Office de la Propriété Intellectuelle du Canada
Un organisme d'Industrie Canada

Canadian Intellectual Property Office
An agency of Industry Canada

CA 2900465 A1 2014/10/09

(21) 2 900 465

(12) DEMANDE DE BREVET CANADIEN
CANADIAN PATENT APPLICATION

(13) A1

(86) **Date de dépôt PCT/PCT Filing Date:** 2014/03/10
(87) **Date publication PCT/PCT Publication Date:** 2014/10/09
(85) **Entrée phase nationale/National Entry:** 2015/08/05
(86) **N° demande PCT/PCT Application No.:** US 2014/022358
(87) **N° publication PCT/PCT Publication No.:** 2014/164410
(30) **Priorité/Priority:** 2013/03/11 (US13/792,321)

(51) **Cl.Int./Int.Cl.** ***C12N 9/92*** (2006.01),
C12P 7/10 (2006.01), ***C12P 7/16*** (2006.01),
C12P 7/18 (2006.01)

(71) **Demandeur/Applicant:**
E. I. DU PONT DE NEMOURS AND COMPANY, US

(72) **Inventeurs/Inventors:**
CHEN, ZHONGQIANG, US;
KELLY, KRISTEN J., US;
YE, RICK W., US

(74) **Agent:** TORYS LLP

(54) **Titre : XYLOSE-ISOMERASES BACTERIENS ACTIFS DANS LES CELLULES DE LEVURE**
(54) **Title: BACTERIAL XYLOSE ISOMERASES ACTIVE IN YEAST CELLS**

(57) **Abrégé/Abstract:**
Specific polypeptides were identified as bacterial xylose isomerases that are able to provide xylose isomerase activity in yeast cells. The xylose isomerase activity can complete a xylose utilization pathway so that yeast can use xylose in fermentation, such as xylose in biomass hydrolysate.

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
9 October 2014 (09.10.2014)

WIPO | PCT

(10) International Publication Number
WO 2014/164410 A1

(51) International Patent Classification:
C12N 9/92 (2006.01) *C12P 7/16* (2006.01)
C12P 7/10 (2006.01) *C12P 7/18* (2006.01)

(21) International Application Number:
PCT/US2014/022358

(22) International Filing Date:
10 March 2014 (10.03.2014)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
13/792,321 11 March 2013 (11.03.2013) US

(71) Applicant: **E. I. DU PONT DE NEMOURS AND COMPANY** [US/US]; 1007 Market Street, Wilmington, Delaware 19898 (US).

(72) Inventors: **CHEN, Zhongqiang**; 115 Congressional Drive, Apt. D, Wilmington, Delaware 19807 (US). **KELLY, Kristen, J.**; 2510 Foulk Woods Road, Wilmington, Delaware 19810 (US). **YE, Rick, W.**; 27 Piersons Ridge, Hockessin, Delaware 19707 (US).

(74) Agent: **FELTHAM, S., Neil**; E. I. du Pont de Nemours and Company, Legal Patent Records Center, Chestnut Run Plaza 721/2340, 974 Centre Road, PO Box 2915 Wilmington, Delaware 19805 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, 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, IR, IS, JP, KE, KG, 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, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

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

Published:

— *with international search report (Art. 21(3))*

— *with sequence listing part of description (Rule 5.2(a))*

WO 2014/164410 A1

(54) Title: BACTERIAL XYLOSE ISOMERASES ACTIVE IN YEAST CELLS

(57) Abstract: Specific polypeptides were identified as bacterial xylose isomerases that are able to provide xylose isomerase activity in yeast cells. The xylose isomerase activity can complete a xylose utilization pathway so that yeast can use xylose in fermentation, such as xylose in biomass hydrolysate.

TITLE

BACTERIAL XYLOSE ISOMERASES ACTIVE IN YEAST CELLS

This application claims the benefit of United States National Application 13/792321, filed March 11, 2013, which is incorporated by reference in its entirety.

FIELD OF THE INVENTION

The invention relates to the field of genetic engineering of yeast. More specifically, a group of xylose isomerases are identified that are active in yeast cells engineered for their expression.

BACKGROUND OF THE INVENTION

Currently yeasts are the organism or choice for the fermentative production of ethanol. Most common is the use of *Saccharomyces cerevisiae*, in processes using hexoses obtained from grains or mash as the carbohydrate source. Use of hydrolysate prepared from cellulosic biomass as a carbohydrate source for fermentation is desirable, as this is a readily renewable resource that does not compete with the food supply. After glucose, the second most abundant sugar in cellulosic biomass is xylose, a pentose. *Saccharomyces cerevisiae* is not naturally capable of metabolizing xylose, but can be engineered to metabolize xylose with expression of xylose isomerase activity to convert xylose to xylulose, and additional pathway engineering.

Success in expressing heterologous bacterial xylose isomerase enzymes that are active in yeast has been limited. Some specific xylose bacterial isomerase sequences have been reported to provide xylose isomerase activity for a xylose utilization pathway in yeast. For example as US 7,622,284 discloses a yeast cell expressing a xylose isomerase from *Piromyces* sp. US 2012/0184020 dislcoses eukaryotic cells expressing a xylose isomerase isloagted from *Ruminococcus flavefaciens*. Simlilarly WO2011078262 disclose several xylose isomerases from each of *Reticulitermes speratus* and *Mastotermes darwiniensis* and proteins with

high sequence identities to these, and their expression in eukaryotic cells. WO212009272 discloses constructs and fungal cells containing a xylose isomerase from *Abiotrophia defectiva* and others with sequence identity to it.

There remains a need for additional engineered yeast cells that express xylose isomerase activity for successful utilization of xylose, thereby allowing effective use of sugars obtained from cellulosic biomass during fermentation.

SUMMARY OF THE INVENTION

The invention provides recombinant yeast cells that are engineered to express a polypeptide that provides xylose isomerase activity.

Accordingly, the invention provides a recombinant yeast cell comprising a heterologous nucleic acid molecule encoding a polypeptide having xylose isomerase activity and amino acid sequence with at least about 85% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NOs:1, 3, 5, and 7.

In another aspect the invention provides a method for producing a yeast cell that has xylose isomerase activity comprising:

a) providing a yeast cell;

b) introducing a heterologous nucleic acid molecule encoding a polypeptide having xylose isomerase activity and amino acid sequence with greater than 95% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NOs:1, 3, 5, and 7;

wherein a yeast cell having xylose isomerase activity is produced.

