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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
1 July 2010 (01.07.2010)

PCT

(10) International Publication Number
WO 2010/074577 A1

(51) International Patent Classification:
C12N 1/14 (2006.01) *C12P 7/16* (2006.01)
C12N 1/16 (2006.01) *C12P 7/42* (2006.01)
C12N 1/18 (2006.01) *C12P 7/52* (2006.01)
C12N 1/19 (2006.01) *C12P 7/54* (2006.01)
C12P 7/08 (2006.01) *C12P 7/56* (2006.01)
C12P 7/10 (2006.01) *C12P 7/48* (2006.01)

(21) International Application Number:
PCT/NL2009/050803

(22) International Filing Date:
24 December 2009 (24.12.2009)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
08172919.6 24 December 2008 (24.12.2008) EP
61/140,795 24 December 2008 (24.12.2008) US

(71) Applicant *(for all designated States except US)*: **Royal Nedalco B.V.** [NL/NL]; Van Konijnenburgweg 100, NL-4600 AA Bergen op Zoom (NL).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **Teunissen, , Aloysius Wilhelmus Rudolphus Hubertus** [NL/NL]; Tjaskerlaan 28, NL-3052 HP Rotterdam (NL). **De Bont, Johannes Adrianus Maria** [NL/NL]; Lawickse Allee 68, NL-6707 AK Wageningen (NL).

(74) Agent: **Ketelaars, Maarten**; J.W. Frisolaan 13, NL-2517 JS The Hague (NL).

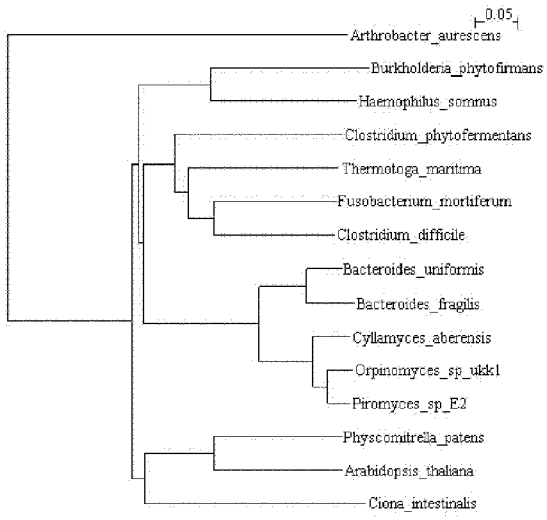
(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG,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

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, 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, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

[Continued on next page]

(54) Title: XYLOSE ISOMERASE GENES AND THEIR USE IN FERMENTATION OF PENTOSE SUGARS

Fig.1



(57) Abstract: The present invention relates to eukaryotic cells which have the ability to isomerise xylose directly into xylulose. The cells have acquired this ability by transformation with nucleotide sequences coding for a xylose isomerase that has one or more specific sequence elements typical for isomerases having the ability of functional expression in yeasts, such as e.g. xylose isomerases obtainable froma bacterium of the genera Clostridiumand Fusobacteriumor a tunicate form the genus *Ciona*.The cell preferablyis ayeast or a filamentous fungus, more preferably a yeast is capable of anaerobic alcoholic fermentation. The may further comprise one or more genetic modifications that increase the flux of the pentose phosphate pathway, reduce unspecific aldose reductase activity, confer to the cell the ability to increase the specific xylulose kinase activity, convert L-arabinose into D-xylulose 5-phosphate, increase transport of at least one of xylose and arabinose into the host cell, decrease sensitivity to catabolite repression, increase tolerance to ethanol, osmolarity or organic acids; and/or reduce production of by-products. The cell preferably is a cell that has the ability to produce a fermentation product such as ethanol, lactic acid, 3-hydroxy-propionic acid,acrylic acid, acetic acid, succinic acid, citric acid, amino acids, 1,3-propane-diol, ethylene, glycerol, β-lactam antibiotics and cephalosporins. The invention further relates to processes for producing these fermentation products wherein a cell of the invention is used to ferment xylose into the fermentation products.

WO 2010/074577 A1

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))*

Xylose isomerase genes and their use in fermentation of pentose sugars

Field of the invention

The present invention relates to the use of nucleic acid sequences encoding xylose isomerases in the transformation of a eukaryotic microbial host cell to confer to the host cell the ability of isomerising xylose to xylulose. The transformed host cell is used in a process for the production of ethanol and other fermentation products by fermentation of a pentose-containing medium.

Background of the invention

Economically viable ethanol production from the hemicellulose fraction of plant biomass requires the simultaneous conversion of both pentoses and hexoses at comparable rates and with high yields. Yeasts, in particular *Saccharomyces spp.*, are the most appropriate candidates for this process since they can grow fast on hexoses, both aerobically and anaerobically. Furthermore they are much more resistant to the toxic environment of lignocellulose hydrolysates than (genetically modified) bacteria.

Although wild-type *S. cerevisiae* strains can slowly metabolise the pentose sugar xylulose, they are not capable of metabolising xylose. Already in the 1980's it was suggested that metabolic engineering of yeasts for xylose utilization should be based on the introduction of xylose isomerase (XI, EC 5.3.1.5) rather than expressing heterologous xylose reductase and xylitol dehydrogenase to convert xylose into xylulose. Unfortunately, all attempts of introducing a bacterial xylose isomerase in *S. cerevisiae* have failed to produce a functionally expressed xylose isomerase with the notable exception of the *T. thermophilus* isomerase. This enzyme was functionally expressed in *S. cerevisiae* but only very low activities were observed at growth permitting temperatures. This situation drastically changed when a newly discovered xylose isomerase from the anaerobic fungus *Piromyces Sp.*E2 was introduced in *S. cerevisiae* and high levels of enzyme activities were observed enabling this strain to grow anaerobically and produce ethanol from xylose (WO 03/062340 and WO 06/009434). Such yeast strains for the first time provided specific rates of xylose consumption and ethanol formation that are compatible with ethanol production at a commercial scale.

Since the discovery of the functional expression of the *Piromyces* xylose isomerase in yeast several reports have appeared of functional expression in yeasts of other xylose isomerases, all of which share more than 70% amino acid sequence identity with the *Piromyces* enzyme, such e.g. the bacterial xylose isomerase from *Bacteroides* (WO 04/099381; WO 06/009434; WO 09/109633), and the fungal xylose isomerases from *Cyllamyces* (WO 04/099381) and *Orpinomyces* (Madhavan et al., 2008, DOI 10.1007/s00253-008-1794-6).

However, prior to December 24, 2008 no reports have issued of functional expression in yeasts of xylose isomerases having less than 70% amino acid sequence identity with the *Piromyces* enzyme. More recently, in February 2009, Brat et al. (2009, Appl. Environ. Microbiol. 75: 2304–2311) published functional expression in the yeast *S. cerevisiae* of a xylose isomerise from the anaerobic bacterium *Clostridium phytofermentans*, the amino acid sequence of which shares only 52% identity with that of the *Piromyces* enzyme.

To date some 450 xylose isomerase amino acid sequences are publicly available in Genbank and other sequence databases, including the xylose isomerase sequences of *Piromyces, Cyllamyces aberensis, Physcomitrella patens, Arabidopsis thaliana, Haemophilus somnus, Ciona intestinalis, Clostridium difficile, Thermatoga maritime, Bacteroides fragilis, Burkholderia phytofirmans, Arthrobacter aurescens* and *Fusobacterium_mortiferum*.

There is, however, still a need in the art for nucleotide sequences encoding other xylose isomerases that may be used to transform host cells like *S. cerevisiae* to confer to them the ability of isomerising xylose to xylulose, so as to enable the use of thus transformed host cell in processes for the production of ethanol or other fermentation products by fermentation of pentose-containing feedstock.

Description of the invention

Definitions

The enzyme "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 vice versa. The enzyme is also known as a D-xylose ketoisomerase. Some xylose isomerases are also capable of catalysing the conversion between D-glucose and D-fructose and are therefore sometimes referred to as glucose isomerase. Xylose isomerases require

magnesium as cofactor. Xylose isomerases of the invention may be further defined by their amino acid sequence as herein described below. Likewise xylose isomerases may be defined by the nucleotide sequences encoding the enzyme as well as by nucleotide sequences hybridising to a reference nucleotide sequence encoding a xylose isomerase as herein described below. A unit (U) of xylose isomerase activity is herein defined as the amount of enzyme producing 1 nmol of xylulose per minute, in a reaction mixture containing 50 mM phosphate buffer (pH 7.0), 10 mM xylose and 10 mM $MgCl_2$, at 37°C. Xylulose formed was determined by the method of Dische and Borenfreund (1951, J. Biol. Chem. 192: 583-587) or by HPLC as described in the Examples.

Sequence identity 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. 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. "Similarity" between two amino acid sequences is determined by comparing the amino acid sequence and its conserved amino acid substitutes of one polypeptide to the sequence of a second polypeptide. "Identity" and "similarity" can be readily calculated by known methods. The terms “sequence identity” or “sequence similarity” means that two (poly)peptide or two nucleotide sequences, when optimally aligned, preferably over the entire length (of at least the shortest sequence in the comparison) and maximizing the number of matches and minimizes the number of gaps such as by the programs ClustalW (1.83), GAP or BESTFIT using default parameters, share at least a certain percentage of sequence identity as defined elsewhere herein. GAP uses the Needleman and Wunsch global alignment algorithm to align two sequences over their entire length, maximizing the number of matches and minimizes the number of gaps. Generally, the GAP default parameters are used, with a gap creation penalty = 50 (nucleotides) / 8 (proteins) and gap extension penalty = 3 (nucleotides) / 2 (proteins). For nucleotides the default scoring matrix used is nwsgapdna and for proteins the default scoring matrix is Blosum62 (Henikoff & Henikoff, 1992, PNAS 89, 915-919). A preferred multiple alignment program for aligning protein sequences of the invention is ClustalW (1.83) using a blosum matrix and default settings (Gap opening penalty:10; Gap extension penalty: 0.05). It is clear than when RNA sequences are said to be essentially similar or have a certain degree of sequence identity with DNA sequences,

thymine (T) in the DNA sequence is considered equal to uracil (U) in the RNA sequence. Sequence alignments and scores for percentage sequence identity may be determined using computer programs, such as the GCG Wisconsin Package, Version 10.3, available from Accelrys Inc., 9685 Scranton Road, San Diego, CA 92121-3752 USA or the open-source software Emboss for Windows (current version 2.7.1-07).Alternatively percent similarity or identity may be determined by searching against databases such as FASTA, BLAST, etc.

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 GCG program package (Devereux, J., et al., Nucleic Acids Research 12 (1):387 (1984)), BestFit, BLASTP, BLASTN, and FASTA (Altschul, S. F. et al., J. Mol. Biol. 215:403-410 (1990). The BLAST X program is publicly available from NCBI and other sources (BLAST Manual, Altschul, S., et al., NCBI NLM NIH Bethesda, MD 20894; Altschul, S., et al., J. Mol. Biol. 215:403-410 (1990). The well-known Smith Waterman algorithm may also be used to determine identity.

Preferred parameters for polypeptide sequence comparison include the following: Algorithm: Needleman and Wunsch, J. Mol. Biol. 48:443-453 (1970); Comparison matrix: BLOSSUM62 from Hentikoff and Hentikoff, Proc. Natl. Acad. Sci. USA. 89:10915-10919 (1992); Gap Penalty: 12; and Gap Length Penalty: 4. A program useful with these parameters is publicly available as the "Ogap" program from Genetics Computer Group, located in Madison, WI. The aforementioned parameters are the default parameters for amino acid comparisons (along with no penalty for end gaps).

Preferred parameters for nucleic acid comparison include the following: Algorithm: Needleman and Wunsch, J. Mol. Biol. 48:443-453 (1970); Comparison matrix: matches=+10, mismatch=0; Gap Penalty: 50; Gap Length Penalty: 3. Available as the Gap program from Genetics Computer Group, located in Madison, Wis. Given above are the default parameters for nucleic acid comparisons.

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; Ile 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.

Nucleotide sequences encoding xylose isomerases of the invention may also be defined by their capability to hybridise with the nucleotide sequences of SEQ ID NO. 2, respectively, under moderate, or preferably under stringent hybridisation conditions. 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 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%.

A "nucleic acid construct" or "nucleic acid vector" is herein understood to mean a man-made nucleic acid molecule resulting from the use of recombinant DNA technology. The term "nucleic acid construct" therefore does not include naturally occurring nucleic acid molecules although a nucleic acid construct may comprise (parts of) naturally occurring nucleic acid molecules. The terms "expression vector" or expression construct" refer to nucleotide sequences that are capable of affecting expression of a gene in host cells or host organisms compatible with such sequences. These expression vectors typically include at least suitable transcription regulatory sequences and optionally, 3' transcription termination signals. Additional factors necessary or helpful in effecting expression may also be present, such as expression enhancer elements. The expression vector will be introduced into a suitable host cell and be able to effect expression of the coding sequence in an in vitro cell culture of the host cell. The expression vector will be suitable for replication in the host cell or organism of the invention.

