



(43) International Publication Date
11 December 2014 (11.12.2014)

(10) International Publication Number
WO 2014/195522 A2

(51) International Patent Classification:
C07K 14/395 (2006.01)

(21) International Application Number:
PCT/EP2014/063070

(22) International Filing Date:
20 June 2014 (20.06.2014)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
13197988.2 18 December 2013 (18.12.2013) EP
14160772.1 19 March 2014 (19.03.2014) EP
14164270.2 10 April 2014 (10.04.2014) EP

(71) Applicants: **DSM IP ASSETS B.V.** [NL/NL]; Het Overloon 1, NL-6411 TE Heerlen (NL). **RIJKSUNIVERSITEIT GRONINGEN** [NL/NL]; Broerstraat 5, NL-9712 CP Groningen (NL).

(72) Inventors: **KLAASSEN, Paul**; P.O. Box 4, NL-6100 AA Echt (NL). **DE WAAL, Paulus Petrus**; P.O. Box 4, NL-6100 AA Echt (NL). **DE JONG, René Marcel**; P.O. Box 4, NL-6100 AA Echt (NL). **DRIESSEN, Arnold Jacob Mattieu**; P.O.Box 11103, NL-9700 CC Groningen (NL). **NIJLAND, Jeroen Gerben**; P.O.Box 11103, NL-9700 CC Groningen (NL). **SHIN, Hyun Yong**; P.O.Box 11103, NL-9700 CC Groningen (NL).

(74) Agent: **KLEIBORN, Paul Erik**; DSM Intellectual Property, P.O. Box 4, NL-6100 AA Echt (NL).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, 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:

- upon request of the applicant, before the expiration of the time limit referred to in Article 21(2)(a)
- without international search report and to be republished upon receipt of that report (Rule 48.2(g))
- with sequence listing part of description (Rule 5.2(a))

(54) Title: POLYPEPTIDES WITH PERMEASE ACTIVITY

(57) Abstract: The invention relates to a polypeptide having one or more substitution corresponding to T89C/Y, V187M/Q/C/G/Y, A191 I/K/D/C/E/N/W/R/P, Y214L/P, Q215D/E/S/M/L/C/F/A, I218W, T219V/L/R/E, I222G/Y/M/P/S/W, Y226M/T/P/Q, Q338R/I/D/E/F/G/P, M339Y/E/N, Q341V/Y/S/H, Q342N/H/R/C/F/G/I/K/L/P/S/T/W, L343Y/Q/I/C/E/K/N/P/S, N346G/I, N347S/L/K/G/TA/, Y350E/S/Q, G373R/K/L/IWY, N376K/Q/A/C/D/E/I/M/P, T380Y/G/M/A/S/D/E/I/K/L/W, S383Y/WA/C/E/I/L/Q/R, F446H/D/I, T448R/K/H/D/E/W, T449F/H/A/G/I/L/P/QA/, T451W/Y/S, W455D/K/E/P/Q, N478Q/I/F/G, or W479S/I/F/D/E/H/M/N/R/T/Y of SEQ ID NO:59 wherein the polypeptide is member of the Major Facilitator Superfamily (MFS).



WO 2014/195522 A2

POLYPEPTIDES WITH PERMEASE ACTIVITY

Field of the invention

5 The invention is directed to novel polypeptides and to recombinant organisms expressing the polypeptides. In an embodiment, the present invention relates to novel permease polypeptides with altered sugar specificity and/or sugar transport activity.

Background of the invention

10 The plasma membrane of yeast cells and other eukaryotes is a complex bio-membrane, consisting of two layers of phospholipids, with a plethora of proteins embedded in it. Many molecules may cross the plasma membrane by diffusion and osmosis or with the aid of specific transport systems.

Transport systems allow the uptake of nutrients and ions, export of products of metabolism and undesired or harmful substances. Different mechanisms exist. Primary active transporters drive solute accumulation or extrusion by using for instance ATP hydrolysis. Secondary carriers, belonging to the Major Facilitator Superfamily (MFS) 20 transporters, facilitate the transport of one or more molecular species across the membrane in response to chemi-osmotic gradients. In the yeast *Saccharomyces cerevisiae*, 186 MFS proteins have been identified (Nelissen, 1997) in strain S288c.

An example of such a carrier is the Hxt1 protein, involved in hexose transport in *Saccharomyces cerevisiae*.

25 Permeases are membrane transport proteins, a class of multipass transmembrane proteins that facilitate the diffusion of a specific molecule, herein specifically one or more sugar, in or out of the cell by passive transport. In contrast, active transporters couple molecule transmembrane transport with an energy source such as ATP or a favorable ion gradient.

30 The terms permease, facilitator, transporter or transport protein or related terms are all describing proteins with multiple membrane spanning domains that exhibit a function in transporting molecules across a membrane. This transport can be brought about by different mechanisms: uniport (transport of one molecule), symport (simultaneous co-transport of two different molecules in the same direction), antiport (simultaneous transport of two molecules in opposite directions) and facilitated diffusion. 35

The family of sugar transporters in yeast consists of 30-40 members (34 members in strain S288c (Nelissen, 1997)). The sugar transporters can be divided in five clusters: hexose permeases (*HXT*-genes, *GAL2*), disaccharide permeases, myo-inositol permeases, sugar receptors and a final cluster of transporters of which the substrate is unknown.

Lignocellulosic biomass, an attractive alternative feedstock for the production of liquid transportation fuels, consists of several different sugars. The hexose fraction of lignocellulose, mainly glucose, can in principle be readily fermented by non-recombinant versions of the yeast *S. cerevisiae*. However, this organism is not able to metabolize the pentose sugars, such as xylose and arabinose, into ethanol.

Methods of creating microorganisms that are able to metabolize pentose sugars are known in the art. For instance, in WO/2009/109630 the construction of expression cassettes and the transformation of *S. cerevisiae* cells into pentose fermenting cells by expressing xylose isomerase are illustrated.

Native pentose-utilizing organisms exist but are lacking well-developed genetic tools and/or low product tolerances, which limit their suitability as hosts for lignocellulosic conversion processes.

As a consequence, efforts have focused on the introduction of pentose conversion pathways in the yeast *S. cerevisiae*, which is still the organism of choice in the ethanol industry, in order to enable pentose fermentation.

Despite the vast amount of progress achieved in the past years, the transport of pentose sugars is still considered to be (one of) the rate-limiting step in pentose metabolism.

Pentose transport in *S. cerevisiae* is mediated by the different members of the hexose transporter (Hxt) family. Hxt4, Hxt5, Hxt7 and Gal2 have been described as the main xylose transporters in *S. cerevisiae* (Hamacher *et al*, 2002), and Gal2 is also known to mediate arabinose transport (Becker. *et al*, 2003). However, the affinity for the respective pentose sugars is approximately 10 to 100 times lower than for the respective hexose sugars. The lack of a dedicated xylose or arabinose transporter in recombinant yeast cells thus limits the capacity for co-utilization of hexoses and pentoses in sugar mixtures, and prohibits a high pentose catabolic flux. As a consequence, conversion of biomass sugars may be considered bi-phasic: in the first phase, a relatively fast conversion of hexoses (glucose) takes place, while in the second phase, which starts

when the hexoses have been exhausted from the medium, pentose fermentation commences, but at a far lower rate as compared to the rate of hexose conversion.

It is therefore a long-felt desire to express pentose-specific sugar transporters, i.e. no glucose interference (pentose specificity) and high affinity to pentose, in an otherwise unchanged transporter landscape, in order to maintain the ability to convert hexoses at approximately the same level.

One way of solving this problem is to screen for heterologous sugar (pentose) transporters which are pentose specific and have a (moderately) high affinity for pentose. However, such efforts have been with limited success so far. Only a few have been shown to be able to facilitate pentose transport upon expression in *S. cerevisiae*, but all favour glucose above xylose (Young *et al*, 2011, and references therein), as is the case with the *S. cerevisiae* Hxt-proteins, as indicated above.

Another approach is to re-engineer hexose transporters to pentose transporters. For instance, the works by Kasahara *et al* (2000; 2009; 2010) indicate which residues in several sugar transporters play a key role in the determination of the substrate affinity to the natural substrate.

Mutant hexose transporters that are able to transport pentose sugars more efficiently are known in the art. For instance, in WO/2012/049173, the isolation of mutant hexose transporter genes from cultures of pentose-fermenting *S. cerevisiae* cells is described.

In *S. cerevisiae*, the permease GAL2 transports galactose across the cell membrane. It is also known as a transporter of glucose across the membrane.

Summary of the invention

An object of the invention is to provide novel permease polypeptides with altered, in particular improved, sugar specificity. Another object of the invention is to provide recombinant strains expressing the permease polypeptide that have improved uptake of the molecule that the permease transports across the cell membrane. Another object is to provide a permease polypeptide that has a improved capacity for transport of C5 sugars, in particular xylose compared to a parent polypeptide. Another object is to provide a permease polypeptide that has reduced transport activity for C6 sugar, in particular glucose, compared to a parent polypeptide. Another object is to provide a method to identify mutations in other related permease polypeptides that have a

beneficial effect on the improved capacity for transport of xylose or reduced transport activity for glucose.

One or more of these objects are attained according to the invention. According to the present invention, there is provided a polypeptide having one or more substitution
 5 corresponding to T89C/Y, V187M/Q/C/G/Y, A191I/K/D/C/E/N/W/R/P, Y214L/P, Q215D/E/S/M/L/C/F/A, I218W, T219V/L/R/E, I222G/Y/M/P/S/W, Y226M/T/P/Q, Q338R/I/D/E/F/G/P, M339Y/E/N, Q341V/Y/S/H, Q342N/H/R/C/F/G/I/K/L/P/S/T/W, L343Y/Q/I/C/E/K/N/P/S, N346G/I, N347S/L/K/G/T/V, Y350E/S/Q, G373R/K/L/M/Y, N376K/Q/A/C/D/E/I/M/P, T380Y/G/M/A/S/D/E/I/K/L/W, S383Y/W/V/C/E/I/L/Q/R,
 10 F446H/D/I, T448/R/K/H/D/E/W, T449F/H/A/G/I/L/P/Q/V, T451W/Y/S, W455D/K/E/P/Q, N478Q/I/F/G, or W479S/I/F/D/E/H/M/N/R/T/Y of SEQ ID NO:59 wherein the polypeptide is member of the Major Facilitator Superfamily (MFS) as defined herein after.

In an embodiment the polypeptide has pentose transport activity.

A polypeptide according to the invention having one or more of these mutations
 15 has an advantageous sugar consumption and/or fermentation product production. This this will be described in more detail below and will be illustrated by examples 1-5 below.

Brief description of the drawings

20 FIG. 1 shows results of aerobic shake flask cultures hexose transporter mutants on Verduyn-urea + 15 g l⁻¹ glucose + 20 g l⁻¹ xylose. (A) Optical density measurements at 600 nm wavelength, (B) glucose concentrations (g l⁻¹), (C) xylose concentrations during the culture period.

25 FIG. 2 shows results of aerobic shake flask cultures hexose transporter mutants on Verduyn-urea + 20 g l⁻¹ xylose. (A) Optical density measurements at 600 nm wavelength, (B) xylose concentrations (g l⁻¹).

FIG. 3 shows the strain construction scheme for strain YD01227

30 FIG. 4 shows results of aerobic shake flask cultures of quadruple hexose kinase mutant (col 2) and reference strain RN1014 on Verduyn-urea + 100 g l⁻¹ glucose + 20 g l⁻¹ xylose. A) Optical density measurements at 600 nm wavelength, (B) glucose concentrations (g l⁻¹), (C) xylose concentrations during the culture period.

FIG. 5 shows results of micro-well plate cultures on glucose - xylose mixtures. Growth characteristics of reference strain RN1014, and quadruple hexokinase mutants YD01227 and colony 2 (Col 2) with equal genotype on Verduyn-urea supplemented with mixtures of glucose and xylose in the following concentrations: (A) 60 + 20, (B) 100 + 20, (C) 15 + 5, and (D) 25 g l⁻¹ + 5 g l⁻¹, respectively. Every 15 minutes, OD600 was measured automatically by Bioscreen C apparatus. Data points are the average of measurements in triplicate.

FIG. 6 shows a plasmid map pRN993.

FIG. 7 shows the relationships between residual xylose and glucose concentrations (A), or residual xylose concentrations and growth (OD600; B) in the Glucose Transport Activity Counter Screen (GTAC screen), see examples. Residual xylose concentrations (g l⁻¹) were plotted against the residual (A) glucose concentration (g l⁻¹) and (B) OD600 at 96 hours. (Inset B) linear regression analysis of shows a good correlation of OD600 and xylose. Each data point represents the average of sugar concentrations of 1 to 3 replicates, or of multiple measurements of RN1001 control strain or medium samples per part of the screen.

FIG. 8 shows growth and xylose consumption of Gal2-N376-mutant variants in the GTAC screen. (A) OD600 measurements at time points 24, 72 and 96 hours, and (B) residual xylose concentrations (g l⁻¹) measured for the Gal2-N376-mutant variants, RN1001 control strain and empty samples (YD10 medium-only) during the GTAC Screen. Each column represents the average OD600 or residual xylose concentration of 3 transformants; error bars represent the standard error of the mean. Asterisks indicate amino acids with large hydrophobic side chains with a clear effect.

FIG. 9 shows growth profiles of TOP15 Xylose Transport Activity (XTA) Screening (see examples). OD600 values of the three sampling points for the TOP15 strains/variants from (A) Part A and (B) Part B of the XTA Screen. Each column represents the average OD600 of 3 transformants; error bars represent the standard error of the mean.

FIG. 10 shows an alignment of permease protein sequences.

FIG. 11 shows plot of amino acid van der Waals volume (Å³) (Y-axis) against amino acid hydrophobicity (Δt_R) (X-axis). Advantageous amino acids (reduced glucose

transport activity) for position corresponding to 376 in SEQ ID NO: 59 are within the area defined by the short striped lines and for position 339 by the long strips lines. Values for van der Waals volume used herein are described in: http://www.proteinsandproteomics.org/content/free/tables_1/table08.pdf. The corresponding literature is N. J. Darby, Thomas E. Creighton, Protein Structure (1993) Oxford University Press. The values for hydrophobicity (Δt_R) of amino acids are herein used as described in <http://onlinelibrary.wiley.com/doi/10.1002/psc.310010507/pdf>. The reference corresponding to this is Monera, O.D. et al, Journal of Peptide science Vol. 1 319-329 (1995).

FIG. 12 shows (A) Growth curve of evolved strain RN1053-X2 and wild-type strain RN1053 on Verduyn-urea-his supplemented with 2% xylose. Growth curves were expressed as units of optical density (OD) measured at 600 nm wavelength (OD₆₀₀) over time (h) (B) Expression level of *HXT8-HXT17* in the strain RN1053-X2. Expression level relative to *ACT1* levels in each sample, was expressed as Normalised Fold Expression Level; expression levels were normalized against RN1053 mRNA levels of the specific *HXT* gene (RN1053 expression level set to 1).

FIG. 13 (A-B) shows. (A) *HXT11* Expression level and (B) Growth of *HXT11* knockout strains (KO1 to -6) compared to RN1053-X2-empty (1053-X2) and RN1053-empty. Growth was expressed as units of optical density (OD) measured at 600 nm wavelength (OD₆₀₀).

FIG. 13 (C-D) shows (C) Plasmid map of pRS313-P7T7, (D) Growth of RN1053-X2 transformed with empty vector control (1053-X), and of four transformants after introduction of pRS313-P7T7-inverseHXT11 (iHXT1 to -4) on Verduyn-urea supplemented with 2% xylose. Growth was expressed as units of optical density (OD) measured at 600 nm wavelength (OD₆₀₀)

FIG. 14 shows growth profiles of RN1053-HXT11 (Hxt11, closed diamonds), RN1053-HXT12 (Hxt12, open triangles), RN1053-empty vector control (1053; triangles), RN1041-empty vector control (1041; open squares) on Verduyn-urea supplemented with (A) mixture of glucose and xylose (2% each) or (B) 2% xylose. Growth curves were expressed as units of optical density (OD) measured at 600 nm wavelength (OD₆₀₀) over time (h).

FIG. 15 shows xylose uptake studies on RN1053-HXT11 (HXT11), RN1053-HXT2 (Hxt2), RN1053-empty (empty) with or without increasing concentrations of competing glucose (0-500 mM).

FIG. 16 shows (A) Growth of RN1053 expressing chimeric Hxt11p-GFP protein on 2% xylose at 16 h (initial cell density was OD 0.5). (B) Fluorescence microscopy of RN1053-HXT11 and RN1041-HXT11 grown on Verduyn-urea supplemented with glucose or xylose.

5 FIG. 17 shows fermentation profiles of sugar consumption and ethanol production of (A) RN1041-empty plasmid and (B) RN1053-HXT11 on Verduyn-urea supplemented with 80 g l⁻¹ glucose and 40 g l⁻¹ xylose.

FIG. 18 shows growth profiles (OD600)(A), residual xylose (B) and glucose (C) concentrations (g l⁻¹) after 96 hours GTAC-screen with YD01227-GAL2 (n=3), YD01227-HXT2 (n=3), YD01227-HXT11 (n=3) transformants. Medium and RN1001 samples were added as controls.

FIG. 19 (A-B). shows growth profiles of shake flask cultures of (A) YD01227-empty, YD01227-HXT11 and YD01227-mHXT11(N366D) transformants (n=2) on Verduyn-urea supplemented with 15% glucose and 1% xylose and, (B) of RN1053-
15 transformants on Verduyn-urea with 2% xylose (n=3).

FIG. 19 (C-E) shows ¹⁴C-radiolabeled sugar uptake profiles by RN1053-HXT11 and RN1053-mHXT11(N366D). (A) ¹⁴C-glucose uptake, and ¹⁴C-xylose uptake in the absence (D) and presence (E) of increasing unlabeled glucose (0-500 mM) concentrations. Sugar uptake was expressed as nmol mg dry weight (DW) of yeast per
20 minute (min).

FIG. 19 (C-F) shows fermentation profiles of RN1041-empty (closed diamonds; 1041+empty), RN1053-HXT11 (closed triangles; 1053+HXT11), RN1053-mHXT11 (N366D) (closed squares; 1053+epHXT11) on Verduyn-urea + 100 g l⁻¹ glucose and 60 g l⁻¹ xylose. (C) Growth curves (expressed by OD600 measurements) over time (h).
25 Constituents measured in the fermentation broth over time (h) such as (D) glucose, (E) xylose and (F) ethanol.

FIG. 19 (J-L) shows fermentation profiles on Verduyn-urea supplemented with 80 g l⁻¹ glucose and 40 g l⁻¹ xylose of strain RN1053-HXT11 (J and L-grey dashed line) and RN1053-mHXT11(N366D) (K and L-black line). (L) Combined figure of both
30 fermentation profiles illustrated in (J) and (K).

FIG. 20 shows the scheme for the glucose/xylose ratio in the Verduyn-urea-his medium during chemostat cultivation (days) of YD01227;

FIG. 21 shows growth profiles of shake flask cultures on Verduyn-urea-his supplemented with 1% xylose without (closed triangles; 1+0) or with (3% glucose, stripe,

1+3; 6% glucose, closed circles, 1+6; 10% glucose, closed diamonds, 1+10) increasing glucose concentrations with YD01227-evo (A) and YD01227-ori (B). (C) ^{14}C -xylose uptake study of YD01227-ori (1227-ori, closed squares) or YD01227-evo (1227-evo, closed diamonds) with or without increasing glucose concentrations (0-500mM).

5 FIG. 22 shows (A) mRNA expression profile of YD01227-ori (white bars), YD01227-evoB on Verduyn-urea-his supplemented with a sugar mixture in a ratio xylose:glucose 1:3 (gray bars) and 1:10 (black bars) . Data were expressed as normalized expression (relative to *ACT1* and relative to YD01227-ori, which was set to 1).

10 FIG. 22 (B-D) Uptake experiments with Xylose (B) and Glucose (C). Uptake was measured in nmol/mg.DW.min in RN1053-Hxt3-6 (diamonds), RN1053-N367I (triangles) and 1053-emp (squares). (D) Uptake of 50mM ^{14}C Xylose in the presence of 0, 50, 100, 200 and 500mM Glucose in the RN1053-Hxt3-6 (diamonds) and RN1053-N367I strain (squares).

15 FIG. 23 Fluorescence images of strain RN1053 expressing GFP fusion proteins of Hxt36 (A) and Hxt36-N367I (B). Images were analyzed on a Nikon Eclipse-Ti microscope. (C) Total amount of GFP fluorescence (in AU/OD600) in both strains.

FIG. 24 Growth of the YD01227 strain containing vectors expressing HXT36 transporters with all possible amino acid substitutions at position 367. Cells were grown on 1% xylose and 10% xylose, and YD01227 transformed with the empty vector pRS313-P7T7 was used as a control.

FIG. 25 Uptake experiments to determine the K_m and V_{max} for glucose (panel A) and xylose (panel B). Uptake was measured in nmol/mgDW.min in the RN1053-HXT36 strain (diamonds), the RN1053-HXT36-N367I strain (squares) and in the RN1053-HXT36-N367A strain (triangles). The uptake levels of the RN1053-empty strain were, for both sugars, subtracted from the RN1053-HXT36, RN1053-HXT36-N367I and RN1053-HXT36-N367A strains.

FIG. 26 Growth of the RN1053 strain expressing *HXT36* (A), *HXT36*-N367I (B) and *HXT36*-N367A (C) on 5 g L⁻¹ D-glucose and 5 g L⁻¹ D-xylose. The residual D-glucose (diamonds), D-xylose (squares) and ethanol (circles) were measured in g/l.

FIG. 27 Characterization of xylose specificity of HXT11-N366X mutants. (A) Maximal Exponential growth rate (1/h) of HXT11-N366X mutants in the strain YD01227. Maximal exponential growth rate (1/h) for N366X mutants expressed in RN1053 on glucose (B) or xylose (C).

FIG. 28 Fluorescence images of strain RN1053 expressing GFP fusion proteins of HXT11 N366X mutants grown on 2% maltose. Letters in left upper hand of picture depict the amino acid on position 366 of the respective Hxt11 variant.

FIG. 29 Fermentations on Verduyn-urea supplemented with xylose (40 g L⁻¹) and glucose (71,8 g L⁻¹) of (A) RN1053 HXT11-N366M, (B) RN1053 HXT11-N366T, (C) RN1001 and (D) RN1053 HXT11-N366N. Symbols: glucose (◆), xylose (■), ethanol (▲), cell density (●). (A), (B), (C) and (D) are the panels of FIG. 29.

Brief description of the sequence listing

SEQ ID NO: 1: primer 5034-kanf
 SEQ ID NO: 2: primer 5035-kanr
 SEQ ID NO: 3: primer 5116-lf2
 SEQ ID NO: 4: primer 5118-lr2
 SEQ ID NO: 5: primer 5115-lf1
 SEQ ID NO: 6: primer 5117-lr1
 SEQ ID NO: 7: pRN201; TOPO-BLUNT-loxP-kanMX-loxP
 SEQ ID NO: 8: pRN251; TOPO-BLUNT-loxP-hphMX-loxP
 SEQ ID NO: 9: pRN365; TOPO-BLUNT-loxP-natMX-loxP
 SEQ ID NO: 10: primer 115-natf
 SEQ ID NO: 11: primer 116-natr
 SEQ ID NO: 12: pRN447; TOPO-BLUNT-loxP-zeoMX-loxP
 SEQ ID NO: 13: primer 28-H3f
 SEQ ID NO: 14: primer 29-H3r
 SEQ ID NO: 15: pRN247 (TOPO- BLUNT-HIS3:loxPkanMXloxP)
 SEQ ID NO: 16: primer 201-Hx2uf
 SEQ ID NO: 17: primer 202-Hx2ur
 SEQ ID NO: 18: primer 203-Hx2df
 SEQ ID NO: 19: primer 204-Hx2dr
 SEQ ID NO: 20: primer 205-Hx3uf
 SEQ ID NO: 21: primer 206-Hx3ur
 SEQ ID NO: 22: primer 210-Hx4df
 SEQ ID NO: 23: primer 211-Hx4dr
 SEQ ID NO: 24: primer 212-Hx5uf

- SEQ ID NO: 25: primer 213-Hx5ur
SEQ ID NO: 26: primer 229-Hx7df
SEQ ID NO: 27: primer 230-Hx7dr
SEQ ID NO: 28: primer 243-Gal2ufn
5 SEQ ID NO: 29: primer 244-Gal2urn
SEQ ID NO: 30: primer 233-Ga2df
SEQ ID NO: 31: primer 234-Ga2dr
SEQ ID NO: 32: pRN485; TOPO-BLUNT-GAL2:loxPzeoMXloxP
SEQ ID NO: 33: pRN566; TOPO-BLUNT-HXT367:loxP-hphMX-loxP
10 SEQ ID NO: 34: pRN569; TOPO-BLUNT-HXT514:loxP-natMX-loxP
SEQ ID NO: 35: pRN635; TOPO-BLUNT-HXT2:loxP-kanMX-loxP
SEQ ID NO: 36: primer 281-Hx3inr2
SEQ ID NO: 37: primer 323-Hx7inr1
SEQ ID NO: 38: primer Hx4inr2
15 SEQ ID NO: 39: primer Hx5inf
SEQ ID NO: 40: primer 324-Ga2inf1
SEQ ID NO: 41: primer 325-Ga2inr1 SEQ ID NO 42: primer 289-Hx2inf
SEQ ID NO: 43: primer 290-Hx2inr
SEQ ID NO: 44: primer 838-Glk1-psuc227f
20 SEQ ID NO: 45: primer 834- Hxk2-psuc227f
SEQ ID NO: 46: primer 645-pSUC227r
SEQ ID NO: 47: primer 839-Glk1-psuc225r
SEQ ID NO: 48: primer 835-Hxk2-psuc225r
SEQ ID NO: 49: primer 646-pSUC225f
25 SEQ ID NO: 50: primer 846-Hxk1_loxP_f
SEQ ID NO: 51: primer 847-Hxk1_loxP_r
SEQ ID NO: 52: primer 848-Gal1_loxP_f
SEQ ID NO: 53: primer 849-Gal1_loxP_r
SEQ ID NO: 54: pRN774; TOPO-BLUNT-loxP-hphMX-loxP (loxP sites in opposite
30 orientation)
SEQ ID NO: 55: pRN775; TOPO-BLUNT-loxP-natMX-loxP (loxP sites in opposite
orientation)
SEQ ID NO: 56: *WT-GAL2* DNA sequence SEQ ID NO: 57 pRN993; XbaI site (TCTAGA)
and BssHII site (GCGCGC).

SEQ ID NO: 58: pDB1250; WT-GAL2 expression vector for screening; XbaI site (TCTAGA) and BssHII site (GCGCGC).

SEQ ID NO: 59: WT Gal2p amino acid sequence

SEQ ID NO: 60: pRN187 (pSH65-derived CRE recombinase expression vector)

5 SEQ ID NO: 61: pRN486 (TOPO-BLUNT-his3::loxP-natMX-loxP)

SEQ ID NO: 62: Primer ActinF (Real time PCR)

SEQ ID NO: 63: Primer ActinR (Real time PCR)

SEQ ID NO: 64: Primer HXT1F (Real time PCR)

SEQ ID NO: 65: Primer HXT1R (Real time PCR)

10 SEQ ID NO: 66: Primer HXT2F (Real time PCR)

SEQ ID NO: 67: Primer HXT2R (Real time PCR)

SEQ ID NO: 68: Primer HXT3F (Real time PCR)

SEQ ID NO: 69: Primer HXT3R (Real time PCR)

SEQ ID NO: 70: Primer HXT4F (Real time PCR)

15 SEQ ID NO: 71: Primer HXT4R (Real time PCR)

SEQ ID NO: 72: Primer HXT5F (Real time PCR)

SEQ ID NO: 73: Primer HXT5R (Real time PCR)

SEQ ID NO: 74: Primer HXT7F (Real time PCR)

SEQ ID NO: 75: Primer HXT7R (Real time PCR)

20 SEQ ID NO: 76: Primer HXT8F ((Real time PCR)

SEQ ID NO: 77: Primer HXT8R (Real time PCR)

SEQ ID NO: 78: Primer HXT9F (Real time PCR)

SEQ ID NO: 79: Primer HXT9R (Real time PCR)

SEQ ID NO: 80: Primer HXT10F (Real time PCR)

25 SEQ ID NO: 81: Primer HXT10R (Real time PCR)

SEQ ID NO: 82: Primer HXT11F (Real time PCR)

SEQ ID NO: 83: Primer HXT11R (Real time PCR)

SEQ ID NO: 84: Primer HXT12F (Real time PCR)

SEQ ID NO: 85: Primer HXT12R (Real time PCR)

30 SEQ ID NO: 86: Primer HXT13F (Real time PCR)

SEQ ID NO: 87: Primer HXT13R (Real time PCR)

SEQ ID NO: 88: Primer HXT14F (Real time PCR)

SEQ ID NO: 89: Primer HXT14R (Real time PCR)

SEQ ID NO: 90: Primer HXT15F (Real time PCR)

- SEQ ID NO: 91: Primer HXT15R (Real time PCR)
SEQ ID NO: 92: Primer HXT16F (Real time PCR)
SEQ ID NO: 93: Primer HXT16R (Real time PCR)
SEQ ID NO: 94: Primer HXT17F (Real time PCR)
5 SEQ ID NO: 95: Primer HXT17R (Real time PCR)
SEQ ID NO: 96: Primer GAL2F (Real time PCR)
SEQ ID NO: 97: Primer GAL2R (Real time PCR)
SEQ ID NO: 98: Primer KOP11* for KO HXT11
SEQ ID NO: 99: Primer KOT11* for KO HXT11
10 SEQ ID NO: 100: Primer iHXT11F (Inverse HXT11)
SEQ ID NO: 101: Primer iHXT11R (Inverse HXT11)
SEQ ID NO: 102: Primer HXT11F (Cloning)
SEQ ID NO: 103: Primer HXT12F (Cloning)
SEQ ID NO: 104: Primer HXT11/12R (Cloning)
15 SEQ ID NO: 105: Primer HXT1 XbaI (Cloning)
SEQ ID NO: 106: Primer R HXT1 Cfr9I (Cloning)
SEQ ID NO: 107: Primer F HXT2 XbaI (Cloning)
SEQ ID NO: 108: Primer R HXT2 Cfr9I (Cloning)
SEQ ID NO: 109: Primer F HXT3 XbaI (Cloning)
20 SEQ ID NO: 110: Primer R HXT6 Cfr9I (Cloning)
SEQ ID NO: 111: Primer F HXT4 XbaI (Cloning)
SEQ ID NO: 112: Primer R HXT4RN Cfr9I (Cloning)
SEQ ID NO: 113: Primer F HXT5 XbaI (Cloning)
SEQ ID NO: 114: Primer R HXT5 Cfr9I (Cloning)
25 SEQ ID NO: 115: Primer F HXT7 XbaI (Cloning)
SEQ ID NO: 116: Primer R HXT7 Cfr9I (Cloning)
SEQ ID NO: 117: Plasmid pRS313-P7T7
SEQ ID NO: 118: Plasmid pRS313-P7t7-HXT11+GFP
SEQ ID NO: 119: DNA sequence of HXT11 ORF *Saccharomyces cerevisiae*
30 SEQ ID NO: 120: DNA sequence of HXT2 ORF *Saccharomyces cerevisiae*
SEQ ID NO: 121: DNA sequence of GAL2 ORF *Saccharomyces cerevisiae*
SEQ ID NO: 122: DNA sequence of HXT3-6 ORF *Saccharomyces cerevisiae*
SEQ ID NO: 123: Hxt11p amino acid sequence *Saccharomyces cerevisiae*
SEQ ID NO: 124: Hxt2p amino acid sequence *Saccharomyces cerevisiae*

SEQ ID NO: 125: Gal2p amino acid sequence *Saccharomyces cerevisiae*

SEQ ID NO: 126: Hxt3-6 amino acid sequence *Saccharomyces cerevisiae*

SEQ ID NO: 127: F HXT36 Bcui

SEQ ID NO: 128: R HXT36 367NNN

5 SEQ ID NO: 129: F HXT36 367NNN

SEQ ID NO: 130: R HXT36 BamHI

SEQ ID NO: 131: R HXT36 BamHI-stop

SEQ ID NO: 132: F GFP BamHI

SEQ ID NO: 133: R GFP ClaI

10 SEQ ID NO: 134: F HXT11 XbaI

SEQ ID NO: 135: R HXT11 BamHI

SEQ ID NO: 136: F HXT11 366NNN

SEQ ID NO: 137: R HXT11 366NNN

SEQ ID NO: 138: F HXT11 N366F

15 SEQ ID NO: 139: R HXT11 N366F

SEQ ID NO: 140: F HXT11 N366E

SEQ ID NO: 141: R HXT11 N366E

SEQ ID NO: 142: F HXT11 N366K

SEQ ID NO: 143: R HXT11 N366K

20 SEQ ID NO: 144: F HXT11 N366M

SEQ ID NO: 145: R HXT11 N366M

SEQ ID NO: 146: F HXT11 N366W

SEQ ID NO: 147: R HXT11 N366W

SEQ ID NO: 148: F HXT11 N366Y

25 SEQ ID NO: 149: R HXT11 N366Y

30

Detailed description of the invention

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

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

5 The invention relates to a method of identifying amino acid positions in permease polypeptides, preferably hexose permease polypeptides, more preferably hexose permease polypeptides from yeast and fungi, even more preferably in *Saccharomyces cerevisiae* Hxt or Gal2 permease polypeptides, which can be mutated to alter the sugar specificity of the permease.

10 The permeases belong to the Major Facilitator Superfamily (MFS). This is defined hereinbelow. Cellular transport systems allow the uptake of essential nutrients and ions, and excretion of products of metabolism and deleterious substances. In addition, transport systems play a role in the communication between cells and the environment. Also, they are an essential part of the cell system to yield or consume energy-supplying molecules, such as ATP.

15 The transport of solutes by primary active transporters is energy-driven in the first place, such as by energy supplied from ATP hydrolysis, photon absorption, electron flow, substrate decarboxylation, or methyl transfer. If charged molecules are pumped in one direction as a consequence of the consumption of a primary cellular energy source, an electrochemical potential is the result. The resulting chemiosmotic gradient can then be
20 used to drive the transport of additional molecules via secondary carrier transporters which facilitate the transport of one or more molecules across the membrane, with or without the cotransport of protons or sodium ions.

The last two decades the existence of a multitude of previously unknown protein families of primary and secondary transporters has been clarified by the emergence of
25 genomics strategies and making use of the many performed biochemical and molecular genetics studies. The two main transporter families of which proteins were found throughout all living organism are of the ATP-binding cassette (ABC) superfamily and the major facilitator superfamily (MFS), also known as the uniporter-symporter-antiporter family. Whereas ABC family permeases consist of multiple components and are primary
30 active transporters, capable of transporting both small molecules and macromolecules only after generating energy through ATP hydrolysis, the MFS transporters consist of a single polypeptide of a secondary carrier which facilitates transport of small solutes in response to a chemiosmotic ion gradient. ABC superfamily and MFS proteins account for almost half of the solute transporters encoded within the microbe genomes (reviewed

by Pao *et al*, 1998, Microbiol Mol Biol Rev.; **62** pp.1–34, and Saier *et al*, 1999, J Mol Microbiol Biotechnol, **1** pp.257-279).

Suitable permease polypeptide sequences can contain one or more of the following motifs:

- 5 a) G-R-x(3)-G-x(3)-G-x(11)-E-x(5)-[LIVM]-R-G-x(12)-[GA];
- b) R-x(14)-G-x(2)-Y-x(2)-[YF]-[YF]-[GSAL] and/or
- c) V-x(15)-[GNR]-[RH]-R-x(2)-[LM]-x(2)-[GA]

Motif (a) is corresponds to residues 179-221 in Gal2; motif (b) is corresponds to
10 residues 330-353 in Gal2; motif (c) is corresponds to residues 375-399 in Gal2.

The claimed method comprises modeling a permease polypeptide sequence onto the published crystal structure of the xylose- or glucose-bound *Escherichia coli* xylose permease Xyle (respectively, PDB code 4GBY & 4GBZ in the PDB database, <http://www.pdb.org>) to identify the amino acid positions in the channel of the permease
15 that directly interact with the bound sugar (called the first-shell residues in the art), and the residues that interact with the first shell residues (called the second shell residues in the art). Suitable modeling software to construct such models are YASARA, Prime (Schrodinger Inc.) or MODELLER using the default settings. Alternatively, the sugar-specificity-altering first and second shell amino acid positions in a permease polypeptide
20 sequence can be identified by a global pairwise alignment of the permease sequence with the Gal2 sequence SEQ ID NO: 59 using the NEEDLE protocol described below. An example alignment for Gal2 and Hxt's from *Saccharomyces cerevisiae* is given in Figure 10, which shows how alignment can be used to identify the corresponding amino acid positions in the different yeast Hxt's. The amino acid positions herein thus refer to SEQ
25 ID NO: 59 that describes Gal2 or to corresponding amino acid positions in other polypeptides, in particular other permease polypeptides. For example, the corresponding position of the position N376 in Gal2 (SEQ ID NO: 59) in Hxt1 is N370, in Hxt2 N361, in Hxt3 N367, in Hxt4SC N376, in Hxt4RN N376, in Hxt5 N391, in Hxt6/7 N370, in Hxt8 N372, in Hxt9 N366, in Hxt10 N354, in Hxt11 N366, in Hxt12 N256, in Hxt13 N363, in
30 Hxt14 N387, in Hxt15 N366, in Hxt16 N366 and in Hxt17 N363. Similarly, the corresponding position of N346 in Gal2 (SEQ ID NO:9) in Hxt1 is D340, in Hxt2 N331, in Hxt3 D337, in Hxt4SC D346, in Hxt4RN D346, in Hxt5 D361, in Hxt6/7 D340, in Hxt8 D342, in Hxt9 D336, in Hxt10 C324, in Hxt11 D336, in Hxt12 D226, in Hxt13 E333, in Hxt14 I357, in Hxt15 E336, in Hxt16 E336 and in Hxt17 E333. This can be similarly done

for other MFS Superfamily transporters, so that corresponding positions in these polypeptides corresponding to the positions in SEQ ID NO: 59 can be obtained.

A person skilled in the art can subsequently mutate the identified amino acid positions in the permease polypeptide to all other 19 amino acids, and screen for improved C5 sugar uptake and/or reduced C6 sugar uptake of the mutant permease, as described in Example 4 and 5.

For instance, for a polypeptide having a mutation at a position corresponding to one or more position corresponding to N346 of SEQ ID NO: 59, the mutations at the positions corresponding to N346 may be a substitution with C, P, G, A, V, L, I, M, F, W, Y, H, S, T, N, Q, D, E, K, R or a deletion. X may be any amino acid, X(2) means two X.

Herein, Gal2 is a facilitated diffusion transporter required for both the high-affinity galactokinase-dependent and low-affinity galactokinase-independent galactose transport processes. It belongs to the major facilitator superfamily, sugar transporter (TC 2.A.1.1) family. "Permease polypeptide", is also designated herein as "polypeptide permease" or "polypeptide". "Permease polypeptide polynucleotide", is herein a polynucleotide that encodes the permease polypeptide.

In an embodiment of the invention, the permease polypeptide has at least 35%, at least 40%, at least 45%, at least 50%, at least 60%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99% sequence identity with SEQ ID NO: 59.

Herein mutations are indicated by one letter amino acids and positions of these amino acids. For example, A6 herein indicates an amino acid (one letter code) at a certain position in SEQ ID NO: 59, here A (Alanine) at position 6 of the protein. A6 (L/N/Q/G/V/I/Y/S/E/K) indicates herein mutation of amino acid at a certain position, here A (Alanine) at position 6 of the protein is exchanged for any of L (Leucine), N (Asparagine), Q (Glutamine), G (Glycine), V (Valine), I (Isoleucine), Y (Tyrosine), S (Serine), E (Glutamic acid) or K (Lysine).

In an embodiment, the polypeptide has xylose transport activity.

In an embodiment, the polypeptide has one or more substitution corresponding to V187M/Q/C/G/Y, Q215D/E/S/M/L/C/F/A, I218W, I222G/Y/M/P/S/W, Y226M/T/P/Q, M339Y/E/N, Q342N/H/R/F/G/I/K/L/P/S/T/W, N346G/I, Y350E/S/Q, G373R/K/L/M/Y, N376K/Q/A/C/D/E/I/M/P, S383Y/W/V/C/E/I/L/Q/R, F446H/D/I, T448/R/K/H/D/E/W, T449F/H/A/G/I/L/P/Q/V, or W479S/I/F/D/E/H/M/N/R/T/Y of SEQ ID NO: 59.

In a further embodiment, the polypeptide has one or more substitution corresponding to T89C, V187M/Q, A191I/K/D, Y214L/P, Q215D/E/S, I218W, I222G/Y/M, Y226M/T, Q338R/I, M339Y, Q341V/Y/S, Q342N/H/R, L343Y/Q/I, N346G, N347S/L/K, Y350E, G373R, N376K/Q/, T380Y/G/M, S383Y/W/V, F446H/D, T448/R/K/H, T449F/H, N478Q, or W479S/I/F of SEQ ID NO: 59.

In a further embodiment, the polypeptide has one or more of the substitutions corresponding to V187M/Q, Q215D/E/S, I218W, I222G/Y/M, Y226M/T, M339Y, Q342N/H/R, N346G, Y350E, G373R, N376K/Q/, S383Y/W/V, F446H/D, T448/R/K/H, T449F/H, or W479S/I/F of SEQ ID NO: 59.

In an embodiment, the polypeptide has one or more of the substitutions corresponding to V187M, M339Y, I218W, Q342N, N376K, G373R or W479S of SEQ ID NO: 59.

In an embodiment the polypeptide has reduced glucose transport activity compared to the polypeptide having SEQ ID NO: 59. In an embodiment the polypeptide has increased xylose transport activity compared to the polypeptide having SEQ ID NO: 59.

The permease polypeptide of the invention may have one or more alternative and/or additional activities other than that of sugar permease activity.

As set out above, a permease polypeptide of the invention will typically have sugar permease activity. However, a permease polypeptide of the invention may have one or more of the activities set out above in addition to or alternative to that activity.

Polynucleotide sequence

With the permease polypeptide and its amino acid sequence as disclosed herein, the skilled person may determine suitable polynucleotides that encode the permease polypeptide.

In an embodiment the polynucleotide is a variant polynucleotide having at least 35%, at least 40%, at least 45%, at least 50%, at least 60%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99% sequence identity to SEQ ID NO: 56, and encodes the polypeptide as described in claims 1 to 11.

The invention therefore provides polynucleotide sequences comprising the gene encoding the permease polypeptide, as well as its coding sequence.

The polynucleotides of the invention may be isolated or synthesized. The permease polypeptides and permease polypeptide polynucleotides herein may be synthetic polypeptides, respectively polynucleotides. The synthetic polynucleotides may be optimized in codon use, preferably according to the methods described in
5 WO2006/077258 and/or PCT/EP2007/055943, which are herein incorporated by reference. PCT/EP2007/055943 addresses codon-pair optimization.

The term refers to a polynucleotide molecule, which is a ribonucleic acid (RNA) or deoxyribonucleic acid (DNA) molecule, either single stranded or double stranded. A polynucleotide may either be present in isolated form, or be comprised in recombinant
10 nucleic acid molecules or vectors, or be comprised in a host cell.

The word "polypeptide" is used herein for chains containing more than seven amino acid residues. All oligopeptide and polypeptide formulas or sequences herein are written from left to right and in the direction from amino terminus to carboxy terminus. The one-letter code of amino acids used herein is commonly known in the art.

15 By "isolated" polypeptide or protein is intended a polypeptide or protein removed from its native environment. For example, recombinantly produced polypeptides and proteins expressed in host cells are considered isolated for the purpose of the invention as are native or recombinant polypeptides which have been substantially purified by any suitable technique such as, for example, the single-step purification method disclosed in
20 Smith and Johnson, Gene 67:31-40 (1988).

The polynucleotides of the present invention, such as a polynucleotide encoding the permease polypeptide can be isolated or synthesized using standard molecular biology techniques and the sequence information provided herein.

25 The polynucleotide encoding the permease polypeptide of the invention can be amplified using cDNA, mRNA or alternatively, genomic DNA, as a template and appropriate oligonucleotide primers according to standard PCR amplification techniques. The nucleic acid so amplified can be cloned into an appropriate vector and characterized by DNA sequence analysis.

30

Transformation

The polynucleotides according to the invention may be expressed in a suitable host. The invention thus relates to a transformed host cell. In an embodiment, the host cell may be transformed with a nucleic acid construct that comprises a polynucleotide

that encodes the polypeptide according to the invention defined before. Therefore standard transformation techniques may be used.

The transformed host cell comprises at least one polynucleotide that encodes a mutant MFS polypeptide as described herein before.

5

Co-consumption

In an embodiment the transformed host is capable of co-consumption of glucose and at least one pentose. This pentose may be arabinose or xylose, in an embodiment it is xylose. Co-consumption (or co-fermentation) of two substrates is defined herein as a simultaneous uptake and intracellular conversion of two different carbon sources (e.g. xylose and glucose), at an appreciable level. Said carbon sources are simultaneously converted into products, such as e.g. biomass, ethanol, glycerol, and the like.

Co-consumption of a cell is herein quantified and expressed as co-consumption index. The co-consumption index is herein the co-consumption index for glucose and xylose and is calculated as the sum over the time interval of 0-24 hours (measured at 0, 8, 12, 14, 16, 18, 20, 22 and 24 hours) of the absolute difference of the glucose uptake rate (Q_g) and the specific uptake rate (Q_x), expressed as grams of sugar consumed per time unit, in an anaerobic batch culture fermentation at 1.0 g/l dry yeast pitch, 30 degrees C temperature and wherein the fermentation medium contains 71.8 grams of glucose per liter and 40.0 grams xylose per liter, at the start of the fermentation. See examples 16-18.