SEQUENCE DESCRIPTIONS

The invention can be more fully understood from the following detailed description and the accompanying sequence descriptions which form a part of this application.

The following sequences conform with 37 C.F.R. 1.821-1.825 ("Requirements for Patent Applications Containing Nucleotide Sequences

and/or Amino Acid Sequence Disclosures - the Sequence Rules") and are consistent with World Intellectual Property Organization (WIPO) Standard ST.25 (2009) and the sequence listing requirements of the EPO and PCT (Rules 5.2 and 49.5(a-bis), and Section 208 and Annex C of the Administrative Instructions). The symbols and format used for nucleotide and amino acid sequence data comply with the rules set forth in 37 C.F.R. §1.822.

Table 1 SEQ ID NOs for xylose isomerase polypeptides, and coding regions that are codon optimized for expression in *S. cerevisiae*

Strain	SEQ ID NO: amino acid	SEQ ID NO: nucleotide codon opt.
Lachnospiraceae bacterium ICM7	1	2
Lachnospiraceae bacterium oral taxon 107 str. F0167	3	4
Lachnospiraceae bacterium oral taxon 082 str. F0431	5	6
Eubacterium saburreum DSM 3986	7	8
Ruminococcus champanellensis 18P13	9	10
Ruminococcus flavefaciens FD-1	11	*nd
Abiotrophis defectiva	12	*nd
Leptotrichia goodfellowii F0264	13	14
Sebaldella termitidis ATCC 33386	15	16

*nd = not designed

SEQ ID NO:17 is the nucleotide sequence of the pHR81 vector containing the ILVp-xylA(Hm1)-ILV5t chimeric gene.

SEQ ID NO:18 is the nucleotide sequence of P5 Integration Vector.

SEQ ID NO:19 is the nucleotide sequence of a URA3 deletion scar.

SEQ ID NO:20 is the nucleotide sequence of the upstream ura3Δ post deletion region.

SEQ ID NO:21 is the nucleotide sequence of the downstream ura3Δ post deletion region.

SEQ ID NO:22 is the nucleotide sequence of the upstream his3Δ post deletion region.

SEQ ID NO:23 is the nucleotide sequence of the downstream his3Δ post deletion region.

SEQ ID NO:24 is the nucleotide sequence of pJT254.

DETAILED DESCRIPTION

The following definitions may be used for the interpretation of the claims and specification:

As used herein, the terms “comprises,” “comprising,” “includes,” “including,” “has,” “having,” “contains” or “containing,” or any other variation thereof, are intended to cover a non-exclusive inclusion. For example, a composition, a mixture, process, method, article, or apparatus that comprises a list of elements is not necessarily limited to only those elements but may include other elements not expressly listed or inherent to such composition, mixture, process, method, article, or apparatus. Further, unless expressly stated to the contrary, “or” refers to an inclusive or and not to an exclusive or. For example, a condition A or B is satisfied by any one of the following: A is true (or present) and B is false (or not present), A is false (or not present) and B is true (or present), and both A and B are true (or present).

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

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

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

The term “xylose isomerase” refers to an enzyme that catalyzes the interconversion of D-xylose and D-xylulose. Xylose isomerases (XI) belong to the group of enzymes classified as EC 5.3.1.5.

The terms “xylose utilization pathway” refers to a metabolic pathway comprising genes encoding enzymes sufficient to convert xylose to a target chemical. In the situation where the target chemical is ethanol such a pathway typically comprises genes encoding the following enzymes: xylulokinase (XKS1) , transaldolase (TAL1), transketolase 1 (TKL1), D-ribulose -5-phosphate 3-epimerase (RPE1), and ribose 5-phosphate ketol-isomerase (RKI1). Elements of this pathway may be native or heterologous to the host cell.

The term "gene" refers to a nucleic acid fragment that expresses a specific protein or functional RNA molecule, which may optionally include regulatory sequences preceding (5' non-coding sequences) and following (3' non-coding sequences) the coding sequence. "Native gene" or "wild type gene" refers to a gene as found in nature with its own regulatory

sequences. "Chimeric gene" refers to any gene that is not a native gene, comprising regulatory and coding sequences that are not found together in nature. Accordingly, a chimeric gene may comprise regulatory sequences and coding sequences that are derived from different sources, or regulatory sequences and coding sequences derived from the same source, but arranged in a manner different than that found in nature. "Endogenous gene" refers to a native gene in its natural location in the genome of an organism. A "foreign" gene refers to a gene not normally found in the host organism, but that is introduced into the host organism by gene transfer. Foreign genes can comprise native genes inserted into a non-native organism, or chimeric genes.

The term "promoter" or "Initiation control regions" refers to a DNA sequence capable of controlling the expression of a coding sequence or functional RNA. In general, a coding sequence is located 3' to a promoter sequence. Promoters may be derived in their entirety from a native gene, or be composed of different elements derived from different promoters found in nature, or even comprise synthetic DNA segments. It is understood by those skilled in the art that different promoters may direct the expression of a gene in different tissues or cell types, or at different stages of development, or in response to different environmental conditions. Promoters which cause a gene to be expressed in most cell types at most times are commonly referred to as "constitutive promoters".

The term "expression", as used herein, refers to the transcription and stable accumulation of coding (mRNA) or functional RNA derived from a gene. Expression may also refer to translation of mRNA into a polypeptide. “Overexpression” refers to the production of a gene product in transgenic organisms that exceeds levels of production in normal or non-transformed organisms.

The term "transformation" as used herein, refers to the transfer of a nucleic acid fragment into a host organism, resulting in genetically stable inheritance. The transferred nucleic acid may be in the form of a plasmid maintained in the host cell, or some transferred nucleic acid may be integrated into the genome of the host cell. Host organisms containing the

transformed nucleic acid fragments are referred to as "transgenic" or "recombinant" or "transformed" organisms.

The terms "plasmid" and "vector" as used herein, refer to an extra chromosomal element often carrying genes which are not part of the central metabolism of the cell, and usually in the form of circular double-stranded DNA molecules. Such elements may be autonomously replicating sequences, genome integrating sequences, phage or nucleotide sequences, linear or circular, of a single- or double-stranded DNA or RNA, derived from any source, in which a number of nucleotide sequences have been joined or recombined into a unique construction which is capable of introducing a promoter fragment and DNA sequence for a selected gene product along with appropriate 3' untranslated sequence into a cell.