As used herein, the term "promoter" or "transcription regulatory sequence" refers to a nucleic acid fragment that functions to control the transcription of one or more coding sequences, and is located upstream with respect to the direction of transcription of the transcription initiation site of the coding sequence, and is structurally identified by the presence of a binding site for DNA-dependent RNA polymerase, transcription initiation sites and any other DNA sequences, including, but not limited to transcription factor binding sites, repressor and activator protein binding sites, and any other sequences of nucleotides known to one of skill in the art to act directly or indirectly to regulate the amount of transcription from the promoter. A "constitutive" promoter is a promoter that is active in most tissues under most physiological and developmental

conditions. An "inducible" promoter is a promoter that is physiologically or developmentally regulated, e.g. by the application of a chemical inducer.

The term "selectable marker" is a term familiar to one of ordinary skill in the art and is used herein to describe any genetic entity which, when expressed, can be used to select for a cell or cells containing the selectable marker. The term "reporter" may be used interchangeably with marker, although it is mainly used to refer to visible markers, such as green fluorescent protein (GFP). Selectable markers may be dominant or recessive or bidirectional.

As used herein, the term "operably linked" refers to a linkage of polynucleotide elements in a functional relationship. A nucleic acid is "operably linked" when it is placed into a functional relationship with another nucleic acid sequence. For instance, a transcription regulatory sequence is operably linked to a coding sequence if it affects the transcription of the coding sequence. Operably linked means that the DNA sequences being linked are typically contiguous and, where necessary to join two protein encoding regions, contiguous and in reading frame.

The terms "protein" or "polypeptide" are used interchangeably and refer to molecules consisting of a chain of amino acids, without reference to a specific mode of action, size, 3-dimensional structure or origin.

"Fungi" (singular fungus) are herein understood as heterotrophic eukaryotic microorganism that digest their food externally, absorbing nutrient molecules into their cells. Fungi are a separate kingdom of eukaryotic organisms and include yeasts, molds, and mushrooms. The terms fungi, fungus and fungal as used herein thus expressly includes yeasts as well as filamentous fungi.

The term "gene" means a DNA fragment comprising a region (transcribed region), which is transcribed into an RNA molecule (e.g. an mRNA) in a cell, operably linked to suitable regulatory regions (e.g. a promoter). A gene will usually comprise several operably linked fragments, such as a promoter, a 5' leader sequence, a coding region and a 3'nontranslated sequence (3'end) comprising a polyadenylation site. "Expression of a gene" refers to the process wherein a DNA region which is operably linked to appropriate regulatory regions, particularly a promoter, is transcribed into an RNA, which is biologically active, i.e. which is capable of being translated into a biologically active protein or peptide.

The term "homologous" when used to indicate the relation between a given (recombinant) nucleic acid or polypeptide molecule and a given host organism or host cell, is understood to mean that in nature the nucleic acid or polypeptide molecule is produced by a host cell or organisms of the same species, preferably of the same variety or strain. If homologous to a host cell, a nucleic acid sequence encoding a polypeptide will typically (but not necessarily) be operably linked to another (heterologous) promoter sequence and, if applicable, another (heterologous) secretory signal sequence and/or terminator sequence than in its natural environment. It is understood that the regulatory sequences, signal sequences, terminator sequences, etc. may also be homologous to the host cell. In this context, the use of only "homologous" sequence elements allows the construction of "self-cloned" genetically modified organisms (GMO's) (self-cloning is defined herein as in European Directive 98/81/EC Annex II). When used to indicate the relatedness of two nucleic acid sequences the term "homologous" means that one single-stranded nucleic acid sequence may hybridize to a complementary single-stranded nucleic acid sequence. The degree of hybridization may depend on a number of factors including the amount of identity between the sequences and the hybridization conditions such as temperature and salt concentration as discussed later.

The term "heterologous" when used with respect to a nucleic acid (DNA or RNA) or protein refers to a nucleic acid or protein that does not occur naturally as part of the organism, cell, genome or DNA or RNA sequence in which it is present, or that is found in a cell or location or locations in the genome or DNA or RNA sequence that differ from that in which it is found in nature. Heterologous nucleic acids or proteins are not endogenous to the cell into which it is introduced, but has been obtained from another cell or synthetically or recombinantly produced. Generally, though not necessarily, such nucleic acids encode proteins that are not normally produced by the cell in which the DNA is transcribed or expressed. Similarly exogenous RNA encodes for proteins not normally expressed in the cell in which the exogenous RNA is present. Heterologous nucleic acids and proteins may also be referred to as foreign nucleic acids or proteins. Any nucleic acid or protein that one of skill in the art would recognize as heterologous or foreign to the cell in which it is expressed is herein encompassed by the term heterologous nucleic acid or protein. The term heterologous also applies to non-

natural combinations of nucleic acid or amino acid sequences, i.e. combinations where at least two of the combined sequences are foreign with respect to each other.

Detailed description of the invention

To date some 450 xylose isomerase amino acid sequences are publicly available in Genbank and other sequence databases. Among them are a number of amino acid sequences of xylose isomerases that are known for the ability of functional expression in yeasts, including e.g. xylose isomerases from anaerobic fungi such *Piromyces, Cyllamyces*, and *Orpinomyces*, as well as bacterial xylose isomerases from *Bacteroides*, all of which share more than 70% amino acid sequence identity with the *Piromyces* enzyme. The present inventors have now surprisingly found amino acid sequences of xylose isomerases that are not related to the *Piromyces* enzyme in the sense that they share less than 70% amino acid sequence identity with the amino acid sequence of *Piromyces* enzyme, and that nonetheless have the ability of functional (i.e. active) expression in yeasts. Furthermore, the present inventors have identified a number of amino acid sequence elements that are shared among all xylose isomerases with the ability of functional expression in yeasts. Functional expression of a xylose isomerase in a yeast is herein understood as expression of a codon-optimised coding sequence for a xylose isomerase from a glycolytic promoter on a 2μ-based plasmid in *S. cerevisiae* that allows the detectable growth of the yeast on xylose as sole carbon source, preferably under anaerobic conditions with production of ethanol at the expense of xylose, more preferably with at least one of a growth rate, biomass and ethanol yield that is at least 10, 20, 50 or 80% of that achieved with a codon-optimised *Piromyces* xylose isomerase coding sequence under otherwise identical conditions. Preferably functional expression is expression that allows the detectable growth of the yeast on xylose as sole carbon source at a temperature which is lower than 35, 33, 30 or 28°C and at a temperature which is higher than 20, 22, or 25°C.

In a first aspect the present invention relates to a transformed host cell that has the ability of isomerising xylose to xylulose. The ability of isomerising xylose to xylulose is conferred to the host cell by transformation of the host cell with a nucleic acid construct comprising a nucleotide sequence encoding a xylose isomerase. The transformed host cell's ability to isomerise xylose into xylulose is understood to mean

the direct isomerisation of xylose, in a single reaction catalysed by a xylose isomerase, to xylulose, as opposed to the two step conversion of xylose into xylulose via a xylitol intermediate as catalysed by xylose reductase and xylitol dehydrogenase, respectively.

In one embodiment the nucleotide sequence encoding the xylose isomerase is selected from the group consisting of:

(a) a nucleotide sequence encoding a polypeptide with xylose isomerase activity, which polypeptide comprises an amino acid sequence that has at least 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 95, 96, 97, 98, or 99% sequence identity with the amino acid sequence of SEQ ID NO. 1 (*C. difficile*);

(b) a nucleotide sequence encoding a polypeptide with xylose isomerase activity, which polypeptide comprises an amino acid sequence that has at least 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 95, 96, 97, 98, or 99% sequence identity with the amino acid sequence of SEQ ID NO. 2 (*Ciona*);

(c) a nucleotide sequence encoding a polypeptide with xylose isomerase activity, which polypeptide comprises an amino acid sequence that has at least 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 95, 96, 97, 98, or 99% sequence identity with the amino acid sequence of SEQ ID NO. 35 (*F. mortiferum*);

(d) a nucleotide sequence the complementary strand of which hybridises to a nucleotide sequence of (a), (b) or (c); and,

(e) a nucleotide sequence the sequence of which differs from the sequence of a nucleotide sequence of (d) due to the degeneracy of the genetic code.

In one embodiment the nucleotide sequence encoding the xylose isomerase encodes an amino acid sequence comprising one or amino acid sequence elements that are shared among xylose isomerases with the ability of functional expression in yeasts. In this embodiment the nucleotide sequence encoding the xylose isomerase encodes an amino acid sequence comprising one or amino acid sequence elements selected from the group consisting of:

(a) a methionine residue at position 91;

(b) the amino acid sequence TGIKLL at positions 134-139;

(c) a phenylalanine residue at position 230;

(d) the amino acids phenylalanine and lysine at positions 264 and 265, respectively;

(e) the amino acid sequence TLAGH at positions 274-278;

(f) the amino acid sequence RYASF at positions 387-391;

(g) a glycine residue at position 394; and,

(h) an alanine residue at position 431.

In this embodiment the encoded xylose isomerase may comprise at least 1, 2, 3, 4, 5, 6, 7 or all 8 of the elements (a) - (h) in all possible combinations. Thus, in this embodiment the encoded xylose isomerase may comprise the following element or combinations of elements: (a); (b); (c); (d); (e); (f); (g); (h); (a) and (b); (a) and (c); (a) and (d); (a) and (e); (a) and (f); (a) and (g); (a) and (h); (b) and (c); (b) and (d); (b) and (e); (b) and (f); (b) and (g); (b) and (h); (c) and (d); (c) and (e); (c) and (f); (c) and (g); (c) and (h); (d) and (e); (d) and (f); (d) and (g); (d) and (h); (e) and (f); (e) and (g); (e) and (h); (f) and (g); (f) and (h); (g) and (h); (a), (b) and (c); (a), (b) and (d); (a), (b) and (e); (a), (b) and (f); (a), (b) and (g); (a), (b) and (h); (a), (c) and (d); (a), (c) and (e); (a), (c) and (f); (a), (c) and (g); (a), (c) and (h); (a), (d) and (e); (a), (d) and (f); (a), (d) and (g); (a), (d) and (h); (a), (e) and (f); (a), (e) and (g); (a), (e) and (h); (a), (f) and (g); (a), (f) and (h); (a), (g) and (h); (b), (c) and (d); (b), (c) and (e); (b), (c) and (f); (b), (c) and (g); (b), (c) and (h); (b), (d) and (e); (b), (d) and (f); (b), (d) and (g); (b), (d) and (h); (b), (e) and (f); (b), (e) and (g); (b), (e) and (h); (b), (f) and (g); (b), (f) and (h); (b), (g) and (h); (c), (d) and (e); (c), (d) and (f); (c), (d) and (g); (c), (d) and (h); (c), (e) and (f); (c), (e) and (g); (c), (e) and (h); (c), (f) and (g); (c), (f) and (h); (c), (g) and (h); (d), (e) and (f); (d), (e) and (g); (d), (e) and (h); (d), (f) and (g); (d), (f) and (h); (d), (g) and (h); (e), (f) and (g); (e), (f) and (h); (e), (g) and (h); (f), (g) and (h); (a), (b), (c) and (d); (a), (b), (c) and (e); (a), (b), (c) and (f); (a), (b), (c) and (g); (a), (b), (c) and (h); (a), (b), (d) and (e); (a), (b), (d) and (f); (a), (b), (d) and (g); (a), (b), (d) and (h); (a), (b), (e) and (f); (a), (b), (e) and (g); (a), (b), (e) and (h); (a), (b), (f) and (g); (a), (b), (f) and (h); (a), (b), (g) and (h); (a), (c), (d) and (e); (a), (c), (d) and (f); (a), (c), (d) and (g); (a), (c), (d) and (h); (a), (c), (e) and (f); (a), (c), (e) and (g); (a), (c), (e) and (h); (a), (c), (f) and (g); (a), (c), (f) and (h); (a), (c), (g) and (h); (a), (d), (e) and (f); (a), (d), (e) and (g); (a), (d), (e) and (h); (a), (d), (f) and (g); (a), (d), (f) and (h); (a), (d), (g) and (h); (a), (e), (f) and (g); (a),

