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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
11 September 2009 (11.09.2009)

PCT

(10) International Publication Number
WO 2009/109633 A1

(51) **International Patent Classification**:
C12N 9/92 (2006.01) *C12N 15/52* (2006.01)
C12P 7/10 (2006.01)

(21) **International Application Number**:
PCT/EP2009/052623

(22) **International Filing Date**:
5 March 2009 (05.03.2009)

(25) **Filing Language**: English

(26) **Publication Language**: English

(30) **Priority Data**:
08102407.7 7 March 2008 (07.03.2008) EP

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

(72) **Inventors; and**
(75) **Inventors/Applicants** *(for US only)*: **KLAASSEN, Paul** [NL/NL]; Chico Mendesring 111, 3315 TN Dordrecht (NL). **LAAN, VAN DER,, Jan Metske** [NL/NL]; Leursebaan 364, 4839 AP Breda (NL). **GIELESEN, Bianca Elisabeth Maria** [NL/NL]; G.a. Brederolaan 69b, 3141 CD Maassluis (NL). **SUYLEKOM, VAN, Gijsberdina Pieternella** [NL/NL]; Bieselaar 42, 5109 BB 's Gravenmoer (NL).

(74) **Agents**: **KLEIBORN, Paul Erik** et al.; Dsm Intellectual Property, Delft Office (600-0240), P.o. Box 1, 2600 MA Delft (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, 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, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, 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, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— *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))*

WO 2009/109633 A1

(54) **Title**: A PENTOSE SUGAR FERMENTING CELL

(57) **Abstract**: The invention relates to a cell which comprises a nucleotide sequence encoding a xylose isomerase, wherein the amino acid sequence of the xylose isomerase has at least about 70% sequence identity to the amino acid sequence set out in SEQ ID NO: 3 and wherein the nucleotide sequence is heterologous to the host. A cell of the invention may be used in a process for producing a fermentation product, such as ethanol. Such a process may comprise fermenting a medium containing a source of xylose with a cell of the invention such that the cell ferments xylose to the fermentation product.

A PENTOSE SUGAR FERMENTING CELL

Field of the invention

The present invention relates to a cell which is capable of isomerising xylose to xylulose. The invention also relates to a process in which such cells are used for the production of a fermentation product, such as ethanol.

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 not 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 CO_2 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, the second-most-common sugar, after glucose, is 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 produce ethanol using xylose as a carbon source. Also, no naturally-occurring organisms are known which can ferment xylose to ethanol with both a high ethanol yield and a 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.

Summary of the invention

According to the invention, there is provided a cell that is capable of fermentation, such as alcoholic fermentation, and of using xylose as a carbon source. Such a cell comprises a nucleotide sequence encoding a xylose isomerase, wherein the amino acid sequence of the xylose isomerase has at least about 70% sequence identity to the amino acid sequence set out in SEQ ID NO: 3 and wherein the nucleotide sequence is heterologous to the host. Such a cell produces a higher amount of ethanol when using xylose as a carbon source as compared to the wild type filamentous fungus.

The invention also provides:

- a process for producing a fermentation product which process comprises fermenting a medium containing a source of xylose with a cell of the invention such that the cell ferments xylose to the fermentation product;
- a process for producing a fermentation product which process comprises fermenting a medium containing at least a source of xylose and a source of L-arabinose with a cell as defined of the invention which is also capable of utilizing L-arabinose such that the cell ferments xylose and L-arabinose to the fermentation product; and
- a process for producing a fermentation product which process comprises fermenting a medium containing at least a source of xylose and a source of L-arabinose with a cell of the invention and a cell able to use L-arabinose, whereby each cell ferments xylose and/or arabinose to the fermentation product.

The invention further provides the use of a cell of the invention in a process for the production of a fermentation product.

Brief description of the drawings

Figure 1 sets out the plasmid map of pYISIT4-XKS1-xylA (Baun CpO) encoding xylose isomerase from *Bacteroides uniformis* ATCC 8492 for expression in *Saccharomyces cerevisiae*. CpO denotes codon pair optimized.

Figure 2 sets out a physical map of plasmid pPWT080, the sequence of which is given in SEQ ID no. 4.

Figure 3 sets out a physical map of the wild-type GRE3-locus (panel a) and a one copy integration of PWT080 in the GRE3-locus (panel b, showing where the primers bind and panel c, showing where the *RKI1*-probe binds)

Figure 4 sets out an Autoradiogram showing the correct integration of one copy of the plasmid pPWT080 in CEN.PK113-7D;

Panel a: *Xcm*I-digestion of chromosomal DNA preparations, hybridized with the *RKI1*-probe. Lane 1: CEN.PK113-7D; lane 2: BIE104F1; lane 3: BIE104P1

Panel b: *Psi*I-digestion of chromosomal DNA preparations, hybridized with the *RKI1*-probe. Lane 1: CEN.PK113-7D; lane 2: BIE104F1; lane 3: BIE104P1

Δ*GRE3::*PPP stands for the replacement of the coding region of the *GRE3*-gene by the cassette containing the genes *TAL1, TKL1, RKI1* and *RPE1* under control of strong constitutive promoters, *ΔGRE3::*[*TPI1p-TAL1-ADH1p-TKL1-PGI1p-RPE1-ENO1p-RKI1*].

Figure 5 sets out a physical map of the *GRE3*-locus, where the coding region of the *GRE3*-gene was replaced by the integration of the PPP-genes *TAL1, TKL1, RKI1* and *RPE1.* Panel a shows the where the primers of SEQ ID 5 and 6 bind, panel b shows where the *RKI1*-probe binds.

Figure 6 sets out a physical map of plasmid pYI#SIT4

Figure 7 sets out a physical map of plasmid pPWT007

Figure 8 sets out a physical map of plasmid pPWT042

Figure 9 sets out a physical map of the wild-type *SIT4*-locus (panel a) and a one copy integration of PWT080 in the *SIT4*-locus (panel b, showing where the primers bind)

Figure 10 sets out a growth curve of BIE104P1Y9 on 2% xylose as sole carbon source, after several precultivations, and of the reference strain without one copy of pPWT042 integrated in the genome. Events indicated in the graph by numbers (1): transfer to YNB 1% glucose + 1% xylose; (2): transfer to YNB 0.1% glucose + 2% xylose; (3) transfer to YNB 2% xylose; (4) transfer to YNB 2% xylose (only BIE104P1Y9).

Figure 11 sets out a growth curve of the reference strain BIE104P1 and a xylose metabolizing strain, BIE104P1Y9.

Figure 12 sets out the xylose and glucose consumption and ethanol production in time of strains BIE104P1 precultured on glucose (panel a), BIE104P1Y9 precultured on glucose (panel b) and BIE104P1Y9 precultured on xylose (panel c).

Figure 13 sets out a physical map of plasmid pPWT018.

Figure 14 sets out a physical map of plasmid pPWT006.

Figure 15 sets out a Southern blot autoradiogram. Chromosomal DNA of wild-type strain CEN.PK113-7D (lane 1) and BIE104A2 (lane 2) was digested with Both EcoRI and HindIII. The blot was hybridized with a specific *SIT2*-probe.

Figure 16 sets out physical maps of the wild-type *SIT2*-locus (panel a) and after introduction of the ara-genes by integration of plasmid pPWT018, followed by intramolecular recombination leading to the loss of vector and selectable marker sequences (panel b). The hybridization of the probe is indicated.

Figure 17 sets out a graphic representation of growth curves of strain BIE104A2P1Y9 on different media. Panel a: strain BIE104A2P1Y9 grown on galactose, followed by events indicated in the graph by numbers (1) transfer to 1% arabinose + 1% xylose and (2) transfer to 2% xylose + 0.2% arabinose. Panel b: strain BIE104A2P1Y9 grown on glucose, followed by (1) transfer to 1% arabinose + 1% xylose and (2) transfer to 2% xylose + 0.2% arabinose.

Brief description of the sequence listing

SEQ ID NO: 1 sets out the wild-type xylose isomerase sequence from *Bacteroides uniformis* ATCC 8492. Genbank accession no. AAYH02000036.

SEQ ID NO: 2 sets out a codon optimized sequence derived from SEQ ID NO: 1.

SEQ ID NO: 3 sets out the amino acid sequence of xylose isomerase from *Bacteroides uniformis* ATCC 8492.

SEQ ID NO: 4 sets out the sequence of plasmid pPWT080.

SEQ ID NO: 5 sets out the sequence of forward primer.

SEQ ID NO: 6 sets out the sequence of reverse primer.

SEQ ID NO: 7 sets out the sequence of the forward multifunctional primer for diagnostic PCR.

SEQ ID NO: 8 sets out the sequence of reverse multifunctional primer for diagnostic PCR.

SEQ ID NO: 9 sets out the sequence of forward primer RKI1-probe.

SEQ ID NO: 10 sets out the sequence of reverse primer RKI1-probe.

SEQ ID NO: 11 sets out the sequence of forward primer kanMX-cassette.

SEQ ID NO: 12 sets out the sequence of reverse primer kanMX-cassette.

SEQ ID NO: 13 sets out the sequence of forward primer.

SEQ ID NO: 14 sets out the sequence of reverse primer.

SEQ ID NO: 15 sets out the sequence of forward multifunctional primer for diagnostic PCR.

SEQ ID NO: 16 sets out the sequence of reverse multifunctional primer for diagnostic PCR.

SEQ ID NO: 17 sets out the sequence of sequence of plasmid pPWT018

SEQ ID NO: 18 sets out the sequence of forward primer integration pPWT018.

SEQ ID NO: 19 sets out the sequence of reverse primer integration pPWT018.

SEQ ID NO: 20 sets out the sequence of forward primer SIT2-probe.

SEQ ID NO: 21 sets out the sequence of reverse primer SIT2-probe.

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 allows.

The invention relates to a cell which comprises a nucleotide sequence encoding a xylose isomerase, wherein the amino acid sequence of the xylose isomerase has at least about 70% identity to the amino acid sequence set out in SEQ ID NO: 3 and wherein the nucleotide sequence is heterologous to the host.

The presence of the nucleotide sequence encoding a xylose isomerase confers on the cell the ability to isomerise xylose to xylulose.

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 cell of the invention is defined with reference to a xylose isomerase having the amino acid sequence of SEQ ID NO: 3 or a sequence having at least about 70% sequence identity thereto. Likewise, a cell of the invention may be defined with reference to a xylose isomerase be a nucleotide sequence which encoding such an amino acid sequence.

SEQ ID NO: 3 sets out the amino acid sequence of xylose isomerase from *Bacteroides uniformis* ATCC 8492. A cell of the invention comprises a nucleotide sequence encoding a xylose isomerase having the amino acid of SEQ ID NO: 3 or one which has at least about 70% sequence identity thereto.

Preferably, a cell according to the present invention is a cell comprising a nucleotide sequence encoding a xylose isomerase having a sequence which has at least about 75%, preferably at least about 80%, at least about 85%, at least about 90%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98% or at least about 99% or at least 75%, at least 80%, at least 85%, at least 90%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98% or at least 99% sequence identity with the amino acid sequence of SEQ ID NO:3. However, a cell according according to the present invention may comprise a nucleotide sequence encoding a xylose isomerase having a sequence which has at least about 50%, at least about 55%, at least about 60% or at least about 70% sequence identity with the amino acid sequence set out in SEQ ID NO: 3.

Sequence identity (or sequence similarity) is herein defined as a relationship between two or more amino acid (polypeptide or protein) sequences or two or more nucleic acid (polynucleotide) sequences, as determined by comparing the sequences. Usually, sequence identities or similarities are compared, typically over the whole length of the sequences compared. However, sequences may be compared over shorter comparison windows. In the art, "identity" also means the degree of sequence relatedness between amino acid or nucleic acid sequences, as the case may be, as determined by the match between strings of such sequences.

Preferred methods to determine identity are designed to give the largest match between the sequences tested. Methods to determine identity and similarity are codified in publicly available computer programs. Preferred computer program methods to determine identity and similarity between two sequences include e.g. the BestFit, BLASTP, BLASTN, and FASTA (Altschul, S. F. et al., J. Mol. Biol. 215:403-410 (1990), publicly available from NCBI and other sources (BLAST Manual, Altschul, S., et al., NCBI NLM NIH Bethesda, MD 20894). Preferred parameters for amino acid sequences comparison using BLASTP are gap open 11.0, gap extend 1, Blosum 62 matrix. Preferred parameters for nucleic acid sequences comparison using BLASTP are gap open 11.0, gap extend 1, DNA full matrix (DNA identity matrix).

Optionally, in determining the degree of amino acid similarity, the skilled person may also take into account so-called "conservative" amino acid substitutions, as will be clear to the skilled person.

Conservative amino acid substitutions refer to the interchangeability of residues having similar side chains. For example, a group of amino acids having aliphatic side chains is glycine, alanine, valine, leucine, and isoleucine; a group of amino acids having aliphatic-hydroxyl side chains is serine and threonine; a group of amino acids having amide-containing side chains is asparagine and glutamine; a group of amino acids having aromatic side chains is phenylalanine, tyrosine, and tryptophan; a group of amino acids having basic side chains is lysine, arginine, and histidine; and a group of amino acids having sulphur-containing side chains is cysteine and methionine.

Preferred conservative amino acids substitution groups are: valine-leucine-isoleucine, phenylalanine-tyrosine, lysine-arginine, alanine- valine, and asparagine-glutamine. Substitutional variants of the amino acid sequence disclosed herein are those in which at least one residue in the disclosed sequences has been removed and

a different residue inserted in its place. Preferably, the amino acid change is conservative. Preferred conservative substitutions for each of the naturally occurring amino acids are as follows: Ala to ser; Arg to lys; Asn to gln or his; Asp to glu; Cys to ser or ala; Gln to asn; Glu to asp; Gly to pro; His to asn or gln; He to leu or val; Leu to ile or val; Lys to arg; gln or glu; Met to leu or ile; Phe to met, leu or tyr; Ser to thr; Thr to ser; Trp to tyr; Tyr to trp or phe; and, Val to ile or leu.

A nucleotide sequence encoding an enzyme which catalyses the conversion of xylose to xylulose according to the invention may also be defined by its capability to hybridise with the nucleotide sequences encoding the enzyme having the sequence set out in SEQ ID NO: 3 or a sequence having at least about 70% sequence identity therewith, under moderate, or preferably under stringent hybridisation conditions.

Formally, such nucleotide sequences hybridize with the reverse complement of the nucleotide sequences which encode the enzyme having the sequence set out in SEQ ID NO: 3 or a sequence having at least about 70% sequence identity therewith, for examples sequences which hybridize with the reverse complement of SEQ ID NOs: 1 or 2.

Stringent hybridisation conditions are herein defined as conditions that allow a nucleic acid sequence of at least about 25, preferably about 50 nucleotides, 75 or 100 and most preferably of about 200 or more nucleotides, to hybridise at a temperature of about 65°C in a solution comprising about 1 M salt, preferably 6 x SSC (sodium chloride, sodium citrate) or any other solution having a comparable ionic strength, and washing at 65°C in a solution comprising about 0.1 M salt, or less, preferably 0.2 x SSC or any other solution having a comparable ionic strength. Preferably, the hybridisation is performed overnight, i.e. at least for 10 hours and preferably washing is performed for at least one hour with at least two changes of the washing solution. These conditions will usually allow the specific hybridisation of sequences having about 90% or more sequence identity.

