE252	<i>None</i>	In fermentors

In order to illustrate the improvements with respect to the conversion of sugar mixtures that were achieved during the development of strain BIE272, a performance test was executed in the AFM (Halotec, Vennendaal, the Netherlands). To this end, single colony isolates of the strains BIE104, BIE201, BIE252 and BIE272 were cultivated in 100 ml Verduyn medium with 2% glucose as the carbon source, for 24 hours at 30°C and 280 rpm.

The cells were harvested by centrifugation and cultivations for CO₂ production were performed at 33°C and pH 4.2 in 200 ml Verduyn medium supplemented with 50 g/l glucose, 50 g/l xylose, 35 g/l arabinose, 10 g/l galactose and 5 g/l mannose in the AFM (temperature 33°C, stirrer speed 250 rpm, fermentation time minimally 72 hours). The CO₂ production was constantly monitored at intervals, and samples were taken for analysis (optical density at 600 nm using a spectrophotometer; ethanol, glycerol and residual sugars by NMR).

25

The CO₂ evolution profiles are set out in figure 13, 14, 15 and 16. The total amount of CO₂ that was produced during the experiment, which lasted 71 hours and 25 minutes, is set out in table 10.

- 5 Table 10 *Total amount of CO₂ produced by strains BIE104, BIE201, BIE252 and BIE272 in Verduyn medium containing 50 g/l glucose, 50 g/l xylose, 35 g/l arabinose, 10 g/l galactose and 5 g/l mannose, in about 72 hours.*

Strain	Amount of CO ₂ produced (ml)
BIE104	2478
BIE201	4208
BIE252	5108
BIE272	6066

10

In figure 13, all four strains are shown in one graph. In the figures 14, 15 and 16, a pairwise comparison of two strains at the time is made.

15 In figure 14, strain BIE104 is compared to BIE201. From this sugar mixture, strain BIE104 can only ferment glucose and mannose, while strain BIE201 ferments glucose, mannose, galactose and arabinose. This resulted in a different CO₂ production rate profile (figure 14) and an increase of 70% in the total amount of CO₂ produced.

20 In figure 15, strains BIE201 and BIE252 were compared. The ability to convert xylose, which is new in strain BIE252 relative to BIE201, next to arabinose and the hexoses, yielded a higher CO₂ production rate (figure 15) and a higher total CO₂ production (+ 106% relative to BIE104 and + 21% relative to BIE201).

25 In figure 16, it is shown that strain BIE272 showed a higher conversion rate of sugars into ethanol and carbondioxide. At the end of the experiment (71 hours and 25 minutes), strain BIE272 had produced 145% more CO₂ relative to BIE104, and 19% more relative to strain BIE252.

In figures 17, 18, 19 and 20, the sugar consumption and ethanol formation are set out for strains BIE104, BIE201, BIE252 and BIE272 respectively.

Strain BIE104 (figure 17) only consumes the glucose and mannose. The strain is not capable of converting the pentoses xylose and arabinose, since this is a non-transformed strain. Also galactose is not converted, for under fermentative conditions, the energy charge was probably too low to allow synthesis of the Leloir proteins for galactose utilization, as described by van den Brink et al (van den Brink et al (2009) Energetic limits to metabolic flexibility: responses of *Saccharomyces cerevisiae* to glucose-galactose transitions. *Microbiology* 155(Pt 4):1340-50). The yield on dosed sugars amounts 0.14 grams of ethanol per gram sugar, in 72 hours.

Strain BIE201 (figure 18) is capable of converting glucose, mannose, arabinose and galactose. Xylose is not fermented, since the pathway for xylose fermentation was not introduced in this strain. In 72 hours, arabinose is almost completely fermented in this experiment, while the hexoses, including galactose, were completely converted before 36 hours after the start of the experiment. The yield on dosed sugars amounts 0.25 grams of ethanol per gram sugar, in 72 hours.

Strain BIE252 (figure 19) is capable of fermenting glucose, xylose, mannose, arabinose and galactose. In 72 hours, xylose and arabinose are almost completely fermented in this experiment, while the hexoses glucose, mannose and galactose were already exhausted before 36 hours after the start of the experiment. The yield on dosed sugars amounts 0.36 grams of ethanol per gram sugar, in 72 hours.

Strain BIE272 (figure 20) is capable of fermenting glucose, xylose, mannose, arabinose and galactose. In 72 hours, all sugars were fermented rapidly and completely, except for arabinose, which was fermented almost completely. The yield on dosed sugars amounts 0.42 grams of ethanol per gram sugar, in 72 hours.

The fermentation characteristics of strains BIE104, BIE201, BIE252 and BIE272 are summarized in the table below.

Table 11 Fermentation characteristics of strains BIE104, BIE201, BIE252 and BIE272 in Verduyn medium containing 5% glucose, 5% xylose, 3.5% arabinose, 1% galactose and 0.5 % mannose. The yield is expressed as grams of ethanol per gram dosed sugar, calculated over the whole fermentation (72 hours). The productivity is expressed as grams of ethanol per liter per hour, calculated over the time period indicated in the table.

Strain	Yield (g EtOH/g sugar)	Productivity 0 – 24 h (g EtOH/l, h)	Productivity 24 - 48 h (g EtOH/l, h)	Productivity 0 – 48 h (g EtOH/l, h)	Productivity 0 – 72 h (g EtOH/l, h)
BIE104	0.14	0.91	0.03	0.46	0.30
BIE201	0.25	1.01	0.47	0.74	0.53
BIE252	0.36	1.13	0.90	1.01	0.75
BIE272	0.42	1.69	0.91	1.30	0.90

As is clear from table 11, not only the yield was increased in case of strain BIE272 in the overall process, but also the productivity at several time intervals. In the first 24 hours of the fermentation, the productivity was increased by 50% in case of strain BIE272, as compared to strain BIE252.

In addition to the increased yield and productivity of strain BIE272 relative to the other strains tested, the consumption rate of xylose and arabinose, in the presence of glucose, was increased as well.

At time point 6.2 hours, the culture of strain BIE252 contained 19.8 g glucose per liter (110 mM). At time point 23.3 hours, the glucose concentration was 7.1 g/l (39 mM). Both concentrations are repressing concentrations (glucose or catabolite repression), at which the use of other carbon sources than glucose is actively prevented. Surprisingly, during this time period, the arabinose concentration decreased from 30.8 g/l to 24.3 g/l and the xylose concentration decreased from 42.8 g/l to 32.9 g/l. So, co-consumption of glucose, xylose and arabinose took place under these conditions.

In case of strain BIE272, the following decreases were observed from time point 6.2 hours until 23.3 hours: glucose from 22.4 g/l to 4.8 g/l (from 124 mM to 27 mM),

arabinose from 33.7 g/l to 22.4 g/l and xylose from 46.2 g/l to 22.6 g/l. Also in this strain, co-consumption of xylose, arabinose and glucose took place.

The consumption rates of arabinose and xylose, in the presence of glucose, were calculated as grams of pentose consumed per hour per gram dry yeast biomass. The values are presented in table 12.

Table 12 Consumption rates of arabinose and xylose in the presence of glucose of strains BIE252 and BIE272

Strain	Arabinose consumption rate (g/g,h)	Xylose consumption rate (g/g,h)
BIE252	0.13	0.19
BIE272	0.20	0.41

10

The ability to consume the pentose arabinose and xylose in the presence of repressing glucose concentrations was further improved in strain BIE272.

CLAIMS

1. Yeast cell belonging to the genus *Saccharomyces* having introduced into its genome at
5 least one *xylA* gene and at least one of each of *araA*, *araB* and *araD* genes and that is
capable of consuming a mixed sugar mixture comprising glucose, xylose and
arabinose, wherein the cell co-consumes glucose and arabinose, has genetic
variations obtained during adaptive evolution and has a specific xylose consumption
rate in the presence of glucose that is 0.25 g xylose/h, g DM or more.
10
2. Yeast cell according to claim 1, wherein the yeast cell is *Saccharomyces cerevisiae*.
3. Yeast cell according to claim 1 or 2, wherein the specific xylose consumption rate in
the presence of glucose is 0.35 g xylose/h, g DM or more.
15
4. Yeast cell according to any of claims 1 to 3, wherein the specific xylose consumption
rate in the presence of glucose is 0.25-0.60 g xylose/h, g DM.
5. Yeast cell according to any of claims 1 to 4, wherein the copy number of the *araA*,
20 *araB* and *araD* genes is three or four each.
6. Yeast cell according to any of claims 1 to 5, wherein the copy number of *xylA* is about
9 or 10.
- 25 7. Yeast cell according to any of claims 1 to 6, having one or more of the single
nucleotide polymorphism chosen from the group consisting of mutations G1363T in the
SSY1 gene, A512T in *YJR154w* gene, A1186G in *CEP3* gene, A436C in *GAL80* gene
and A113G in *PMR1* gene.
- 30 8. Yeast cell according to claim 6, which has a single polymorphism A436C in *GAL80*
gene.

9. Yeast cell according to claim 6, which also has a single nucleotide polymorphism A1186G in *CEP3* gene.
10. Yeast cell according to claim 6 or 7, which has a single nucleotide polymorphism A113G in *PMR1* gene.
11. Yeast cell according to any of claims 1-10, wherein the yeast cell has a yield of 0.40 g ethanol/g sugars or more or about 0.42 g ethanol/g sugars.
12. Yeast cell according to any of claims 1-10, wherein the yeast cell has a productivity of 1.20 g EtOH/l, h or more or about 1.69 g EtOH/l, h, measured in the interval of 0-24h after start of the fermentation.
13. A polypeptide having the sequence SEQ ID NO: 8 and variant polypeptides thereof wherein one or more of the other positions may have mutation of the aminoacid with an aminoacid that is an existing conserved aminoacid in the SPCA family.
14. Process for the production of one or more fermentation product from a sugar composition comprising glucose, galactose, arabinose and xylose, wherein the sugar composition is fermented with a yeast cell according to any of claims 1 to 8.
15. Process according to claim 10, wherein the sugar composition is produced from lignocellulosic material by:
- a) pretreatment of one or more lignocellulosic material to produce pretreated lignocellulosic material;
 - b) enzymatic treatment of the pretreated lignocellulosic material to produce the sugar composition.
16. Process according to claims 10 or 11, wherein the fermentation is conducted anaerobically.
17. Process according to any of claims 10-12, wherein the fermentation product is selected from the group consisting of ethanol, n-butanol, isobutanol, lactic acid, 3-hydroxy-

propionic acid, acrylic acid, acetic acid, succinic acid, fumaric acid, malic acid, itaconic acid, maleic acid, citric acid, adipic acid, an amino acid, such as lysine, methionine, tryptophan, threonine, and aspartic acid, 1,3-propane-diol, ethylene, glycerol, a β -lactam antibiotic and a cephalosporin, vitamins, pharmaceuticals, animal feed supplements, specialty chemicals, chemical feedstocks, plastics, solvents, fuels, including biofuels and biogas or organic polymers, and an industrial enzyme, such as a protease, a cellulase, an amylase, a glucanase, a lactase, a lipase, a lyase, an oxidoreductases, a transferase or a xylanase.