In an embodiment, the co-consumption index of the transformed host cell is 27 g/h or less, 25 g/h or less, 23 g/h or less, 20 g/h or less, 18 g/h or less, 16 g/h or less, 14 g/h or less, or 12 g/h or less,

25

In an embodiment, the transformed host cell comprises a heterologous nucleotide that encodes a polypeptide as defined before and described in claim 1 or encodes a polypeptide having substitution F85S/V, T89V/L, V187I, Y226A, Q338N, M339S/A, N346A/D, Y350I, G373N/S, N376S, or T449S of sequence ID NO: 59.

30

In an embodiment, the transformed host is transformed with a polynucleotide that encodes a polypeptide that is a mutant of a polypeptide that is native in the untransformed host cell.

In an embodiment, the polypeptide that is native in the untransformed host cell is a member of the the Major Facilitator Superfamily (MFS) transporters, in an embodiment a hexose transporter polypeptide..

In an embodiment, the polypeptide that is native in the untransformed host
5 cell is a transporter polypeptide chosen from the list consisting of Gal2, Hxt1, Hxt2, Hxt3, Hxt4, Hxt5, Hxt6, Hxt7, Hxt8, Hxt9, Hxt10, Hxt11, Hxt12, Hxt13, Hxt14, Hxt15, Hxt16 and Hxt17. In an embodiment, in the polypeptide of the invention has not the amino acid residue X (X may be any amino acid) at a given position A (A may be any position in the polypeptide) in SEQ ID NO:59, where X is a mutation in SEQ ID NO: 59, when X is
10 native at the to A corresponding position in a second MFS family polypeptide.

In an embodiment, the polypeptide has not the amino acid residue S, that corresponds to F85S in SEQ ID NO:59, in the corresponding position in HXT8 (i.e. not 80S in HXT8).

In an embodiment, the polypeptide has not the amino acid residue A, that
15 corresponds to V187A in SEQ ID NO:59, in the corresponding position in HXT8 (i.e. not 183A in HXT8), in HXT9 (i.e. not 177A in HXT9), in HXT 13 (i.e. not 174A in HXT13) in HXT 15 (i.e. not 177A in HXT15) in HXT16 (i.e. not 177A in HXT16) or in HXT17 (i.e. not 174A in HXT17).

In an embodiment, the polypeptide has not the amino acid residue V, that
20 corresponds to I218V in SEQ ID NO:59, in the corresponding position in HXT8 (i.e. not 214V in HXT8), in HXT9 (i.e. not 208V in HXT9), in HXT 10 (i.e. not 196V in HXT10) in or in HXT 14 (i.e. not 215V in HXT14).

In an embodiment, the polypeptide has not the amino acid residue N, that corresponds to Q338N in SEQ ID NO:59, in the corresponding position in HXT8 (i.e. not
25 334N in HXT8), or in HXT9 (i.e. not 328N in HXT9).

In an embodiment, the polypeptide has not the amino acid residue S, that corresponds to M339S in SEQ ID NO:59, in the corresponding position in HXT1 (i.e. not 333S in HXT1), in HXT2 (i.e. not 324S in HXT2), in HXT3 (i.e. not 330S in HXT3), in HXT4 (i.e. not 339S in HXT4), in HXT5 (i.e not 354S in HXT5), in HXT6 or in HXT7 (i.e.
30 not 333S in HXT6 or HXT7), in HXT8 (i.e. not 335S in HXT8), in HXT9 (i.e. not 329S in HXT9), in HXT10 (i.e. not 317S in HXT10) or in HXT11 (i.e. not 329S in HXT11).

In another embodiment, the polypeptide has not the amino acid residue D, that corresponds to N346D in SEQ ID NO:59, in the corresponding position in HXT1 (i.e. not 340D in HXT1), in HXT3 (i.e. not 337D in HXT3), in HXT4 (i.e. not 346D in HXT4), in

HXT5 (i.e. not 361D in HXT5), in HXT6 or HXT7 (i.e. not 340D in HXT6 or HXT7), in HXT8 (i.e. not 342D in HXT8), in HXT 9 (i.e. not 336D in HXT9), or in HXT11 (i.e. not 336D in HXT11).

5 In an embodiment, the polypeptide has not the amino acid residue S, that corresponds to N346I in SEQ ID NO:59, in the corresponding position in HXT14 (i.e. not 357I in HXT14),

10 In an embodiment, the polypeptide has not 352Y or 376S or 446Y or 504F in GAL2. In an embodiment, the polypeptide has not 200A or 209V or 324I or 388M or 389W or 392V or 414Q or 415N or 449V or 471I in HXT3). In an embodiment, the polypeptide has not 66F or 66S or 126C or 126R or 295M or 310T or 379F or 388W in GLUT1).

In an embodiment, the polypeptide has not the amino acid residue I, that corresponds to N346I in SEQ ID NO:59, in the corresponding position in HXT14 (i.e. not 357I in HXT14),

15 In another embodiment the polypeptide has not the amino acid residue S, that corresponds to T380S in SEQ ID NO:59 in HXT11 (i.e. not 370S in HXT11).

In another embodiment the polypeptide has not the amino acid residue S, that corresponds to F350S in SEQ ID NO:59 in HXT14 (i.e. not 384S in HXT14).

20 In an embodiment, the polypeptide has not the amino acid residue S, that corresponds to T449S in SEQ ID NO:59, in the corresponding position in HXT2 (i.e. not 434S in HXT2),

In an embodiment, the polypeptide has not the amino acid residue G, that corresponds to A451G in SEQ ID NO:59, in the corresponding position in HXT8 (i.e. not 447G in HXT8), or in HXT9 (i.e. not 441G in HXT9).

25 In an embodiment the polypeptide has not the amino acid residue Q, that corresponds to A451Q in SEQ ID NO:59 in HXT14 (i.e. not 464Q in HXT14).

30 In an embodiment, the polypeptide is a mutant HXT3-6 and has one or more substitutions in HXT3-6, In an embodiment thereof, the polypeptide has substitutions corresponding to N367A/C/D/F/G/I/L/M/S/T/V of SEQ ID NO: 126. In an embodiment, the polypeptide has substitutions corresponding to N367A/I of SEQ ID NO: 126. In an embodiment, the polypeptide has substitutions corresponding to N367A of SEQ ID NO: 126.

In an embodiment, the polypeptide is a mutant HXT11 and has one or more substitutions in HXT11, In an embodiment thereof, the polypeptide has substitutions

corresponding to N366A/C/D/F/G/I/L/M/S/T/V of SEQ ID NO: 123. In an embodiment the polypeptide is a mutant HXT11. In an embodiment the polypeptide has substitutions corresponding to N366/F/I/L/M/T/V of SEQ ID NO: 123. In an embodiment, the polypeptide has substitutions corresponding to N366D/M/T of SEQ ID NO: 123. In an embodiment, the polypeptide has substitutions corresponding to N366M/T of SEQ ID NO: 123. In an embodiment the polypeptide has substitutions corresponding to N366M/T or N366T of SEQ ID NO: 123.

In an embodiment the polypeptide has not 333S in HXT6. In an embodiment the polypeptide has not 376S in GAL2.

10 The invention further relates to a nucleic acid construct comprising the polynucleotide as described before, e.g. a vector.

Another aspect of the invention thus pertains to vectors, including cloning and expression vectors, comprising a polynucleotide of the invention encoding a permease polypeptide protein or a functional equivalent thereof and methods of growing, transforming or transfecting such vectors in a suitable host cell, for example under conditions in which expression of a permease of the invention occurs. As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked.

Polynucleotides of the invention can be incorporated into a recombinant replicable vector, for example a cloning or expression vector. The vector may be used to replicate the nucleic acid in a compatible host cell. Thus in a further embodiment, the invention provides a method of making polynucleotides of the invention by introducing a polynucleotide of the invention into a replicable vector, introducing the vector into a compatible host cell, and growing the host cell under conditions which bring about replication of the vector. The vector may be recovered from the host cell. Suitable host cells are described below.

It will be appreciated by those skilled in the art that the design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression of protein desired, etc. The vectors, such as expression vectors, of the invention can be introduced into host cells to thereby produce proteins or peptides, encoded by nucleic acids as described herein. The vectors, such as recombinant expression vectors, of the invention can be designed for expression of permease polypeptide proteins in prokaryotic or eukaryotic cells.

For example, permease polypeptides can be expressed in bacterial cells such as *E. coli*, insect cells (using baculovirus expression vectors), filamentous fungi, yeast cells or mammalian cells. Suitable host cells are discussed further in Goeddel, *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, CA
5 (1990). Representative examples of appropriate hosts are described hereafter.

Appropriate culture mediums and conditions for the above-described host cells are known in the art.

For most filamentous fungi and yeast, the vector or expression construct is preferably integrated in the genome of the host cell in order to obtain stable
10 transformants. However, for certain yeasts also suitable episomal vectors are available into which the expression construct can be incorporated for stable and high level expression, examples thereof include vectors derived from the 2 μ and pKD1 plasmids of *Saccharomyces* and *Kluyveromyces*, respectively, or vectors containing an AMA sequence (e.g. AMA1 from *Aspergillus*). In case the expression constructs are integrated
15 in the host cells genome, the constructs are either integrated at random loci in the genome, or at predetermined target loci using homologous recombination, in which case the target loci preferably comprise a highly expressed gene.

Accordingly, expression vectors useful in the present invention include chromosomal-, episomal- and virus-derived vectors e.g., vectors derived from bacterial
20 plasmids, bacteriophage, yeast episome, yeast chromosomal elements, viruses such as baculoviruses, papova viruses, vaccinia viruses, adenoviruses, fowl pox viruses, pseudorabies viruses and retroviruses, and vectors derived from combinations thereof, such as those derived from plasmid and bacteriophage genetic elements, such as cosmids and phagemids.

25 When the polypeptide according to the invention is to be secreted from the host cell into the cultivation medium, an appropriate signal sequence can be added to the polypeptide in order to direct the *de novo* synthesized polypeptide to the secretion route of the host cell. The person skilled in the art knows to select an appropriate signal sequence for a specific host.

30 The vector may further include sequences flanking the polynucleotide giving rise to RNA which comprise sequences homologous to eukaryotic genomic sequences or viral genomic sequences. This will allow the introduction of the polynucleotides of the invention into the genome of a host cell.

An integrative cloning vector may integrate at random or at a predetermined target locus in the chromosome(s) of the host cell into which it is to be integrated.

The vector system may be a single vector, such as a single plasmid, or two or more vectors, such as two or more plasmids, which together contain the total DNA to be introduced into the genome of the host cell.

The vector may contain a polynucleotide of the invention oriented in an antisense direction to provide for the production of antisense RNA.

Vector DNA can be introduced into prokaryotic or eukaryotic cells via conventional transformation or transfection techniques. As used herein, the terms "transformation" and "transfection" are intended to refer to a variety of art-recognized techniques for introducing foreign nucleic acid (e.g., DNA) into a host cell, including calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, transduction, infection, lipofection, cationic lipid-mediated transfection or electroporation. Suitable methods for transforming or transfecting host cells can be found in Sambrook, et al. (*Molecular Cloning: A Laboratory Manual*, 2nd, ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989), Davis et al., *Basic Methods in Molecular Biology* (1986) and other laboratory manuals.

As indicated before, the invention provides an isolated polypeptide having the amino acid sequence according to SEQ ID NO: 59 with the mutations indicated in claim 1.

The permease polypeptide according to the invention can be recovered and purified from recombinant cell cultures by methods known in the art. Most preferably, high performance liquid chromatography ("HPLC") is employed for purification.

Polypeptides of the present invention include naturally purified products, products of chemical synthetic procedures, and products produced by recombinant techniques from a prokaryotic or eukaryotic host, including, for example, bacterial, yeast, higher plant, insect and mammalian cells. Depending upon the host employed in a recombinant production procedure, the polypeptides of the present invention may be glycosylated or may be non-glycosylated. In addition, polypeptides of the invention may also include an initial modified methionine residue, in some cases as a result of host-mediated processes.

The invention also features biologically active fragments of the polypeptides according to the invention.

Provided also are host cells, comprising a polynucleotide or vector of the invention. The polynucleotide may be heterologous to the genome of the host cell. The term "heterologous", usually with respect to the host cell, means that the polynucleotide does not naturally occur in the genome of the host cell or that the polypeptide is not
5 naturally produced by that cell.

In another embodiment, the invention features cells, e.g., transformed host cells or recombinant host cells that contain a nucleic acid encompassed by the invention. A "transformed cell" or "recombinant cell" is a cell into which (or into an ancestor of which) has been introduced, by means of recombinant DNA techniques, a nucleic acid
10 according to the invention. Both prokaryotic and eukaryotic cells are included, e.g., bacteria, fungi, yeast, and the like, especially preferred are yeast cells including e.g. *Saccharomyces*, for example *Saccharomyces cerevisiae*.

A host cell can be chosen that modulates the expression of the inserted sequences, or modifies and processes the gene product in a specific, desired fashion.
15 Such modifications (e.g., glycosylation) and processing (e.g., cleavage) of protein products may facilitate optimal functioning of the protein.

Various host cells have characteristic and specific mechanisms for post-translational processing and modification of proteins and gene products. Appropriate cell lines or host systems familiar to those of skill in the art of molecular biology and/or
20 microbiology can be chosen to ensure the desired and correct modification and processing of the foreign protein expressed. To this end, eukaryotic host cells that possess the cellular machinery for proper processing of the primary transcript, glycosylation, and phosphorylation of the gene product can be used. Such host cells are well known in the art.

25 If desired, a cell as described above may be used to in the preparation of a polypeptide according to the invention. Such a method typically comprises cultivating a host cell (e. g. transformed or transfected with an expression vector as described above) under conditions to provide for expression (by the vector) of a coding sequence encoding the polypeptide, and optionally recovering the expressed polypeptide.
30 Polynucleotides of the invention can be incorporated into a recombinant replicable vector, e. g. an expression vector. The vector may be used to replicate the nucleic acid in a compatible host cell. Thus in a further embodiment, the invention provides a method of making a polynucleotide of the invention by introducing a polynucleotide of the invention into a replicable vector, introducing the vector into a compatible host cell, and

growing the host cell under conditions which bring about the replication of the vector. The vector may be recovered from the host cell.

The vectors may be transformed or transfected into a suitable host cell as described above to provide for expression of a polypeptide of the invention. This process
5 may comprise culturing a host cell transformed with an expression vector as described above under conditions to provide for expression by the vector of a coding sequence encoding the polypeptide.

Herein standard isolation, hybridization, transformation and cloning techniques are used (e. g., as described in Sambrook, J., Fritsh, E. F., and Maniatis, T. Molecular
10 Cloning: A Laboratory Manual.2nd, ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989).

Homology & Identity

Amino acid or nucleotide sequences are said to be homologous when exhibiting
15 a certain level of similarity. Two sequences being homologous indicate a common evolutionary origin. Whether two homologous sequences are closely related or more distantly related is indicated by “percent identity” or “percent similarity”, which is high or low respectively. Although disputed, to indicate “percent identity” or “percent similarity”, “level of homology” or “percent homology” are frequently used interchangeably.

20 A comparison of sequences and determination of percent identity between two sequences can be accomplished using a mathematical algorithm. The skilled person will be aware of the fact that several different computer programs are available to align two sequences and determine the homology between two sequences (Kruskal, J. B. (1983) An overview of sequence comparison In D. Sankoff and J. B. Kruskal, (ed.), Time warps, string edits and macromolecules: the theory and practice of sequence comparison, pp.
25 1-44 Addison Wesley). The percent identity between two amino acid sequences can be determined using the Needleman and Wunsch algorithm for the alignment of two sequences. (Needleman, S. B. and Wunsch, C. D. (1970) J. Mol. Biol. 48, 443-453). The algorithm aligns amino acid sequences as well as nucleotide sequences. The
30 Needleman-Wunsch algorithm has been implemented in the computer program NEEDLE. For the purpose of this invention the NEEDLE program from the EMBOSSE package was used (version 2.8.0 or higher, *EMBOSS: The European Molecular Biology Open Software Suite* (2000) Rice,P. Longden,I. and Bleasby,A. Trends in Genetics 16, (6) pp276—277, <http://emboss.bioinformatics.nl/>). For protein sequences, EBLOSUM62

is used for the substitution matrix. For nucleotide sequences, EDNAFULL is used. Other matrices can be specified. The optional parameters used for alignment of amino acid sequences are a gap-open penalty of 10 and a gap extension penalty of 0.5. The skilled person will appreciate that all these different parameters will yield slightly different results but that the overall percentage identity of two sequences is not significantly altered when
5 using different algorithms.

Global Homology Definition

The homology or identity is the percentage of identical matches between the two full sequences over the total aligned region including any gaps or extensions. The
10 homology or identity between the two aligned sequences is calculated as follows: Number of corresponding positions in the alignment showing an identical amino acid in both sequences divided by the total length of the alignment including the gaps. The identity defined as herein can be obtained from NEEDLE and is labelled in the output of
15 the program as "IDENTITY".

Longest Identity Definition

The homology or identity between the two aligned sequences is calculated as follows: Number of corresponding positions in the alignment showing an identical amino
20 acid in both sequences divided by the total length of the alignment after subtraction of the total number of gaps in the alignment. The identity defined as herein can be obtained from NEEDLE by using the NOBRIEF option and is labelled in the output of the program as "longest-identity".

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

The sugar composition

The sugar composition according to the invention comprises glucose, arabinose
30 and xylose. Any sugar composition may be used in the invention that suffices those criteria. Optional sugars in the sugar composition are galactose and mannose. In a preferred embodiment, the sugar composition is a hydrolysate of one or more lignocellulosic material. Lignocellulose herein includes hemicellulose and hemicellulose parts of biomass. Also lignocellulose includes lignocellulosic fractions of biomass.

Suitable lignocellulosic materials may be found in the following list: orchard primings, chaparral, mill waste, urban wood waste, municipal waste, logging waste, forest thinnings, short-rotation woody crops, industrial waste, wheat straw, oat straw, rice straw, barley straw, rye straw, flax straw, soy hulls, rice hulls, rice straw, corn gluten feed, oat hulls, sugar cane, corn stover, corn stalks, corn cobs, corn husks, switch grass, miscanthus, sweet sorghum, canola stems, soybean stems, prairie grass, gamagrass, foxtail; sugar beet pulp, citrus fruit pulp, seed hulls, cellulosic animal wastes, lawn clippings, cotton, seaweed, trees, softwood, hardwood, poplar, pine, shrubs, grasses, wheat, wheat straw, sugar cane bagasse, corn, corn husks, corn hobs, corn kernel, fiber from kernels, products and by-products from wet or dry milling of grains, municipal solid waste, waste paper, yard waste, herbaceous material, agricultural residues, forestry residues, municipal solid waste, waste paper, pulp, paper mill residues, branches, bushes, canes, corn, corn husks, an energy crop, forest, a fruit, a flower, a grain, a grass, a herbaceous crop, a leaf, bark, a needle, a log, a root, a sapling, a shrub, switch grass, a tree, a vegetable, fruit peel, a vine, sugar beet pulp, wheat midlings, oat hulls, hard or soft wood, organic waste material generated from an agricultural process, forestry wood waste, or a combination of any two or more thereof.

An overview of some suitable sugar compositions derived from lignocellulose and the sugar composition of their hydrolysates is given in table 1. The listed lignocelluloses include: corn cobs, corn fiber, rice hulls, melon shells, sugar beet pulp, wheat straw, sugar cane bagasse, wood, grass and olive pressings.

Table 1: Overview of sugar compositions from lignocellulosic materials. Gal=galactose, Xyl=xylose, Ara=arabinose, Man=mannose, Glu=glucose, Rham=rhamnose.

Lignocellulosic material	Gal	Xyl	Ara	Man	Glu	Rham	Sum
Corn cob a	10	286	36		227	11	570
Corn cob b	131	228	160		144		663
Rice hulls a	9	122	24	18	234	10	417
Rice hulls b	8	120	28		209	12	378
Melon Shells	6	120	11		208	16	361
Sugar beet pulp	51	17	209	11	211	24	523
Wheat straw Idaho	15	249	36		396		696

Corn fiber	36	176	113		372		697
Cane Bagasse	14	180	24	5	391		614
Corn stover	19	209	29		370		626
Athel (wood)	5	118	7	3	493		625
Eucalyptus (wood)	22	105	8	3	445		583
CWR (grass)	8	165	33		340		546
JTW (grass)	7	169	28		311		515
MSW	4	24	5	20	440		493
Reed Canary Grass Veg	16	117	30	6	209	1	379
Reed Canary Grass Seed	13	163	28	6	265	1	476
Olive pressing residu	15	111	24	8	329		487

It is clear from table 1 that in these lignocelluloses a high amount of sugar is presence in de form of glucose, xylose, arabinose and galactose. The conversion of glucose, xylose, arabinose and galactose to fermentation product is thus of great economic importance. Also mannose is present in some lignocellulose materials be it usually in lower amounts than the previously mentioned sugars. Advantageously therefore also mannose is converted by the transformed host cell.

The transformed host cell

In an embodiment, the transformed host cell may comprise one or more copies of xylose isomerase gene and/or one or more copies of xylose reductase and/or xylitol dehydrogenase, and two to ten copies of *araA*, *araB* and *araD*, genes, wherein these genes are integrated into the cell genome.

In one embodiment, the transformed host cell comprises genes, for example the above xylose isomerase gene and/or one or more copies of xylose reductase and/or xylitol dehydrogenase, and two to ten copies of *araA*, *araB* and *araD*, genes, are integrated into the transformed host cell genome.

The number of copies may be determined by the skilled person by any known method. In the examples, a suitable method is described.

In an embodiment, the transformed host cell is able to ferment glucose, arabinose, xylose and galactose.

In an embodiment, the cell is capable of converting 90% or more glucose, xylose arabinose, galactose and mannose available, into a fermentation product. In an embodiment, cell is capable of converting 91% or more, 92% or more, 94% or more, 95% or more, 96% or more, 97% or more, 98% or more or 100% of all glucose, xylose
5 arabinose, galactose and mannose available, into a fermentation product.

In one embodiment of the invention the transformed host cell is able to ferment one or more additional sugar, preferably C5 and/or C6 sugar e.g. mannose. In an embodiment of the invention the transformed host cell comprises one or more of: a *xyIA*-gene, *XYL1* gene and *XYL2* gene and/or *XKS1*-gene, to allow the transformed host cell
10 to ferment xylose; deletion of the aldose reductase (*GRE3*) gene; overexpression of PPP-genes *TAL1*, *TKL1*, *RPE1* and *RKI1* to allow the increase of the flux through the pentose phosphate pathway in the cell.

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

In an embodiment the transformed host cell is inhibitor tolerant. Inhibitor
30 tolerance is resistance to inhibiting compounds. The presence and level of inhibitory compounds in lignocellulose may vary widely with variation of feedstock, pretreatment method hydrolysis process. Examples of categories of inhibitors are carboxylic acids, furans and/or phenolic compounds. Examples of carboxylic acids are lactic acid, acetic acid or formic acid. Examples of furans are furfural and hydroxy- methylfurfural.

Examples of phenolic compounds are vanillin, syringic acid, ferulic acid and coumaric acid. The typical amounts of inhibitors are for carboxylic acids: several grams per liter, up to 20 grams per liter or more, depending on the feedstock, the pretreatment and the hydrolysis conditions. For furans: several hundreds of milligrams per liter up to several grams per liter, depending on the feedstock, the pretreatment and the hydrolysis conditions.

For phenolics: several tens of milligrams per liter, up to a gram per liter, depending on the feedstock, the pretreatment and the hydrolysis conditions.

The transformed host cells according to the invention may be inhibitor tolerant, i.e. they can withstand common inhibitors at the level that they typically have with common pretreatment and hydrolysis conditions, so that the transformed host cells can find broad application, i.e. it has high applicability for different feedstock, different pretreatment methods and different hydrolysis conditions.

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

In an embodiment, the transformed host cell is marker-free. As used herein, the term "marker" refers to a gene encoding a trait or a phenotype which permits the selection of, or the screening for, a host cell containing the marker. Marker-free means that markers are essentially absent in the transformed host cell. Being marker-free is particularly advantageous when antibiotic markers have been used in construction of the transformed host cell and are removed thereafter. Removal of markers may be done using any suitable prior art technique, e.g. intramolecular recombination. A suitable method of marker removal is illustrated in the examples.

A transformed host cell may be able to convert plant biomass, celluloses, hemicelluloses, pectins, starch, starch derivatives, for example into fermentable sugars. Accordingly, a transformed host cell may express one or more enzymes such as a cellulase (an endocellulase or an exocellulase), a hemicellulase (an endo- or exo-xylanase or arabinase) necessary for the conversion of cellulose into glucose monomers and hemicellulose into xylose and arabinose monomers, a pectinase able to convert

pectins into glucuronic acid and galacturonic acid or an amylase to convert starch into glucose monomers.

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

In an embodiment, the transformed host cell is a cell that is naturally capable of alcoholic fermentation, preferably, anaerobic alcoholic fermentation. A transformed host cell preferably has a high tolerance to ethanol, a high tolerance to low pH (i.e. capable of growth at a pH lower than about 5, about 4, about 3, or about 2.5) and towards organic and/or a high tolerance to elevated temperatures.

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

Construction of the transformed host cell

According to an embodiment, the genes may be introduced in the host cell by introduction into a host cell:

- a) a cluster consisting of the genes *araA*, *araB* and *araD* under control of a strong constitutive promoter
- b) a cluster consisting of PPP-genes *TAL1*, *TKL1*, *RPE1* and *RKI1*, optionally under control of strong constitutive promoter; and deletion of an aldose reductase gene;
- c) a cluster consisting of a *xylA*-gene and a *XKS1*-gene under control of strong constitutive promoter;
- d) a construct comprising a *xylA* gene under control of a strong constitutive promoter, which has the ability to integrate into the genome on multiple loci;

and adaptive evolution to produce the transformed host cell. The above cell may be constructed using recombinant expression techniques.

Recombinant expression

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

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

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

The nucleic acid construct may be maintained episomally and thus comprise a sequence for autonomous replication, such as an autosomal replication sequence. A suitable episomal nucleic acid construct may e.g. be based on the yeast 2 μ or pKD1 plasmids (Gleer *et al.*, 1991, Biotechnology 9: 968-975), or the AMA plasmids (Fierro *et al.*, 1995, Curr Genet. 29:482-489). Alternatively, each nucleic acid construct may be integrated in one or more copies into the genome of the cell. Integration into the cell's genome may occur at random by non-homologous recombination but preferably, the nucleic acid construct may be integrated into the cell's genome by homologous recombination as is well known in the art (see e.g. WO90/14423, EP-A-0481008, EP-A-0635 574 and US 6,265,186).

Most episomal or 2 μ plasmids are relatively unstable in yeast, being lost in approximately 10⁻² or more cells after each generation. Even under conditions of selective growth, only 60% to 95% of the cells retain the episomal plasmid. The copy number of most episomal plasmids ranges from 20-100 per cell of cir⁺ hosts. However, the plasmids are not equally distributed among the cells, and there is a high variance in

the copy number per cell in populations. Strains transformed with integrative plasmids are extremely stable, even in the absence of selective pressure. However, plasmid loss can occur at approximately 10^{-3} to 10^{-4} frequencies by homologous recombination between tandemly repeated DNA, leading to looping out of the vector sequence. Preferably, the vector design in the case of stable integration is thus, that upon loss of the selection marker genes (which also occurs by intramolecular, homologous recombination) that looping out of the integrated construct is no longer possible. Preferably the genes are thus stably integrated. Stable integration is herein defined as integration into the genome, wherein looping out of the integrated construct is no longer possible. Preferably selection markers are absent. Typically, the enzyme encoding sequence will be operably linked to one or more nucleic acid sequences, capable of providing for or aiding the transcription and/or translation of the enzyme sequence.

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

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

The promoter that could be used to achieve the expression of a nucleotide sequence coding for an enzyme according to the present invention, may be not native to the nucleotide sequence coding for the enzyme to be expressed, i.e. a promoter that is heterologous to the nucleotide sequence (coding sequence) to which it is operably linked. The promoter may, however, be homologous, i.e. endogenous, to the host cell.

Promoters are widely available and known to the skilled person. Suitable examples of such promoters include e.g. promoters from glycolytic genes, such as the phosphofructokinase (PFK), triose phosphate isomerase (TPI), glyceraldehyde-3 - phosphate dehydrogenase (GPD, TDH3 or GAPDH), pyruvate kinase (PYK), phosphoglycerate kinase (PGK) promoters from yeasts or filamentous fungi; more

details about such promoters from yeast may be found in (WO 93/03159). Other useful promoters are ribosomal protein encoding gene promoters, the lactase gene promoter (LAC4), alcohol dehydrogenase promoters (ADH1, ADH4, and the like), and the enolase promoter (ENO). Other promoters, both constitutive and inducible, and enhancers or upstream activating sequences will be known to those of skill in the art. The promoters used in the host cells of the invention may be modified, if desired, to affect their control characteristics. Suitable promoters in this context include both constitutive and inducible natural promoters as well as engineered promoters, which are well known to the person skilled in the art. Suitable promoters in eukaryotic host cells may be *GAL7*, *GAL10*, or *GAL1*, *CYC1*, *HIS3*, *ADH1*, *PGL*, *PH05*, *GAPDH*, *ADC1*, *TRP1*, *URA3*, *LEU2*, *ENO1*, *TPI1*, and *AOX1*. Other suitable promoters include *PDC1*, *GPD1*, *PGK1*, *TEF1*, and *TDH3*.

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

The transcription termination sequence further preferably comprises a polyadenylation signal.

Optionally, a selectable marker may be present in a nucleic acid construct suitable for use in the invention. As used herein, the term "marker" refers to a gene encoding a trait or a phenotype which permits the selection of, or the screening for, a host cell containing the marker. The marker gene may be an antibiotic resistance gene whereby the appropriate antibiotic can be used to select for transformed cells from among cells that are not transformed. Examples of suitable antibiotic resistance markers include e.g. dihydrofolate reductase, hygromycin-B-phosphotransferase, 3'-O-phosphotransferase II (kanamycin, neomycin and G418 resistance). Antibiotic resistance markers may be most convenient for the transformation of polyploid host cells. Also non-antibiotic resistance markers may be used, such as auxotrophic markers (*URA3*, *TRP1*, *LEU2*) or the *S. pombe* *TPI* gene (described by Russell P R, 1985, Gene 40: 125-130).

In a preferred embodiment the host cells transformed with the nucleic acid constructs are marker gene free. Methods for constructing recombinant marker gene free microbial host cells are disclosed in EP-A-O 635 574 and are based on the use of bidirectional markers such as the *A. nidulans amdS* (acetamidase) gene or the yeast *URA3* and *LYS2* genes.
5 Alternatively, a screenable marker such as Green Fluorescent Protein, lacL, luciferase, chloramphenicol acetyltransferase, beta-glucuronidase may be incorporated into the nucleic acid constructs of the invention allowing to screen for transformed cells.

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

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

Alternatively, overexpression of enzymes in the host cells of the invention may be achieved by using a promoter that is not native to the sequence coding for the enzyme to be overexpressed, i.e. a promoter that is heterologous to the coding sequence to which it is operably linked. Although the promoter preferably is heterologous to the coding
25 sequence to which it is operably linked, it is also preferred that the promoter is homologous, i.e. endogenous to the host cell. Preferably the heterologous promoter is capable of producing a higher steady state level of the transcript comprising the coding sequence (or is capable of producing more transcript molecules, i.e. mRNA molecules, per unit of time) than is the promoter that is native to the coding sequence. Suitable
30 promoters in this context include both constitutive and inducible natural promoters as well as engineered promoters.

In an embodiment, the transformed host cell is markerfree, which means that no auxotrophic or dominant markers, in particular antibiotic resistance markers, are present in the genome or extra-chromosomally.

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

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

Adaptation

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

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

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

Fitness

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

Genetic changes

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

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

The adaptive evolution

The transformed host cells may in their preparation be subjected to adaptive evolution. A transformed host cell may be adapted to sugar utilisation by selection of mutants, either spontaneous or induced (e.g. by radiation or chemicals), for growth on the desired sugar, preferably as sole carbon source, and more preferably under anaerobic conditions. Selection of mutants may be performed by techniques including serial transfer of cultures as e.g. described by Kuyper *et al.* (2004, FEMS Yeast Res. 4: 655-664) or by cultivation under selective pressure in a chemostat culture. E.g. in a preferred host cell at least one of the genetic modifications described above, including modifications obtained by selection of mutants, confer to the host cell the ability to grow on the xylose as carbon source, preferably as sole carbon source, and preferably under anaerobic conditions. When XI is used as gene to convert xylose, preferably the cell produce essentially no xylitol, e.g. the xylitol produced is below the detection limit or e.g. less than about 5, about 2, about 1, about 0.5, or about 0.3 % of the carbon consumed on a molar basis.

Adaptive evolution is also described e.g. in Wisselink H.W. *et al* (2007) Appl. Environ. Microbiol. 73:4881–4891.

In one embodiment of adaptive evolution a regimen consisting of repeated batch cultivation with repeated cycles of consecutive growth in different media is applied, e.g. three media with different compositions (glucose, xylose, and arabinose; xylose and arabinose. See Wisselink *et al.* (2009) Applied and Environ. Microbiol. 75: 907–914.

Yeast transformation and genetic stability

Genetic engineering, i.e. transformation of yeast cells with recombinant DNA, became feasible for the first time in 1978 [Beggs, 1978; Hinnen *et al.*, 1978]. Recombinant DNA technology in yeast has established itself since then. A multitude of different vector constructs are available. Generally, these plasmid vectors, called shuttle vectors, contain genetic material derived from *E.coli* vectors consisting of an origin of replication and a selectable marker (often the β lactamase gene, *ampR*), which enable them to be propagated in *E.coli* prior to transformation into yeast cells. Additionally, the shuttle vectors contain a selectable marker for selection in yeast. Markers can be genes encoding enzymes for the synthesis of a particular amino acid or nucleotide, so that cells carrying the corresponding genomic deletion (or mutation) are complemented for auxotrophy or autotrophy. Alternatively, these vectors contain heterologous dominant resistance markers, which provides recombinant yeast cells (i.e. the cells that have taken up the DNA and express the marker gene) resistance towards certain antibiotics, like g418 (Geneticin), hygromycinB or phleomycin. In addition, these vectors may contain a sequence of (combined) restriction sites (multiple cloning site or MCS) which will allow to clone foreign DNA into these sites, although alternative methods exist as well.

Traditionally, four types of shuttle vectors can be distinguished by the absence or presence of additional genetic elements:

- Integrative plasmids (YIp) which by homologous recombination are integrated into the host genome at the locus of the marker or another gene, when this is opened by restriction and the linearized DNA is used for transformation of the yeast cells. This generally results in the presence of one copy of the foreign DNA inserted at this particular site in the genome.

- Episomal plasmids (YEp) which carry part of the 2 μ plasmid DNA sequence necessary for autonomous replication in yeast cells. Multiple copies of the transformed plasmid are propagated in the yeast cell and maintained as episomes.
- Autonomously replicating plasmids (YRp) which carry a yeast origin of replication (ARS, autonomously replicated sequence) that allows the transformed plasmids to be propagated several hundred-fold.
- CEN plasmids (YCp) which carry in addition to an ARS sequence a centromeric sequence (derived from one of the nuclear chromosomes) which normally guarantees stable mitotic segregation and usually reduces the copy number of self-replicated plasmid to just one.

These plasmids are being introduced into the yeast cells by transformation. Transformation of yeast cells may be achieved by several different techniques, such as permeabilization of cells with lithium acetate (Ito et al, 1983) and electroporation methods.

In commercial application of recombinant microorganisms, plasmid instability is the most important problem. Instability is the tendency of the transformed cells to lose their engineered properties because of changes to, or loss of, plasmids. This issue is discussed in detail by Zhang et al (Plasmid stability in recombinant *Saccharomyces cerevisiae*. Biotechnology Advances, Vol. 14, No. 4, pp. 401-435, 1996). Strains transformed with integrative plasmids are extremely stable, even in the absence of selective pressure (Sherman, F. http://dbb.urmc.rochester.edu/bs/sherman_f/yeast/9.html and references therein).

The heterologous DNA is usually introduced into the organism in the form of extra-chromosomal plasmids (YEp, YCp and YRp). Unfortunately, it has been found with both bacteria and yeasts that the new characteristics may not be retained, especially if the selection pressure is not applied continuously. This is due to the segregational instability of the hybrid plasmid when recombinant cells grow for a long period of time. This leads to population heterogeneity and clonal variability, and eventually to a cell population in which the majority of the cells has lost the properties that were introduced by transformation. If vectors with auxotrophic markers are being used, cultivation in rich media often leads to rapid loss of the vector, since the vector is only retained in minimal media. The alternative, the use of dominant antibiotic resistance markers, is often not compatible with production processes. The use of antibiotics may not be desired from a

registration point of view (the possibility that trace amounts of the antibiotic end up in the end product) or for economic reasons (costs of the use of antibiotics at industrial scale).

Loss of vectors leads to problems in large scale production situations. Alternative
5 methods for introduction of DNA do exist for yeasts, such as the use of integrating
plasmids (YIp). The DNA is integrated into the host genome by recombination, resulting
in high stability. (Caunt, P. (1988) Stability of recombinant plasmids in yeast. Journal of
Biotechnology 9: 173 – 192). We have found that an integration method using the host
transposons are a good alternative. In an embodiment genes may be integrated into the
10 transformed host cell genome. Initial introduction (i.e. before adaptive evolution) of
multiple copies be executed in any way known in the art that leads to introduction of the
genes. In an embodiment, this may be accomplished using a vector with parts
homologous to repeated sequences (transposons), of the host cell. When the host cell is
a yeast cell, suitable repeated sequences are the long terminal repeats (LTR) of the Ty
15 element, known as delta sequence. Ty elements fall into two rather similar subfamilies
called Ty1 and Ty2. These elements are about 6 kilobases (kb) in length and are
bounded by long terminal repeats (LTR), sequences of about 335 base pairs (Boeke JD
et al., (1988) The *Saccharomyces cerevisiae* Genome Contains Functional and
Nonfunctional Copies of Transposon Ty1. Molecular and Cellular Biology 8: 1432-1442).
20 In the fully sequenced *S. cerevisiae* strain, S288c, the most abundant transposons are
Ty1 (31 copies) and Ty2 (13 copies) (Gabriel A, Daprich J, Kunkel M, Gresham D, Pratt
SC, *et al.* (2006) Global mapping of transposon location. PLoS Genet 2: e212). These
transposons consist of two overlapping open reading frames (ORFs), each of which
encode several proteins. The coding regions are flanked by the aforementioned, nearly
25 identical LTRs. Other, but less abundant and more distinct Ty elements in *S. cerevisiae*
comprise Ty3, Ty4 and Ty5. For each family of full-length Ty elements there are an order
of magnitude more solo LTR elements dispersed through the genome. These are
thought to arise by LTR–LTR recombination of full-length elements, with looping out of
the internal protein encoding regions.

30 The retrotransposition mechanism of the Ty retrotransposon has been exploited
to integrate multiple copies throughout the genome (Boeke *et al.*, 1988; Jacobs *et al.*,
1988). The long terminal repeats (LTR) of the Ty element, known as delta sequences,
are also good targets for integration by homologous recombination as they exist in about
150-200 copies that are either Ty associated or solo sites (Boeke, 1989; Kingsman and

Kingsman, 1988). (Parekh R.N. (1996). An Integrating Vector for Tunable, High Copy, Stable Integration into the Dispersed Ty DELTA Sites of *Saccharomyces cerevisiae*. Biotechnol. Prog. 12:16-21). By adaptive evolution, the number of copies may change.

5 The host cell

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

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

15 Yeasts may either grow by budding of a unicellular thallus or may grow by fission of the organism. A preferred yeast as a transformed host cell may belong to the genera *Saccharomyces*, *Kluyveromyces*, *Candida*, *Pichia*, *Schizosaccharomyces*, *Hansenula*, *Kloeckera*, *Schwanniomyces* or *Yarrowia*. Preferably the yeast is one capable of anaerobic fermentation, more preferably one capable of anaerobic alcoholic fermentation.

20 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.

25 The filamentous fungi of the suitable for use as a cell of the present invention are morphologically, physiologically, and genetically distinct from yeasts. Filamentous fungal cells may be advantageously used since most fungi do not require sterile conditions for propagation and are insensitive to bacteriophage infections. Vegetative growth by filamentous fungi is by hyphal elongation and carbon catabolism of most filamentous fungi is obligately aerobic. Preferred filamentous fungi as a host cell may belong to the genus *Aspergillus*, *Trichoderma*, *Humicola*, *Acremonium*, *Fusarium* or *Penicillium*. More preferably, the filamentous fungal cell may be a *Aspergillus niger*, *Aspergillus oryzae*, a *Penicillium chrysogenum*, or *Rhizopus oryzae* cell.

30 In one embodiment the host cell may be yeast.

Preferably the host is an industrial host, more preferably an industrial yeast. An industrial host and industrial yeast cell may be defined as follows. The living

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

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

araA, *araB* and *araD* genes

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

Organisms, for example *S. cerevisiae* strains, able to produce ethanol from L-arabinose may be produced by modifying a cell introducing the *araA* (L-arabinose
25 isomerase), *araB* (L-ribulokinase) and *araD* (L-ribulose-5-P4-epimerase) genes from a suitable source. Such genes may be introduced into a transformed host cell in order that it is capable of using arabinose. Such an approach is given is described in WO2003/095627. *araA*, *araB* and *araD* genes from *Lactobacillus plantarum* may be used and are disclosed in WO2008/041840. The *araA* gene from *Bacillus subtilis* and the
30 *araB* and *araD* genes from *Escherichia coli* may be used and are disclosed in EP1499708. In another embodiment, *araA*, *araB* and *araD* genes may be derived from at least one of the genus *Clavibacter*, *Arthrobacter* and/or *Gramella*, in particular one of *Clavibacter michiganensis*, *Arthrobacter aurescens*, and/or *Gramella forsetii*, as disclosed in WO 2009011591.

PPP-genes

A transformed host cell may comprise one or more genetic modifications that increases the flux of the pentose phosphate pathway. In particular, the genetic modification(s) may lead to an increased flux through the non-oxidative part of the pentose phosphate pathway. A genetic modification that causes an increased flux of the non-oxidative part of the pentose phosphate pathway is herein understood to mean a modification that increases the flux by at least a factor of about 1.1, about 1.2, about 1.5, about 2, about 5, about 10 or about 20 as compared to the flux in a strain which is genetically identical except for the genetic modification causing the increased flux. The flux of the non-oxidative part of the pentose phosphate pathway may be measured by growing the modified host on xylose as sole carbon source, determining the specific xylose consumption rate and subtracting the specific xylitol production rate from the specific xylose consumption rate, if any xylitol is produced. However, the flux of the non-oxidative part of the pentose phosphate pathway is proportional with the growth rate on xylose as sole carbon source, preferably with the anaerobic growth rate on xylose as sole carbon source. There is a linear relation between the growth rate on xylose as sole carbon source ($\mu_{\max}$) and the flux of the non-oxidative part of the pentose phosphate pathway. The specific xylose consumption rate (Q_s) is equal to the growth rate (μ) divided by the yield of biomass on sugar (Y_{xs}) because the yield of biomass on sugar is constant (under a given set of conditions: anaerobic, growth medium, pH, genetic background of the strain, etc.; i.e. $Q_s = \mu / Y_{xs}$). Therefore the increased flux of the non-oxidative part of the pentose phosphate pathway may be deduced from the increase in maximum growth rate under these conditions unless transport (uptake is limiting).

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

In a preferred host cell, the genetic modification comprises overexpression of at least one enzyme of the (non-oxidative part) pentose phosphate pathway. Preferably the enzyme is selected from the group consisting of the enzymes encoding for ribulose-5-phosphate isomerase, ribulose-5-phosphate epimerase, transketolase and transaldolase. Various combinations of enzymes of the (non-oxidative part) pentose phosphate pathway may be overexpressed. E.g. the enzymes that are overexpressed may be at least the enzymes ribulose-5-phosphate isomerase and ribulose-5-phosphate epimerase; or at least the enzymes ribulose-5-phosphate isomerase and transketolase; or at least the enzymes ribulose-5-phosphate isomerase and transaldolase; or at least the enzymes ribulose-5-phosphate epimerase and transketolase; or at least the enzymes ribulose-5-phosphate epimerase and transaldolase; or at least the enzymes transketolase and transaldolase; or at least the enzymes ribulose-5-phosphate epimerase, transketolase and transaldolase; or at least the enzymes ribulose-5-phosphate isomerase, transketolase and transaldolase; or at least the enzymes ribulose-5-phosphate isomerase, ribulose-5-phosphate epimerase, and transaldolase; or at least the enzymes ribulose-5-phosphate isomerase, ribulose-5-phosphate epimerase, and transketolase. In one embodiment of the invention each of the enzymes ribulose-5-phosphate isomerase, ribulose-5-phosphate epimerase, transketolase and transaldolase are overexpressed in the host cell. More preferred is a host cell in which the genetic modification comprises at least overexpression of both the enzymes transketolase and transaldolase as such a host cell is already capable of anaerobic growth on xylose. In fact, under some conditions host cells overexpressing only the transketolase and the transaldolase already have the same anaerobic growth rate on xylose as do host cells that overexpress all four of the enzymes, i.e. the ribulose-5-phosphate isomerase, ribulose-5-phosphate epimerase, transketolase and transaldolase. Moreover, host cells overexpressing both of the enzymes ribulose-5-phosphate isomerase and ribulose-5-phosphate epimerase are preferred over host cells overexpressing only the isomerase or only the epimerase as overexpression of only one of these enzymes may produce metabolic imbalances.