Moderate conditions are herein defined as conditions that allow a nucleic acid sequences of at least 50 nucleotides, preferably of about 200 or more nucleotides, to hybridise at a temperature of about 45°C in a solution comprising about 1 M salt, preferably 6 x SSC or any other solution having a comparable ionic strength, and washing at room temperature in a solution comprising about 1 M salt, preferably 6 x SSC or any other solution having a comparable ionic strength. Preferably, the

hybridisation is performed overnight, i.e. at least for 10 hours, and preferably washing is performed for at least one hour with at least two changes of the washing solution. These conditions will usually allow the specific hybridisation of sequences having up to 50% sequence identity. The person skilled in the art will be able to modify these hybridisation conditions in order to specifically identify sequences varying in identity between 50% and 90%.

To increase the likelihood that the introduced enzyme is expressed in active form in a cell of the invention, the corresponding encoding nucleotide sequence may be adapted to optimise its codon usage to that of the chosen yeast cell. Several methods for codon optimisation are known in the art. A preferred method to optimise codon usage of the nucleotide sequences to that of the yeast is a codon pair optimization technology as disclosed in WO2006/077258 and/or WO2008/000632. WO2008/000632 addresses codon-pair optimization. Codon-pair optimisation is a method wherein the nucleotide sequences encoding a polypeptide are modified with respect to their codon-usage, in particular the codon-pairs that are used, to obtain improved expression of the nucleotide sequence encoding the polypeptide and/or improved production of the encoded polypeptide. Codon pairs are defined as a set of two subsequent triplets (codons) in a coding sequence.

As a simple measure for gene expression and translation efficiency, herein, the Codon Adaptation Index (CAI), as described in Xuhua Xia, Evolutionary Bioinformatics 2007,: 3 53-58, is used. The index uses a reference set of highly expressed genes from a species to assess the relative merits of each codon, and a score for a gene is calculated from the frequency of use of all codons in that gene. The index assesses the extent to which selection has been effective in moulding the pattern of codon usage. In that respect it is useful for predicting the level of expression of a gene, for assessing the adaptation of viral genes to their hosts, and for making comparisons of codon usage in different organisms. The index may also give an approximate indication of the likely success of heterologous gene expression. In the codon pair optimized genes according to the invention, the CAI is 0.6 or more, 0.7 or more, 0.8 or more, 0.85 or more, 0.87 or more 0.90 or more, 0.95 or more, or about 1.0.

In a cell of the invention, the xylose isomerase is typically heterologous to the cell. That is to say, the xylose isomerase has a sequence which does not naturally occur in the cell in question as part of the organism, cell, genome DNA or RNA

sequence in which it is present. That is to say, the xylose isomerase is exogenous to the cell or does not occur naturally in the cell. Accordingly, a nucleotide sequence encoding a xylose isomerase is typically expressed or is capable of being expressed in active form in the transformed host cell.

A cell of the invention is thus a cell that comprises, i.e. has been transformed with, a nucleic acid construct comprising the nucleotide sequence encoding the xylose isomerase as defined above. The nucleic acid construct comprising the xylose isomerase coding sequence preferably is capable of expression of the xylose isomerase in the host cell.

Methods for expressing a heterologous xylose isomerase sequence in a cell are well known to those skilled in the art.

Accordingly, a 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.

Techniques for the recombinant expression of xylose isomerase 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", 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.

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 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} 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 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 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 xylose isomerase, for instance by multiple copies of a nucleotide construct or by use of construct which has multiple copies of the xylose isomerase sequence.

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 2μ 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 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).

Typically, the xylose isomerase 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 xylose isomerase sequence.

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 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 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.

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 xylose isomerase 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 xylose isomerase 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). Although the of antibiotic resistance markers may be most convenient for the transformation of polyploid host cells, preferably however, non- antibiotic resistance markers are 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, centromers, telomers 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.

Preferably, the xylose isomerase is expressed in the cytosol. Cytosolic expression may be achieved by deletion or modification of a mitochondrial or peroxisomal targeting signal.

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 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 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*, *Acremoniurra*, *Fusarium* or *Penicillium*. More preferably, the filamentous fungal cell may be a *Aspergillus niger, Aspergillus oryzae,* a *Penicillium chrysogenum,* or *Rhizopus oryzae* cell.

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*.

A cell of the invention may be able to convert plant biomass, celluloses, hemicelluloses, pectins, rhamnose, galactose, fucose, 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.

A cell of the invention is preferably is a host capable of active or passive xylose transport into the cell.

Preferably, a cell of the invention:

is capable of active glycolysis; and/or

shows flux through the pentose phosphate pathway; and/or

displays xylulose kinase activity so that the xylulose isomerised from xylose may be metabolised to pyruvate.

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 ß-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.

The nucleotide sequence encoding a xylose isomerase is typically expressed or is capable of being expressed in active form in the transformed host cell. Thus, expression of the nucleotide sequence in the host cell produces an active xylose isomerase, typically with a specific activity of at least about 10 U xylose isomerase activity per mg protein at about 30°C, preferably at least about 20, at least about 25, at least about 30, at least about 50, at least about 100, at least about 200, at least about 300, at least about 500, at least about 750 or at least about 1000 U per mg at about 30°C. The specific activity of the xylose isomerase expressed in the transformed host cell is herein defined as the amount of xylose isomerase activity units per mg protein of cell free lysate of the host cell, e.g. a yeast cell free lysate. Determination of the xylose isomerase activity, amount of protein and preparation of the cell free lysate are as described herein. Preferably, expression of the nucleotide sequence encoding the xylose isomerase in the host cell produces a xylose isomerase with a K_m for xylose that is less than 50, 40, 30 or 25 mM, more preferably, the K_m for xylose is about 20 mM or less.

A cell of the invention may comprise one ore 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 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 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 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 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 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, 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, 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.

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-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 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-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 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 <-> 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 <-> D-erythrose 4-phosphate + D-fructose 6-phosphate and vice versa. The enzyme is also known as dihydroxyacetonetransferase; 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*.

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, preferably under conditions where xylose or xylose and glucose are available as carbon sources, more preferably as major carbon sources (i.e. more than 50% of the available carbon source consists of xylose or xylose and glucose), most preferably as sole carbon sources. Suitable promoters in this context include both constitutive and inducible natural promoters as well as engineered promoters. A preferred promoter for use in the present invention will in addition be insensitive to catabolite (glucose) repression and/or will preferably not require xylose for induction. Promotors having these characteristics 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 (ADHI, 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.

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 host 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. Overexpression of an enzyme is thus preferably determined by measuring the level of the enzyme's specific activity in the host cell using appropriate enzyme assays as described herein. Alternatively, overexpression of the enzyme may be determined indirectly by quantifying the specific steady state level of enzyme protein, e.g. using antibodies specific for the enzyme, or by quantifying the specific steady level of the mRNA coding for the enzyme. The latter may particularly be suitable for enzymes of the pentose phosphate pathway for which enzymatic assays are not easily feasible as substrates for the enzymes are not commercially available. 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.

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 ATP + D-xylulose = ADP + 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.

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. 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 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.

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.

In the host cells of the invention, genetic modification that reduces unspecific aldose reductase activity in the host cell may be combined with any of the modifications increasing the flux of the pentose phosphate pathway and/or with any of the modifications increasing the specific xylulose kinase activity in the host cells as described above. This is not, however, essential.

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 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 or xylulose to xylitol. Unspecific aldose reductases catalyse the reaction:

aldose + NAD(P)H + H^+ ↔ alditol + $NAD(P)^+$

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.

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 hybridising to a reference nucleotide sequence encoding an aldose reductase.

A cell of the invention may be adapted to xylose utilisation by selection of mutants, either spontaneous or induced (e.g. by radiation or chemicals), for growth on xylose, preferably on xylose as sole carbon source, and more preferably under anaerobic conditions. Selection of mutants may be performed by techniques including serial passaging 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. In a preferred host cell of the 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 xylose as carbon source, preferably as sole carbon source, and preferably under anaerobic conditions. Preferably the modified host 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.

A cell of the invention may have the ability to grow on xylose as sole carbon source at a rate of at least about 0.05, about 0.1, about 0.2, about 0.25 or about 0.3 h^{-1} under aerobic conditions, or, if applicable, at a rate of at least about 0.03, about 0.05, about 0.07, about 0.08, about 0.09, about 0.1, about 0.12, about 0.15 or about 0.2 h^{-1} under anaerobic conditions. Preferably the modified host cell has the ability to grow on a mixture of glucose and xylose (in a 1:1 weight ratio) as sole carbon source at a rate of at least about 0.05, about 0.1, about 0.2, about 0.25 or about 0.3 h^{-1} under aerobic conditions, or, if applicable, at a rate of at least about 0.03, about 0.05, about 0.1, about 0.12, about 0.15, or about 0.2 h^{-1} under anaerobic conditions.

A cell of the invention may have a specific xylose consumption rate of at least about 200, about 250, about 300, about 346, about 350, about 400, about 500, about 600, about 750, or about 1000 mg xylose/g cells/h. A cell of the invention may have a yield of fermentation product (such as ethanol) on xylose that is at least about 40, about 50, about 55, about 60, about 70, about 80, about 85, about 90, about 95 about 98 or about 99% of the host cell's yield of fermentation product (such as ethanol) on glucose. More preferably, the yield of a fermentation product (such as ethanol) of a cell of the invention on xylose may be equal to the cell's yield of fermentation product (such as ethanol) on glucose. Likewise, the cell's biomass yield on xylose may be at least about 40, about 50, about 55, about 60, about 70, about 80, about 85, about 90, about 95, about 98 or about 99% of the host cell's biomass yield on glucose. More preferably, the cell's biomass yield on xylose, may be equal to the host cell's biomass yield on glucose. It is understood that in the comparison of yields on glucose and xylose both yields are compared under aerobic conditions or both under anaerobic conditions.

A cell of the invention may be capable of using arabinose. A cell of the invention may, 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 is order that it is capable of using arabinose. Such an approach is described in WO2003/095627.

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 ß-lactam antibiotic or a cephalosporin. A preferred modified host 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 modified host 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 modified host cell. In the latter case the modified host 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 modified host cell ferments both the xylose and glucose, preferably simultaneously in which case preferably a modified host 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 modified host 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.

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 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 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 modified host 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.

A preferred process is a process for the production of a ethanol, whereby the process comprises the steps of: (a) fermenting a medium containing a source of xylose with a modified host cell as defined above, whereby the host cell ferments xylose to ethanol; and optionally, (b) recovery of the ethanol. The fermentation medium may also comprise a source of glucose that is also fermented to ethanol. In the process the volumetric ethanol productivity is preferably at least about 0.5, about 1.0, about 1.5, about 2.0, about 2.5, about 3.0, about 5.0 or about 10.0 g ethanol per litre per hour. 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, such as a product selected from the group consisting of 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 and a cephalosporin. The process preferably comprises fermenting a medium containing a source of xylose with a modified host cell as defined herein above, whereby the host cell ferments xylose to the fermentation product.

The invention also provides a process for producing a fermentation product, such as a product selected from the group consisting of 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 and a cephalosporin. The process preferably comprises fermenting a medium containing at least a source of xylose and a source of L-arabinose with a cell as defined above which is able to use both of xylose and L-arabinose such that the cell ferments xylose and L-arabinose to the fermentation product.

The invention also provides a process for producing a fermentation product, such as a product selected from the group consisting of 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 and a cephalosporin. The process preferably comprises fermenting a medium containing at least a source of xylose and a source of L-arabinose with a cell as defined above and a cell able to use L-arabinose, whereby each cell ferments xylose and/or arabinose to the fermentation product.

A process of the invention may also comprise recovery of the fermentation product. The medium with which the process is carried out may also contain a source of glucose.

The 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.

The following Examples illustrate the invention:

EXAMPLES

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

Xylose isomerise activity (as determined in examples 1 and 2)

Xylose isomerase activity may be assayed at 37°C in a reaction mixture containing 50 mM phosphate buffer (pH 7.0), 10 mM xylose, 10 mM $MgCl_2$ and a suitable amount of cell-free extract. The amount of xylulose formed may be determined by the cysteine-carbazole method (Goldstein and McCusker, Yeast 15, 1541-1553, 1999). Alternatively, xylose isomerase activity is assayed at 30°C using the enzyme assay of Kersters-Hildersson *et al.* (Kinetic characterization of D-xylose isomerases by enzymatic assays using D-sorbitol dehydrogenase. Enz. Microb. Technol. 9 (1987) 145-148). The *in vitro* activity of xylose isomerase in the cell-free extracts of transformed *S. cerevisiae* strains is dependent on bivalent cations (Mg_2^+ or Co_2^+).

Transformation of *S. cerevisiae*

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

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.

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	
step 5	4'	72°C	
step 6	30"	20°C	

Sample pretreatment for xylose isomerase activity determinations (general herein and in example 3)

0.5 ml of 0.1 M MOPS buffer (pH 7.5) was added to the cell pellet of an overnight culture. The cells were resuspended and transferred to a 2 ml Eppendorf tube which already contained 0.5 g of glassbeads with a diameter of 0.4-0.5 mm. All samples were vigorously shaked in an Eppendorf tube shaker (IKA VIBRAX-VXR) for 20 min at 4°C, at maximal speed. The extract was centrifuged for 5 minutes at 14000 rpm and 4°C. The supernatant, which is the cell free extract, was transferred into a fresh Eppendorf tube.

Assay conditions xylose isomerase activity assay (general herein and as determined in example 3).

The following method is a modified version of the method described by Dische-Borenfreud (J. Biol. Chem. (1951) 192, 2, 583-587). One (1.0) ml of the substrate mix (100 mM MOPS pH 7.5, 10 mM $MgCl_2$, 10 mM D-xylose) was mixed with 50 µl (diluted) cell free extract, in duplicate, on ice. Subsequently the reaction tubes were placed in a 50°C water bath for 30 minutes. In addition, the reactions were carried out at 30°C, also in duplicate. The reaction was stopped by placing the reaction tubes on ice water, followed by addition of 0.2 ml 1.67% L-cysteine monohydrate hydrochloride (Merck) solution. The mixture is then well mixed by vortexing. Subsequently, 6 ml of H_2SO_4 solution (190 ml water with 450 ml 95-97% concentrated H_2SO_4) was added, immediately followed by 0.2 ml of 0.12% (w/v) carbazole (Merck), dissolved in ethanol. This final mixture was mixed well by vortexing and left at room temperature for 60 min. The absorption is measured at 560 nm using plastic cuvettes.

D(+)-fructose, which is also a ketose, was used as a reference. To this end, approximately 1000 mg D-fructose was weighed accurately and dissolved in 0.1 M

MOPs buffer, pH 7.5 in a 50 ml volumetric flask. A series of dilutions was made ranging from approximately 2 to 20 µmole/ml. 50 µl of these fructose solutions were used in the assay as described above and the absorption at 560 nm was used to make a calibration curve. The activity of the samples was calculated by relating the absorbance at 560 nm to the calibration curve.

The protein concentration of the sample was determined according to a modified protocol of the Bradford method, using the Coomassie Plus Protein Assay (Thermo Scientific). The specific activity of xylose isomerase is expressed as nmol /mg protein.min.

