



(86) Date de dépôt PCT/PCT Filing Date: 2015/10/22  
(87) Date publication PCT/PCT Publication Date: 2016/04/28  
(45) Date de délivrance/Issue Date: 2023/10/03  
(85) Entrée phase nationale/National Entry: 2017/03/10  
(86) N° demande PCT/PCT Application No.: EP 2015/074522  
(87) N° publication PCT/PCT Publication No.: 2016/062823  
(30) Priorité/Priority: 2014/10/22 (EP14189927.8)

(51) Cl.Int./Int.Cl. *C07K 14/395* (2006.01),  
*C12N 15/81* (2006.01), *C12P 7/08* (2006.01),  
*C12P 7/12* (2006.01)  
(72) Inventeurs/Inventors:  
BOLES, ECKHARD, DE;  
DIETZ, HEIKO, DE;  
FARWICK, ALEXANDER, FR;  
SCHADEWEG, VIRGINIA, DE;  
OREB, MISLAV, DE  
(73) Propriétaire/Owner:  
BUTALCO GMBH, CH  
(74) Agent: GOWLING WLG (CANADA) LLP

(54) Titre : VARIANTS DE TRANSPORTEUR DE GAL2 ET LEURS UTILISATIONS  
(54) Title: VARIANTS OF GAL2 TRANSPORTER AND THEIR USES

(57) Abrégé/Abstract:

The present invention relates to polypeptides which are Gal2 variants comprising at least one amino acid substitution at a position corresponding to T354, and optionally further amino acid substitution(s). The present invention further relates to nucleic acid molecules encoding the polypeptides and to host cells containing said nucleic acid molecules. The present invention further relates to a method for the production of bioethanol and/or other bio-based compounds, comprising the expression of said nucleic acid molecules, preferably in said host cells. The present invention also relates to the use of the polypeptides, nucleic acids molecule or host cells for the production of bioethanol and/or other bio-based compounds, and/or for the recombinant fermentation of biomaterial containing pentose(s), preferably D-xylose and/or L-arabinose.

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

(19) World Intellectual Property  
Organization  
International Bureau



(10) International Publication Number  
**WO 2016/062823 A1**

(43) International Publication Date  
28 April 2016 (28.04.2016)

## (51) International Patent Classification:

*C07K 14/395* (2006.01) *C12P 7/08* (2006.01)  
*C12N 15/81* (2006.01) *C12P 7/12* (2006.01)

## (21) International Application Number:

PCT/EP2015/074522

## (22) International Filing Date:

22 October 2015 (22.10.2015)

## (25) Filing Language:

English

## (26) Publication Language:

English

## (30) Priority Data:

14189927.8 22 October 2014 (22.10.2014) EP

(71) Applicant: BUTALCO GMBH [CH/CH]; Christen und Zobrist, Treuhand AG, Achereggstrasse 10, CH-6362 Stansstad (CH).

(72) Inventors: BOLES, Eckhard; Kranichsteiner Str. 112a, 64289 Darmstadt (DE). DIETZ, Heiko; Am Vogelherd 3, 81476 Munich (DE). FARWICK, Alexander; 5, Rue Horace Vernet, F-59110 La Madeleine (FR). SCHADEWEG, Virginia; Alt-Niedereschbach 32, 60437 Frankfurt (DE). OREB, Mislav; Hedderheimer Kirchstraße 37, 60439 Frankfurt (DE).

(74) Agent: BOEHMERT & BOEHMERT ANWALTS-PARTNERSCHAFT MBB; Grahn, Sibylla, Hollerallee 32, 28209 Bremen (DE).

(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, 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, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

## Published:

- with international search report (Art. 21(3))
- with sequence listing part of description (Rule 5.2(a))

(54) Title: VARIANTS OF GAL2 TRANSPORTER AND THEIR USES

(57) Abstract: The present invention relates to polypeptides which are Gal2 variants comprising at least one amino acid substitution at a position corresponding to T354, and optionally further amino acid substitution(s). The present invention further relates to nucleic acid molecules encoding the polypeptides and to host cells containing said nucleic acid molecules. The present invention further relates to a method for the production of bioethanol and/or other bio-based compounds, comprising the expression of said nucleic acid molecules, preferably in said host cells. The present invention also relates to the use of the polypeptides, nucleic acids molecule or host cells for the production of bioethanol and/or other bio-based compounds, and/or for the recombinant fermentation of biomaterial containing pentose(s), preferably D-xylose and/or L-arabinose.



WO 2016/062823 A1

## Variants of Gal2 transporter and their uses

The present invention relates to polypeptides which are Gal2 variants comprising at least one amino acid substitution at a position corresponding to T354, and optionally further amino acid substitution(s). The present invention further relates to nucleic acid molecules encoding the polypeptides and to host cells containing said nucleic acid molecules. The present invention further relates to a method for the production of bioethanol and/or other bio-based compounds, comprising the expression of said nucleic acid molecules, preferably in said host cells. The present invention also relates to the use of the polypeptides, nucleic acids molecule or host cells for the production of bioethanol and/or other bio-based compounds, and/or for the recombinant fermentation of biomaterial containing pentose(s), preferably D-xylose and/or L-arabinose.

## BACKGROUND OF THE INVENTION

The beer, wine and baking yeast *Saccharomyces cerevisiae* has already been used for centuries for the production of bread, wine and beer owing to its characteristic of fermenting sugar to ethanol and carbon dioxide. In biotechnology, *S. cerevisiae* is used particularly in ethanol production for industrial purposes, in addition to the production of heterologous proteins. Ethanol is used in numerous branches of industry as an initial substrate for syntheses. Ethanol is gaining increasing importance as an alternative fuel, due to the increasingly scarce presence of oil, the rising oil prices and continuously increasing need for petrol worldwide. Moreover, *S. cerevisiae* is also used for the production of other biofuels or valuable biochemical compounds like isobutanol, succinic acid, farnesen/farnesan or artemisinin.