(e), (f) and (h); (a), (e), (g) and (h); (a), (f), (g) and (h); (b), (c), (d) and (e); (b), (c), (d) and (f); (b), (c), (d) and (g); (b), (c), (d) and (h); (b), (c), (e) and (f); (b), (c), (e) and (g); (b), (c), (e) and (h); (b), (c), (f) and (g); (b), (c), (f) and (h); (b), (c), (g) and (h); (b), (d), (e) and (f); (b), (d), (e) and (g); (b), (d), (e) and (h); (b), (d), (f) and (g); (b), (d), (f) and (h); (b), (d), (g) and (h); (b), (e), (f) and (g); (b), (e), (f) and (h); (b), (e), (g) and (h); (b), (f), (g) and (h); (c), (d), (e) and (f); (c), (d), (e) and (g); (c), (d), (e) and (h); (c), (d), (f) and (g); (c), (d), (f) and (h); (c), (d), (g) and (h); (c), (e), (f) and (g); (c), (e), (f) and (h); (c), (e), (g) and (h); (c), (f), (g) and (h); (d), (e), (f) and (g); (d), (e), (f) and (h); (d), (e), (g) and (h); (d), (f), (g) and (h); (e), (f), (g) and (h); (a), (b), (c), (d) and (e); (a), (b), (c), (d) and (f); (a), (b), (c), (d) and (g); (a), (b), (c), (d) and (h); (a), (b), (c), (e) and (f); (a), (b), (c), (e) and (g); (a), (b), (c), (e) and (h); (a), (b), (c), (f) and (g); (a), (b), (c), (f) and (h); (a), (b), (c), (g) and (h); (a), (b), (d), (e) and (f); (a), (b), (d), (e) and (g); (a), (b), (d), (e) and (h); (a), (b), (d), (f) and (g); (a), (b), (d), (f) and (h); (a), (b), (d), (g) and (h); (a), (b), (e), (f) and (g); (a), (b), (e), (f) and (h); (a), (b), (e), (g) and (h); (a), (b), (f), (g) and (h); (a), (c), (d), (e) and (f); (a), (c), (d), (e) and (g); (a), (c), (d), (e) and (h); (a), (c), (d), (f) and (g); (a), (c), (d), (f) and (h); (a), (c), (d), (g) and (h); (a), (c), (e), (f) and (g); (a), (c), (e), (f) and (h); (a), (c), (e), (g) and (h); (a), (c), (f), (g) and (h); (a), (d), (e), (f) and (g); (a), (d), (e), (f) and (h); (a), (d), (e), (g) and (h); (a), (d), (f), (g) and (h); (a), (e), (f), (g) and (h); (b), (c), (d), (e) and (f); (b), (c), (d), (e) and (g); (b), (c), (d), (e) and (h); (b), (c), (d), (f) and (g); (b), (c), (d), (f) and (h); (b), (c), (d), (g) and (h); (b), (c), (e), (f) and (g); (b), (c), (e), (f) and (h); (b), (c), (e), (g) and (h); (b), (c), (f), (g) and (h); (b), (d), (e), (f) and (g); (b), (d), (e), (f) and (h); (b), (d), (e), (g) and (h); (b), (d), (f), (g) and (h); (b), (e), (f), (g) and (h); (c), (d), (e), (f) and (g); (c), (d), (e), (f) and (h); (c), (d), (e), (g) and (h); (c), (d), (f), (g) and (h); (c), (e), (f), (g) and (h); (d), (e), (f), (g) and (h); (a), (b), (c), (d), (e) and (f); (a), (b), (c), (d), (e) and (g); (a), (b), (c), (d), (e) and (h); (a), (b), (c), (d), (f) and (g); (a), (b), (c), (d), (f) and (h); (a), (b), (c), (d), (g) and (h); (a), (b), (c), (e), (f) and (g); (a), (b), (c), (e), (f) and (h); (a), (b), (c), (e), (g) and (h); (a), (b), (c), (f), (g) and (h); (a), (b), (d), (e), (f) and (g); (a), (b), (d), (e), (f) and (h); (a), (b), (d), (e), (g) and (h); (a), (b), (d), (f), (g) and (h); (a), (b), (e), (f), (g) and (h); (a), (c), (d), (e), (f) and (g); (a), (c), (d), (e), (f) and (h); (a), (c), (d), (e), (g) and (h); (a), (c), (d), (f), (g) and (h); (a), (c), (e), (f), (g) and (h); (a), (d), (e), (f), (g) and (h); (b), (c), (d), (e), (f) and (g); (b), (c), (d), (e), (f) and (h); (b), (c), (d), (e), (g) and (h); (b), (c), (d), (f), (g) and (h); (b), (c), (e), (f), (g) and (h); (b), (d), (e), (f), (g) and (h); (c), (d), (e), (f), (g) and (h); (a), (b), (c),

(d), (e), (f) and (g); (a), (b), (c), (d), (e), (f) and (h); (a), (b), (c), (d), (e), (g) and (h); (a), (b), (c), (d), (f), (g) and (h); (a), (b), (c), (e), (f), (g) and (h); (a), (b), (d), (e), (f), (g) and (h); (a), (c), (d), (e), (f), (g) and (h); and finally, (b), (c), (d), (e), (f), (g) and (h). Furthermore, amino acid sequences in positions corresponding to those of amino acid sequences in (b), (d), (e) and (f) may differ preferably in no more than 1, 2, or 3 amino acid from the amino acid sequences in (b), (d), (e) and (f). Preferably, element (b) at least consists of I at position 136; element (d) at least consists of F at position 264; element (f) at least consists of an F or a Y at position 391. In one preferred embodiment, the encoded xylose isomerase comprises at least one of elements (a), (f), (g) and (h), More preferably, the encoded xylose isomerase in addition comprises at least one of elements (b) and (c) and most preferably the encoded xylose isomerase in addition comprises at least one of elements (d) and (e). In one preferred embodiment, the encoded xylose isomerase comprises at least element (a), more preferably in addition the isomerise comprises at least elements (b), (c) and/or (g), still more preferably in addition the isomerise comprises at least elements (d), (f) and/or (h), most preferably the isomerise additionally comprises element (e).

Amino acid positions of the above sequence elements (a) – (h) refer to positions in the reference amino acid sequence of the *Piromyces* xylose isomerase of SEQ ID NO: 3. In amino acid sequences of the invention other than SEQ ID NO: 3, preferably, the amino acid positions of sequence elements (a) – (h) are present in amino acid positions corresponding to the positions of sequence elements (a) – (h) in SEQ ID NO: 3, preferably in a ClustalW (1.83) sequence alignment using default settings. The skilled person will know how to identify corresponding amino acid positions in xylose isomerase amino acid sequences other than SEQ ID NO: 3 using amino acid sequence alignment algorithms as defined hereinabove. An example of such an alignment is depicted in Table 4, which shows a Clustal W (1.83) multiple sequence alignment of xylose isomerase amino acid sequences from organisms indicated on the left. The shaded amino acids the Piromyces sequence (SEQ ID NO: 3) in Table 4 indicate sequence elements that are shared among xylose isomerases that have the ability of functional expression in yeast.

In one embodiment the nucleotide sequence encoding the xylose isomerase encodes an amino acid sequence that is not one or more of SEQ ID NO: 3-7. Preferably the nucleotide sequence encoding the xylose isomerase does not have an amino acid

sequence that has more than (or has an amino acid sequence that has less than) 99, 98, 97, 96, 95, 94, 93, 92, 91, 90, 89, 88, 87, 86, 85, 84, 83, 82, 81, 80, 79, 78, 77, 76, 75, 74, 73, 72, 71 or 70% sequence identity with at least one amino acid sequence selected from the group consisting of SEQ ID NO: 3-7.

In one embodiment the nucleotide sequence encoding the xylose isomerase encodes an amino acid sequence that is not one or more of SEQ ID NO: 35. Preferably the nucleotide sequence encoding the xylose isomerase does not have an amino acid sequence that has more than (or has an amino acid sequence that has less than) 99, 98, 97, 96, 95, 94, 93, 92, 91, 90, 89, 88, 87, 86, 85, 84, 83, 82, 81, 80, 79, 78, 77, 76, 75, 74, 73, 72, 71 or 70% sequence identity with at least one amino acid sequence selected from the group consisting of SEQ ID NO: 35.

In one embodiment the nucleotide sequence encoding the xylose isomerase encodes an amino acid sequence that is not an amino acid sequences selected from the group consisting of:

(a) the *Piromyces* xylose isomerase disclosed in WO 03/062340;

(b) the *Bacteroides thetaiotaomicron* xylose isomerases disclosed in WO 04/099381 and in WO 06/009434);

(c) the *Cyllamyces* xylose isomerase disclosed in WO 04/099381; and,

(d) the *Orpinomyces* xylose isomerase disclosed in Madhavan et al. (2008, *supra*).

Preferably the nucleotide sequence encoding the xylose isomerase does not have an amino acid sequence that has more than (or has an amino acid sequence that has less than) 99, 98, 97, 96, 95, 94, 93, 92, 91, 90, 89, 88, 87, 86, 85, 84, 83, 82, 81, 80, 79, 78, 77, 76, 75, 74, 73, 72, 71 or 70% sequence identity with at least one amino acid sequence selected from the group consisting of:

(a) the *Piromyces* xylose isomerase disclosed in WO 03/062340;

(b) the *Bacteroides thetaiotaomicron* xylose isomerases disclosed in WO 04/099381 and in WO 06/009434);

(c) the *Cyllamyces* xylose isomerase disclosed in WO 04/099381; and,

(d) the *Orpinomyces* xylose isomerase disclosed in Madhavan et al. (2008, *supra*).

In one embodiment the nucleotide sequence encoding the xylose isomerase encodes an amino acid sequence that is not an amino acid of a xylose isomerase from an anaerobic fungus of the Family *Neocallimastigaceae*, such as a fungus from a genus

selected from the group consisting of: *Anaeromyces*, *Caecomyces*, *Cyllamyces*, *Neocallimastix*, *Orpinomyces*, *Piromyces*, and *Ruminomyces*.

In one embodiment the nucleotide sequence encoding the xylose isomerase encodes an amino acid sequence that is not an amino acid of a xylose isomerase from a bacterium from the genus *Bacteroides* or *Parabacteroides*.

The nucleotide sequences of the invention encode a novel class of xylose isomerases that may be functionally expressed in eukaryotic microbial host cells of the invention as defined below. The nucleotide sequences of the invention preferably encode xylose isomerases that naturally occurs in certain fungi, bacteria and tunicate.

A preferred nucleotide sequence of the invention thus encodes a xylose isomerase with an amino acid sequence that is identical to that of a xylose isomerase that is obtainable from (or naturally occurs in) a bacterium of the Family *Clostridiaceae*, more preferably a bacterium of the genus *Clostridium*, of which *C. difficile* is most preferred.

A preferred nucleotide sequence of the invention thus encodes a xylose isomerase with an amino acid sequence that is identical to that of a xylose isomerase that is obtainable from (or naturally occurs in) a bacterium of the Family *Fusobacteriaceae*, more preferably a bacterium of the genus *Fusobacterium*, of which *F. mortiferum* is most preferred.

A preferred nucleotide sequence of the invention thus encodes a xylose isomerase with an amino acid sequence that is identical to that of a xylose isomerase that is obtainable from (or naturally occurs in) a tunicate, preferably a tunicate of the Family *Cionidae*, more preferably a tunicate of the genus *Ciona*, of which *C. intestinales* is most preferred.

In another embodiment the nucleotide sequence encoding the xylose isomerase is selected from the group consisting of:

(a) a nucleotide sequence encoding a polypeptide with xylose isomerase activity, which polypeptide comprises an amino acid sequence that has at least 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 95, 96, 97, 98, or 99% sequence identity with the amino acid sequence of at least one of SEQ ID NO. 8, 9, 11 or 13;

(b) a nucleotide sequence the complementary strand of which hybridises to a nucleotide sequence of (a) or (b); and,

(c) a nucleotide sequence the sequence of which differs from the sequence of a nucleotide sequence of (d) due to the degeneracy of the genetic code.

It is however understood that nucleotide sequences encoding engineered forms of any of the xylose isomerases defined above and that comprise one or more amino acid substitutions, insertions and/or deletions as compared to the corresponding naturally occurring xylose isomerases but that are within the ranges of identity or similarity as defined herein are expressly included in the invention. Therefore, in one embodiment the nucleotide sequence of the invention encodes a xylose isomerase amino acid sequence comprising a xylose isomerise signature sequence as defined by Meaden et al. (1994, Gene, 141: 97-101): VXW[GP]GREG[YSTA] (present at positions 187-195) and [LIVM]EPKPX[EQ]P (present at positions 232-239), wherein "X" can be any amino acid and wherein amino acids in brackets indicates that one of the bracketed amino acids can be present at that position in the signature sequence. A xylose isomerase amino acid sequence of the invention further preferably comprises the conserved amino acid residues His-102, Asp-105, and Asp-340, which constitute a triad directly involved in catalysis, Lys-235 plays a structural as well as a functional catalytic role, and Glu-233, which is involved in binding of the magnesium (Vangrysperre et al., 1990, Biochem. J. 265: 699-705; Henrick et al., J. Mol. Biol. 208: 129-157; Bhosale et al., 1996 Microbiol. Rev. 60: 280-300). Amino acid positions of the above signature sequences and conserved residues refer to positions in the reference amino acid sequence of the *Piromyces* xylose isomerase of SEQ ID NO: 3. In amino acid sequences of the invention other than SEQ ID NO: 3, preferably, the amino acid positions of the above signature sequences and conserved residues are present in amino acid positions corresponding to the positions of the signature sequences and conserved residues in SEQ ID NO: 3, preferably in a ClustalW (1.83 or 1.81) sequence alignment using default settings. The skilled person will know how to identify corresponding amino acid positions in xylose isomerase amino acid sequences other than SEQ ID NO: 3 using amino acid sequence alignment algorithms as defined hereinabove. An example of such an alignment is depicted in Table 4. In addition, to date some 450 amino acid sequences of xylose isomerases are known in the art and new ones are added continuously being added. Sequence alignments of SEQ ID NO: 3 and the xylose isomerase sequences of the invention with these known and new xylose isomerase amino acid sequences will indicate further conserved regions and amino acid positions,

the conservation of which are important for structure and enzymatic activity. These regions and positions will tolerate no or only conservative amino acid substitutions. Amino acid substitutions outside of these regions and positions are unlikely to greatly affect xylose isomerase activity.