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

The term "selectable marker" means an identifying factor, usually an antibiotic or chemical resistance gene, that is able to be selected for based upon the marker gene's effect, i.e., resistance to an antibiotic, wherein the effect is used to track the inheritance of a nucleic acid of interest and/or to identify a cell or organism that has inherited the nucleic acid of interest.

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

approaches the frequency of preferred codon usage of the host cell.

The term “codon-optimized” as it refers to genes or coding regions of nucleic acid molecules for transformation of various hosts, refers to the alteration of codons in the gene or coding regions of the nucleic acid molecules to reflect the typical codon usage of the host organism without altering the polypeptide encoded by the DNA.

The term “carbon substrate” or “fermentable carbon substrate” refers to a carbon source capable of being metabolized by microorganisms. A type of carbon substrate is “fermentable sugars” which refers to oligosaccharides and monosaccharides that can be used as a carbon source by a microorganism in a fermentation process.

The term “lignocellulosic” refers to a composition comprising both lignin and cellulose. Lignocellulosic material may also comprise hemicellulose.

The term “cellulosic” refers to a composition comprising cellulose and additional components, which may include hemicellulose and lignin.

The term “saccharification” refers to the production of fermentable sugars from polysaccharides.

The term “pretreated biomass” means biomass that has been subjected to thermal, physical and/or chemical pretreatment to increase the availability of polysaccharides in the biomass to saccharification enzymes.

“Biomass” refers to any cellulosic or lignocellulosic material and includes materials comprising cellulose, and optionally further comprising hemicellulose, lignin, starch, oligosaccharides and/or monosaccharides. Biomass may also comprise additional components, such as protein and/or lipid. Biomass may be derived from a single source, or biomass can comprise a mixture derived from more than one source; for example, biomass could comprise a mixture of corn cobs and corn stover, or a mixture of grass and leaves. Biomass includes, but is not limited to, bioenergy crops, agricultural residues, municipal solid waste, industrial solid waste, sludge from paper manufacture, yard waste, wood and forestry waste. Examples of biomass include, but are not limited to, corn

cobs, crop residues such as corn husks, corn stover, corn grain fiber, grasses, beet pulp, wheat straw, wheat chaff, oat straw, barley straw, barley hulls, hay, rice straw, rice hulls, switchgrass, miscanthus, cord grass, reed canary grass, waste paper, sugar cane bagasse, sorghum bagasse, sorghum stover, soybean stover, components obtained from milling of grains, trees, branches, roots, leaves, wood chips, sawdust, palm waste, shrubs and bushes, vegetables, fruits, flowers, and animal manure.

"Biomass hydrolysate" refers to the product resulting from saccharification of biomass. The biomass may also be pretreated or pre-processed prior to saccharification.

The term "heterologous" means not naturally found in the location of interest. For example, a heterologous gene refers to a gene that is not naturally found in the host organism, but that is introduced into the host organism by gene transfer. For example, a heterologous nucleic acid molecule that is present in a chimeric gene is a nucleic acid molecule that is not naturally found associated with the other segments of the chimeric gene, such as the nucleic acid molecules having the coding region and promoter segments not naturally being associated with each other.

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

The term "target compound" or "target chemical" refers to a compound made by a microorganism via an endogenous or recombinant biosynthetic pathway which is able to metabolize a fermentable carbon source to produce the target compound.

The term "percent identity", as known in the art, is a relationship between two or more polypeptide sequences or two or more polynucleotide sequences, as determined by comparing the sequences. In the art, "identity" also means the degree of sequence relatedness between polypeptide or polynucleotide sequences, as the case may be, as

determined by the match between strings of such sequences. "Identity" and "similarity" can be readily calculated by known methods, including but not limited to those described in: 1.) Computational Molecular Biology (Lesk, A. M., Ed.) Oxford University: NY (1988); 2.) Biocomputing: Informatics and Genome Projects (Smith, D. W., Ed.) Academic: NY (1993); 3.) Computer Analysis of Sequence Data, Part I (Griffin, A. M., and Griffin, H. G., Eds.) Humania: NJ (1994); 4.) Sequence Analysis in Molecular Biology (von Heinje, G., Ed.) Academic (1987); and 5.) Sequence Analysis Primer (Gribskov, M. and Devereux, J., Eds.) Stockton: NY (1991).

Preferred methods to determine identity are designed to give the best match between the sequences tested. Methods to determine identity and similarity are codified in publicly available computer programs. Sequence alignments and percent identity calculations may be performed using the MegAlign program of the LASERGENE bioinformatics computing suite (DNASTAR Inc., Madison, WI).

Multiple alignment of the sequences is performed using the "Clustal method of alignment" which encompasses several varieties of the algorithm including the "Clustal V method of alignment" corresponding to the alignment method labeled Clustal V (described by Higgins and Sharp, *CABIOS*. 5:151-153 (1989); Higgins, D.G. et al., *Comput. Appl. Biosci.,* 8:189-191 (1992)) and found in the MegAlign v8.0 program of the LASERGENE bioinformatics computing suite (DNASTAR Inc.). For multiple alignments, the default values correspond to GAP PENALTY=10 and GAP LENGTH PENALTY=10. Default parameters for pairwise alignments and calculation of percent identity of protein sequences using the Clustal method are KTUPLE=1, GAP PENALTY=3, WINDOW=5 and DIAGONALS SAVED=5. For nucleic acids these parameters are KTUPLE=2, GAP PENALTY=5, WINDOW=4 and DIAGONALS SAVED=4. After alignment of the sequences using the Clustal V program, it is possible to obtain a "percent identity" by viewing the "sequence distances" table in the same program.

Additionally the “Clustal W method of alignment” is available and corresponds to the alignment method labeled Clustal W (described by Higgins and Sharp, *CABIOS*. 5:151-153 (1989); Higgins, D.G. et al., *Comput. Appl. Biosci.* 8:189-191(1992); Thompson, J.D. et al, Nucleic Acid Research, 22 (22): 4673-4680, 1994) and found in the MegAlign v8.0 program of the LASERGENE bioinformatics computing suite (DNASTAR Inc.). Default parameters for multiple alignment (stated as protein/nucleic acid (GAP PENALTY=10/15, GAP LENGTH PENALTY=0.2/6.66, Delay Divergen Seqs(%)=30/30, DNA Transition Weight=0.5, Protein Weight Matrix=Gonnet Series, DNA Weight Matrix=IUB). After alignment of the sequences using the Clustal W program, it is possible to obtain a “percent identity” by viewing the “sequence distances” table in the same program. Sequence identities referred to herein shall always be considered to have been determined according to the parameters set forth above unless otherwise noted.