Example 1

Expression of xylose isomerase from Bacteroides uniformis ATCC 8492 *in Saccharomyces cerevisiae*

1.1.1 Construction of xylose isomerase expression vector

Xylose isomerase [E.C. 4.2.1.2], GenBank accession number AAYH02000036 (SEQ ID NO: 1) from *Bacteroides uniformis* ATCC 8492 was analysed for the codon usage. The codon use was optimized as described in WO2006/077258 and WO2008/000632 (SEQ ID NO: 2).

The gene according to SEQ ID NO: 2 was cloned in front of the *TPI1*-promoter of *S.cerevisiae*. In order to prevent potential inefficient expression of the xylose isomerase, the following sequence was placed in front of the coding sequence:

ACTAGTAAAAACACATACATAAACTAAAAATG,

showing the start codon underlined.

A *SpeI* restriction site ACTAGT) was introduced in the strong, constitutive *TPI1*-promoter, changing the sequence

TCTTGCTTAAATCTATAACTACAAAAAACACATACATAAACTAAAAATG

(original *TPI1* promoter) into

TCTTGCTTAAATCTATA**ACTAGT**AAAAACACATACATAAACTAAAAATG .

This allows for operably linking the codon optimized xylose isomerase coding sequence to the *TPI1*-promoter.

In addition, the termination codon TAA was changed into TAAG, which is the most efficient termination codon in yeast. Convenient restriction sites were added to facilitate cloning. The sequence is synthesized by GeneArt AG (Regensburg, Germany).

The final yeast expression construct pYISIT4-XKS1-xylA (Baun CpO) is set out in Figure 1.

1.2 Yeast transformation

S. cerevisiae strain CEN.PK113-7D (*MATa URA3 HIS3 LEU2 TRP1 MAL2-8 SUC2*) and a derivative of CEN.PK113-7D, in which the *GRE3*-gene was replaced by the genes of the non-oxidative part of the pentose phosphate pathway (see above) (*MATa URA3 HIS3 LEU2 TRP1 MAL2-8 SUC2 GRE3::*[*TPI1p-TAL1_ADH1p-TKL1_PGI1p-RPE1_ENO1p-RKI1*]) are transformed with the construct pYISIT4-XKS1-xylA (Baun CpO). Transformation mixtures are plated on Yeast Carbon Base (YCB) w/o ammonium sulphate (Difco), 40mM KPi (pH 6.8) and 5mM acetamide. Untransformed cells cannot grow on this medium.

Transformants are characterized using PCR techniques and/or Southern blotting techniques.

Example 2

Growth of transformed yeast strains on xylose

2.1 Medium composition

Growth experiments: *Saccharomyces cerevisiae* strains are grown on medium having the following composition: 0.67% (w/v) yeast nitrogen base and either glucose, galactose or xylose, or a combination of these substrates (see below). For agar plates the medium is supplemented with 2% (w/v) bacteriological agar.

Ethanol production: Shake-flask cultivations were performed at 30°C in a synthetic medium (Verduyn *et al.*, Yeast 8:501-517, 1992). The pH of the medium was

adjusted to 6.0 with 2 M KOH prior to sterilisation. For solid synthetic medium, 1.5% of agar was added.

Pre-cultures were prepared by inoculating 100 ml medium containing the appropriate sugar in a 500-ml shake flask with a frozen stock culture. After incubation at 30°C in an orbital shaker (200 rpm), this culture was used to inoculate either shake-flask cultures. The synthetic medium for anaerobic cultivation was supplemented with 0.01 g l-1 ergosterol and 0.42 g l-1 Tween 80 dissolved in ethanol (Andreasen and Stier. J. Cell Physiol. 41:23-36, 1953; and Andreasen and Stier. J. Cell Physiol. 43:271-281, 1954).

2.2 Growth experiments

Saccharomyces cerevisiae strain CEN.PK113-7D or the derivative constitutively expressing the PPP (see Example 1), transformed with pYISIT4-XKS1-xylA (Baun CpO), are grown on agar plates with 2% glucose as carbon source. When colonies are visible, single colonies are used to inoculate liquid medium with 100 mM xylose, 100 mM glucose and 100 mM galactose as carbon sources, or combinations thereof. Growth is monitored by measuring the increase in optical density at 600 nm on a LKB Ultrospec K spectrophotometer.

2.3 Ethanol production

Saccharomyces cerevisiae strain CEN.PK113-7D or the derivative constitutively expressing the PPP (see Example 1), transformed with pYISIT4-XKS1-xylA (Baun CpO), are grown on agar plates with 2% glucose as carbon source. When colonies were visible, single colonies are used to inoculate a synthetic medium (Verduyn et al., supra). Mixtures of glucose, xylose and or galactose are added to the medium as a carbon source, ranging from 0 to 50 grams per liter. Growth is monitored by measuring the increase in optical density at 600 nm on a LKB Ultrospec K spectrophotometer. Ethanol production and sugar consumption in time are monitored by HPLC and/or NMR analysis.

Example 3

3.1 Introduction of four constitutively expressed genes of the non-oxidative pentose phosphate pathway

Saccharomyces cerevisiae BIE104P1, expressing the genes *TAL1*, *TKL1*, *RKI1* and *RPE1* constitutively, was obtained by transforming CEN.PK113-7D (*MATa URA3 HIS3 LEU2 TRP1 MAL2-8 SUC2*) with plasmid pPWT080 (figure 2). To a large extent, plasmid pPWT080 was constructed by using synthetic DNA, synthesized by GeneArt AG (Regensburg, Germany). The sequence of plasmid pPWT080 is set out in SEQ ID 4. In short, plasmid pPWT080 consists of the promoter region of the *GRE3*-gene, followed by the four PPP-genes *TAL1*, *TKL1*, *RKI1* and *RPE1* under control of strong constitutive promoters, and the 3' non-coding sequences of the *GRE3*-gene, as set out in figure 2. As selectable markers, the kanMX-gene conferring resistance to G418 and the *Aspergillus amdS*-gene allowing the transformants to grow in acetamide as sole nitrogen source are present on this plasmid. Upon integration, followed by intramolecular recombination, the markers are lost and the integration of this construct leads to inactivation of the coding region of the *GRE3*-gene and the overexpression of the genes *TAL1*, *TKL1*, *RPE1* and *RKI1*.

Prior to the transformation of CEN.PK113-7D, pPWT080 was linearized using the restriction enzyme *Sfi*I (New England Biolabs), according to the instructions provided by the supplier. Transformation mixtures were plated on YPD (per liter: 10 grams of yeast extract, 20 grams per liter peptone, 20 grams per liter dextrose, 20 grams of agar) containing 100 µg G418 (Sigma Aldrich) per ml.

After two to four days, colonies appeared on the plates, whereas the negative control (i.e. no addition of DNA in the transformation experiment) resulted in blank YPD/G418-plates.

The integration of plasmid pPWT080 is directed to the *GRE3*-locus. Transformants were characterized using PCR and Southern blotting techniques.

PCR reactions, which are indicative for the correct integration of one copy of plasmid pPWT080, were performed with the primers indicated by SEQ ID 5 and 6, and 6 and 7 (see figure 3). With the primer pairs of SEQ ID 5 and 6, the correct integration

at the *GRE3*-locus was checked. If plasmid pPWT080 was integrated in multiple copies (head-to-tail integration), the primer pair of SEQ ID 6 and 7 will give a PCR-product. If the latter PCR product is absent, this is indicative for a one copy integration.

In order to verify the correct one copy integration in transformants identified as such using the above described PCR technique, a Southern blot analysis was performed. To this end, the chromosomal DNA was isolated from the wild-type strain CEN.PK113-7D and transformants using standard molecular biology techniques. The chromosomal DNA was digested with the restriction enzymes *Xcm*I and *Psi*I, electroforesed over a 0.7% agarose gel and the DNA was transferred to a nylon membrane (Hybond N+, Amersham Pharmacia Biotech) according to the instructions of the manufacturer.

As a probe for detecting the correct integration of the plasmid pPWT080, a probe derived from the *RKI1*-gene, present in plasmid pPWT080, was used. The probe was made by using the primers of SEQ ID 9 and 10 and plasmid pPWT080 as a template. The labeling of the probe and the subsequent hybridization and washing procedures were performed as suggested by the supplier of the ECL Direct Labeling and Detection System (GE Life Sciences).

The autoradiogram, as presented in figure 4, shows correct integration of one copy of plasmid pPWT080, in accordance with the expected hybridisation pattern as can be deduced from figure 3 (panel c). The strain was designated BIE104F1.

In order to be able to introduce the genes encoding xylose isomerase and xylulokinase (section 3.2), it is necessary to remove the selection markers introduced by the integration of plasmid pPWT080. The design of plasmid pPWT080 was such, that upon integration of pPWT080 in the chromosome, homologous sequences are in close proximity of each other. This design allows the selectable markers to be lost by spontaneous intramolecular recombination of these homologous regions. The removal of the markers from the strain results in a marker free strain that is more stable in its use, than a strain containing markers. More specifically, the promoter region of the *GRE3*-gene and the 3' non-coding region of the *GRE3*-gene are duplicated after integration of one copy of pPWT080 at the *GRE3*-locus of *S. cerevisiae*. Upon vegetative growth, intramolecular recombination will take place, although at low

frequency. The frequency of this recombination depends on the length of the homology and the locus in the genome (unpublished results). Upon sequential transfer of a subfraction of the culture to fresh medium, intramolecular recombinants will accumulate in time.

To this end, strain BIE104F1 was cultured in YPD-2% glucose, starting from a colony isolate. 25 µl of an overnight culture was used to inoculate fresh YPD-2% glucose medium. After five serial transfers, the optical density of the culture was determined and cells were diluted to a concentration of approximately 5000 per ml. 100 µl of the cell suspension was plated on Yeast Carbon Base medium (Difco) containing 30 mM KPi (pH 6.8), 0.1% $(NH_4)_2SO_4$, 40 mM fluoro-acetamide (Amersham) and 1.8% agar (Difco). Cells identical to cells of strain BIE104F1, i.e. without intracellular recombination, still contain the *amd*S-gene. To those cells, fluoro-acetamide is toxic. These cells will not be able to grow and will not form colonies on a medium containing fluoro-acetamide. However, if intramolecular recombination has occurred, BIE104F1-variants that have lost the selectable markers will be able to grow on the fluoro-acetamide medium, since they are unable to convert fluoro-acetamide into growth inhibiting compounds. Those cells will form colonies on this agar medium.

The thus obtained fluoro-acetamide resistant colonies were subjected to PCR analysis using primers of SEQ ID 5 and 6, and 7 and 8. Primers of SEQ ID 5 and 6 will give a band if recombination of the selectable markers has taken place as intended, as set out in figure 5. As a result, the coding region of the *GRE3*-gene is replaced by the four genes *TKL1, TAL1, RKI1* and *RPE1*. In that case, a PCR reaction using primers of SEQ ID 7 and 8 should not result in a PCR product, since primer 7 primes in a region that should be out-recombined (see figure 3, panel b). If a band is obtained with these primers, this is indicative for the presence of the complete plasmid pPWT080 in the genome, so no recombination has taken place.

If primers of SEQ ID 5 and 6 do not result in a PCR product, recombination has taken place, but in such a way that the complete plasmid pPWT080 has recombined out of the genome. Not only were the selectable markers lost, but also the four PPP-genes. In fact, wild-type yeast has been retrieved.

Isolates that exhibited the expected PCR results, were subjected to Southern blot analysis (vide supra). The result is presented in figure 4. One of the strains that showed the correct pattern of bands on the Southern blot (as can be deduced from figure 3) is the strain designated as BIE104P1.

3.2 Introduction of constitutively expressed genes encoding xylose isomerase and xylulokinase

Plasmid pYISIT4-XKS1-xylA (Baun CpO), as set out in figure 1, was improved in order to allow for G418 selection of the transformants. To this end, a 4630 bp insert containing the *xylA*-gene under control of the *TPI1*-promoter and the *XKS1*-gene under control of the *TDH1*-promoter was excised from plasmid pYISIT4-XKS1-xylA (Baun) (figure 1), using the restriction enzymes *Mlu*I and *Sac*II.

Plasmid pYI#SIT4, as set out in figure 6, was digested with restriction enzyme *Acc65*I.

The kanamycin-resistance marker (kanMX) present on plasmid p427TEF (Dualsystems Biotech AG), allowing selection in *E. coli* (kanamycin) and *S. cerevisiae* (G418) was isolated by PCR using primers of SEQ ID 11 and 12. The sequence of primer of SEQ ID 12 was designed in such a way that the *Mlu*I-site in the kanMX-fragment was lost, which keeps the *Mlu*I-site in the resulting plasmid (pPWT007, see below) unique. The PCR product was subcloned in the pCRII-TOPO vector using the Zero Blunt® TOPO PCR Cloning Kit for Sub-cloning (Invitrogen). Correct clones were used to excise the kanMX-resistance marker using the restriction enzyme *Acc65*I. Ligation of this fragment with the digested plasmid pYI#SIT4 resulted in pPWT007, which is set out in figure 7.

Plasmid pPWT007 was cleaved with the restriction enzymes *Mlu*I and *Sac*II. After clean-up of this vector, the above described 4630 bp *Mlu*I-*Sac*II fragment of pYISIT4-XKS1-xylA (Baun) was ligated. The resulting plasmid is called pPWT042, which is set out in figure 8.

Strain BIE104P1 (*MATa URA3 HIS3 LEU2 TRP1 MAL2-8 SUC2 ΔGRE3::*[*TPI1p-TAL1-ADH1p-TKL1-PGI1p-RPE1-ENO1p-RKI1*]) (see section 3.1) was transformed with plasmid pPWT042. Prior to the transformation of BIE104P1, pPWT042 was linearized using the restriction enzyme *Sfi*I, according to the instructions

provided by the supplier. Transformation mixtures were plated on YPD (per liter: 10 grams of yeast extract, 20 grams per liter peptone, 20 grams per liter dextrose, 20 grams of agar) containing 100 µg G418 (Sigma Aldrich) per ml.

After two to four days, colonies appeared on the plates, whereas the negative control (i.e. no addition of DNA in the transformation experiment) resulted in blank YPD/G418-plates.

Upon digestion of plasmid pPWT042 with *Sfi*I, its integration is directed to the *SIT4*-locus (Gottlin-Ninfa and Kaback (1986) Molecular and Cellular Biology Vol. 6, No. 6, 2185-2197) in the genome. Transformants were characterized using PCR and Southernblotting techniques.

PCR reactions, using Phusion® DNA polymerase (Finnzymes), which are indicative for the correct integration of one copy of plasmid pPWT042, were performed with the primers indicated by SEQ IDs 13 and 14, and 14 and 15.

As set out in figure 9, with primer pair SEQ ID 13 and 14, the correct integration at the *SIT4*-locus was checked. The correct integration of the plasmid in the *SIT4*-locus may also be checked with primer pair SEQ ID 15 and 16 (figure 9). If plasmid pPWT042 was integrated in multiple copies (head-to-tail integration), the primer pair of SEQ ID 14 and 15 will give a PCR-product. If the latter PCR product is absent, this is indicative for one copy integration of plasmid pPWT042

A strain with one copy of plasmid pPWT042 integrated into the genome was designated BIE104P1Y9.