In order to make possible a favourably-priced and efficient biofuel production, the use of biomass containing lignocellulose, such as for example straw, waste from the timber industry and agriculture and the organic component of everyday household waste, presents itself as an initial substrate. Firstly, said biomass is very convenient and secondly is present in large quantities. The three major components of lignocellulose are lignin, cellulose and hemicellulose. Hemicellulose, which is the second most frequently occurring polymer after

cellulose, is a highly branched heteropolymer. It consists of pentoses (L-arabinose, D-xylose), uronic acids (4-O-methyl-D-glucuronic acid, D-galacturonic acid) and hexoses (D-mannose, D-galactose, L-rhamnose, D-glucose). Although hemicellulose can be hydrolyzed more easily than cellulose, it contains the pentoses L-arabinose and D-xylose, which can normally not be converted by the yeast *S.cerevisiae*.

In order to be able to use pentoses for fermentations, these must firstly enter the cell through the plasma membrane. Although *S.cerevisiae* is not able to metabolize D-xylose or L-arabinose, it can take up D-xylose or L-arabinose into the cell. However, *S. cerevisiae* does not have a specific transporter. The transport takes place by means of the hexose transporters. The affinity of the transporters to D-xylose is, however, distinctly lower than to D-glucose (Kotter and Ciriacy, 1993). In yeasts which are able to metabolize D-xylose, such as for example *P. stipitis*, *C. shehatae* or *P. tannophilus* (Du Preez et al., 1986), there are both unspecific low-affinity transporters, which transport D-glucose, and also specific high-affinity proton symporters only for D-xylose (Hahn-Hagerdal et al., 2001).

In earlier experiments, some yeasts were found, such as for example *Candida tropicalis*, *Pachysolen tannophilus*, *Pichia stipitis*, *Candida shehatae*, which by nature ferment D-xylose or L-arabinose or can at least assimilate it. However, these yeasts lack entirely the capability of fermenting L-arabinose and D-xylose to ethanol, or they only have a very low ethanol yield (Dien et al., 1996). Moreover, very little is yet known about the uptake of D-xylose and L-arabinose. In the yeast *C. shehatae* one assumes a proton symport (Lucas and Uden, 1986). In *S. cerevisiae*, it is known from the galactose permease Gal2 that it can transport D-xylose but also transports L-arabinose, which is very similar in structure to D-galactose. (Kou et al., 1970). Most hexose transporters can mediate uptake of D-xylose.

Alcoholic fermentation of pentoses in biotechnologically modified yeast strains of *S.cerevisiae*, wherein inter alia various genes of the yeast strain *Pichia stipitis* were used for the genetic modification of *S.cerevisiae*, was described in recent years particularly in connection with the fermentation of xylose. The engineering concentrated here particularly on the introduction of the genes for the initial xylose assimilation from *Pichia stipitis*, a xylose-fermenting yeast, into *S.cerevisiae*, i.e. into a yeast which is traditionally used in the ethanol production from hexose (Jin et al. 2004).

Jeppson et al. (2006) describe xylose fermentation by *S.cerevisiae* by means of the introduction of a xylose metabolic pathway which is either similar to that in the yeasts *Pichia stipitis* and *Candida shehatae*, which naturally use xylose, or is similar to the bacterial metabolic pathway.

Katahira et al. (2006) describe sulphuric acid hydrolysates of lignocellulose biomass such as wood chips, as an important material for the production of fuel bioethanol. In this study, a recombinant yeast strain was constructed, which is able to ferment xylose and celooligosaccharides. For this, various genes were integrated into this yeast strain and namely for the inter-cellular expression of xylose reductase and xylitol dehydrogenase from *Pichia stipitis* and xylulokinase from *S.cerevisiae* and for the presentation of beta-glucosidase from *Aspergillus aculeatus* on the cell surface. In the fermentation of sulphuric acid hydrolysates of wood chips, xylose and celooligosaccharides were fully fermented by the recombinant strain after 36 hours.

Pitkanen et al. (2005) describe the obtaining and characterizing of xylose chemostat isolates of a *S.cervisiae* strain, which over-expresses genes of *Pichia stipitis* coding for xylose reductase and xylitol dehydrogenase and the gene which codes endogenous xylulokinase. The isolates were obtained from aerobic chemostat cultures on xylose as the single or major carbon source. Under aerobic conditions on minimal medium with 30 g/l xylose, the growth rate of the chemostat isolates was 3 times higher than that of the original strain ( $0.15\text{ h}^{-1}$  compared with  $0.05\text{ h}^{-1}$ ). The xylose uptake rate was increased almost two-fold. The activities of the key enzymes of the pentose phosphate metabolic pathway (transketolase, transaldolase) were increased two-fold, whilst the concentrations of their substrates (pentose-5-phosphates, sedoheptulose-7-phosphate) were lowered accordingly.

Brat et al. (2009) screened nucleic acid databases for sequences encoding putative xylose isomerases and finally were able to clone and successfully express a highly active new kind of xylose isomerase from the anaerobic bacterium *Clostridium phytofermentans* in *S. cerevisiae*. Heterologous expression of this enzyme confers on the yeast cells the ability to metabolize D-xylose and to use it as the sole carbon and energy source.

Demeke et al. (2013) developed an expression cassette containing 13 genes including *C. phytofermentans xylA*, encoding D-xylose isomerase, and enzymes of the pentose phosphate

pathway and inserted the cassette in two copies in the genome of the industrial *S. cerevisiae* strain Ethanol Red. Subsequent EMS mutagenesis, genome shuffling and selection in D-xylose-enriched lignocellulose hydrolysate, followed by multiple rounds of evolutionary engineering in complex medium with D-xylose, gradually established highly efficient D-xylose fermentation.

Becker and Boles (2003) describe the engineering and the selection of a laboratory strain of *S. cerevisiae* which is able to use L-arabinose for growth and for fermenting it to ethanol. This was possible due to the over-expression of a bacterial L-arabinose metabolic pathway, consisting of *Bacillus subtilis* AraA and *Escherichia coli* AraB and AraD and simultaneous over-expression of yeast galactose permease transporting L-arabinose in the yeast strain. Molecular analysis of the selected strain showed that the predetermining precondition for a use of L-arabinose is a lower activity of L-ribulokinase. However, inter alia, a very slow growth is reported from this yeast strain.