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

Standard recombinant DNA and molecular cloning techniques used herein are well known in the art and are described by Sambrook, J. and Russell, D., Molecular Cloning: A Laboratory Manual, Third Edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY (2001); and by Silhavy, T. J., Bennan, M. L. and Enquist, L. W., Experiments with Gene Fusions, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY (1984); and by Ausubel, F. M. et. al., Short Protocols in Molecular Biology, 5th Ed. Current Protocols, John Wiley and Sons, Inc., N.Y., 2002. Additional methods used here are in *Methods in Enzymology*, Volume 194, *Guide to Yeast Genetics and Molecular and Cell Biology* (Part A, 2004, Christine Guthrie and Gerald R. Fink (Eds.), Elsevier Academic Press, San Diego, CA).

The present invention relates to engineered yeast strains that have xylose isomerase enzyme activity. A challenge for engineering yeast to utilize xylose, which is the second most predominant sugar obtained from cellulosic biomass, is to produce sufficient xylose isomerase activity in the yeast cell. Xylose isomerase catalyzes the conversion of xylose to xylulose, which is the first step in a xylose utilization pathway. Applicants have found that expression of specific xylose isomerase polypeptides provides xylose isomerase activity in the yeast cell, while expression of other xylose isomerase polypeptides does not provide activity. A yeast cell expressing xylose isomerase activity provides a host cell for expression of a complete xylose utilization pathway, thereby engineering a yeast cell that can produce a target compound, such as ethanol, butanol, or 1,3-propanediol, using xylose derived from lignocellulosic biomass as a carbon source.

Yeast Host Cells

Yeast cells of the invention are those that comprise a functional bacterial xylose isomerase and a capable of the production of a target compound. Preferred target compounds are those of commercial value including but not limited to ethanol, butanol, or 1,3-propanediol.

Any yeast cells that either produce a target chemical, or can be engineered to produce a target chemical, may be used as host cells

herein. Examples of such yeasts include, but are not limited to, yeasts of the genera *Kluyveromyces*, *Candida*, *Pichia*, *Hansenula*, *Schizosaccharomyces*, *Kloeckera*, *Schwammiomyces*, *Yarrowia*, and *Saccharomyces*.

Yeast cells of the invention comprising an active bacterial xylose isomerase may be engineered according to methods well known in the art. For example yeast cell that have the native ability to produce ethanol from C6 sugars may be transferred with genes comprising C5 metabolic pathways including the bacterial xylose isomerase of the invention. Such cells may be capable of either aerobic or anaerobic fermentive ethanol production.

In other embodiments yeast cells may be engineered to express a pathway for synthesis of butanol or 1,3-propanediol. Engineering of pathways for butanol synthesis (including isobutanol, 1-butanol, and 2-butanol) have been disclosed, for example in US 8,206,970, US 20070292927, US 20090155870, US 7,851,188, and US 20080182308, which are incorporated herein by reference. Engineering of pathways for 1,3-propanediol have been disclosed in US 6,514,733, US 5,686,276, US 7,005,291, US 6,013,494, and US 7,629,151, which are incorporated herein by reference.

For utilization of xylose as a carbon source, a yeast cell is engineered for expression of a complete xylose utilization pathway. Engineering of yeast such as *S. cerevisiae* for production of ethanol from xylose is described in Matsushika et al. (Appl. Microbiol. Biotechnol. (2009) 84:37-53) and in Kuyper et al. (FEMS Yeast Res. (2005) 5:399-409). In one embodiment, in addition to engineering a yeast cell as disclosed herein to have xylose isomerase activity, the activities of other pathway enzymes are increased in the cell to provide the ability to grow on xylose as a sole carbon source. Typically the activity levels of five pentose pathway enzymes are increased: xylulokinase (XKS1) , transaldolase (TAL1), transketolase 1 (TKL1), D-ribulose -5-phosphate 3-epimerase (RPE1), and ribose 5-phosphate ketol-isomerase (RKI1). Any method known to one skilled in the art for increasing expression of a gene may be

used. For example, as described herein in Example 1, these activities may be increased by expressing the host coding region for each protein using a highly active promoter. Chimeric genes for expression are constructed and are integrated into the yeast genome. Alternatively, heterologous coding regions for these enzymes may be expressed in the yeast cell to obtain increased enzyme activities. For additional methods for engineering yeast capable of metabolizing xylose see for example US7622284B2, US8058040B2, US 7,943,366 B2, WO2011153516A2, WO2011149353A1, WO2011079388A1, US20100112658A1, US20100028975A1, US20090061502A1, US20070155000A1, WO2006115455A1, US20060216804A1 and US8129171B2

In one embodiment the present yeast cell has xylose isomerase activity as described below, and additional genetic engineering to provide a complete xylose utilization pathway as described above. These cells are able to grow in medium containing xylose as the sole carbon source. More typically, these cells are grown in medium containing xylose as well as other sugars such as glucose and arabinose. This allows effective use of the sugars found in a hydrolysate medium that is prepared from cellulosic biomass by pretreatment and saccharification.

Xylose Isomerase

Expression of xylose isomerases in yeast cells has been problematic; in particular, many bacterial xylose isomerases have been found to have little to no activity when expressed in yeast cells. In the present recombinant yeast cell, xylose isomerase activity is provided by expression of a heterologous nucleic acid molecule encoding a polypeptide having an amino acid sequence with at least about 85% sequence identity to an amino acid sequence of SEQ ID NO:1, 3, 5, or 7. These sequences were identified by BLAST searching of the GenBank database (National Center for Biotechnology Information (NCBI); Benson et al. Nucleic Acids Research , 2011 Jan; 39 (Database issue):D32-7) using xylose isomerase sequences from *Ruminococcus flavefaciens* FD-1 (SEQ ID NO:11) and from *Ruminococcus champanellensis* 18P13 (SEQ

ID NO:9). SEQ ID NO:11 is identical to the *Ruminococcus flavefaciens* xylose isomerase of SEQ ID NO:31 in US 2012/0184020.