3.3 Growth experiments

Single colony isolates of strains BIE104P1 and BIE104P1Y9 were used to inoculate YNB-medium (Difco) supplemented with 2% glucose. The inoculated flasks were incubated for approximately 16 hours at 30 °C and 280 rpm. The optical density at 600 nm of the overnight cultures was determined. YNB-medium supplemented with 1% glucose and 1% xylose was inoculated with the overnight cultures at a starting OD600 of 0.2. Cells were grown overnight at 30°C and 280 rpm. Subsequently, YNB medium containing 2% xylose and 0.1% glucose were inoculated at a starting OD600 of 0.2.

The minute amount of glucose present in the latter medium was consumed rapidly by both strains. Upon transfer to YNB with 2% xylose as sole carbon source, at a starting OD600 of 0.2, only BIE104P1Y9 was able to grow on this medium after a very long lag phase of approximately 4 weeks. If the optical density at 600 nm reached a value of at least 2.0, the cells were transferred to a flask with fresh YNB-medium containing 2% xylose, at a starting OD600 of 0.2.

This was repeated a number of times, as is set it in figure 10. The graph clearly shows that strain BIE104P1Y9 grows rapidly and efficiently on a mineral medium containing 2% xylose as sole carbon source, while a reference strain, missing the integrated plasmid pPWT042, is not capable of doing so.

3.4 Xylose isomerase activity

Single colony isolates of strains BIE104P1 and BIE104P1Y9 were used to inoculate YPD 2% glucose. The inoculated flasks were incubated for approximately 16 hours at 30 °C and 280 rpm. The optical density at 600 nm of the overnight cultures was determined. Cells were harvested by centrifugation. The pellet was washed once with 0.1 M MOPS (3-(N-morpholino)propanesulfonic acid; Sigma) buffer, pH 7.5 and frozen at -20°C until the analysis was performed.

The results of the analysis are summarized in the table below.

Strain	XI-activity at 30°C (nmol/mg protein.min)	XI-activity at 50°C (nmol/mg protein.min)
Reference strain BIE104P1	<20	<20
BIE104P1Y9	110	640

The values are the average of two independent experiments.

3.5 Ethanol production

Single colony isolates of strains BIE104P1 and BIE104P1Y9 were used to inoculate Verduyn-medium (Verduyn *et al.*, Yeast 8:501-517, 1992) supplemented with

2% glucose as sole carbon source. In addition, strain BIE104P1Y9 was inoculated in Verduyn-medium with 2% xylose as sole carbon source. The inoculated flasks were incubated for approximately 64 hours at 30 °C and 280 rpm. The optical density at 600 nm of the cultures was determined. The cells were harvested by centrifugation and the cell pellet was washed with sterile milliQ water (Millipore).

Fresh Verduyn-medium supplemented with 2% glucose and 2% xylose was inoculated with the three precultures described above. The amount of cells inoculated was such that the initial OD600 was 0.2. The flasks were closed with waterlocks, ensuring anaerobic growth conditions after the oxygen was exhausted from the medium and head space.

The flasks were incubated for 72 hours at 30°C and 280 rpm. Samples were taken at 23, 47 and 71 hours for analysis. The following analyses were performed: OD600 determination, NMR analysis (xylose, glucose, ethanol, acetic acid and glycerol). The results are shown in figures 11 and 12 and the tale below. The data represent the residual amount of sugars at the indicated (glucose and xylose in grams per liter) and the formation of (by-)products (ethanol, glycerol and acetic acid).

In figure 11, the development of the optical density at 600 nm (OD600) in time is shown. The reference strain, BIE104P1, reaches its maximum OD600 before or at 23 h after the start of the experiment. Apparently, at or before the 23h time point, the glucose has been exhausted from the medium (figure 12, panel a). Also, the ethanol production has reached its maximum at the moment the glucose has been consumed by this yeast strain. Both growth and ethanol production seize, because this strain cannot utilize and ferment xylose, for it misses the necessary active proteins (i.e. a xylose isomerase and overexpressed xylulokinase).

Strain BIE104P1Y9 however, in which the xylose isomerase derived from *Bacteroides uniformis* and the native xylulokinase are overexpressed, is capable of growing on and fermenting xylose into ethanol (figures 11 and 12). After approximately 1 day of anaerobic cultivation, strain BIE104P1Y9 already consumed some xylose, while all glucose was already consumed. Subsequently, the residual amount of glucose was fermented into ethanol, as is apparent from figure 12 (panel b and c) and the table

below. By-product formation (actetic acid and glycerol) is low, as is apparent from the table below.

The results are not (significantly) influenced by the precultures (glucose or xylose), as is apparent from the results presented in figures 11 and 12.

BIE104P1 pregrown on glucose

Time (h)	Glucose	Xylose	Glycerol	Acetic Acid	Ethanol
0	19,5	19,6	0,0	0,0	0,7
23	0,0	20,8	0,5	0,3	9,0
47	0,0	20,7	0,7	0,7	8,8
71	0,0	19,9	0,7	0,9	8,4

BIE104P1Y9 pregrown on glucose

Time (h)	Glucose	Xylose	Glycerol	Acetic Acid	Ethanol
0	19,5	19,6	0,0	0,0	0,7
23	0,0	16,6	0,6	0,5	11,4
47	0,0	6,3	0,7	0,8	14,3
71	0,0	1,7	0,4	1,1	16,8

BIE104P1Y9 pregrown on xylose

Time (h)	Glucose	Xylose	Glycerol	Acetic Acid	Ethanol
0	19,5	19,6	0,0	0,0	0,7
23	0,7	17,7	0,0	0,5	11,3
47	0,0	6,1	0,7	0,8	14,9
71	0,0	1,1	0,5	1,1	16,7

All values are given in grams per litre.

Based on these results, a Qs of 363 mg xylose per gram biomass, per hour was calculated (time interval 23-47 hours; optical density of 30 equals to 6 grams of dry matter per litre), in case of strain BIE104P1Y9 pregrown on xylose.

Example 4

4.1 *Introduction of the genes araA, araB and araD into the genome of S. cerevisiae*

Plasmid pPWT018, as set out in figure 13, was constructed as follows: vector pPWT006 (figure 14), consisting of a SIT2-locus (Gottlin-Ninfa and Kaback (1986) Molecular and Cell Biology vol. 6, no. 6, 2185-2197) and the markers allowing for selection of transformants on the antibiotic G418 and the ability to grow on acetamide (vide supra), was digested with the restriction enzymes BsiWI and MluI. The genes encoding arabinose isomerase (araA), L-ribulokinase (araB) and L-ribulose-5-phosphate-4-epimerase (araD) from Lactobacillus plantarum, as disclosed in patent application WO2008/041840, were synthesized by GeneArt AG (Regensburg, Germany). One large fragment was synthesized, harbouring the three ara-genes mentioned above, under control of (or operably linked to) strong promoters from S. cerevisiae, i.e. the TDH3-promoter controlling the expression of the araA-gene, the ENO1-promoter controlling the araB-gene and the PGI1-promoter controlling the araD-gene. This fragment was surrounded by the unique restriction enzymes Acc65I and MluI. Cloning of this fragment into pPWT006 digested with MluI and BsiWI, resulted in plasmid pPWT018 (figure 13). The sequence of plasmid pPWT018 is set out in SEQ ID 17.

CEN.PK113-7D (MATa URA3 HIS3 LEU2 TRP1 MAL2-8 SUC2) was transformed with plasmid pPWT018, which was previously linearized with SfiI (New England Biolabs), according to the instructions of the supplier. A synthetic SfiI-site was designed in the 5'-flank of the SIT2-gene (see figure 13). Transformation mixtures were plated on YPD-agar (per liter: 10 grams of yeast extract, 20 grams per liter peptone, 20 grams per liter dextrose, 20 grams of agar) containing 100 µg G418 (Sigma Aldrich) per ml.

After two to four days, colonies appeared on the plates, whereas the negative control (i.e. no addition of DNA in the transformation experiment) resulted in blank YPD/G418-plates.

The integration of plasmid pPWT018 is directed to the SIT2-locus. Transformants were characterized using PCR and Southern blotting techniques.

PCR reactions, which are indicative for the correct integration of one copy of plasmid pPWT018, were performed with the primers indicated by SEQ ID 18 and 15, and 15 and 14 (see figure 3). With the primer pairs of SEQ ID 18 and 15, the correct integration at the SIT2-locus was checked. If plasmid pPWT018 was integrated in multiple copies (head-to-tail integration), the primer pair of SEQ ID 15 and 14 will give a PCR-product. If the latter PCR product is absent, this is indicative for one copy integration of pPWT018. A strain in which one copy of plasmid pPWT018 was integrated in the SIT2-locus was designated BIE104R2.

In order to be able to transform the yeast strain with other constructs, it is necessary to remove the selectable markers. The design of plasmid pPWT018 was such, that upon integration of pPWT018 in the chromosome, homologous sequences are in close proximity of each other. This design allows the selectable markers to be lost by spontaneous intramolecular recombination of these homologous regions.

Upon vegetative growth, intramolecular recombination will take place, although at low frequency. The frequency of this recombination depends on the length of the homology and the locus in the genome (unpublished results). Upon sequential transfer of a subfraction of the culture to fresh medium, intramolecular recombinants will accumulate in time.

To this end, strain BIE104R2 was cultured in YPD-medium (per liter: 10 grams of yeast extract, 20 grams per liter peptone, 20 grams per liter dextrose), starting from a single colony isolate. 25 µl of an overnight culture was used to inoculate fresh YPD medium. After at least five of such serial transfers, the optical density of the culture was determined and cells were diluted to a concentration of approximately 5000 per ml. 100 µl of the cell suspension was plated on Yeast Carbon Base medium (Difco) containing 30 mM KPi (pH 6.8), 0.1% $(NH_4)_2SO_4$, 40 mM fluoro-acetamide (Amersham) and 1.8% agar (Difco). Cells identical to cells of strain BIE104R2, i.e. without intracellular recombination, still contain the amdS-gene. To those cells, fluoro-acetamide is toxic. These cells will not be able to grow and will not form colonies on a medium containing fluoro-acetamide. However, if intramolecular recombination has occurred, BIE104R2-variants that have lost the selectable markers will be able to grow on the fluoro-

acetamide medium, since they are unable to convert fluoro-acetamide into growth inhibiting compounds. Those cells will form colonies on this agar medium.

The thus obtained fluoro-acetamide resistant colonies were subjected to PCR analysis using primers of SEQ ID 18 and 15, and 14 and 19. Primers of SEQ ID 18 and 5 will give a band if recombination of the selectable markers has taken place as intended. As a result, the cassette with the genes araA, araB and araD under control of the strong yeast promoters have been integrated in the SIT2-locus of the genome of the host strain. In that case, a PCR reaction using primers of SEQ ID 14 and 19 should not result in a PCR product, since primer 14 primes in a region that should be out-recombined. If a band is obtained with the latter primers, this is indicative for the presence of the complete plasmid pPWT018 in the genome, so no recombination has taken place.

If primers of SEQ ID 18 and 15 do not result in a PCR product, recombination has taken place, but in such a way that the complete plasmid pPWT018 has recombined out of the genome. Not only were the selectable markers lost, but also the ara-genes. In fact, wild-type yeast has been retrieved.

Isolates that showed PCR results in accordance with one copy integration of pPWT018 were subjected to Southern blot analysis. The chromosomal DNA of strains CEN.PK113-7D and the correct recombinants were digested with EcoRI and HindIII (double digestion). A SIT2-probe was prepared with primers of SEQ ID 20 and 21, using pPW018 as a template. The result of the hybridisation experiment is shown in figure 15. The expected hybridisation pattern may be deduced from the physical maps as set out in figure 16 (panels a and b).

In the wild-type strain, a band of 2.35 kb is observed, which is in accordance with the expected size (figure 16, panel a). Upon integration and partial loss by recombination of the plasmid pPWT018, a band of 1.06 kb was expected (figure 16, panel b). Indeed, this band is observed, as shown in figure 15 (lane 2).

One of the strains that showed the correct pattern of bands on the Southern blot (as can be deduced from figure 15) is the strain designated as BIE104A2.

4.2 *Introduction of four constitutively expressed genes of the non-oxidative pentose phosphate pathway*

Saccharomyces cerevisiae BIE104A2, expressing the genes araA, araB and araD constitutively, was transformed with plasmid pPWT080 (figure 2). The procedure and results were already described in Example 3 (section 3.1). In short, BIE104A2 was transformed with SfiI-digested pPWT080. Transformation mixtures were plated on YPD-agar (per liter: 10 grams of yeast extract, 20 grams per liter peptone, 20 grams per liter dextrose, 20 grams of agar) containing 100 µg G418 (Sigma Aldrich) per ml.

After two to four days, colonies appeared on the plates, whereas the negative control (i.e. no addition of DNA in the transformation experiment) resulted in blank YPD/G418-plates.

The integration of plasmid pPWT080 is directed to the GRE3-locus. Transformants were characterized using PCR and Southern blotting techniques, as described in Example 3, section 3.1.

A transformant showing correct integration of one copy of plasmid pPWT080, in accordance with the expected hybridisation pattern, was designated BIE104A2F1.

In order to be able to introduce the genes encoding xylose isomerase and xylulokinase (section 3.2), it is necessary to remove the selection markers introduced by the integration of plasmid pPWT080. To this end, strain BIE104A2F1 was cultured in YPD-medium, starting from a colony isolate. 25 µl of an overnight culture was used to inoculate fresh YPD-medium. After five serial transfers, the optical density of the culture was determined and cells were diluted to a concentration of approximately 5000 per ml. 100 µl of the cell suspension was plated on Yeast Carbon Base medium (Difco) containing 30 mM KPi (pH 6.8), 0.1% $(NH_4)_2SO_4$, 40 mM fluoro-acetamide (Amersham) and 1.8% agar (Difco). Fluoro-acetamide resistant colonies were subjected to PCR analysis and, in case of correct PCR-profiles, Southern blot analysis (section 3.1 of Example 3). One of the strains that showed the correct pattern of bands on the Southern blot is the strain designated as BIE104A2P1.

4.3 *Introduction of constitutively expressed genes encoding xylose isomerase and xylulokinase*

Strain BIE104A2P1 (MATa URA3 HIS3 LEU2 TRP1 MAL2-8 SUC2 SIT2::[TDH3-araA, ENO1-araB, PGI1-araD] ΔGRE3::[TPI1p-TAL1, ADH1p-TKL1, PGI1p-RPE1, ENO1p-RKI1]) was transformed with plasmid pPWT042. Prior to the transformation of BIE104A2P1, pPWT042 was linearized using the restriction enzyme SfiI, according to the instructions provided by the supplier. Transformation mixtures were plated on YPD-agar (per liter: 10 grams of yeast extract, 20 grams per liter peptone, 20 grams per liter dextrose, 20 grams of agar) containing 100 µg G418 (Sigma Aldrich) per ml.

After two to four days, colonies appeared on the plates, whereas the negative control (i.e. no addition of DNA in the transformation experiment) resulted in blank YPD/G418-plates.

Upon digestion of plasmid pPWT042 with SfiI, its integration is directed to the SIT4-locus (Gottlin-Ninfa and Kaback (1986) Molecular and Cellular Biology Vol. 6, No. 6, 2185-2197) in the genome. Transformants were characterized using PCR and Southernblotting techniques, as described in Example 3 (section 3.2).

A strain with one copy of plasmid pPWT042 integrated into the genome was designated BIE104A2P1Y9.