The enzyme "ribulose 5-phosphate epimerase" (EC 5.1.3.1) is herein defined as an enzyme that catalyses the epimerisation of D-xylulose 5-phosphate into D-ribulose 5-phosphate and vice versa. The enzyme is also known as phosphoribulose epimerase; erythrose-4-phosphate isomerase; phosphoketopentose 3-epimerase; xylulose phosphate 3-epimerase; phosphoketopentose epimerase; ribulose 5-phosphate 3-

epimerase; D-ribulose phosphate-3-epimerase; D-ribulose 5-phosphate epimerase; D-ribulose-5-P 3-epimerase; D-xylulose-5-phosphate 3-epimerase; pentose-5-phosphate 3-epimerase; or D-ribulose-5-phosphate 3-epimerase. A ribulose 5-phosphate epimerase may be further defined by its amino acid sequence. Likewise a ribulose 5-phosphate epimerase may be defined by a nucleotide sequence encoding the enzyme as well as by a nucleotide sequence hybridising to a reference nucleotide sequence encoding a ribulose 5-phosphate epimerase. The nucleotide sequence encoding for ribulose 5-phosphate epimerase is herein designated *RPE1*.

The enzyme "ribulose 5-phosphate isomerase" (EC 5.3.1.6) is herein defined as an enzyme that catalyses direct isomerisation of D-ribose 5-phosphate into D-ribulose 5-phosphate and vice versa. The enzyme is also known as phosphopentosisomerase; phosphoriboisomerase; ribose phosphate isomerase; 5-phosphoribose isomerase; D-ribose 5-phosphate isomerase; D-ribose-5-phosphate ketol-isomerase; or D-ribose-5-phosphate aldose-ketose-isomerase. A ribulose 5-phosphate isomerase may be further defined by its amino acid sequence. Likewise a ribulose 5-phosphate isomerase may be defined by a nucleotide sequence encoding the enzyme as well as by a nucleotide sequence hybridising to a reference nucleotide sequence encoding a ribulose 5-phosphate isomerase. The nucleotide sequence encoding for ribulose 5-phosphate isomerase is herein designated *RKI1*.

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

The enzyme "transaldolase" (EC 2.2.1.2) is herein defined as an enzyme that catalyses the reaction: sedoheptulose 7-phosphate + D-glyceraldehyde 3-phosphate $\leftrightarrow$ D-erythrose 4-phosphate + D-fructose 6-phosphate and vice versa. The enzyme is also known as dihydroxyacetone transferase; dihydroxyacetone synthase; formaldehyde transketolase; or sedoheptulose-7-phosphate :D-glyceraldehyde-3-phosphate glyceronetransferase. A transaldolase may be further defined by its amino acid

sequence. Likewise a transaldolase may be defined by a nucleotide sequence encoding the enzyme as well as by a nucleotide sequence hybridising to a reference nucleotide sequence encoding a transaldolase. The nucleotide sequence encoding for transketolase from is herein designated *TAL1*.

5

Xylose Isomerase or xylose reductase genes

According to the invention, one or more copies of one or more xylose isomerase gene and/or one or more xylose reductase and xylitol dehydrogenase are introduced into the genome of the host cell. The presence of these genetic elements confers on the cell
10 the ability to convert xylose by isomerisation or reduction.

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

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

20 Accordingly, such a transformed host cell is capable of isomerising xylose to xylulose. The ability of isomerising xylose to xylulose is conferred on the host cell by transformation of the host cell with a nucleic acid construct comprising a nucleotide sequence encoding a defined xylose isomerase. A transformed host cell isomerises xylose into xylulose by the direct isomerisation of xylose to xylulose.

25 A unit (U) of xylose isomerase activity may herein be defined as the amount of enzyme producing 1 nmol of xylulose per minute, under conditions as described by Kuyper *et al.* (2003, FEMS Yeast Res. 4: 69-78).

The Xylose isomerase gene may have various origin, such as for example *Piromyces* sp. as disclosed in WO2006/009434. Other suitable origins are *Bacteroides*,
30 in particular *Bacteroides uniformis* as described in PCT/EP2009/52623, *Bacillus*, in particular *Bacillus stearothermophilus* as described in PCT/EP2009/052625.

In another embodiment, one or more copies of one or more xylose reductase and xylitol dehydrogenase genes are introduced into the genome of the host cell. In this embodiment the conversion of xylose is conducted in a two step conversion of xylose

into xylulose via a xylitol intermediate as catalysed by xylose reductase and xylitol dehydrogenase, respectively. In an embodiment thereof xylose reductase (XR), xylitol dehydrogenase (XDH), and xylokinase (XK) may be overexpressed, and optionally one or more of genes encoding NADPH producing enzymes are up-regulated and one or
5 more of the genes encoding NADH consuming enzymes are up-regulated, as disclosed in WO 2004085627.

XKS1 gene

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

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

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

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

understood that these levels of overexpression may apply to the steady state level of the enzyme's activity, the steady state level of the enzyme's protein as well as to the steady state level of the transcript coding for the enzyme.

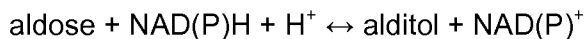
5 Aldose reductase (*GRE3*) gene deletion

In the embodiment, where XI is used as gene to convert xylose, it may be advantageous to reduce aldose reductase activity. A transformed host cell may therefore comprise one or more genetic modifications that reduce unspecific aldose reductase activity in the host cell. Preferably, unspecific aldose reductase activity is
10 reduced in the host cell by one or more genetic modifications that reduce the expression of or inactivates a gene encoding an unspecific aldose reductase. Preferably, the genetic modification(s) reduce or inactivate the expression of each endogenous copy of a gene encoding an unspecific aldose reductase in the host cell (herein called *GRE3* deletion). Transformed host cells may comprise multiple copies of genes encoding unspecific
15 aldose reductases as a result of di-, poly- or aneu-ploidy, and/or the host cell may contain several different (iso)enzymes with aldose reductase activity that differ in amino acid sequence and that are each encoded by a different gene. Also in such instances preferably the expression of each gene that encodes an unspecific aldose reductase is reduced or inactivated. Preferably, the gene is inactivated by deletion of at least part of
20 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
25 reduced in the host cell is a nucleotide sequence encoding a polypeptide with aldose reductase activity.

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

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



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

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

Bioproducts production

Over the years suggestions have been made for the introduction of various organisms for the production of bio-ethanol from crop sugars. In practice, however, all
15 major bio-ethanol production processes have continued to use the yeasts of the genus *Saccharomyces* as ethanol producer. This is due to the many attractive features of *Saccharomyces* species for industrial processes, i. e. , a high acid-, ethanol-and osmo-tolerance, capability of anaerobic growth, and of course its high alcoholic fermentative capacity. Preferred yeast species as host cells include *S. cerevisiae*, *S. bulderi*, *S.*
20 *barnetti*, *S. exiguus*, *S. uvarum*, *S. diastaticus*, *K. lactis*, *K. marxianus* or *K fragilis*.

A transformed host cell may be a cell suitable for the production of ethanol. A transformed host cell may, however, be suitable for the production of fermentation products other than ethanol

Such non-ethanolic fermentation products include in principle any bulk or fine
25 chemical that is producible by a eukaryotic microorganism such as a yeast or a filamentous fungus.

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

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

Lignocellulose

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

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

Pretreatment

Before enzymatic treatment, the lignocellulosic material may be pretreated. The pretreatment may comprise exposing the lignocellulosic material to an acid, a base, a solvent, heat, a peroxide, ozone, mechanical shredding, grinding, milling or rapid depressurization, or a combination of any two or more thereof. This chemical pretreatment is often combined with heat-pretreatment, e.g. between 150-220 °C for 1 to 30 minutes.

Enzymatic hydrolysis

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

Fermentation

The fermentation process may be an aerobic or an anaerobic fermentation process. An anaerobic fermentation process is herein defined as a fermentation process run in the absence of oxygen or in which substantially no oxygen is consumed,

preferably less than about 5, about 2.5 or about 1 mmol/L/h, more preferably 0 mmol/L/h is consumed (i.e. oxygen consumption is not detectable), and wherein organic molecules serve as both electron donor and electron acceptors. In the absence of oxygen, NADH produced in glycolysis and biomass formation, cannot be oxidised by oxidative phosphorylation. To solve this problem many microorganisms use pyruvate or one of its derivatives as an electron and hydrogen acceptor thereby regenerating NAD⁺.

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

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

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

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

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

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

An oxygen-limited fermentation process is a process in which the oxygen consumption is limited by the oxygen transfer from the gas to the liquid. The degree of oxygen limitation is determined by the amount and composition of the ingoing gasflow as well as the actual mixing/mass transfer properties of the fermentation equipment used. Preferably, in a process under oxygen-limited conditions, the rate of oxygen

consumption is at least about 5.5, more preferably at least about 6, such as at least 7 mmol/L/h. A process of the invention may comprise recovery of the fermentation product.

In a preferred process the cell ferments both the xylose and glucose, preferably simultaneously in which case preferably a cell is used which is insensitive to glucose repression to prevent diauxic growth. In addition to a source of xylose (and glucose) as carbon source, the fermentation medium will further comprise the appropriate ingredient required for growth of the cell. Compositions of fermentation media for growth of microorganisms such as yeasts are well known in the art

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

SSF mode

For Simultaneous Saccharification and Fermentation (SSF) mode, the reaction time for liquefaction/hydrolysis or presaccharification step is dependent on the time to realize a desired yield, i.e. cellulose to glucose conversion yield. Such yield is preferably as high as possible, preferably 60% or more, 65% or more, 70% or more, 75% or more 80% or more, 85% or more, 90% or more, 95% or more, 96% or more, 97% or more, 98% or more, 99% or more, even 99.5% or more or 99.9% or more.

According to the invention very high sugar concentrations in SHF mode and very high product concentrations (e.g. ethanol) in SSF mode are realized. In SHF operation the glucose concentration is 25g/L or more, 30 g/L or more, 35g/L or more, 40 g/L or more, 45 g/L or more, 50 g/L or more, 55 g/L or more, 60 g/L or more, 65 g/L or more, 70 g/L or more, 75 g/L or more, 80 g/L or more, 85 g/L or more, 90 g/L or more, 95 g/L or more, 100 g/L or more, 110 g/L or more, 120g/L or more or may e.g. be 25g/L-250 g/L, 30g/L-200g/L, 40g/L-200 g/L, 50g/L-200g/L, 60g/L-200g/L, 70g/L-200g/L, 80g/L-200g/L, 90 g/L, 80g/L-200g/L.

Product concentration in SSF mode

In SSF operation, the product concentration (g/L) is dependent on the amount of glucose produced, but this is not visible since sugars are converted to product in the SSF, and product concentrations can be related to underlying glucose concentration by

multiplication with the theoretical maximum yield (Yps max in gr product per gram glucose)

The theoretical maximum yield (Yps max in gr product per gram glucose) of a fermentation product can be derived from textbook biochemistry. For ethanol, 1 mole of glucose (180 gr) yields according to normal glycolysis fermentation pathway in yeast 2 moles of ethanol ($=2 \times 46 = 92$ gr ethanol. The theoretical maximum yield of ethanol on glucose is therefore $92/180 = 0.511$ gr ethanol/gr glucose.

For Butanol (MW 74 gr/mole) or iso butanol, the theoretical maximum yield is 1 mole of butanol per mole of glucose. So Yps max for (iso-)butanol = $74/180 = 0.411$ gr (iso-)butanol/gr glucose.

For lactic acid the fermentation yield for homolactic fermentation is 2 moles of lactic acid (MW = 90 gr/mole) per mole of glucose. According to this stoichiometry, the Yps max = 1 gr lactic acid/gr glucose.

For other fermentation products a similar calculation may be made.

SSF mode

In SSF operation the product concentration is 25g * Yps g/L /L or more, 30 * Yps g/L or more, 35g * Yps /L or more, 40 * Yps g/L or more, 45 * Yps g/L or more, 50 * Yps g/L or more, 55 * Yps g/L or more, 60 * Yps g/L or more, 65 * Yps g/L or more, 70 * Yps g/L or more , 75 * Yps g/L or more, 80 * Yps g/L or more, 85 * Yps g/L or more, 90 * Yps g/L or more, 95 * Yps g/L or more, 100 * Yps g/L or more, 110 * Yps g/L or more, 120g/L * Yps or more or may e.g. be 25 * Yps g/L-250 * Yps g/L, 30 * Yps gl/L-200 * Yps g/L, 40 * Yps g/L-200 * Yps g/L, 50 * Yps g/L-200 * Yps g/L, 60 * Yps g/L-200 * Yps g/L, 70 * Yps g/L-200 * Yps g/L, 80 * Yps g/L-200 * Yps g/L, 90 * Yps g/L , 80 * Yps g/L-200 * Yps g/L

Accordingly, the invention provides a method for the preparation of a fermentation product, which method comprises:

- a. degrading lignocellulose using a method as described herein; and
 - b. fermenting the resulting material,
- thereby to prepare a fermentation product.

Fermentation product

The fermentation product of the invention may be any useful product. In one embodiment, it is a product selected from the group consisting of ethanol, n-butanol, isobutanol, lactic acid, di-terpene, glycosylated di-terpene, 3-hydroxy-propionic acid, acrylic acid, acetic acid, succinic acid, fumaric acid, malic acid, itaconic acid, maleic acid, citric acid, adipic acid, an amino acid, such as lysine, methionine, tryptophan, threonine, and aspartic acid, 1,3-propane-diol, ethylene, glycerol, a β -lactam antibiotic and a cephalosporin, vitamins, pharmaceuticals, animal feed supplements, specialty chemicals, chemical feedstocks, plastics, solvents, fuels, including biofuels and biogas or organic polymers, and an industrial enzyme, such as a protease, a cellulase, an amylase, a glucanase, a lactase, a lipase, a lyase, an oxidoreductases, a transferase or a xylanase. For example the fermentation products may be produced by cells according to the invention, following prior art cell preparation methods and fermentation processes, which examples however should herein not be construed as limiting. n-butanol may be produced by cells as described in WO2008121701 or WO2008086124; lactic acid as described in US2011053231 or US2010137551; 3-hydroxy-propionic acid as described in WO2010010291; acrylic acid as described in WO2009153047.

Recovery of the fermentation product

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

The following examples illustrate the invention:

EXAMPLES

METHODS

Molecular biology techniques and chemicals. Restriction enzymes and T4 DNA ligase were acquired from Fermentas. Antibiotics hygromycin (HG), phleomycin (phleo) and geneticin (G418) were acquired from Invivogen. pYL16 and nourseothricin (nour) were acquired from Werner Bioagents. Ampicillin and kanamycin were acquired from Sigma-Aldrich.

For PCR amplifications, Phusion® High-Fidelity DNA Polymerase was used (Finnzymes). PCR fragments were sub-cloned using the TOPO® TA Cloning® Kit or the Zero Blunt® TOPO® PCR Cloning Kit (both from Life Technologies). Oligonucleotides used for strain construction were purchased from Sigma-Aldrich.

Plasmids were amplified and maintained in chemically competent TOP10 cells (TOPO® TA Cloning® Kit, Life Technologies) following manufacturer's instructions. Plasmids were isolated from *E. coli* mini cultures using the GeneJET™ Plasmid Miniprep Kit (Fermentas). Genomic DNA was isolated from yeast using the YeaStar™ Genomic DNA Kit (ZymoResearch) following manufacturer's instructions.

Standard molecular biology and yeast genetics techniques were conducted according to textbooks including Sambrook *et al.* (1989) and Ausubel *et al.* (1995).

Strains and maintenance. For storage of the strains used in this work (Table 2), shake flask cultures were performed in rich medium (YP), consisting of 10 g l⁻¹ yeast extract (Oxoid) and 20 g l⁻¹ peptone (BD Difco), supplemented with either 2% glucose (YPD), 2% maltose (YPM), or 3% xylose (YPX). Cultures were maintained at 30 °C in an orbital shaker until cultures reached stationary growth phase. After adding glycerol to 30% (v/v), samples from shake-flask cultures were stored in 2 ml aliquots at -80 °C.

Table 2: Strains used or prepared herein

Strain	Genotype
RN1001	<i>Mat a, ura3-52, leu2-112, gre3::loxP, loxP-Ptpi:TAL1, loxP-Ptpi::RK11, loxP-Ptpi-TKL1, loxP-Ptpi-RPE1, delta::Padh1XKSTTcyc1-LEU2, delta::URA3-Ptpi-xylA-Tcyc1</i>
RN1014	RN1001 + <i>in vivo</i> engineering on xylose and acetic acid
RN1041	RN1001 <i>his3::loxP</i>

Strain	Genotype
RN1053	RN1041 <i>hxt2::loxP-kanMX-loxP</i> , <i>hxt367::loxP-hphMX-loxP</i> , <i>hxt145::loxP-natMX-loxP</i> , <i>gal2::loxP-zeoMX-loxP</i>
YD01227	RN1014 <i>glk1::lox72</i> ; <i>hxx1::loxP</i> ; <i>hxx2::lox72</i> ; <i>gal1::loxP</i> ; <i>his3::loxPnatMXloxP</i>

OD600 and HPLC analysis in shake flask culture. Shake flask cultures were sampled regularly during culture. For OD600 measurements, cultures were diluted appropriately for accurate measurement and optical density was measured at 600 nm wavelength in a Perkin Elmer Spectrophotometer λ 2 instrument. Remaining sample was filtrated to separate medium from yeast.

The filtrate was inserted into the appropriate vials for HPLC analysis. The concentrations of glucose, xylose, glycerol, acetic acid and ethanol in the medium were determined using a Shimadzu HPLC system. The system is equipped with column oven CTO-10A-vp and Autoinjector SIL-10AD-vp with a guard column (Bio-Rad H cartridge, Bio-Rad) and an Aminex HPX-87H column (300 x 7.8 mm; Bio-Rad). Elution took place at 80 °C with 5 mM H₂SO₄ at 0.6 mL/min. The eluate was monitored using a Refractive Index detector RID-10A (Shimadzu).

Microwell plate culture for growth curve profiling. For micro-well cultivation of strains, the Bioscreen C (Growth Curves Ltd.) was used. Overnight pre-cultures were pelleted, washed with demi water and diluted in demi water to twice the desired OD600 for inoculation. Medium was prepared in twice the concentration as desired. In one well of a honeycomb wellplate, 150 μ l medium was mixed with 150 μ l cell suspension. Measurements were conducted in triplicate. Settings for the Bioscreen C were maintained at 30°C incubation T, measurements every 15 min, shaking at type Continuous, amplitude Maximum, and speed Normal. Shaking was set to stop 5 sec before measurement.

Automated transformation and colony picking. For the generation of transformation of a saturation mutagenesis library into the model strains shake-flask cultures were performed in either YPM for RN1053, or YPX for YD01227 (see below). Yeast cells were pelleted and, subsequently, used in an automated transformation protocol based on Schiestl and Gietz (1989). Transformation mixes were plated on selection medium consisting of yeast nitrogen base (Sigma-Aldrich; 6.7 g l⁻¹), agar (BD Biosciences; 15 g l⁻¹), supplemented with either 2% maltose (RN1053 transformations) or 3% xylose (YD01227 transformations). Transformation plates were incubated at 30

°C, and after colony formation, colonies were re-plated using an automated process transferring colonies to 96 well microtiter plates (MTP) containing the above-referred selection media. MTPs with transformants were incubated at 30 °C until clear growth was observed.

5 **NMR analysis.** For the quantification of glucose, xylose, glycerol, acetic acid and ethanol in the sample, 100 µl sample is transferred accurately into a suitable vial. Subsequently 100 µl internal standard solution, containing maleic acid (20 g/l), EDTA (40 g/l) and trace amounts of DSS (4,4-dimethyl-4-silapentane-1-sulfonic acid) in D₂O, and 450 µL D₂O is added.

10 1D ¹H NMR spectra are recorded on a Bruker Avance III 700 MHz, equipped with a cryo-probe, using a pulse program with water suppression (power corresponding to 3 Hz) at a temperature of 27°C.

The analyte concentrations are calculated based on the following signals (δ relative to DSS):

- 15
- α-glucose peak at 5.22 ppm (d, 0.38 H, J = 4 Hz),
 - α-xylose peak at 5.18 ppm (d, 0.37 H, J = 4 Hz),
 - glycerol peak at 3.55 ppm (dd, 2 H, J_{1,2} = 6 Hz and J_{1a,1b} = 12 Hz)
 - acetic acid peak at 1.91 ppm (s, 3 H)
 - ethanol peak at 1.17 ppm (t, 3 H, J = 7Hz)

20

 - The signal user for the standard:
 - Maleic acid peak at 6.05 ppm (s, 2H)

Example 1 - Hexose transporter gene deletions

25 **Deletion cassettes construction.** Primers used in plasmid constructions are shown in Table 3; generated plasmids are shown in Table 4. Schemes with restriction sites used for cloning and sites used to release deletion constructs from the plasmid backbone are shown in Table 5.

30 Table 3: Primers (oligonucleotides) used in the examples

SEQ ID NO:	Internal code	Primer	Sequence (5'→3')	Gene(s)	Purpose
2	5034	Kanf	AAGCTTGCCTCGTCCCCGCC	kanMX	Amplification kanMX
3	5035	Kanr	GTCGACACTGGATGGCGGCG	kanMX	Amplification kanMX
4	5116	If2	ATTCTAGTAACGGCCGCCAGTGTGCTGGAATTCGCCCTTAAGCTTGCCCTCGT	loxP	Part of loxP flank

SEQ ID NO:	Internal code	Primer	Sequence (5'→3')	Gene(s)	Purpose
			CCCCGCCG		
5	5118	lr2	CATACATTATACGAAGTTATGCGCGCTCTAGATATCGTCGACACTGGATGGC GGCG	loxP	Part of loxP flank
6	5115	lf1	ATCCGGACGTACGTATAACTTCGTATAGCATACATTATACGAAGTTATTCTAG TAACGGCCGCCA	loxP	Reamplification/full loxP flank
10	5117	lr1	TCATGACGTCTCGAGGCCATAACTTCGTATAGCATACATTATACGAAGTTAT GCGCGCT	loxP	Reamplification/full loxP flank
11	115	Natf	ACATGTAAAATGACCACTCTTGACGACACGGC	nat1	Amplification nat1
13	116	Natr	CAGTACTAGGGGCCAGGGCATGCTC	nat1	Amplification nat1
14	28	H3f	TGTACATCCGGAATTCAGATTGGTGAGCGCTAGGAGTCACTGCC	HIS3	-
16	29	H3r	CTCAGTATTTACACCCGATATGATCCGTCG	HIS3	-
17	201	Hx2uf	GACTAGTACCGGTGTTTTCAAAACCTAGCAACCCC	HXT2	Upstream flank
18	202	Hx2ur	CGTACGCGTCTTCCGGAAGGGTACCATCAGATTTTCATTGACC	HXT2	Upstream flank
19	203	Hx2df	GAAGACACTCGAGACGTCTTTGTCTGTGAAACCAAGGGC	HXT2	Downstream flank
20	204	Hx2dr	GTCGACGGGCCCTTATGTTGGTCTGTTTAGTATGGCCG	HXT2	Downstream flank
21	205	Hx3uf	AAGCGGCCGCACTAGTACCGGTGAAACAACCTCAATAACGATGTGGGAC	HXT3	Upstream flank
22	206	Hx3ur	ATCCGGACGTCTTCTCAAGAAATCAGTTTGGGCGACG	HXT3	Upstream flank
23	210	Hx4df	AGAAGACGCTCGAGACGTCCCTTATGGGAAGAAGGTGTTTTGCC	HXT4	Downstream flank
24	211	Hx4dr	ATGGATCCTAGGGGTTCTTGACAGTAAATGCG	HXT4	Downstream flank
25	212	Hx5uf	AAGCGGCCGCACTAGTACATGTGAACCTGAAACGCTCATCAAGGC	HXT5	Upstream flank
26	213	Hx5ur	TTCGTACGCGTCTTCCGAGTAACATGAAACCAGAGTACCACG	HXT5	Upstream flank
27	229	Hx7df	AGAAGACCTCGAGACGTCCGACGCTGAAGAAATGACTCACG	HXT7	Downstream flank
28	230	Hx7dr	AGTCGACGGATCCGTAATTTTCTCTTTAAGTGACGGGCG	HXT7	Downstream flank
29	243	Gal2ufn	AAGCGGCCGCACTAGTACCGGTGATCTATATTCGAAAGGGGCGG	GAL2	Upstream flank
30	244	Gal2urn	AACGTACGTCCGATCATTAGAATACTTTTGAGATTGTGCGCT	GAL2	Upstream flank
31	233	Ga2df	AGAAGACCTCGAGACGTCTTACCTTGAAATCTGAAGGCTGG	GAL2	Downstream flank

SEQ ID NO:	Internal code	Primer	Sequence (5'→3')	Gene(s)	Purpose
36	234	Ga2dr	GTGGATCCTAGGTAAACGGTACGAGAAAAGCTCCG	GAL2	Downstream flank
42	281	Hx3inr2	GCTCTTTTCACGGAGAAATTCGGG	HXT3-6-7	Integration check
43	289	Hx2inf	TCTTCGGGAAGCTAGATAGGTGGC	HXT2	Integration check
38	290	Hx2inr	GAAGTAATCAGCCACAATACGCC	HXT2	Integration check
39	299	Hx4inr2	CCATACATTTGTCTGACTCAAGCGC	HXT5-1-4	Integration check
37	317	Hx5inf	GGGTTAATTAGTTTTAGGGGCACGG	HXT5-1-4	Integration check
40	323	Hx7inr1	GATGAGAATCCTTGGCAACCGC	HXT3-6-7	Integration check
41	324	Ga2inf1	TCAATTCGGAAGCTTCCTCCGG	GAL2	Integration check
44	325	Ga2inr1	CAGTGATAGTTTGGTTCGAGCGG	GAL2	Integration check
45	838	Glk1-psuc227f	ATGTCATTTCGACGACTTACACAAAGCCACTGAGAGAGCGGTATCCAGGCC GTCGACCTCGAGTACCGTTCG	GLK1	Hexokinase flank/Bipartite cassette
46	834	Hxk2-psuc227f	GCCAGAAAGGGTCCATGGCCGATGTGCCAAAGGAATTGATGCAACAAATCC GTCGACCTCGAGTACCGTTCG	HXK2	Hexokinase flank/Bipartite cassette
47	645	pSUC227r	GCAATTCGGCTATACGTAAC		Bipartite cassette
48	839	Glk1-psuc225r	CAATCTCAAGTGACCTTCTCTCACCTCGGCACCCAAGGGTGACAAGCC GGATCCTACCGTTCGTATAGC	GLK1	Hexokinase flank/Bipartite cassette
49	835	Hxk2-psuc225r	GCCAGAAAGGGTCCATGGCCGATGTGCCAAAGGAATTGATGCAACAAATCC GTCGACCTCGAGTACCGTTCG	HXK2	Hexokinase flank/Bipartite cassette
50	646	pSUC225f	CGTTCATCATGGAAAATAGC		Bipartite cassette
51	846	Hxk1_loxP_f	ATGGTTCATTTAGGTCCAAGAAACACAGGCTAGAAAGGGTTCATGGCCG GATCCACTAGCATAACTTCG	HXK1	Hexokinase flank/DRM cassette
52	847	Hxk1_loxP_r	ATGGTTCATTTAGGTCCAAGAAACACAGGCTAGAAAGGGTTCATGGCCG GATCCACTAGCATAACTTCG	HXK1	Hexokinase flank/DRM cassette
53	848	Gal1_loxP_f	ATGACTAAATCTCATTGAGAAGAAGTGATTGTACCTGAGTTCAATTCTAGCGG ATCCACTAGCATAACTTCG	GAL1	Hexokinase flank/DRM cassette
54	849	Gal1_loxP_r	TTATAATTCATATAGACAGCTGCCAATGCTGGTTTAGAGACGATGATGTTG GGCCGCCAGTGTGATGG	GAL1	Hexokinase flank/DRM cassette

Table 4: Plasmids used in the strain construction

Number	Construct	Purpose	SEQ ID
--------	-----------	---------	--------

			NO:
pRN201	pCR-BLUNT- <i>loxP</i> - <i>kanMX</i> - <i>loxP</i>	Dominant resistance marker	8
pRN251	pCR-BLUNT- <i>loxP</i> - <i>hphMX</i> - <i>loxP</i>	Dominant resistance marker	9
pRN365	pCR-BLUNT- <i>loxP</i> - <i>natMX</i> - <i>loxP</i>	Dominant resistance marker	10
pRN447	pCR-BLUNT- <i>loxP</i> - <i>zeoMX</i> - <i>loxP</i>	Dominant resistance marker	14
pRN247	pCR-BLUNT- <i>his3:loxP</i> - <i>kanMX</i> - <i>loxP</i>	<i>HIS3</i> deletion construct	16
pRN485	pCR-BLUNT- <i>gal2:loxP</i> - <i>zeoMX</i> - <i>loxP</i>	<i>GAL2</i> deletion construct	33
pRN566	pCR-BLUNT- <i>hxt367:loxP</i> - <i>hphMX</i> - <i>loxP</i>	<i>HXT3-HXT6-HXT7</i> cluster deletion construct	34
pRN569	pCR-BLUNT- <i>hxt514:loxP</i> - <i>natMX</i> - <i>loxP</i>	<i>HXT5-HXT1-HXT4</i> cluster deletion construct	35
pRN635	pCR-BLUNT- <i>hxt2:loxP</i> - <i>kanMX</i> - <i>loxP</i>	<i>HXT2</i> deletion construct	36
pRN993	pRN978-P _{HXT7(-491)} - <i>GAL2</i> -T _{ADHI}	<i>GAL2</i> expression vector	58
pDB1250	pRN978-P _{HXT7(-491)} -syn.t.wt- <i>GAL2</i> -T _{ADHI}	Synthetic wild-type <i>GAL2</i> expression vector	59
pRN187	pCRE- <i>zeoMX</i> (based on pSH65)	CRE recombinase expression vector	60
pRN486	pCR-BLUNT- <i>HIS3::loxP</i> natMXloxP	<i>HIS3</i> deletion construct	61

Table 5: Cloning scheme

Construct	Fragment	Cloning Sites	Release Sites
pRN247	<i>HIS3</i> upstream	<i>SacI</i> - <i>DraI</i>	<i>XhoI</i> - <i>BsrGI</i>
	<i>loxP</i> - <i>kanMX</i> - <i>loxP</i>	<i>StuI</i> - <i>BsiWI</i>	
	<i>HIS3</i> downstream	<i>BsiWI</i> - <i>ApaI</i>	
pRN485	<i>GAL2</i> upstream	<i>NotI</i> - <i>BsiWI</i>	<i>BamHI</i> - <i>SpeI</i> - <i>PmlI</i>
	<i>loxP</i> - <i>zeoMX</i> - <i>loxP</i>	<i>BsiWI</i> - <i>XhoI</i>	
	<i>GAL2</i> downstream	<i>XhoI</i> - <i>BamHI</i>	
pRN566	<i>HXT3</i> upstream	<i>SpeI</i> - <i>BsiWI</i>	<i>BamHI</i> - <i>AgeI</i> - <i>NaeI</i>
	<i>loxP</i> - <i>hphMX</i> - <i>loxP</i>	<i>BsiWI</i> - <i>XhoI</i>	
	<i>HXT7</i> downstream	<i>XhoI</i> - <i>BamHI</i>	
pRN569	<i>HXT5</i> upstream	<i>NotI</i> - <i>BspEI</i>	<i>BamHI</i> - <i>NotI</i> - <i>ApaI</i>
	<i>loxP</i> - <i>natMX</i> - <i>loxP</i>	<i>BspEI</i> - <i>XhoI</i>	
	<i>HXT4</i> downstream	<i>XhoI</i> - <i>BamHI</i>	
pRN635	<i>HXT2</i> upstream	<i>SpeI</i> - <i>BsiWI</i>	<i>AgeI</i> - <i>NotI</i> - <i>BspHI</i>
	<i>loxP</i> - <i>kanMX</i> - <i>loxP</i>	<i>BsiWI</i> - <i>XhoI</i>	

Construct	Fragment	Cloning Sites	Release Sites
	<i>HXT2</i> downstream	<i>XhoI-EcoRI</i>	

The *kanMX* marker was amplified from the plasmid pFA6-*kanMX4* (http://www-sequence.stanford.edu/group/yeast_deletion_project/kanmx4.txt) using primers SEQ ID NO's 1 and 2. Subsequently, the *kanMX* marker was floxed through adding *loxP* flanks by PCR amplification with primers SEQ ID NO's 3 and 4. Re-amplification was done with primers SEQ ID NO's 5 and 6. The resulting *loxP-kanMX-loxP* fragment was cloned in pCR-BLUNT resulting in pRN201 (SEQ ID NO: 7).

For the construction of pRN251 (SEQ ID NO: 8), *hphMX* was isolated from pGRE3:*hphMX* (Kuyper *et al*, 2005). To delete a *MluI* site as appropriate restriction site in the vicinity of *hphMX*, pGRE3:*hphMX* was cut with *Eco32I* and re-ligated. Subsequently, *hphMX* was cloned as *XhoI-MluI* fragment into pRN201 digested with *SaII* and *MluI* to replace *kanMX*.

For the construction of pRN365 (SEQ ID NO: 9), the *Streptomyces noursei nat1* gene was PCR-amplified from pYL16 (Werner Bioagents) using primers with SEQ ID NO's 10 and -11. The *PstI-ScaI nat1* fragment together with the *Acc65I-NcoI* pRN201-fragment were cloned into pRN201, already linearized with *Acc65I* and *ScaI*, in order to replace *kanR* for *nat1*.

For the construction of pRN447 (SEQ ID NO: 12), pRN201 was digested with *PmlI*. This served two ends. Firstly, the *Streptoalloteichus hindustanus ble* (zeocin or phleomycin resistance gene) ORF was isolated, and secondly, after re-ligation of the *PmlI*-digested pRN201 an *NcoI* site was deleted. Subsequently, *ble* as *NcoI-PmlI* fragment and part of pRN201 as *BamHI-NcoI* vector fragment were cloned into the re-ligated pRN201 (missing *ble*), digested with *BamHI* and *ScaI* resulting in pRN447.

For the *HIS3* deletion construct, primers SEQ ID13 and -14 were used to amplify the *HIS3* locus from yeast genomic DNA. Sites used to cut out the *HIS3* flanks and to ligate these to the floxed *kanMX* marker are shown in Table 5. The ligation product was digested with *SacI* and *ApaI* and cloned into pCR-BLUNT digested with *SacI* and *ApaI*. The resulting plasmid is pRN247 (SEQ ID NO: 15).

For the deletion of the eight main hexose transporters (*HXT1-7* and *GAL2*) in *S. cerevisiae*, four deletion constructs were generated (see Table 4). Each deletion construct contained a different floxed dominant resistance marker. For each *HXT* gene 400-700 bp flanks were amplified using the primers listed in Table 3 (SEQ ID NO's 16-

31) using RN1001 genomic DNA as template. The upstream flank, the dominant resistance marker and the downstream marker were ligated using the fragments and cloning sites listed under Table 5. The ligations were amplified using the forward primers of the upstream flank and the reverse primer of the downstream flank (primer combinations SEQ ID NO:'s 16+19, SEQ ID NO:'s 20+27, SEQ ID NO:'s 24+23, and SEQ ID NO:'s 28+31). The fused PCR fragments were cloned into pCR-BLUNT to obtain pRN485, pRN566, pRN569, pRN635 (SEQ ID NO:'s 32-35, respectively). To obtain high yields of plasmid DNA, the plasmids were isolated from 50 mL *E. coli* cultures using NucleoBond® Xtra Midi kit (Bioké, Leiden, the Netherlands). Before transformation to yeast, deletion constructs were released from plasmid backbone by digestion with the release restriction sites listed in Table 5.

Strain construction. The xylose-fermenting strain RN1001 was made histidine auxotroph by the insertion of *loxP-kanMX-loxP* (released from pRN247; SEQ NO ID15) at the *HIS3* locus. Subsequently, the marker was removed through transient expression of plasmid pRN187 (derived from pSH65 expressing galactose-inducible *cre* recombinase; SEQ ID NO 60). Introduction of pRN187 was selected on phleo and *CRE* recombinase expression was induced on YP-medium supplemented with galactose. The resulting *his3:loxP* strain was named RN1041. The hexose transporters were deleted in the following order: 1) *HXT3-HXT6-HXT7* cluster, 2) *HXT5-HXT1-HXT4* cluster, 3) *GAL2*, 4) *HXT2*. The deletion constructs were linearized or released from the plasmid backbone by cutting with the enzyme combinations listed in Table 5 and these were integrated in the genome of RN1041. All transformations were plated on yeast extract (10 g/L), peptone (20g/L) agar (15g/L) medium supplemented with 20g/L maltose. Maltose was added to the medium, because the uptake of this disaccharide goes via an alternative transport system than the glucose transport system (Wieczorke *et al*, 1999). With each deletion of a (cluster of) *HXT* gene(s), an additional marker was inserted in the order: 1) *hphMX*, 2) *natMX*, 3) *zeoMX*, 4) *kanMX*. With each inserted additional marker the respective antibiotic was additionally supplemented to the medium in the following order: 1) HG, 2) HG and nour, 3) HG, nour and phleo, 4) HG, nour, phleo and G418. After integration of all four deletion constructs, a single colony was isolated under selection of all four antibiotics. Correct integrations were verified by PCR analysis on genomic DNA isolates. Primers outside of the integration site were used (combinations SEQ ID NO:'s

36+37, SEQ ID NO:s 38+39, SEQ ID NO:s 40+41, SEQ ID NO:s 42+43; sequences listed in Table 3).

Strain characterization. To characterize the (intermediate) hexose transporter strains, shake flask cultures were performed. Cultures were inoculated at OD₆₀₀=0.1. The resulting strain, RN1053 (Δ hxt1-7;gal2-mutant RN1041; see Table 2 for exact genotype), showed a retarded growth pattern on Verduyn-urea (mineral medium according to Verduyn using urea as nitrogen source; Luttik *et al*, 2001) supplemented with 0.2 g l⁻¹ histidine (Sigma-Aldrich; Verduyn-urea-his; to complement for the histidine auxotrophy) and 15 g l⁻¹ glucose and 20 g l⁻¹ xylose only starting to grow slowly on glucose only after 60 hours; interestingly, when glucose was present in the medium xylose was finished as well after 150 hrs (Figure 1) indicating that one or more of the cryptic hexose transporter genes (*HXT8-17*) was induced on glucose and facilitated xylose transport (Figure 1). However, on xylose as sole carbon source RN1053 did not grow on Verduyn-urea (+ 20 g l⁻¹ % xylose during the culturing period (Figure 2) indicating the strain is useable as model strain for testing putative xylose transporters. RN10153 was further maintained on YPM.

Example 2 – Hexokinase gene deletions

Deletion cassettes construction. For deletion of hexokinase genes oligonucleotides were designed (SEQ ID NO:s in Table 3) comprised of 60 nucleotide flanking sequences homologous to the hexokinase gene locus and of 20 nucleotides homologous to a floxed dominant resistance marker cassette. The oligonucleotides were used to amplify the deletion constructs. Subsequent PCR products were column filter-purified (Fermentas GeneJet Kit) and used for transformations experiments. Three types of deletion cassettes were used: firstly, for *GLK1* and *HXK2* deletions a bipartite system was used. One fragment consisted of a lox66 site, KanMX, *GAL1* promoter upstream of *CRE*, and the 5'-part of *CRE* (*CRE1*) amplified from pSUC227 with one gene-specific primer (SEQ ID NO: 44 for *GLK1* and SEQ ID NO: 45 for *HXK2*) and one pSUC227-specific primer (SEQ ID NO: 46); the second fragment consisted of the 3'-part of *CRE* (*CRE2*) with overlap on *CRE1*, and a lox71 site, amplified from pSUC225 with again one gene-specific primers (SEQ ID NO: 47 for *GLK1* and SEQ ID48 for *HXK2*) and one pSUC225-specific primer SEQ ID NO: 49. Through homologous recombination the two

fragments integrate as lox66-kanMX-CRE-lox71 at the hexokinase locus (pSUC225 and pSUC227 sequences and method provided in PCT/EP2013/055047). Secondly, for *HXK1* (primers SEQ ID NO:s 50-51) and *GAL1* (primers SEQ ID NO:s 52-53) deletions, a floxed dominant resistance marker (DRM) was amplified with flanking sequences homologous to the respective hexokinase to replace the coding region at the locus; as templates for the PCR amplifications of the DRM cassettes pRN774 (loxP-hphMX-loxP; SEQ ID NO: 54) and pRN775 (loxP-natMx-loxP; SEQ ID NO: 55) were used, respectively. Thirdly, for *HIS3* deletion to allow for complementation of the auxotrophic phenotype by transporter episomal plasmids, a similar construct with *HIS3*-homologous flanks was integrated as was used to generate RN1041 (RN1001-*his3*:loxP, in strain family RN1053; see above Example 1). In this case the construct bore natMX as dominant resistance marker instead of kanMX (SEQ ID NO 61).

Strain construction. For the generation of a strain incapable of hexose metabolism but capable of hexose transport, hexokinase gene deletions were made in the xylose-fermenting strain RN1014 (Table 2; Figure 3 for deletion scheme).

As mentioned, in the case of *GLK1* and *HXK2*, the disruption cassettes were bipartite. Through homologous recombination the two fragments integrate as lox66-kanMX-CRE-lox71 at the hexokinase locus. The integration was selected on YPD supplemented with G418. The disruption cassettes for *HXK1* and *GAL1* consisted of one fragment: either loxP-natMX-loxP or loxP-hphMX-loxP, respectively. RN1014 was transformed with the purified PCR products and the integration was selected on the appropriate antibiotic. Additionally, *HIS3* was disrupted with a similar construct (SEQ ID NO: 61) used for the generation of RN1053. In this case natMX was the selection marker instead of kanMX.

The genes were deleted in the following order: 1) *GLK1*, 2) *HXK1*, 3) *GAL1*, 4) *HXK2* and 5) *HIS3*. After the deletion of *GLK1* and *HXK1* both markers were recycled by galactose-induced Cre-mediated recombination. After deletion of *HXK2* the intermediate strain was maintained on xylose-containing rich medium (YPX). After *HIS3* deletion the integrated hphMX and kanMX markers were removed by galactose-induced CRE recombination. To ensure growth of the strain, 2% xylose was added to YP 2% galactose + nourseothricin (YPGX). Selection on nourseothricin ensured maintenance of the natMX marker at the *HIS3* locus leaving a selection trait to be used possibly later on. After single colony isolation, the strain was verified for its deletions and delta sequence profile by colony PCR, and named YD01227.

Strain characterization. In aerobic shake flask culture experiments using Verduyn-urea-his, another colony with the same quadruple hexokinase knockout (KO) genotype (col 2) as YD01227 was characterized in a pre-screen for its ability to consume xylose in the absence and presence an excess of glucose (10%), and for its ability to consume glucose (Figure 4). Cultures were inoculated at OD600=0.1. Both RN1014 and the quadruple hexokinase KO are able to grow on and consume xylose (data not shown). And expectedly, the quadruple hexokinase KO does neither grow on nor consumes glucose, whereas RN1014 does (data not shown). Furthermore, the excess of glucose (10%) prevents the growth on and consumption of xylose for at least 96 hours, whereas RN1014 utilizes xylose (Figure 4).

In Bioscreen C experiments, YD01227 and the aforementioned col 2 were screened for growth on Verduyn-urea-his supplemented with different sugars and sugar mixtures. Cultures were inoculated at OD600=0.05. YD01227 grew on xylose but was not able to grow on glucose, maltose or galactose (data not shown). The glucose-xylose mixtures were screened to support a choice of medium composition suitable for the screen for pentose-specific transporters. As seen in Figure 5, with a ratio of glucose:xylose of 5:1 showed the optimal inhibition of growth on xylose for strain YD01227. This was the case for both a high (10:2), as for a low sugar load (2.5: 0.5). YD01227 was further maintained on YPX for storage and handling.

Example 3. GAL2 saturation mutagenesis library and other constructs.

A synthetic DNA construct for wild-type (WT) *GAL2* was ordered at GeneArt (SEQ ID NO: 56; Invitrogen) and was used as template for site-directed mutagenesis. The synthetic WT *GAL2* DNA construct was cloned into pRN993 (SEQ ID NO: 57) as *XbaI-BssHII* fragment exchanging another ORF for the synthetic WT *GAL2* construct to generate pDB1250 (SEQ ID NO: 58). pDB1250 is a yeast shuttle vector based on pRS313, bearing as most prominent features besides the synthetically made *GAL2* ORF: 1) a *HIS3* expression cassette to complement the histidine auxotrophy, 2) a CEN.ARSH to maintain 1-2 copies (low copy number) of the expression vector in yeast cells, 3) truncated *HXT7* promoter (-491 bp) resulting in medium expression levels of downstream ORF, 4) *ADH1* terminator, and 5) ampicillin resistance gene (*amp^r*) for selection in *E. coli* TOP10 cells (see above) for cloning purposes (Figure 6). The saturation mutagenesis library for at least 13 non-wild-type amino acid changes on 30 amino acid positions in Gal2p was ordered from Invitrogen Life Technologies .

(<http://www.invitrogen.com/site/us/en/home/Products-and-Services/Applications/Cloning/gene-synthesis/directed-evolution/GeneArt-Site-Saturation-Mutagenesis.html>); for positions and amino acid changes to the wild-type Gal2p amino acid sequence (SEQ ID NO: 59) see Table 6.