Wiedemann and Boles (2008) show that expressing of the codon-optimized genes of L-arabinose isomerase from *Bacillus licheniformis* and L-ribulokinase and L-ribulose-5-P 4-epimerase from *Escherichia coli* strongly improved L-arabinose conversion rates.

Farwick et al. (2014) developed a new system for screening and engineering of pentose transporters which are no longer inhibited by glucose. This system was based on a D-xylose-fermenting yeast strain having deletions of all hexose-transporters and all hexo-/glucokinases (*hxt<sup>0</sup> hxk<sup>0</sup>* strain). D-glucose can no longer be used as a carbon source but interferes with D-xylose utilization at transport level. As a result, mutant transporters that allow D-xylose uptake in the presence of increasing concentrations of D-glucose could easily be selected. Using this system in evolutionary engineering and mutagenesis approaches the authors were able to generate specific D-xylose transporters from *S. cerevisiae* hexose transporters. Some of these mutant transporters had an exchange at a position corresponding to N376 of the galactose transporter Gal2. However, although they proved resistant against glucose most of them had a reduced uptake rate for xylose.

WO 2008/080505 A1 discloses an arabinose transporter from *Pichia stipitis*, which enables yeast cells to take up L-arabinose. EP 11 001 841.3 discloses a specific arabinose transporter of the plant *Arabidopsis thaliana* for the construction of pentose-fermenting yeasts.

WO 2012/049170 A2 and WO 2012/049173 A1 disclose pentose and glucose fermenting yeast cells which contain and express among other nucleic acids, a polypeptide with arabinose permease activity comprising a mutation in position T219 to asparagine or N376 to serine of Gal2 which renders the transporter resistant against the inhibitory effect of glucose.

There still exists a need in the art for specific pentose transporters, in particular specific D-xylose transporters, which have a higher affinity and/or higher activity for pentoses, in particular combined with glucose resistance, which allow to specifically take up D-xylose and/or L-arabinose into cells, such as yeast cells, with high uptake rates even at low pentose concentrations, and therefore to promote the utilization and fermentation of pentoses, in particular D-xylose and/or L-arabinose, and in particular in the simultaneous presence of glucose.

It is thus an object of the present invention to provide improved and/or more specific transporters, which transport pentose(s), such as D-xylose and/or L-arabinose with higher activities and/or higher affinities.

#### SUMMARY OF THE INVENTION

According to the present invention this object is solved by a polypeptide, comprising at least one amino acid substitution at a position corresponding to T354 of the amino acid sequence of SEQ ID NO: 1,

wherein the polypeptide has at least 60%, or preferably at least 70% or 80% or 90% or 95% sequence identity with the amino acid sequence of SEQ ID NO: 1, and wherein the polypeptide has an *in vitro* and/or *in vivo* pentose transport function.

According to the present invention this object is solved by a nucleic acid molecule coding for a polypeptide of the present invention.

According to the present invention this object is solved by a host cell, containing a nucleic acid molecule of the present invention and preferably expressing said nucleic acid molecule, wherein said host cell is preferably a fungus cell and more preferably a yeast cell, such as *Saccharomyces* species, *Kluyveromyces* sp., *Hansenula* sp., *Pichia* sp. or *Yarrowia* sp.

According to the present invention this object is solved by a method for the production of bioethanol and/or other bio-based compounds, comprising the expression of a nucleic acid molecule according to the present invention, preferably in a host cell according to the present invention.

According to the present invention this object is solved by using a polypeptide according to the present invention, a nucleic acid molecule according to the present invention, or a host cell according to the present invention for the production of bioethanol and/or other bio-based compounds,  
and/or for the recombinant fermentation of biomaterial containing pentose(s), preferably D-xylose and/or L-arabinose.

#### DESCRIPTION OF THE PREFERRED EMBODIMENTS OF THE INVENTION

Before the present invention is described in more detail below, it is to be understood that this invention is not limited to the particular methodology, protocols and reagents described herein as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention which will be limited only by the appended claims. Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art.

##### *Gal2 variants*

As discussed above, the present invention provides Gal2 variants.

In particular, the present invention provides a polypeptide, comprising at least one amino acid substitution at a position corresponding to T354 of the amino acid sequence of SEQ ID NO: 1.

The polypeptide of the present invention has at least 60%, or preferably at least 70% or 80% or 90% or 95% sequence identity with the amino acid sequence of SEQ ID NO: 1, and has an *in vitro* and/or *in vivo* pentose transport function.

Preferably, the polypeptide of the present invention is Gal2 of *Saccharomyces cerevisiae*.

SEQ ID NO: 1 is the wild-type protein or amino acid sequence of Gal2 of CEN.PK2-1C and CEN.PK113-7D.

Gal2 of CEN.PK2-1C is a protein of 574 amino acids.

The polypeptides, preferably the Gal2 variants, according to the invention comprise at least one amino acid substitution at a position corresponding to T354 of the amino acid sequence of SEQ ID NO: 1 or of an amino acid sequence, which is at least 60% identical, preferably at least 70% identical, more preferably at least 80% identical, even more preferably at least 90% identical, yet more preferably 95% identical, and yet more preferably 99% identical to the amino acid sequence of SEQ ID NO: 1.

As used herein, the term "at a position corresponding to" means the respective position in SEQ ID No: 1 which, however, in related polypeptide chains can have another relative position number. The equivalent substitution can be determined by comparing a position in both sequences, which may be aligned for the purpose of comparison. The relative position of the amino acid can vary due to different length of the related polypeptide, or deletions or additions of amino acids in the related polypeptide.

The polypeptides, preferably the Gal2 variants, according to the invention have an *in vitro* and/or *in vivo* pentose transport function, in particular an *in vitro* and/or *in vivo* D-xylose and/or L-arabinose transport function.