4.4 *Growth experiments*

Single colony isolates of strains BIE104A2P1Y9 were used to inoculate YNB-medium (Difco) supplemented with 2% glucose or 2% galactose. The inoculated flasks were incubated at 30°C and 280 rpm until the optical density at 600 nm reached a value of at least 2.0.

YNB-medium supplemented with 1% arabinose and 1% xylose was inoculated with the overnight cultures at a starting OD600 of 0.2. Cells were grown at 30°C and 280 rpm. The optical density at 600 nm was monitored regularly. When the optical density reached a value larger than 2.0, an aliquot of the culture was transferred to

fresh YNB medium containing 2% xylose and 0.2% arabinose. The amount of cells added was such that the starting OD600 of the culture was 0.2.

The optical density was monitored regularly. The results are shown in figure 17, panel a (precultures on galactose) and panel b (precultures on glucose).

The results clearly show that the strains are capable of utilizing both arabinose and xylose.

CLAIMS

1. A cell which comprises a nucleotide sequence encoding a xylose isomerase, wherein the amino acid sequence of the xylose isomerase has at least about 70% sequence identity to the amino acid sequence set out in SEQ ID NO: 3 and wherein the nucleotide sequence is heterologous to the host.

2. A cell according to claim 1 which is a eukaryotic cell.

3. A cell according to claim 1 or 2 which is a yeast cell.

4. A cell according to claim 3 which is a yeast cell of the genus *Saccharomyces*, *Kluyveromyces*, *Candida*, *Pichia*, *Schizosaccharomyces*, *Hansenula*, *Klockera*, *Schwanniomyces* or *Yarrowia*.

5. A cell according to claim 4, wherein the yeast cell is of the species *S. cerevisiae, S. bulderi, S. barnetti, S. exiguus, S. uvarum, S. diastaticus, K. lactis, K. marxianus* or *K. fragilis*.

6. A cell according to claim 1 or 2 which is a filamentous fungal cell.

7. A cell according to claim 6 wherein the filamentous fungal cell is of the genus *Aspergillus, Penicillium, Rhizopus, Trichoderma, Humicola, Acremonium* or *Fusarium*

8. A cell according to claim 7, wherein the filamentous fungus cell is of the species *Aspergillus niger, Aspergillus oryzae, Penicillium chrysogenum,* or *Rhizopus oryzae*.

9. A cell according to any one of the preceding claims, wherein the cell comprises one or more genetic modifications resulting in:
   a. an increase in transport of xylose in the cell;
   b. an increase in xylulose kinase activity;

c. an increase in flux through the pentose phosphate pathway;
d. a decrease in aldose reductase activity;
e. a decrease in sensitivity to catabolite repression;
f. an increase in tolerance to ethanol, osmolarity or organic acids; or
g. a reduced production of by-products.

10. A cell according to claim 9, wherein the one or more genetic modifications result in overexpression of at least one gene encoding an enzyme of the non-oxidative part of the pentose phosphate pathway.

11. A cell according to claim 10, wherein the gene is a gene encoding a ribulose-5-phosphate isomerase, a ribulose-5-phosphate epimerase, a transketolase or a transaldolase.

12. A cell according to claim 10 or 11, wherein the one ore more genetic modifications result in overexpression of at least the genes encoding a transketolase and a transaldolase.

13. A cell according to any one of claims 9 to 12, wherein the one or more genetic modifications result in overexpression of a gene encoding a xylulose kinase.

14. A cell according to any one of claims 8 to 13, wherein the gene that is overexpressed is a gene which is endogenous to the cell.

15. A cell according to any one of claims 9 to 14, wherein the one or more genetic modifications result in a decrease in unspecific aldose reductase activity in the cell.

16. A cell according to claim 15, wherein the one or more genetic modifications reduce the expression of an endogenous gene which encodes an unspecific aldose reductase or reduce the activity of the said unspecific aldose reductase.

17. A cell according to claim 16, wherein the gene is inactivated by deletion of at least part of the gene or by disruption of the gene.

18. A cell according to claim 16 or 17, wherein the expression of each gene in the cell that encodes an unspecific aldose reductase is reduced.

19. A cell according to any one of the preceding claims which has the ability to use L-arabinose.

20. A cell according to any of claims 9 to 19, wherein the genes *TAL1*, *TKL1*, *RPE1* and *RKI1* are overexpressed.

21. A cell according to any of claims 9 to 20, wherein the coding region of the *GRE3*-gene is inactivated by replacement of the coding region with a nucleotide sequence comprising the genes *TAL1*, *TKL1*, *RPE1* and *RKI1*.

22. A cell according to any of claims 9 to 21, wherein the genes *araA, araB and araD* from *Lactobacillus plantarum* are expressed.

23. A cell according to any of claims 9 to 22, wherein all expressed genes are constitutively expressed or constitutively overexpressed.

24. A cell according to claim 23 wherein one or more constitutively expressed or constitutively overexpressed genes are stably integrated into the genome of the cell.

25. A cell according to claim 24, wherein all constitutively expressed or constitutively overexpressed genes are stably integrated into the genome of the cell.

26. A process for producing a fermentation product which process comprises fermenting a medium containing a source of xylose with a cell according to any one of the preceding claims such that the cell ferments xylose to the fermentation product.

27. A process for producing a fermentation product which process comprises fermenting a medium containing at least a source of xylose and a source of L-arabinose with a cell as defined in any of claims 19 to 23 such that the cell ferments xylose and L-arabinose to the fermentation product.

28. A process for producing a fermentation product which process comprises fermenting a medium containing at least a source of xylose and a source of L-arabinose with a cell as defined in any one of the claims 1 to 18 and a cell able to use L-arabinose, whereby each cell ferments xylose and/or arabinose to the fermentation product.

29. A process according to any one of claims 26 to 28, which comprises recovering the fermentation product.

30. A process according to any one of claims 26 to 29 wherein the medium also contains a source of glucose.

31. A process according to any one of claims 28 to 30, wherein the fermentation product is 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, butanol, a β-lactam antibiotic and a cephalosporin.

32. A process according to any one of claims 26 to 31, wherein the process is anaerobic.

33. A process according to any one of claims 26 to 32, wherein the process is aerobic, preferably performed under oxygen limited conditions.

34. Use of a cell of the invention in a process for the production of a fermentation product.

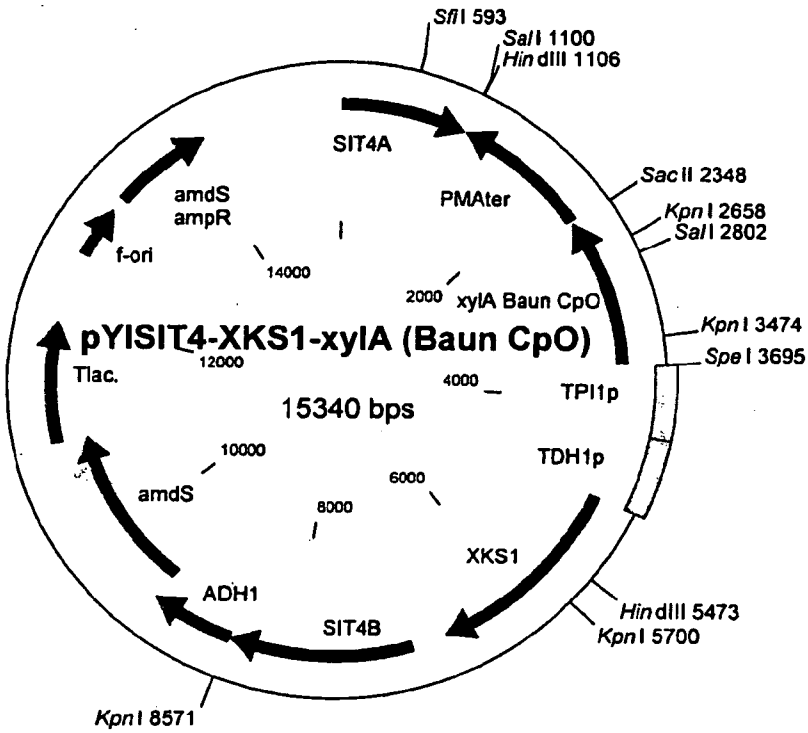

SfiI 593
SalI 1100
HindIII 1106
SIT4A
SacII 2348
PMAter
KpnI 2658
SalI 2802
amdS
ampR
f-ori
14000
2000
xylA Baun CpO
pYISIT4-XKS1-xylA (Baun CpO)
KpnI 3474
SpeI 3695
12000
Tlac.
4000
TPI1p
15340 bps
10000
TDH1p
6000
amdS
8000
XKS1
ADH1
HindIII 5473
KpnI 5700
SIT4B
KpnI 8571

Figure 1

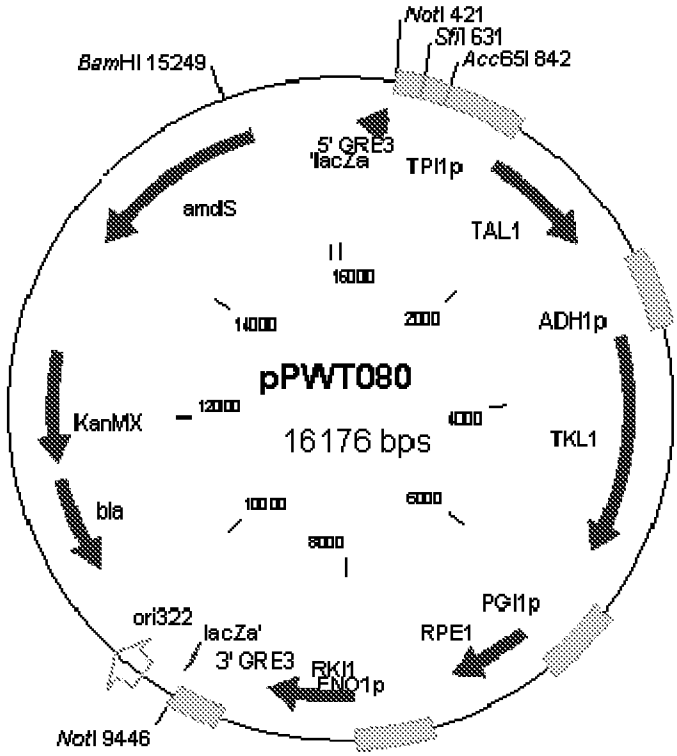
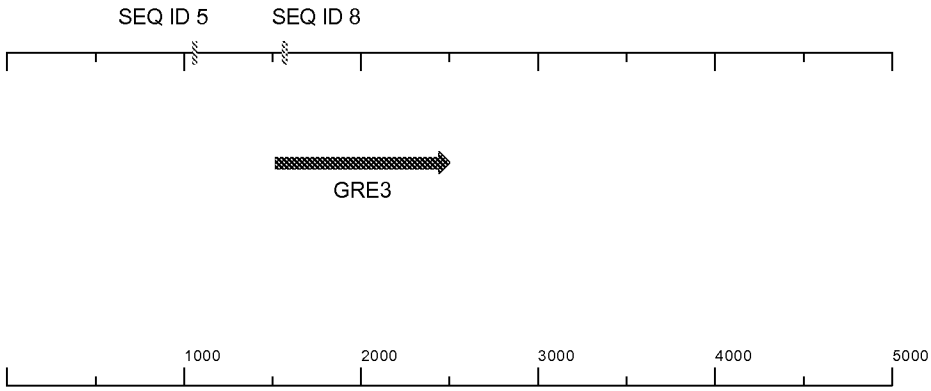
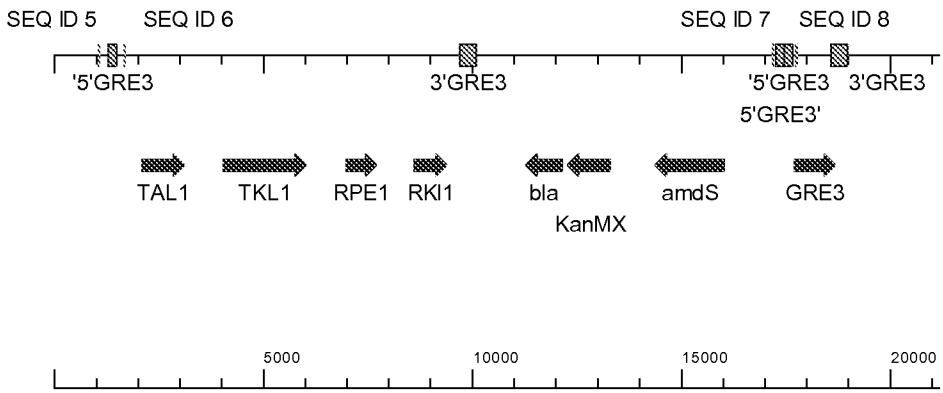

NotI 421
SfiI 631
Acc65I 842
BamHI 15249
5' GRE3
'lacZa
TPI1p
amdS
TAL1
16000
14000
2000
ADH1p
pPWT080
KanMX
12000
4000
16176 bps
TKL1
bla
10000
6000
8000
PGI1p
ori322
lacZa'
RPE1
3' GRE3
RKI1
ENO1p
NotI 9446