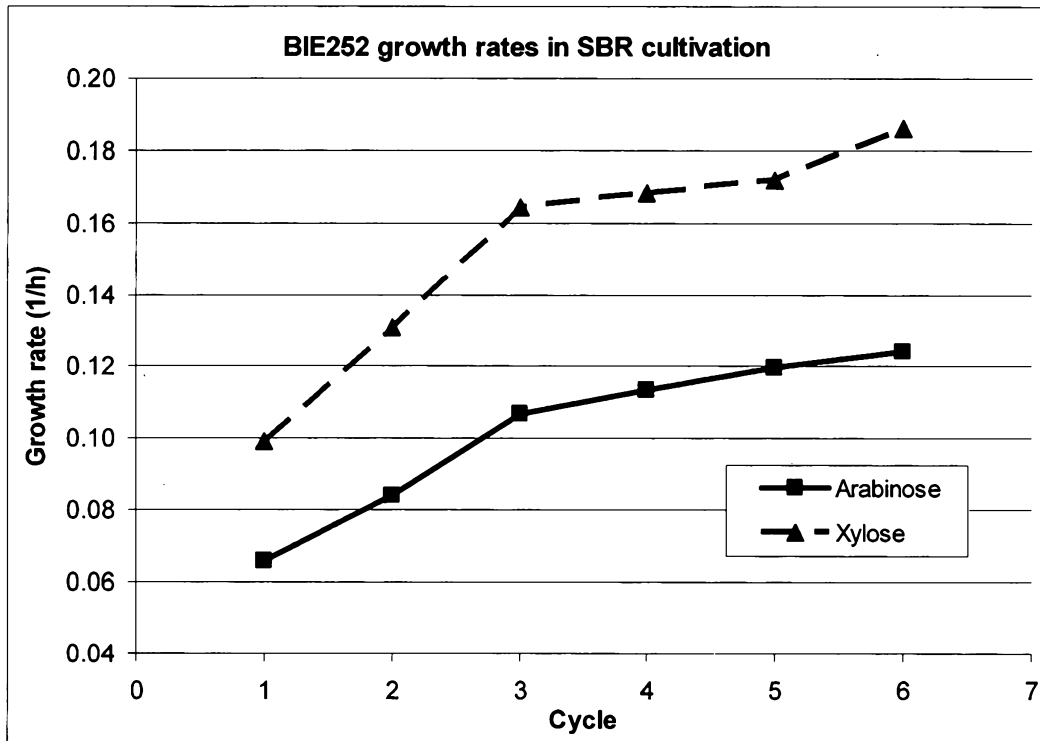


Fig. 1

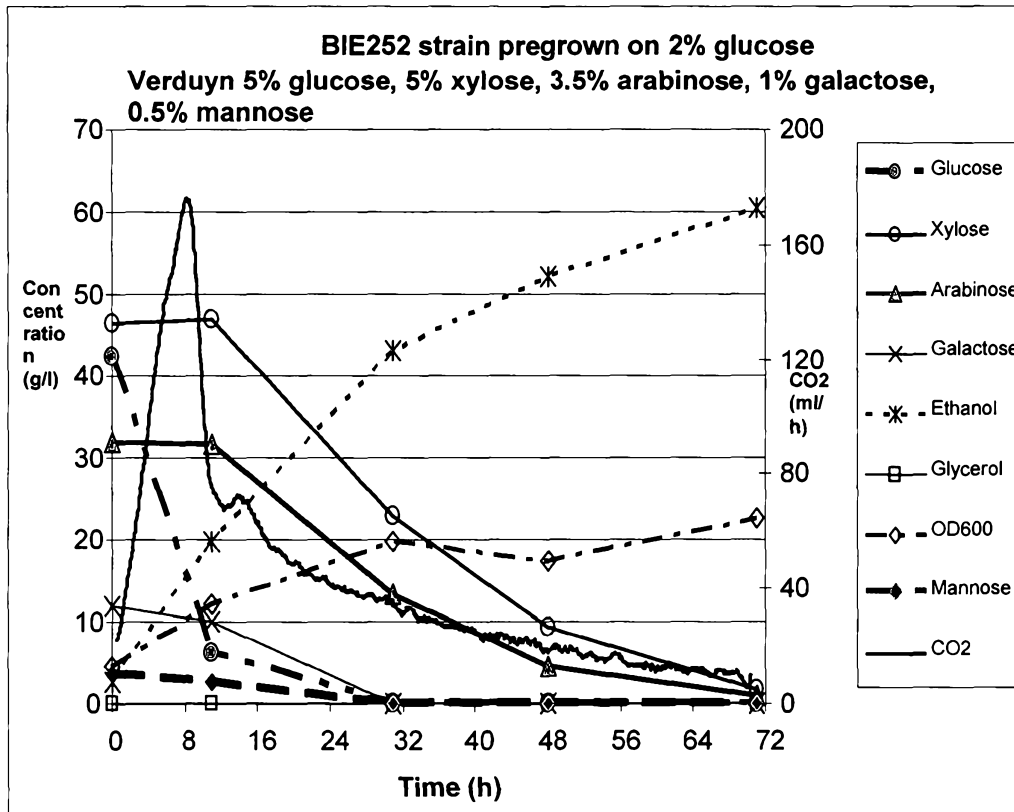


Fig. 2

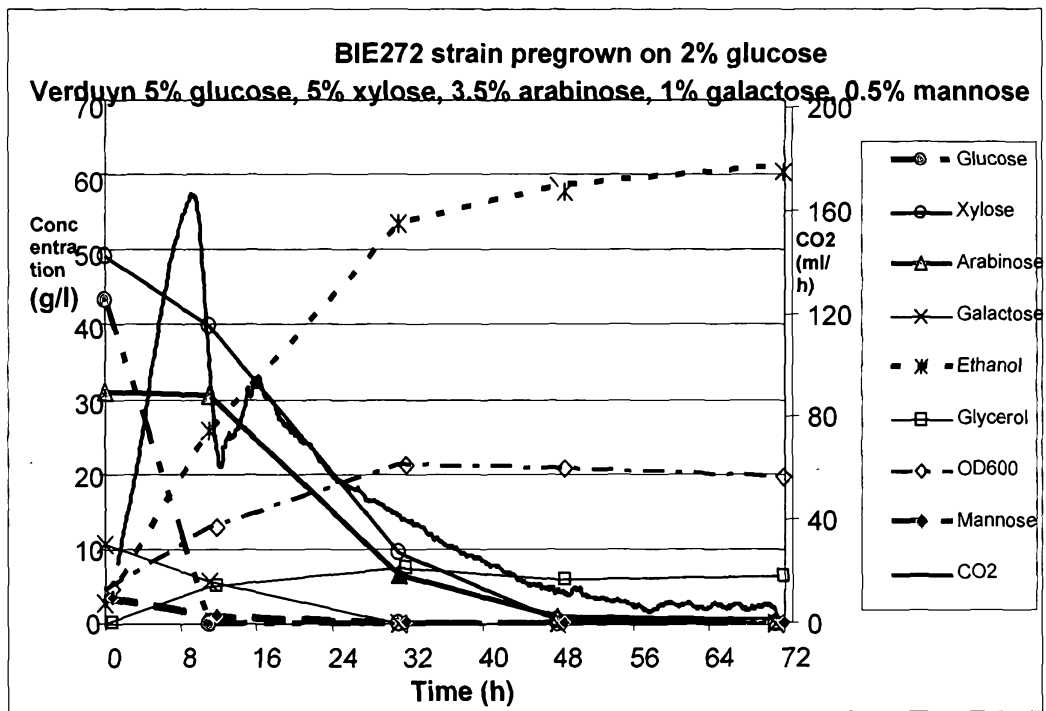


Fig. 3

4/22

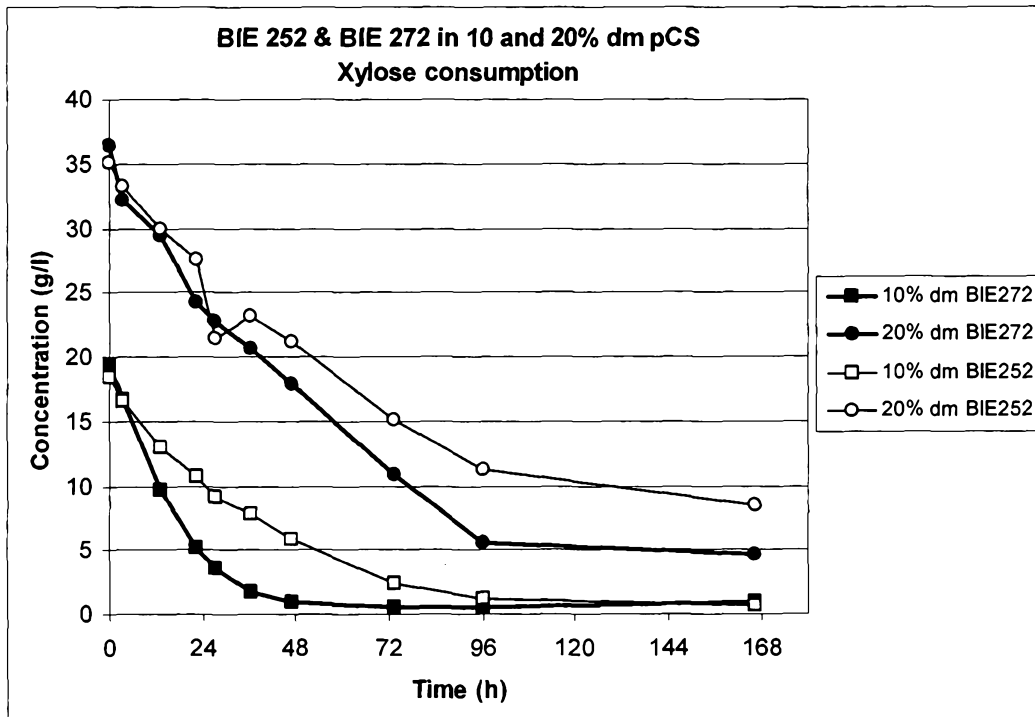


Fig. 4

5/22

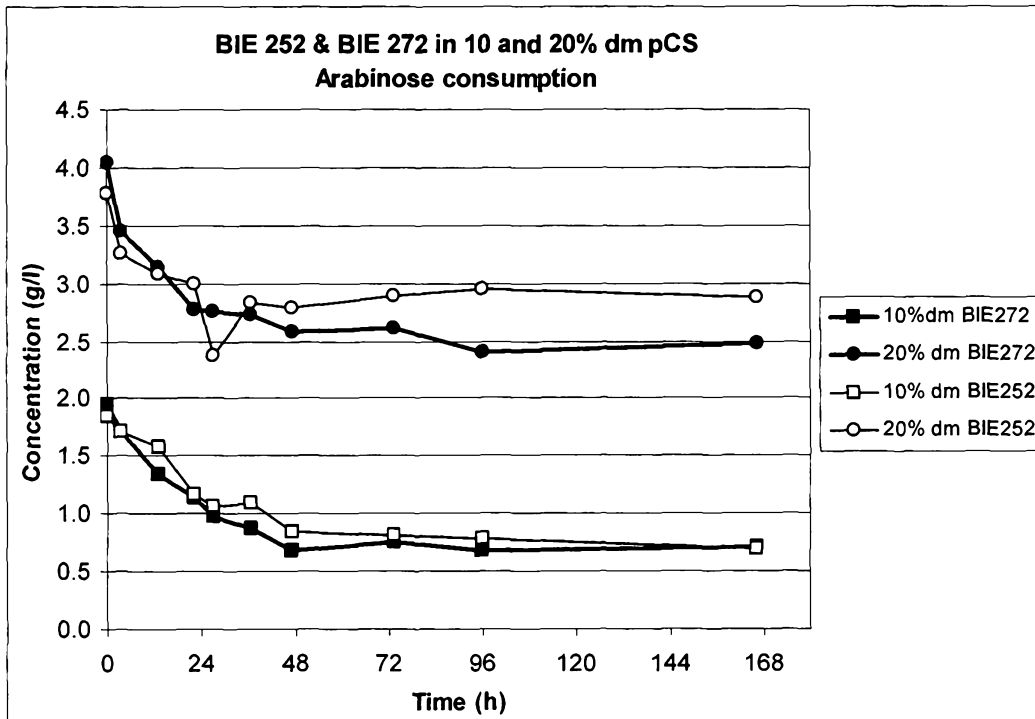


Fig. 5

6/22

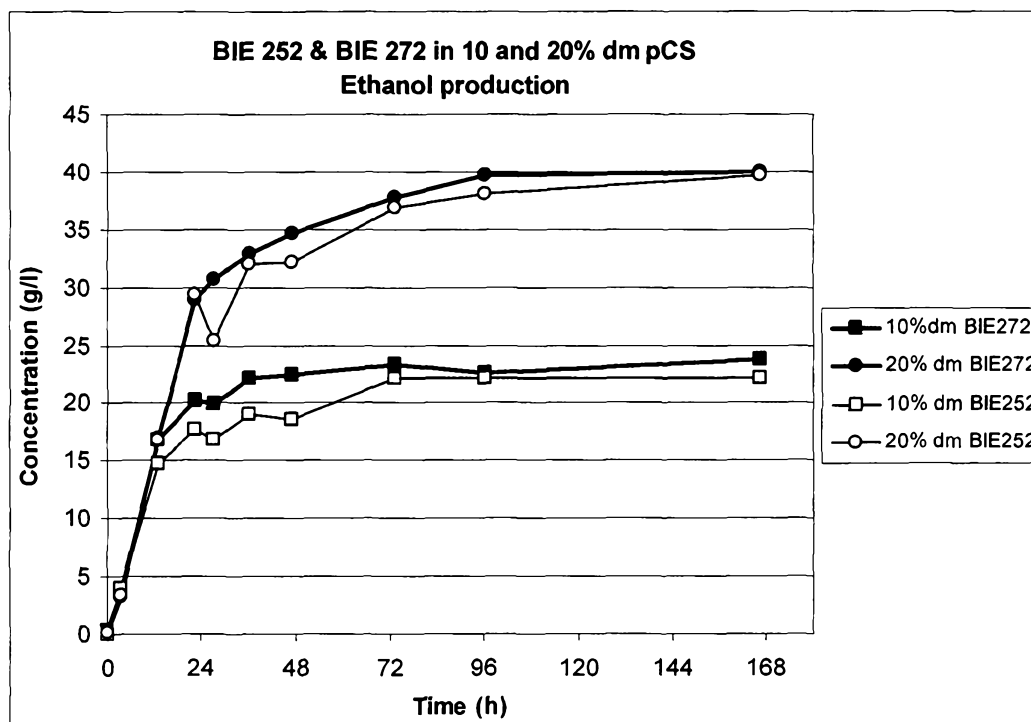