5

Table 6: GAL2 Single Site Saturation Mutagenesis Library

# sites	Gal2		# non wt AA	Non wt AA
	Position	WT AA		
1	85	F	15	A, C, D, E, G, H, K, M, N, P, Q, R, S, T, V
2	89	T	18	A, C, D, E, F, G, H, I, K, L, M, N, P, Q, R, S, V, W, Y
3	187	V	17	A, C, D, E, F, G, H, I, K, L, M, Q, R, S, T, W, Y
4	191	A	19	C, D, E, F, G, H, K, L, M, N, P, Q, R, S, T, V, W, Y
5	214	Y	15	A, C, G, I, K, L, M, N, P, Q, R, S, T, V, W
6	215	Q	16	A, C, D, E, F, G, H, I, L, M, N, R, S, V, W, Y
7	218	I	15	A, C, D, E, G, H, K, L, M, N, R, S, T, V, W
8	219	T	15	A, C, D, E, F, G, I, K, L, M, N, P, Q, R, S, V, W
9	222	I	15	A, C, D, E, G, H, K, M, P, R, S, T, V, W, Y
10	226	Y	19	A, C, D, E, F, G, H, I, K, L, M, N, P, Q, R, S, T, V, W
11	338	Q	18	A, C, D, E, F, G, H, I, L, M, N, P, R, S, T, V, W, Y
12	339	M	19	A, C, D, E, F, G, H, I, K, L, N, P, Q, R, S, T, V, W, Y
13	341	Q	13	D, E, H, I, K, L, M, N, R, S, T, V, Y
14	342	Q	19	A, C, D, E, F, G, H, I, K, L, M, N, P, R, S, T, V, W, Y
15	343	L	19	A, C, D, E, F, G, H, I, K, M, N, P, Q, R, S, T, V, W, Y
16	346	N	19	A, C, D, E, F, G, H, I, K, L, M, P, Q, R, S, T, V, W, Y
17	347	N	13	A, D, E, G, H, I, K, L, Q, R, S, T, V
18	350	F	14	D, E, H, I, K, L, M, N, Q, R, S, T, V, Y
19	373	G	19	A, C, D, E, F, H, I, K, L, M, N, P, Q, R, S, T, V, W, Y
20	376	N	18	A, C, D, E, F, G, H, I, K, L, M, P, Q, R, S, T, V, W, Y
21	380	T	18	A, C, D, E, F, G, I, K, L, M, N, P, Q, R, S, V, W, Y
22	383	S	17	A, C, D, E, F, G, I, K, L, M, N, Q, R, T, V, W, Y
23	444	F	19	A, C, D, E, G, H, I, K, L, M, N, P, Q, R, S, T, V, W, Y
24	446	Y	13	D, E, H, I, K, L, M, N, Q, R, S, T, V
25	448	T	18	A, C, D, E, F, G, H, I, K, L, M, P, Q, R, S, V, W, Y
26	449	T	19	A, C, D, E, F, G, H, I, K, L, M, N, P, Q, R, S, V, W, Y
27	451	A	18	C, D, F, G, H, I, K, L, M, N, P, Q, R, S, T, V, W, Y
28	455	W	18	A, C, D, E, F, G, I, K, L, M, N, P, Q, R, S, T, V, Y
29	478	N	16	A, C, D, F, G, H, I, K, L, P, Q, R, V, W, Y
30	479	W	18	C, D, E, F, G, H, I, K, L, M, N, P, Q, R, S, T, V, Y

10 Resulting constructs were inserted by custom cloning at GeneArt (Invitrogen, Regensburg, Germany) in pDB1250 (SEQ ID NO: ID58).

Example 4: Glucose Transport Activity Counter-screening

15 **Aim.** Using the Glucose Transport Activity Counter (GTAC)-screen, i.e. transforming hexokinase-mutant strain YD01227 as host to introduce GAL2 variants and screening the resulting transformants on medium for growth on and consumption of xylose in the presence of a 5 times higher amount of glucose, mutations can be identified that favour xylose above in the presence of a surplus glucose in the Gal2p

variant (i.e. higher affinity for xylose than for glucosea reduction or more preferably full removal of glucose transport capability, while keeping xylose transport capability more or less intact).

Transformation and colony picking. YD01227 was transformed with a total of 497 constructs, each one bearing a *GAL2* mutant. One construct with wild-type *GAL2* sequence (pDB1250; SEQ ID NO: 58) was included as control. For each transformation, 3 colonies, when available, were re-plated to agar medium MTPs amounting to 1450 transformants. The 1450 transformants were screened in three parts. For each part of the screening, wild-type *GAL2* was included as control.

Pre-culture and screen. For pre-culture, transformants were transferred by automated process from selection agar medium MTPs to 96 well half-deepwell plates (96HDWP) containing liquid selection medium consisting of mineral medium (according to Verduyn, with urea as nitrogen source) supplemented with 3% xylose. The 96 HDWPs were cultured for 3 days of pre-culture in an orbital shaker at 30 °C and 750 rpm. Subsequently, for each well 20 µl of culture was transferred by automated process to 24 deepwell plates (24 DWPs) containing 2.5 ml Verduyn-urea supplemented with the sugar mixture glucose:xylose in a ratio 5:1 with the following concentrations: 10 g l⁻¹ glucose and 2 g l⁻¹ xylose (YD10-medium). For each of the three sampling points a series of 24DWPs was inoculated. On each 24DWP, one RN1001 growth control was inoculated to have an indication of plate variation effects. After 24 hours, 72 hours and 96 hours automatic sampling and transfer to 96HDWP was conducted for automated OD-measurement at 600 nm wavelength.

After the last OD measurement after 96 hours, cells were pelleted after centrifugation and 100 µl supernatant was collected by automated process for flow-NMR analysis residual constituents in the medium after culture (see above for method description). For each construct and each time point, the measured residual glucose and xylose concentrations for the different replicates were averaged. In order to compare the different parts of the screen, a relative value was calculated based on the difference to the wild-type residual xylose concentration measured in the particular part of the GTAC screen, according to the following formula:

$$\text{RelXyl} = \frac{(\text{residual xylose}_{\text{Gal2-wildtype}} - \text{residual xylose}_{\text{Gal2-mutant}})}{\text{residual xylose}_{\text{Gal2-wildtype}}} \times 100\%$$

Results. Whereas RN1001 displayed complete consumption of both glucose and xylose, YD01227 transformants of wild-type and mutagenesis library *GAL2* constructs did not consume glucose and displayed a spectrum of residual xylose concentrations representing their individual ability to consume xylose in the presence of glucose (Figure 7A). As shown in Figure 7B, the xylose consumption displayed a high correlation ($R^2=0.93$) with the growth measurements (OD600). Since the screen was conducted in aerobic conditions, and little ethanol formation was shown (data not shown) and growth displayed high correlation with the xylose consumption, the residual xylose concentrations were used as main parameter to compare *GAL2* variants with wild-type. The comparison of all tested mutant *GAL2* variants versus wild-type *GAL2* on xylose consumption (average RelXyl) at 96 hours is listed in Table 7. All mutations affecting glucose transport to the benefit of xylose transport have RelXyl>0 (Table 7). The TOP positions to target for a second round of mutagenesis were sorted based on their average RelXyl score per mutation; specific mutations were sorted on preference based on RelXyl score as well (Table 8).

Table 7. Relative OD600 (RelOD600) and xylose consumption (RelXyl) of GTAC Screen. Wild-type *GAL2* construct pDB1250 was set to 0. RelOD600 values are the averages of 1-3 replicates and are relative values compared to wild-type.

Pos	wtAA	MutAA	RelXyl
85	F	A	1,45
		C	0,66
		D	-1,89
		E	-13,09
		G	-1,06
		H	1,76
		K	-20,81
		M	7,03
		N	1,96
		P	2,83
		Q	37,68
		R	-1,93
		S	19,81
		T	50,51
		V	32,76
89	T	A	11,68

Pos	wtAA	MutAA	RelXyl
		C	13,22
		D	19,68
		F	-9,97
		H	25,59
		I	-17,32
		K	-24,84
		L	12,38
		M	-8,38
		N	9,88
		Q	-16,03
		R	-16,11
		S	-11,44
		V	41,06
		Y	-2,89
187	V	A	26,84
		C	9,98
		D	3,76
		E	6,21
		F	14,53
		G	10,01
		H	8,32
		I	6,69
		K	4,09
		L	4,12
		M	1,85
		Q	10,17
		R	3,05
		S	7,54
		T	10,41
		W	1,81
		Y	1,22
191	A	C	4,70
		D	12,43
		E	-2,61
		F	5,49
		G	15,83
		H	2,83
		I	11,93
		K	1,94
		L	-8,25
		M	-14,37

Pos	wtAA	MutAA	RelXyl
		N	20,35
		P	14,89
		Q	2,94
		R	-1,09
		S	3,09
		T	10,61
		V	4,52
		W	2,49
		Y	1,93
214	Y	A	-12,81
		C	-5,85
		G	18,41
		I	-12,07
		K	-4,66
		L	35,33
		M	23,92
		N	32,71
		P	18,44
		Q	39,94
		R	25,50
		S	14,91
		V	55,48
		W	47,35
215	Q	A	2,01
		C	9,90
		D	13,62
		F	10,12
		G	-9,62
		H	-18,18
		I	17,72
		L	24,63
		M	21,59
		N	-11,47
		S	2,14
		V	-8,22
		W	-20,47
		Y	-8,07
218	I	A	17,98
		C	5,52
		D	4,64
		E	8,44

Pos	wtAA	MutAA	RelXyl
		G	4,62
		H	9,88
		K	27,86
		L	9,73
		M	5,80
		N	24,31
		R	-0,73
		S	24,96
		T	-2,65
		V	7,69
		W	-2,10
219	T	A	52,55
		C	25,05
		D	18,04
		E	-14,58
		F	43,86
		G	46,45
		I	-11,07
		K	-4,44
		M	-18,28
		N	7,85
		Q	-22,81
		R	-24,62
		S	4,90
		W	-16,77
222	I	A	11,37
		C	13,45
		D	20,91
		E	3,35
		G	3,75
		H	9,73
		K	-4,92
		M	4,39
		P	0,97
		S	1,85
		T	-1,59
		V	8,42
		W	-8,04
		Y	7,87
226	Y	A	63,63
		C	62,07

Pos	wtAA	MutAA	RelXyl
		D	81,68
		E	70,43
		G	39,06
		H	5,10
		I	32,79
		K	-1,86
		L	42,48
		M	74,04
		N	62,81
		P	49,78
		Q	56,23
		R	65,53
		S	14,97
		T	7,27
		V	-11,15
		W	44,48
338	Q	A	38,18
		C	52,66
		D	33,17
		E	46,93
		F	46,65
		G	13,80
		H	26,28
		I	28,32
		L	33,08
		M	-2,53
		N	22,26
		P	3,96
		R	39,83
		S	3,96
		T	11,84
		V	24,84
		W	35,55
		Y	19,90
339	M	A	50,56
		C	30,36
		D	46,71
		E	10,48
		F	63,75
		G	73,03
		H	67,99

Pos	wtAA	MutAA	RelXyl
		I	40,96
		K	71,92
		L	67,43
		N	83,73
		P	12,77
		Q	77,74
		R	70,25
		S	81,65
		T	57,02
		V	82,35
		W	-23,02
		Y	37,43
341	Q	D	-20,08
		E	-3,38
		H	-22,19
		L	46,71
		M	23,17
		N	11,77
		R	-5,18
		S	15,78
		T	28,67
		Y	19,78
342	Q	A	61,47
		C	64,63
		D	48,62
		E	34,90
		F	45,61
		G	14,10
		H	46,28
		I	16,81
		K	41,75
		L	37,38
		M	27,70
		N	32,30
		P	44,88
		R	23,95
		S	68,48
		T	50,54
		V	24,39
		W	59,43
		Y	76,45

Pos	wtAA	MutAA	RelXyl
343	L	A	9,67
		C	8,74
		D	9,41
		E	2,39
		F	15,49
		G	11,87
		H	16,96
		I	10,67
		K	40,15
		M	16,94
		N	14,03
		P	26,08
		Q	8,73
		R	29,11
		S	20,28
		T	14,38
		V	-6,06
		W	7,76
		Y	20,43
346	N	A	21,99
		C	17,11
		D	8,88
		E	-8,32
		F	9,91
		G	8,42
		H	26,01
		I	17,12
		K	17,48
		L	6,48
		M	13,19
		Q	19,93
		R	7,94
		S	23,37
		T	23,19
		V	47,09
		W	45,58
		Y	41,70
347	N	A	44,64
		D	26,67
		E	25,36
		G	-22,71

Pos	wtAA	MutAA	RelXyl
		H	11,18
		I	27,33
		K	-22,05
		L	7,79
		Q	-14,96
		R	-15,36
		S	-15,56
		T	-19,33
		V	-28,12
350	Y	D	2,93
		E	-13,15
		H	-19,99
		I	-10,76
		K	3,00
		L	40,49
		M	5,01
		N	13,57
		Q	0,66
		R	57,58
		S	15,25
		T	21,64
		V	13,39
		Y	-7,36
373	G	A	25,90
		C	36,56
		D	36,22
		E	41,01
		F	15,15
		H	6,96
		I	9,10
		K	20,72
		L	51,94
		M	35,03
		N	38,49
		P	35,75
		Q	5,50
		R	8,35
		S	11,62
		T	10,67
		V	21,91
		W	22,96

Pos	wtAA	MutAA	RelXyl
		Y	31,04
376	N	A	36,70
		C	78,63
		D	19,23
		E	20,15
		F	92,62
		H	39,30
		I	99,46
		K	7,53
		L	93,69
		M	97,33
		P	51,70
		Q	24,24
		R	27,86
		S	59,39
		T	99,01
		V	96,57
		W	13,51
		Y	43,95
380	T	A	53,06
		C	42,34
		D	37,28
		E	-29,92
		F	54,67
		G	34,22
		I	47,40
		K	42,12
		L	39,25
		M	74,78
		N	49,79
		P	43,28
		Q	43,31
		R	42,31
		S	42,03
		V	42,53
		W	51,16
		Y	25,09
383	S	A	57,21
		C	67,78
		D	29,42
		E	54,55

Pos	wtAA	MutAA	RelXyl
		F	52,71
		G	39,71
		I	25,88
		K	49,19
		L	37,72
		M	52,64
		N	58,14
		Q	83,94
		R	61,78
		T	71,85
		V	19,38
		W	65,32
		Y	62,76
444	F	A	32,41
		C	14,64
		D	13,21
		E	16,38
		G	12,78
		H	25,62
		I	31,95
		K	28,79
		L	39,43
		M	17,30
		N	5,43
		P	15,65
		Q	19,63
		R	0,06
		S	10,94
		T	7,44
		V	37,29
		W	2,94
		Y	32,56
446	F	D	10,80
		E	10,63
		H	16,93
		I	8,65
		K	-3,40
		L	22,78
		M	18,62
		N	14,99
		Q	34,25

Pos	wtAA	MutAA	RelXyl
		R	-8,44
		S	-20,59
		T	-16,07
		V	-9,48
448	T	A	76,95
		C	58,53
		D	66,10
		E	53,25
		F	61,94
		G	65,05
		H	67,00
		I	58,01
		K	54,82
		L	63,49
		M	77,64
		P	64,39
		Q	58,15
		R	33,80
		S	59,40
		V	53,50
		W	56,69
		Y	47,73
449	T	A	57,64
		C	50,44
		D	55,89
		E	37,14
		F	92,65
		G	61,00
		H	82,47
		I	71,71
		K	79,30
		L	81,03
		M	89,29
		N	84,60
		P	37,06
		Q	81,19
		R	82,14
		S	83,15
		V	74,98
		W	80,75
		Y	79,80

Pos	wtAA	MutAA	RelXyl
451	T	C	10,57
		D	9,68
		E	15,39
		F	13,38
		G	43,34
		H	7,65
		I	17,96
		K	8,59
		L	11,00
		M	5,68
		N	14,56
		P	11,89
		Q	8,63
		R	4,46
		S	13,08
		T	2,10
		V	4,99
		W	15,25
		Y	1,62
455	W	A	1,31
		C	1,19
		D	-6,13
		E	19,68
		F	-1,57
		G	0,73
		I	-2,70
		K	6,91
		L	19,08
		M	19,92
		N	19,10
		P	14,76
		Q	3,03
		R	-14,86
		S	8,05
		T	19,75
		V	10,66
		Y	7,30
478	N	A	-11,71
		C	-1,14
		D	-3,08
		F	30,35

Pos	wtAA	MutAA	RelXyl
		G	8,29
		H	-15,02
		I	10,06
		K	9,95
		L	-3,27
		P	30,98
		Q	4,37
		R	11,41
		V	21,91
		W	3,25
		Y	4,98
479	W	C	13,20
		D	6,12
		E	11,20
		F	1,41
		G	1,34
		H	6,90
		I	1,50
		K	5,54
		L	7,80
		M	5,47
		N	2,20
		P	7,07
		Q	15,50
		R	1,63
		S	18,50
		T	21,51
		V	5,89
		Y	8,63

Table 8. Gal2p positions and mutations identified in the GTAC screen are ordered on preference. SCORE TOP HIT is the RelXyl score for the amino acid substitution within a position yielding the clearest improvement compared to wild-type Gal2p This allows a sorting of the most influential mutations and relevant positions to target in Gal2p to eliminate glucose affinity.

	GTAC	Mutations					
	Position	wt AA	Most preferred	Preferred	Least preferred	Inactive	SCORE TOP HIT

	GTAC	Mutations					
Most preferred	376	N	FILMTV	ACHPSY	DEKQRW		99,46
	449	T	FM	HIKLNQRSVWY	ACDEGP		92,65
	383	S	CQRTWY	AEFKMN	DGILV		83,94
	339	M	NQSV	AFGHKLRT	CDEIPY	W	83,73
	226	Y	ACDMNR	EQ	GHILPSTW	KV	81,68
	448	T	AM	CDEFGHIKLPQSV W	RY		77,64
	342	Q	ACSWY	DFHKPT	EGILMNRV		76,45
	380	T	AFMW	CDGIKLNQRSV	Y	E	74,78
	350	F	LR	NSTV	DKMQ	EHY	57,58
	214	Y	VW	LNQ	GMPRS	ACIK	55,48
	338	Q	CEF	ADHILRW	GNPSTVY	M	52,66
	219	T	AFG	CD	NS	EIKMQR W	52,55
	373	G	L	ACDEMNPVWY	FHIKQRST		51,94
	85	F	QTV	S	ACHMNP	DEGKR	50,51
Preferred	346	N	VWY	AHST	CDFGIKLMQR	E	47,09
	341	Q	L	MT	NSY	DEHR	46,71
	347	N	A	DEI	HL	GKQRST V	44,64
	451	A	G	CDEFILNPSW	HKMQRTVY		43,34
	89	T	DHV	ACL	N	FIKMQR SY	41,06
	343	L	K	PRSY	ACDEFGHIMNQT W	V	40,15
	444	F	AILVY	HK	CDEGMNPQSTW	R	39,43
	446	F	Q	HLMN	DEI	KRSTV	34,25
	478	N	FP	V	GIKQRWY	ACDHL	30,98
Least preferred	218	I	KNS	AHL	CDEGMV	RTW	27,86
	187	V	A	CFGQT	DEHIKLMRSWY		26,84
	215	Q	LM	CDF	AIS	GHNWY	24,63
	479	W	QST	CE	DFGHIKLMNPRVY		21,51
	222	I	D	ACH	EGMPVY	KTW	20,91
	191	A	N	DGIPT	CFHKQSVWY	ELMR	20,35
	455	W	ELMNT	PV	ACGKQSY	DFIR	19,92

The most prominent mutations were found at position 376 when the wild-type amino acid Asn was exchanged for amino acids with large hydrophobic side chains such as Phe, Ile, Leu, Met, Thr, or Val; these variants facilitated clear growth and almost full xylose consumption and almost no transport activity for glucose in the GTAC screen (Figure 8).

We have found that advantageous amino acids at position 376 have a van der Waals volume of 85 to 138 Å³ and a hydrophobicity of 10 to 100 Δt_R, or in an embodiment a van der Waals volume of 100 to 138 Å³ and a hydrophobicity of 10 to 100 Δt_R. A further important embodiment is specific amino acids at position 339, where we have found an amino acid that has a hydrophobicity of -30 to 10 Δt_R results in mutants with reduced glucose transport activity. This is illustrated by figure 11.

The values for van der Waals volume (Å³) for amino acids are herein used as described in: http://www.proteinsandproteomics.org/content/free/tables_1/table08.pdf. The corresponding literature is N. J. Darby, Thomas E. Creighton, Protein Structure (1993) Oxford University Press. The values for hydrophobicity (Δt_R) of amino acids are herein used as described in <http://onlinelibrary.wiley.com/doi/10.1002/psc.310010507/pdf>. The reference corresponding to this is Monera, O.D. et al, Journal of Peptide science Vol. 1 319-329 (1995).

Example 5: Xylose Transport Activity Screening

Aim. Using the Xylose Transport Activity (XTA) screen, *i.e.* using hexose transporter-mutant strain RN1053 as host to introduce *GAL2* variants and screening the resulting transformants on medium for growth on low xylose (1 g l⁻¹) concentrations, mutations can be identified that increase the xylose transport activity in the Gal2p variant. Activity of transport can be defined by more than one parameter, for instance, the affinity of the transporter (expressed by the Michaelis constant, *i.e.* K_m), or the rate of the transporter (expressed as V_{max}). It is also possible that a mutation increases the expression of the transporter, and thus improves xylose transport activity in the host cell.

Transformation and colony picking. RN1053 was transformed with a total of 468 constructs, each one bearing a *GAL2* mutant. One construct with wild-type *GAL2* sequence (SEQ. ID58) was included as control. For each transformation, 1-3 colonies were re-plated to agar medium MTPs amounting to 1229 transformants. The 1229 transformants were screened in two parts. For each part of the screening, wild-type *GAL2* was included as control.

Pre-culture and screen. For the pre-culture, transformants resulting from automated transformation and colony picking were transferred by automated process from selection agar medium MTPs to 96HDWP containing in each well 250 µl Verduyn-urea supplemented with 2% maltose. After 3 days of pre-culture in an orbital shaker at 30 °C and 750 rpm, 5 µl of culture was transferred to three different 96HDWP containing 250 µl Verduyn-urea supplemented with 1 g l⁻¹ xylose (RN01-medium), each 96 HDWP representing a sampling point. On each plate (24 DWP or 96 HDWP) at least one RN1001 growth control was inoculated to have an indication of plate-to-plate effects. After 24 hours, 72 hours and 96 hours a series of 96HDWPs was harvested for automated OD-measurement at 600 nm. For each construct and each time point the OD600 values for the different replicates was averaged. In order to compare the different parts of the screen, a relative value was calculated based on the difference to the wild-type OD600 value measured in the particular part of the XTA screen, according to the following formula:

$$\text{RelOD600} = \frac{(\text{OD600}_{\text{Gal2-mutant}} - \text{OD600}_{\text{Gal2-wildtype}})}{\text{OD600}_{\text{Gal2-wildtype}}} \times 100\%$$

Results. In both parts of the XTA screen RN1001 is part of the TOP15 positions based on the growth profile compared to wild-type Gal2p; RN1001 has the full complement of hexose transporters and the presence of RN1001 in the TOP15 in both parts of the XTA Screen (Figure 9), indicates that one or more endogenous hexose transporters in yeast facilitate high affinity xylose uptake in contrast with the wild-type Gal2 RN1053-transformants which displayed poorer growth profiles. The mutant variants with the single amino acid changes present in the TOP15 displayed an increased affinity to xylose compared to wild-type Gal2p and developed towards the RN1001 growth profile or even showed similar or improved growth characteristics on low xylose concentrations. Interestingly, when aligning amino acid sequences of the *S. cerevisiae* hexose transporter family, some of the Gal2p mutations found in the TOP15, e.g. N346D and M339S, are to be found in the wild-type sequences of Hxt2 (S324) and Hxt11 (D336 and S329), respectively. The comparison of all tested mutant GAL2 variants versus wild-type GAL2 on growth (average RelOD600) at 96 hours is listed in Table 8. All mutations with RelOD600>0 are proposed to have a positive effect on the xylose affinity.

Table 9. Average RelOD600 values for the screened Gal2p mutants in the XTA Screen.

Wild-type GAL2 construct pDB1250 was set to 0. RelOD600 values are the averages of 1-3 replicates and are relative values compared to wild-type.

5

Pos	wtAA	MutAA	RelOD600
85	F	A	-2,50
		C	-1,81
		D	-32,03
		E	-37,55
		G	-51,06
		H	-35,08
		K	-27,75
		N	-6,14
		P	-23,27
		Q	-0,27
		R	-18,30
		S	7,30
		T	-6,07
		V	2,40
89	T	A	-1,12
		C	19,39
		E	-13,89
		I	-20,86
		K	-13,89
		L	1,12
		M	-9,48
		N	-8,70
		Q	-4,57
		R	-10,12
		V	3,11
		W	-43,47
		Y	0,01
187	V	A	-10,62
		C	14,34
		D	-5,07
		E	-2,94
		F	-6,14
		G	14,77
		H	-2,79
		I	15,55

Pos	wtAA	MutAA	RelOD600
		L	10,08
		M	36,83
		Q	26,77
		R	-11,36
		S	-28,01
		T	-10,54
		W	-17,20
		Y	11,95
191	A	C	5,10
		D	7,05
		E	0,58
		F	-1,81
		G	-4,01
		H	-12,93
		I	10,13
		K	7,11
		M	-4,82
		N	4,60
		P	1,90
		Q	-8,47
		R	3,34
		S	-19,15
		T	-2,00
		V	-9,91
		W	4,10
		Y	-15,24
214	Y	A	-3,04
		C	-12,04
		G	-6,92
		I	-6,92
		K	-4,15
		L	6,02
		M	-26,59
		N	-9,60
		P	5,35
		Q	-26,44
		R	-29,77
		S	-37,90
		W	-21,91
215	Q	A	1,54
		C	5,92

Pos	wtAA	MutAA	RelOD600
		D	25,93
		E	20,81
		F	3,28
		I	-5,50
		L	8,24
		M	11,19
		R	-48,74
		S	16,66
		V	-21,07
		W	-39,22
		Y	-24,30
218	I	A	-15,52
		C	-13,03
		D	-19,50
		E	-15,31
		G	-9,12
		H	-9,12
		K	-10,19
		L	-15,45
		M	-24,27
		N	-23,20
		R	-13,39
		S	-13,60
		T	-15,95
		V	-4,00
		W	40,30
219	T	A	-6,71
		C	-20,53
		D	-21,69
		E	3,47
		F	-17,01
		G	-35,80
		I	-47,23
		L	3,78
		M	-20,00
		N	-13,93
		P	-18,27
		Q	-14,25
		R	7,17
		S	-18,08
		V	11,19

Pos	wtAA	MutAA	RelOD600
		W	-15,13
222	I	A	-7,91
		C	-6,92
		D	-12,54
		E	-16,80
		G	28,78
		H	-13,50
		K	-48,17
		M	20,18
		P	11,01
		R	-0,62
		S	3,03
		T	10,25
		V	-12,55
		W	11,38
		Y	23,82
226	Y	A	23,92
		C	-6,78
		D	-12,89
		E	5,10
		F	-28,04
		G	-4,64
		H	-11,26
		I	-12,46
		K	-10,55
		M	24,16
		N	-4,86
		P	20,39
		Q	0,33
		R	-9,19
		S	-5,78
		T	23,80
		W	-5,71
338	Q	A	4,10
		C	10,88
		D	7,99
		E	3,59
		F	0,45
		G	5,29
		H	11,63
		I	12,95

Pos	wtAA	MutAA	RelOD600
		L	0,58
		M	4,79
		N	14,12
		P	2,21
		R	15,03
		S	-17,02
		T	-26,76
		V	-11,33
		W	-6,92
		Y	-17,30
339	M	A	28,42
		C	20,32
		D	-0,09
		E	18,97
		F	4,17
		G	-20,00
		H	-14,46
		I	-17,55
		K	-11,75
		L	21,03
		N	4,25
		P	-4,64
		Q	-3,65
		R	-7,91
		S	54,52
		T	10,79
		V	-11,90
		W	6,56
		Y	38,66
341	Q	D	-10,62
		E	-8,13
		H	2,89
		K	-21,66
		L	-22,79
		M	-11,30
		N	-13,31
		R	-10,76
		S	7,16
		T	-3,08
		V	18,89
		Y	9,37

Pos	wtAA	MutAA	RelOD600
342	Q	A	-15,66
		C	2,04
		D	-12,54
		E	-9,23
		F	33,22
		G	42,01
		H	45,21
		I	34,38
		K	12,92
		L	28,91
		M	7,99
		N	45,31
		P	-7,15
		R	44,36
		S	22,25
		T	20,62
		V	-16,07
		W	32,62
		Y	-21,41
343	L	A	15,98
		C	7,16
		D	-12,82
		E	9,65
		F	-6,07
		G	-6,63
		H	-9,83
		I	19,04
		K	2,40
		M	-0,73
		N	0,97
		P	4,53
		Q	19,53
		R	-10,62
		S	10,72
		T	-6,46
		W	0,69
		Y	29,99
346	N	A	26,36
		C	15,73
		D	69,10
		F	10,15

Pos	wtAA	MutAA	RelOD600
		G	24,01
		H	-4,93
		I	0,76
		K	-1,66
		P	-0,80
		Q	-5,28
		R	-7,84
		V	-44,97
		Y	-3,51
347	N	D	5,29
		E	-24,11
		G	7,59
		H	-4,57
		I	-3,08
		K	9,65
		L	14,52
		Q	6,55
		R	-1,93
		S	18,92
		T	2,61
		V	6,09
350	Y	D	-2,08
		E	39,52
		I	-28,19
		L	33,56
		M	7,09
		Q	6,95
		R	-40,48
		S	7,93
		T	-6,33
		V	-15,81
		Y	-9,27
373	G	A	-51,50
		E	-50,72
		F	-47,17
		H	-37,99
		I	2,84
		K	15,40
		L	18,36
		M	14,21
		N	21,06

Pos	wtAA	MutAA	RelOD600
		Q	14,02
		R	32,18
		S	18,92
		T	-11,04
		V	-6,71
		W	-5,20
		Y	1,21
376	N	A	10,19
		C	24,67
		D	8,93
		E	10,19
		F	-13,02
		H	-13,78
		I	14,59
		K	47,76
		L	-9,63
		M	6,29
		P	29,04
		Q	35,51
		R	-23,42
		S	44,42
		V	-42,40
		W	-15,74
		Y	-21,28
380	T	A	3,56
		C	8,37
		D	24,95
		E	14,21
		F	-17,33
		G	29,22
		I	20,81
		K	14,34
		L	23,45
		M	29,16
		N	-50,25
		P	-7,46
		Q	-15,32
		R	-0,55
		S	-6,02
		V	-6,21
		W	14,78

Pos	wtAA	MutAA	RelOD600
		Y	29,48
383	S	A	16,05
		C	3,32
		D	-2,94
		E	1,97
		F	-15,10
		G	-4,50
		I	2,40
		K	-47,86
		L	3,59
		M	-2,56
		N	-44,72
		Q	14,40
		R	21,37
		T	-28,26
		V	24,58
		W	26,71
		Y	27,84
444	F	C	-6,49
		D	-2,08
		E	-16,16
		G	1,54
		H	-16,38
		I	-20,29
		K	-24,06
		L	-10,19
		M	-7,20
		N	-9,76
		P	-18,79
		Q	-27,18
		S	-14,45602
		W	-16,94
		Y	-18,08
446	F	D	19,61
		E	-36,17
		H	30,98
		I	0,83
		L	-40,55
		M	-2,58
		N	-10,90
		Q	-35,54

Pos	wtAA	MutAA	RelOD600
		R	-37,62
		S	-2,94
		T	-39,41
		V	-47,79
448	T	A	10,57
		C	0,08
		D	15,78
		E	19,55
		F	-6,65
		G	16,91
		H	25,64
		I	-0,18
		K	26,65
		L	22,31
		M	0,77
		P	12,58
		Q	14,08
		R	29,22
		S	11,19
		V	
		W	8,09
		Y	-2,01
449	T	A	4,67
		C	6,59
		D	-2,58
		E	-2,08
		F	16,48
		G	11,00
		H	16,26
		I	9,51
		K	-3,51
		L	11,29
		M	-15,59
		N	-7,13
		P	8,73
		Q	13,21
		R	-10,33
		S	15,48
		V	14,41
		W	-5,78
		Y	-5,57

Pos	wtAA	MutAA	RelOD600
451	T	C	-33,23
		D	-23,20
		F	-4,75
		G	-16,94
		H	-41,76
		I	-40,26
		K	-8,91
		L	-23,73
		N	-6,07
		P	-6,14
		Q	-8,22
		R	-6,21
		S	2,90
		T	-11,42
		V	-5,89
		W	4,10
		Y	3,91
455	W	C	-14,75
		D	8,18
		E	4,22
		F	-2,31
		G	-10,67
		I	5,45
		K	8,12
		L	-5,96
		M	6,39
		N	-9,22
		P	0,83
		Q	0,95
		R	-48,55
		S	-4,82
		T	-9,44
		V	-2,06
		Y	-19,81
478	N	A	-25,74
		C	-14,75
		D	-9,91
		F	6,86
		G	0,01
		H	-13,12
		I	8,81

Pos	wtAA	MutAA	RelOD600
		K	-24,30
		P	-20,47
		Q	14,40
		R	-27,44
		W	-13,40
		Y	6,77
479	W	C	-12,96
		D	15,20
		E	18,40
		F	24,87
		H	1,83
		I	48,83
		K	8,58
		L	-3,79
		M	7,87
		N	8,05
		P	-42,40
		Q	-33,73
		R	13,42
		S	57,22
		T	15,48
		V	-12,54
		Y	12,07

5 Table 10. Gal2p positions and mutations identified in the XTA screening are ordered on preference. SCORE TOP HIT is the RelXyl score for the amino acid substitution within a position yielding the clearest improvement compared to wild-type Gal2p This allows a sorting of the most influential mutations and relevant positions to target in Gal2p to increase xylose transport activity.

	XTA	Mutations					
	Position	wt AA	Most preferred	Preferr ed	Least preferred	Inactive	SCORE TOP HIT
Most	346	N	D	AG	CF	HIKPQRVY	69,10
	479	W	S	I	DEFHKMN RTY	CLPQV	57,22

	XTA	Mutations					
	339	M	SY	ACEL	FNTW	DGHIKPQRV	54,52
	342	Q	GHN R	FILW	CKMST	ADEPVY	47,76
	376	N	KS	CPQ	ADEIM	FHLRVWY	47,76
Preferred	218	I	W			ACDEGHKLMN RSTV	40,30
	350	F	E	L	MQS	DIRTVY	39,52
	187	V	MQ	CGI	LY	ADEFHRSTW	36,83
	373	G	R	KLMN QS	IY	AEFHTVW	32,18
	446	F	H	D		EILMNQRSTV	30,98
Least preferred	343	L	Y	AIQ	CEKSP	DFGHMNRW	29,99
	380	T	GM Y	DIL	ACEKW	FNPQRSV	29,48
	448	T	HKR	EL	ADGPQSW	MCFIVY	29,22
	222	I	G	MY	PTW	ACDEHKRSV	28,78
	383	S	VWY	AQR	CEIL	DFGKMNT	27,84
	215	Q	DE	S	ACFL	IMRVWY	25,93
	226	Y	AMPT		E	CDFGHIKNQR SW	24,16
	89	T	C		LVY	AEIKMNQRW	19,39
	347	N	S	KL	DGQTV	EHIR	18,92
	449	T	FH	QSV	ACGILP	DEKMNRWY	16,48
	338	Q	NR	CDHI	AEFGLMP	STVWY	15,03
	478	N	Q	FIY		ACDGHKPRW	14,40
	191	A	I	DK	CENPRW	FGHMQSTVY	10,13
	341	Q	V	SY	H	DEKLMNRT	9,37
	455	W	DK	EIM		CFGLNPQRST VY	8,18
	85	F	S	V		ACDEGHKNPQ RT	7,30
	219	T	V	R	EL	ACDFGIMNPQ SW	7,17
	214	Y	LP			ACGIKMNQRS	6,02

	XTA	Mutations					
						W	
	451	A	WY	S		CDFGHIKLNP QRTV	4,10
	444	F			G	CDEHIKLMNP QSWY	1,54

EXAMPLES 6 to 15**Methods**

5

Molecular biology techniques and chemicals. Restriction enzymes and T4 DNA ligase were acquired from Fermentas (Fisher Scientific, Landsmeer, the Netherlands). Primers used in the studies (SEQ ID NO: 1-55) are indicated in Table 11. Standard molecular biology and yeast genetics techniques were conducted according to textbooks including Sambrook *et al.* (2001; Molecular cloning : a laboratory manual, third edition, Cold Spring Harbor Press, Cold Spring Harbor, New York, NY, USA) and Ausubel *et al.* (1995; Current Protocols in Molecular Biology).

10

15

20

25

PCR amplification and cloning. For PCR amplifications, Phusion® High-Fidelity PCR Master Mix with HF buffer was used (Finnzymes; Fisher Scientific, Landsmeer, the Netherlands). Primers used for cloning and sequencing are indicated in Table 11 (SEQ ID NO: 37-55). *HXT2*, *HXT3-6*, and *HXT11* PCR fragments were cloned into the yeast expression vector pRS313-P7T7 (SEQ ID NO 56), based on the shuttle vector pRS313 (Sikorski & Hieter, 1989, A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in *Saccharomyces cerevisiae*. Genetics, 122, pp. 19-27). Construct pRS313-P7T7 bears the truncated *HXT7* promoter (Hauf *et al.*, 2000, Simultaneous genomic overexpression of seven glycolytic enzymes in the yeast *Saccharomyces cerevisiae*. Enzyme Microb Technol, 26:688-698) and the *HXT7* terminator. In between the promoter and the terminator, a multiple cloning site (MCS) exists. Resulting PCR fragments, were digested by restriction enzyme combinations (as further indicated in the Examples below) and cloned by standard molecular biology techniques into the yeast expression vector. *HXT11* was also cloned into pRS313-P7T7-GFP (resulting in SEQ ID NO: 57) for localization studies of the GFP-tagged Hxt11 protein using fluorescence microscopy (Example 5). pDB1250 contained the *GAL2* ORF

(SEQ ID NO:60) between the truncated *HXT7* promoter and the *ADH1* terminator (for reference of sequence pDB1250 see SEQ ID NO: 58. Plasmids were amplified and maintained in DH5 α cells following manufacturer's instructions. Plasmids were isolated from *E. coli* mini cultures using the GeneElute plasmid Miniprep kit (Sigma-Aldrich, Zwijndrecht, the Netherlands). Gene/Protein sequences and constructs used and generated during these studies are listed in Table 12.

Table 11: oligonucleotides used in the examples

Name	Sequence (5' $\rightarrow$ 3')	purpose	SEQ ID NO
ActinF	GGATTCTGAGGTTGCTGCTTTGG	Real time PCR	62
ActinR	GAGCTTCATCACCAACGTAGGAG	Real time PCR	63
HXT1F	TGTTCTCTGTACACCGTTGACCG	Real time PCR	64
HXT1R	AGATCATACAGTTACCAGCACCC	Real time PCR	65
HXT2F	CTTCGCATCCACTTTCGTG	Real time PCR	66
HXT2R	AATCATGACGTTACCGGCAGCC	Real time PCR	67
HXT3F	GAAGCTAGAGCTGCTGGTTCAGC	Real time PCR	68
HXT3R	ACAACGACATAAGGAATTGGAGCC	Real time PCR	69
HXT4F	ATGGAGAGTTCCATTAGGTCTAGG	Real time PCR	70
HXT4R	ATAACAGCTGGATCGTCTGCGC	Real time PCR	71
HXT5F	TTGCTATGTCGTCTATGCCTCTG	Real time PCR	72

HXT5R	AGATAAGGACATAGGCAACGGG	Real time PCR	73
HXT7F	GGGTGCTGCATCCATGACTGC	Real time PCR	74
HXT7R	ACAACGACATAAGGAATTGGAGCC	Real time PCR	75
HXT8F	GTACTACTATCTTCAAATCTGTCGG	Real time PCR	76
HXT8R	CTTGTGACGCCAACGGAGGCG	Real time PCR	77
HXT9F	CCATTGAGAGGTTTGGACGCCG	Real time PCR	78
HXT9R	ACACAATCATACAGTTACCGGCG	Real time PCR	79
HXT10F	GGAATGCAAGACTCTTTCGAGAC	Real time PCR	80
HXT10R	CTAGTGACGCCAACGGTGGCG	Real time PCR	81
HXT11F	GCCACTCAATGGAGAGTCGGC	Real time PCR	82
HXT11R	CAACTAGCAAGGCTGGATCGTC	Real time PCR	83
HXT12F	CACCATCTTCAAATCTGTCGGTC	Real time PCR	84
HXT12R	CAATCATACAGTTACCGGCACCC	Real time PCR	85
HXT13F	CCCTCATGGCCAGGACGGTC	Real time PCR	86
HXT13R	TTGCCATAACCAGTTGCATGCAG	Real time PCR	87
HXT14F	GCCTTAGTAGTGTACTGCATCGGT	Real time PCR	88
HXT14R	TGATACGTAGATACCATGGAGCC	Real time	89

		PCR	
HXT15F	GAGGCCTGTGTCTCCATCGCC	Real time PCR	90
HXT15R	CACAAGAATACCTGTGATCAAACG	Real time PCR	91
HXT16F	CAAGGAAGTATAGTAATACTGCGC	Real time PCR	92
HXT16R	TTGGCGATGGAGACACAGGCC	Real time PCR	93
HXT17F	TAACACTGCACAATGGAGAGTCC	Real time PCR	94
HXT17R	TGAGTACCCATGGATCCTCTGG	Real time PCR	95
GAL2F	TCAATGGAGAGTTCCATTAGGGC	Real time PCR	96
GAL2R	CTGGACGGCAGGATCCTCTGG	Real time PCR	97
KOP11*	AATAATCATTGCACAATTTAGTTCTAAACGCTTTTGTTA T TACTCAATATCCGTTTTAAGAGCTTGGTGAGCGCTAG GAGTC	KO HXT11	98
KOT11*	TCGTCAATTTTTTTTTTTGCTTTTTTACCAATTTACCGA AAACTAGAAGAGAGAGTTCAAGAGAAAAAAAAGAAAA GCAAAAAGAAAAAGGAAAGCGCGC	KO HXT11	99
iHXT11F	GGCCTCTAGATCAGCTGGAAAAGAACCTCTTGTAAT TG	Inverse HXT11	100
iHXT11R	GCTAGGATCCATGTCAGGTGTTAATAATACATCCGCA AATG	Inverse HXT11	101
HXT11F	GGCCTCTAGAATGTCAGGTGTTAATAATACATCCGC	Cloning	102
HXT12F	GGCCTCTAGAATGGGTTTGATTGTCTCAATATTCAAC	Cloning	103
HXT11/1 2R	CGATGGATCCTCAGCTGGAAAAGAACCTCTTGTAAT TG	cloning	104
HXT1 Xbal	GCATTCTAGAATGAATTCAACTCCCGATCTAATATC	cloning	105

R HXT1 Cfr9i	TGCATCCCGGGTTATTTCTGCTAAACAACTCTTGTA	cloning	106
F HXT2 Xbai	GTCCTCTAGAATGTCTGAATTCGCTACTAGCCG	cloning	107
R HXT2 Cfr9i	CATCGCCCGGGTTATTCCTCGGAACTCTTTTTCTTT TG	cloning	108
F HXT3 Xbal	GCATTCTAGAATGAATTCAACTCCAGATTTAATATCTC	cloning	109
R HXT6 Cfr9I	CATCGCCCGGGTTATTTGGTGCTGAACATTCTCTTG	cloning	110
F HXT4 Xbal	GTCCTCTAGAATGTCTGAAGAAGCTGCCTATCAAG	cloning	111
R HXT4RN Cfr9I	TATCGCCCGGGTTAATTAAGTACCTACTTTTTTCCGA	cloning	112
F HXT5 Xbal	GTCCTCTAGAATGTCGGAAGTTGAAAACGCTCATC	cloning	113
R HXT5 Cfr9I	GCATCCCGGGTTATTTTTCTTTAGTGAACATCCTTTTA TA	cloning	114
F HXT7 Xbal	GTCCTCTAGAATGTCACAAGACGCTGCTATTGCA	cloning	115
R HXT7 Cfr9I	CATCGCCCGGGTTATTTGGTGCTGAACATTCTCTTG	cloning	116

* In *italics* the HXT11 flanking sequence, and underlined the HIS3 sequence

RNA extraction and cDNA synthesis. Total RNA was isolated from yeast cells in exponential phase by a glass-bead disruption/Trizol extraction procedure. Yeast pellets from 2 ml of cell culture were mixed with 0.2 ml of glass beads (diameter, 0.45 mm) and 0.9 ml of Trizol with 125 µl chloroform, and disrupted in a Fastprep FP120 (Bio-101, Thermo Savant, California, USA) by a 45-second burst at speed 6. 1 µg of total RNA was used to synthesize cDNA using the iScript Kit (Bio-Rad, Veenendaal, the Netherlands).

Real-time PCR. The *HXT1-HXT17* and *GAL2*-specific real-time PCRs were performed, respectively, using SensiMix SYBR & Fluorescein kit (GC Biotech, Alphen

aan den Rijn, the Netherlands) and the MyiQ iCYCLER Real Time PCR instrument (BIO-Rad, Veenendaal, the Netherlands). Each 25- μ l reaction contained 12.5 μ l of SYBR green Master Mix, 4 μ l cDNA, 0.5 μ l of each primer (10 nM) and 7.5 μ l of sterile deionized water. The PCR conditions were 10 min at 95 °C followed by 39 cycles of amplification (15 sec at 95 °C, 30 sec at 60 °C, 30 sec at 72 °C). The primers shown in Table 11 were utilized to amplify the cDNA fragments by PCR amplification (SEQ ID NO's 1-36).

Error prone PCR. Error-prone PCR experiments were performed following the indications provided by the DNA Taq polymerase (Thermo Fischer) using 10 ng of template in 100 μ l of PCR mix containing 5.5 mM MgCl₂ and 0.15 mM MnCl₂.