Figure 2

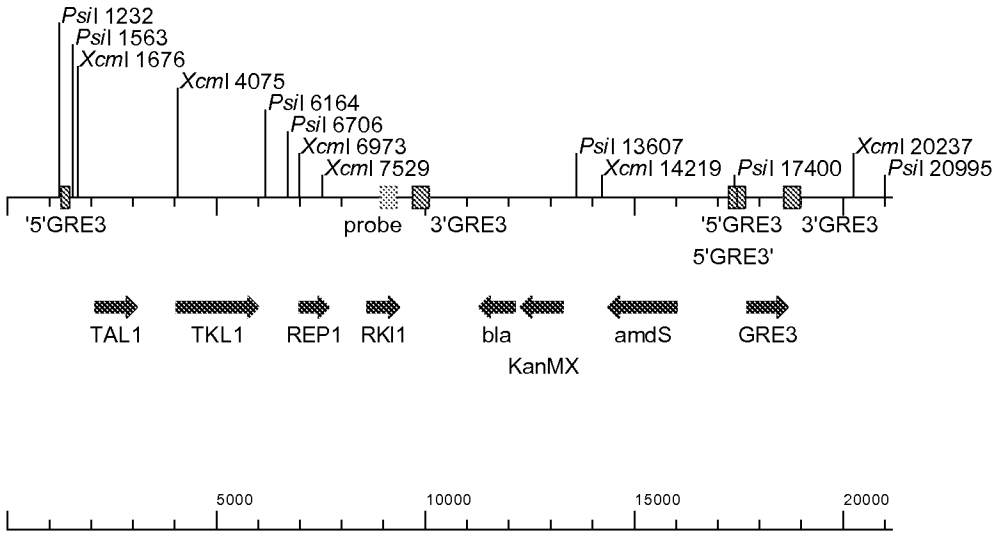


Panel a



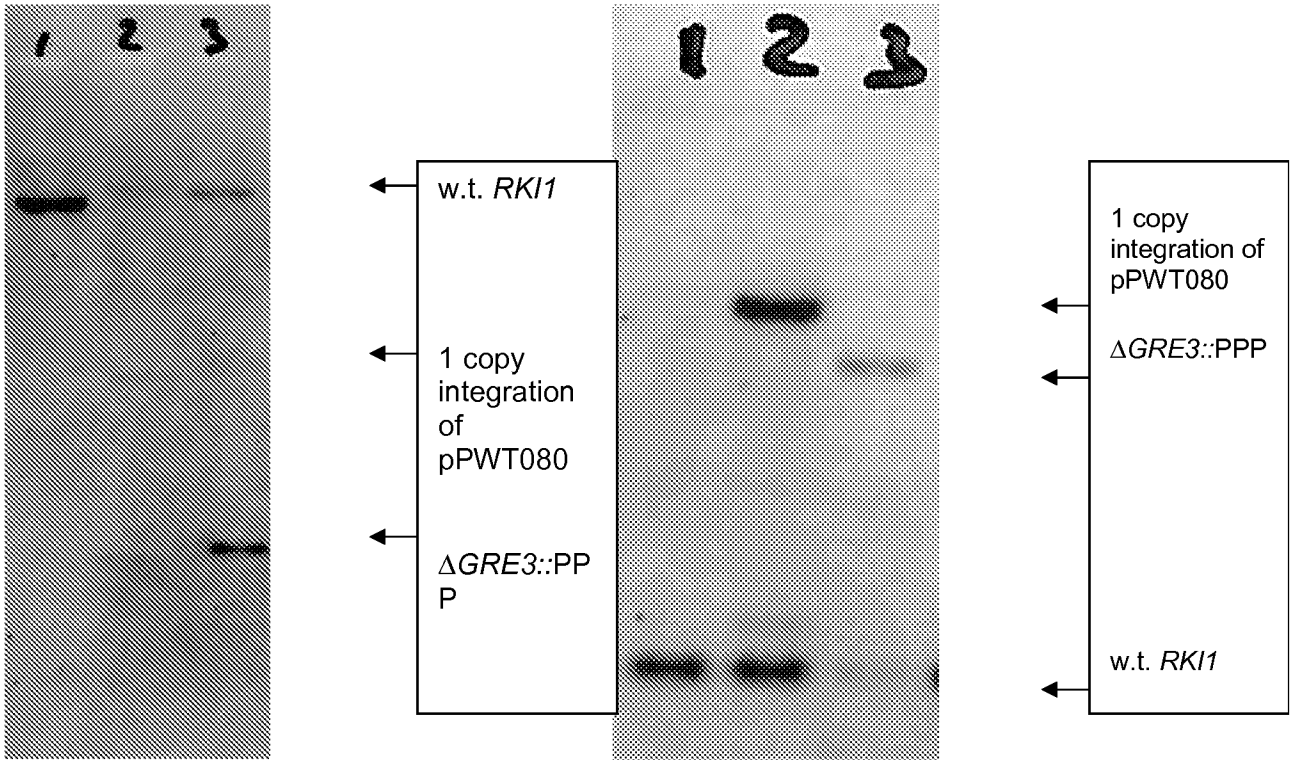
Panel b

Figure 3



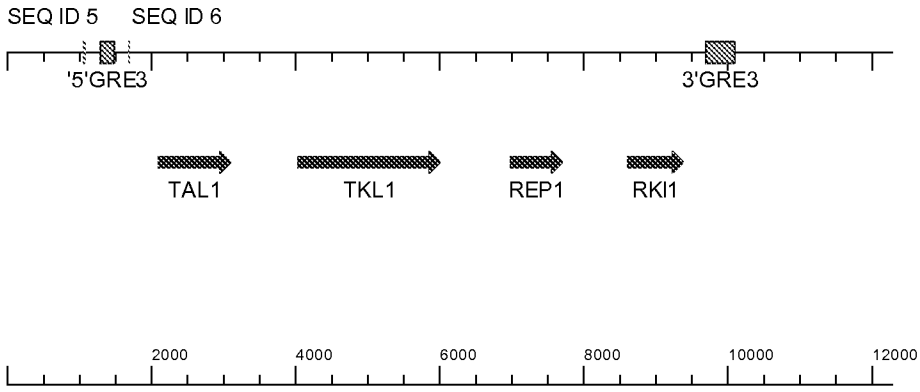
Panel c

Figure 3

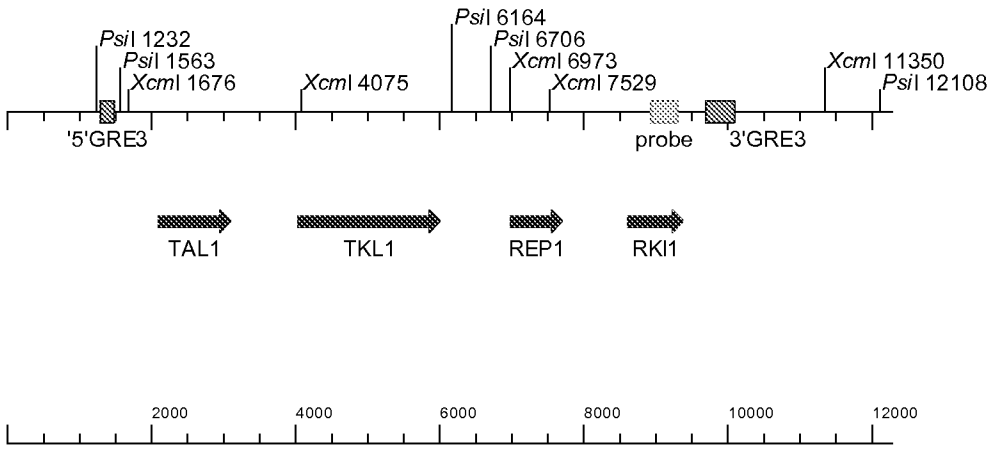


Panel a Panel b

Figure 4



Panel a



Panel b

Figure 5

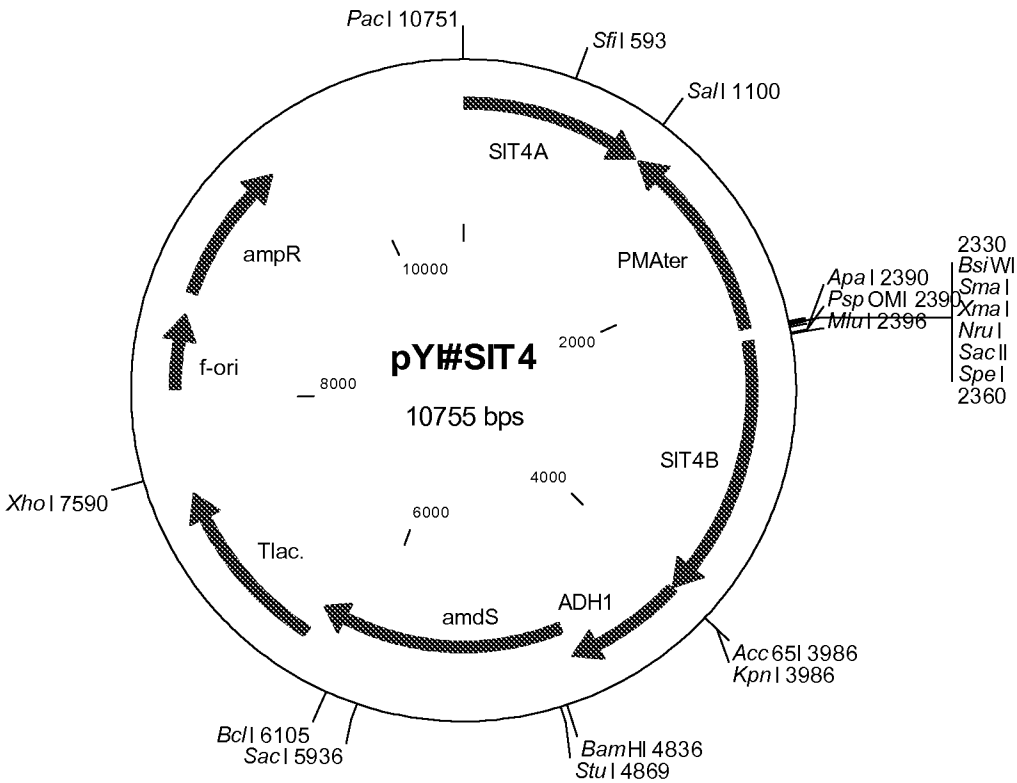

Pac I 10751
Sfi I 593
Sal I 1100
SIT4A
ampR
10000
PMAter
2330
Bsi WI
Sma I
Xma I
Nru I
Sac II
Spe I
2360
Apa I 2390
Psp OMI 2390
Mlu I 2396
2000
pYI#SIT4
f-ori
8000
10755 bps
SIT4B
4000
Xho I 7590
6000
Tlac.
ADH1
amdS
Acc 65I 3986
Kpn I 3986
Bcl I 6105
Sac I 5936
BamHI 4836
Stu I 4869

Figure 6

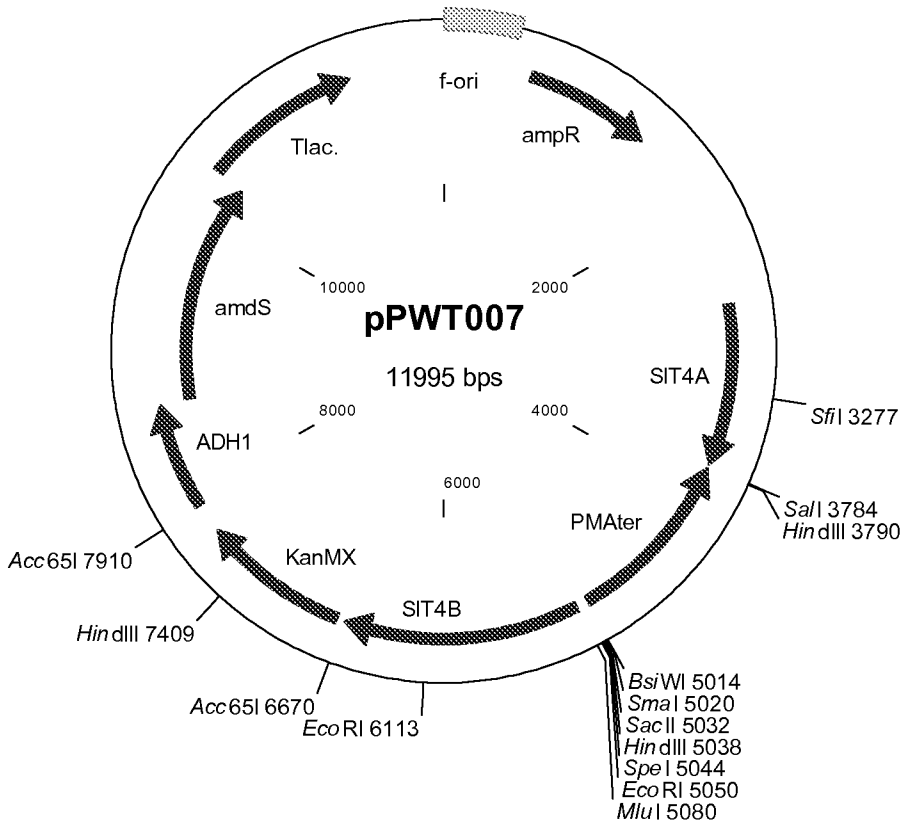

f-ori
ampR
Tlac.
10000
2000
amdS
pPWT007
11995 bps
SIT4A
8000
4000
ADH1
6000
PMAter
KanMX
SIT4B
Sfi I 3277
Sal I 3784
Hin dIII 3790
Acc65I 7910
Hin dIII 7409
Acc65I 6670
Eco RI 6113
Bsi WI 5014
Sma I 5020
Sac II 5032
Hin dIII 5038
Spe I 5044
Eco RI 5050
Mlu I 5080

Figure 7

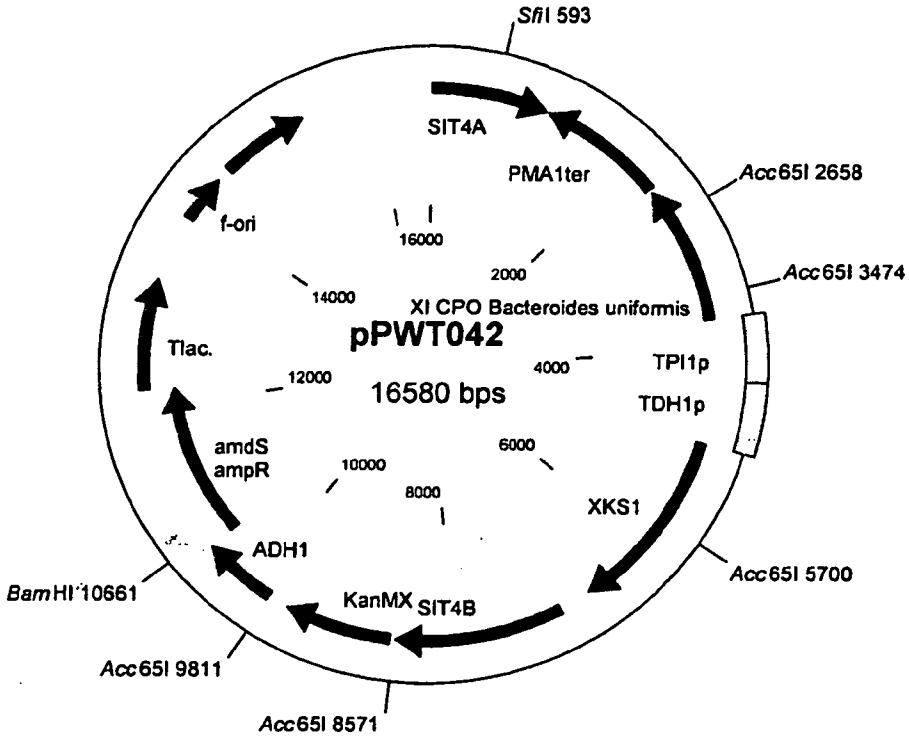
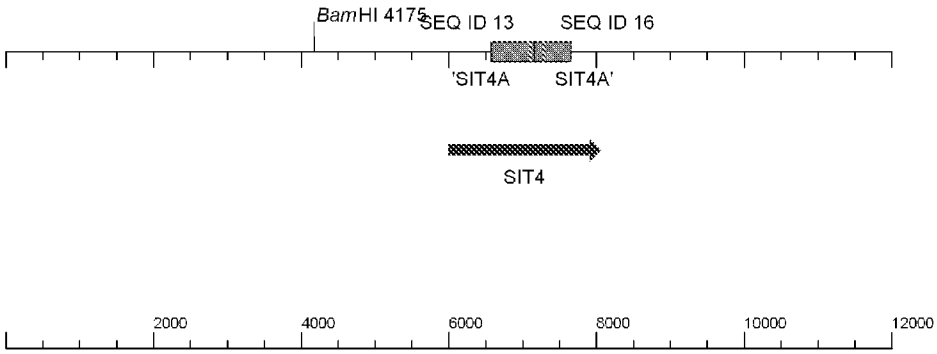
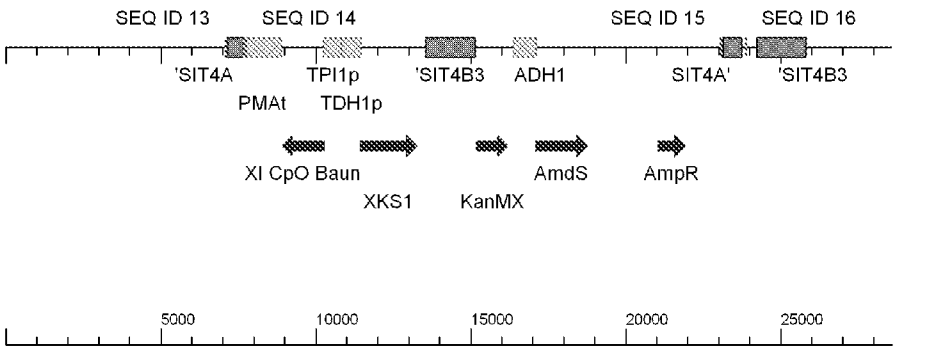