Fig. 6

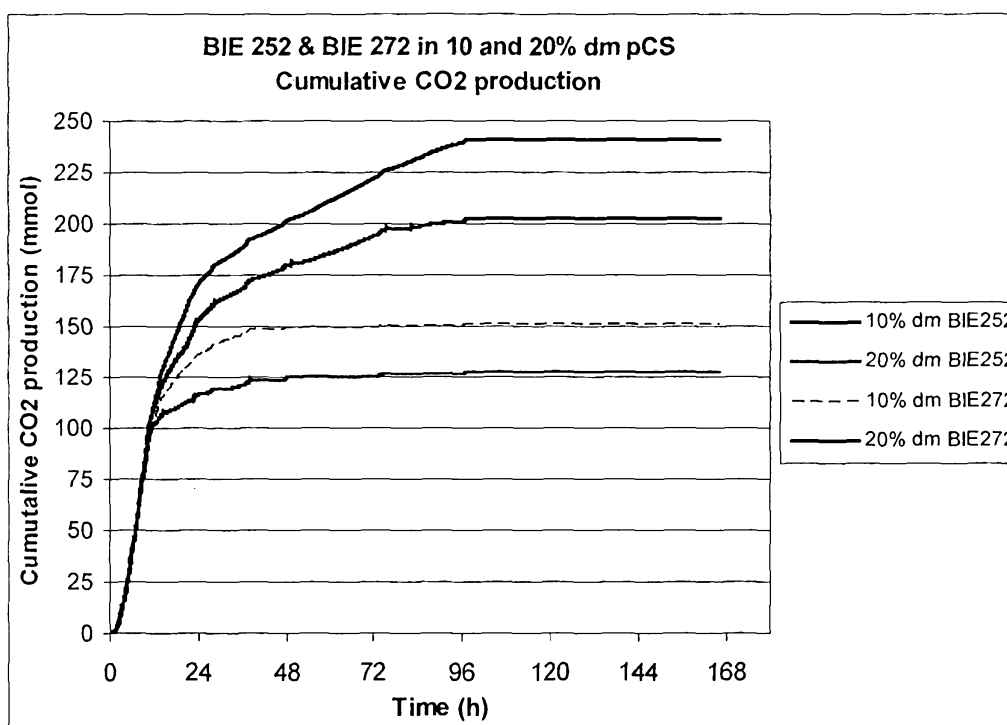


Fig. 7

8/22

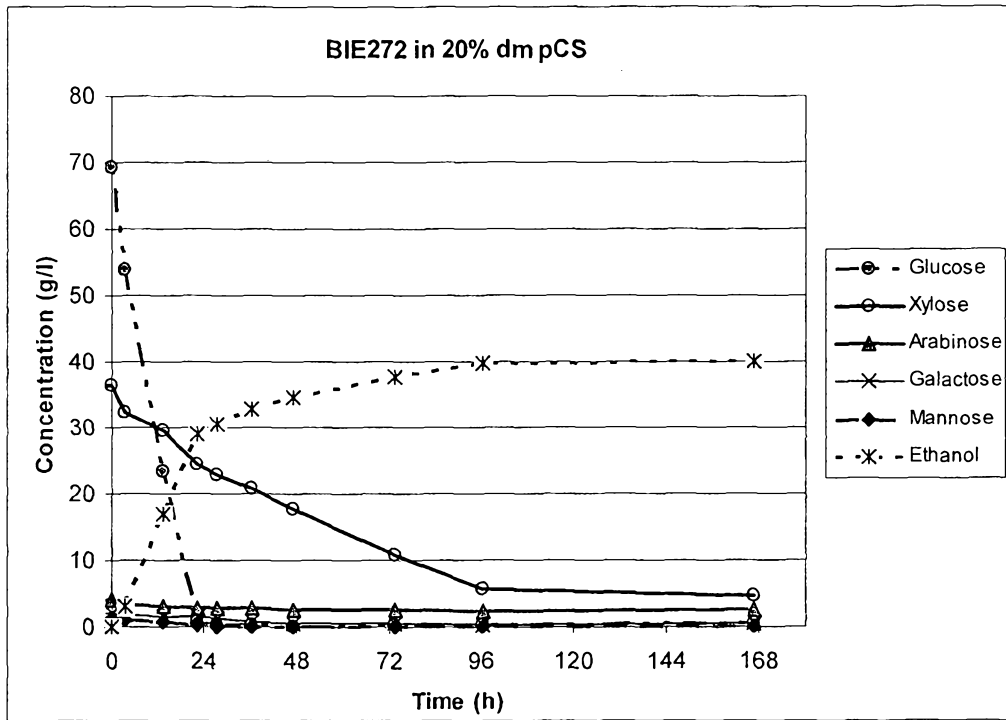


Fig. 8

9/22

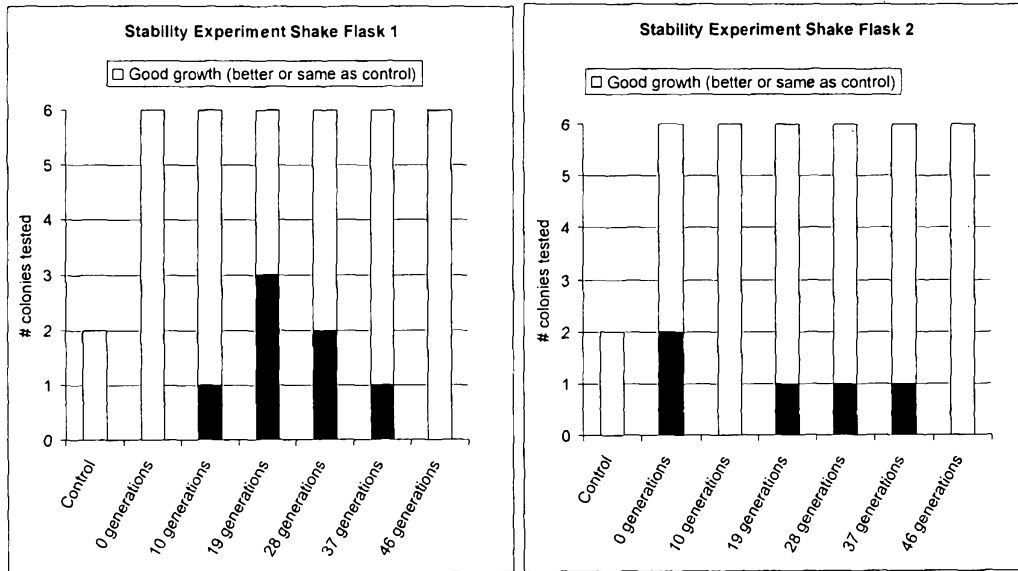


Fig. 9

10/22

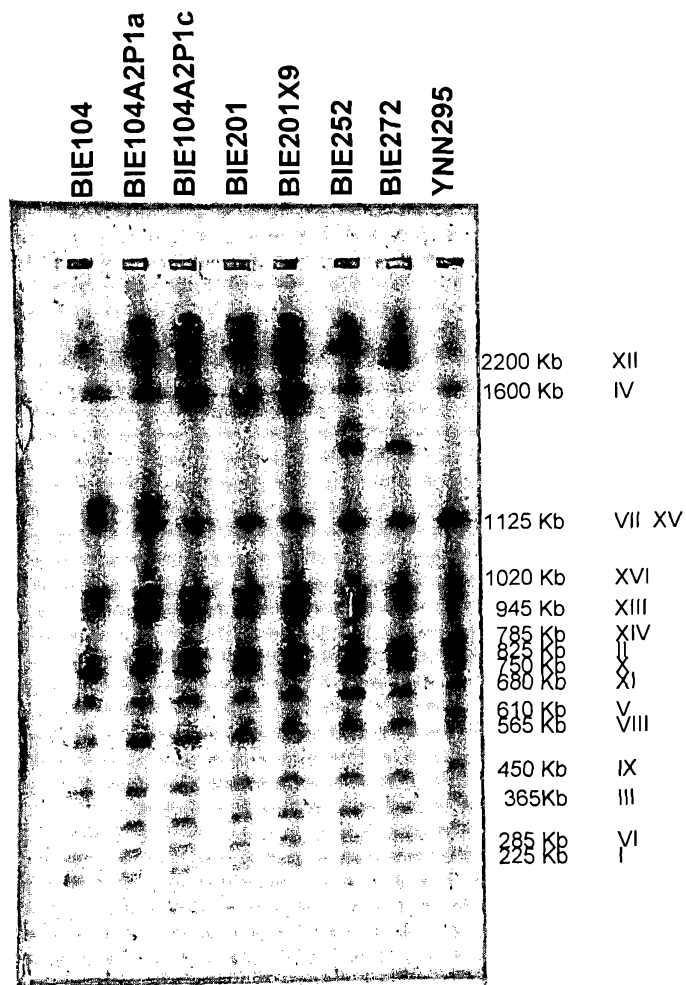


Fig. 10

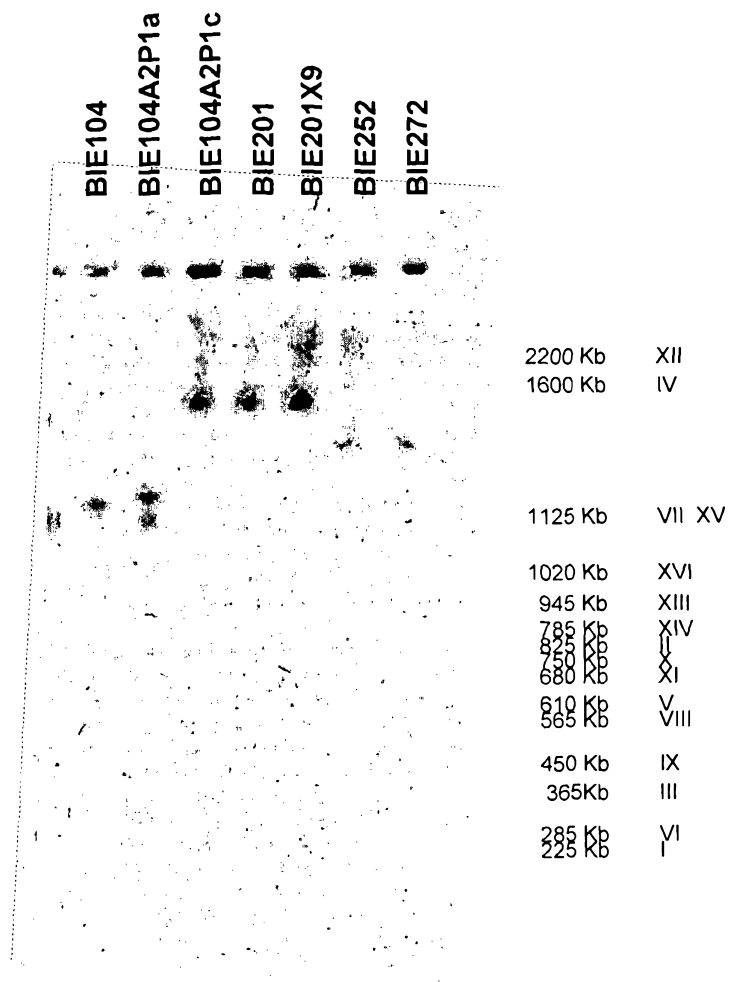


Fig. 11

12/22

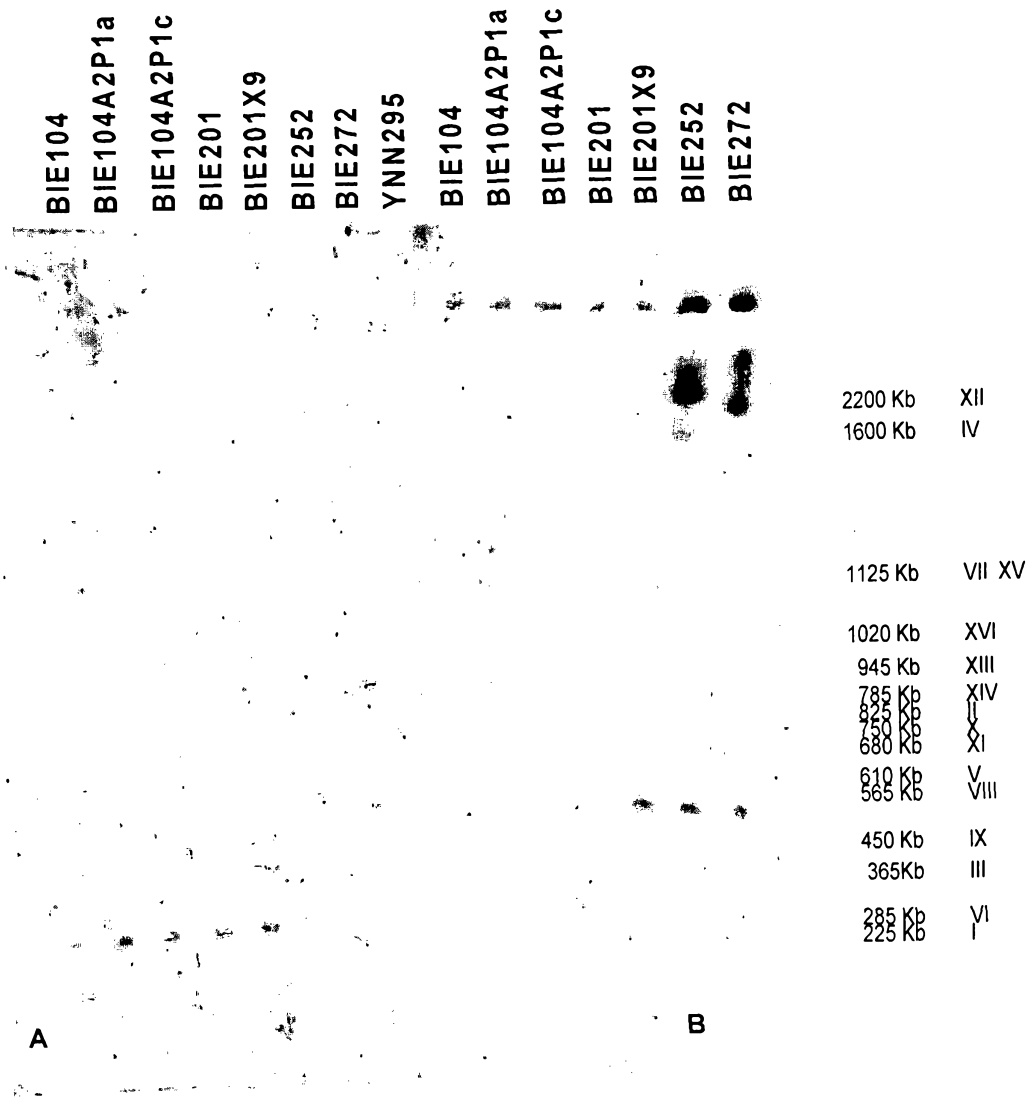