Strain maintenance, cultivation and evolutionary engineering. Strains generated in these studies are listed in Table 13. For storage of the strains used in this work (Table 2), shake flask cultures were performed in rich medium (YP), consisting of 10 g l⁻¹ yeast extract (Oxoid) and 20 g l⁻¹ peptone (BD Difco), supplemented with either 2% glucose (YPD), 2% maltose (YPM; in case of RN1053-derivatives), or 3% xylose (YPX; in case of YD01227-derivatives). Cultures were maintained at 30 °C in an orbital shaker until cultures reached stationary growth phase. After adding glycerol to 30% (v/v), samples from shake-flask cultures were stored in 2 ml aliquots at -80 °C.

For strain characterizations and evolutionary engineering, cultivations were conducted using mineral medium according to Verduyn using urea as nitrogen source (Verduyn-urea Luttik *et al*, 2001, J Bacteriol, 182:501-517) at pH 4.5 supplemented with the desired sugar (mixtures). In cultures of the model strains RN1053 or YD01227, Verduyn-urea was supplemented with 0.2 g l⁻¹ histidine (Sigma-Aldrich; Verduyn-urea-his) to complement for the histidine auxotrophy.

Cultivations of strains for characterization of growth and sugar consumption profiles were conducted in shake flasks, a Bioscreen C reader using honeycomb well plates (Growth Curves Ltd, represented by Thermo Fisher Scientific BV, Breda, the Netherlands). Cultures were maintained at a temperature of 30°C. Specifically for the purpose of evolutionary engineering, chemostat cultures were grown in a 3 L stirred tank bioreactor (Applikon, Schiedam, the Netherlands) filled with 500 ml of Verduyn-urea-his supplemented with the required carbon sources at a temperature of 30°C. Starting Dissolved Oxygen (DO) setpoint was 5%, stirring was performed at 400 rpm and the starting OD₆₀₀ was 0.2.

Table12: plasmids and gene/protein sequences

<u>Plasmid/gene/protein</u>	<u>Sequence type</u>	<u>SEQ ID NO or reference</u>	<u>Example</u>
pRS313-P7T7 (empty vector control)	Artificial DNA	117	7, 8, 9, 10, 11, 13, 15
pRS313-P7TA-GAL2 (pDB1250)	Artificial DNA		12
pRS313-P7T7-iHXT11	Artificial DNA		7
pRS313-P7T7-HXT11	Artificial DNA		8, 9, 10
pRS313-P7T7-HXT12	Artificial DNA		8
pRS313-P7T7-HXT2	Artificial DNA		9, 10
pRS313-P7T7-HXT11-GFP	Artificial DNA	57	10
pRS313-P7T7-mHXT11(N366D)	Artificial DNA		11
pRS313-P7T7-HXT36	Artificial DNA		15
pRS313-P7T7-mHXT36(N367I)	Artificial DNA		15
<i>HXT11</i>	DNA <i>S. cerevisiae</i>	58	7, 8, 9, 10, 11, 12
<i>HXT2</i>	DNA <i>S. cerevisiae</i>	59	9, 10
<i>GAL2</i>	DNA <i>S. cerevisiae</i>	60	12
<i>HXT3-6</i>	DNA <i>S. cerevisiae</i>	61	15
Hxt11p	Protein, <i>S. Cerevisiae</i>	62	7, 8, 9, 10, 11, 12

Hxt2p	Protein, S. <i>Cerevisiae</i>	63	9
Gal2p	Protein, S. <i>Cerevisiae</i>	64	10
Hxt36p	Protein, S. <i>Cerevisiae</i>	65	15

Table 13: Strains used or prepared herein

<u>Strain</u>	<u>Genotype</u>	<u>Example</u>
RN1001	<i>Mat a, ura3-52, leu2-112, gre3::loxP, loxP-Ptpi:TAL1, loxP-Ptpi::RKI1, loxP-Ptpi-TKL1, loxP-Ptpi-RPE1, delta::Padh1XKS1Tcyc1-LEU2, delta::URA3-Ptpi-xylA-Tcyc1</i>	10
RN1014	RN1001 + <i>in vivo</i> engineering on xylose and acetic acid	Reference
RN1041	RN1001 <i>his3::loxP</i>	WO2013081456
RN1041-empty	RN1041, pRS313-P7T7 (empty vector control)	8, 11
RN1053	RN1041 <i>hxt2::loxP-kanMX-loxP, hxt367::loxP-hphMX-loxP, hxt145::loxP-natMX-loxP, gal2::loxP-zeoMX-loxP</i>	6, 7
RN1053-X2	RN1053 single colony selected on YPX after chemostat evolutionary engineering on 2% xylose	6, 7
RN1053-X2- <i>hxt11Δ</i>	RN1053-X2, <i>hxt11::HIS3</i>	7
RN1053-empty	RN1053, pRS313-P7T7 (empty vector control)	8, 9, 10, 15
RN1053-HXT2	RN1053, pRS313-P7T7-HXT2	9
RN1053-HXT3-6	RN1053, pRS13-P7T7-HXT3-6	15
RN1053-mHXT3-6(N367I)	RN1053, pRS13-P7T7-mHXT3-6(N367I)	15
RN1053-HXT11	RN1053, pRS13-P7T7-HXT11	8, 9, 10, 11, 13
RN1053-mHXT11-	RN1053, pRS313-P7T7-mHXT11(N366D)	13

N366D		
RN1053-iHXT11	RN1053-pRS313-P7T7-iHXT11	7
RN1053-HXT11-GFP	RN1053, pRS313-P7T7-HXT11-GFP	10
RN1053-HXT12	RN1053, pRS13-P7T7-HXT12	8
YD01227 (ori)	RN1014 <i>glk1::lox72; hxx1::loxP; hxx2::lox72; gal1::loxP; his3::loxP</i> natMXloxP	14
YD01227-evoA, -B, and -C	YD01227, three single colonies selected on plates with 1% xylose + 10% glucose after chemostat evolutionary engineering runs (see Example 14)	14
YD01227-empty	YD01227, pRS313-P7T7 (empty vector control)	13
YD01227-HXT2	YD01227, pDB1162 (pRS313-P7T7-HXT2)	12
YD01227-GAL2	YD01227, pDB1250 (pRS313-P7TA-GAL2)	12
YD01227-HXT11	YD01227, pDB1152 (pRS313-P7T7-HXT11)	12, 13
YD01227-mHXT11(N366D)	YD01227, pRS313-P7T7-mHXT11(N366D)	13

Analytical methods. Cell growth was monitored by optical density (OD) at 600 nm using UV-visible spectrophotometer (Novaspec PLUS). The concentrations of glucose, xylose, ethanol were measured in supernatant of cultures (separated from cell pellet after centrifugation at 4000 rpm for 5 min) by High Performance Liquid Chromatography (Shimadzu, Kyoto, Japan) using an Aminex HPX-87H column (Bio-Rad) and a refractive index detector (Shimadzu, Kyoto, Japan). The temperature of the column and detector was maintained at 65°C. The mobile phase was 0.005 N H₂SO₄ at a flow rate of 0.55 ml/min.

Sugar Uptake measurement. The uptake of radio-labeled xylose was measured as follows: the cells were grown for 24 hours in shake flasks in Verduyn-urea supplemented with 2% xylose and 0.05% maltose and were collected by centrifugation (3,000 rpm, 3 min, 20°C), washed and re-suspended in Verduyn-urea. [14C] xylose or [14C] glucose (CAMPRO scientific, Veenendaal, the Netherlands) stocks were added to the cell suspension. The reaction was stopped, for xylose after 1 minute and for glucose after 15 seconds, by addition of 5 ml of 0.1M lithium chloride, and the cell suspension was filtered (0.45 µm HV membrane filter, Millipore, France). The filters were washed

with another 5 ml of lithium chloride and counted with Liquid Scintillation Counter in the emulsifier scintillator plus (Perkin-Elmer, USA). Uptake experiments for YD01227-ori and YD01227-evo were done with 0.5, 2, 6, 20, 40mM xylose or 0.1, 0.4, 1.5, 6, 20, 80mM glucose. Glucose competition studies for RN1053-HXT3-6 and RN1053-mHXT3-6(N367I), RN1053-HXT11 and RN1053-mHXT11 (N366D) were performed with [14C] xylose stock and with unlabeled glucose. The final xylose and glucose concentrations were 50 mM and 50-500 mM, respectively.

Fluorescence microscopy. Plasmid pRS313-P7T7-HXT11-GFP (SEQ ID NO 57) was transformed into strain RN1053 and RN1041. Fresh colonies were inoculated into minimal medium with xylose or glucose. A fresh liquid cell culture taken in exponential growth phase (at an optical density of 10 at 600 nm) was subjected to fluorescence microscopy under a Nikon Eclipse-Ti microscope equipped with a 100× oil immersion objective, a filter set for GFP, and a Nikon DS-5Mc cooled camera. We routinely examined at least 100 cells per sample, and each experiment was replicated at least three times.

Automated transformation and colony picking. For the generation of transformation of a saturation mutagenesis library into YD01227, shake-flask cultures were performed in either YPM for RN1053, or YPX for YD01227 (see below). Yeast cells were pelleted and, subsequently, used in an automated transformation protocol based on Gietz,R.D. and Woods,R.A. (Gietz,R.D. and Woods,R.A. 2006, Yeast transformation by the LiAc/SS Carrier DNA/PEG method. Methods Mol. Biol., 313:107-120). Transformation mixtures were plated on selection medium consisting of yeast nitrogen base (Sigma-Aldrich; 6.7 g l⁻¹), agar (BD Biosciences; 15 g l⁻¹), supplemented with either 2% maltose (RN1053 transformations) or 3% xylose (YD01227 transformations). Transformation plates were incubated at 30 °C, and after colony formation, colonies were re-plated using an automated process transferring colonies to 96 well microtiter plates (MTP) containing the above-referred selection media. MTPs with transformants were incubated at 30 °C until clear growth was observed.

NMR analysis. For the quantification of glucose, xylose, glycerol, acetic acid and ethanol in the sample, 100 µl sample is transferred accurately into a suitable vial. Subsequently 100 µl internal standard solution, containing maleic acid (20 g/l), EDTA (40 g/l) and trace amounts of DSS (4,4-dimethyl-4-silapentane-1-sulfonic acid) in D₂O, and 450 µL D₂O is added.

1D ^1H NMR spectra are recorded on a Bruker Avance III 700 MHz, equipped with a cryo-probe, using a pulse program with water suppression (power corresponding to 3 Hz) at a temperature of 27°C.

The analyte concentrations are calculated based on the following signals (δ relative to DSS):

- α -glucose peak at 5.22 ppm (d, 0.38 H, $J = 4$ Hz),
- α -xylose peak at 5.18 ppm (d, 0.37 H, $J = 4$ Hz),
- glycerol peak at 3.55 ppm (dd, 2 H, $J_{1,2} = 6$ Hz and $J_{1a,1b} = 12$ Hz)
- acetic acid peak at 1.91 ppm (s, 3 H)
- ethanol peak at 1.17 ppm (t, 3 H, $J = 7$ Hz)
- The signal user for the standard:
- Maleic acid peak at 6.05 ppm (s, 2H)

EXAMPLE 6

ELEVATED EXPRESSION OF *S. CEREVISIAE* HXT11 IN EVOLVED XYLOSE TRANSPORT-NEGATIVE STRAIN WITH RESTORED ABILITY TO GROW ON XYLOSE

RN1053 chemostat culture. Uptake of xylose in *Saccharomyces cerevisiae* is facilitated by a distinct subgroup of hexose transporters (Hamacher *et al*, 2002, Characterization of the xylose-transporting properties of yeast hexose transporters and their influence on xylose utilization. Microbiology, 148: 2783-2788). We have constructed a deletion mutant with xylose-fermenting capabilities (RN1053) which lacks the major hexose transporters *HXT1-HXT7* and *GAL2*. As a result of the deletions, RN1053 is not able to utilize xylose (as described in Example 1). *Saccharomyces cerevisiae* possesses more HXT genes than the eight major ones (*HXT8-17*) of which not much is known rather than their expression is low or negligible (Sedlak & Ho, 2004, Characterization of the effectiveness of hexose transporters for transporting xylose during glucose and xylose co-fermentation by a recombinant *Saccharomyces yeast*. Yeast, 21, pp. 671-684) or that they have been connected to physiological roles other than sugar transport (Nourani *et al* 1997 Multiple-drug-resistance phenomenon in the yeast *Saccharomyces cerevisiae*: involvement of two hexose transporters. Mol Cell Biol, 17: 5453-5460). Therefore we refer to them as the cryptic *HXT* genes. An attempt was made to select for possible spontaneous mutations in cryptic HXT loci resulting in improved expression or

possibly improved affinity for xylose. In order to do so, strain RN1053 was grown in an anaerobic, xylose-limited chemostat culture at a dilution rate of 0.05 h⁻¹. The evolved strains by chemostat cultivation, RN1053-X2, was isolated on 2% xylose plate. A single colony was selected for cultivation in aerobic shake-flask on Verduyn-urea-his supplemented with 2% xylose for 96 h for comparison to the original RN1053 strain. The growth curves of RN1053-X2 and RN1053 strain were represented in Fig. 12A. The evolved strain RN1053-X2 was able to grow on 2% xylose, whereas the original strain RN1053 did not grow on the xylose medium. A reason could be that one or more of the cryptic *HXT*-genes (*HXT8-HXT17*) was up-regulated on xylose medium during the chemostat culture facilitating the uptake of xylose.

Expression profiling evolved RN1053-X2. The expression patterns of *HXT8-HXT17* in the evolved strain RN1053-X2 and original strain RN1053 were compared during batch cultivations on 2% xylose medium. The transcription level of *HXT11* and *HXT12* were dramatically increased by up to 8-fold in the evolved strain RN1053-X2 from the beginning of exponential phase (Fig. 12B) and reached a maximum level at day 3, compared to original strain RN1053. The chemostat was effective for the evolution of mutants to enhance xylose consumption by expressing *HXT11* and/or *HXT12* transporter genes in strain RN1053. However, expression level of other *HXT*-genes (i.e. *HXT8-HXT10*, and *HXT13-HXT17*) were repressed on xylose medium, compared to wild-type of RN1053.

The first limiting step of xylose metabolism is its transport across the plasma membrane (Kahar P, Taku K, Tanaka S, 2011. Enhancement of xylose uptake in 2-deoxyglucose tolerant mutant of *Saccharomyces cerevisiae*. J. Biosci. Bioeng. 111:557-63). It was reported that expression of *HXT* transporter genes of *S. cerevisiae* was regulated in response to different levels of extracellular glucose (Ozcan & Johnston, 1999. Function and regulation of yeast hexose transporters. Microbiol. Mol. Biol. Rev. 63:554-569). Possibly, mutations in the promoters or regulatory genes are the basis of enhanced expression of the cryptic *HXT* genes in the RN1053-X2 strain evolved in the chemostat culture on xylose.

EXAMPLE 7

KNOCKING OUT/DOWN *HXT11* IN EVOLVED RN1053 ABOLISHES NEWLY ACQUIRED ABILITY TO GROW ON XYLOSE

Knockout and silencing of HXT11 in the strain RN1053-X2. For the deletion construct of HXT11, P_{HIS3}-HIS3-T_{HIS3} expression cassette was amplified from the plasmid template pRS313-P7T7 (SEQ ID NO 117) using oligonucleotides KOP11 (SEQ ID NO 98) and KOT11 (SEQ ID NO 99) consisting of *HXT11* flanking sequences for integration at HXT11 locus and HIS3 sequence to amplify the expression cassette.

In order to determine whether Hxt11p was responsible for the spontaneous growth on xylose, we deleted *HXT11* in the evolved strain RN1053-X2 by succession of one-step gene deletion complementing for the *HIS3* marker upon deletion of *HXT11*. *HXT11* mRNA levels decreased considerably after knockout construct was introduced (Figure 13a). When *HXT11* was disrupted in the strain RN1053-X2, the strain lost its newly acquired ability to grow on xylose medium (Figure 13b). In addition, the *HXT11* expression level of knockout strains decreased to 60%. Most probably, still *HXT11* levels were measured since also expressed *HXT12* transcripts in the RN1053-X2 strain are about 98% homologous to *HXT11* (Figure 13a).

Antisense RNA technique is very useful for the repression of translation of a target protein by scavenging target mRNA in microorganisms. Translation repression can be promoted by antisense sequences that hybridize to messenger RNA. The antisense mechanism involves ribosome interference, in which the ribosome cannot bind to the nucleotides of the mRNA (Park *et al* 2001. Antisense-mediated inhibition of arginase [CAR1] gene expression in *Saccharomyces cerevisiae*. J. Biosci. Bioeng. 92:481-484). This method is more convenient for inhibiting gene expression than the gene disruption method. For expression of inverse *HXT11* (iHXT11), iHXT11 was amplified using primers iHXT11F (SEQ ID NO: 100) and iHXT11R (SEQ ID NO: 101) and cloned as *Bam*HI-*Xba*I fragment inversely between the truncated *HXT7* promoter and *HXT7* terminator in the pRS313-P7T7 vector (Figure 13c; SEQ ID NO: 117). The construct was introduced into RN1053-X2 to express antisense HXT11 RNA which hinders the translation of *HXT11* mRNA into protein. The iHXT11 was overexpressed in the strain RN1053-X2, and the strain lost its ability to grow on xylose medium (Figure 13d).

These knock-out/knock-down experiments clearly indicate that the evolved strain RN1053-X2 started to consume xylose due to the increase in *HXT11* expression and the probable functional expression of Hxt11p on the yeast membrane.

EXAMPLE 8**OVEREXPRESSING *HXT11* IN RN1053 RESTORES GROWTH ON XYLOSE**

Over-expression *HXT11* and *HXT12* in RN1053. To determine whether strains
5 expressing *HXT11* or *HXT12* are capable of growth on xylose, both *HXT*-genes were
expressed individually in the original RN1053 strain which is incapable to grow on xylose
because of the deletion of the eight major hexose transporters (*HXT1-HXT7* and *GAL2*).
For these experiments RN1041-empty, RN1053-empty, RN1053-*HXT11* and RN1053-
10 *HXT12* were used (see Table 13). RN1053-*HXT11* and RN1053 *HXT12* were
constructed in the following manner: the open-reading frames for the *HXT11* and *HXT12*
genes were PCR-amplified from cDNA of the wild-type RN1053 using primers HXT11F
(SEQ ID NO: 102), HXT12F (SEQ ID NO: 103) and HXT11/12R (SEQ ID NO: 104); PCR
fragments were sequenced and were found 100% homologous to the respective
CEN.PK113-7D gene sequences (*Saccharomyces* Genome Database,
15 www.yeastgenome.org; *HXT11* sequence SEQ ID NO; 119); the PCR fragments were
cut using restriction enzymes *Xba*I and *Bam*HI and cloned in yeast expression vector
pRS313-P7T7 (SEQ ID NO: 117; Figure 13C; Table 12). The *HXT11* and *HXT12*
expression construct of the transporters were transformed into RN1053 using a standard
yeast genetic technique according to the Gietz method (Gietz and Woods 2006, Yeast
20 transformation by the LiAc/SS Carrier DNA/PEG method. Methods Mol.Biol 313 pp. 107-
120). Transformants isolated from single colonies resulted in strains RN1053-*HXT11* and
RN1053-*HXT12*. Both strains were inoculated into maltose medium, followed by cultures
in shake flask on liquid media containing 2% xylose and/or 2% glucose. Only RN1053-
HXT11 and RN1041 displayed clear growth within 48 hours, whereas RN1053-empty
25 and RN1053-*HXT12* displayed hardly any growth in this period (Figure 14a). The growth
of RN1053-*HXT11* started earlier and displayed faster growth than that of reference
strain RN1041-empty (with wild-type *HXT* background) in xylose medium (Figure 14b).
However, the introduction of *HXT12* sequence in RN1053 did not allow growth on xylose
medium. In the *Saccharomyces* genome database reference strain S288C, *HXT12* is
30 considered a pseudogene because of a frame shift mutation (www.yeastgenome.org;
ORFs YIL170W and YIL171W).

This experiment showed clearly that the spontaneous growth on xylose of
RN1053-X2 was caused by higher expression of *HXT11* and that *HXT11* is an efficient
xylose transporter, if expressed.

EXAMPLE 9**Hxt11p FACILITATES XYLOSE TRANSPORT WITH INTRINSIC HIGHER LEVEL OF XYLOSE SPECIFICITY THAN Hxt2p**

5

Xylose uptake by Hxt11p. To determine whether *HXT11* is capable of xylose transport in the absence and presence of glucose, uptake studies using radiolabeled ¹⁴C-xylose were performed. For these experiments the strains RN1053-empty, RN1053-HXT11 and RN1053-HXT2 were used (see Table13). *HXT2* is known for its xylose transport capabilities (Saloheimo *et al*, 2007, Xylose transport studies with xylose-utilizing *Saccharomyces cerevisiae* strains expressing heterologous and homologous permeases. Appl Microbiol Biotechnol. 74:1041-1052; Sedlak and Ho, 2004, Characterization of the effectiveness of hexose transporters for transporting xylose during glucose and xylose co-fermentation by a recombinant *Saccharomyces yeast*. Yeast, 21:671-684). Construction of RN1053-HXT11 was described in the previous example. RN1053-HXT2 was constructed in the following manner: *HXT2* (SEQ ID NO: 120) was PCR-amplified from RN1001 genome (Table 12) using primers F HXT2 XbaI (SEQ ID NO:107) and R HXT2 Cfr9I (SEQ ID NO: 108). Resulting PCR fragment was sequence-verified and cloned into pRS313-P7-T7 (SEQ ID NO: 117) as *XbaI-Cfr9I* fragment. RN1053 was transformed with resulting construct pRS313-P7T7-HXT2 (; see Table 12) generating a transformant derived from a single colony named strain RN1053-HXT2 (Table 13). To determine if Hxt11p increased xylose uptake in the strain RN1053, xylose uptake was measured using ¹⁴C in the cells expressing the transporters. In the presence of glucose, strains expressing *HXT11* accumulated up to 50% more xylose compared to strains expressing *HXT2* in the presence of glucose (Figure 15). In addition, xylose uptake of RN1053-HXT11 was increased up to 4.5-fold, compared to RN1053-HXT2 (Figure 15). These experiments show that Hxt11p is capable of transporting xylose across the yeast plasma membrane. Even more, Hxt11p harnesses an intrinsic higher specificity towards xylose that could not be easily competed by glucose as compared to previously identified xylose transporters within the yeast genome such as Hxt2p.

EXAMPLE 10**Hxt11p IS FUNCTIONALLY EXPRESSED IN THE YEAST PLASMA MEMBRANE**

Functional expression of *HXT11* in RN1053. Detection of the Hxt11p by immunoblot analysis will not completely demonstrate the functional expression of the protein. For functional expression of a hexose transporter protein, it must reside in the plasma membrane. In order to monitor the expression and targeting of Hxt11p, a chimeric Hxt11p-GFP protein was engineered and examined by fluorometry and fluorescence microscopy. For these studies RN1053-HXT11-GFP, RN1041-HXT11-GFP, RN1053-HXT11 and RN1053-empty were used (Table 13). For the construction of the HXT11-GFP expressing strains, RN1053 and RN1041 were transformed with the expression vector bearing this chimeric Hxt11-GFP protein (pRS313-P7T7-HXT11-GFP; SEQ ID NO: 118; Table 12). The Hxt11p+GFP fusion protein is a functional hexose transporter: it restored growth on 2% xylose to the strain RN1053 (Figure 16a). In addition, Hxt11p+GFP fluorescence of RN1053-HXT11-GFP and RN1041-HXT11-GFP was localized to the plasma membrane on xylose or glucose medium (Figure 16b). These experiments show that Hxt11p-GFP is functional as xylose transporter and that Hxt11p is expressed functionally in the plasma membrane.

EXAMPLE 11**HXT11 EXPRESSION IN RN1053 FACILITATES FASTER CO-FERMENTATION OF GLUCOSE AND XYLOSE IN INDUSTRIALLY RELEVANT CONCENTRATIONS**

Fermentation profile of RN1053-HXT11 in glucose-xylose mixtures. To determine the fermentation behavior on a glucose-xylose mixture of a strain functionally expressing HXT11 in a *HXT* deletion background (*hxt1-7*, *gal2*) in comparison to a strain expressing the wild-type *HXT* landscape, fermentations were conducted on Verduyn-urea supplemented with industrially relevant glucose-xylose concentrations (80 and 40 g l⁻¹ respectively). For these fermentations RN1053-HXT11 and RN1041-empty were used (see Table 13).

As shown in Figure 17a, xylose of RN1041 was not co-fermented with glucose; whereas the degree of glucose-xylose co-fermentation was enhanced in strain RN1053-HXT11 as shown in Figure 17b. Furthermore, xylose fermentation was completely exhausted in HXT11-RN1053, whereas 10 g/L of xylose was still remaining in RN1041 at

40 h; in addition, the xylose consumption of RN1053-HXT11 was dramatically increased at 2% residual glucose, and ethanol concentration was increased from 63.92 g/L to 72.33 g/L at the end of the fermentation (Figure 17ab).

This example shows that a strain in which the major hexose transporters are deleted but expressing *HXT11* displays faster xylose utilization than a strain with a wild-type hexose transporter background without *HXT11*.

EXAMPLE 12

WILD-TYPE *HXT11* EXPRESSION SUPPORTS GROWTH ON XYLOSE IN GLUCOSE TRANSPORT COUNTER-ACTIVITY SCREEN

Aim. Using the Glucose Transport Activity Counter (GTAC)-screen, i.e. transforming hexokinase-mutant strain YD01227 as host to introduce transporters and screening the resulting transformants on medium for growth on and consumption of xylose in the presence of a 5 times higher amount of glucose, xylose transporters can be identified that exhibit an improved ability to transport xylose in the presence of a surplus of glucose (i.e. a higher affinity for xylose than for glucose; a reduction or more preferably full removal of glucose transport capability, while keeping xylose transport capability at least at the same level).

Transformation and colony picking. YD01227 was transformed with constructs bearing wild-type *GAL2* (pDB1250, EPA-29355;), *HXT2* (pDB1162; see Example 9 for construction plasmid) and *HXT11* (pRS313-P7T7-HXT11, see for construction Example 8) constructs. For each transformation, 3 colonies, when available, were re-plated using automated colony picking to YNB + 3% xylose agar medium MTPs and grown at 30°C until visible colony growth was visible in the agar puncture.

Pre-culture and screen. For pre-culture, transformants were transferred by automated process from selection agar medium MTPs to 96 well half-deepwell plates (96HDWP) containing liquid selection medium consisting of mineral medium (according to Verduyn, with urea as nitrogen source) supplemented with 3% xylose. The 96 well HDWPs were incubated for 3 days (pre-culture) in an orbital shaker at 30 °C and 750 rpm. Subsequently, for each sample, 20 µl of culture was transferred to 24 well deepwell plates (24 DWPs) containing 2.5 ml Verduyn-urea supplemented with the sugar mixture glucose:xylose in a ratio 5:1 with the following concentrations: 10 g l⁻¹ glucose and 2 g l⁻¹ xylose (YD10-medium). For each of the three sampling points a series of 24 well DWPs

was inoculated. On each 24DWP, one RN1001 growth control was inoculated to have an indication of plate variation effects between different 24DWPs. After 24 hours, 72 hours and 96 hours automatic sampling, transfer to 96HDWP was conducted for automated OD-measurement at 600 nm wavelength. After the last OD measurement after 96 hours, cells were pelleted after centrifugation and 100 µl supernatant was collected for flow-NMR analysis of residual sugars and ethanol formation in the medium, after incubation (see above for method description).

Results. Whereas RN1001 displayed growth from 24 hours onward (Figure 18a) and complete consumption of both glucose and xylose (Figure 18b, 18c), YD01227 transformants expressing *GAL2*, *HXT2* and *HXT11* constructs did not consume glucose (Figure 18c) due to the deletion of hexo- and glucokinase activities. The latter strains also displayed diverse growth profiles (Figure 18a) over time and diverse residual xylose concentrations representing their individual ability to consume xylose (Figure 18b) in the presence of glucose. YD01227 expressing the *HXT11* construct (YD01227-HXT11) displayed clear growth at 72 hours (Figure 18a) and substantial xylose consumption with $4.83 \pm 1.26 \text{ g l}^{-1}$ (n=3) xylose residing in the medium after 96 hours from the $22.7 \pm 0.12 \text{ g l}^{-1}$ (n=43) present in the medium at the beginning of the experiment; YD01227-GAL2 and YD01227-HXT2 displayed hardly any growth and not much xylose was consumed in comparison to YD01227-HXT11, with respectively 18.34 ± 0.39 and $21.29 \pm 0.32 \text{ g l}^{-1}$ xylose residing in the medium after 96 hours of incubation (Figure 18b).

This example shows that *HXT11* supports considerable biomass formation and xylose consumption in the presence of glucose indicating that *HXT11* expression in the hexokinase mutant YD01227 enables a more xylose-specific component in the sugar transport than exerted by *GAL2* or *HXT2* expression, both proven xylose transporters in the *S. cerevisiae* transportome (Saloheimo *et al*, 2007, Xylose transport studies with xylose-utilizing *Saccharomyces cerevisiae* strains expressing heterologous and homologous permeases. Appl Microbiol Biotechnol., 74:1041-1052; Sedlak and Ho, 2004, Characterization of the effectiveness of hexose transporters for transporting xylose during glucose and xylose co-fermentation by a recombinant *Saccharomyces yeast*. Yeast 21:671-684; Hamacher *et al*, 2002, Characterization of the xylose-transporting properties of yeast hexose transporters and their influence on xylose utilization. Microbiology, 148:2783-2788).

EXAMPLE 13**IMPROVED *HXT11* MUTANT OBTAINED FROM *HXT11* ERROR PRONE LIBRARY
SCREENED IN YD01227**

To improve xylose uptake in the presence of glucose, error-prone PCR was conducted to generate a *HXT11* mutant library encoding variants of Hxt11p which were screened for competitive xylose transport in the presence of glucose in the YD01227 strain (see Table 3 herein). The concentration of Mn²⁺ in the PCR reaction mixture was 0.15 mM, and it was added following the standard error-prone PCR protocol (Cirino, P.C *et al.* 2003. Generating mutant libraries using error-prone PCR. *Methods Mol. Biol.*, 231, pp. 3-9) using primers HXT11F (SEQ ID NO: 41), and HXT11/12R (SEQ ID NO: 43). The library was cloned into pRS313-P7T7 as *XbaI-BamHI* fragments. Strain YD01227 was transformed with the HXT11-library and three thousand transformants were screened on Verduyn-urea medium supplemented with a 1:15 ratio of xylose (1%) and glucose (15%) in 96-well plate format using Synergy MX (BioTek Instruments, Inc, USA). From these 3000 mutants, eight mutants were obtained out-performing wild-type Hxt11p on screening medium. A representative clone was shown in Figure 19a. The plasmids were isolated from YD01227 by using protocol according to Chowdhury (Chowdhury, K. 1991. One step 'miniprep' method for the isolation of plasmid DNA. *Nucl. Acids Res* 19, pp. 2792), and re-sequenced. All eight mutants were found to carry a plasmid bearing mutant sequences which translated into a protein containing the same mutation in Hxt11 (wild-type amino acid sequence Hxt11p SEQ ID NO: 62) leading to amino acid change at position 366 Asn (N) into Asp (D). In shake flask experiments the N366D mutant (YD01227-mHXT11[N366D]; see Table 13) also displayed faster growth in the screening medium compared to YD01227-HXT11, as shown in Figure 19a. pRS313-P7T7-mHXT11(N366D) construct was re-transformed to RN1053 resulting in RN1053-mHXT11(N366D). The RN1053 expressing *HXT11*-N366D mutant was not affected with respect to xylose utilization on 2% xylose medium, compared to RN1053 expressing the wild-type *HXT11* gene, since similar growth curves were observed (Figure 19b).

Sugar uptake of strain RN1053 with Hxt11 and N366D mutant. Xylose and glucose transport kinetics were determined in the xylose utilizing *S. cerevisiae* strain RN1053 expressing the transporters Hxt11 and mHxt11p-(N366D). The measured xylose transport rate was plotted against glucose or xylose concentration. The affinity for glucose of RN1053-mHXT11(N366D) was strongly decreased up to 2-fold compared to

RN1053-HXT11, whereas the affinity for xylose of RN1053-mHXT11(N366D) was also slightly decreased up to 1.2-fold, compared to RN1053-HXT11 (Figure 19c, 19d). In addition, in the presence of increasing concentrations of glucose strains expressing the N366D mutant accumulated up to 75% more xylose compared to strains expressing Hxt11 (Figure 19e).

Fermentation of xylose in the presence of glucose. Fermentation experiments of sugar mixtures were performed comparing RN1053-mHXT11(N366D) to RN1053-HXT11 and RN1041-empty.

In a first experiment on Verduyn-urea supplemented with 100 g/L of glucose and 60 g/L of xylose, as shown in Figure 19f, 19g, 19h, and 19i glucose utilization, biomass formation (OD600), and ethanol production of RN1053-mHXT11(N366D) were delayed during the fermentation, compared to RN1041 and RN1053-HXT11. It seems that glucose consumption by RN1053-mHXT11(N366D) was strongly decreased at early exponential phase compared to RN1053-HXT11 (Figure 19g). In addition, xylose consumption of RN1053-mHXT11(N366D) was not decreased (Figure 19h) although the cell density of RN1053-mHXT11(N366D) was lower than that of RN1041, and RN1053-HXT11 (Figure 19f).

In a second fermentation experiment on Verduyn-urea supplemented with lower sugar concentrations (80 g l⁻¹ of glucose and 40 g l⁻¹ of xylose) comparing RN1053-mHXT11(N366D) to RN1053-HXT11, a clear difference in sugar consumption profile was observed (Figure 19j, 19k, 19l). Whereas the glucose consumption profile displayed a faster glucose consumption rate during the phase that the glucose is declining rapidly (between 0 and 35 hours; $q_{\text{gluc(RN1053-HXT11)}}$ 2.22 g/l/h vs. $q_{\text{gluc(RN1053-mHXT11[N366D])}}$ 2.04 g/l/h, 8.5% decline in q_{gluc}), whereas the xylose consumption rate has greatly increased during that same time window ($q_{\text{xyI(RN1053-HXT11)}}$ of 0.23 g/l/h vs. $q_{\text{xyI(RN1053-mHXT11[N366D])}}$ of 0.38 g/l/h, 65% incline in q_{xyI}). During the phase that mainly xylose was fermented (35-73 hrs) the xylose consumption rates were almost identical ($q_{\text{xyI(RN1053-HXT11)}}$ of 0.36 g/l/h vs. $q_{\text{xyI(RN1053-mHXT11[N366D])}}$ of 0.37 g/l/h). The maximal xylose consumption rate was higher and reached earlier in the fermentation for RN1053-mHXT11(N366D) (0.65 g/l/h at 23 hrs) than for RN1053-HXT11 (0.50 g/l/h at 35 hrs). At the end of the fermentation run (period of 72 hrs typical for industrial fermentations) the RN1053-HXT11 consumed 51% of the xylose whereas RN1053-mHXT11(N366D) consumed 63%. The ethanol titers measured at the end of the fermentation were 4.4% higher for the RN1053 expressing the N366D mutant than for RN1053 expressing wild-type *HXT11*. Considering that the

input of sugars into the fermentation was almost identical, one could imagine that higher yields were obtained from this glucose-xylose mixture with sugar concentrations typical for industrially relevant batch fermentations, and more specifically higher yields from the xylose fraction.

5 These fermentation experiments indicate that compared to its wild-type reference sequence the presence of a xylose-specific transporter variant engineered from *S. cerevisiae* hexose transporter *HXT11* on the membrane increases the xylose consumption rate during the glucose phase, where the glucose consumption rate declined somewhat, and that in the end higher ethanol yields were obtained on a typical
10 glucose-xylose mixture typical for relevant industrially relevant hydrolysates.

EXAMPLE 14

EVOLVED HEXOKINASE MUTANT CONSUMES XYLOSE IN PRESENCE OF GLUCOSE

15 **Evolutionary engineering of strain YD01227.** For the evolutionary engineering, strain YD01227 was used for evolutionary engineering on glucose-xylose mixtures to evolve for xylose assimilation in the presence of glucose aiming at isolating spontaneous mutants in hexose transport or the regulation of the expression and/or activity of hexose
20 transport.

 YD01227 was inoculated for 16 hours in shake flask with Verduyn-urea-his supplemented with 2% xylose. At the start of the evolutionary engineering, YD01227 was diluted to an OD600 of 0.2 and DO setpoint at 5% in Verduyn-urea-his containing 1% xylose, 3% glucose. Carbondioxide outflow was monitored. At various time points
25 samples were taken for analysis of glucose and xylose concentrations. It was determined whether strain YD01227 was growing solely on xylose or on both glucose and xylose. The glucose to xylose ratio, at the start of the evolutionary engineering, was kept to low to a ratio (glucose 3%, xylose 1%), which still allowed YD01227 growth at the beginning of the experiment. However, this was at significantly lower growth rates if
30 compared to growth on only 1% xylose. In this setup the strain consumes the xylose which leads to higher glucose:xylose ratios and, therefore, a drop in growth rate. When the CO₂ production dropped, additional xylose (5 ml of 50% xylose to 500ml fermentor volume) was added to maintain growth. At an OD600 of 20, which was reached on average after 5-6 days, the culture was diluted into fresh Verduyn-urea-his in a higher

glucose to xylose ratio, if growth rates had significantly improved in the previous cycle. In total, in a time frame of 27 days, the strain was serial diluted 5 times in 1%Xyl/3%Glc (1:3), 1.5%Xyl/9%Glc (1:6), 1%Xyl/8%Glc (1:8), 1%Xyl/10%Glc (1:10) and 0,57%Xyl/10%Glc (1:15), respectively. Before inoculation in 1% xylose and 8% glucose the setpoint for DO was lowered into 0% (anaerobic growth) in order to maintain a lower growth rate. In figure 20 the scheme for the glucose/xylose ratio in the Verduyn-urea-his medium during chemostat cultivation (days) of YD01227 is given. After 27 days samples were taken of the evolved YD01227 strain and plated 1% xylose and 10% glucose with the original YD01227 as negative control. Whereas the original YD01227 showed only small colonies the colonies of the evolved YD01227 strain were 10-15 times larger.

Xylose growth and Xylose uptake with/without glucose competition. After re-streaking the evolved YD01227 strain on a 1% xylose and 10% glucose plate, three colonies (EvoA, EvoB and EvoC) were analyzed for growth in shake flasks on 1% xylose in the presence of respectively 0%, 3%, 6% and 10% glucose. YD01227 EvoB had the highest growth rate on xylose at 6% en 10% glucose and was compared with the original YD01227 strain (Figure 21a, 21b). In the original YD01227 strain (YD01227 ORI in figure 21b) the growth rate is already partly inhibited at 3% glucose and completely inhibited at 6% and 10% glucose however the YD01227 EvoB strain shows only minor inhibition in growth rate at all glucose concentrations and this seems unrelated to the amount of glucose added (Figure 21a). The same two strains were used in a ¹⁴C xylose (50mM) uptake experiment in which the xylose uptake is inhibited by glucose (Figure 21c). The uptake of xylose without the addition of glucose in both strains is the same however as soon as glucose was added to both strains the uptake of xylose in the YD01227 EvoB strain was not as inhibited as the uptake in the original YD01227 strain. In a 1:10 ratio the xylose uptake in the original YD01227 strain is completely abolished whereas in the YD01227 EvoB strain still 5 nmol/mgDW.min is taken up.

EXAMPLE 15**SINGLE NUCLEOTIDE POLYMORPHISM IN *HXT3-6* CHIMERA ALLOWS FOR
XYLOSE CONSUMPTION IN THE PRESENCE OF GLUCOSE IN EVOLVED
HEXOKINASE MUTANT**

5

Expression of the *HXTs* in the evolved YD01227 strain. The expression levels of *HXT1-17* and *GAL2* in the evolved YD01227 EvoB and original YD01227 strains were compared during batch cultivations on Verduyn-urea-his containing 1% xylose and 3% glucose (Figure 22a). Primers (SEQ ID NO: 2-37) were used in the real time PCR characterization of the expression of *HXTs* in YD01227 and evolved derivative. Furthermore, the expression levels in the YD01227 EvoB strain were also analyzed on Verduyn-urea-his containing 1% xylose and 10% glucose. The absolute C(t) values (data not shown) show that the *HXT1-7* genes are the only *HXT* genes that are intermediately or highly expressed of which the *HXT3-6* chimera (specific deletion in strain lineage intragenic and intergenic *HXT3* and *HXT6* sequences resulting in one *HXT36* chimeric sequence) has the highest expression levels. None of the analyzed sugar transporters is up-regulated in the YD01227 EvoB strain compared to the original YD01227 strain. The up-regulation seen in the *HXT1* gene in YD01227 EvoB 1% xylose and 10% glucose is caused by the high glucose concentration in this sample which is quite well known (Ozcan & Johnston, 1995, Three different regulatory mechanisms enable yeast hexose transporter (*HXT*) genes to be induced by different levels of glucose. Mol Cell Biol 15, pp. 1564-1572). This high glucose concentration also leads to the down-regulation of *HXT2* and *HXT7* in YD01227 EvoB. Both the up-regulation in *HXT1* and down-regulation of *HXT2* and *HXT7*, are described in literature (Boles & Hollenberger 1997, Kinetic characterization of individual hexose transporters of *Saccharomyces cerevisiae* and their relation to the triggering mechanisms of glucose repression , FEMS Microbiol Rev 21, pp. 85-111).

20

25

30

Sequencing of the highly expressed *HXT* genes. *HXT1-7* were amplified from cDNA which was isolated from the YD01227 EvoB culture on 1% xylose and 3% glucose using the primers SEQ ID NO: 49-60 . The PCR products from these genes were sequenced. No mutations were revealed in *HXT1*, *HXT2* and *HXT4*, one silent mutation in *HXT5* and *HXT7* and a mutation leading to amino acid change at position 367 (Asn into Ile; N367I) in chimera *HXT3-6*. Somewhere in the YD01227 strain lineage a deletion occurred between the neighboring loci of *HXT3* and *HXT6* in which intragenic and

intergenic sequences were deleted (part of 3' part of HXT3 ORF, *HXT3* terminator, HXT6 promoter, HXT6 ORF) resulting in one HXT36 chimeric sequence which is in frame and can be expressed as mRNA and translated into functional protein. The translated chimeric protein Hxt3-6p of which the first 438 amino acids are identical to the CEN.PK Hxt3p amino acid sequence, whereas the 130 amino acids towards the C-terminus are identical to Hxt6p in CEN.PK. Genomic rearrangements in the HXT3-6-7 locus have been documented in the past, e.g. HXT6/7 chimeric sequences resulting from chemostat cultures on low glucose concentrations, and are proposed to be caused by homologous recombination due to the highly homologous stretches of sequences in this cluster of hexose transporters (Brown et al 1998. Multiple duplications of yeast hexose transport genes in response to selection in a glucose-limited environment. *Mol Biol Evol* 15:931-942). The N367I point mutation is located in trans membrane domain (TMD) 8 which is known to contain residues responsible for the affinity for glucose (Kasahara & Kasahara, 2003. Transmembrane segments 1, 5, 7 and 8 are required for high-affinity glucose transport by *S. cerevisiae* Hxt2 transporter. *Biochem J.* 372:247-252, reconfirmed by this patent application).

EXAMPLES 16-18

METHODS

Methods not mentioned in the section of Examples 16 and 17 were already described in section Examples 6-15. Oligonucleotides used in the studies described in Example 16 and 17 are depicted in Table 14. The saturated mutagenesis of position N367 in HXT36 was done using PCR with Phusion® High-Fidelity PCR Master Mix with HF buffer using primer pairs F HXT36 *BcuI* (SEQ ID NO 127)/R HXT36 367NNN (SEQ ID NO 128) and F HXT36 367NNN (SEQ ID NO 129)/R HXT36 *BamHI* (SEQ ID NO 130). The fragments of 1119 and 623 base pairs were subsequently used in an overlap PCR using the outside primers F HXT36 *BcuI* and R HXT36 *BamHI* and cloned into pRS313P7T7 using *BcuI* and *BamHI*. Sequencing of 48 *E. coli* clones yielded N367S (tcc), N367P (ccc), N367G (ggg), N367Y (tac), N367A (gcc), N367H (cac), N367R (agg),

N367F (ttt), N367E (gag), and N367V (gtg). The remaining 8 amino acids at position 367 were amplified and cloned with overlap PCR using specific primers in which the NNN was replaced by tta (L), tgt (C), tgg (W), atg (M), act (T), aag (K), gat (D) and cag (Q).

The carboxyl-terminal GFP fusions with HXT36 and HXT36-N367I mutant were made by amplification of the corresponding genes with the Phusion® High-Fidelity PCR Master Mix (HF buffer) using primers F HXT36 *BcuI* (SEQ ID NO 127) and R HXT36 *Bam*HI-stop (SEQ ID NO 131). The GFP gene itself was amplified with F GFP *Bam*HI (SEQ ID NO 132) and R GFP *Cl*aI (SEQ ID NO 133). *HXT36* and *HXT36*-N367I were digested with the restriction enzymes *BcuI* and *Bam*HI and GFP was digested with *Bam*HI and *Cl*aI. The *HXT36* genes were separately ligated in a two-fragment ligation together with GFP into pRS313-P7T7 which was cut with *BcuI* and *Cl*aI.

For the fermentations of Example 18, strains were grown, *in duplo*, in 50 ml Schott bottles filled to the rim with 68 ml fermentation medium containing 0.5% D-glucose and 0.5% D-xylose and kept completely closed at 30 °C in a water bath. Stirring speed, with a magnet stirrer, was 200 rpm and the strains were inoculated at a starting OD600 of approximately 8.0. At regular intervals samples were taken for OD600 measurements and HPLC analysis.

Table 14. Oligonucleotides used in cloning and sequencing.