SfiI 593
SIT4A
PMA1ter
Acc65I 2658
Acc65I 3474
TPI1p
TDH1p
XKS1
Acc65I 5700
SIT4B
KanMX
Acc65I 8571
Acc65I 9811
BamHI 10661
ADH1
amdS
ampR
Tlac.
f-ori
16000
2000
4000
6000
8000
10000
12000
14000
XI CPO Bacteroides uniformis
pPWT042
16580 bps

Figure 8



Panel a



Panel b

Figure 9

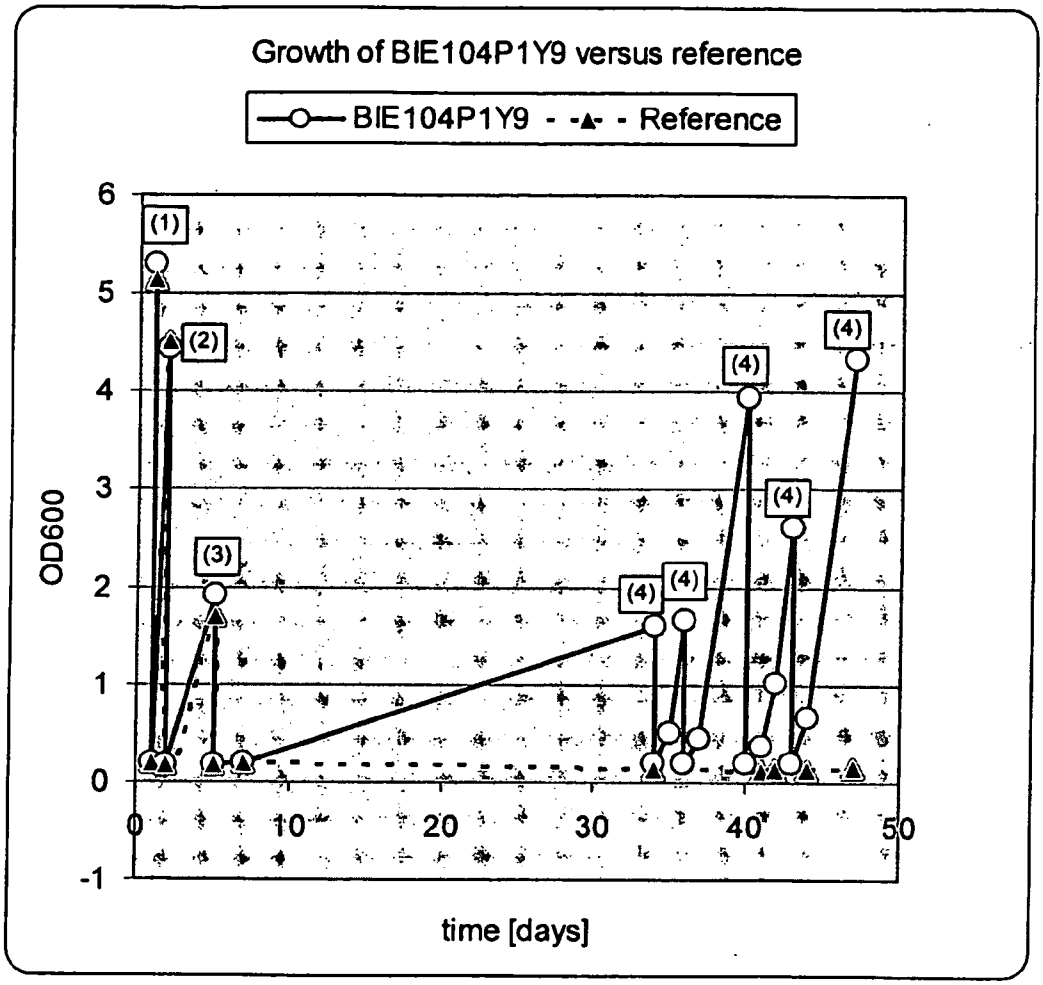

Growth of BIE104P1Y9 versus reference
BIE104P1Y9
Reference
OD600
time [days]
(1)
(2)
(3)
(4)
(4)
(4)
(4)
(4)
6
5
4
3
2
1
0
-1
0
10
20
30
40
50

Figure 10

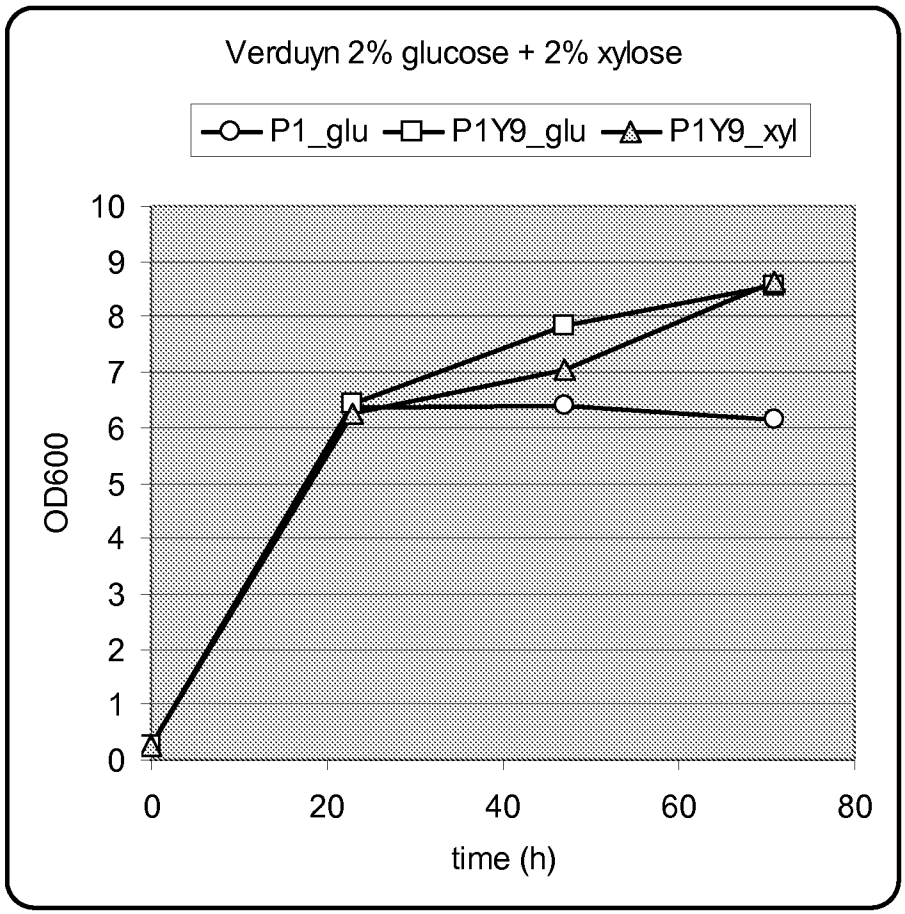
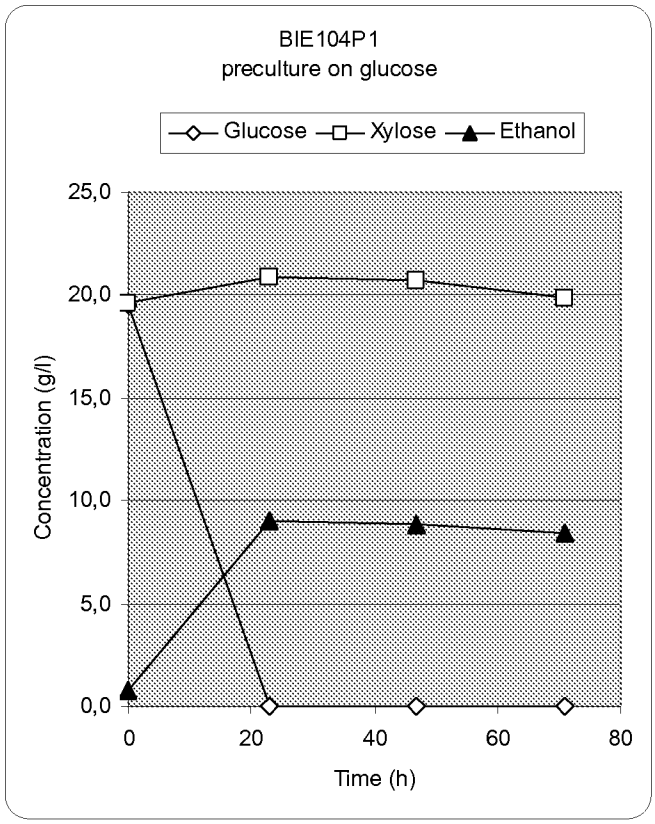
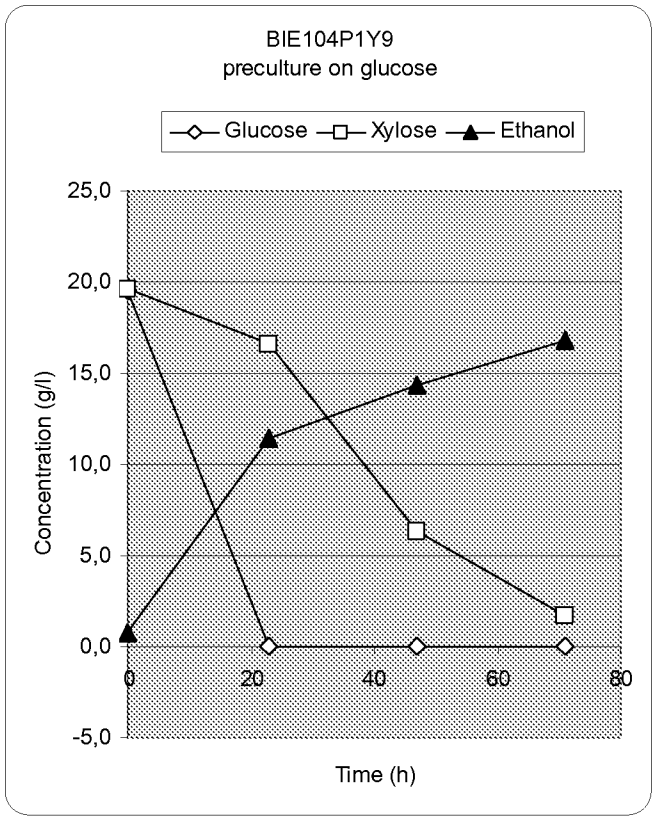

Verduyn 2% glucose + 2% xylose
P1_glu
P1Y9_glu
P1Y9_xyl
OD600
0
1
2
3
4
5
6
7
8
9
10
0
20
40
60
80
time (h)

Figure 11



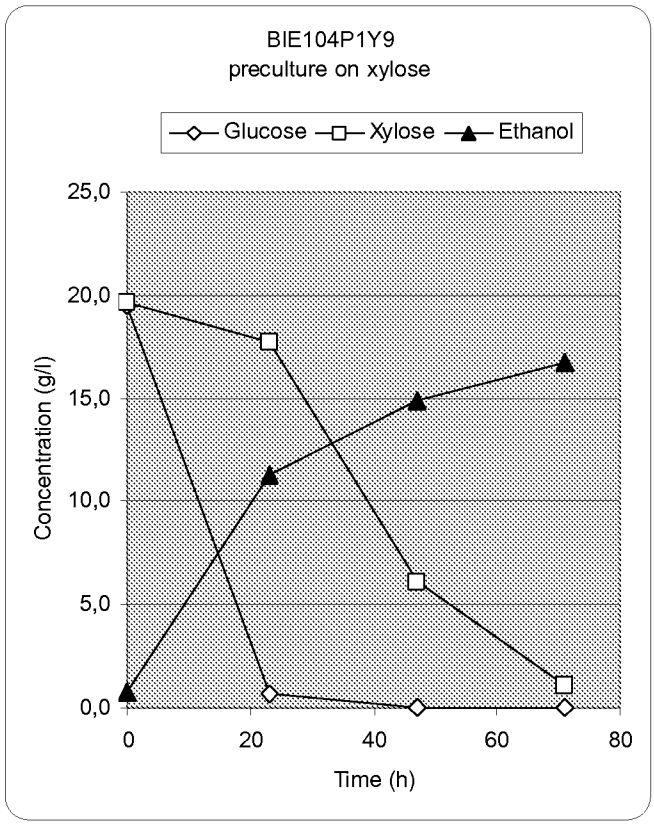
Panel a

Figure 12



Panel b

Figure 12



Panel c

Figure 12

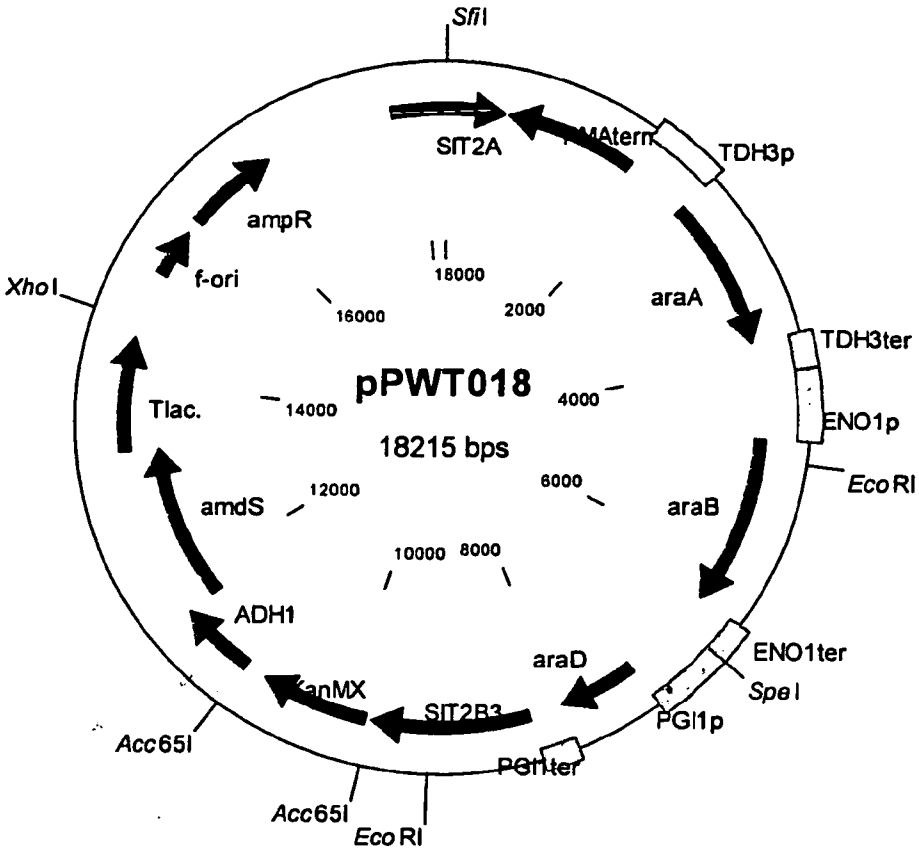

SfiI
SIT2A
TDH3p
ampR
f-ori
XhoI
araA
TDH3ter
pPWT018
18215 bps
ENO1p
EcoRI
Tlac.
amdS
araB
ADH1
araD
ENO1ter
SpeI
PGI1p
Acc65I
EcoRI