The nucleotide sequence encodes a xylose isomerase that is preferably expressed in active form in the transformed host cell. Thus, expression of the nucleotide sequence in the host cell produces a xylose isomerase with a specific activity of at least 10 U xylose isomerase activity per mg protein at 25°C, preferably at least 20, 25, 30, 50, 100, 200 or 300 U per mg at 25°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 in the Examples. Preferably, expression of the nucleotide sequence 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.

The nucleotide sequences of the invention, encoding polypeptides with xylose isomerase activity, are obtainable from genomic and/or cDNA of a fungus, yeast or bacterium that belongs to a phylum, class or genus as described above, using method for isolation of nucleotide sequences that are well known in the art per se (see e.g. Sambrook and Russell (2001) "Molecular Cloning: A Laboratory Manual (3rd edition), Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, New York). The nucleotide sequences of the invention are e.g. obtainable in a process wherein a) degenerate PCR primers (such as those in SEQ ID NO.'s 14 and 15) are used on genomic and/or cDNA of a suitable organism (e.g. a fungus, bacterium or tunicate as indicated above) to generate a PCR fragment comprising part of the nucleotide sequences encoding the polypeptides with xylose isomerase activity; b) the PCR fragment obtained in a) is used as probe to screen a cDNA and/or genomic library of the organism; and c) producing a cDNA or genomic DNA comprising the nucleotide sequence encoding a polypeptide with xylose isomerase activity.

To increase the likelihood that the xylose isomerase is expressed at sufficient levels and in active form in the transformed host cells of the invention, the nucleotide sequence encoding these enzymes, as well as other enzymes of the invention (see

below), are preferably adapted to optimise their codon usage to that of the host cell in question. The adaptiveness of a nucleotide sequence encoding an enzyme to the codon usage of a host cell may be expressed as codon adaptation index (CAI). The codon adaptation index is herein defined as a measurement of the relative adaptiveness of the codon usage of a gene towards the codon usage of highly expressed genes in a particular host cell or organism. The relative adaptiveness (w) of each codon is the ratio of the usage of each codon, to that of the most abundant codon for the same amino acid. The CAI index is defined as the geometric mean of these relative adaptiveness values. Non-synonymous codons and termination codons (dependent on genetic code) are excluded. CAI values range from 0 to 1, with higher values indicating a higher proportion of the most abundant codons (see Sharp and Li , 1987, Nucleic Acids Research 15: 1281-1295; also see: Jansen et al., 2003, Nucleic Acids Res. 31(8):2242-51). An adapted nucleotide sequence preferably has a CAI of at least 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8 or 0.9. Most preferred are the sequences as listed in SEQ ID NO's: 16, 17 and 38, which have been codon optimised for expression in *S. cerevisiae* cells.

The host cell to be transformed with a nucleic acid construct comprising a nucleotide sequence encoding a xylose isomerase of the invention preferably is a eukaryotic microbial host, more preferably a fungal host cell, such as a yeast or filamentous fungal host cell. Preferably the host cell is a cultured cell. The host cell of the invention, preferably is a host capable of active or passive pentose (xylose and preferably also arabinose) transport into the cell. The host cell preferably contains active glycolysis. The host cell may further preferably contains an endogenous pentose phosphate pathway and may contain endogenous xylulose kinase activity so that xylulose isomerised from xylose may be metabolised to pyruvate. The host further preferably contains enzymes for conversion of a pentose (preferably through pyruvate) to a desired fermentation product such as ethanol, lactic acid, 3-hydroxy-propionic acid, acrylic acid, acetic acid, succinic acid, citric acid, amino acids, 1,3-propane-diol, ethylene, glycerol, β-lactam antibiotics and cephalosporins. A particularly preferred host cell is a host cell that is naturally capable of alcoholic fermentation, preferably, anaerobic alcoholic fermentation. The host cell further preferably has a high tolerance to ethanol, a high tolerance to low pH (i.e. capable of growth at a pH lower than 5, 4, or 3) and towards organic acids like lactic acid, acetic acid or formic acid and sugar degradation products such as furfural and hydroxy-methylfurfural, and a high tolerance

to elevated temperatures. Any of these characteristics or activities of the host cell may be naturally present in the host cell or may be introduced or modified by genetic modification, preferably by self cloning or by the methods of the invention described below. A suitable cell is a cultured cell, a cell that may be cultured in fermentation process e.g. in submerged or solid state fermentation. Particularly suitable cells are eukaryotic microorganism like e.g. fungi, however, most suitable for use in the present inventions are yeasts or filamentous fungi.

Yeasts are herein defined as eukaryotic microorganisms and include all species of the subdivision Eumycotina (Yeasts: characteristics and identification, J.A. Barnett, R.W. Payne, D. Yarrow, 2000, 3rd ed., Cambridge University Press, Cambridge UK; and, The yeasts, a taxonomic study, C.P. Kurtzman and J.W. Fell (eds) 1998, 4th ed., Elsevier Science Publ. B.V., Amsterdam, The Netherlands) that predominantly grow in unicellular form. Yeasts may either grow by budding of a unicellular thallus or may grow by fission of the organism. Preferred yeasts as host cells belong to the genera *Saccharomyces*, *Kluyveromyces*, *Candida*, *Pichia, Schizosaccharomyces*, *Hansenula*, *Kloeckera*, *Schwanniomyces*, and *Yarrowia*. Preferably the yeast is capable of anaerobic fermentation, more preferably anaerobic alcoholic fermentation. 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 fungal host cells include *S. cerevisiae*, *S. exiguus*, *S. bayanus*, *K. lactis*, *K. marxianus* and *Schizosaccharomyces pombe*.

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 present invention are morphologically, physiologically, and genetically distinct from yeasts. Vegetative growth by filamentous fungi is by hyphal elongation and carbon catabolism of most filamentous fungi is obligately aerobic. Preferred filamentous fungi as host cells belong to the genera *Aspergillus*, *Trichoderma*, *Humicola*, *Acremonium*, *Fusarium*, and *Penicillium*.

In a transformed host cell of the invention, the nucleotide sequence encoding the xylose isomerase as defined above is preferably operably linked to a promoter that causes sufficient expression of the nucleotide sequences in the cell to confer to the cell the ability to convert xylose into xylulose. More preferably the promoter causes sufficient expression of the nucleotide sequences to confer to the cell the ability to grow on xylose as sole carbon and/or energy source, most preferably under anaerobic conditions. Suitable promoters for expression of the nucleotide sequence as defined above include promoters that are insensitive to catabolite (glucose) repression and/or that do require xylose for induction. Promoters 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 (*PPK*), triose phosphate isomerase (*TPI*), glyceraldehyde-3-phosphate dehydrogenase (*GPD, TDH3* or *GAPDH*), pyruvate kinase (*PYK*), phosphoglycerate kinase (*PGK*), glucose-6-phosphate isomerase promoter (*PGI1*) promoters from yeasts or filamentous fungi; more details about such promoters from yeast may be found in (WO 93/03159). Other useful promoters are ribosomal protein encoding gene promoters, the lactase gene promoter (*LAC4*), alcohol dehydrogenase promoters (*ADH1, ADH4*, and the like), the enolase promoter (*ENO*), the hexose(glucose) transporter promoter (*HXT7*), and the cytochrome c1 promoter (*CYC1*). Other promoters, both constitutive and inducible, and enhancers or upstream activating sequences will be known to those of skill in the art. Preferably the promoter that is operably linked to nucleotide sequence as defined above is homologous to the host cell.

The transformed host cell of the invention further preferably comprises xylulose kinase activity so that xylulose isomerised from xylose may be metabolised to pyruvate. Preferably, the cell contains endogenous xylulose kinase activity. More preferably, a cell of the invention comprises a genetic modification that increases the specific xylulose kinase activity. Preferably the genetic modification 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 cell or may be a xylulose kinase that is heterologous to the cell. A nucleotide sequence that may be used for overexpression of xylulose kinase in the cells of the invention is e.g. the xylulose kinase gene from *S. cerevisiae* (*XKS1*) as described by Deng and Ho (1990, Appl. Biochem. Biotechnol. 24-25: 193-199). Another preferred

xylulose kinase is a xylose kinase that is related to the xylulose kinase from *Piromyces* (*xylB*; see WO 03/0624430). This *Piromyces* xylulose kinase is actually more related to prokaryotic kinase than to all of the known eukaryotic kinases such as the yeast kinase. The eukaryotic xylulose kinases have been indicated as non-specific sugar kinases, which have a broad substrate range that includes xylulose. In contrast, the prokaryotic xylulose kinases, to which the *Piromyces* kinase is most closely related, have been indicated to be more specific kinases for xylulose, i.e. having a narrower substrate range. In the cells of the invention, a xylulose kinase to be overexpressed is overexpressed by at least a factor 1.1, 1.2, 1.5, 2, 5, 10 or 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 further preferably comprises a genetic modification that increases the flux of the pentose phosphate pathway as described in WO 06/009434. In particular, the genetic modification causes an increased flux of 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 1.1, 1.2, 1.5, 2, 5, 10 or 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 as described in WO 06/009434.

Genetic modifications that increase the flux of the pentose phosphate pathway may be introduced in the cells of the invention 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 cell of the invention, 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 3-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 3-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 3-epimerase and transketolase; or at least the enzymes ribulose-5-phosphate 3-epimerase and transaldolase; or at least the enzymes transketolase and transaldolase; or at least the enzymes ribulose-5-phosphate 3-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 3-epimerase, and transaldolase; or at least the enzymes ribulose-5-phosphate isomerase, ribulose-5-phosphate 3-epimerase, and transketolase. In one embodiment of the invention each of the enzymes ribulose-5-phosphate isomerase, ribulose-5-phosphate 3-epimerase, transketolase and transaldolase are overexpressed in the cell of the invention. Preferred is a cell in which the genetic modification comprises at least overexpression of the enzyme transaldolase. More preferred is a 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 we have found that cells overexpressing only the transketolase and the transaldolase already have the same anaerobic growth rate on xylose as do cells that overexpress all four of the enzymes, i.e. the ribulose-5-phosphate isomerase, ribulose-5-phosphate 3-epimerase, transketolase and transaldolase. Moreover, cells of the invention overexpressing both of the enzymes ribulose-5-phosphate isomerase and ribulose-5-phosphate 3-epimerase are preferred over cells overexpressing only the isomerase or only the 3-epimerase as overexpression of only one of these enzymes may produce metabolic imbalances.

There are various means available in the art for overexpression of enzymes in the cells of the invention. In particular, an enzyme may be overexpressed by increasing the copynumber of the gene coding for the enzyme in the cell, e.g. by integrating additional copies of the gene in the 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. The coding sequence used for overexpression of the enzymes preferably is homologous to the host cell of the invention. However, coding sequences that are heterologous to the host cell of the invention may likewise be applied.

Alternatively overexpression of enzymes in the 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 cell of the invention. 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 promoters as described above for expression of the nucleotide sequences encoding xylose isomerases as defined above.

A further preferred cell of the invention comprises a genetic modification that reduces unspecific aldose reductase activity in the 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 modifications reduce or inactivate the expression of each endogenous copy of a gene encoding an unspecific aldose reductase that is capable of reducing an aldopentose, including, xylose, xylulose and arabinose, in the cell's genome. A given cell may comprise multiple copies of genes encoding unspecific aldose reductases as a result of di-, poly- or aneu-ploidy, and/or a 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 cell of the invention and amino acid sequences of such aldose reductases are described in WO 06/009434 and include e.g. the (unspecific) aldose reductase genes of *S. cerevisiae GRE3* gene (Träff et al., 2001, Appl. Environm. Microbiol. 67: 5668-5674) and orthologues thereof in other species.

In a further preferred embodiment, the transformed cell of the invention that has the ability of isomerising xylose to xylulose as described above, in addition has the ability to convert L-arabinose into D-xylulose 5-phosphate as e.g. described in Wisselink et al. (2007, AEM Accepts, published online ahead of print on 1 June 2007; Appl. Environ. Microbiol. doi:10.1128/AEM.00177-07) and in EP 1 499 708. The ability of to converting L-arabinose into D-xylulose 5-phosphate is preferably conferred to the cell by transformation with a nucleic acid construct(s) comprising nucleotide sequences encoding a) an arabinose isomerase; b) a ribulokinase, preferably a L-ribulokinase a xylose isomerase; and c) a ribulose-5-P-4-epimerase, preferably a L-ribulose-5-P-4-epimerase. Preferably, in the cells of the invention, the ability to convert L-arabinose into D-xylulose 5-phosphate is the ability to convert L-arabinose into D-xylulose 5-phosphate through the subsequent reactions of 1) isomerisation of arabinose into ribulose; 2) phosphorylation of ribulose to ribulose 5-phosphate; and, 3) epimerisation of ribulose 5-phosphate into D-xylulose 5-phosphate. Suitable nucleotide sequences encoding arabinose isomerases, a ribulokinases and ribulose-5-P-4-epimerases may be obtained from *Bacillus subtilis*, *Escherichia coli* (see e.g. EP 1 499 708), *Lactobacilli*, e.g. *Lactobacillus plantarum* (see e.g. Wisselink et al. *supra*), or species of *Clavibacter*, *Arthrobacter* and *Gramella*, of which preferably *Clavibacter michiganensis, Arthrobacter aurescens* and *Gramella forsetii*.