SEQ ID NO	Name	Sequence (5' → 3')
127	F HXT36 <i>BcuI</i>	GCATACTAGTATGAATTCAACTCCAGATTTAATATCC
128	R HXT36 367NNN	CAACAAGTAGAGAAGAAAnnnGACGACACCG
129	F HXT36 367NNN	CGGTGTCGTCnnnTTCTTCTCTACTTGTTG
130	R HXT36 <i>Bam</i> HI	ACGTGGATCCTTATTTGGTGCTGAACATTCTCTTGT
131	R HXT36 <i>Bam</i> HI-stop	CCATGGATCCTTTGGTGCTGAACATTCTCTTGTAC
132	F GFP <i>Bam</i> HI	AAAGGATCCATGGTGAGCAAGGGCGAGGAGC
133	R GFP <i>Cl</i> aI	AAAATCGATTTACTTGTACAGCTCGTCC

EXAMPLE 16**DECREASE IN V_{MAX} OF HXT3-6 N367I IS NOT THE RESULT OF DECREASED
EXPRESSION OF MUTANT AS SHOWN BY GFP-TAGGING STUDIES**

5 To ensure that this lower V_{max} is not due to a decrease in expression, chimeras
were made in which GFP was fused to the C-terminus of the HXT36 and the HXT36-
N367I mutant. Both fusions were transformed to the RN1053 strain, and fluorescence
imaging revealed that the proteins are uniformly distributed over the plasma membrane
(Fig. 23A and 23B). Since the same levels of GFP were recorded, Hxt36p and Hxt36p-
10 N367I are expressed to similar extents. (Fig 23C).

EXAMPLE 17**SATURATION MUTAGENESIS ON POSITION ASPARAGIN-367 IN *HXT3-6*
CHIMERA FOR THE EXPLORATION OF THE SEQUENCE SPACE REVEALS
15 *ALANINE AS POTENT RESIDUE IN ENHANCED TRANSPORTER WITH ENHANCED
XYLOSE TRANSPORT CAPACITY***

To explore the sequence space of position N367 (corresponding to position N376
in SEQ ID NO: 59), all amino acid substitutions were individually introduced into the
20 *HXT36* gene. The individual HXT36-N367X mutants were transformed to the YD01227
hexokinase deletion strain and tested for growth on minimal medium containing 1 % D-
xylose and 10 % D-glucose. The transformant bearing the original HXT36-N367I mutant
showed an OD_{600} of 0.56 after 24 hrs whereas the HXT36 wild-type was unable to grow
under these conditions (Fig. 24). The fastest growing transformant beared the Hxt36p
25 N367A mutant, which reached an OD_{600} of almost 2. Also the transformants expressing
the HXT36 mutants the other nonpolar aliphatic amino acid substitutions (glycine, valine,
leucine and methionine) where able to grow on D-xylose in the presence of 10% D-

glucose. On the other hand, the phenylalanine and histidine mutants showed a reduced growth rate whereas strong polar and charged amino acid substitutions did not support growth (Fig. 24). These data show that the N367 (corresponding to N376 in Gal2p and N366 in Hxt11p) is a critical residue in determining the specificity of Hxt36p for glucose
5 versus xylose.

The Hxt36p N367A and N367I mutants were analyzed further to determine their transporter kinetics. Herein, the transporters were expressed in strain RN1053 that is equipped with a low background glucose transport activity. The K_m and V_{max} for D-glucose uptake by HXT36 was about 6 mM and 32 nmol/mgDW.min, respectively (Table
10 15 and Fig 25). Remarkably, the Hxt36p N367I mutant was completely defective in D-glucose uptake, while the affinity for D-xylose uptake was improved 2.7-fold (i.e., from 108 to 40 mM) compared to Hxt36p (Table 15). The mutation, however, also caused a near to 3-fold decrease in the V_{max} for D-xylose uptake.

Also the transport activity of the Hxt36p-N367A mutant was examined which
15 showed the fastest growth on D-xylose. This transporter still showed some glucose uptake although with a very poor K_m (171 mM versus 6 mM). Compared to the N367I mutant, the N367A mutation caused both an improvement of the K_m and V_{max} values for D-xylose uptake to 25 mM and 15.3 nmol/mgDW.min, respectively.

20

EXAMPLE 18

CO-FERMENTATION OF D-GLUCOSE AND D-XYLOSE BY AN ENGINEERED S. CEREVISIAE STRAIN WITH ALTERED TRANSPORT CHARACTERISTICS

In order to investigate co-fermentation of D-glucose and D-xylose, the RN1053 strain harboring the wild-type Hxt36, Hxt36-N367I and Hxt36p-N367A, were grown on 5
25 g L⁻¹ D-glucose/ 5 g L⁻¹ D-xylose at a higher industrially relevant starting OD₆₀₀ of approximately 8.0. Sugar consumption and ethanol concentrations were followed

through time (Fig. 26). The strain containing the Hxt36-N367I transporter grows on D-xylose but because of the severe D-glucose uptake defect (Fig. 26B), it only shows some background level of D-glucose consumption that is similar to that of the original RN1053 strain without any re-introduced transporter (data not shown). The strain harboring the Hxt36-N367A mutant showed an improved D-glucose and D-xylose co-consumption (Fig. 26C) as compared to the strain containing the Hxt36 wild-type transporter (Fig. 26A). Moreover, also the total sugar consumption increased because of the co-consumption of glucose and xylose by the Hxt36p-N367A-expressing RN1053 yielding a higher ethanol concentration (2.83 g L^{-1}) than wild-type Hxt36p-expressing RN1053 (2.67 g L^{-1}) after 9 hours of fermentation.

Table 15. K_m and V_{max} values for D-glucose and D-xylose uptake by Hxt36p transporters expressed in strain RN1053 .

	K_m (mM)		V_{max} (nmol/mgDW.min)	
	Glucose	Xylose	Glucose	Xylose
HXT36	6.13 ± 0.02	107.9 ± 12.1	31.7 ± 0.07	32.9 ± 3.1
HXT36-N367I	- ^a	39.8 ± 5.6	- ^a	12.1 ± 1.63
HXT36-N367A	170.7 ± 37.8	24.9 ± 3.4	37.2 ± 4.4	15.3 ± 0.2

^a Could not be determined

EXAMPLES 19 and 20

Material and Methods

Methods not mentioned in the section of Examples 19 and 20 were already described in sections of Examples 6-18. Oligonucleotides used in the studies described in Example 19 and 20 are depicted in Table 16.

Mutagenesis of N366X. The saturated mutagenesis of position N366 in HXT11 was done using PCR with Phusion® High-Fidelity PCR Master Mix with HF buffer using primer pairs F HXT11 XbaI/R HXT11 366NNN and F HXT11 366NNN/R HXT11 BamHI (see Table 16). The fragments of 1113 and 591 base pairs were subsequently used in an overlap PCR using the outside primers F HXT11 XbaI and R HXT11 BamHI and cloned into pRS313P7T7 using *XbaI* and *BamHI*. Sequencing of 48 *E. coli* clones yielded N367S (tct), N367P (cca), N367G (ggt), N367A (gcc), N367H (cac), N367R (cgc), N367L (ttg), N367C (tgt), N367T (acg), N367D (gat), N367Q (caa), and N367V (gtg). The remaining 6 amino acids at position 366 were amplified and cloned as mentioned above with overlap PCR using specific primers in which the NNN was replaced by ttt (F), gag (E), tgg (W), atg (M), aaa (K), and N367Y (tat).

Using the generated *HXT11* variant sequences generated by the saturation mutagenesis PCR, *HXT11*-N366x-GFP-fusion constructs to study cellular localization of *HXT11* variants were prepared similarly (with oligonucleotides, restriction sites and pRS313-P7T7-GFP backbone vector) as described in the methods for Example 10 (p. 96).

Fermentation experiments. Yeast cultures were pre-cultured in mineral medium containing 2% maltose. Cells at mid-exponential phase were harvested and inoculated after washing twice with sterilized water. Fermentation experiments were performed using 100 ml of mineral medium containing 7% glucose and 4% xylose in 120 ml bottle at 30°C with an initial OD600 of 5 under oxygen limited conditions. Stirring speed, with a magnetic stirrer, was 200 rpm. All of the bottle fermentation experiments were repeated independently.

Table 16. Primers used for saturation mutagenesis of *S. cerevisiae* HXT11.

Name	Sequence (5' → 3')
F HXT11 XbaI	GGCCTCTAGAATGTCAGGTGTTAATAATACATCCGC
R HXT11 BamHI	CGATGGATCCTCAGCTGGAAAAGAACCTCTTGTAATTG
F HXT11 366NNN	CGGTGTGGTTnnnTTTTTCTCTTCATTC
R HXT11 366NNN	GAATGAAGAGAAAAA nnnAACCACACCG
F HXT11 N366F	CGGTGTGGTTtttTTTTTCTCTTCATTC
R HXT11 N366F	GAATGAAGAGAAAAAaaaAACCACACCG
F HXT11 N366E	CGGTGTGGTTgagTTTTTCTCTTCATTC
R HXT11 N366E	GAATGAAGAGAAAAAActcAACCACACCG
F HXT11 N366K	CGGTGTGGTTaaaTTTTTCTCTTCATTC
R HXT11 N366K	GAATGAAGAGAAAAAAttAACCACACCG
F HXT11 N366M	CGGTGTGGTTatgTTTTTCTCTTCATTC
R HXT11 N366M	GAATGAAGAGAAAAAcatAACCACACCG
F HXT11 N366W	CGGTGTGGTTtggTTTTTCTCTTCATTC
R HXT11 N366W	GAATGAAGAGAAAAAaccaAACCACACCG
F HXT11 N366Y	CGGTGTGGTTtatTTTTTCTCTTCATTC
R HXT11 N366Y	GAATGAAGAGAAAAAataAACCACACCG

n is any nucleotide

5

EXAMPLE 19**N366 MUTATIONS IN HXT11 IMPROVE XYLOSE UTILIZATION IN THE PRESENCE OF GLUCOSE.**

The N366D mutation in Hxt11 improved xylose uptake in the presence of glucose because of a reduction in the glucose transport affinity. In order to assess the importance of this position, N366 was replaced with each of the other 19 amino acids to generate a series of N366X mutants. The corresponding genes were expressed in strain YD01227 and evaluated for their ability to utilize 1 % xylose in the presence of 10 % glucose using 96-well plates. Here, 10 % glucose was used instead of 15 % to increase the sensitivity of the assay in order to discriminate between the performance of the various mutants. Growth rates on xylose in the presence of a 10-fold excess glucose

was improved when N366 was substituted by a methionine (M) or threonine (T) residue (Figure 27A), and was up to 3-fold higher than the N366D substitution. Amino acids with a positive charged or bulky hydrophobic side chain did not support growth on xylose under the screening condition. The mutants were also tested for growth on glucose and expressed in strain RN1053. The N366M and N366T Hxt11 mutants showed similar growth on glucose or as compared to the wild-type Hxt11 (Figure 27B and 27C). In addition, the expression level of all individual mutants was determined using GFP-tagged Hxt11 proteins (Figure 28). All of the Hxt11-GFP proteins were highly expressed on membrane, except that with the GFP-tagged N366W and N366Y Hxt11 proteins, a cytosolic and likely vacuolar localization seems apparent. This suggests protein-misfolding and faulty targeting to the membrane by these mutants explaining the low activity in the growth experiments with xylose and glucose. However, the wild-type and N366M and N366T Hxt11 mutant protein localized to the plasma membrane. Overall, these data indicate that N366 is a critical residue in determining the specificity of Hxt11 for xylose versus glucose.

The xylose and glucose transport kinetics via Hxt11 and the N366T and N366M Hxt11 mutants were determined for the genes expressed in the xylose utilizing *S. cerevisiae* strain RN1053. Compared to the wild-type Hxt11, the affinity for glucose transport by N366T and N366M Hxt11 was reduced up to 5 and 4-fold, respectively. In contrast, the affinity for xylose by these mutants was improved by up to 2-fold (Table 17) relative to Hxt11. The $V_{\max}$ for glucose uptake by the N366T Hxt11 mutant was increased by about 40% as compared to wild-type Hxt11, while the $V_{\max}$ was unchanged for the N366M Hxt11 protein. Importantly, the $V_{\max}$ for xylose of the mutants remained largely unchanged compared to Hxt11.

Table 17. K_m and V_{max} values for D-glucose and D-xylose uptake by Hxt11 transporters expressed in strain RN1053.

	K_m (mM)		V_{max} (nmol/mg DW.min)	
	Glucose	Xylose	Glucose	Xylose
Hxt11	33.4 ± 2.1	84.2 ± 10.0	82.3 ± 3.8	44.5 ± 1.7
Hxt11-N366D	87.0 ± 6.4	106.7 ± 21.7	98.9 ± 5.7	45.5 ± 1.0
Hxt11-N366T	194.4 ± 47.9	46.7 ± 2.7	125.6 ± 3.7	40.1 ± 2.4
Hxt11-N366M	144.9 ± 36.0	50.1 ± 9.7	75.3 ± 8.6	34.2 ± 3.4
Hxt2	n.d.	51.2 ± 0.1	n.d.	12.5 ± 0.2

n.d., not determined

5

EXAMPLE 20

CO-FERMENTATION OF D-GLUCOSE AND D-XYLOSE BY ENGINEERED S.

CEREVISIAE STRAINS EXPRESSING HXT11 VARIANTS

Because of the marked effects of the N366 mutations in Hxt11 on glucose
 10 transport without interfering with xylose transport, the mutants were further examined for
 their ability to co-metabolize xylose and glucose under industrial conditions, i.e., 7%
 glucose and 4% xylose, respectively. Herein, the mutants were expressed in strain
 RN1053, and growth was compared to the Hxt11 wild-type and strain RN1001 containing
 a full complement of endogenous transporters. The two mutants supported a near to
 15 perfect co-consumption of glucose and xylose (Figure 29A and 29B) in contrast to the
 RN1001 strain (Figure 29C) and RN1053 wild-type Hxt11 (Figure 29D) and that showed
 delayed consumption of xylose. These data demonstrate that the mutagenesis of the
 N366 position of Hxt11 yields mutants that mediate a balanced uptake of glucose and
 xylose thereby supporting co-consumption of the hexose and pentose sugars. Best
 20 results were obtained for mutant N366T HXT11.

Co-consumption

Co-consumption of a cell is herein quantified and expressed as co-consumption index.

The co-consumption index was herein the co-consumption index for glucose and xylose

5 and was calculated as the sum over the time interval of 0-24 hours (it was measured at 0, 8, 12, 14, 16, 18, 20, 22 and 24 hours) of the absolute difference of the glucose uptake rate (Q_g) and the xylose uptake rate (Q_x), expressed as grams of sugar consumed per time unit. The fermentation was an anaerobic batch culture fermentation at 1.0 g/l dry yeast pitch, 30 degrees C temperature and wherein the fermentation
10 medium contains 71.8 grams of glucose per liter and 40.0 grams xylose per liter, at the start of the fermentation. A low value for co-consumption index indicates high co-consumption, a high value less co-consumption.

15 These fermentation data and calculations for the the strains RN1001, RN1053 HXT11, RN1053 HXT11 (N366M) and RN1053 HXT11 (N366T) are given in table 18.

Table 18: Fermentation data and calculation of co-consumption index for the the strains RN1001, RN1053 HXT11, RN1053 HXT11 (N366M) and RN1053 HXT11 (N366T)

DS68616										
Time	Average									
	Glucose	Xylose		Qg (g/h)	Qx (g/h)		abs(Qg- Qx)	sum(abs(Qg- Qx))		corr(glucose, xylose)
0	71.8	39.9								
8	46.9	39.8		3.12	0.02		3.10	29.1		0.69
12	31.0	38.7		3.97	0.27		3.70			
14	18.4	37.8		6.31	0.46		5.85			
16	5.42	35.9		6.47	0.94		5.53			
18	0	29.1		2.71	3.39		0.68			
20	0	20.7		0	4.25		4.25			
22	0	15.4		0	2.632		2.63			
24	0	8.60		0	3.40		3.40			
HXT11										
	Average									
	Glucose	Xylose		Qg (g/h)	Qx (g/h)		abs(Qg- Qx)	sum(abs(Qg- Qx))		corr(glucose, xylose)
0	71.8	39.9								
8	55.4	40.0		2.06	0.00		2.06	25.0		0.84
12	41.1	38.6		3.55	0.345		3.21			
14	25.5	35.2		7.84	1.70		6.15			
16	12.5	32.3		6.49	1.47		5.02			
18	2.85	26.0		4.82	3.12		1.70			
20	0	19.4		1.42	3.34		1.91			
22	0	14.3		0	2.51		2.51			
24	0	9.44		0	2.45		2.45			
HXT11 (N366M)										
	Average									
	Glucose	Xylose		Qg (g/h)	Qx (g/h)		abs(Qg- Qx)	sum(abs(Qg- Qx))		corr(glucose, xylose)
0	71.8	39.9								
8	60.2	38.2		1.45	0.220		1.23	11.2		0.977
12	52.5	35.6		1.93	0.644		1.29			

14	42.7	31.3		4.87	2.15		2.72				
16	35.6	28.8		3.55	1.23		2.32				
18	29.4	26.0		3.09	1.44		1.66				
20	23.3	22.3		3.05	1.82		1.24				
22	18.5	18.7		2.42	1.83		0.592				
24	14.1	14.5		2.18	2.07		0.117				
HXT11(N366T)											
	Average										
	Glucose	Xylose		Qg (g/h)	Qx (g/h)		abs(Qg- Qx)		sum(abs(Qg- Qx))		corr(glucose, xylose)
0	71.8	39.9									
8	59.1	36.7		1.59	0.409		1.18		11.8		0.977
12	47.7	33.5		2.84	0.798		2.05				
14	37.5	28.6		5.13	2.41		2.71				
16	28.6	24.7		4.42	1.99		2.43				
18	21.3	20.7		3.68	1.97		1.72				
20	14.2	15.9		3.55	2.44		1.11				
22	8.87	11.1		2.65	2.38		0.268				
24	4.94	6.56		1.96	2.26		0.3				

The co-consumption index of the strains RN1001, RN1053 HXT11, RN1053 HXT11 (N366M) and RN1053 HXT11 (N366T) was determined as in table 18. Summarizing, the results for co-consumption index are given in table 19.

5

Table 19: Co-consumption index of strains

Strain	Co-consumption index (sum abs (Qg*-Qx*) (g/h))
RN1001	29.1
RN1053 HXT11	25.0
RN1053 HXT11 (N366M)	11.2
RN1053 HXT11 (N366T)	11.8

REFERENCES

Ausubel *et al.* 1995 Current Protocols in Molecular Biology, John Wiley & Sons, Inc.

5 Becker J, Boles E. 2003 A modified *Saccharomyces cerevisiae* strain that consumes L-Arabinose and produces ethanol. *Appl Environ Microbiol.* 69: 4144-4150.

 Hamacher T, Becker J, Gárdonyi M, Hahn-Hägerdal B, Boles E. 2002 Characterization of the xylose-transporting properties of yeast hexose transporters and their influence on xylose utilization. *Microbiology* 148: 2783-2788.

10 Kasahara T, Kasahara M. 2000 Three aromatic amino acid residues critical for galactose transport in yeast Gal2 transporter. *J Biol Chem.* 275: 4422-4428.

 Kasahara T, Maeda M, Boles E, Kasahara M. 2009 Identification of a key residue determining substrate affinity in the human glucose transporter GLUT1. *Biochim Biophys Acta.* 1788: 1051-1055.

15 Kasahara T, Kasahara M. 2010 Identification of a key residue determining substrate affinity in the yeast glucose transporter Hxt7: a two-dimensional comprehensive study. *J Biol Chem.* 285: 26263-26268.

 Kuyper M, Hartog MM, Toirkens MJ, Almering MJ, Winkler AA, van Dijken JP, Pronk JT 2005 Metabolic engineering of a xylose-isomerase-expressing *Saccharomyces cerevisiae* strain for rapid anaerobic xylose fermentation. *FEMS Yeast Research* 5: 399-409.

 Luttik MA, Kötter P, Salomons FA, van der Klei IJ, van Dijken JP, Pronk JT 2000 The *Saccharomyces cerevisiae* *ICL2* gene encodes a mitochondrial 2-methylisocitrate lyase involved in propionyl-coenzyme A metabolism. *Journal of Bacteriology* 182: 7007-7013.

25 Nelissen B, De Wachter R, Goffeau A. 1997 Classification of all putative permeases and other membrane plurispanners of the major facilitator superfamily encoded by the complete genome of *Saccharomyces cerevisiae*. *FEMS Microbiol Rev.* 21: 113-134.

30 Sambrook *et al.* 1989 Molecular Cloning, a Laboratory Manual

 Schiestl RH and Gietz RD 1989 High efficiency transformation of intact yeast cells using single stranded nucleic acids as a carrier. *Current Genetics* 16: 339-346.

Wieczorke R, Krampe S, Weierstall T, Freidel K, Hollenberg CP, Boles E 1999
Concurrent knock-out of at least 20 transporter genes is required to block uptake of
hexoses in *Saccharomyces cerevisiae*. FEBS Letters 464: 123-128.

5 Young E, Poucher A, Comer A, Bailey A, Alper H. 2011 Functional survey for
heterologous sugar transport proteins, using *Saccharomyces cerevisiae* as a host. Appl
Environ Microbiol. 77: 3311-3319

CLAIMS

1. Polypeptide having one or more substitution corresponding to T89C/Y,
5 V187M/Q/C/G/Y, A191I/K/D/C/E/N/W/R/P, Y214L/P, Q215D/E/S/M/L/C/F/A,
I218W, I222G/Y/M/P/S/W, Y226M/T/P/Q, Q338R/I/D/E/F/G/P, M339Y/E/N,
Q341V/Y/S/H, Q342N/H/R/C/F/G/I/K/L/P/S/T/W, L343Y/Q/I/C/E/K/N/P/S, N346G/I,
N347S/L/K/G/T/V, Y350E/S/Q, G373R/K/L/M/Y, N376K/Q/A/C/D/E/I/M/P,
T380Y/G/M/A/S/D/E/I/K/L/W, S383Y/W/V/C/E/I/L/Q/R, F446H/D/I,
10 T448/R/K/H/D/E/W, T449F/H/A/G/I/L/P/Q/V, T451W/Y/S, W455D/K/E/P/Q,
N478Q/I/F/G, or W479S/I/F/D/E/H/M/N/R/T/Y of SEQ ID NO:59 wherein the
polypeptide is member of the Major Facilitator Superfamily (MFS).
2. Polypeptide according to claim 1, wherein the polypeptide has pentose transport
15 activity.
3. Polypeptide according to claim 2, wherein the polypeptide has xylose transport
activity.
- 20 4. Polypeptide according to any of claims 1 to 3, wherein the polypeptide has one or
more substitution corresponding to V187M/Q/C/G/Y, Q215D/E/S/M/L/C/F/A,
I218W, I222G/Y/M/P/S/W, Y226M/T/P/Q, M339Y/E/N,
Q342N/H/R/F/G/I/K/L/P/S/T/W, N346G/I, Y350E/S/Q, G373R/K/L/M/Y,
N376K/Q/A/C/D/E/I/M/P, S383Y/W/V/C/E/I/L/Q/R, F446H/D/I, T448/R/K/H/D/E/W,
25 T449F/H/A/G/I/L/P/Q/V, or W479S/I/F/D/E/H/M/N/R/T/Y of SEQ ID NO: 59.
5. Polypeptide according to any of claims 1 to 4, wherein the polypeptide has one or
more substitution corresponding to T89C, V187M/Q, A191I/K/D, Y214L/P,
Q215D/E/S, I218W, I222G/Y/M, Y226M/T, Q338R/I, M339Y, Q341V/Y/S,
30 Q342N/H/R, L343Y/Q/I, N346G, N347S/L/K, Y350E, G373R, N376K/Q/
T380Y/G/M, S383Y/W/V, F446H/D, T448/R/K/H, T449F/H, N478Q, or W479S/I/F
of SEQ ID NO: 59.
6. Polypeptide according to any of claims 1 to 5, wherein the polypeptide has one or

more of the substitutions corresponding to V187M/Q, Q215D/E/S, I218W, I222G/Y/M, Y226M/T, M339Y, Q342N/H/R, N346G, Y350E, G373R, N376K/Q/, S383Y/W/V, F446H/D, T448/R/K/H, T449F/H, or W479S/I/F of SEQ ID NO: 59.

- 5 7. Polypeptide according to any of claims 1 to 6, wherein the polypeptide has one or more of the substitutions corresponding to V187M, M339Y, I218W, Q342N, N376K G373R or W479S of SEQ ID NO: 59.
8. Polypeptide according to any of claims 1 to 7, wherein the polypeptide has reduced glucose transport activity compared to the polypeptide having SEQ ID NO: 59.
- 10 9. Polypeptide according to claim 8, wherein the polypeptide has increased xylose transport activity compared to the polypeptide having SEQ ID NO: 59.
10. Polypeptide according to any of claims 1 to 9, comprising sequences that contain one or more of the following amino acid motifs:
 - (a) G-R-x(3)-G-x(3)-G-x(11)-E-x(5)-[LIVM]-R-G-x(12)-[GA] ;
 - (b) R-x(14)-G-x(2)-Y-x(2)-[YF]-[YF]-[GSAL] and/or
 - (c) V-x(15)-[GNR]-[RH]-R-x(2)-[LM]-x(2)-[GA]
 wherein the fragments in (a) to (c) in the polypeptide correspond to the positions in

20 SEQ ID NO: 59.
11. Polypeptide according to any of claims 1 to 10 comprising a motif G-R-x(3)-G-x(3)-G-x(11)-E-x(5)-[LIVM]-R-G-x(12)-[GA].
- 25 12. Polynucleotide having at least 50% identity to SEQ ID NO: 56, encoding the polypeptide according to any of claims 1 to 11.
13. Nucleic acid construct comprising the polynucleotide of claim 12.
- 30 14. Host cell transformed with the nucleic acid construct of claim 13.
15. Transformed host cell according to claim 14, which is yeast.
16. Transformed host cell according to claim 15, which belongs to the genus

Saccharomyces, in particular the species *Saccharomyces cerevisiae*.

17. Transformed host cell comprising a heterologous nucleotide that encodes a polypeptide according to claim 1 or encodes a polypeptide having substitution
5 F85S/V, T89V/L, V187I, Y226A, Q338N, M339S/A, N346A/D, Y350I, G373N/S, N376S, or T449S of sequence ID NO: 59.
18. Transformed host according to claim 17, where in the polynucleotide encodes a polypeptide that is a mutant of a polypeptide that is native in the untransformed
10 host cell.
19. Transformed host cell according to claim 18, wherein the polypeptide that is native in the untransformed host cell is a member of the the Major Facilitator Superfamily (MFS) transporters.
15
20. Transformed host cell according to claim 18 or 19, wherein the polypeptide that is native in the untransformed host cell is a hexose transporter polypeptide.
21. Transformed host cell according to any of claims 18 to 20, wherein the polypeptide
20 that is native in the untransformed host cell is a hexose transporter polypeptide.
22. Transformed host cell according to claim 21, wherein the polypeptide that is native in the untransformed host cell is a transporter polypeptide chosen from the list consisting of Gal2, Hxt1, Hxt2, Hxt3, Hxt4, Hxt5, Hxt6, Hxt7, Hxt8, Hxt9, Hxt10,
25 Hxt11, Hxt12, Hxt13, Hxt14, Hxt15, Hxt16 and Hxt17.
23. Process for the degradation of ligno-cellulosic or hemi-cellulosic material, wherein ligno-cellulosic or hemi-cellulosic material is contacted with an enzyme composition, wherein one or more sugar is produced, and wherein the produced
30 sugar is fermented to give a fermentation product, wherein the fermentation is conducted with a transformed host cell of any of claims 14 to 22.
24. Process according to claim 23, wherein the fermentation product is one or more of ethanol, butanol, lactic acid, di-terpene, glycosylated di-terpene, a plastic, an

organic acid, a solvent, an animal feed supplement, a pharmaceutical, a vitamin, an amino acid, an enzyme or a chemical feedstock.