SEQ ID NOs:1, 3, 5, and 7 are the amino acid sequences of bacterial xylose isomerases from *Lachnospiraceae bacterium* ICM7 (called herein Hm1), *Lachnospiraceae bacterium oral taxon* 107 str. F0167 (called herein Hm2), *Lachnospiraceae bacterium oral taxon* 082 str. F0431 (called herein Hm3), and *Eubacterium saburreum* DSM 3986 (called herein Hm4), respectively. The identities of these four sequences to the *Ruminococcus flavefaciens* FD-1 (SEQ ID NO:11) and *Ruminococcus champanellensis* 18P13 (SEQ ID NO:9) sequences are between 60.9% and 62.6% as given in Table 2. The identities of these four sequences to a hypothetical protein from *Abiotrophis defectiva* ATCC 49176 (SEQ ID NO:12; Accession #ZP 04453767), which is identical to SEQ ID NO:2 of WO 2102/009272 and is identified therein as *Abiotrophia defectiva* xylose isomerase, are between 71.7% and 73.2% as given in Table 2.

Expression of a nucleic acid molecules encoding Hm1, Hm2, Hm3, and Hm4 in *S. cerevisiae* was found herein (Example 3) to allow growth in medium containing xylose as the sole sugar, of a *S. cerevisiae* strain containing a xylose utilization pathway but lacking xylose isomerase activity. Xylose was utilized and ethanol was produced by the yeast cells. Thus expression of each of HM1, Hm2, Hm3, and Hm4 provided xylose isomerase activity to complete the xylose utilization pathway in the yeast cells. Among Hm1, Hm2, Hm3, and Hm4 the sequence identities are in the range of 92.2% to 95.7% as given in Table 2.

Any polypeptide having xylose isomerase activity and having at least about 85% identity to any of SEQ ID NO:1, 3, 5, and 7 may be expressed in the present yeast cell. In various embodiments the polypeptide may have amino acid sequence identity of about 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or up to 100% to any of SEQ ID NO:1, 3, 5, and 7.

When transformed with the xylose isomerase of the invention a *S. cerevisiae* demonstrated increase growth, xylose utilization and ethanol yield when grown in xylose containing medium. Xylose isomerase

proteins, having as much as 95% identity to SEQ ID NO:1, 3, 5, and 7 did not have the same effect, suggesting that the ability of the enzyme to be active in a yeast host may not be sequence dependent. Specifically, sequences from *Leptotrichia goodfellowii* F0264 (called herein Oral-2; SEQ ID NO:13) and *Sebaldella termitidis* ATCC 33386 (called herein Term-1; SEQ ID NO:15) have similar sequence identities to the *Ruminococcus flavefaciens* FD-1 (SEQ ID NO:11) and *Ruminococcus champanellensis* 18P13 (SEQ ID NO:9) xylose isomerases (see Table 2) as Hm1, Hm2, Hm3, and Hm4, but did not provide xylose isomerase activity in yeast cells as tested in Example 3 herein.

Table 2 Comparison of xylose isomerase amino acid sequence identities

	Hm1	Hm2	Hm3	Hm4	Oral2	Term 1	R. f. XI	R. c. XI
Hm1								
Hm2	95.5							
Hm3	93.6	92.3						
Hm4	92.9	93.2	95.7					
Leptotrichia goodfellowii F0264 (Oral2)	57.4	57.2	56.1	57.0				
Sebaldella termitidis ATCC 33386 (Term1)	55.8	55.1	54.9	55.8	85.2			
XI from *R. flavefaciens*	62.2	61.5	60.9	60.9	61.7	59.9		
XI from *R. champanellensis*	62.6	61.9	61.8	62.2	60.7	61.2	77.4	
XI from *A. defectiva*	73.2	72.5	71.7	72.1	57.4	54.9	61.9	61.0

The present amino acid sequences that provide xylose isomerase activity in yeast cells are not native to yeast cells, thus their encoding

nucleic acid sequences are heterologous to yeast cells. For expression, nucleic acid molecules encoding the present polypeptides may be designed using codon optimization for the desired yeast cell, as is well known to one skilled in the art. For example, for expression of HM1, Hm3, Hm5, or Hm7 in *Saccharomyces cerevisiae*, nucleic acid molecules named xylA(Hm1) (SEQ ID NO:2), xylA(Hm2) (SEQ ID NO:4), xylA(Hm3) (SEQ ID NO:6), and xylA(Hm4) (SEQ ID NO:8) were designed using codon-optimization for expression *S. cerevisiae*.

Methods for gene expression in yeasts are known in the art (see for example *Methods in Enzymology*, Volume 194, *Guide to Yeast Genetics and Molecular and Cell Biology* (Part A, 2004, Christine Guthrie and Gerald R. Fink (Eds.), Elsevier Academic Press, San Diego, CA). Expression of genes in yeast typically requires a promoter, operably linked to the coding region of interest, and a transcriptional terminator. A number of yeast promoters can be used in constructing expression cassettes for genes encoding the desired proteins, including, but not limited to constitutive promoters FBA1, GPD1, ADH1, GPM, TPI1, TDH3, PGK1, ILV5p, and the inducible promoters GAL1, GAL10, and CUP1. Suitable transcription terminators include, but are not limited to FBAt, GPDt, GPMt, ERG10t, GAL1t, CYC1t, ADH1t, TAL1t, TKL1t, ILV5t, and ADHt.

Suitable promoters, transcriptional terminators, and coding regions may be cloned into *E. coli*-yeast shuttle vectors, and transformed into yeast cells. These vectors allow strain propagation in both *E. coli* and yeast strains.

Typically the vector contains a selectable marker and sequences allowing autonomous replication or chromosomal integration in the desired host. Typically used plasmids in yeast are shuttle vectors pRS423, pRS424, pRS425, and pRS426 (American Type Culture Collection, Rockville, MD), which contain an *E. coli* replication origin (e.g., pMB1), a yeast 2μ origin of replication, and a marker for nutritional selection. The selection markers for these four vectors are His3 (vector pRS423), Trp1 (vector pRS424), Leu2 (vector pRS425) and Ura3 (vector pRS426). Additional vectors that may be used include pHR81 (ATCC #87541) and

pRS313 (ATCC #77142). Construction of expression vectors with chimeric genes encoding the desired proteins may be performed by either standard molecular cloning techniques in *E. coli* or by the gap repair recombination method in yeast.