A further preferred transformed host cell according to the invention may comprises further genetic modifications that result in one or more of the characteristics selected from the group consisting of (a) increased transport of xylose and/or arabinose into the cell; (b) decreased sensitivity to catabolite repression; (c) increased tolerance to ethanol, osmolarity or organic acids; and, (d) reduced production of by-products. By-products are understood to mean carbon-containing molecules other than the desired

fermentation product and include e.g. xylitol, arabinitol, glycerol and/or acetic acid. Any genetic modification described herein may be introduced by classical mutagenesis and screening and/or selection for the desired mutant, or simply by screening and/or selection for the spontaneous mutants with the desired characteristics. Alternatively, the genetic modifications may consist of overexpression of endogenous genes and/or the inactivation of endogenous genes. Genes the overexpression of which is desired for increased transport of arabinose and/or xylose into the cell are preferably chosen form genes encoding a hexose or pentose transporter. In *S. cerevisiae* and other yeasts these genes include *HXT1, HXT2, HXT4, HXT5, HXT7* and *GAL2*, of which *HXT7, HXT5* and *GAL2* are most preferred (see Sedlack and Ho, Yeast 2004; 21: 671–684). Another preferred transporter for expression in yeast is the glucose transporter encoded by the *P. stipitis SUT1* gene (Katahira et al., 2008, Enzyme Microb. Technol. 43: 115-119). Similarly orthologues of these transporter genes in other species may be overexpressed. Other genes that may be overexpressed in the cells of the invention include genes coding for glycolytic enzymes and/or ethanologenic enzymes such as alcohol dehydrogenases. Preferred endogenous genes for inactivation include hexose kinase genes e.g. the *S. cerevisiae HXK2* gene (see Diderich et al., 2001, Appl. Environ. Microbiol. 67: 1587-1593); the *S. cerevisiae MIG1* or *MIG2* genes; genes coding for enzymes involved in glycerol metabolism such as the *S. cerevisiae* glycerol-phosphate dehydrogenase 1 and/or 2 genes; or (hybridising) orthologues of these genes in other species. Other preferred further modifications of host cells for xylose fermentation are described in van Maris et al. (2006, Antonie van Leeuwenhoek 90:391–418), WO2006/009434, WO2005/023998, WO2005/111214, and WO2005/091733. Any of the genetic modifications of the cells of the invention as described herein are, in as far as possible, preferably introduced or modified by self cloning genetic modification.

In a preferred transformed host cell according to the invention, the nucleic acid construct confers to the host cell the ability to grow on xylose as carbon/energy source, preferably as sole carbon/energy source, and preferably under anaerobic conditions, i.e. conditions as defined herein below for anaerobic fermentation process. Preferably, when grown on xylose as carbon/energy source the transformed host produces essentially no xylitol, e.g. the xylitol produced is below the detection limit or e.g. less than 5, 2, 1, 0.5, or 0.3% of the carbon consumed on a molar basis. Preferably, in case carbon/energy source also includes arabinose, the cell produces essentially no

arabinitol, e.g. the arabinitol produced is below the detection limit or e.g. less than 5, 2, 1, 0.5, or 0.3 % of the carbon consumed on a molar basis.

A transformed host cell of the invention preferably has the ability to grow on xylose as sole carbon/energy source at a rate of at least 0.01, 0.02, 0.05, 0.1, 0.2, 0,25 or 0,3 h^{-1} under aerobic conditions, or, more preferably, at a rate of at least 0.005, 0.01, 0.02, 0.05, 0.08, 0.1, 0.12, 0.15 or 0.2 h^{-1} under anaerobic conditions. A cell of the invention preferably has the ability to grow on a mixture of glucose and xylose (in a 1:1 weight ratio) as sole carbon/energy source at a rate of at least 0.01, 0.02, 0.05, 0.1, 0.2, 0,25 or 0,3 h^{-1} under aerobic conditions, or, more preferably, at a rate of at least 0.005, 0.01, 0.02, 0.05, 0.08, 0.1, 0.12, 0.15 or 0.2 h^{-1} under anaerobic conditions. Thus, in a preferred transformed host cell according to the invention, the nucleic acid construct confers to the host cell the ability to anaerobically ferment xylose as sole carbon source in a process wherein ultimately pyruvate is used as an electron (and hydrogen acceptor) and is reduced to fermentation products such as ethanol, lactic acid, 3-hydroxy-propionic acid, acrylic acid, acetic acid, succinic acid, citric acid, amino acids, 1,3-propane-diol, ethylene, glycerol, β-lactam antibiotics and cephalosporins.

Preferably, a cell of the invention has a specific xylose consumption rate of at least 200, 300, 400, 600, 700, 800, 900 or 1000 mg h-1 (g dry weight)-1. Preferably, a cell of the invention has a yield of fermentation product (such as ethanol) on xylose that is at least 20, 40, 50, 60, 80, 90, 95 or 98% of the cell's yield of fermentation product (such as ethanol) on glucose. More preferably, the modified host cell's yield of fermentation product (such as ethanol) on xylose is equal to the host cell's yield of fermentation product (such as ethanol) on glucose. Likewise, the modified host cell's biomass yield on xylose is preferably at least 55, 60, 70, 80, 85, 90, 95 or 98% of the host cell's biomass yield on glucose. More preferably, the modified host cell's biomass yield on xylose is 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.

In another aspect the invention relates to a process for producing a fermentation product selected from the group consisting of ethanol, lactic acid, 3-hydroxy-propionic acid, acrylic acid, acetic acid, succinic acid, citric acid, amino acids, 1,3-propane-diol, ethylene, glycerol, β-lactam antibiotics and cephalosporins. The process preferably comprises the steps of: a) fermenting a medium containing a source of xylose, and

optionally arabinose, with a cell as defined hereinabove, whereby the cell ferments xylose, and optionally arabinose, to the fermentation product, and optionally, b) recovery of the fermentation product.

In addition to a source of xylose the carbon source in the fermentation medium may also comprise a source of glucose. The skilled person will further appreciate that the fermentation medium may further also comprise other types of carbohydrates such as e.g. in particular a source of arabinose. The sources of xylose, glucose and arabinose may be xylose, glucose and arabinose as such (i.e. as monomeric sugars) or they may be in the form of any carbohydrate oligo- or polymer comprising xylose, glucose and/or arabinose units, such as e.g. lignocellulose, arabinans, xylans, cellulose, starch and the like. For release of xylose, glucose and/or arabinose units from such carbohydrates, appropriate carbohydrases (such as arabinases, xylanases, glucanases, amylases, cellulases, glucanases 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 preferably during the fermentation. This, in turn, will prevent repression of systems required for metabolism and transport of non-glucose sugars such as xylose and arabinose. In a preferred process the modified host cell ferments both the xylose and glucose, and optionally arabinose, 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 eukaryotic microorganisms such as yeasts and filamentous fungi are well known in the art.

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 5, 2.5 or 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, as well as non-ethanol fermentation products such as lactic acid, 3-hydroxy-propionic acid, acrylic acid, acetic acid, succinic acid, citric acid, amino acids, 1,3-propane-diol, ethylene, glycerol, butyric acid, caproate, butanol, glyoxylate, β-lactam antibiotics and cephalosporins. Anaerobic processes of the invention are preferred over aerobic processes because anaerobic processes do not require investments and energy for aeration and in addition, anaerobic processes produce higher product yields than aerobic processes. Alternatively, the fermentation process of the invention may be run under aerobic oxygen-limited conditions. Preferably, in an aerobic process under oxygen-limited conditions, the rate of oxygen consumption is at least 5.5, more preferably at least 6 and even more preferably at least 7 mmol/L/h.

The fermentation process is preferably run at a temperature that is optimal for the modified cells of the invention. Thus, for most yeasts or fungal cells, the fermentation process is performed at a temperature which is less than 42°C, preferably less than 38°C. For yeast or filamentous fungal cells, the fermentation process is preferably performed at a temperature which is lower than 35, 33, 30 or 28°C and at a temperature which is higher than 20, 22, or 25°C.

Preferably in the fermentation processes of the invention, the cells stably maintain the nucleic acid constructs that confer to the cell the ability of isomerising xylose into xylulose, and optionally converting arabinose into D-xylulose 5-phosphate. Preferably in the process at least 10, 20, 50 or 75% of the cells retain the abilities of isomerising xylose into xylulose, and optionally converting arabinose into D-xylulose 5-phosphate after 50 generations of growth, preferably under industrial fermentation conditions.

A preferred fermentation process according to the invention is a process for the production of ethanol, whereby the process comprises the steps of: a) fermenting a medium containing a source of xylose, and optionally arabinose, with a cell as defined hereinabove, whereby the cell ferments xylose, and optionally arabinose, to ethanol, and optionally, b) recovery of the ethanol. The fermentation medium may further be

performed as described above. In the process the volumetric ethanol productivity is preferably at least 0.5, 1.0, 1.5, 2.0, 2.5, 3.0, 5.0 or 10.0 g ethanol per litre per hour. The ethanol yield on xylose and/or glucose and/or arabinose in the process preferably is at least 50, 60, 70, 80, 90, 95 or 98%. The ethanol yield is herein defined as a percentage of the theoretical maximum yield, which, for xylose, glucose and arabinose is 0.51 g. ethanol per g. xylose, glucose or arabinose.

A further preferred fermentation process according to the invention is a process which comprises fermenting a medium containing a source of xylose and a source of arabinose wherein however two separate strains of cells are used, a first strain of cells as defined hereinabove except that cells of the first strain do not have the ability to converting arabinose into D-xylulose 5-phosphate, which cells of the first strain ferment xylose to the fermentation product; and a second strain of cells as defined hereinabove except that cells of the second strain do not have the ability to (directly) isomerise xylose to xylulose, which cells of the second strain ferment arabinose to the fermentation product. The process optionally comprises the step of recovery of the fermentation product. The cells of the first and second are further as otherwise described hereinabove.

In this document and in its claims, the verb "to comprise" and its conjugations is used in its non-limiting sense to mean that items following the word are included, but items not specifically mentioned are not excluded. In addition, reference to an element by the indefinite article "a" or "an" does not exclude the possibility that more than one of the element is present, unless the context clearly requires that there be one and only one of the elements. The indefinite article "a" or "an" thus usually means "at least one".

All patent and literature references cited in the present specification are hereby incorporated by reference in their entirety.

The following examples are offered for illustrative purposes only, and are not intended to limit the scope of the present invention in any way.

Description of the figure

Figure 1 Phylogenetic tree of xylose isomerases tested for expression in yeast.

Examples

1. Functional expression of xylose isomerases of the invention in yeast

1.1 Host organism

The yeast host strain was RN1000. This strain is a derivative of strain RWB 218 (Kuyper et al., FEMS Yeast Research 5, 2005, 399-409). The plasmid pAKX002 encoding the Piromyces XylA is lost in RN1000. The genotype of the host strain is: *MatA, ura3-52, leu2-112, gre3::hphMX, loxP-Ptpi::TAL1, loxP-Ptpi::RKI1, pUGPtpi-TKL1, pUGPtpi-RPE1,{p415 Padh1XKS1Tcyc1-LEU2}.*

1.2 Expression-constructs with synthetic XI genes

Synthetic codon-optimised (for *Saccharomyces cerevisiae*) XI genes were cloned into a derivative of pRS306 (Sikorski R.S., Hieter P., 1989, "A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in Saccharomyces cerevisiae" Genetics 122:19-27) comprising the *Saccharomyces cerevisiae* the TPI1 promoter (899bp) and the CYC1 terminator (288bp) sequences, using *Xba*I (at the 5'-end of the synthetic genes) and *Bam*HI (at the 3'-end of the synthetic genes) restriction sites. The first three nucleotides in front of the ATG were modified into AAA in order to optimize expression. Table 1 provides a list of the XI sequence that were tested and the corresponding SEQ ID NO's depicting the synthetic sequences. Genes were synthesized by GenScript Corporation (zie www.genscript.com) and delivered cloned in pUC57:

http://www.ncbi.nlm.nih.gov/entrez/viewer.fcgi?db=nuccore&id=2440162.

The TPI promoter was obtained using yeast genomic DNA as template in a PCR with primers: forward: AAACCGGTTTCTTCTTCAGATTCCCTC

reverse: TTAGATCTCTAGATTTATGTATGTGTTTTTTGTAGT.

The CYC1 terminator was obtained using yeast genomic DNA as template in a PCR with primers: forward: AAGAATTCGGATCCCCTTTTCCTTTGTCGA

reverse : AACTCGAGCCTAGGAAGCCTTCGAGCGTC.

1.3 Transformation of the host organism and selection of transformants

RN1000 was transformed with plasmids using the 'Gietz method' (Gietz et al., 1992, Nucleic Acids Res. 1992 Mar 25;20(6):1425.). Primary selection of transformants was done on mineral medium (YNB+2% glucose) via uracil complementation.