Fig. 12

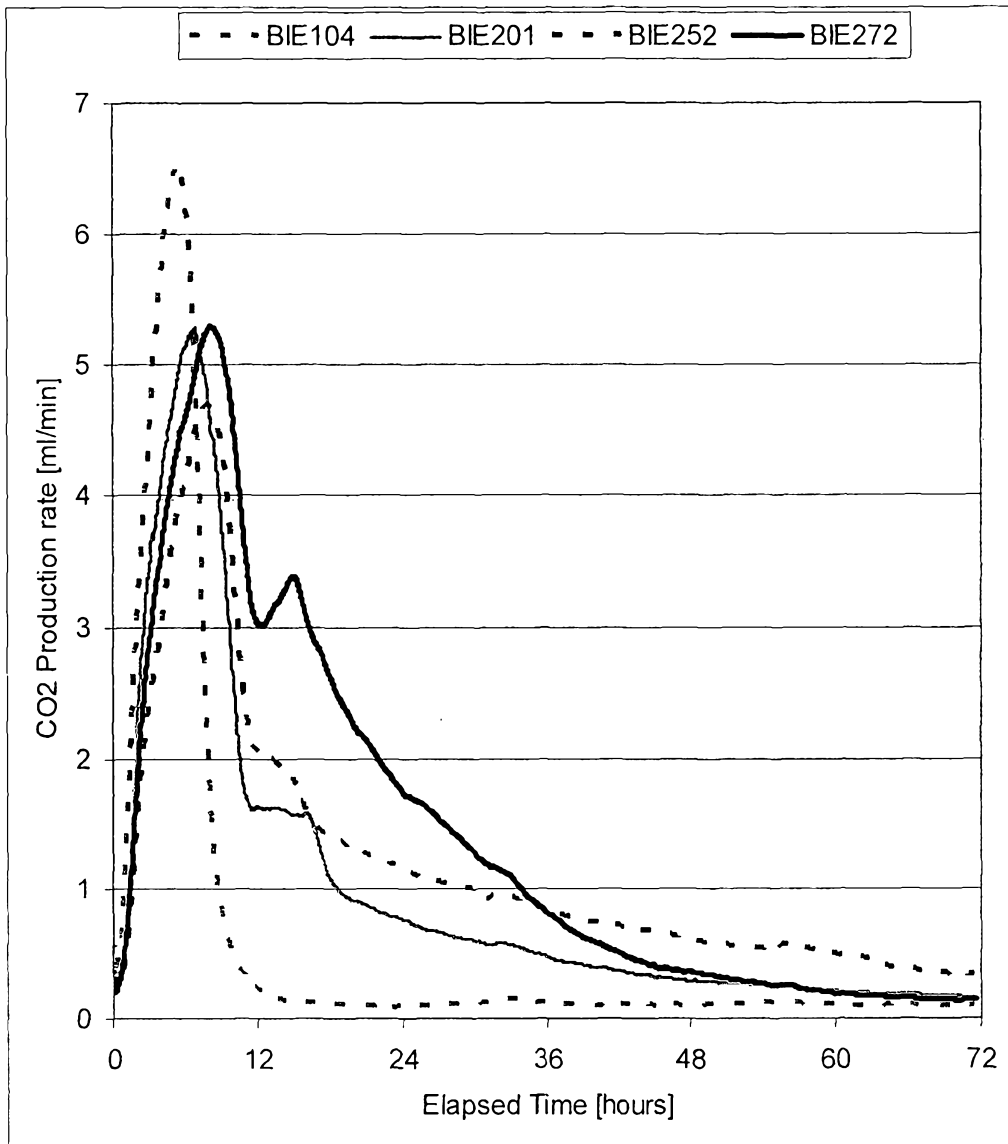


Fig. 13

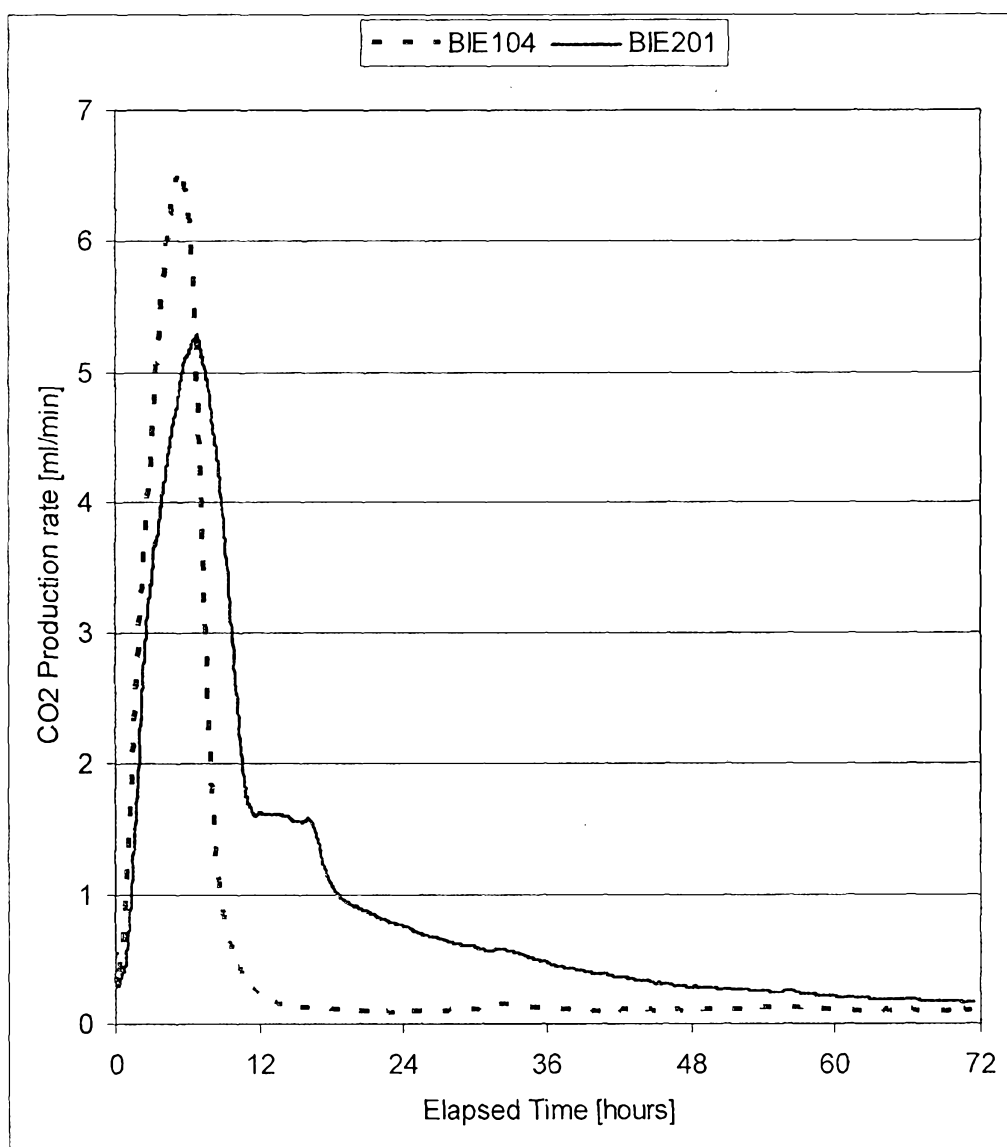


Fig. 14

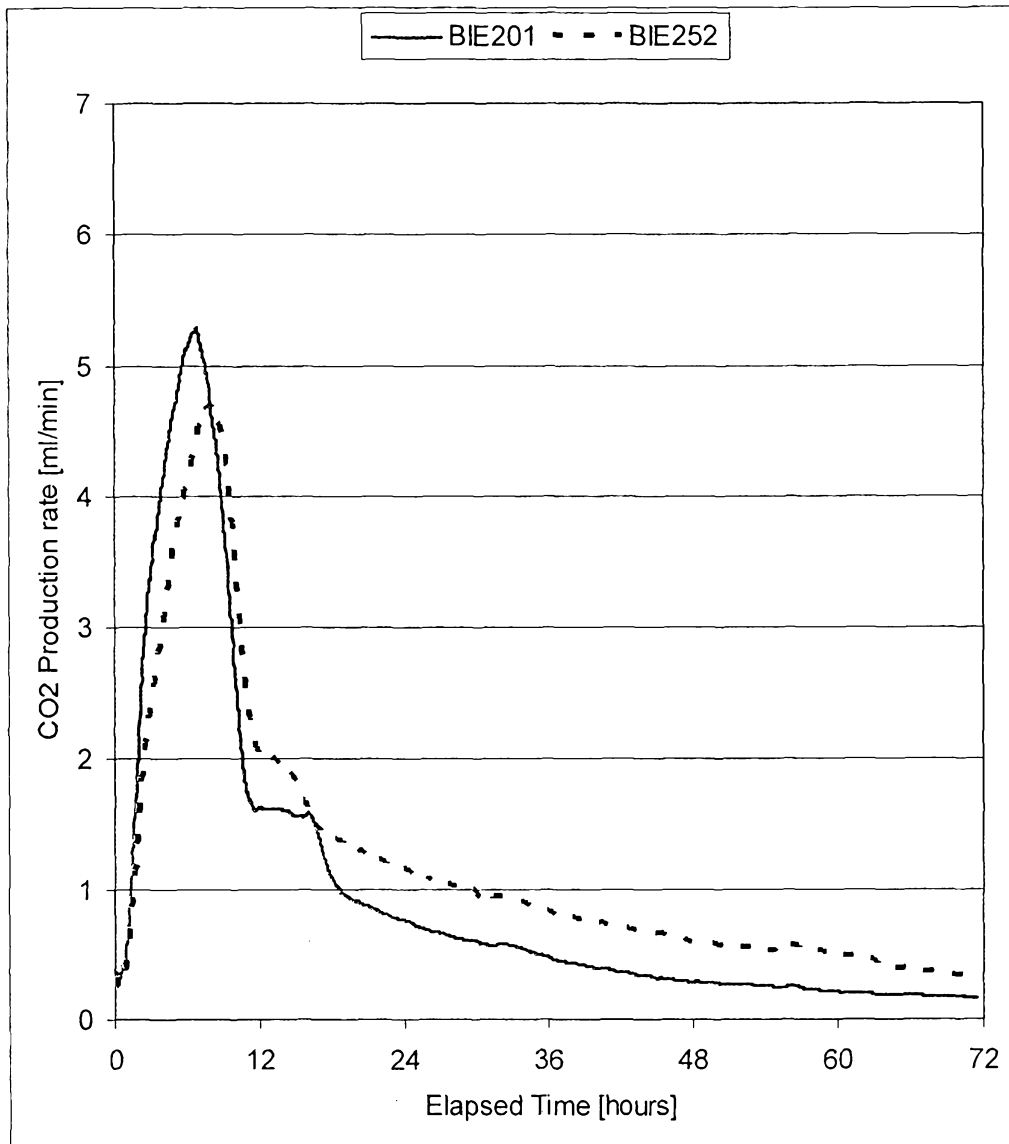


Fig. 15

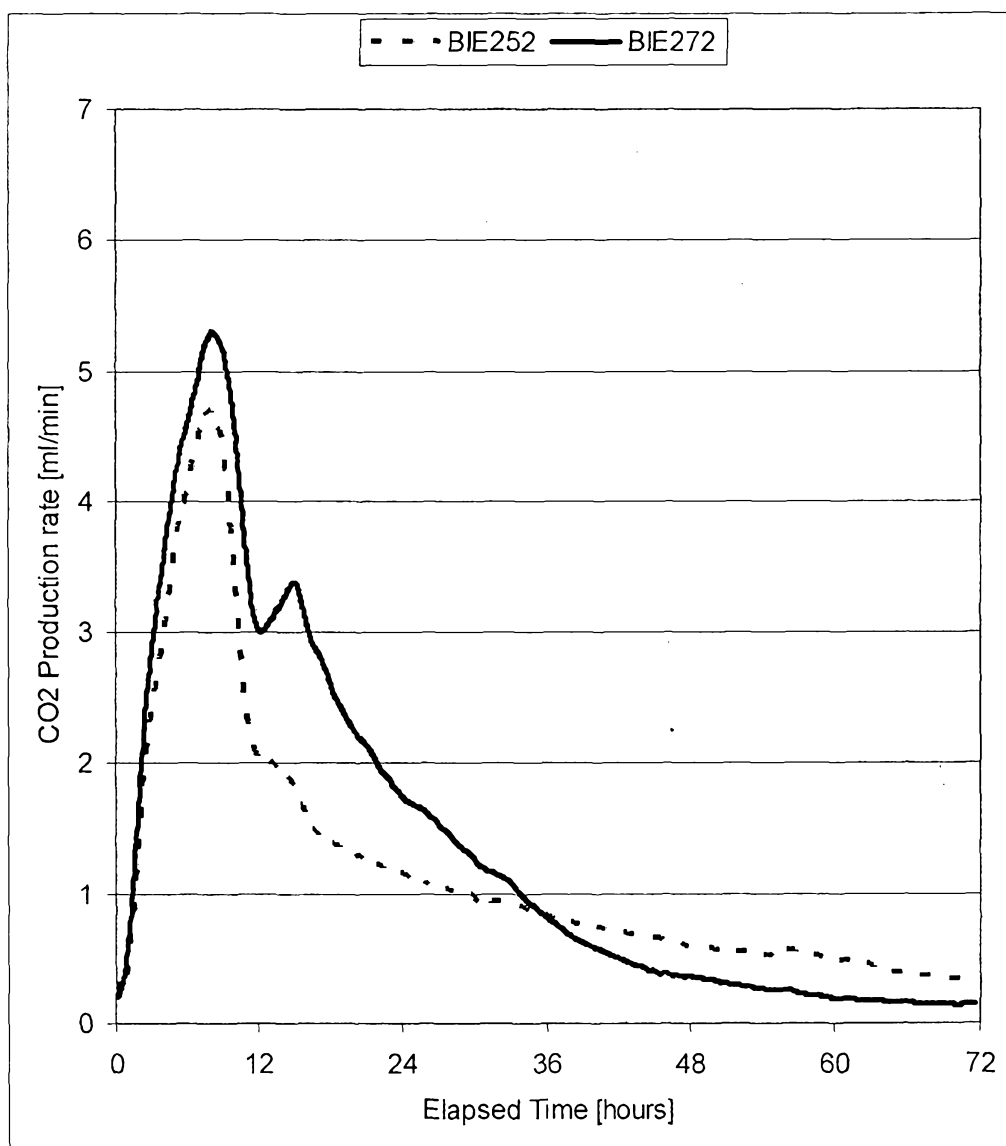


Fig. 16

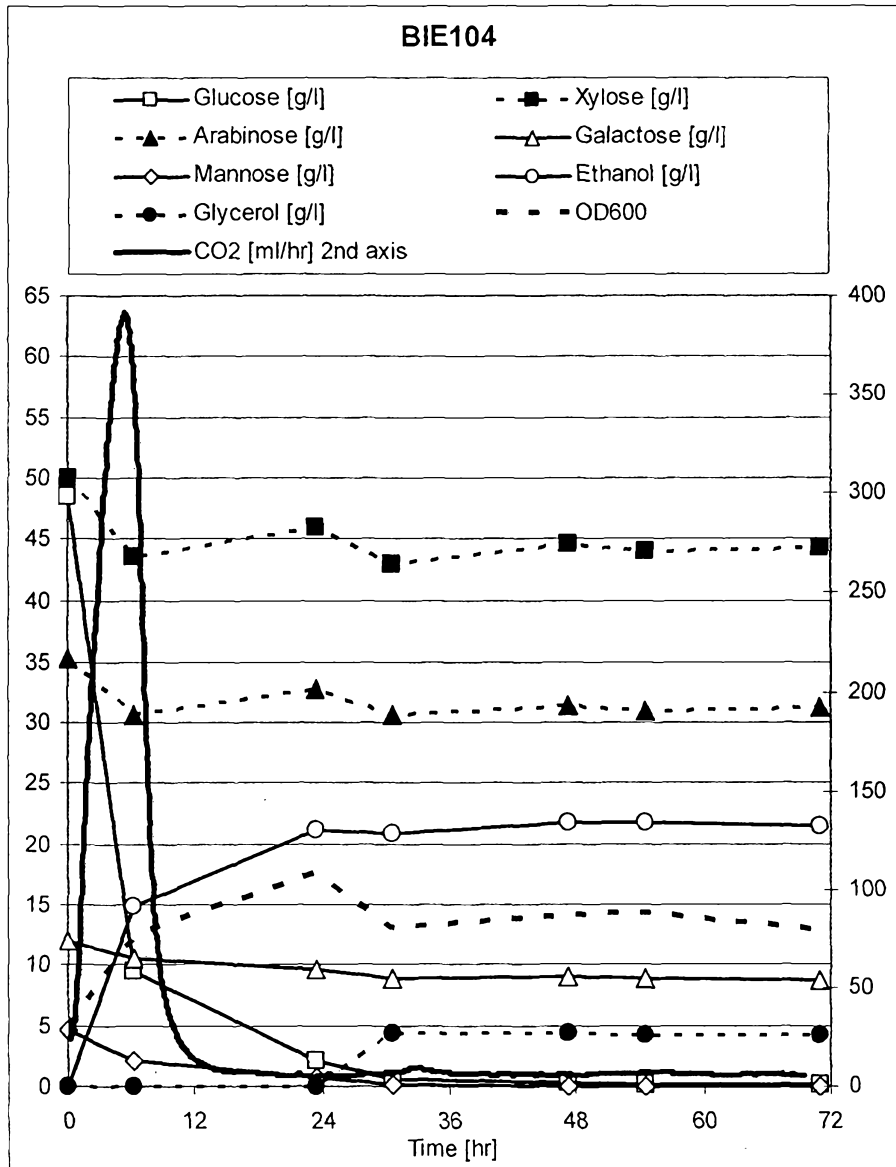


Fig. 17

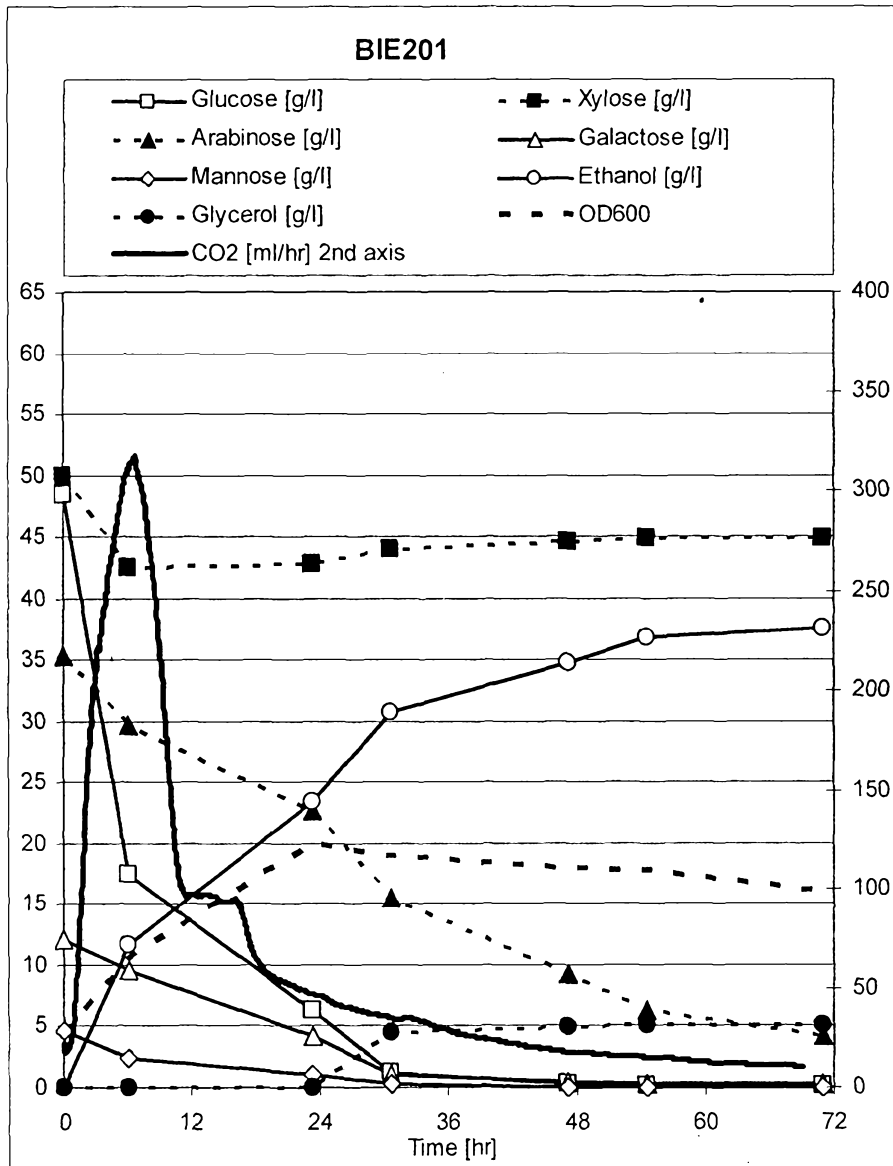