Preferably, the pentose is D-xylose and/or L-arabinose.

As used herein, if not otherwise indicated, the term "xylose" means the same as D-xylose, "arabinose" means the same as L-arabinose, and "glucose" means the same as D-glucose.

As used herein, the term "percent (%) identical" refers to sequence identity between two amino acid sequences. Identity can be determined by comparing a position in both sequences, which may be aligned for the purpose of comparison. When an equivalent position in the compared sequences is occupied by the same amino acid, the molecules are considered to be identical at that position.

As used herein, the term “functional equivalent” refers to amino acid sequences that are not 100% identical to the amino acid sequence of SEQ ID NO. 1 and comprise amino acid additions and/or insertions and/or deletions and/or substitutions and/or exchanges, which do not alter or change the activity or function of the protein as compared to the protein having the amino acid sequence of SEQ ID NO: 1, i.e. an “functional equivalent”, for example, encompasses an amino acid sequence with conservative amino acid substitutions or smaller deletions and/or insertions as long as these modifications do not substantially affect the *in vitro* and/or *in vivo* L-arabinose transport function.

Generally, a person skilled in the art is aware of the fact that some amino acid exchanges in the amino acid sequence of a protein do not have an influence on the (secondary or tertiary) structure, function and/or activity of that protein. Amino acid sequences with such “neutral” amino acid exchanges as compared to the amino acid sequences disclosed herein fall within the scope of the present invention.

In a preferred embodiment, the polypeptides, preferably the Gal2 variants, according to the present invention, comprising at least one amino acid substitution at a position corresponding to T354 of SEQ ID NO. 1, comprise the amino acid substitution T354A.

Preferably, the amino acid substitution at a position corresponding to T354 increase(s) the activity of the *in vitro* and/or *in vivo* pentose transport function compared to the polypeptide without such amino acid substitution(s).

Preferably, the amino acid substitution at a position corresponding to T354 increase(s) the affinity of the polypeptide for pentose(s) compared to the polypeptide without such amino acid substitution(s).

*- further amino acid substitution(s)*

In one embodiment, the polypeptides, preferably the Gal2 variants, according to the present invention comprise further amino acid substitution(s):

preferably amino acid substitution(s) at a position corresponding to V71 of the amino acid sequence of SEQ ID NO: 1.

Said further amino acid substitution is preferably V71I.

The present invention preferably provides the following polypeptides / Gal2 variants:

- T354A
- T354A / V71I

*Nucleic acid molecules*

As discussed above, the present invention provides a nucleic acid molecule, coding for a polypeptide according to the present invention.

In one embodiment, the nucleic acid molecule of the present invention further comprises:

- vector nucleic acid sequences, preferably expression vector sequences,  
and/or
- promoter nucleic acid sequences and terminator nucleic acid sequences,  
and/or
- comprises other regulatory nucleic acid sequence.

In one embodiment, the nucleic acid molecule of the present invention comprises dsDNA, ssDNA, PNA, CNA, RNA or mRNA or combinations thereof.

The nucleic acid molecules according to the invention preferably comprise nucleic acid sequences, which are (except for the addition of the amino acid substitution(s) according to the invention) identical with the naturally occurring nucleic acid sequence or are codon-optimized for the use in a host cell.

The nucleic acid molecule used according to the present invention is preferably a nucleic acid expression construct.

Nucleic acid expression constructs according to the invention are expression cassettes comprising a nucleic acid molecule according to the invention, or expression vectors

comprising a nucleic acid molecule according to the invention or an expression cassette, for example.

A nucleic acid expression construct preferably comprises regulatory sequences, such as promoter and terminator sequences, which are operatively linked with the nucleic acid sequence coding for the polypeptide(s) of the invention.

The nucleic acid expression construct may further comprise 5' and/or 3' recognition sequences and/or selection markers.

#### *Host cells*

As discussed above, the present invention provides host cells containing a nucleic acid molecule according to the present invention.

Preferably, the host cells of the present invention express said nucleic acid molecule.

Preferably, a host cell according to the present invention is a fungus cell and more preferably a yeast cell.

The yeast cell is preferably a member of a genus selected from the group of *Saccharomyces* species, *Kluyveromyces* sp., *Hansenula* sp., *Pichia* sp. or *Yarrowia* sp.

The yeast cell is more preferably a member of a species selected from the group of *S. cerevisiae*, *S. bulderi*, *S. barnetti*, *S. exiguus*, *S. uvarum*, *S. diastaticus*, *K. lactis*, *K. marxianus*, *K. fragilis*, *H. polymorpha*, *P. pastoris* and *Y. lipolytica*, such as *S. cerevisiae*, *K. lactis*, *H. polymorpha*, *P. pastoris* or *Y. lipolytica*.

In a preferred embodiment, the host cell belongs to the species *Saccharomyces cerevisiae*.

When the nucleic acid molecule/sequence coding for the polypeptide (preferably Gal2 variant(s)) of the present invention is expressed in a host cell (preferably a yeast cell), the host cell is imparted the capability to take up D-xylose and/or L-arabinose, which then may be metabolized further. Through this, the cell is able to grow on D-xylose and/or L-arabinose as a carbon source.

Preferably, the host cell (preferably yeast cell) has an increased uptake rate for D-xylose and/or L-arabinose compared to a cell not containing a nucleic acid molecule according to the present invention.

In a preferred embodiment, the host cell (preferably yeast cell) of the present invention further contains

- nucleic acid molecules which code for proteins of a xylose metabolic pathway (preferably xylose isomerase and xylulokinase),

and/or

- nucleic acid molecules which code for proteins of an arabinose metabolic pathway (preferably arabinose isomerase, ribulokinase, ribulose-5-P 4-epimerase).

Such a host cell has preferably

- an increased D-xylose and/or L-arabinose utilization rate

and/or

- a faster growth rate with D-xylose and/or L-arabinose compared to a cell not containing a nucleic acid molecule according to the present invention.

For example, the host cell (preferably yeast cell) of the present invention can further contain nucleic acid molecules which code for proteins of an arabinose metabolic pathway, in particular for arabinose isomerase, ribulokinase, ribulose-5-P 4-epimerase.