The gap repair cloning approach takes advantage of the highly efficient homologous recombination in yeast. Typically, a yeast vector DNA is digested (e.g., in its multiple cloning site) to create a "gap" in its sequence. The "gapped" vector and insert DNAs having sequentially overlapping ends (overlapping with each other and with the gapped vector ends, in the desired order of inserts) are then co-transformed into yeast cells which are plated on the medium containing the appropriate compound mixtures that allow complementation of the nutritional selection markers on the plasmids. The presence of correct insert combinations can be confirmed by PCR mapping using plasmid DNA prepared from the selected cells. The plasmid DNA isolated from yeast can then be transformed into an *E. coli* strain, e.g. *TOP10*, followed by mini preps and restriction mapping to further verify the plasmid construct. Finally the construct can be verified by sequence analysis.

Like the gap repair technique, integration into the yeast genome also takes advantage of the homologous recombination system in yeast. Typically, a cassette containing a coding region plus control elements (promoter and terminator) and auxotrophic marker is PCR-amplified with a high-fidelity DNA polymerase using primers that hybridize to the cassette and contain 40-70 base pairs of sequence homology to the regions 5' and 3' of the genomic area where insertion is desired. The PCR product is then transformed into yeast cells which are plated on medium containing the appropriate compound mixtures that allow selection for the integrated auxotrophic marker. Transformants can be verified either by colony PCR or by direct sequencing of chromosomal DNA.

The present invention provides a method for producing a yeast cell that has xylose isomerase activity following the teachings above. In one embodiment a heterologous nucleic acid molecule encoding a polypeptide having xylose isomerase activity and amino acid sequence with at least

85% sequence identity to any of the amino acid sequences of SEQ ID NO:1, 3, 5, or 7 is introduced into a yeast strain. In various embodiments the amino acid sequence of the polypeptide has at least about about 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or up to 100% to any of SEQ ID NO:1, 3, 5, and 7.

In one embodiment the introduced nucleic acid molecule is a part of a chimeric gene that is introduced into a yeast cell for expression, as described above.

In one embodiment the described nucleic acid molecule is introduced into a yeast cell which has other genetic modifications providing a complete xylose utilization pathway, once the xylose isomerase activity is introduced, as described above for the yeast host cell. Introduction of xylose isomerase activity and the additional genetic modifications may be performed in any order, and/or with two or more of introduction/ modification performed concurrently. These cells are able to grow in medium containing xylose as the sole carbon source. More typically, these cells are grown in medium containing xylose as well as other sugars such as glucose and arabinose. This allows effective use of the sugars found in a hydrolysate medium that is prepared from cellulosic biomass by pretreatment and saccharification.

In further embodiments the described nucleic acid molecule is introduced into a yeast cell which has a metabolic pathway that produces a target chemical. Introduction of xylose isomerase activity and the metabolic pathway may be performed in any order, and/or with two or more genetic modifications performed concurrently. Examples of target compounds include ethanol, butanol, and 1,3-propanediol. Yeast cells containing metabolic pathways for production of target chemicals are described above.

EXAMPLES

The present invention is further defined in the following Examples. It should be understood that these Examples, while indicating preferred embodiments of the invention, are given by way of illustration only. From

the above discussion and these Examples, one skilled in the art can ascertain the essential characteristics of this invention, and without departing from the spirit and scope thereof, can make various changes and modifications of the invention to adapt it to various uses and conditions.

GENERAL METHODS

The meaning of abbreviations is as follows: "kb" means kilobase(s), "bp" means base pairs, "nt" means nucleotide(s), "hr" means hour(s), "min" means minute(s), "sec" means second(s), "d" means day(s), "L" means liter(s), "ml" or "mL" means milliliter(s), "μL" means microliter(s), "μg" means microgram(s), "ng" means nanogram(s), "mg" means milligram(s), "mM" means millimolar, "μM" means micromolar, "nm" means nanometer(s), "μmol" means micromole(s), "pmol" means picomole(s), "XI" is xylose isomerase, "nt" means nucleotide.

Standard recombinant DNA and molecular cloning techniques used here are well known in the art and are described by Sambrook, J., Fritsch, E. F. and Maniatis, T., *Molecular Cloning: A Laboratory Manual*, 2nd ed., Cold Spring Harbor Laboratory: Cold Spring Harbor, NY (1989) (hereinafter "Maniatis"); and by Silhavy, T. J., Bennan, M. L. and Enquist, L. W., *Experiments with Gene Fusions*, Cold Spring Harbor Laboratory: Cold Spring Harbor, NY (1984); and by Ausubel, F. M. et al., *Current Protocols in Molecular Biology*, published by Greene Publishing Assoc. and Wiley-Interscience, Hoboken, NJ (1987), and by *Methods in Yeast Genetics*, 2005, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

HPLC analysis

Cell culture samples were taken at timed intervals and analyzed for EtOH and xylose using either a Waters HPLC system (Alliance system, Waters Corp., Milford, MA) or an Agilent 1100 Series LC; conditions = 0.6 mL/min of 0.01 N H_2SO_4, injection volume = 10 μL, autosampler temperature = 10ºC, column temperature = 65ºC, run time = 25 min, detection by refractive index (maintained at 40ºC). The HPLC column was purchased from BioRad (Aminex HPX-87H, BioRad Inc., Hercules, CA).

Analytes were quantified by refractive index detection and compared to known standards.

Example 1

Up-Regulation of the Native Pentose Pathway in *S. cerevisiae*

In addition to expression of an active xylose isomerase enzyme, a robust pentose pathway is necessary for efficient use of xylose and ethanol production under oxygen-limiting conditions in S. cerevisiae. The pentose pathway consists of five enzymes. In *S. cerevisiae*, these proteins are xylulokinase (XKS1), transaldolase (TAL1), transketolase 1 (TKL1), D-ribulose -5-phosphate 3-epimerase (RPE1), and ribose 5-phosphate ketol-isomerase (RKI1). In order to increase the expression of these proteins, their coding regions from the *S. cerevisiae* genome were cloned for expression under different promoters and integrated in the *S. cerevisiae* chromosome. The *GRE3* locus encoding aldose reductase was chosen for integration. To construct such this strain, the first step was the construction of an integration vector called P5 Integration Vector in GRE3.