Fig. 18

19/22

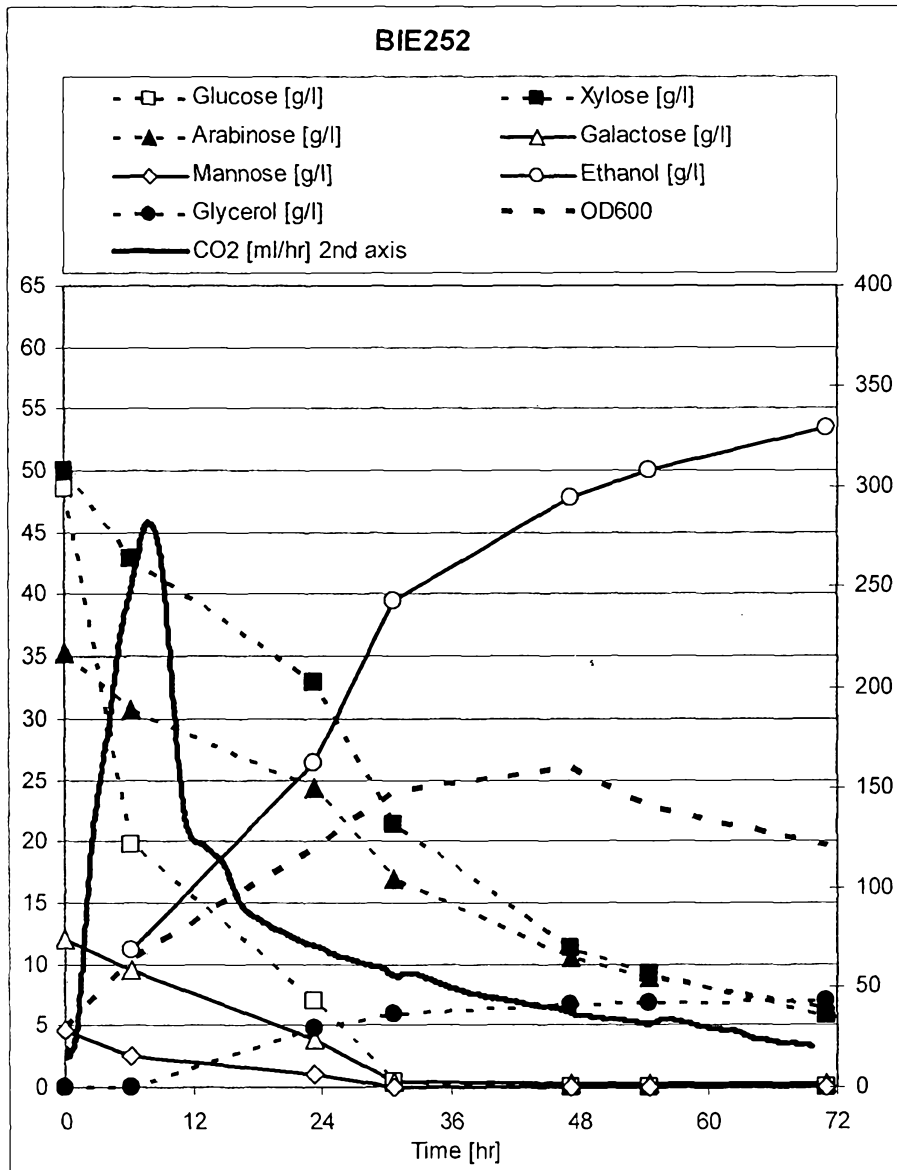


Fig. 19

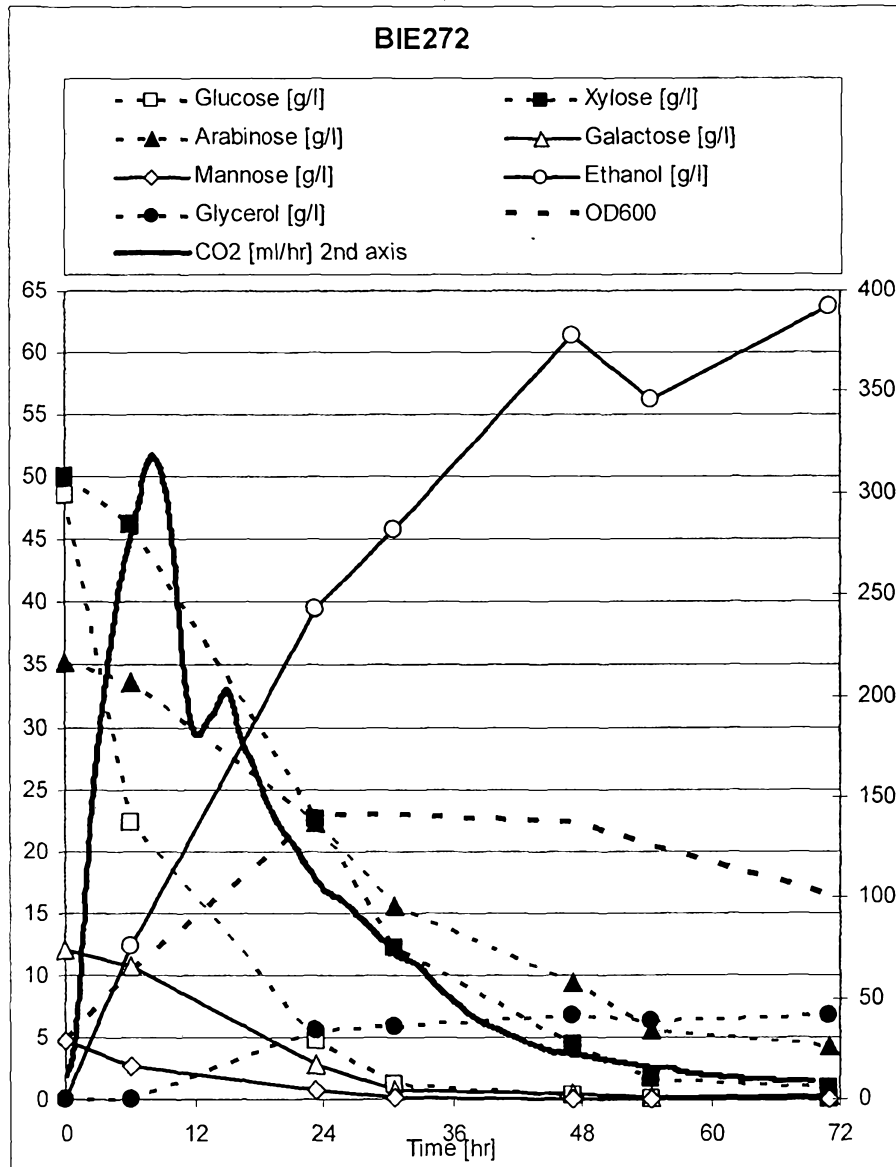


Fig. 20

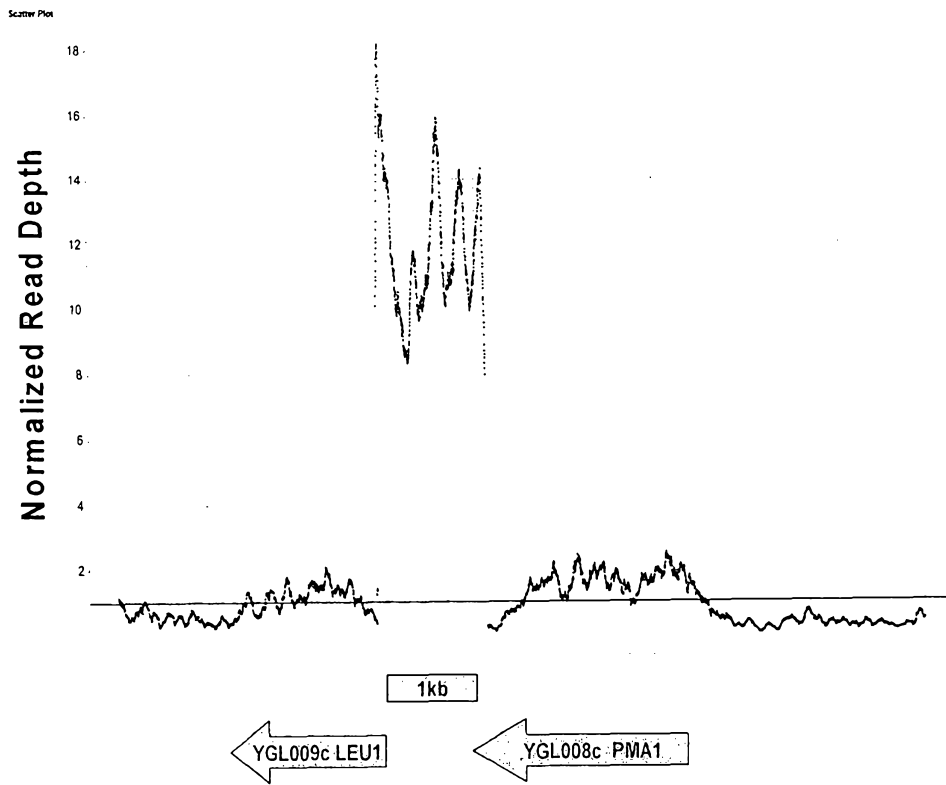


Fig.21

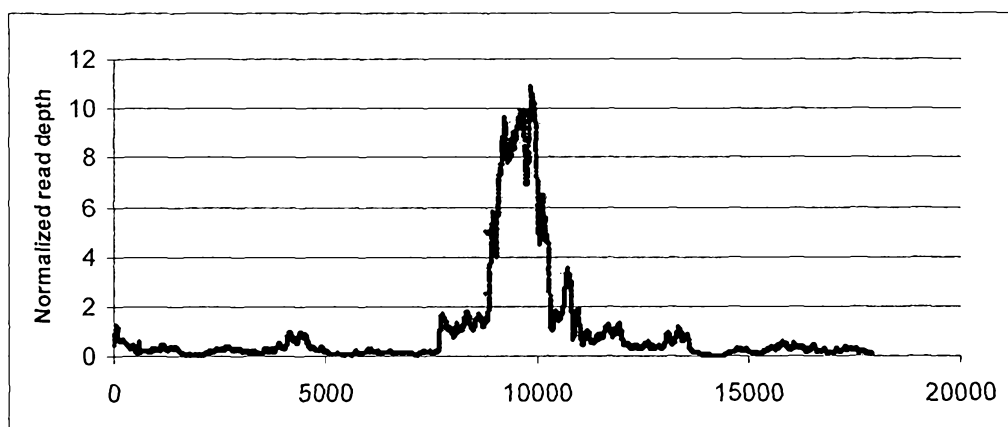


Fig. 22

eolf-seq1.txt
SEQUENCE LISTING

<110> DSM IP Assets B.V.
<120> YEAST CELL CAPABLE OF CONVERTING SUGARS INCLUDING ARABINOSE AND
XYLOSE
<130> 28114-WO-PCT
<160> 8
<170> PatentIn version 3.5
<210> 1
<211> 22
<212> DNA
<213> Artificial sequence
<220>
<223> synthetic DNA, forward primer xylA
<400> 1
caccgtagc cttggcgtaa gc 22

<210> 2
<211> 21
<212> DNA
<213> Artificial sequence
<220>
<223> synthetic DNA, reverse primer xylA
<400> 2
cactttcgaa cacgaattgg c 21

<210> 3
<211> 21
<212> DNA
<213> Artificial sequence
<220>
<223> synthetic DNA, forward primer ACT1
<400> 3
gttacgtcgc cttggacttc g 21

<210> 4
<211> 21
<212> DNA
<213> Artificial sequence
<220>
<223> synthetic DNA, reverse primer ACT1
<400> 4
cggcaatacc tgggaacatg g 21