1.4 Enzyme assays

Xylose isomerase activity is 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. One unit of activity is defined as the amount of enzyme producing 1 nmol of xylulose per min under the assay conditions. Xylulose formed is determined by the method of Dische and Borenfreund (Dische and Borenfreund, 1951, J. Biol. Chem. 192: 583-587) or by HPLC using a Biorad HPX-87N Column operated at 80°C and eluated at 0.6 ml/min using 0.01 M Na_2HPO_4 as the eluens. Xylose and xylulose are detected by a Refractive Index detector at an internal temperature of 60°C.

Specific activity is expressed as units per mg protein. Protein is determined with the Bio-Rad protein reagent (Bio-Rad Laboratories, Richmond, CA, USA) with bovine γ-globulin as a standard.

1.5 Physiological characterisation of the transformed cells

Transformed cells were colony-purified on minimal medium with glucose as sole carbon source several times. Subsequently colony-purified transformed cells are grown in shake flasks in the presence of oxygen on synthetic medium with 2% (w/v) xylose as carbon/energy source. The results are shown in Table 2 where "+" indicates that cells showed significant growth. The sign "-" denotes that no significant growth occurred.

Strains growing at the expense of xylose were subsequently tested for their ability to grow anaerobically at the expense of xylose with a concomitant formation of ethanol. Strains transformed with the xylose isomerases of *Cyllamyces aberensis, Ciona intestinalis, Clostridium difficile, Bacteroides fragilis* and *Fusobacterium_mortiferum* were able to grow anaerobically on xylose with growth rates, biomass and ethanol yields comparable to those of the prior art *Piromyces* enzyme, e.g. comparable to those of RWB218 as previously described by Kuyper et al. (2005, FEMS Yeast Res. 5: 925-934). However, cells transformed with the *C. difficile* XI showed the best performance in xylose fermentation in terms of growth rate.

Similarly the transformed cells of the invention are also capable of mixed substrate utilisation. When the transformed cells are grown in a mixture of glucose and xylose (20 g l^{-1} each) both sugars are completely consumed but glucose was the preferred substrate. Xylose consumption commences only after approximately 80% of the glucose is consumed. The ethanol produced accounted for the consumption of the total of glucose and xylose in each instance of the strains transformed with the xylose

isomerases of *Cyllamyces aberensis, Ciona intestinalis, Clostridium difficile, Bacteroides fragilis* and *Fusobacterium_mortiferum*.

Table 1

Name	length	CAI Initial genes	CAI optimized genes	nucleotide SEQ ID NO	Amino acid SEQ ID NO
*Piromyces*_sp_E2	1314	0.547	0.996	18	3
Cyllamyces_aberensis	1314	0.569	0.989	19	4
Clostridium_difficile	1338	0.107	0.987	16	1
Bacteroides_fragilis	1320	0.104	0.976	20	7
Ciona_intestinalis	1371	0.119	0.993	17	2
Thermotoga_maritima	1335	0.097	0.983	23	10
Haemophilus_somnus	1320	0.145	1.000	24	11
Physcomitrella_patens	1440	0.090	0.988	26	13
Arabidopsis_thaliana	1434	0.154	1.000	25	12
Arthrobacter_aurescens	1188	0.082	0.988	22	9
Burkholderia_phytofirmans	1323	0.065	0.982	21	8
*Orpinomyces*_sp._Ukk1	1314	0.515	Not tested	Not tested	5
Fusobacterium_mortiferum	1311	0.106	0.989	38	35

Table 2

XI source organism	**Growth on 2% xylose in synthetic medium**	**% amino acid identity with *Piromyces* XI**
Piromyces	+	100
Cyllamyces aberensis	+	91
Physcomitrella patens	-	52
Arabidopsis thaliana	-	51
Haemophilus somnus	-	49
Ciona intestinalis	+	47
Clostridium difficile	+	54
Thermatoga maritima	-	51
Bacteroides fragilis	+	78
Burkholderia phytofirmans	-	50
Arthrobacter aurescens	-	17
Fusobacterium_mortiferum	+	51

Table 3 Pairwise amino acid identities between xylose isomerases from various organisms as indicated.

		1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	Piromyces_sp_E2														
2	Bacteroides_fragilis	78,71													5
3	Cyllamyces_aberensis	91,30	77,57												
4	Clostridium_difficile	54,00	52,61	53,78											
5	Ciona_intestinalis	47,83	45,10	45,30	52,36										
6	Fusobacterium_mortiferum	50,92	50,45	51,15	71,79	50,46									
7	Orpinomyces_sp_ukk1	94,51	78,03	91,07	55,38	47,14	50,23								
8	Clostridium_phytofermentans	53,55	53,88	54,92	62,56	46,80	64,68	54,23							
9	Bacteroides_uniformis	81,24	90,41	80,55	55,03	47,49	52,75	80,55	54,79						10
10	Thermotoga_maritima	51,26	51,02	51,72	65,99	50,67	64,91	50,11	59,82	52,28					
11	Haemophilus_somnus	48,74	49,43	49,20	49,89	45,10	52,06	48,74	53,42	50,68	51,03				
12	Arabidopsis_thaliana	51,49	51,71	52,40	53,03	51,10	53,90	51,26	49,77	52,97	55,41	48,75			
13	Physcomitrella_patens	51,72	52,39	52,17	52,81	51,31	52,75	51,03	50,68	53,42	55,86	48,29	70,23		
14	Burkholderia_phytofirmans	49,43	48,74	48,74	48,18	44,77	49,09	48,98	50,23	50,23	49,09	64,69	46,36	47,05	
15	Arthrobacter_aurescens	20,76	20,76	20,76	19,75	20,51	20,25	20,76	21,52	20,25	20,51	19,49	19,24	20,25	20,76

15

Table 4 CLUSTAL W (1.83) multiple sequence alignment 1/5

```
Arabidopsis_thaliana            MKKVEFFMLLLCFIAASSLVSADPPTCPADLGGKCSDSDDWQGDFFPEIP
Physcomitrella_patens           MKALLFSVVLLVAVLSCSGQRVADITCGVDG-SLGSDFQEWEGEFFPNIS
Ciona_intestinalis              ----------------------MSSFAPASGKSDLAEAGSLLTKYPLEVK
Piromyces_sp_E2                 ----------------------------------------MAKEYFPQIQ
Orpinomyces_sp_ukk1             ----------------------------------------MTKEYFPTIG
Cyllamyces_aberensis            ----------------------------------------MVKEYFPAIQ
Bacteroides_fragilis            ----------------------------------------MATKEYFPGIG
Bacteroides_uniformis           ----------------------------------------MATKEYFPGIG
Clostridium_difficile           ----------------------------------------MNEIFKGIG
Fusobacterium_mortiferum        ----------------------------------------MEFFKGID
Thermotoga_maritima             ----------------------------------------MAEFFPEIP
Clostridium_phytofermentans     ----------------------------------------MKNYFPNVP
Haemophilus_somnus              ----------------------------------------MSNYFDKIA
Burkholderia_phytofirmans       ----------------------------------------MSYFEHIP
Arthrobacter_aurescens          --------------------------------------------------

Arabidopsis_thaliana            KIKYE-GPSSKNPLAYRWYNAEEEILGKKMKDWFRFSVAFWHTFRGTGGD
Physcomitrella_patens           YIKYE-GPASHNPLAYKYYNAEELIFGKKMKDWLRFSVAFWHTFRGDGGD
Ciona_intestinalis              KIPYKPDAKVDDVLCFRHYNESEVVMGKPMSDWLRFSVCYWHTFRGTGAD
Piromyces_sp_E2                 KIKFE-GKDSKNPLAFHYYDAEKEVMGKKMKDWLRFAMAWWHTLCAEGAD
Orpinomyces_sp_ukk1             KIRFE-GKDSKNPMAFHYYDAEKEVMGKKMKDWLRFAMAWWHTLCADGAD
Cyllamyces_aberensis            KIKFE-GKDSKNPMAFHYYDAEKEIMGKKMKDWLRFAMAWWHTLCAEGSD
Bacteroides_fragilis            KIKFE-GKDSKNPMAFRYYDAEKMINGRSMKDWLKFAMAWWHTLCAEGGD
Bacteroides_uniformis           KIKFE-GKESKNPMAFRYYDADKVIMGKKMSEWLKFAMAWWHTLCAEGGD
Clostridium_difficile           QIKFE-GVKSNNELAFRYYNPEQVVGNKTMKEHLRFAMSYWHTLCGEGND
Fusobacterium_mortiferum        KVKYE-GVKTNNLLAFAHYNPEEVILGKKMKDHLKFAMSYWHTLTGEGTD
Thermotoga_maritima             KIQFE-GKESTNPLAFRFYDPNEVIDGKPLKDHLKFSVAFWHTFVNEGRD
Clostridium_phytofermentans     EVKYE-GPNSTNPFAFKYYDANKVVAGKTMKEHCRFALSWWHTLCAGGAD
Haemophilus_somnus              KVNYE-GANSTNPFAFKHYNPNEVILGKTVEEHLRLAVCYWHTFCWTGND
Burkholderia_phytofirmans       EIRYE-GPQSDNPLAYRHYDKSKKVLGKTLEEHLRIAVCYWHTFVWPGVD
Arthrobacter_aurescens          ------------------------MTPQPTPQDRFTFGLWTVG-WTGAD
                                                         :::.  * .    * *
```

Table 4 continued 2/5

```
Arabidopsis_thaliana                PFGAATKYWPWEDGTNSVSMAKRRMRANFEFLKKLGVDWWCFHDRDIAPD
Physcomitrella_patens               PFGSPTKRWPWDDGSNSLTVAVRRMRANFEFLKKLGVEKWCFHDRDIAPE
Ciona_intestinalis                  PFGFPTLVRPWDDGTDSIENAERRMRVAFDFMSKLGVKYWTFHDRDIAPE
Piromyces_sp_E2                     QFGGGTKSFPWNEGTDAIEIAKQKVDAGFEIMQKLGIPYYCFHDVDLVSE
Orpinomyces_sp_ukk1                 QFGVGTKSFPWNEGTDPIAIAKQKVDAGFEIMTKLGIEHYCFHDVDLVSE
Cyllamyces_aberensis                QFGPGTKTFPWNEGTDPIEKAKQKVDAGFEIMTKLGIEHYCFHDVDLVDE
Bacteroides_fragilis                QFGGGTKQFPWNGDPDPVQAAKNKMDAGFEFMQKMGIGYYCFHDVDLVTE
Bacteroides_uniformis               QFGGGTKKFPWNGEADKVQAAKNKMDAGFEFMQKMGIEYYCFHDVDLCEE
Clostridium_difficile               PFGVGTVERPWNNITDPIEIAKIKVDAGFEFMSKMGIEYFCFHDRDIAPE
Fusobacterium_mortiferum            PFGNATMDREWN-EYTPMEKAKARVKAGFEFMEKLGLEYFCFHDKDIAPE
Thermotoga_maritima                 PFGDPTAERPWNRFSDPMDKAFARVDALFEFCEKLNIEYFCFHDRDIAPE
Clostridium_phytofermentans         PFGVTTMDRTYGNITDPMELAKAKVDAGFELMTKLGIEFFCFHDADIAPE
Haemophilus_somnus                  MFGAGSLDRSWQKTGDLLVGAKQKAEIAFEFFQKLGVPYYSFHDVDIAPE
Burkholderia_phytofirmans           IFGQGTFRRPWQQAGDAMERAQQKADSAFEFFSKLGTPYYTFHDTDVSPE
Arthrobacter_aurescens              PFGVATR----------PALDPVEAVHKLSELGAYGITFHDNDLVPF
                                     **  :             .      ..   ::.     *** *:

Arabidopsis_thaliana                GTTLEESNKNLDEVIELAKELQKGSKIKPLWGTAQLFLHPRYMHGGATSS
Physcomitrella_patens               GSTLEESNANLDYIVSVAKKLQEGTNIRPLWGTAQLFKHPRYMHGAATSP
Ciona_intestinalis                  GVTLSETNANLDRLAELASQLQGETGIKLLWNTCNLFAHPRYSNGAATNA
Piromyces_sp_E2                     GNSIEEYESNLKAVVAYLKEKQKETGIKLLWSTANVFGHKRYMNGASTNP
Orpinomyces_sp_ukk1                 GNSIEEYESNLKQVVAYLKQKQQETGIKLLWSTANVFGNPRYMNGASTNP
Cyllamyces_aberensis                GKNVEEYEKNLKTIVAYLKEKQKETGIKLLWSTANVFGHKRYMNGASTNP
Bacteroides_fragilis                ADSIEAYEANLKELVAYAKQKQAETGIKLLWGTANVFSHARYMNGAATNP
Bacteroides_uniformis               AETIEEYEANLKEIVAYAKQKQAETGIKLLWGTANVFGHARYMNGAATNP
Clostridium_difficile               GRDLEETNKILDEIVEYIKANMEKTGIKLLWGTANMFGNPRFVHGASTTC
Fusobacterium_mortiferum            AETLEEYHRNLDEIVDLIEEEMKRTGIKLLWGTSNMFSHPRFMHGAATSC
Thermotoga_maritima                 GKTLRETNKILDKVVERIKERMKDSNVKLLWGTANLFSHPRYMHGAATTC
Clostridium_phytofermentans         GDTFEESKKNLFEIVDYIKEKMDQTGIKLLWGTANNFSHPRFMHGASTSC
Haemophilus_somnus                  SNNFKEYLHNFNTIVDILEKKQSETGVKLLWGTANCFTNPRYMSGASTNP
Burkholderia_phytofirmans           GSNLKEYSENFLRITDYLARKQESTGIKLLWGTANLFSHPRYAAGAATSP
Arthrobacter_aurescens              DATASER----DLILKNFKAALAETGLKTPMVTTNLFSHPVFKDGGFTSN
                                               :          : ::    * : * :  :  *. *.
```