Preferred are proteins of the bacterial arabinose metabolic pathway, in particular *E.coli* araB L-ribulokinase, *E.coli* araD L-ribulose-5-P 4-epimerase and *B.licheniformis* araA L-arabinose-isomerase.

In a preferred embodiment, a host cell (preferably yeast cell) according to this invention is modified by the introduction and expression of the genes *araA* (L-arabinose-isomerase), *araB* (L-ribulokinase) and *araD* (L-ribulose-5-P-4-epimerase) and in addition over-expresses a *TAL1* (transaldolase) gene, as described for example by the inventors in EP 1 499 708 B1, and in addition to this contains at least one nucleic acid molecule according to the invention.

Depending on the intended use of the yeast cell, said yeast cell can contain, express or overexpress further nucleic acid sequences coding for further proteins, such as transaldolase TAL1 and/or TAL2, transketolase TKL1 and/or TKL2, D-ribulose-5-phosphate 3-epimerase RPE1, ribose-5-phosphate ketol-isomerase RKI1 or the corresponding sequences from other organisms encoding the same enzyme activities.

For example, the host cell (preferably yeast cell) of the present invention can further overexpress nucleic acid molecules which code for proteins of a xylose metabolic pathway, in particular for xylose isomerase and xylulokinase.

Preferred is *Clostridium phytofermentans* or *Piromyces xylA* xylose isomerase and *S. cerevisiae* XKS1 xylulokinase.

In a preferred embodiment, a host cell (preferably yeast cell) according to this invention is modified by the introduction and/or overexpression of the genes *xylA* (xyloseisomerase), *XKS1* (xylulokinase) and in addition over-expresses a *TAL1* (transaldolase) gene.

Depending on the intended use of the yeast cell, said yeast cell can contain, express or overexpress further nucleic acid sequences coding for further proteins, such as transaldolase TAL1 and/or TAL2, transketolase TKL1 and/or TKL2, D-ribulose-5-phosphate 3-epimerase RPE1, ribose-5-phosphate ketol-isomerase RKI1 or the corresponding sequences from other organisms encoding the same enzyme activities.

#### *Methods and uses for producing bioethanol*

As discussed above, the present invention provides a method for the production of bioethanol and/or other bio-based compounds.

Said method comprises the expression of a nucleic acid molecule according to the present invention, preferably in a host cell according to the present invention.

As discussed above, the present invention provides the use of

- a polypeptide according to the present invention,
- a nucleic acid molecule according to the present invention, or

- a host cell according to the present invention,  
for the production of bioethanol and/or other bio-based compounds,  
and/or for the recombinant fermentation of biomaterial containing pentose(s), preferably D-xylose and/or L-arabinose.

The term “bio-based compounds” or “other bio-based compounds” as used herein refers to chemical compounds and substances, which are obtained from biological materials and raw materials (biomass), particularly by using microorganisms.

The (other) bio-based compounds can be compounds, which are selected from, but not limited to:

lactic acid, acetic acid, succinic acid, malic acid or other organic acids,  
1-butanol, isobutanol, 2-butanol, other alcohols,  
amino acids, alkanes, terpenes, isoprenoids, solvents, pharmaceutical compounds,  
vitamins.

#### *Further description of preferred embodiments*

The inventors have identified Gal2 variants which exhibit

- increased activity of the *in vitro* and/or *in vivo* pentose transport function and/or
- increased affinity for pentose(s)

compared to the wildtype or to a Gal2 polypeptide without the respective amino acid substitution(s).

The inventors have, thus, identified Gal2 variants which, thus, confer a host cell (preferably a yeast cell) the capability to take up D-xylose and/or D-arabinose and, preferably, the capability for an increased uptake for pentose(s), preferably D-xylose and/or D-arabinose.

For this, reference is also to be made to the examples and figures.

#### *- Uptake of L-arabinose and D-xylose*

So that pentose(s), in particular, D-xylose and/or L-arabinose, can be metabolized by *S.cerevisiae*, they must firstly be taken up by the cell.

All hexose transporters tested for the pentose D-xylose have a much higher affinity to hexoses than to D-xylose. For L-arabinose, a similar situation is assumed. Of all strains constructed

hitherto, which can utilize pentoses (D-xylose or L-arabinose), a relatively slow growth is reported. Above all, the slow and poor uptake of the pentoses is named as a reason for this (Becker and Boles, 2003; Richard *et al.*, 2002). In fermentations in a sugar mixture, consisting of D-glucose and D-xylose or D-glucose and L-arabinose, the sugars are not converted simultaneously. Due to the high affinity of the transporters for D-glucose, D-glucose is metabolized at first. A so-called diauxic shift occurs. Only after the D-glucose is exhausted is the pentose converted in a second, distinctly slower growth phase (Kuyper *et al.*, 2005a; Kuyper *et al.*, 2005b). The absence of specific transporters for pentoses is given as an explanation.

#### - Gal2

The hexose galactose is transported by the high affinity transporter Gal2 ( $K_m = 1$  to  $5$  mM) which is equally affine for glucose ( $K_m = 1.5$  to  $1.9$ ) (see e.g. Reifenberger *et al.*, 1997). Like the other structural genes needed for galactose utilization (*GAL1*, galactose kinase; *GAL10*, mutarotase / UDP-glucose-4-epimerase; *GAL7*, galactose-1-phosphate uridyl transferase), its expression is repressed in the presence of glucose and also needs the induction by galactose. Gal2 is also target of catabolite inactivation (Horak and Wolf, 1997). Gal2 is one of only few transporters which can transport L-arabinose. Like most of the other hexose transporters it can transport D-xylose. However, its affinity for L-arabinose and D-xylose is quite low. Therefore, at low D-xylose or L-arabinose concentrations the uptake activity is very low.

Farwick *et al.* (2014) developed mutant transporters that allow D-xylose uptake in the presence of D-glucose. Some of these mutant transporters had an amino acid exchange at a position corresponding to N376 of the galactose transporter Gal2. However, although they proved resistant against glucose most of them had a reduced uptake rate for xylose.