Figure 13

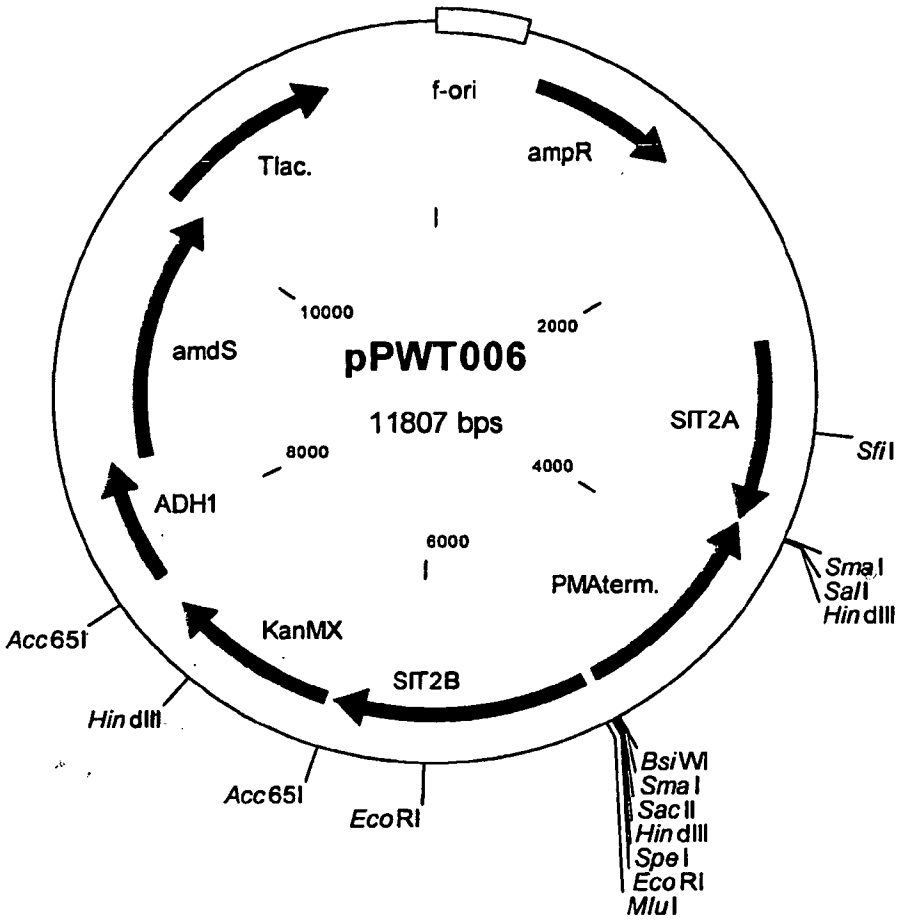

f-ori
ampR
Tlac.
10000
2000
amdS
pPWT006
11807 bps
SIT2A
8000
4000
ADH1
6000
PMAterm.
KanMX
SIT2B
Acc65I
HindIII
Acc65I
EcoRI
BsiWI
SmaI
SacII
HindIII
SpeI
EcoRI
MluI
SfiI
SmaI
SalI
HindIII

Figure 14

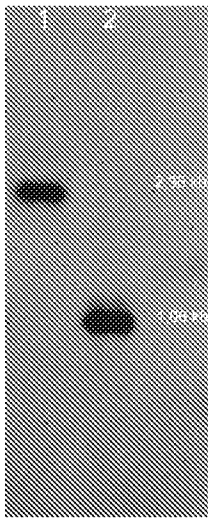
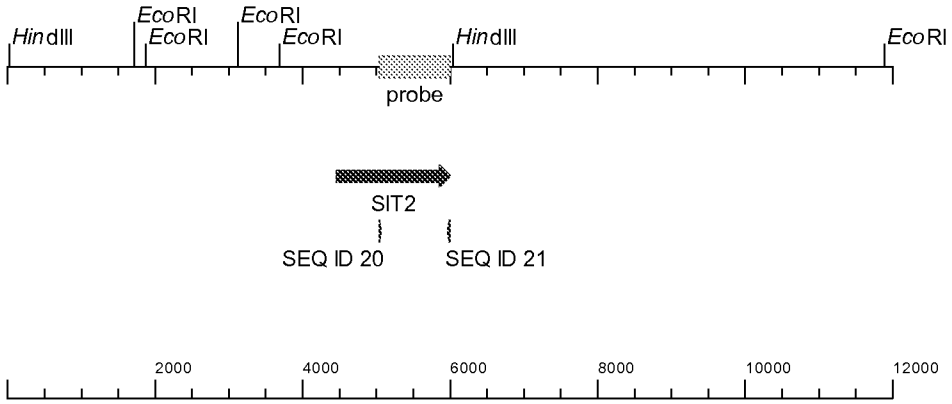
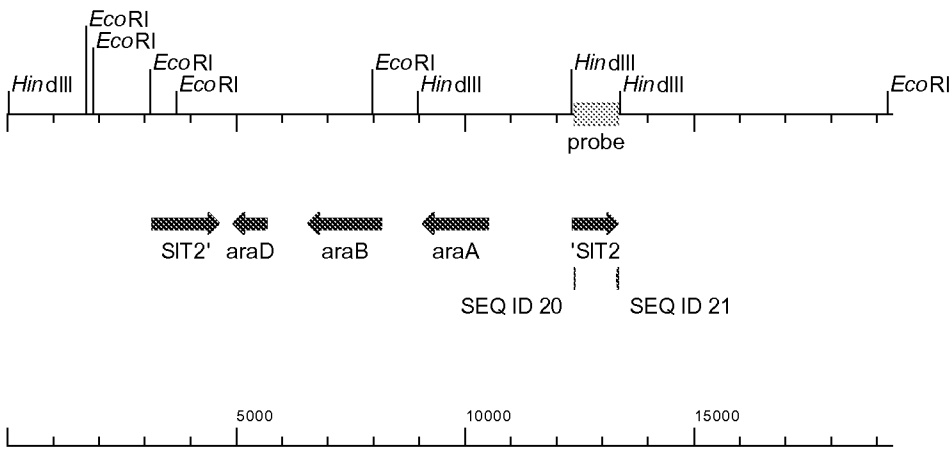


Figure 15

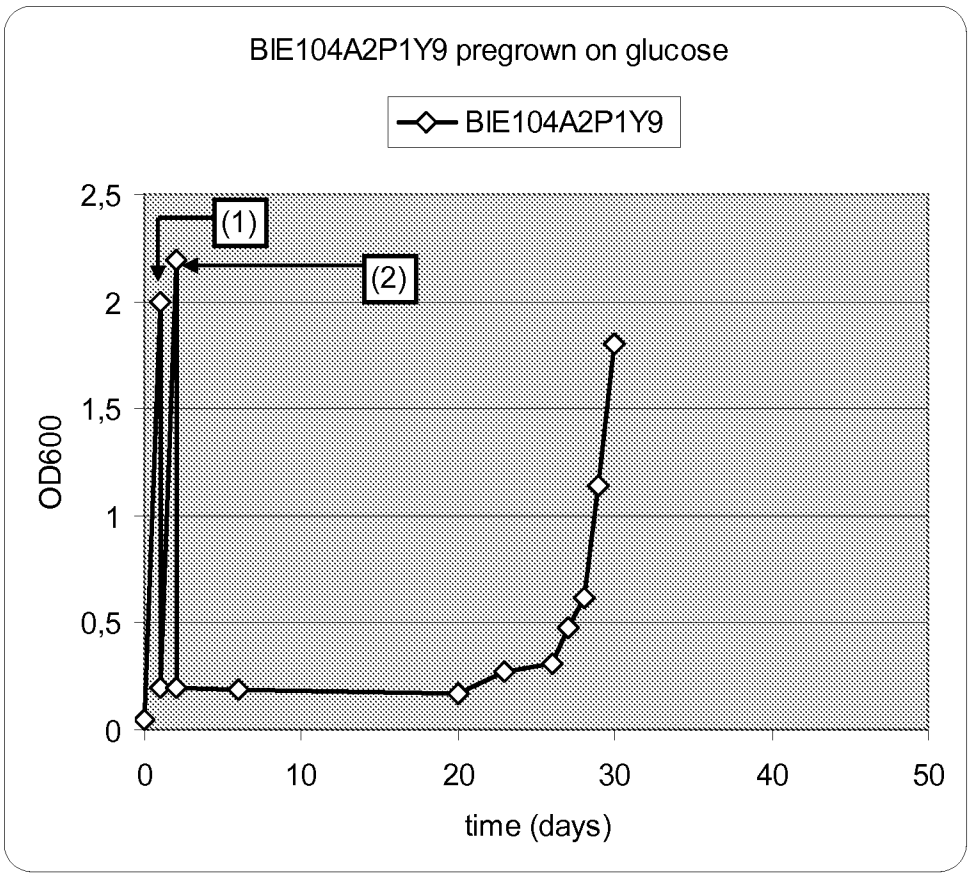


Panel a



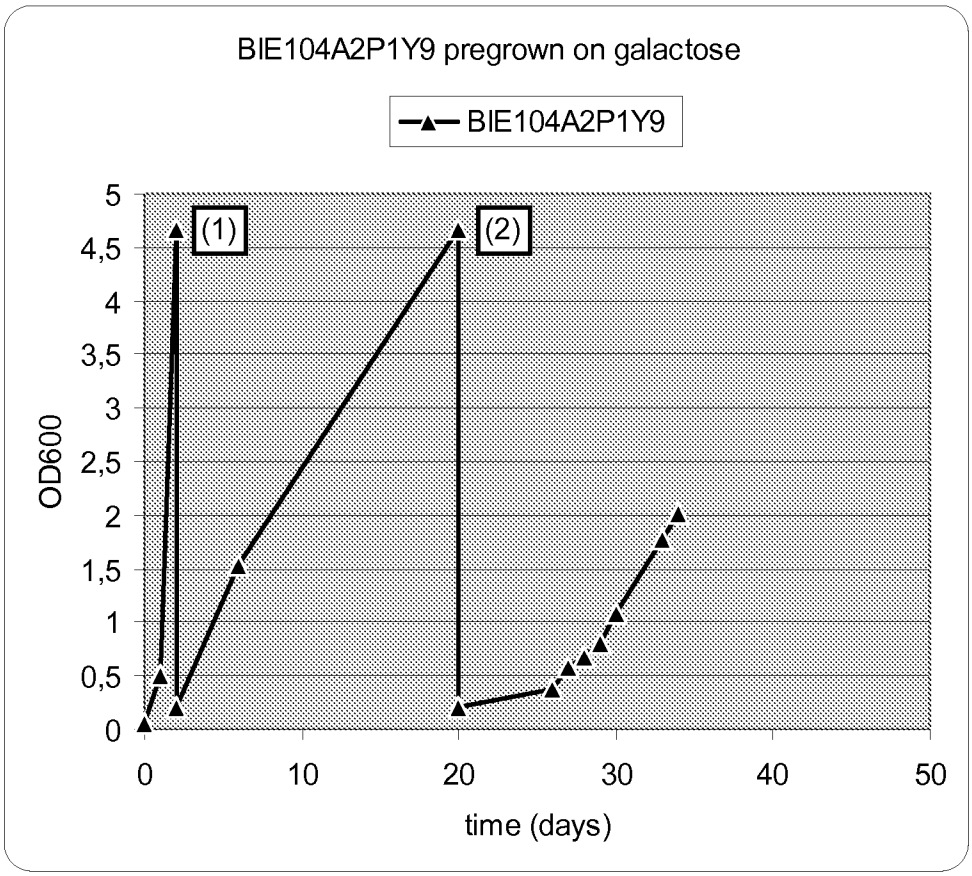
Panel b

Figure 16



Panel a

Figure 17



Panel b

Figure 17

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2009/052623

A. CLASSIFICATION OF SUBJECT MATTER
INV. C12N9/92 C12P7/10 C12N15/52

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
C12N C12P

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

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
EPO-Internal, Sequence Search, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X Y	WO 2006/009434 A1 (UNIV DELFT TECH [NL]; WINKLER AARON ADRIAAN [NL]; KUYPER SIPKO MAARTEN) 26 January 2006 (2006-01-26) pages 11-14; sequence 2 ----	1-18, 23-26, 29-34 19-22, 27-28
Y	DATABASE UniProt [Online] 23 October 2007 (2007-10-23), "RecName: Full=Xylose isomerase; EC=<A HREF="http://srs.ebi.ac.uk/srsbin/cgi-bin/wgetz?[enzyme-ECNumber:5.3.1.5]+-e">5.3.1.5</A>;" XP002530254 retrieved from EBI accession no. UNIPROT:A7UZG5 Database accession no. A7UZG5 the whole document ----	1-34

-/--

[X] Further documents are listed in the continuation of Box C. [X] See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier document but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
5 June 2009	08/07/2009
Name and mailing address of the ISA/ European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Authorized officer Lejeune, Robert

Form PCT/ISA/210 (second sheet) (April 2005)

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2009/052623

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	ANTONIUS J A VAN MARIS ET AL: "Alcoholic fermentation of carbon sources in biomass hydrolysates by Saccharomyces cerevisiae: current status" ANTONIE VAN LEEUWENHOEK, KLUWER ACADEMIC PUBLISHERS, DO, vol. 90, no. 4, 11 October 2006 (2006-10-11), pages 391-418, XP019446684 ISSN: 1572-9699 the whole document	1-34
A	VAN MARIS A J A ET AL: "Development of Efficient Xylose Fermentation in Saccharomyces cerevisiae: Xylose Isomerase as a Key Component" ADVANCES IN BIOCHEMICAL ENGINEERING, BIOTECHNOLOGY, SPRINGER, BERLIN, DE, vol. 108, 1 January 2007 (2007-01-01), pages 179-204, XP008086128 ISSN: 0724-6145 the whole document	1-34
X,P	WO 2008/041840 A1 (DSM IP ASSETS BV [NL]; VAN MARIS ANTON J A [NL]; PRONK JACOBUS THOMAS) 10 April 2008 (2008-04-10) pages 18,19; sequence 15	1-6,26, 34
A	WO 2008/000632 A1 (DSM IP ASSETS BV [NL]; ROUBOS JOHANNES ANDRIES [NL]; PEIJ VAN NOEL NIC) 3 January 2008 (2008-01-03) cited in the application the whole document	1-34

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/EP2009/052623

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2006009434 A1	26-01-2006	BR PI0513437 A	06-05-2008
		CA 2573483 A1	26-01-2006
		JP 2008506383 T	06-03-2008
		US 2008261287 A1	23-10-2008
WO 2008041840 A1	10-04-2008	AU 2007302867 A1	10-04-2008
WO 2008000632 A1	03-01-2008	AU 2007263880 A1	03-01-2008
		CA 2657975 A1	03-01-2008
		EP 2035561 A1	18-03-2009