Table 4 continued 3/5

```
Arabidopsis_thaliana            EVGVYAYAAAQVKKAMEVTHYLGGENYVFWGGREGYQTLLNTDMGRGLDH
Physcomitrella_patens           DVRVYAYAAAQVKKAIEVTKLLGGENYVFWGGREGYQTLLNTDLKKELDH
Ciona_intestinalis              DAHVVAYAAAQVKKSLEIGKKLGAENFVFWGGREGYHTLLNTNVREELDN
Piromyces_sp_E2                 DFDVVARAIVQIKNAIDAGIELGAENYVFWGGREGYMSLLNTDQKREKEH
Orpinomyces_sp_ukk1             DFDVVARAIVQIKNAMDAGIELGAENYVFWGGREGYMSLLNTDQKREKEH
Cyllamyces_aberensis            DFDVVARAIVQIKNAMDAGIELGAENYVFWGGREGYMSLLNTDQKREKEH
Bacteroides_fragilis            DFDVVARAAVQIKNAIDATIELGGTNYVFWGGREGYMSLLNTDQKREKEH
Bacteroides_uniformis           DFDVVARAAIQIKNAIDATIELGGSNYVFWGGREGYMSLLNTDQKREKEH
Clostridium_difficile           NADVYAYAAAQVKKAMEITKYLGGENFVFWGGREGYETLLNTNTELEMDN
Fusobacterium_mortiferum        NADVFAYAAAQTKKALEITKRLNGTGYVFWGGREGYETLLNTDIGLELDN
Thermotoga_maritima             SADVFAYAAAQVKKALEITKELGGEGYVFWGGREGYETLLNTDLGLELEN
Clostridium_phytofermentans     NADVFAYAAAKIKNALDATIKLGGKGYVFWGGREGYETLLNTDLGLELDN
Haemophilus_somnus              NPEVFAWAAAQVFTAMNATQRLGGENYVLWGGREGYETLLNTDLKREREQ
Burkholderia_phytofirmans       DPEVFAFAATQVRHALDATQRLGGDNYVLWGGREGYDTLLNTDLVRERDQ
Arthrobacter_aurescens          DRSIRRFALSKILRNIDLAAELGAETFVMWGGREGSEYDGSKDLSAALDR
                                .  :   *  :    ::    *..  :*:******     ..:     :.

Arabidopsis_thaliana            LARFFEAAVAYKKKIGFKGTLLIEPKPQEPTKHQYDWDAATAANFLRKY-
Physcomitrella_patens           MATFLRSAAEYKKKIGFEGTLLLEPKPQEPTKHQYDWDAATTMGFLKNYG
Ciona_intestinalis              LANFFKMVVAYKKKIGFTGQFLIEPKPKEPSKHQYDYDAMTVIAFLKTY-
Piromyces_sp_E2                 MATMLTMARDYARSKGFKGT[illegible]LIEPKPMEPTKHQYDVDTETAIGFLKAHN
Orpinomyces_sp_ukk1             MATMLTMARDYARSKGFKGTFLIEPKPMEPTKHQYDVDTETVIGFLRAHN
Cyllamyces_aberensis            MAMMLGLARDYARSKGFKGTFLIEPKPMEPTKHQYDVDTETVIGFLRAHG
Bacteroides_fragilis            LAQMLTIARDYGRARGFKGTFLIEPKPMEPTKHQYDVDTETVIGFLKAHG
Bacteroides_uniformis           LAQMLTIARDYARARGFKGTFLIEPKPMEPTKHQYDVDTETVIGFLKAHN
Clostridium_difficile           FARFLQMAVDYAKEIGFTGQFLIEPKPKEPTKHQYDFDTATVLGFLRKYN
Fusobacterium_mortiferum        LARFLQMAVDYAKKIGFEGQFFIEPKPKEPTKHQYDFDTTTVLEFLRKYN
Thermotoga_maritima             LARFLRMAVEYAKKIGFTGQFLIEPKPKEPTKHQYDFDVATAYAFLKNHG
Clostridium_phytofermentans     MARLMKMAVEYGRANGFDGDFYIEPKPKEPTKHQYDFDTATVLAFLRKYG
Haemophilus_somnus              IGRFMQLVVEHKYKIGFQGTLLIEPKPQEPTKHQYDYDVATVYGFLKQFG
Burkholderia_phytofirmans       LARFLHMVVDHAHKIGFKGSLLIEPKPQEPTKHQYDYDVATVHGFLLQHG
Arthrobacter_aurescens          MKEGVDTAAGYIKEKGYGLRIALEPKPNEPRGDIFLPTVGHGLAFIAQLE
                                :   .  .  :    *:   : :**** **  . :   .     *:
```

Table 4 continued 4/5

```
Arabidopsis_thaliana              --GLIDEFKLNIECNHATLSGHTCHHELETARINGLLGNIDANTGDAQTG
Physcomitrella_patens             LSGTNNDYKLNLECNHATLSGHSCHHELETARIYGMLGSVDANTGDAQTG
Ciona_intestinalis                --DLDKDFKLNIEPNHTTLAGHCHEHDVVMASAYNMLGSVDSNTGSPDLG
Piromyces_sp_E2                   ---LDKDFKVNIEVNHATLAGHTFEHELACAVDAGMLGSIDANRGDYQNG
Orpinomyces_sp_ukk1               ---LDKDFKVNIEVNHATLAGHTFEHELACAVDAGMLGSIDANRGDYQNG
Cyllamyces_aberensis              ---LDKDFKINIEVNHATLAGHTFEHELACAVDAGMLGSIDANRGDYQNG
Bacteroides_fragilis              ---LNQDFKVNIEVNHATLAGHTFEHELAVAVDNGMLGSIDANRGDYQNG
Bacteroides_uniformis             ---LDKDFKVNIEVNHATLAGHTFEHELAVAVDNGMLGSIDANRGDYQNG
Clostridium_difficile             ---LDKYFKVNIEANHATLAGHTFQHELNIARINNVLGSIDANQGDLLLG
Fusobacterium_mortiferum          ---LDKYFKMNIEANHATLAGHTFQHELCTARINGVFGSIDANQGDMLLG
Thermotoga_maritima               ---LDEYFKFNIEANHATLAGHTFQHELRMARILGKLGSIDANQGDLLLG
Clostridium_phytofermentans       ---LEKDFKMNIEANHATLAGHTFEHELAMARVNGAFGSVDANQGDPNLG
Haemophilus_somnus                ---LENEIKVNIEANHATLAGHTFQHEIATATSLGIFGSIDANRGDPQLG
Burkholderia_phytofirmans         ---LDKEIRVNIEANHATLAGHSFHHEIATAYALGIFGSVDANRGDPQNG
Arthrobacter_aurescens            HG---DIVGLNPETGHEQMAGLNFTHGIAQALWAGKLFHIDLN-GQRGIK
                                       .   .* * .*  ::*    * :  *   . :  :* * *.

Arabidopsis_thaliana              WDTDQFLTDVGEATMVMMSVIKNGGIAPGGFNFDAKLRRES--TDVEDLF
Physcomitrella_patens             WDTDQFLTDVSEATLIMLSVIKNGGLAPGGFNFDAKLRRES--VDVEDLF
Ciona_intestinalis                WDTDQFPMDVKNATMIMQTVLEQGGLAPGGLNFDCKVRRES--TDVIDMM
Piromyces_sp_E2                   WDTDQFPIDQYELVQAWMEIIRGGGFVTGGTNFDAKTRRNS--TDLEDII
Orpinomyces_sp_ukk1               WDTDQFPIDQYELVQAWMEIIRGGGFVTGGTNFDAKTRRNS--TDLEDII
Cyllamyces_aberensis              WDTDQFPIDQYELVQAWMEIIRGGGFTTGGTNFDAKTRRNS--TDLEDII
Bacteroides_fragilis              WDTDQFPIDNFELTQAMMQIIRNDGLGNGGTNFDAKTRRNS--TDPEDIF
Bacteroides_uniformis             WDTDQFPIDNFELTQAMMQIIRNGGFGNGGTNFDAKTRRNS--TDLEDIF
Clostridium_difficile             WDTDQFPTNIYDATLAMYEVLKQGGIAPGGFNFDSKVRRAS--FEVEDLF
Fusobacterium_mortiferum          WDTDQFPTNVYDAVLAMYETLLAGGFKEGGLNFDAKVRRGS--FEPKDLF
Thermotoga_maritima               WDTDQFPTNIYDTTLAMYEVIKAGGFTKGGLNFDAKVRRAS--YKVEDLF
Clostridium_phytofermentans       WDTDQFPTDVHSATLAMLEVLKAGGFTNGGLNFDAKVRRGS--FEFDDIA
Haemophilus_somnus                WDTDQFPNSVEENTLAMYEILKAGGFTTGGFNFDAKIRRQS--TDPYDLF
Burkholderia_phytofirmans         WDTDQFPNSVEELTLAFYEILKHGGFTTGGMNFDSKVRRQS--VDPEDLF
Arthrobacter_aurescens            YDQDLVFGHGDLTSAFFTVDLLENGFPNGGPKYDGPRHFDYKPSRTDGYD
                                  :* * .             :  .*:  ** ::*   :        .
```

Table 4 continued 5/5

```
Arabidopsis_thaliana          IAHISGMDTMARGLRNAVKILEEGSLSELVRKRYATWDSELGKQIEEGKA
Physcomitrella_patens         IAHISGMDTIARGLRNAAKLLEEGRLTKLVEDRYSSFNSPLGKTIEEGKV
Ciona_intestinalis            IAHVGAMDCFAKALKIAAKIREDGVLGKMKKERYASFGSGLGLKIKTGTA
Piromyces_sp_E2               IAHVSGMDAMARALENAAKLLQESPYTKMKKERYASFDSGIGKDFEDGKL
Orpinomyces_sp_ukk1           IAHISGMDAMARALENAAKLLQESPYCNMKKERYASFDSGIGKDFEDGKL
Cyllamyces_aberensis          IAHISGMDAMARALENAAKLLTESPYKKMKADRYASFDSGMGKDFEDGKL
Bacteroides_fragilis          IAHIAGMDAMARALESAANLLNESPYQKMLSDRYASFDAGKGKEFEEGKL
Bacteroides_uniformis         IAHIAGMDVMARALESAAKLLEESPYKKMLADRYASFDSGKGKEFEDGKL
Clostridium_difficile         LAYIAGMDTFAKGLLIAHKLLEDEVFENFTKERYASFSEGIGKDIVEGKV
Fusobacterium_mortiferum      YAYISGMDTFAKGLKVAAKLIEDGTFEKIKVERYSSYTTGIGKQIVNGEV
Thermotoga_maritima           IGHIAGMDTFALGFKIAYKLAKDGVFDKFIEEKYRSFKEGIGKEIVEGKT
Clostridium_phytofermentans   YGYIAGMDTFALGLIKAAEIIDDGRIAKFVDDRYASYKTGIGKAIVDGTT
Haemophilus_somnus            HAHIGAMDVLALSLKRAAKMIEDQTLQKVVDNRYAGWDQELGQKILNGKA
Burkholderia_phytofirmans     YGHIGAIDNLALAVERAAVLIENDRLDQFKRQRYSGWDAEFGRKISSGDY
Arthrobacter_aurescens        GVWESAKSNMSMYLLLKERALAFRADPEVQEALATSGVFELGEPTLNAGE
                                  .. . ::  .            :.           *     .

Arabidopsis_thaliana          DFEYLEKKAKEF-GEPKVSSAKQELAEMIFQSAM---------
Physcomitrella_patens         GFEELEKISLEA-EEPPITSGKQELAEMIFYSYV---------
Ciona_intestinalis            TLEECDSFIQEN-GEPAKLSGKQEMFEAVLNRYF---------
Piromyces_sp_E2               TLEQVYEYGKKN-GEPKQTSGKQELYEAIVAMYQ---------
Orpinomyces_sp_ukk1           TLEQVYEYGKKN-GEPKVTSGKQELYEAIVAMYQ---------
Cyllamyces_aberensis          TFEQVYEYGKKV-NEPKQTSGKQELYEAIVAMYM---------
Bacteroides_fragilis          SLEELVAYAKAN-GEPKQTSGQQELYEALVNIYSL--------
Bacteroides_uniformis         TLEDLVAYAKAN-GEPKQTSGKQELYEAIVNMYC---------
Clostridium_difficile         GFKELESYALQM-PVIKNKSGRQEMLEAILNRYIYEVDTISNK
Fusobacterium_mortiferum      GFEELSKYALTN-GVKKNSSGRQEMLENILNRYIYE-------
Thermotoga_maritima           DFEKLEEYIIDK-EDIELPSGKQEYLESLLNSYIVKTIAELR-
Clostridium_phytofermentans   SLEELEQYVLTH-SEPVMQSGRQEVLETIVNNILFR-------
Haemophilus_somnus            SLEDLAKIVETQGLAPKPVSGQQEYLENLVNSYLYR-------
Burkholderia_phytofirmans     SLSALAEEAMARGLNPQHASGHQELMENIVNQAIYSGR-----
Arthrobacter_aurescens        TTADLLADASAFDTFNADQAAERSFAFVRLNQLAIEHLLGAR-
                                          :..:.     .
```

Claims

1. A eukaryotic microbial cell comprising nucleotide sequences the expression of which confers to, or increases in the cell the ability to directly isomerise xylose into xylulose, wherein the nucleotide sequence encodes a polypeptide with xylose isomerase activity, which polypeptide comprises an amino acid sequence comprising one or amino acid sequence elements selected from the group consisting of:

(a) a methionine residue at position 91;

(b) the amino acid sequence TGIKLL at positions 134-139;

(c) a phenylalanine residue at position 230;

(d) the amino acids phenylalanine and lysine at positions 264 and 265, respectively;

(e) the amino acid sequence TLAGH at positions 274-278;

(f) the amino acid sequence RYASF at positions 387-391;

(g) a glycine residue at position 394; and,

(h) an alanine residue at position 431;

wherein the polypeptide does not comprise an amino acid sequence having more than 95% sequence identity with one of SEQ ID NO: 3-7.