The polypeptides of the present invention exhibit an increased uptake rate and/or affinity for xylose, in particular in combination with mutations which make them resistant against inhibition by glucose.

Various mutagenesis methods had to be used as well as evolutionary engineering under very specific conditions with engineered yeast strains with a variety of different modifications, to find improved Gal2 derived pentose transporters. These transporters had to be tested in elaborate screening systems, with growth tests and sugar uptake assays. Finally, the mutant

transporters had to be sequenced and from a variety of mutations those had to be elucidated which finally were responsible for the improved properties.

D-xylose accounts for up to 35% of total sugars in xylan-rich lignocellulosic biomass such as hard woods and straw (see Demeke et al. 2013). Biomass with significant amounts of arabinose (source of the data: U.S. Department of Energy):

| <u>Type of biomass</u> | <u>L-arabinose [%]</u> |
|------------------------|------------------------|
| Switchgrass            | 3.66                   |
| Large bothriochloa     | 3.55                   |
| Tall fescue            | 3.19                   |
| Robinia                | 3                      |
| Corn stover            | 2.69                   |
| Wheat straw            | 2.35                   |
| Sugar can bagasse      | 2.06                   |
| Chinese lespedeza      | 1.75                   |
| Sorghum bicolor        | 1.65                   |

The Gal2 variants according to the invention are also of great importance for its utilization.

Possibilities for use of a functional and at the same time specific pentose transporter in the yeast *S.cerevisiae* are firstly the production from lignocellulosic hydrolysates of bioethanol and the production of high-grade precursor products for further chemical syntheses, particularly when pentose concentrations are low and in the simultaneous presence of glucose.

The following list originates from the study “Top Value Added Chemicals From Biomass”.

Here, 30 chemicals were categorized as being particularly valuable, which can be produced from biomass.

| Number of C atoms | Top 30 Candidates                                                                                    |
|-------------------|------------------------------------------------------------------------------------------------------|
| 1                 | hydrogen, carbon monoxide                                                                            |
| 2                 |                                                                                                      |
| 3                 | glycerol, 3-hydroxypropionic acid, lactic acid, malonic acid, propionic acid, serine                 |
| 4                 | acetoin, asparaginic acid, fumaric acid, 3-hydroxybutyrolactone, malic acid, succinic acid, threonin |
| 5                 | arabitol, furfural, glutamic acid, itaconic acid, levulinic acid, proline, xylitol, xylonic acid     |
| 6                 | aconitic acid, citrate, 2,5-furandicarboxylic acid, glucaric acid, lysine, levoglucosan, sorbitol    |

As soon as these chemicals are produced from lignocelluloses by bioconversion (e.g. fermentations with yeasts), it is important to have specific, highly active transporter(s) for the hemicellulose sugars arabinose and xylose.

The following examples and drawings illustrate the present invention without, however, limiting the same thereto.

#### BRIEF DESCRIPTION OF THE DRAWINGS

##### **Figure 1: Growth test of Gal2\_T354A in EBY.VW4000**

The transformants were cultivated in 5 ml SCM-ura with 20 g/l maltose at 30°, washed with water and adjusted to an OD<sub>600</sub> of 1. Thereof it followed a serial dilution. 5 µl were dropped onto the respective media and it was incubated for three days at 30°. As a control for the dilution SCM-ura with 20 g/l maltose was taken. The cells of the mutated transporters were dropped onto SCD-ura with 0.2% and 2% glucose, as well as SCG-ura with 0.2% and 2% D-galactose to test their functionality. Furthermore the wild type of CEN.PK2 and Ethanol Red, as well as Gal2\_ep3.1 were used for comparison. The galactose transporter with the mutation T354A is from HDY.GUF10.

##### **Figure 2: Growth test of Gal2\_T354A in AFY10**

The transformants were cultivated in 5 ml SCE-ura-leu with 2% ethanol at 30°, washed with water and adjusted to an OD<sub>600</sub> of 1. Thereof it followed a serial dilution. 5 µl were dropped onto the respective media and it was incubated for five days at 30°. As a control for the dilution SCE-ura-leu with 2% ethanol was taken. The cells of the mutated transporters were dropped onto SCX-ura-leu with 0.2% and 2% D-xylose, as well as SC-ura-leu with 1% D-xylose with 4% D-glucose and SC-ura-leu with 0.2% D-xylose and 2% D-glucose to test their functionality. Furthermore the wild type of CEN.PK2 and Ethanol Red, as well as Gal2\_ep3.1 were used for comparison. The galactose transporter with the mutation T354A is from HDY.GUF10.

**Figure 3: Drop test of Gal2\_T354A in combination with V71I and L280R in EBY.VW4000 after four days at 30°.** Several dilutions were dropped from left to right (undiluted, 1:10, 1:100, 1:1000).

**Figure 4: Drop test of Gal2\_T354A in combination with V71I and L280R in AFY10 after six days at 30°.** Several dilutions were dropped from left to right (undiluted, 1:10, 1:100, 1:1000).

## EXAMPLES

### Methods

#### **Strains and media**

- Bacteria

*E.coli* SURE (Stratagene)

Full medium LB 1% trypton, 0.5% yeast extract, 0.5% NaCl, pH 7.5 (see Sambrook and Russell, 2001)

For selection on a plasmid-coded antibiotic resistance, 40 µg/ml ampicillin was added to the medium after autoclaving. Solid culture media additionally contained 1.9% agar. The culture took place at 37°C.

- Yeast

CEN.PK2-1C

*MATa leu2-3,112 ura3-52 trp1-289 his3-ΔI MAL2-8<sup>c</sup> SUC2* (EUROSCARF, Frankfurt)

#### Strain EBY.VW4000

EBY.VW4000 (Genotype: *MATa leu2-3,112 ura3-52 trp1-289 his3-ΔI MAL2-8<sup>c</sup> SUC2 Δhxt1-17 Δgal2 stlΔ::loxP agt1Δ::loxP mph2Δ::loxP mph3Δ::loxP*) (Wieczorke *et al.*, 1999)

#### Strain Ethanol Red

available from Lesaffre, Lille, France. Described in Demecke *et al.* (2013).