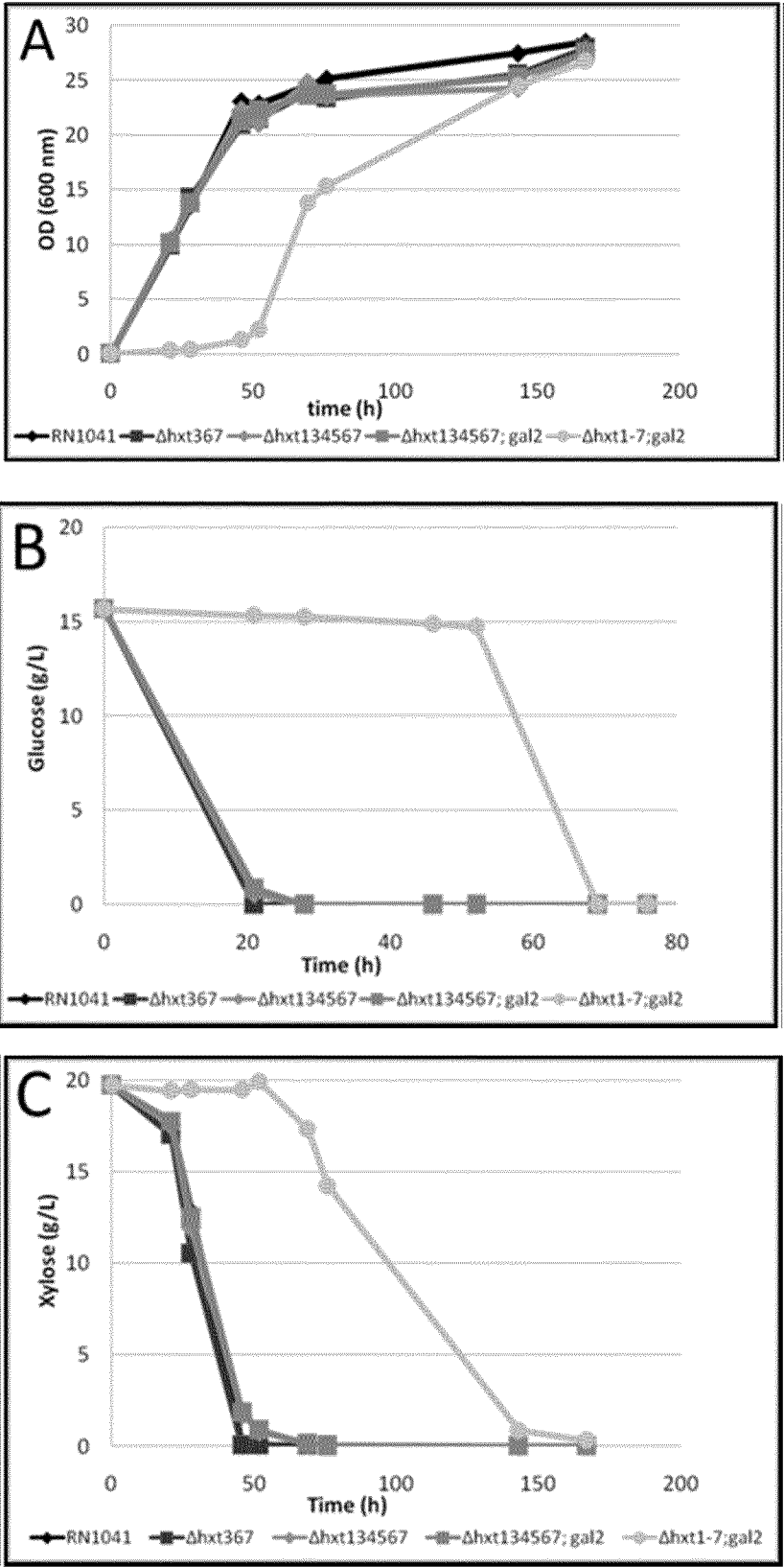


FIG. 1

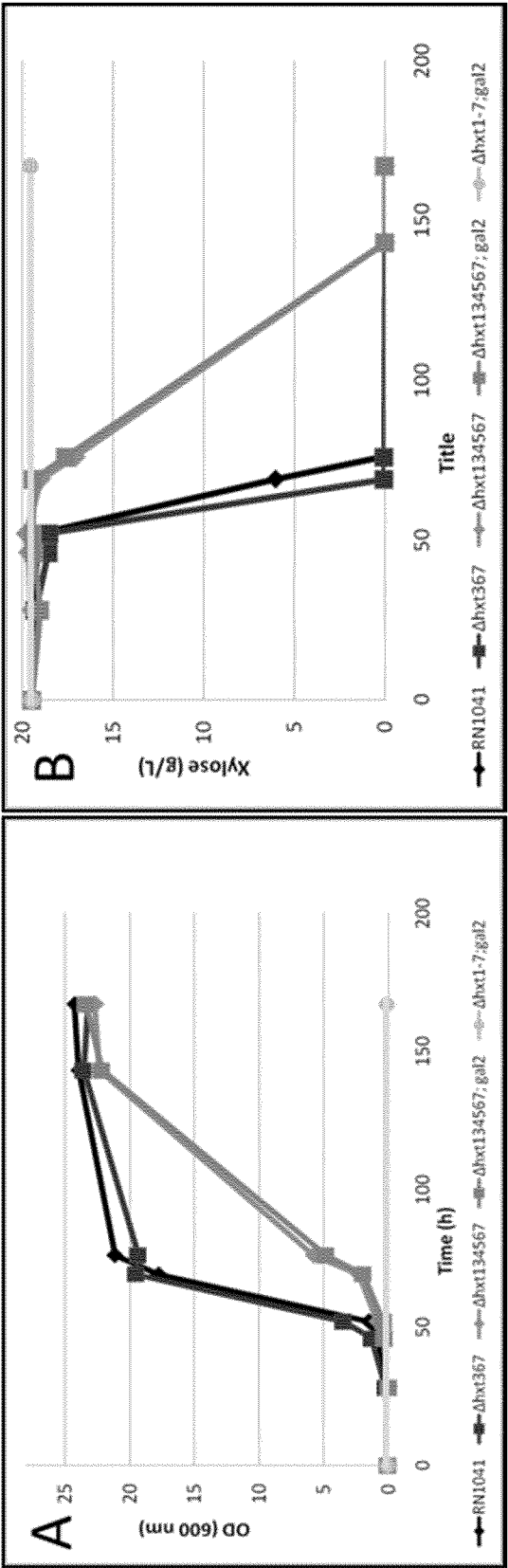


FIG. 2

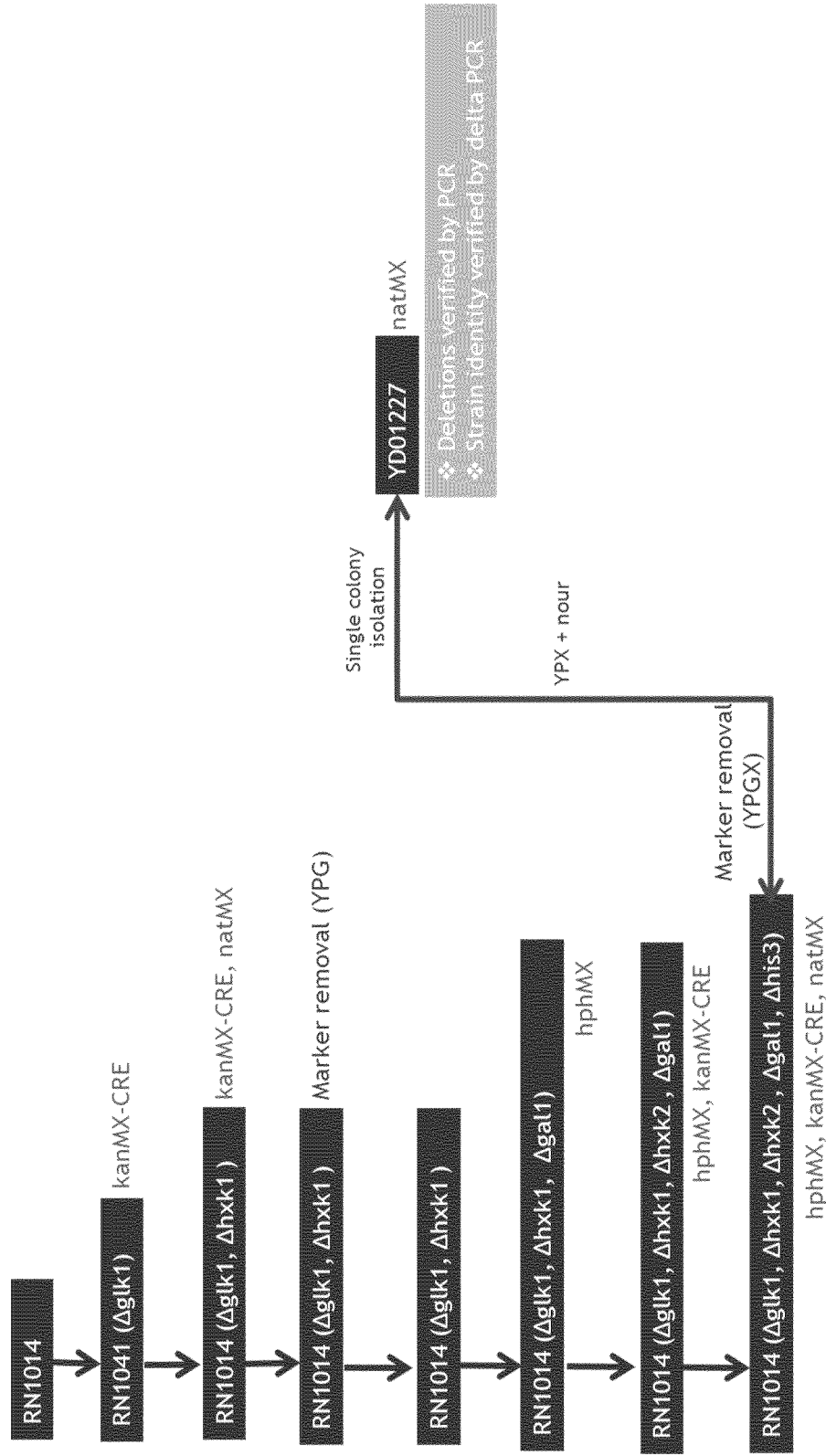


FIG. 3

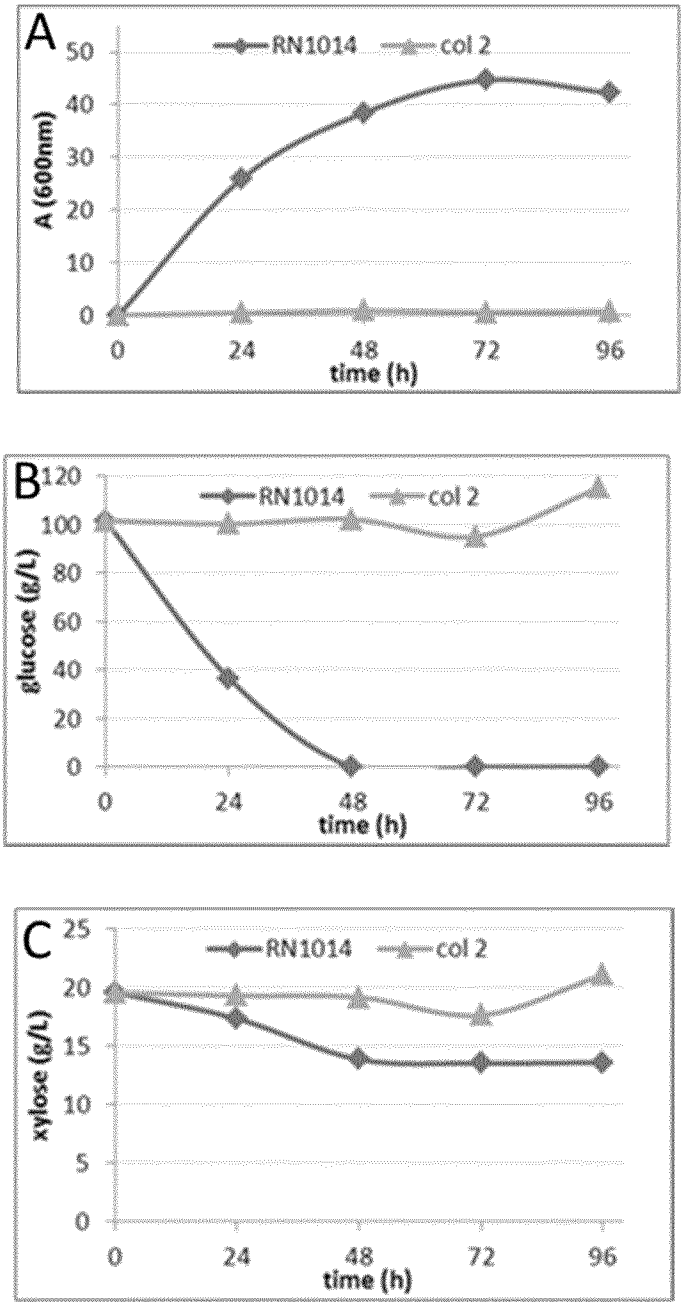


FIG. 4

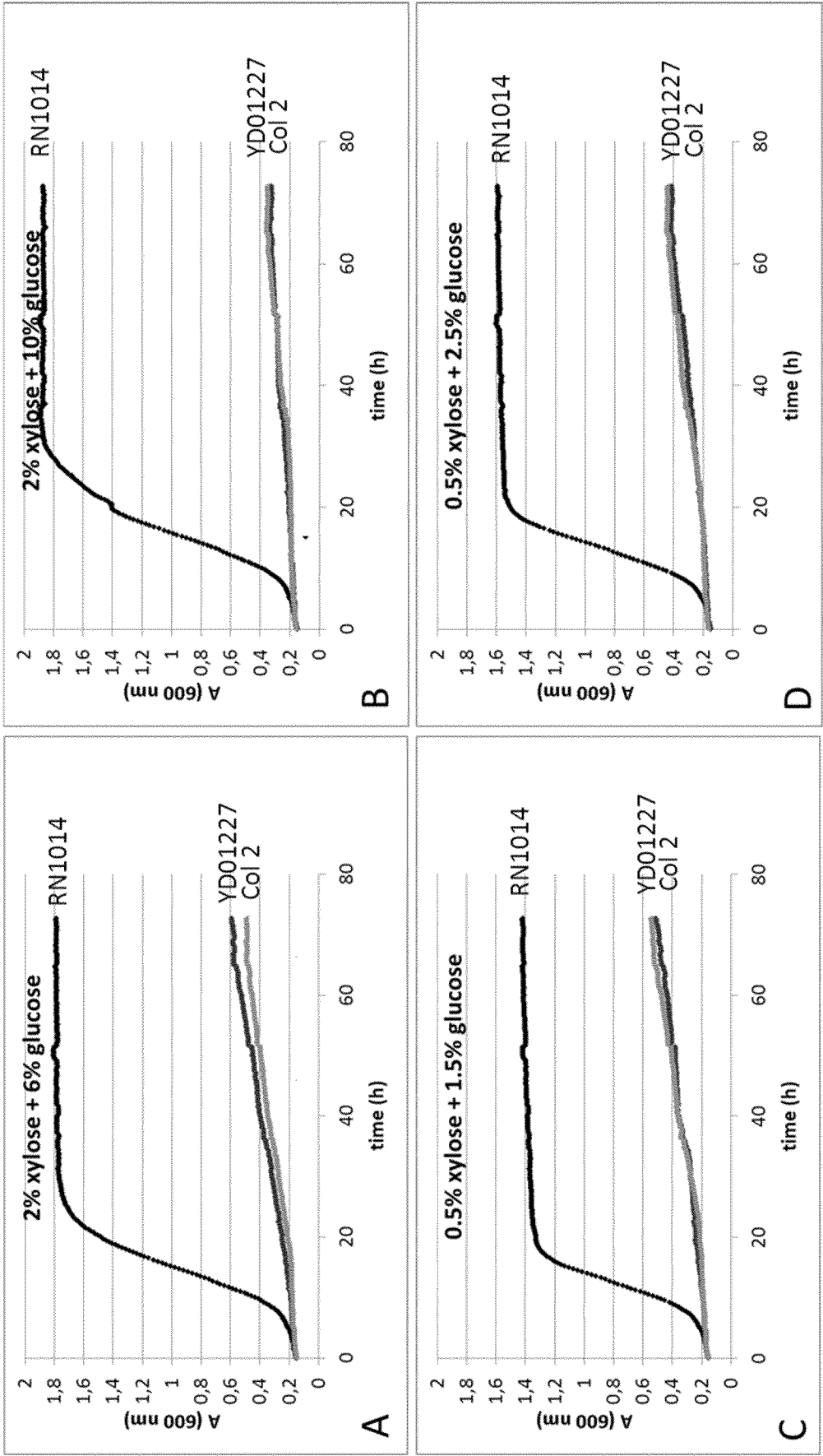


FIG. 5

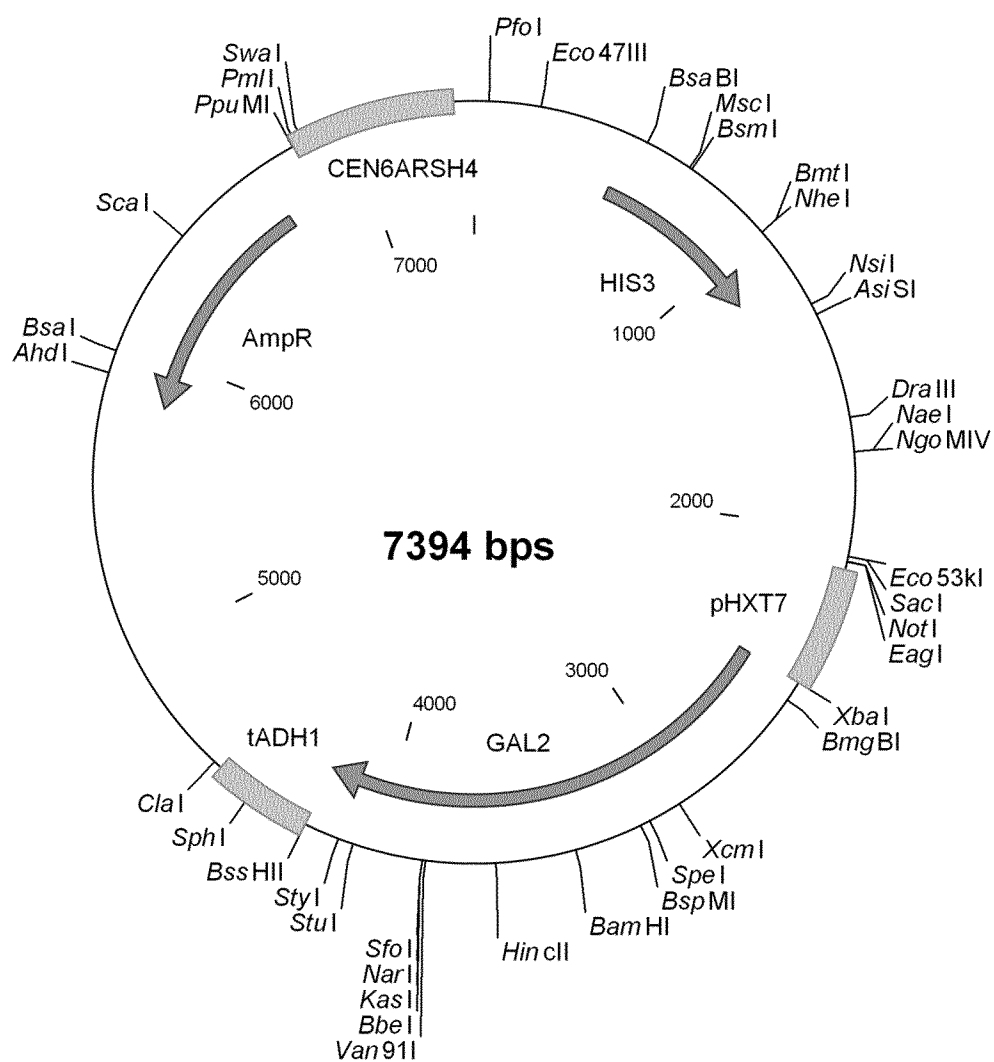


FIG. 6

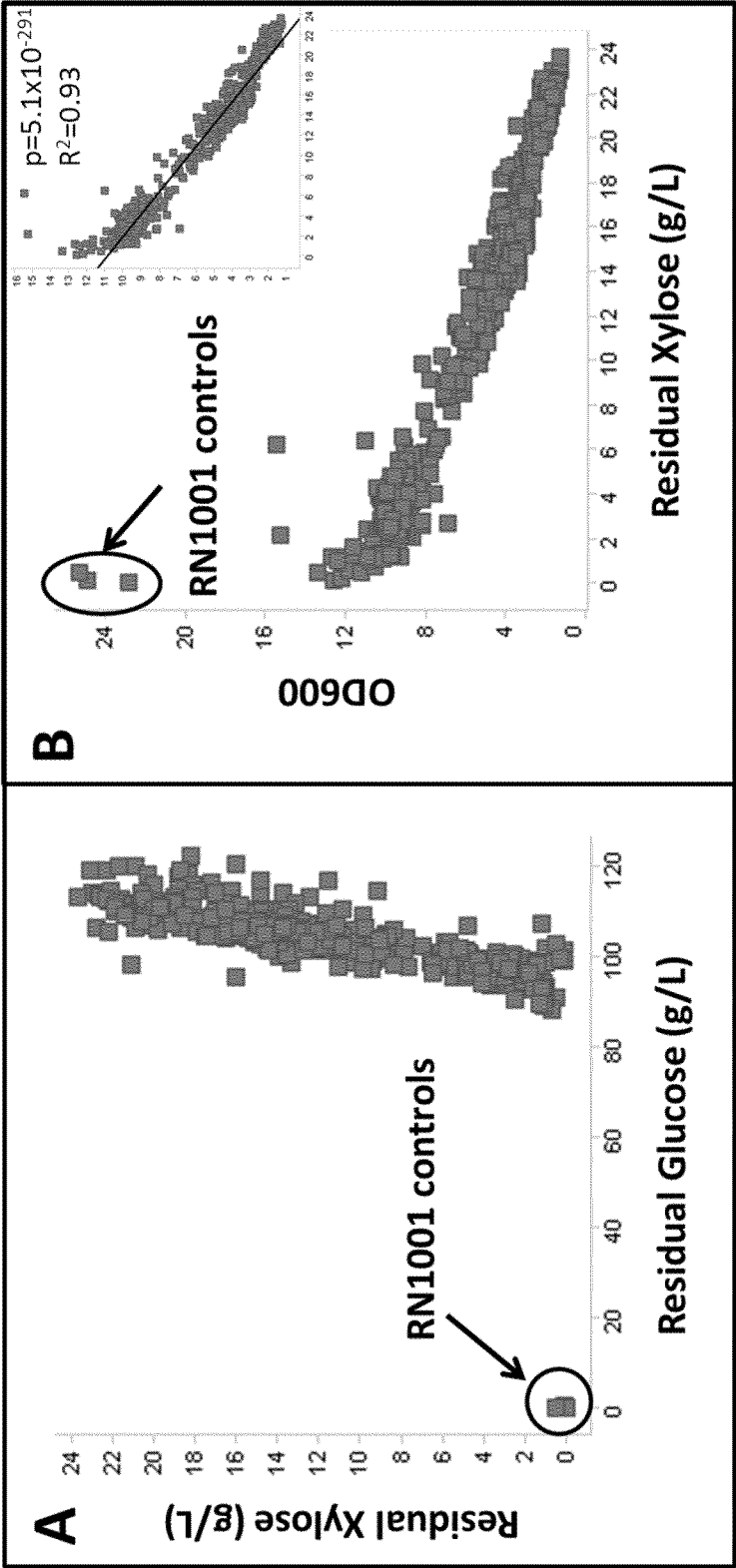


FIG. 7

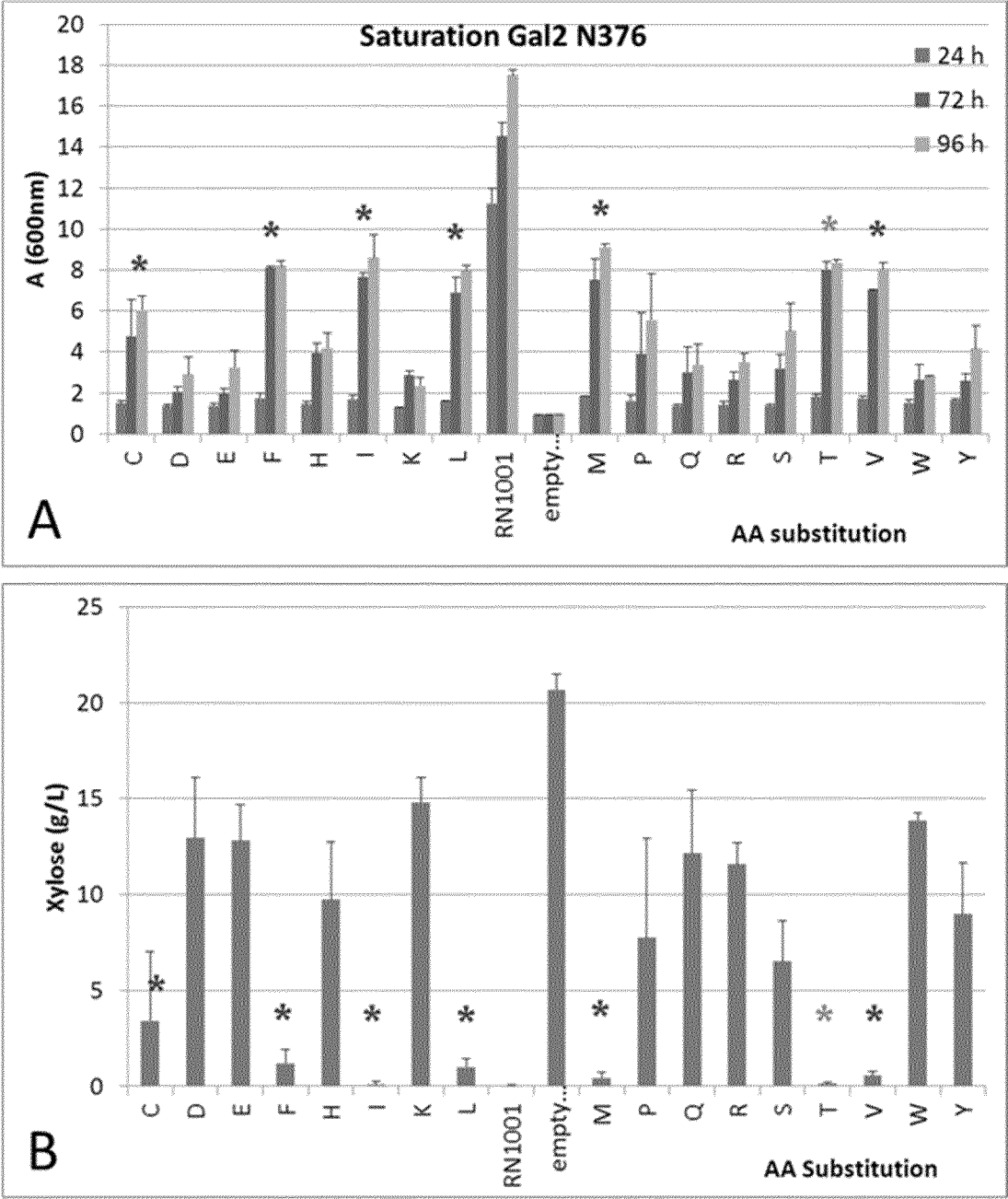


FIG. 8

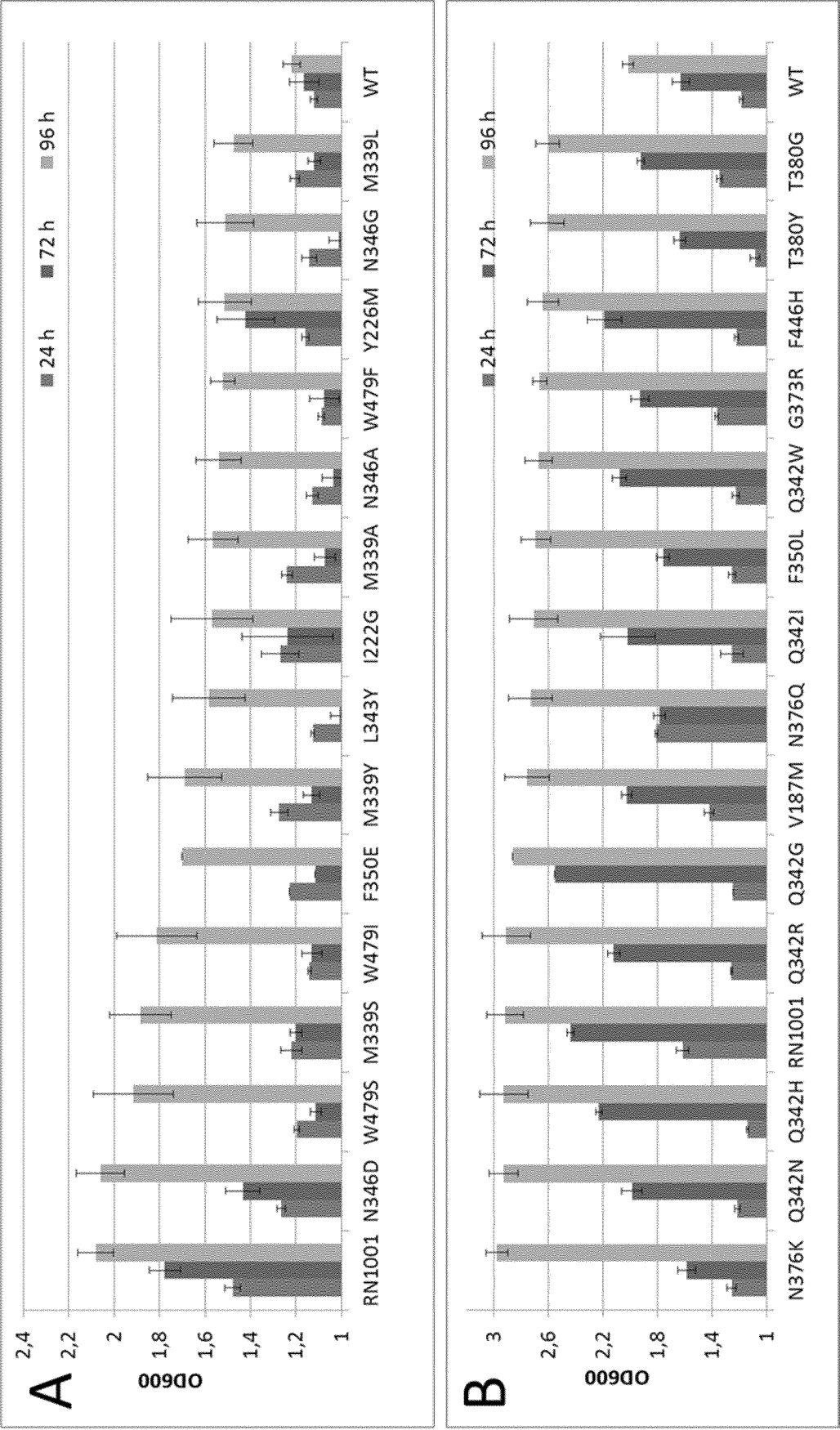


FIG. 9

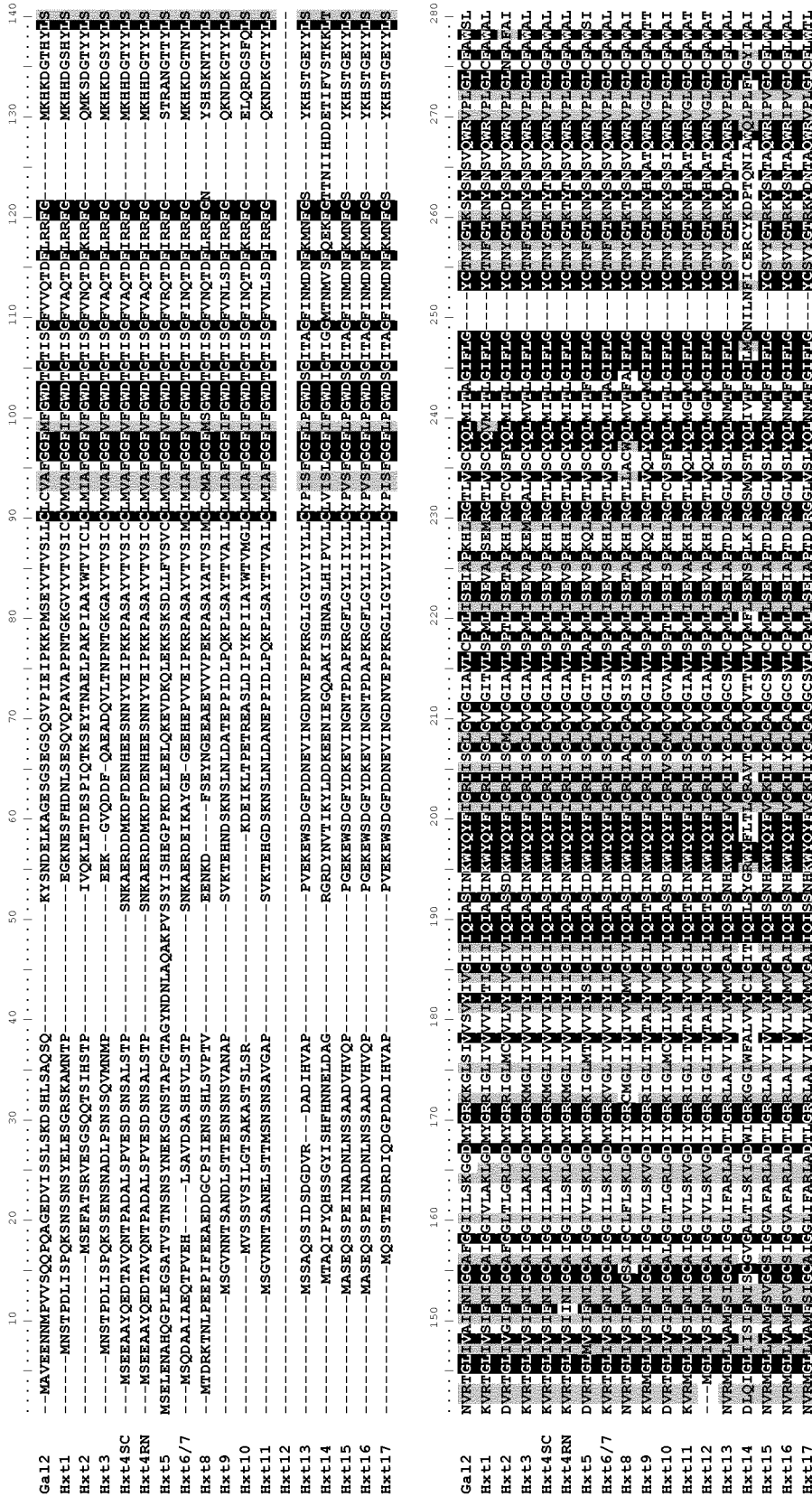


FIG. 10A

[illegible]

FIG. 10C

13/36

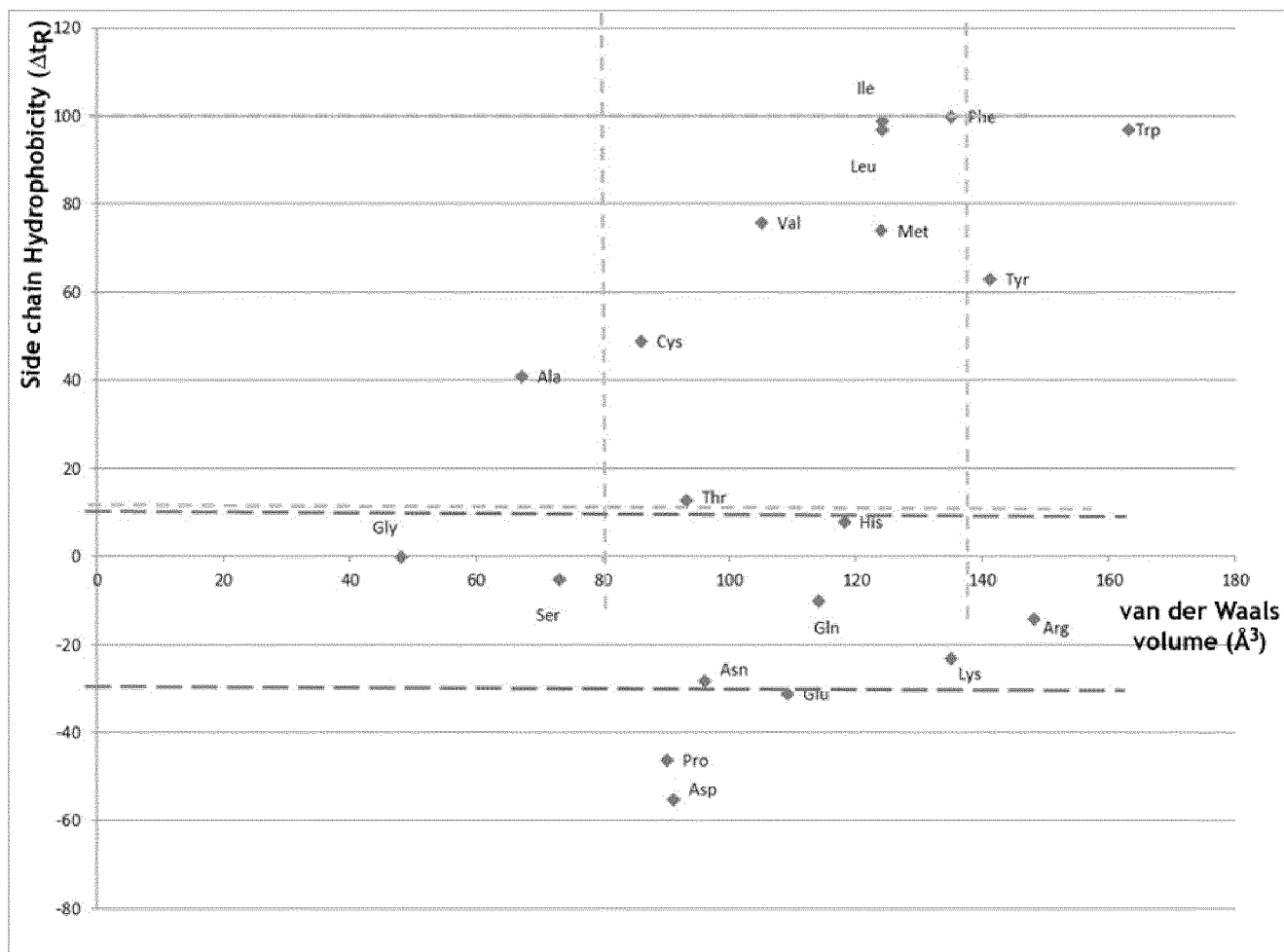


FIG. 11

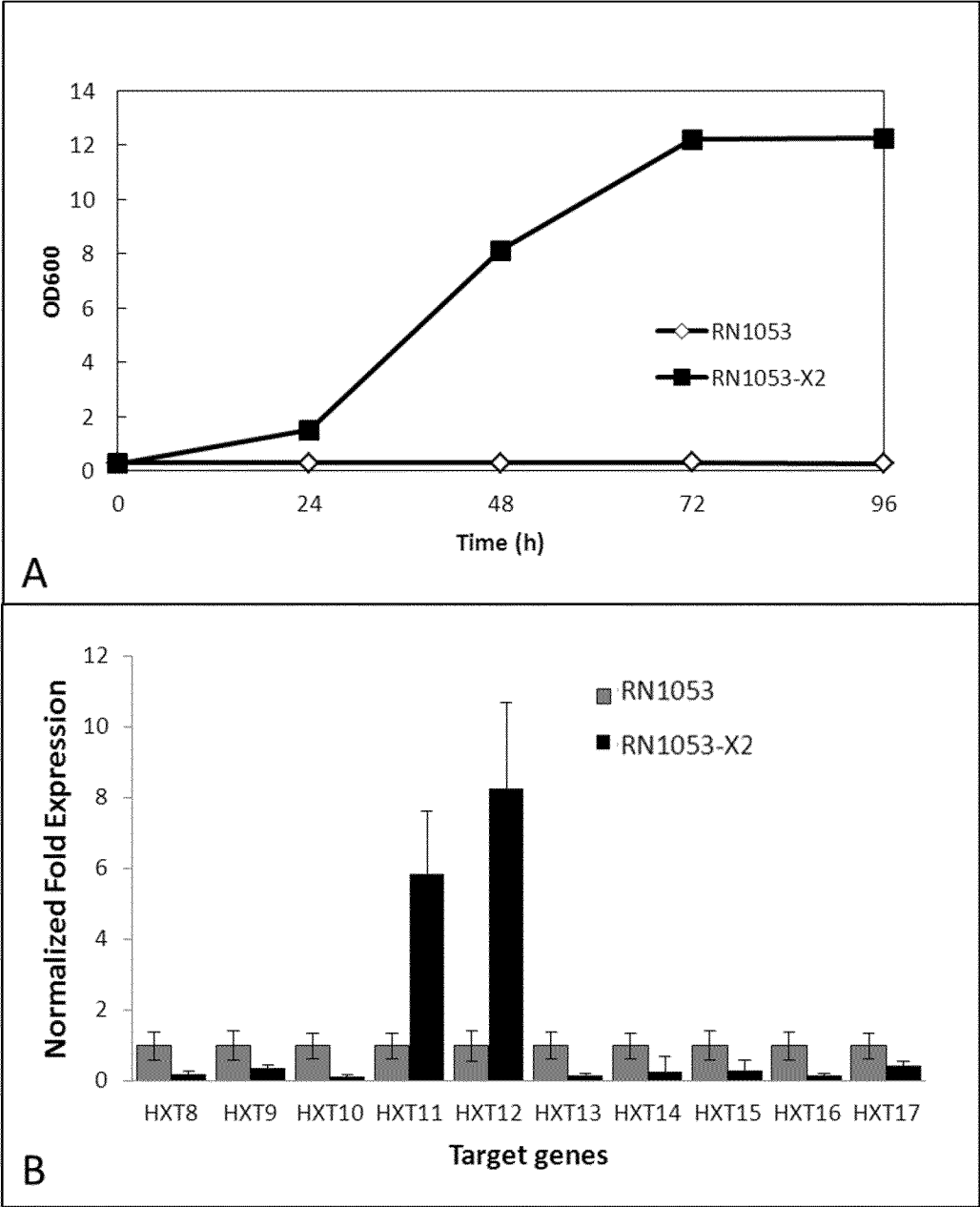


Fig. 12 (A-B)

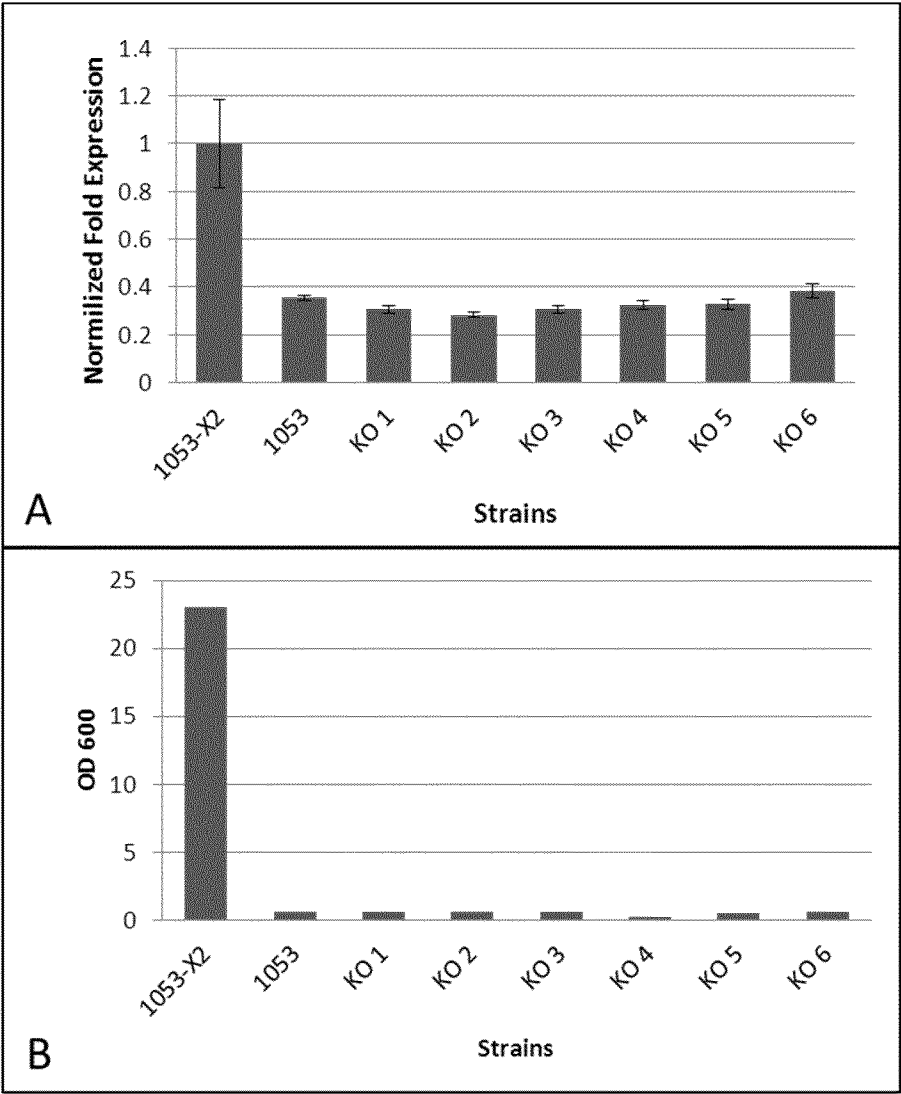


Fig. 13 (A-B)

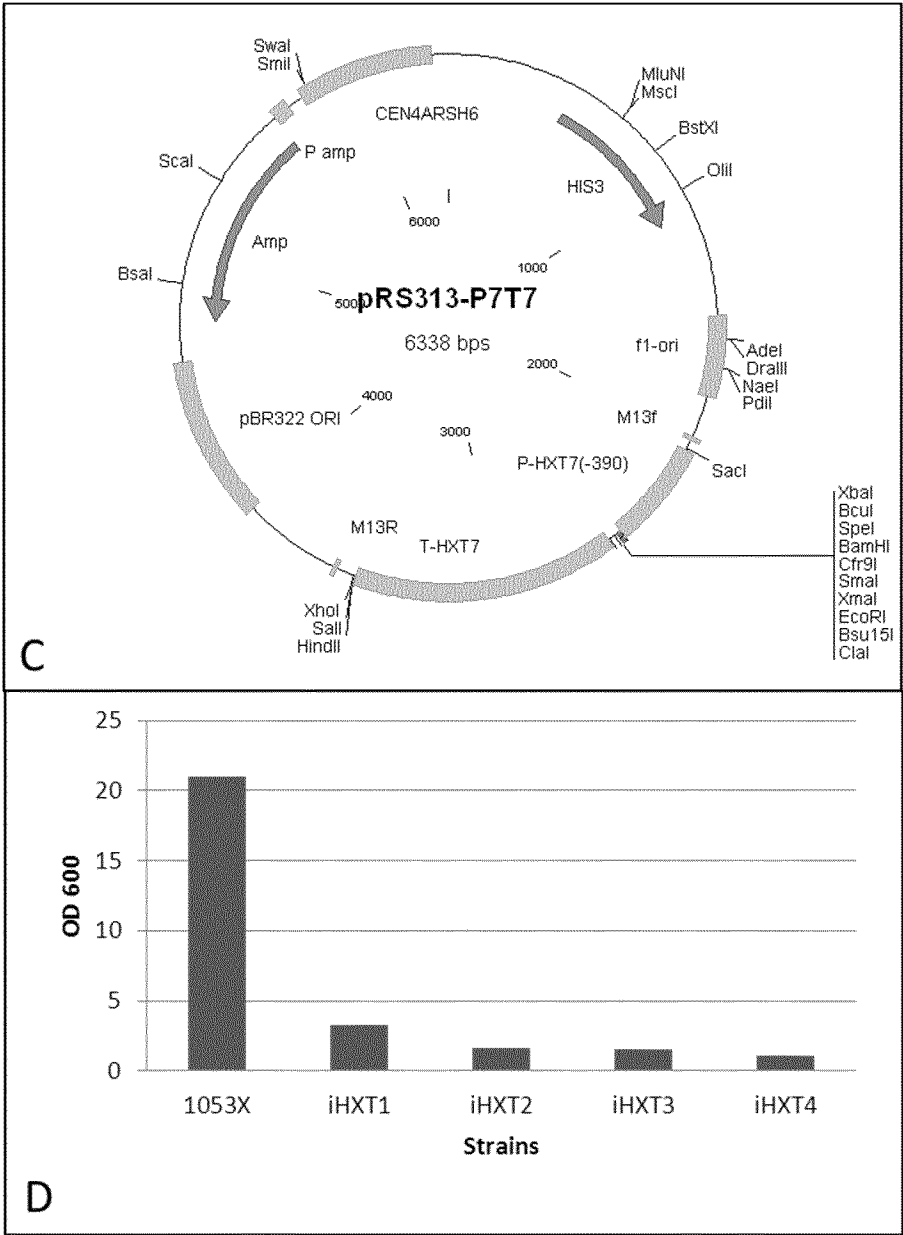


Fig. 13 (C-D)

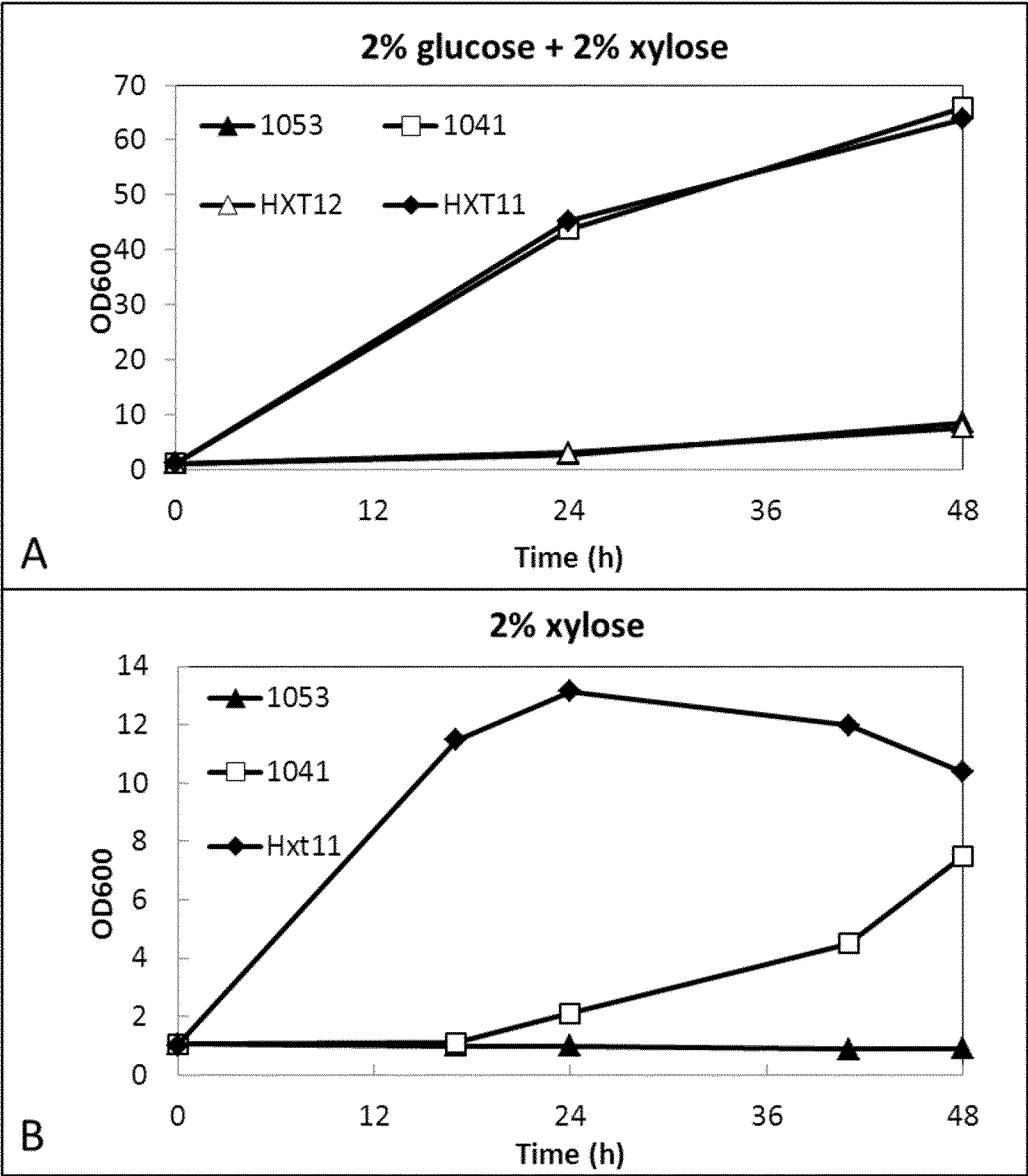


Fig. 14 (A-B)

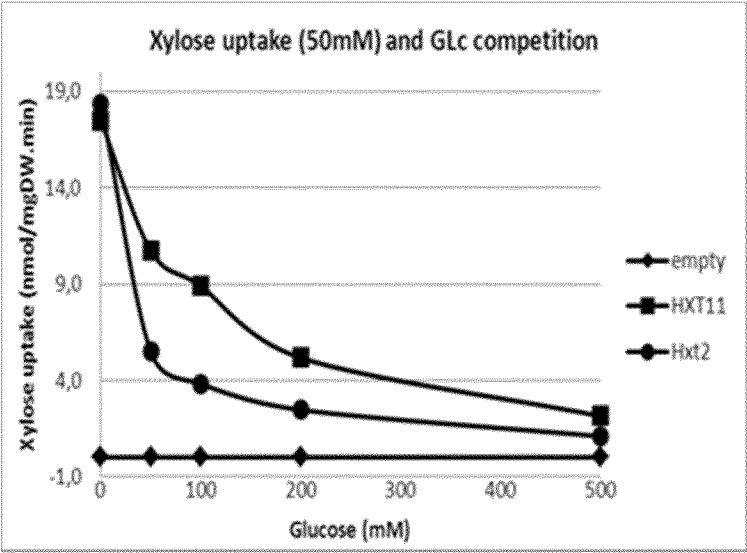


Fig.15

19/36

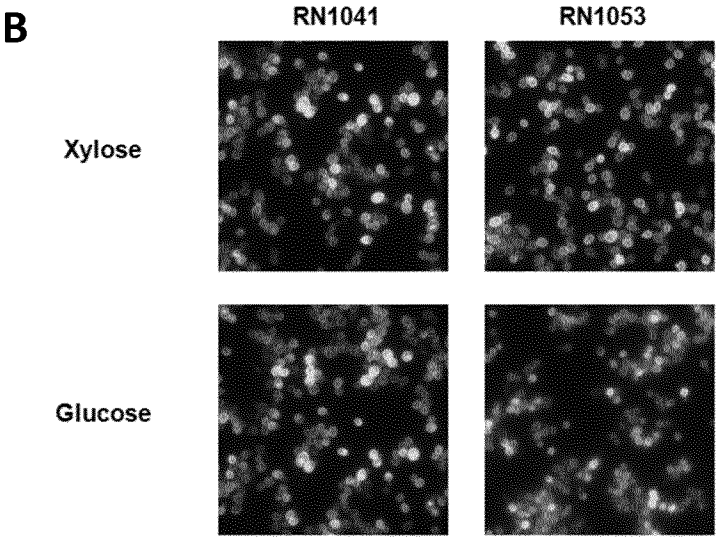
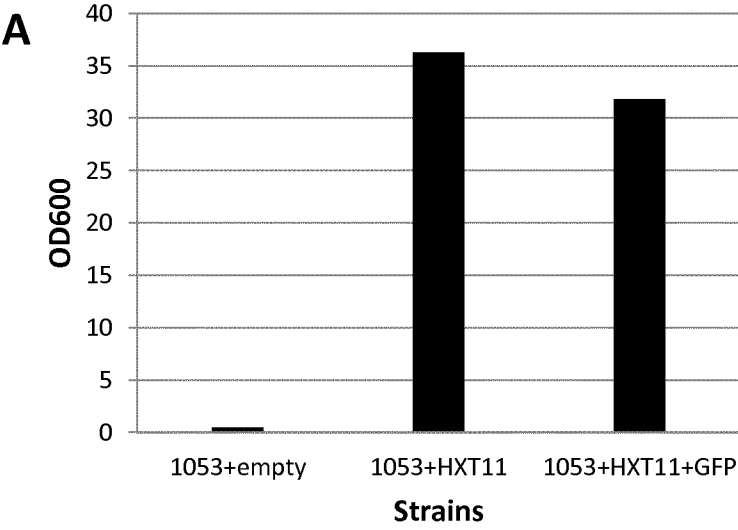


Fig. 16 (A-B)

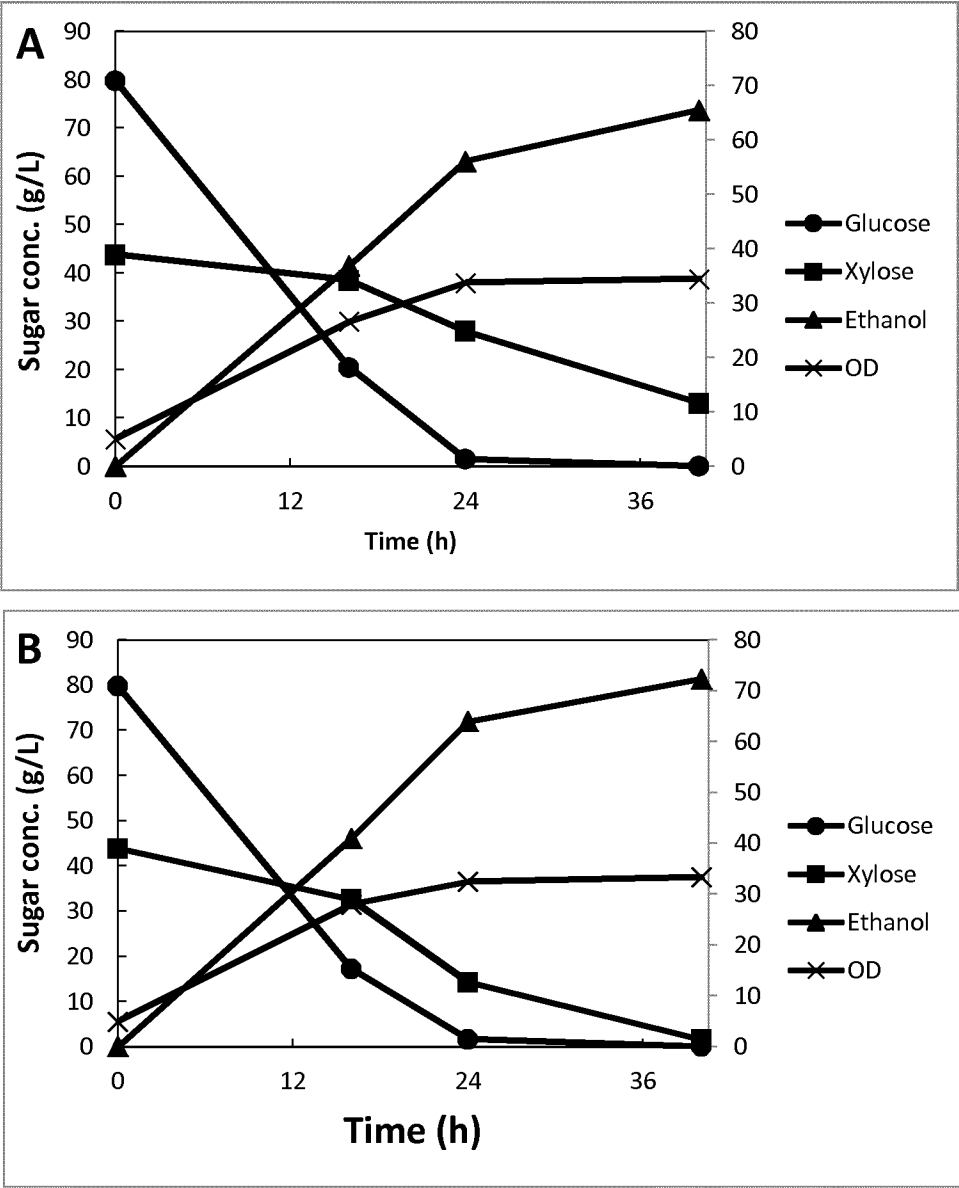


Figure 17 (A-B)

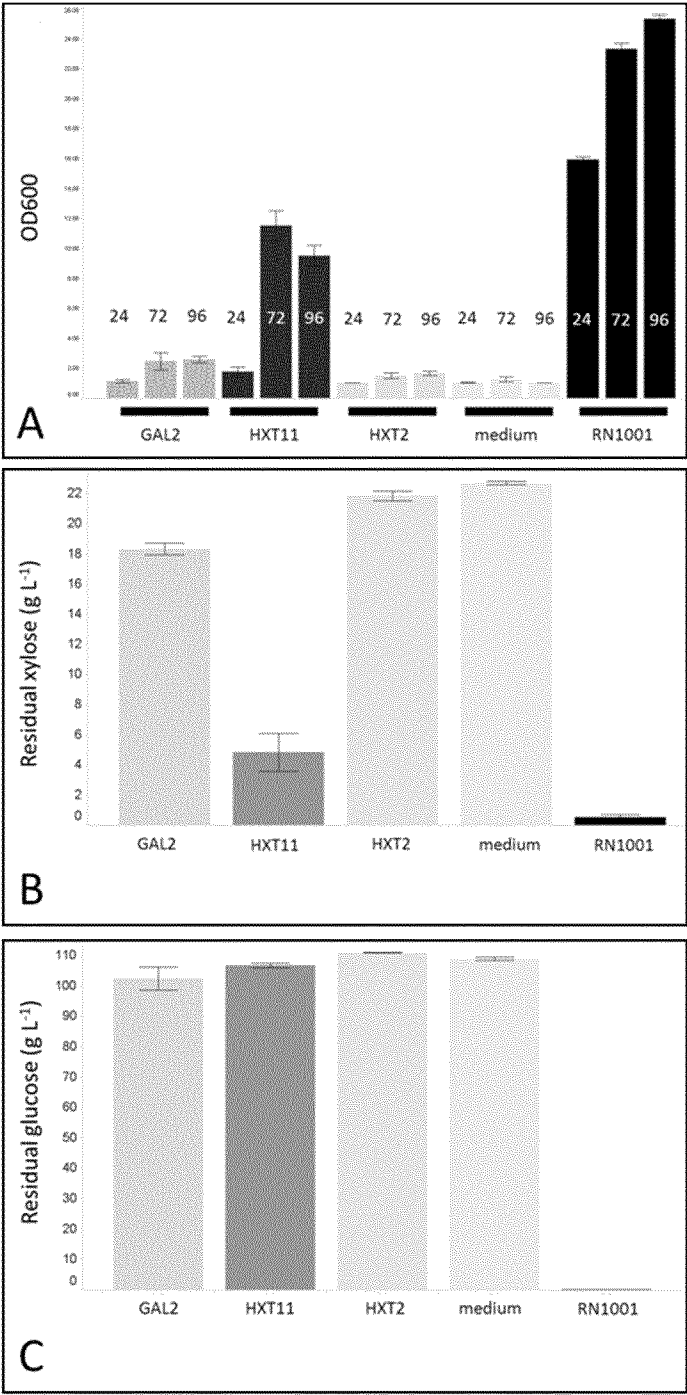


Fig. 18 (A-C)