<210> 5
<211> 22
<212> DNA
<213> Artificial sequence
<220>
<223> synthetic DNA, forward primer PNC1

eolf-seq1.txt

<400> 5
gatagagact ggcacaggat tg 22

<210> 6
<211> 21
<212> DNA
<213> Artificial sequence

<220>
<223> synthetic DNA, reverse primer PNC1

<400> 6
acaatactcc aaagctacac c 21

<210> 7
<211> 950
<212> PRT
<213> Saccharomyces cerevisiae

<400> 7

Met Ser Asp Asn Pro Phe Asn Ala Ser Leu Leu Asp Glu Asp Ser Asn
1 5 10 15

Arg Glu Arg Glu Ile Leu Asp Ala Thr Ala Glu Ala Leu Ser Lys Pro
20 25 30

Ser Pro Ser Leu Glu Tyr Cys Thr Leu Ser Val Asp Glu Ala Leu Glu
35 40 45

Lys Leu Asp Thr Asp Lys Asn Gly Gly Leu Arg Ser Ser Asn Glu Ala
50 55 60

Asn Asn Arg Arg Ser Leu Tyr Gly Pro Asn Glu Ile Thr Val Glu Asp
65 70 75 80

Asp Glu Ser Leu Phe Lys Lys Phe Leu Ser Asn Phe Ile Glu Asp Arg
85 90 95

Met Ile Leu Leu Leu Ile Gly Ser Ala Val Val Ser Leu Phe Met Gly
100 105 110

Asn Ile Asp Asp Ala Val Ser Ile Thr Leu Ala Ile Phe Ile Val Val
115 120 125

Thr Val Gly Phe Val Gln Glu Tyr Arg Ser Glu Lys Ser Leu Glu Ala
130 135 140

Leu Asn Lys Leu Val Pro Ala Glu Cys His Leu Met Arg Cys Gly Gln
145 150 155 160

Glu Ser His Val Leu Ala Ser Thr Leu Val Pro Gly Asp Leu Val His
165 170 175

eolf-seq1.txt

Phe Arg Ile Gly Asp Arg Ile Pro Ala Asp Ile Arg Ile Ile Glu Ala
 180 185 190

Ile Asp Leu Ser Ile Asp Glu Ser Asn Leu Thr Gly Glu Asn Glu Pro
 195 200 205

Val His Lys Thr Ser Gln Thr Ile Glu Lys Ser Ser Phe Asn Asp Gln
 210 215 220

Pro Asn Ser Ile Val Pro Ile Ser Glu Arg Ser Cys Ile Ala Tyr Met
 225 230 235 240

Gly Thr Leu Val Lys Glu Gly His Gly Lys Gly Ile Val Val Gly Thr
 245 250 255

Gly Thr Asn Thr Ser Phe Gly Ala Val Phe Glu Met Met Asn Asn Ile
 260 265 270

Glu Lys Pro Lys Thr Pro Leu Gln Leu Thr Met Asp Lys Leu Gly Lys
 275 280 285

Asp Leu Ser Leu Val Ser Phe Ile Val Ile Gly Met Ile Cys Leu Val
 290 295 300

Gly Ile Ile Gln Gly Arg Ser Trp Leu Glu Met Phe Gln Ile Ser Val
 305 310 315 320

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

Val Thr Leu Ala Leu Gly Val Leu Arg Met Ala Lys Arg Lys Ala Ile
 340 345 350

Val Arg Arg Leu Pro Ser Val Glu Thr Leu Gly Ser Val Asn Val Ile
 355 360 365

Cys Ser Asp Lys Thr Gly Thr Leu Thr Ser Asn His Met Thr Val Ser
 370 375 380

Lys Leu Trp Cys Leu Asp Ser Met Ser Asn Lys Leu Asn Val Leu Ser
 385 390 395 400

Leu Asp Lys Asn Lys Lys Thr Lys Asn Ser Asn Gly Asn Leu Lys Asn
 405 410 415

Tyr Leu Thr Glu Asp Val Arg Glu Thr Leu Thr Ile Gly Asn Leu Cys
 420 425 430

Asn Asn Ala Ser Phe Ser Gln Glu His Ala Ile Phe Leu Gly Asn Pro
 435 440 445

eolf-seq1.txt

Thr Asp Val Ala Leu Leu Glu Gln Leu Ala Asn Phe Glu Met Pro Asp
 450 455 460

Ile Arg Asn Thr Val Gln Lys Val Gln Glu Leu Pro Phe Asn Ser Lys
 465 470 475 480

Arg Lys Leu Met Ala Thr Lys Ile Leu Asn Pro Val Asp Asn Lys Cys
 485 490 495

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

Ser Tyr Leu Lys Ser Lys Gly Lys Lys Thr Glu Lys Leu Thr Glu Ala
 515 520 525

Gln Lys Ala Thr Ile Asn Glu Cys Ala Asn Ser Met Ala Ser Glu Gly
 530 535 540

Leu Arg Val Phe Gly Phe Ala Lys Leu Thr Leu Ser Asp Ser Ser Thr
 545 550 555 560

Pro Leu Thr Glu Asp Leu Ile Lys Asp Leu Thr Phe Thr Gly Leu Ile
 565 570 575

Gly Met Asn Asp Pro Pro Arg Pro Asn Val Lys Phe Ala Ile Glu Gln
 580 585 590

Leu Leu Gln Gly Gly Val His Ile Ile Met Ile Thr Gly Asp Ser Glu
 595 600 605

Asn Thr Ala Val Asn Ile Ala Lys Gln Ile Gly Ile Pro Val Ile Asp
 610 615 620

Pro Lys Leu Ser Val Leu Ser Gly Asp Lys Leu Asp Glu Met Ser Asp
 625 630 635 640

Asp Gln Leu Ala Asn Val Ile Asp His Val Asn Ile Phe Ala Arg Ala
 645 650 655

Thr Pro Glu His Lys Leu Asn Ile Val Arg Ala Leu Arg Lys Arg Gly
 660 665 670

Asp Val Val Ala Met Thr Gly Asp Gly Val Asn Asp Ala Pro Ala Leu
 675 680 685

Lys Leu Ser Asp Ile Gly Val Ser Met Gly Arg Ile Gly Thr Asp Val
 690 695 700

Ala Lys Glu Ala Ser Asp Met Val Leu Thr Asp Asp Asp Phe Ser Thr
 705 710 715 720

eolf-seq1.txt

Ile Leu Thr Ala Ile Glu Glu Gly Lys Gly Ile Phe Asn Asn Ile Gln
 725 730 735

Asn Phe Leu Thr Phe Gln Leu Ser Thr Ser Val Ala Ala Leu Ser Leu
 740 745 750

Val Ala Leu Ser Thr Ala Phe Lys Leu Pro Asn Pro Leu Asn Ala Met
 755 760 765

Gln Ile Leu Trp Ile Asn Ile Leu Met Asp Gly Pro Pro Ala Gln Ser
 770 775 780

Leu Gly Val Glu Pro Val Asp His Glu Val Met Lys Lys Pro Pro Arg
 785 790 795 800

Lys Arg Thr Asp Lys Ile Leu Thr His Asp Val Met Lys Arg Leu Leu
 805 810 815

Thr Thr Ala Ala Cys Ile Ile Val Gly Thr Val Tyr Ile Phe Val Lys
 820 825 830

Glu Met Ala Glu Asp Gly Lys Val Thr Ala Arg Asp Thr Thr Met Thr
 835 840 845

Phe Thr Cys Phe Val Phe Phe Asp Met Phe Asn Ala Leu Ala Cys Arg
 850 855 860

His Asn Thr Lys Ser Ile Phe Glu Ile Gly Phe Phe Thr Asn Lys Met
 865 870 875 880

Phe Asn Tyr Ala Val Gly Leu Ser Leu Leu Gly Gln Met Cys Ala Ile
 885 890 895

Tyr Ile Pro Phe Phe Gln Ser Ile Phe Lys Thr Glu Lys Leu Gly Ile
 900 905 910

Ser Asp Ile Leu Leu Leu Leu Ile Ser Ser Ser Val Phe Ile Val
 915 920 925

Asp Glu Leu Arg Lys Leu Trp Thr Arg Lys Lys Asn Glu Glu Asp Ser
 930 935 940

Thr Tyr Phe Ser Asn Val
 945 950

<210> 8
 <211> 950
 <212> PRT
 <213> Saccharomyces cerevisiae

<400> 8

Met Ser Asp Asn Pro Phe Asn Ala Ser Leu Leu Asp Glu Asp Ser Asn
 Page 5

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

275

280

285

Asp Leu Ser Leu Val Ser Phe Ile Val Ile Gly Met Ile Cys Leu Val
 290 295 300

Gly Ile Ile Gln Gly Arg Ser Trp Leu Glu Met Phe Gln Ile Ser Val
 305 310 315 320

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

Val Thr Leu Ala Leu Gly Val Leu Arg Met Ala Lys Arg Lys Ala Ile
 340 345 350

Val Arg Arg Leu Pro Ser Val Glu Thr Leu Gly Ser Val Asn Val Ile
 355 360 365

Cys Ser Asp Lys Thr Gly Thr Leu Thr Ser Asn His Met Thr Val Ser
 370 375 380

Lys Leu Trp Cys Leu Asp Ser Met Ser Asn Lys Leu Asn Val Leu Ser
 385 390 395 400

Leu Asp Lys Asn Lys Lys Thr Lys Asn Ser Asn Gly Asn Leu Lys Asn
 405 410 415

Tyr Leu Thr Glu Asp Val Arg Glu Thr Leu Thr Ile Gly Asn Leu Cys
 420 425 430

Asn Asn Ala Ser Phe Ser Gln Glu His Ala Ile Phe Leu Gly Asn Pro
 435 440 445

Thr Asp Val Ala Leu Leu Glu Gln Leu Ala Asn Phe Glu Met Pro Asp
 450 455 460

Ile Arg Asn Thr Val Gln Lys Val Gln Glu Leu Pro Phe Asn Ser Lys
 465 470 475 480

Arg Lys Leu Met Ala Thr Lys Ile Leu Asn Pro Val Asp Asn Lys Cys
 485 490 495

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

Ser Tyr Leu Lys Ser Lys Gly Lys Lys Thr Glu Lys Leu Thr Glu Ala
 515 520 525

Gln Lys Ala Thr Ile Asn Glu Cys Ala Asn Ser Met Ala Ser Glu Gly
 530 535 540

Leu Arg Val Phe Gly Phe Ala Lys Leu Thr Leu Ser Asp Ser Ser Thr

Glu Met Ala Glu Asp Gly Lys Val Thr Ala Arg Asp Thr Thr Met Thr
835 840 845

Phe Thr Cys Phe Val Phe Phe Asp Met Phe Asn Ala Leu Ala Cys Arg
850 855 860

His Asn Thr Lys Ser Ile Phe Glu Ile Gly Phe Phe Thr Asn Lys Met
865 870 875 880

Phe Asn Tyr Ala Val Gly Leu Ser Leu Leu Gly Gln Met Cys Ala Ile
885 890 895

Tyr Ile Pro Phe Phe Gln Ser Ile Phe Lys Thr Glu Lys Leu Gly Ile
900 905 910

Ser Asp Ile Leu Leu Leu Leu Leu Ile Ser Ser Ser Val Phe Ile Val
915 920 925

Asp Glu Leu Arg Lys Leu Trp Thr Arg Lys Lys Asn Glu Glu Asp Ser
930 935 940

Thr Tyr Phe Ser Asn Val
945 950