#### Strain HDY.GUF10

Xylose and arabinose consuming industrial *S. cerevisiae* strain derived from Ethanol Red (Dietz 2013).

#### Strain AFY10

EBY.VW4000 *glk1Δ::loxP hxl1Δ::loxP hxl2Δ::loxP ylr446wΔ::loxP pyk2Δ::pPGK1-opt.XKS1-tPGK1 pTPII-TAL1-tTAL1 pTDH3-TKL1-tTKL1 pPFK1-RPE1-tRPE1 pFBA-RKII-tRKII loxP* (Farwick *et al.*, 2014).

#### Strain AFY10X

AFY10 + YE<sub>p</sub>-kanR\_optXI (Farwick *et al.*, 2014).

- synthetic complete selective medium SC

0.67% yeast nitrogen base w/o amino acids and ammonium sulphate, 0.5% ammonium sulphate, 20mM potassium dihydrogenphosphate, pH 6.3, amino acid/nucleobase solution without the corresponding amino acids for the auxotrophy markers of the plasmids used, carbon source in the respectively indicated concentration

Concentration of the amino acids and nucleobases in the synthetic complete medium (Zimmermann, 1975): adenine (0.08 mM), arginine (0.22 mM), histidine (0.25 mM), isoleucine (0.44 mM), leucine (0.44 mM), lysin (0.35 mM), methionine (0.26 mM), phenylalanine (0.29 mM), threonine (0.48 mM), tryptophan (0.19 mM), tyrosin (0.34 mM), uracil (0.44 mM) and valine (0.49 mM). As carbon sources, L-arabinose, D-glucose, D-galactose, D-mannose, ethanol and maltose were used, as indicated.

Solid full and selective media contained in addition 1.9% agar. The culture of the yeast cells took place at 30°C.

### Plasmids

| Plasmid                 | Referenz               |
|-------------------------|------------------------|
| p426H7                  | Becker and Boles, 2003 |
| YEp181_pHXT7-optXI_Clos | Subtil and Boles, 2012 |
| p426H7_GAL2_WT          |                        |
| p426H7_GAL2_ep3.1       |                        |
| p426H7-GAL2_EtOHred     |                        |
| p426H7-GAL2_GUF10       |                        |
| p426H7_GAL2-T354A       |                        |
| p426H7_GAL2-V71I        |                        |

### Preparation of DNA

#### - Isolation of plasmid DNA from *E. coli*

Small-scale preparations of plasmid DNA from *E. coli* cultures were done using the GeneJET™ Plasmid Miniprep Kit (Fisher Scientific) according to manufacturer's instructions. The QUIAGEN Plasmid Maxi Kit was used for large scale preparations.

#### - Isolation of plasmid and genomic DNA from *S. cerevisiae*

For isolation of genomic and plasmid DNA from yeast cells 5 – 10 ml of stationary phase cultures were harvested by centrifugation (1 min, 2000x g) and washed once in 1 ml sterile ddH<sub>2</sub>O. The cell pellet was resuspended in 400 µl YP-buffer 1 by vortexing and then lysed by addition of 400 µl YP-buffer 2, 1/3 to 2/3 Vol glass beads (Ø 0.25-0.5 mm) and 8 minutes of shaking on a VXR basic Vibrax™ (IKA) at 2000 rpm. Cell debris was pelleted by centrifugation (30 sec, 16000x g) and 650 µl of the supernatant transferred to a fresh eppendorf tube. 325 µl of cold YP-buffer 3 were added and the sample was vortexed and then incubated on ice for 10 minutes for precipitation of proteins and other contaminants. The sample was centrifuged (10-15 min, 4°C, 16000x g), 700 µl of the supernatant were transferred to a fresh eppendorf tube and 700 µl isopropanol were added. After mixing vigorously, the sample was incubated for 10 minutes at RT to allow precipitation of the DNA, which was then pelleted by centrifugation

( $\geq 30$  min, RT, 16000x g). The DNA-pellet was washed twice with 500  $\mu$ l of cold ( $-20^{\circ}\text{C}$ ) 70 % (v/v) ethanol, with centrifugation steps of 5 minutes at RT and 16000x g, then dried at RT for 10 minutes and dissolved in 15 – 30  $\mu$ l sterile ddH<sub>2</sub>O depending on the size of the DNA pellet.

*- Determining the DNA concentration*

The DNA concentration is measured by spectral photometry in a wavelength range of 240-300 nm. If the purity of the DNA, determined by the quotient  $E_{260\text{nm}}/E_{280\text{nm}}$  is 1.8, then the extinction  $E_{260\text{nm}} = 1.0$  corresponds to a DNA concentration of 50  $\mu\text{g dsDNA/ml}$  (Sambrook and Russell, 2001).

*- DNA purification of PCR products*

The purification of the PCR products took place with the “QIAquick PCR Purification Kit” of the company Qiagen, according to the manufacturer’s information.

**Digestion of DNA with restriction endonucleases (restriction digestion)**

For the site-specific cleavage of DNA restriction endonucleases from New England Biolabs (NEB) or Fermentas were used with the provided buffers and according to the instructions of the manufacturer. Usually 1-3 units per  $\mu\text{g}$  DNA were used for the reaction, which was incubated for 2-12 hours. This method was used to prepare vectors for recombinational cloning, to confirm correctly assembled plasmids or to specifically degrade a certain plasmid from a mixture.

**Polymerase chain reaction (PCR)**

Different polymerases were used for different PCR experiments in this work. For confirmation of genomic gene deletion or integration the Crimson Taq polymerase (NEB) was used. For amplification of ORFs (for sequencing), genes for recombinational cloning or amplification of integrative cassettes for genomic gene deletion or integration the Phusion or Q5 polymerases (NEB) were used. Composition of PCR reactions and the corresponding PCR program are displayed in the Tables below. Annealing temperatures of primer pairs were calculated with the  $T_m$  calculator tool on the NEB homepage. All PCRs were performed in a Mastercycler gradient (Eppendorf), Piko Thermo Cycler (Finnzymes) or Progene PCR cycler (Techne).