2. A eukaryotic microbial cell according to claim 1, wherein the nucleotide sequence encoding the xylose isomerase is selected from the group consisting of:

(a) a nucleotide sequence encoding a polypeptide with xylose isomerase activity, which polypeptide comprises an amino acid sequence that has at least 45% sequence identity with the amino acid sequence of SEQ ID NO. 1;

(b) a nucleotide sequence encoding a polypeptide with xylose isomerase activity, which polypeptide comprises an amino acid sequence that has at least 45% sequence identity with the amino acid sequence of SEQ ID NO. 2;

(c) a nucleotide sequence encoding a polypeptide with xylose isomerase activity, which polypeptide comprises an amino acid sequence that has at least 45% sequence identity with the amino acid sequence of SEQ ID NO.35;

(d) a nucleotide sequence the complementary strand of which hybridises to a nucleotide sequence of (a) (b) or (c); and,

(e) a nucleotide sequence the sequence of which differs from the sequence of a nucleotide sequence of (d) due to the degeneracy of the genetic code.

3. A cell according to claim 2, wherein the nucleotide sequence encodes an amino acid sequence that is obtainable from a bacterium from the genera *Clostridium* and *Fusobacterium* or from a tunicate from the genus *Conia*.

4. A cell according to any one of the preceding claims, wherein the cell is a yeast or a filamentous fungus of a genus selected from the group consisting of *Saccharomyces*, *Kluyveromyces*, *Candida*, *Pichia*, *Schizosaccharomyces*, *Hansenula*, *Kloeckera*, *Schwanniomyces*, *Yarrowia*, *Aspergillus*, *Trichoderma*, *Humicola*, *Acremonium*, *Fusarium*, and *Penicillium*.

5. A cell according to claim 4, wherein the cell is a yeast is capable of anaerobic alcoholic fermentation.

6. A cell according to claim 5, wherein the yeast belongs to a species selected from the group consisting of *S. cerevisiae*, *S. exiguus*, *S. bayanus*, *K. lactis*, *K. marxianus* and *Schizosaccharomyces pombe*.

7. A cell according to any one of the preceding claims, wherein the nucleotide sequence encoding the polypeptide with xylose isomerase activity is operably linked to a promoter that causes sufficient expression of the nucleotide sequences in the cell to confer to the cell the ability to isomerise xylose into xylulose.

8. A cell according to any one the preceding claims, whereby the cell further comprises a genetic modification that increases the specific xylulose kinase activity.

9. A cell according to any one the preceding claims, wherein the cell comprises a genetic modification that increases the flux of the pentose phosphate pathway.

10. The cell according to any one the preceding claims, whereby the host cell comprises a genetic modification that reduces unspecific aldose reductase activity in the cell.

11. A cell according to any one the preceding claims, wherein the cell exhibits the ability to convert L-arabinose into D-xylulose 5-phosphate.

12. A cell according to any one the preceding claims, wherein the cell comprises at least one further genetic modification that result in a characteristic selected from the group consisting of:

(a) increase transport of at least one of xylose and arabinose into the host cell;

(b) decreased sensitivity to catabolite repression;

(c) increased tolerance to ethanol, osmolarity or organic acids; and,

(d) reduced production of by-products.

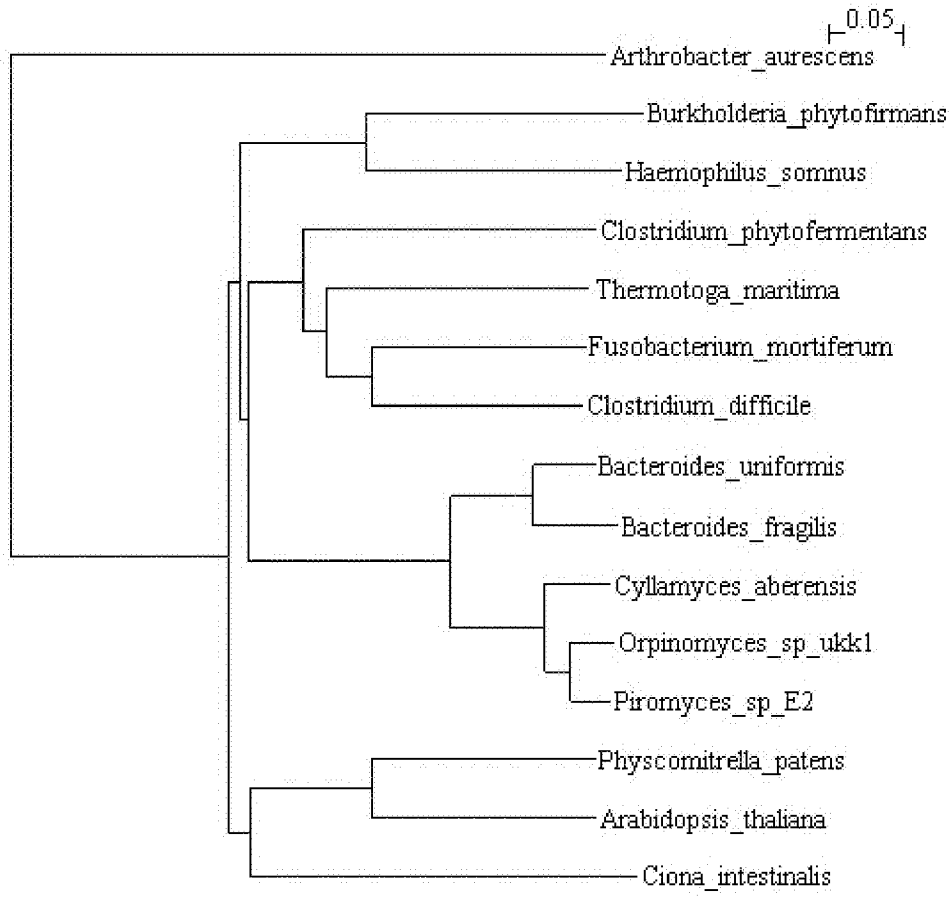
13. A cell according to any one of the preceding claims, wherein the cell has the ability to produce at least one fermentation product selected from the group consisting of ethanol, lactic acid, 3-hydroxy-propionic acid, acrylic acid, acetic acid, succinic acid, citric acid, amino acids, 1,3-propane-diol, ethylene, glycerol, butyric acid, caproate, butanol, glyoxylate, β-lactam antibiotics and cephalosporins.

14. A process for producing a fermentation product selected from the group consisting of ethanol, lactic acid, 3-hydroxy-propionic acid, acrylic acid, acetic acid, succinic acid, citric acid, amino acids, 1,3-propane-diol, ethylene, glycerol, butyric acid, caproate, butanol, glyoxylate, β-lactam antibiotics and cephalosporins, whereby the process comprises the steps of:

(a) fermenting a medium containing a source of xylose, and optionally a source of arabinose, with a cell as defined in any one of claims 1 - 13, whereby the cell ferments xylose to the fermentation product, and optionally,

(b) recovery of the fermentation product.

15. A process according to claim 14, whereby the medium also contains a source of glucose.


0.05
Arthrobacter_aurescens
Burkholderia_phytofirmans
Haemophilus_somnus
Clostridium_phytofermentans
Thermotoga_maritima
Fusobacterium_mortiferum
Clostridium_difficile
Bacteroides_uniformis
Bacteroides_fragilis
Cyllamyces_aberensis
Orpinomyces_sp_ukk1
Piromyces_sp_E2
Physcomitrella_patens
Arabidopsis_thaliana
Ciona_intestinalis

INTERNATIONAL SEARCH REPORT

International application No
PCT/NL2009/050803

A. CLASSIFICATION OF SUBJECT MATTER
INV. C12N1/14 C12N1/16 C12N1/18 C12N1/19 C12P7/08
C12P7/10 C12P7/16 C12P7/42 C12P7/52 C12P7/54
C12P7/56 C12P7/48

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, WPI Data, MEDLINE, Sequence Search

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 2004/099381 A2 (CARGILL DOW LLC [US]; RAJGARHIA VINEET [US]; KOIVURANTA KARI [FI]; PEN) 18 November 2004 (2004-11-18) cited in the application page 30 - page 123; claim 13; sequence 152	1-15
X	WO 2004/044129 A2 (DIVERSA CORP [US]; CALLEN WALTER [US]) 27 May 2004 (2004-05-27) page 44 - page 52; sequence 6	1-2,4-7, 13-15

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

☒ See patent family annex.

* Special categories of cited documents :

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

"E" earlier 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
23 February 2010	09/06/2010

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL – 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Stoyanov, Borislav

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

Continuation of Box II.2

Claims Nos.:

1. The expression "one or amino acid sequence elements" in claim 1 is completely unclear and leaves the skilled reader in doubt how to interpret the claim in relation of the combination of elements (a)-(h). Correspondingly, the claim has been interpreted as reading "one or more amino acid sequence elements" in accordance with the specification page 11, line 10.

2. In present claim 1 no reference sequence has been indicated in relation to which the skilled reader will determine the positions of the amino acid elements (a) to (h). Correspondingly, the claim has been interpreted in view of the specification, page 13, lines 17-22 and the reference sequence was assumed to be SEQ ID NO. 3.

The applicant's attention is drawn to the fact that claims relating to inventions in respect of which no international search report has been established need not be the subject of an international preliminary examination (Rule 66.1(e) PCT). The applicant is advised that the EPO policy when acting as an International Preliminary Examining Authority is normally not to carry out a preliminary examination on matter which has not been searched. This is the case irrespective of whether or not the claims are amended following receipt of the search report or during any Chapter II procedure. If the application proceeds into the regional phase before the EPO, the applicant is reminded that a search may be carried out during examination before the EPO (see EPO Guideline C-VI, 8.2), should the problems which led to the Article 17(2)PCT declaration be overcome.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/NL2009/050803

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☒ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
see FURTHER INFORMATION sheet PCT/ISA/210

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see additional sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying an additional fees, this Authority did not invite payment of additional fees.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:
see additional sheet(s)

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.

☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.

☐ No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

Invention: 1; Claims: 1-15(partially)

A eukaryotic microbial cell comprising a nucleotide sequence encoding a polypeptide with xylose isomerase activity said polypeptide comprising the element as in claim 1(a): a methionine residue at position 91.
(Reference sequence for amino acid numbering is SEQ ID NO. 3 as indicated on page 13, lines 17-22 of the specification).

Inventions: 2-254; Claims: 1-15(partially)

A eukaryotic microbial cell comprising a nucleotide sequence encoding a polypeptide with xylose isomerase activity said polypeptide comprising elements as in claim 1(a)-1(h) in all possible combinations different from element 1(a) alone.
(Reference sequence for amino acid numbering is SEQ ID NO. 3 as indicated on page 13, lines 17-22 of the specification).

INTERNATIONAL SEARCH REPORT

International application No
PCT/NL2009/050803

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2004099381 A2	18-11-2004	AU 2004236724 A1	18-11-2004
		BR PI0409954 A	25-04-2006
		CA 2524387 A1	18-11-2004
		CN 1922198 A	28-02-2007
		EP 1626979 A2	22-02-2006
		JP 2006525029 T	09-11-2006
		ZA 200509697 A	28-03-2007
WO 2004044129 A2	27-05-2004	AU 2003284983 A1	03-06-2004
		BR 0316073 A	27-09-2005
		CA 2504909 A1	27-05-2004
		EP 1578937 A2	28-09-2005
		MX PA05004869 A	19-08-2005
		US 2006281908 A1	14-12-2006