The sequence of the P5 Integration Vector in GRE3 is given as SEQ ID NO:18, and the following numbers refer to nucleotide positions in this vector sequence. Gaps between the given nt numbers include sequence regions containing restriction sites. The TAL1 coding region (15210 to 16217) was expressed with the TPI1 promoter (14615 to 15197) and uses the TAL1t terminator. The RPE1 (13893 to 14609) coding region was expressed with the FBA1 promoter (13290 to 13879) and uses the terminator at the upstream end of the TPI1 promoter. RKI1 coding region (nt 11907 to 12680) was expressed with the TDH3 promoter (11229 to 11900) and uses the GPDt (previously called TDH3t) terminator. The TKL1 coding region (nt 8830 to 10872) was expressed with the PGK1 promoter (nt 8018 to 8817) and uses the TKL1t terminator. The XKS1 coding region (nt 7297 to 5495 to) was expressed with the Ilv5 promoter (nt 8009 to 7310) and uses the ADH terminator. In this integration vector, the URA3 marker (nt 332 to 1135) was flanked by loxP sites (nt 42 to 75

and nt 1513 to 1546) for recycling of the marker. The vector contains integration arms for the GRE3 locus (nt 1549 to 2089 and nt 4566 to 5137). This P5 Integration Vector in GRE3 can be linearized by digesting with the KasI enzyme before integration.

The yeast strain chosen for this study was BP1548 which is a haploid strain derived from prototrophic diploid strain CBS 8272 (Centraalbureau voor Schimmelcultures (CBS) Fungal Biodiversity Centre, Netherlands). This strain is in the CEN.PK lineage of *Saccharomyces cerevisiae* strains. BP1548 contains the MATα mating type and deletions of the URA3 and HIS3 genes.

To produce BP1548, first CBS 8272 was sporulated and a tetrad was dissected to yield four haploid strains using standard procedures (Amberg et al., Methods in Yeast Genetics, 2005). One of the MATα haploids, PNY0899, was selected for further modifications. The URA3 coding sequence (ATG through stop codon) and 130 bp of sequence upstream of the URA3 coding sequence was deleted by homologous recombination using a KanMX deletion cassette flanked by loxP sites, primer binding sites, and homologous sequences outside of the URA3 region to be deleted. After removal of the KanMX marker using the cre recombinase, a 95 bp sequence consisting of a loxP site flanked by the primer binding sites remained as a URA3 deletion scar in the genome (SEQ ID NO:19). This sequence is located in the genome between URA3 upstream sequence (SEQ ID NO:20) and URA3 downstream sequence (SEQ ID NO:21).The HIS3 coding sequence (ATG up to the stop codon) was deleted by homologous recombination using a scarless method. The deletion joins genomic sequences that were originally upstream (SEQ ID NO:22) and downstream (SEQ ID NO:23) of the HIS3 coding sequence. The KasI integration fragment containing all five pentose pathway genes in vector P5 Integration Vector in GRE3 was transformed into the BP1548 strain using the Frozen-EZ Yeast Transformation II Kit from Zymo Research (Irvine, CA). Transformants were selected on synthetic dropout (SD) medium lacking uracil. To recycle the URA3 marker, the CRE recombinase vector pJT254 (SEQ ID NO:24) was transformed into these

integrated strains. This vector was derived from pRS413 and the *cre* coding region (nt 2562 to 3593) was under the control of the *GAL1* promoter (nt 2119 to 2561). Strains that could no longer grow on SD (-uracil) medium were selected. Further passages on YPD medium was used to cure the plasmid pJT257. The resulting strain was designated as C52-79.

Example 2

Selection and Expression of Bacterial Xylose Isomerases

In order to identify candidate bacterial xylose isomerases that may be active when expressed in yeast, we used amino acid sequences of the xylose isomerases from *Ruminococcus flavefaciens* FD-1 (SEQ ID NO:11) and from *Ruminococcus champanellensis* 18P13 (SEQ ID NO:9) in a BLAST search against the GenBank database (National Center for Biotechnology Information (NCBI); Benson et al. Nucleic Acids Research , 2011 Jan; 39 (Database issue):D32-7). From this search, six bacterial xylose isomerases were chosen for testing based on sequence identity. These were the putative xylose isomerases from *Lachnospiraceae bacterium* ICM7 (SEQ ID NO:1), *Lachnospiraceae bacterium* oral taxon 107 str. F0167 (SEQ ID NO:3), *Lachnospiraceae bacterium* oral taxon 082 str. F0431 (SEQ ID NO:5), *Eubacterium saburreum* DSM 3986 (SEQ ID NO:7), *Leptotrichia goodfellowii* F0264 (SEQ ID NO:13), and *Sebaldella termitidis* ATCC 33386 (SEQ ID NO:15). DNA sequences encoding these proteins were synthesized using codon optimization for expression in *S. cerevisiae* and were designated as xylA(Hm1) (SEQ ID NO:2), xylA(Hm2) (SEQ ID NO:4), xylA(Hm3) (SEQ ID NO:6), xylA(Hm4) (SEQ ID NO:8), xylA(Oral-2) (SEQ ID NO:14), and xylA(Term1) (SEQ ID NO:X16), respectively. In addition, a codon-optimized coding region for the *Ruminococcus champanellensis* 18P13 xylose isomerase was synthesized and named xylA-10 (SEQ ID NO:10).

The synthesized xylA coding regions were expressed using a 1,184-nt promoter of the *S. cerevisiae* acetohydroxyacid reductoisomerase gene (ILV5p) and a 635-nt terminator of the *S.*

cerevisiae acetohydroxyacid reductoisomerase gene (ILV5t). The chimeric genes were located between NotI and XhoI sites in a pHR81-based shuttle vector, with the coding region between PmeI and SfiI sites. The pHR81 vector (ATCC #87541) contains a pMB1 origin and an ampicillin resistance (ampR) marker to allow plasmid propagation and selection, respectively, in *E. coli*. In addition, pHR81 has a 2 micron replication origin, a URA3 selection marker, and LEU 2-d for propagation and selection in yeast, which gives high copy number in *S. cerevisiae* when grown in medium lacking leucine, The sequence of the pHR81 vector containing the ILVp-xylA(Hm1)-ILV5t chimeric gene is SEQ ID NO:17. Vectors containing the other coding regions are identical with the exception of the substitution of each separate coding region between ILV5p and ILV5t, between PmeI and SfiI sites. The xylA(Hm1) vector was named pHR81 ilv5p xylA(Hm1), with other vectors having the same name, except substituting the specific xylA coding region designation. These constructs were transformed into the C52-79 strain (Example 1) and transformants were selected on plates containing synthetic glucose medium lacking uracil: 6.7 g/L yeast nitrogen base without amino acids (Amresco, Solon, OH), 0.77 g/L minus ura Drop Out supplement (Clontech Laboratories, Mountain View, CA), 20 g/L glucose. Transformants were then tested for growth and ethanol production.