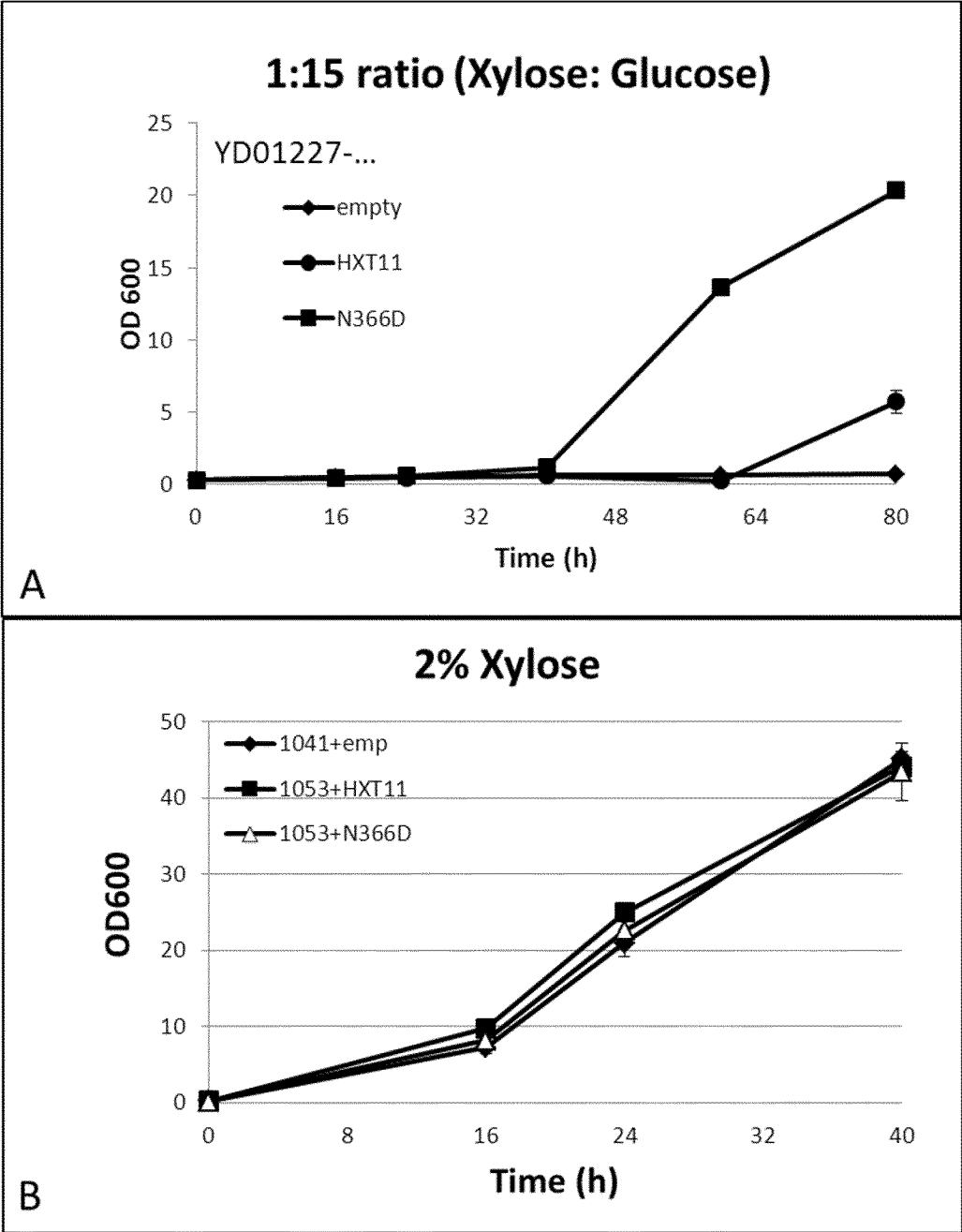


Fig. 19 (A-B)

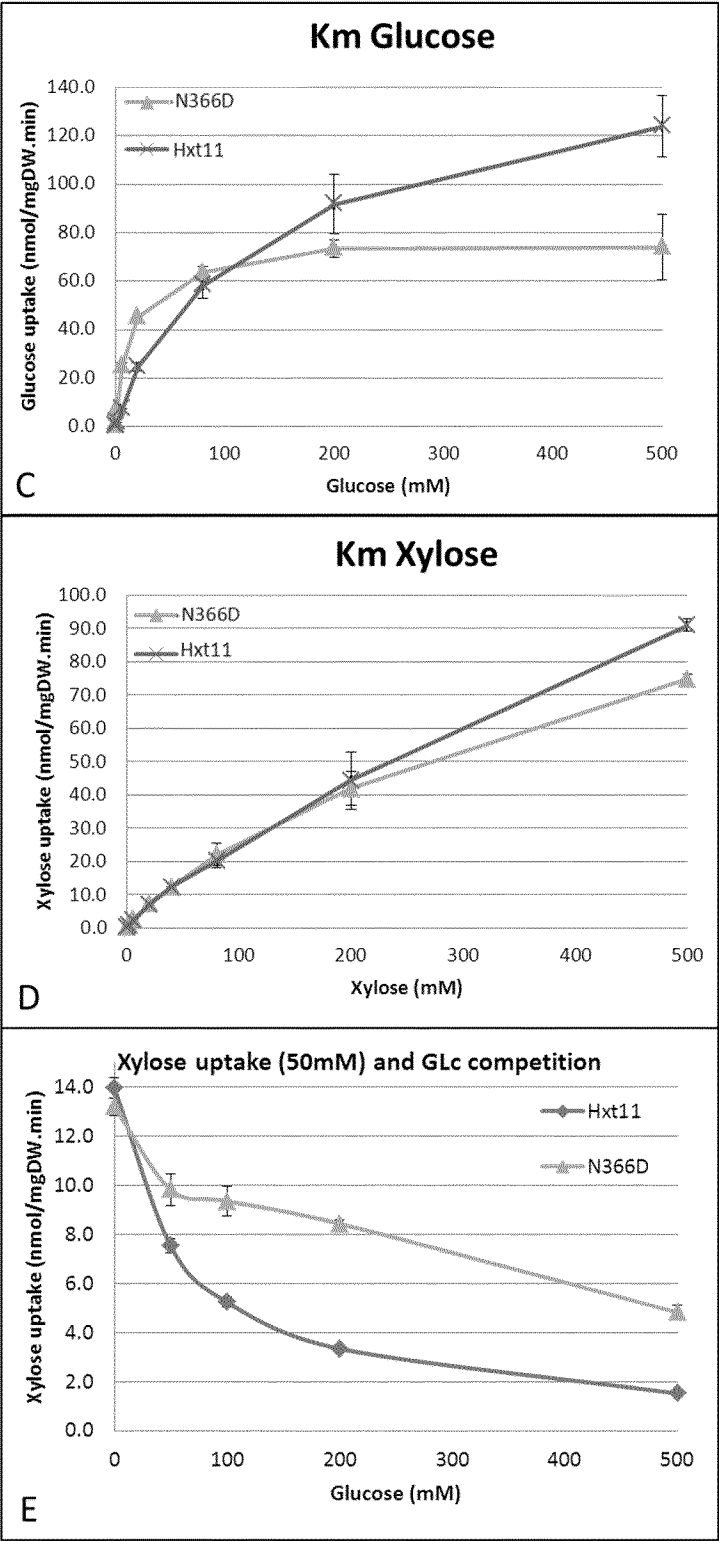


Fig. 19 (C-E)

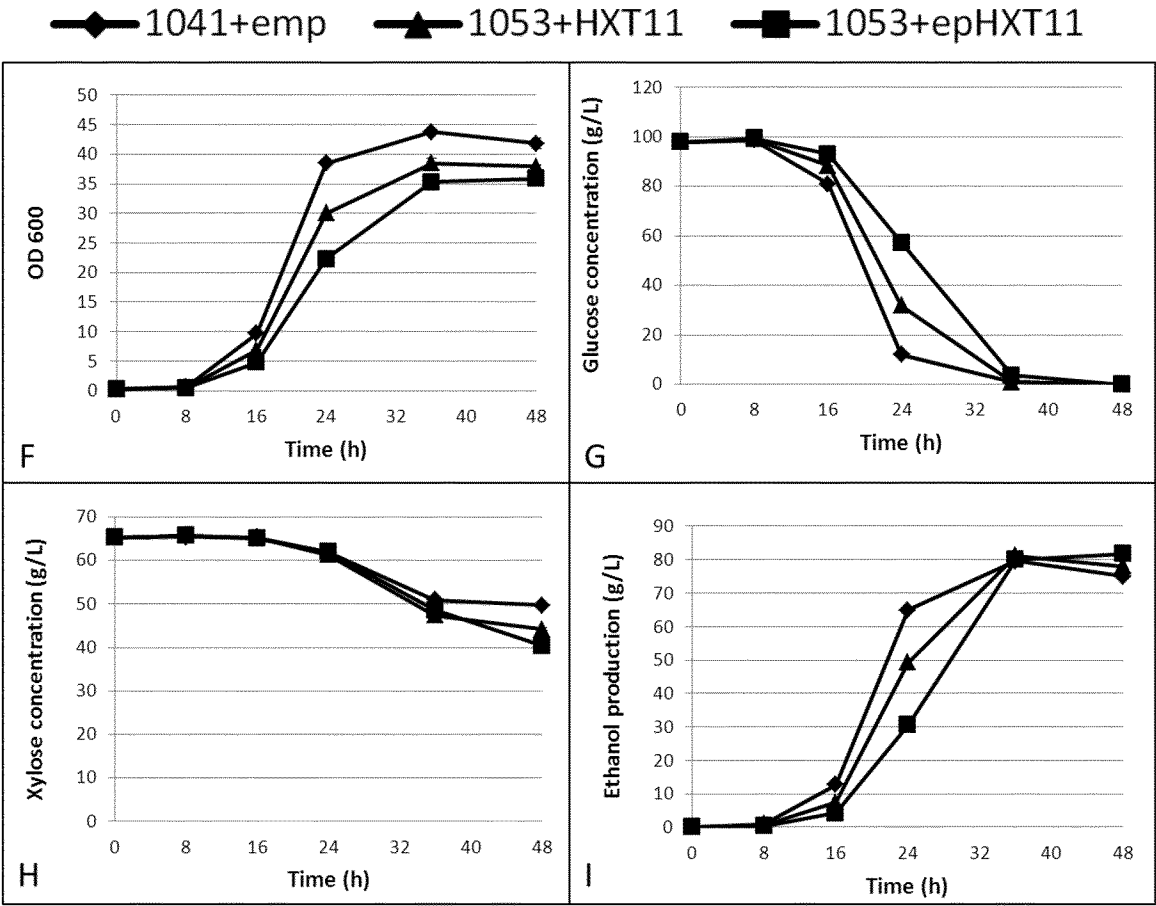


Fig. 19(f-l)

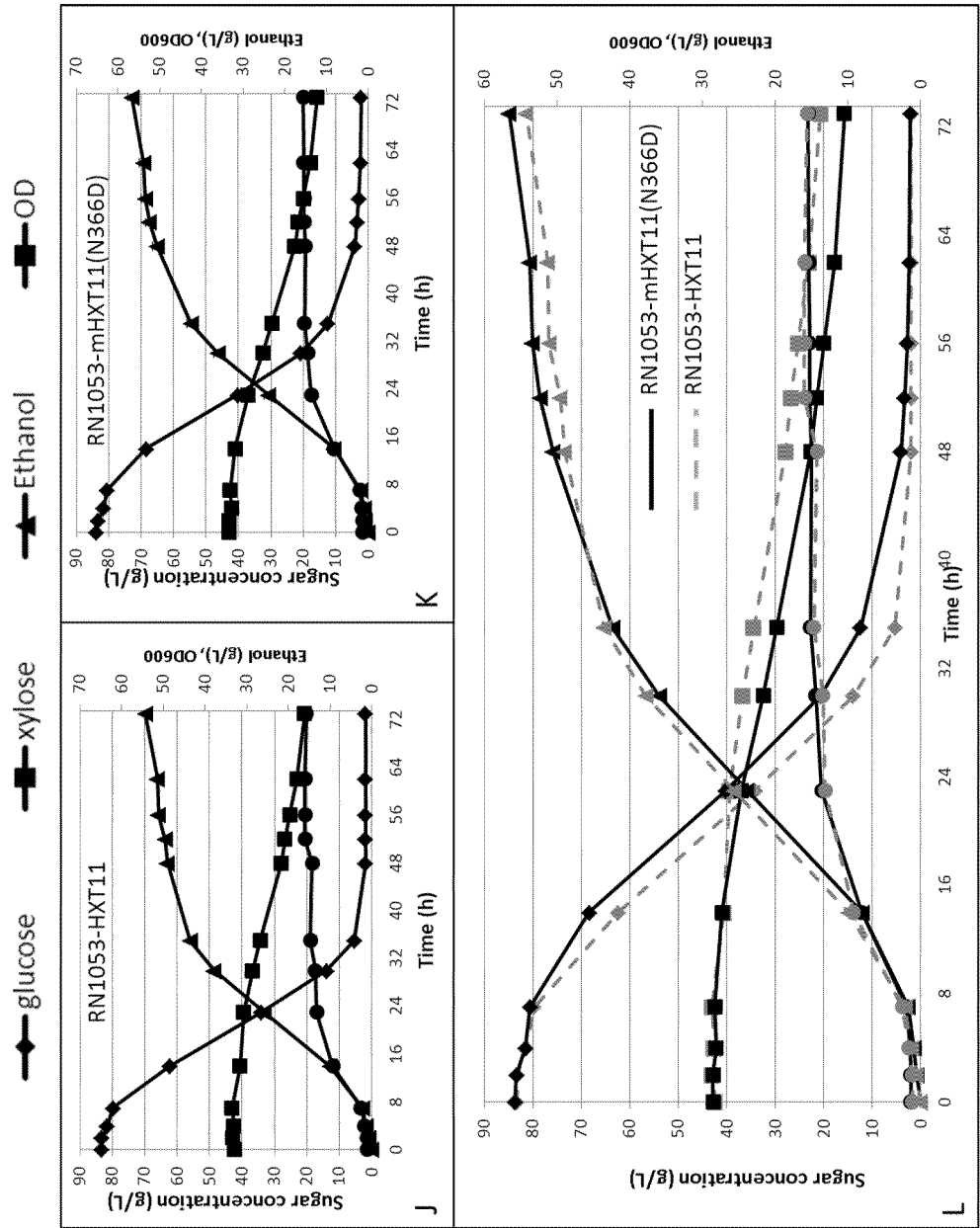


Fig. 19 (J-L)

26/36

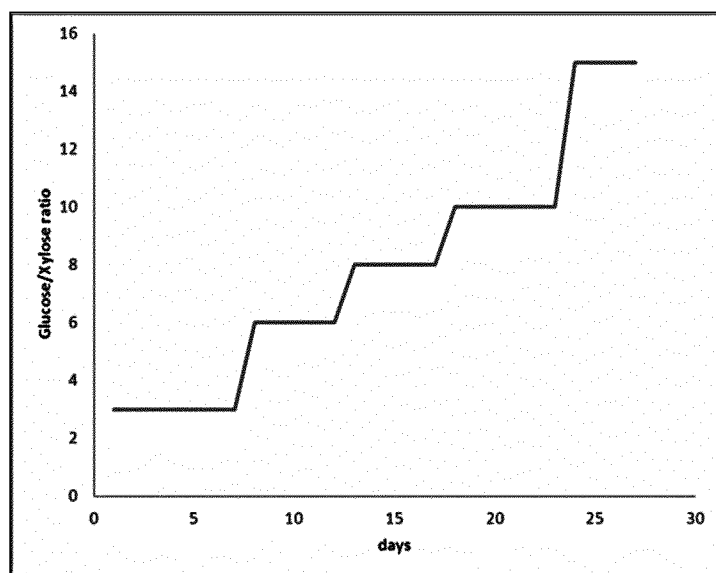


Fig. 20

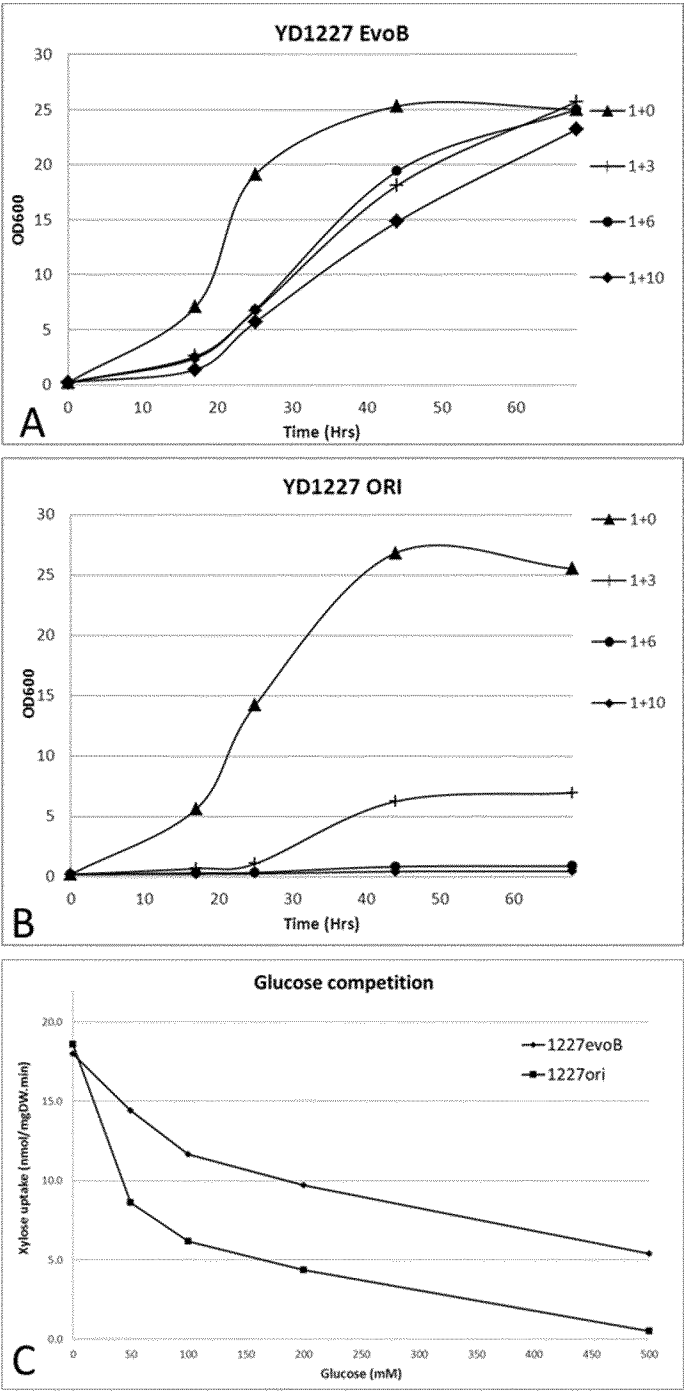


Fig. 21 (A-C)

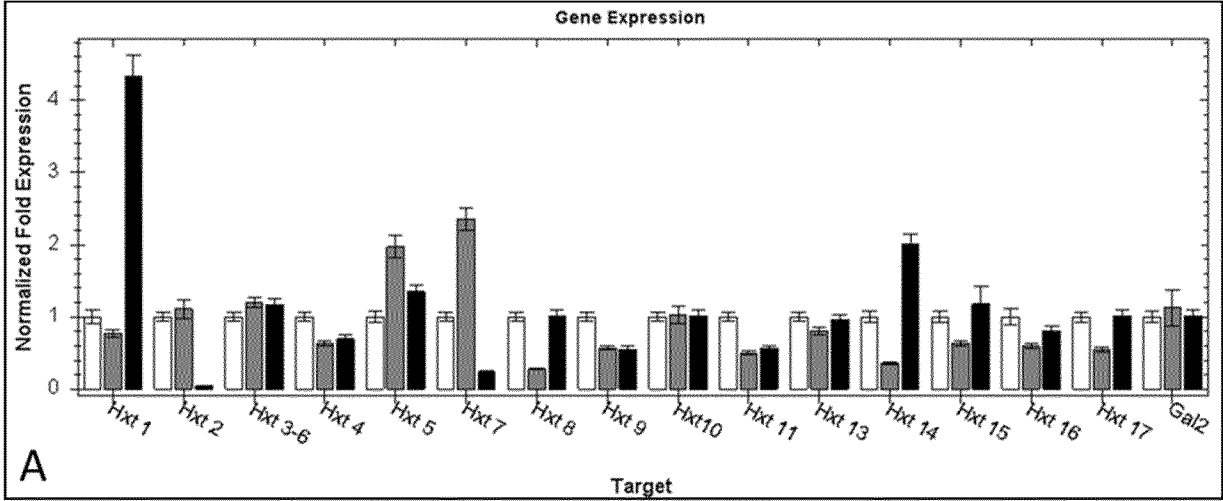


Fig. 22 (A)

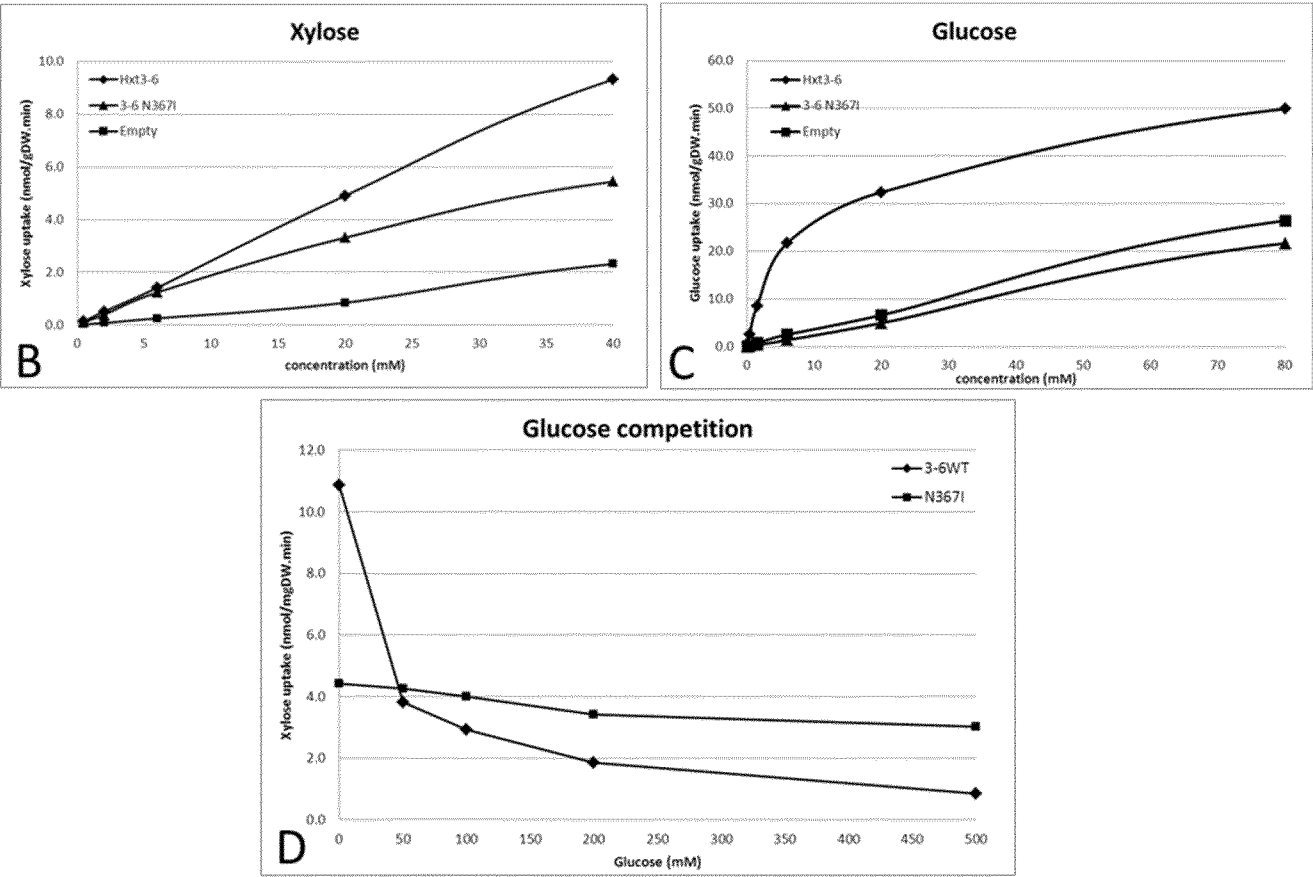


Fig. 22 (B-D)

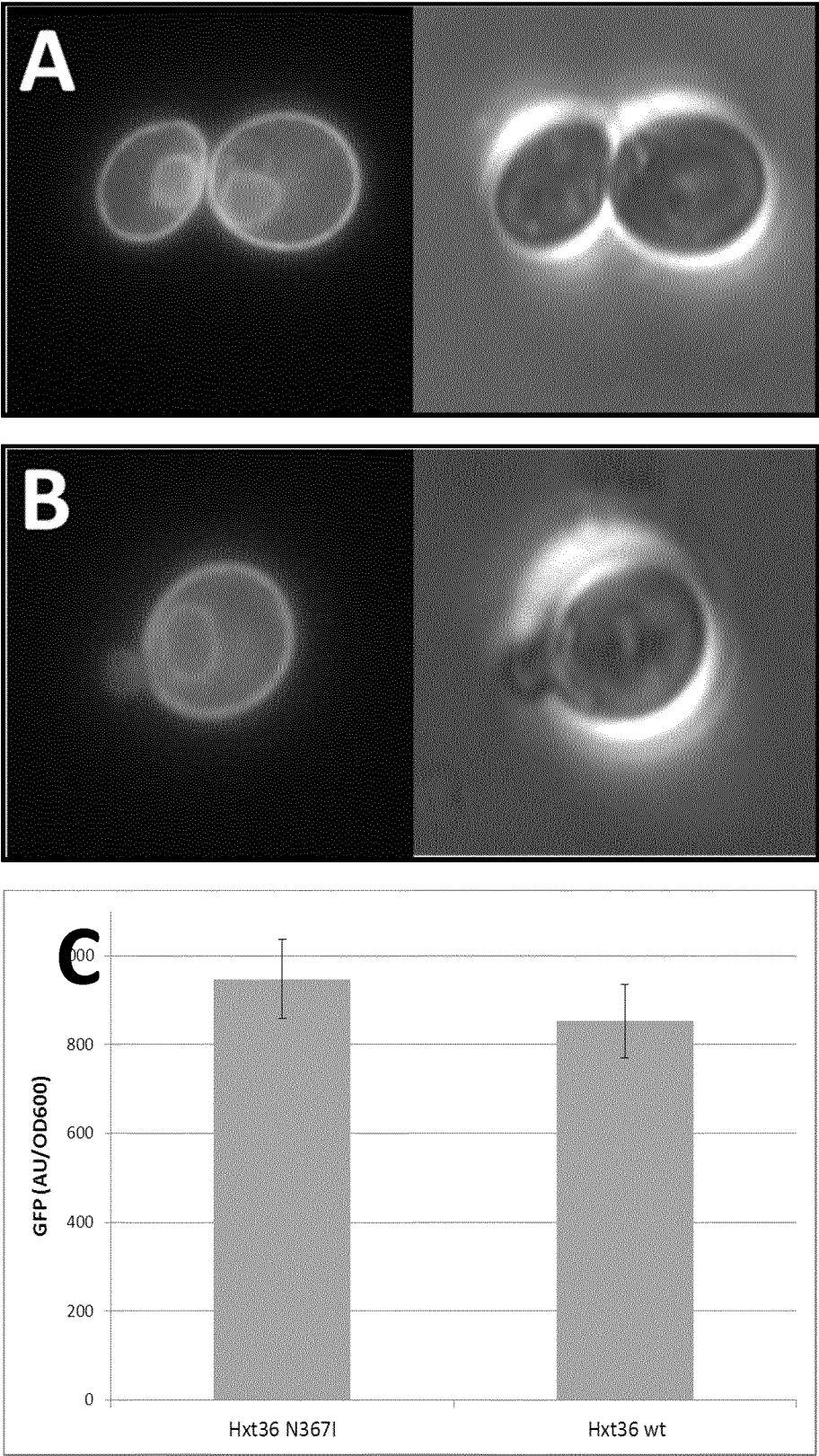


Fig. 23

31/36

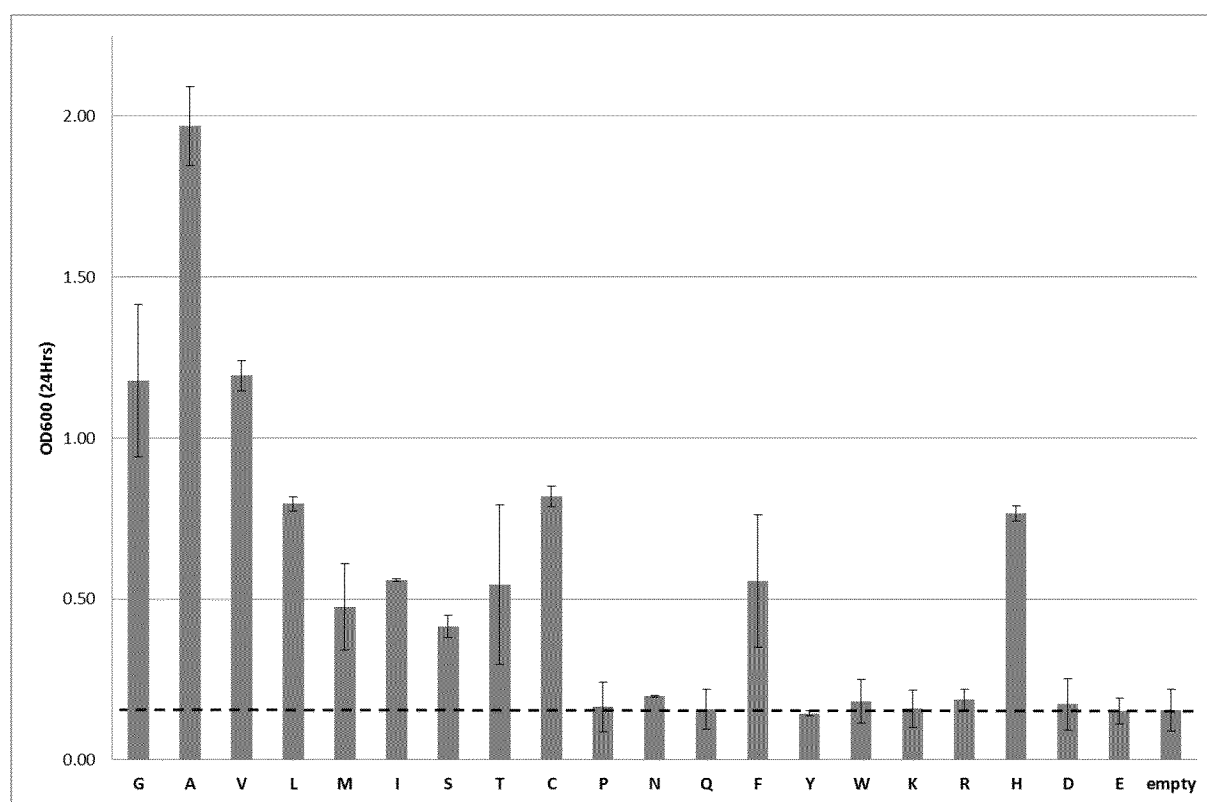


Fig. 24

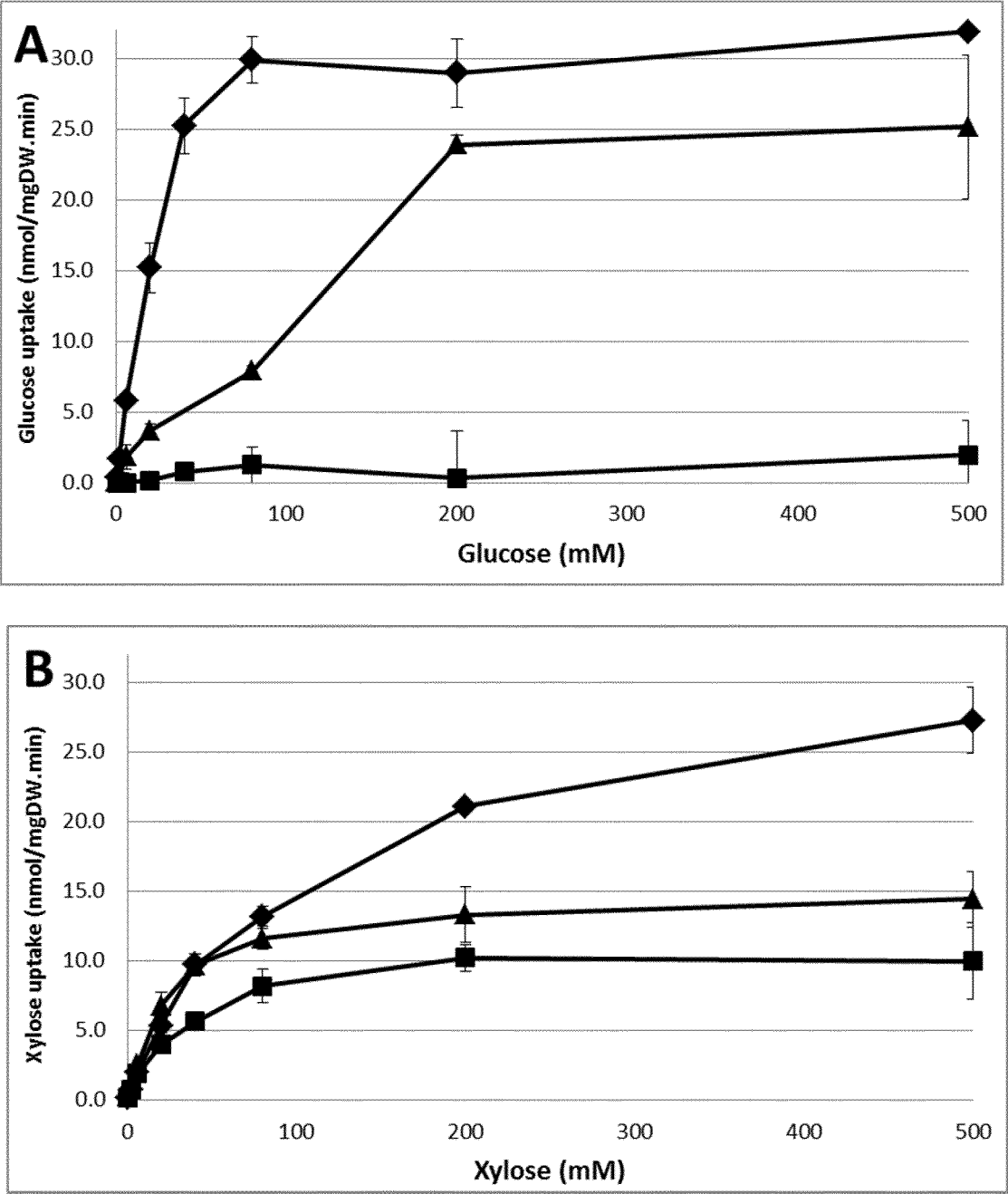


Fig. 25

33/36

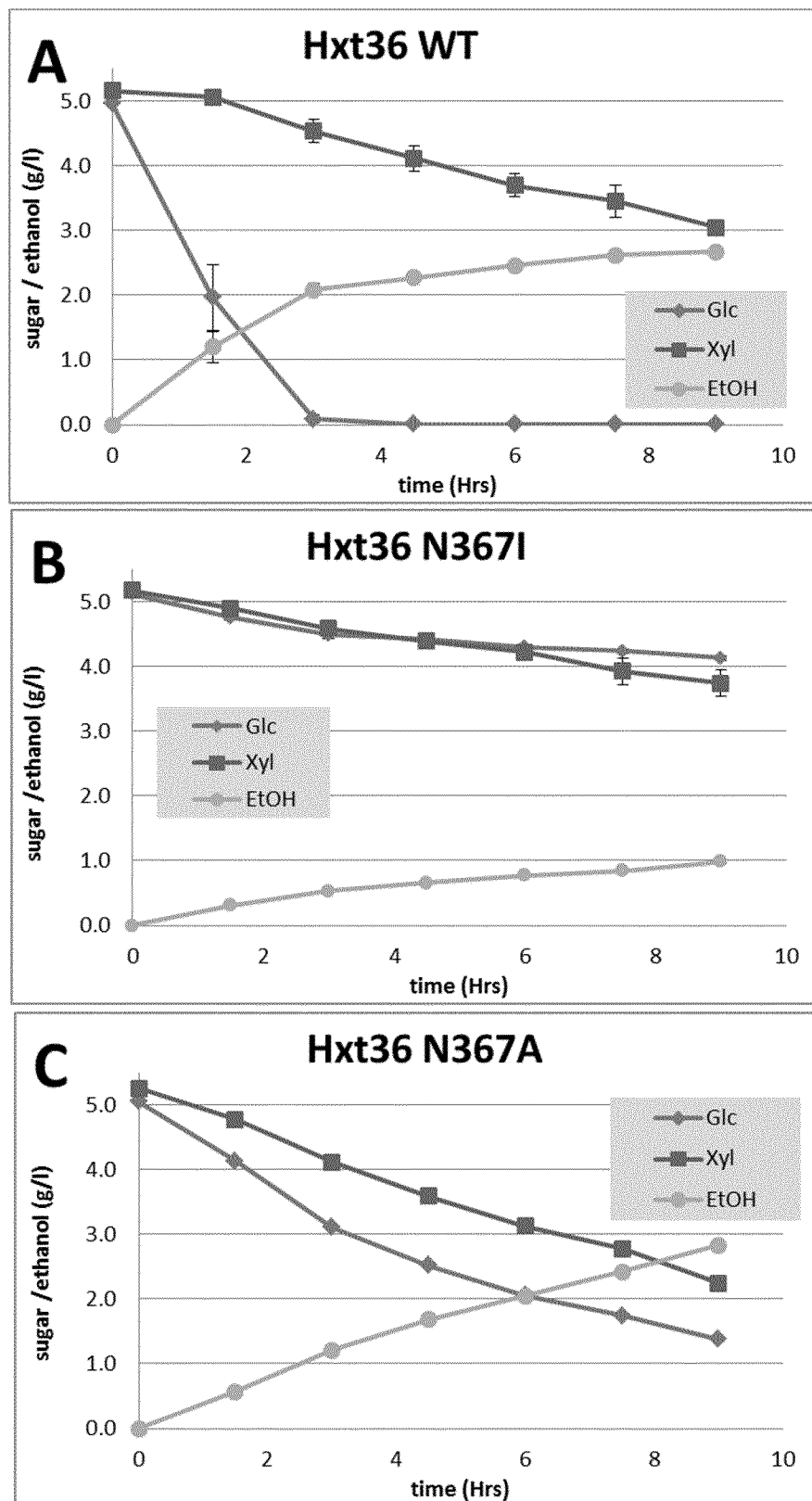


Fig. 26

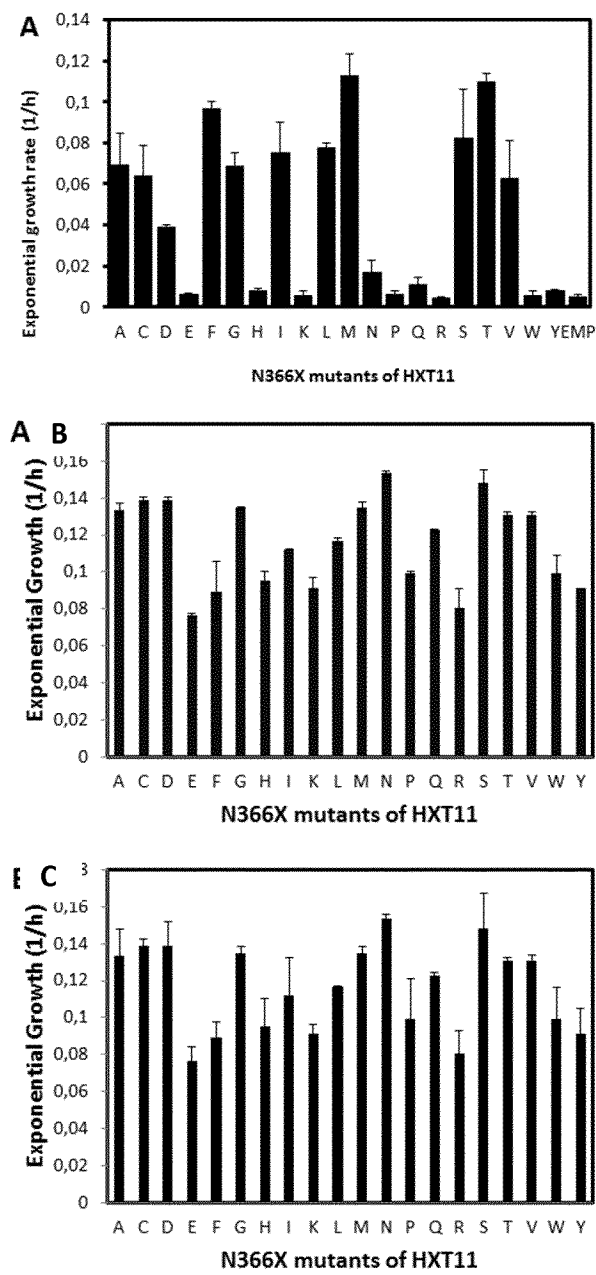


Fig. 27

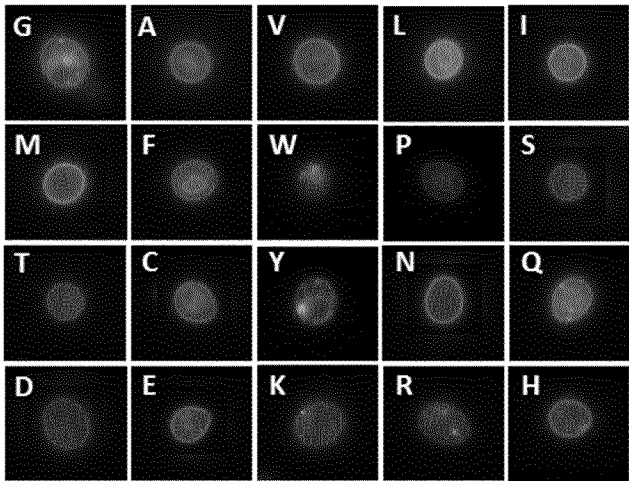


Fig. 28

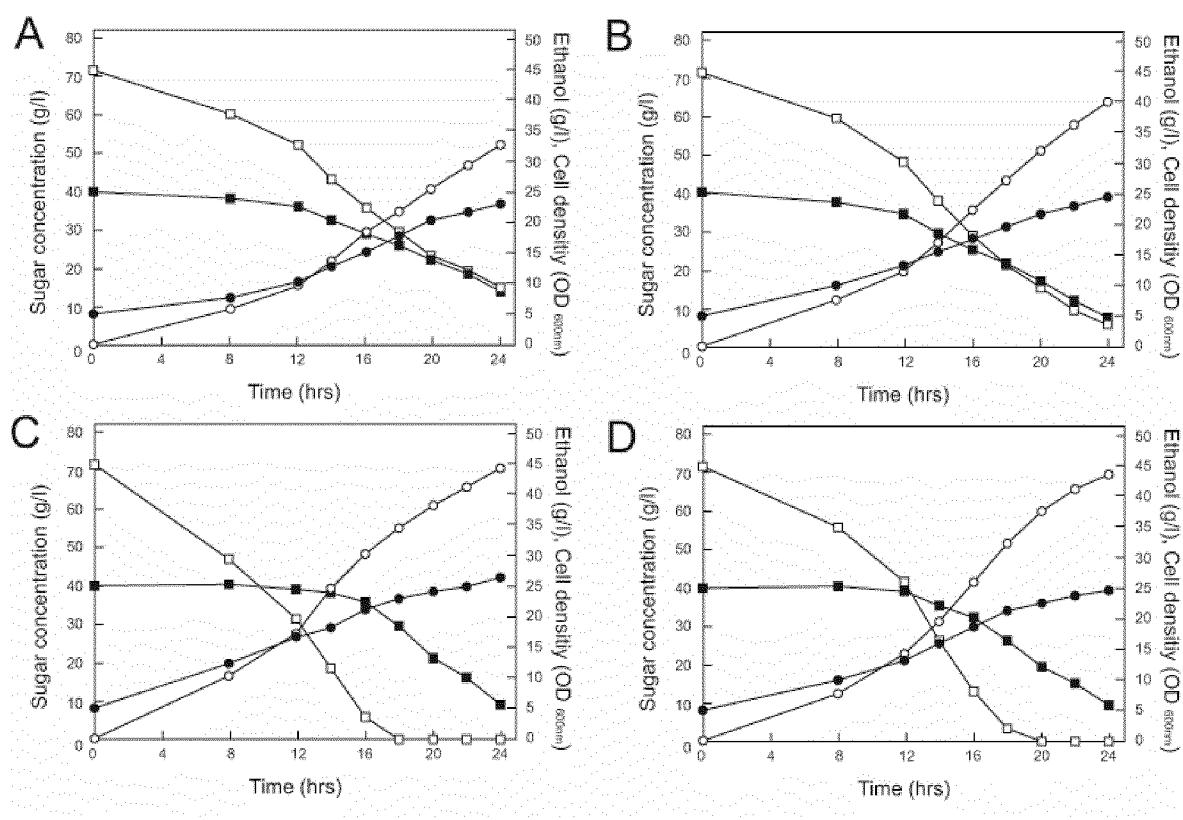


Fig. 29