Example 3

Growth and Ethanol Production in *S. cerevisiae* Containing Different Bacterial Xylose Isomerases

S. *cerevisiae* strain C52-79 (Example 1) lacks the ability to use xylose as the energy and carbon source since it lacks xylose isomerase activity. Yeast strains expressing xylA (Hm1), xylA (Hm2), xylA(Hm3), xylA(Hm4), xylA(Oral-2), xylA(Term1), and xylA-10 chimeric genes were tested in YPX medium (10 g/l yeast extract, 20 g/l peptone, and 40 g/l of xylose). To perform this test, strains were inoculated into 10 ml of YPX medium in 50 ml tissue culture tubes at a starting OD_{600} of 0.5. The lids were tightly closed and the tubes were placed in a 30 ^{o}C rotary shaker set

at a speed of 225 rpm. At different time intervals (24 hr, 44 hr, and 72 hr), samples were taken and the xylose and ethanol concentrations were determined by HPLC analysis as described in General Methods, as well as recording the OD_{600}. Three individual cultures for each strain were grown and analyzed. The results were averaged for each set of 3 replicates. Strains with xylA(Hm1), xylA(Oral-2), xylA(Term-1), and xylA-10 were assayed at the same time. Strains with xylA(Hm2) and xylA(Hm3) were assayed at the same time. The strain with xylA(Hm4) was assayed separately. All of the results are given in Table 3.

Table 3 Growth, xylose consumption, and ethanol production of yeast strains expressing various xylose isomerases

Vector in Strain	OD_{600}		Xylose consumed (g/L)		Ethanol Produced (g/L)	
	After 24 hours					
	Av.	SD	Av.	SD	Av.	SD
pHR81 ilv5p xylA(Hm1)	11.85	0.40	33.86	0.70	13.07	0.29
pHR81 ilv5p xylA(Hm2)	11.60	0.36	32.08	2.34	12.56	0.82
pHR81 ilv5p xylA(Hm3)	10.36	0.20	24.25	1.33	9.51	0.60
pHR81 ilv5p xylA(Hm4)	6.54	0.30	7.65	0.72	2.57	0.30
pHR81 ilv5p xylA(Oral-2)	2.88	0.09	0.52	0.08	0.00	0.00
pHR81 ilv5p xylA(Term-1)	2.35	0.56	0.47	0.07	0.00	0.00
pHR81 ilv5p xylA(xylA-10)	3.00	0.41	0.40	0.13	0.00	0.00
	After 44 hours					
pHR81 ilv5p xylA(Hm1)	12.79	0.51	40.00	0.00	15.82	0.11
pHR81 ilv5p xylA(Hm2)	13.07	0.21	39.92	0.00	15.17	0.33
pHR81 ilv5p xylA(Hm3)	12.48	0.26	39.92	0.00	16.20	0.18
pHR81 ilv5p xylA(Hm4)	11.26	0.84	31.50	2.67	11.65	1.20
pHR81 ilv5p xylA(Oral-2)	2.88	0.09	0.60	0.15	0.00	0.00
pHR81 ilv5p xylA(Term-1)	2.78	0.29	0.04	0.09	0.00	0.00
pHR81 ilv5p xylA(xylA10)	3.22	0.50	0.64	0.12	0.00	0.00
	After 72 hours					
pHR81 ilv5p xylA(Oral-2)	2.70	0.35	0.71	0.24	0.00	0.00
pHR81 ilv5p xylA(Term-1)	2.61	0.17	1.20	0.06	0.00	0.00
pHR81 ilv5p xylA(xylA-10)	3.02	0.13	1.00	0.10	0.00	0.00

As shown in Table 3, yeast strains containing the chimeric gene for expression of Hm1, Hm2, Hm3, and Hm4 consumed xylose and at the same time, produced ethanol when measured at 24 hours. After 44 hours

of incubation essentially all of the xylose was consumed and over 15 g/L of ethanol was produced by strains expressing Hm1, Hm2, and Hm3. For strains expressing Hm4, a majority of the xylose was consumed after 44 hours, producing about 11 g/L of ethanol. These results indicate that Hm1, Hm2, Hm3, and Hm4 were expressed as active xylose isomerase enzymes in *S. cerevisiae*. Strains expressing other Oral-2, Term-1, and XylA10, however, consumed almost no xylose and did not produce ethanol even after 72 hours.

CLAIMS

What is claimed is:

1. A recombinant yeast cell comprising a heterologous nucleic acid molecule encoding a polypeptide having xylose isomerase activity and amino acid sequence with greater than 95% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NOs:1, 3, 5, and 7.

2. The recombinant yeast cell of claim 1 further comprising a complete xylose utilization pathway and having the ability to grow on xylose as a sole carbon source.

3. The recombinant yeast cell of claim 2 further comprising a target compound.

4. The recombinant yeast cell of claim 3 wherein the target compound is selected from the group consisting of ethanol, butanol, and 1,3-propanediol.

5. A method for producing a yeast cell that has xylose isomerase activity comprising:

a) providing a yeast cell;

b) introducing a heterologous nucleic acid molecule encoding a polypeptide having xylose isomerase activity and amino acid sequence with greater than 95% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NOs:1, 3, 5, and 7;

wherein a yeast cell having xylose isomerase activity is produced.

6. The method of claim 5 wherein the heterologous nucleic acid molecule is a part of a chimeric gene.

7. The method of claim 5 wherein the yeast cell having xylose isomerase activity has a complete xylose utilization pathway and grows in a medium

comprising xylose as a carbon source; wherein xylose is utilized.

8. The method of claim 7 wherein the yeast cell comprises a metabolic pathway that produces a target compound.

9. The method of claim 8 wherein the target compound is selected from the group consisting of ethanol, butanol, and 1,3-propanediol.