**Composition of PCR reactions with Crimson Taq polymerase**

| component                             | 15 µl reaction | 25 µl reaction | final concentration |
|---------------------------------------|----------------|----------------|---------------------|
| 5x Crimson <i>Taq</i> reaction buffer | 3 µl           | 5 µl           | 1x                  |
| 2 mM dNTP mix                         | 1.5 µl         | 2.5 µl         | 200 µM each         |
| 10 µM primer (each)                   | 0.3            | 0.5 µl         | 0.2 µM each         |
| template DNA                          | variable       | variable       | variable            |
| Crimson <i>Taq</i> DNA polymerase     | 0.15 µl        | 0.25 µl        | 0.025 U/µl          |
| nuclease-free water                   | to 15 µl       | to 25 µl       |                     |

**PCR program for reactions with Crimson Taq polymerase**

| step                 | temperature (°C) | time     |
|----------------------|------------------|----------|
| initial denaturation | 95               | 1 min    |
|                      | 95               | 22 sec   |
| 30 - 35 cycles       | 45-68            | 35 sec   |
|                      | 68               | 1 min/kb |
| final extension      | 68               | 5 min    |
| hold                 | 4-10             |          |

**Composition of PCR reactions with Phusion or Q5 polymerase**

| component                                 | 25 µl reaction | 50 µl reaction | final conc. |
|-------------------------------------------|----------------|----------------|-------------|
| 5x Phusion HF / Q5 reaction buffer        | 5 µl           | 10 µl          | 1x          |
| 2 mM dNTPs mix                            | 2.5 µl         | 5 µl           | 200 µM each |
| 10 µM primer (each)                       | 0.5 µl         | 1 µl           | 200 µM each |
| template DNA                              | variable       | variable       | variable    |
| Phusion / Q5 High-Fidelity DNA polymerase | 0.25 µl        | 0.5 µl         | 0.02 U/µl   |

|                     |               |               |
|---------------------|---------------|---------------|
| nuclease-free water | to 25 $\mu$ l | to 50 $\mu$ l |
|---------------------|---------------|---------------|

### PCR program for reactions with Phusion or Q5 polymerase

| step                 | temperature (°C) | time                     |
|----------------------|------------------|--------------------------|
| initial denaturation | 98°C             | 1 min                    |
| 15–35 cycles         | 98°C             | 10 sec                   |
|                      | 50–72°C          | 20 sec                   |
|                      | 72°C             | 15 sec/kb (for plasmids) |
|                      |                  | 30 sec/kb (for gDNA)     |
| final extension      | 72°C             | 5 min                    |
| hold                 | 4–10°C           |                          |

### Fusion PCR

A fusion PCR was used for construction the ORF of HXT7-N370F for recombinational cloning. In the first step two overlapping fragments of HXT7 were amplified in two separate Q5 PCR reactions with p426H7\_HXT7 as a template. The PCR reactions were separated in a 1.5 % agarose gel and the correct fragments purified from the corresponding gel pieces. Equal molar amount of both fragments (20 ng minimum) were used in a Q5 PCR reaction without primers. This PCR reaction was run for 6 cycles, before 1  $\mu$ l of forward and reverse primer (from 10  $\mu$ M stocks) were added. The reaction was then run for another 20 cycles.

### error-prone PCR (epPCR)

For generation of random mutagenized ORFs of GAL2 the GeneMorph II Random Mutagenesis Kit (Agilent Technologies<sup>TM</sup>) was applied. The manufacturer's protocol has been followed. The PCR reaction has been run for 33 cycles. The amount of template DNA has been varied to achieve different mutation rates (see Table below) The analysis of epPCR-products revealed that the desired amplification specifications have been met. The PCR fragments were purified and used as templates for a Phusion PCR reaction to extend the fragments' ends with homologous overhangs.

**Amount of template DNA used in different epPCRs and**

| mutation rate         | amount of template (acceptable range) | amplification |       |
|-----------------------|---------------------------------------|---------------|-------|
|                       |                                       | desired       | found |
| medium (4.5–9 mut/kb) | 205 ng (100-500 ng)                   | 10-100        | ≈ 45  |
| high (9–16 mut/kb)    | 23 ng (0.1-100 ng)                    | 100-10000     | ≈ 250 |

**Agarose gel-electrophoresis for DNA or RNA separation**

Fragments in DNA or RNA samples were separated by size using agarose gels with concentrations ranging from 0.7 to 2.0 % (w/v) agarose (Sambrook and Russell, 2001). 1x TAE-buffer was used for preparation of gels and as running buffer. The GeneRuler 1 kb DNA Ladder (Fisher Scientific) was used for sizing of the DNA fragments. DNA samples were mixed with 1/5 Vol of 6x DNA loading dye before loading onto the gel. RNA samples were mixed with the same volume 2x RNA loading dye, incubated at 96°C for 10 min and stored on ice prior to loading. Gels were run at up to 6 - 10 V/cm for 30 to 45 minutes depending on current, gel percentage and expected fragment sizes. DNA and RNA were visualized by UV-light (254 nm) after incubation of the gel in an ethidium bromide bath.

**DNA-purification and DNA-extraction from agarose gels**

To purify DNA (e.g. from PCR reactions or after restriction digestions) and to extract DNA from agarose gels the NucleoSpin® Extract II-Kit (Macherey-Nagel) was used according to manufacturer's instructions.

**DNA sequencing**

Sequencing of DNA samples was done by GATC Biotech AG (Konstanz, Germany). The samples contained 30-100 ng/μl (plasmids) or 10-50 ng/μl (PCR products) of DNA. Suitable primers (10 μM) were sent to GATC Biotech together with the DNA sample.

**Transformation of *E. coli***

*E. coli* cells were transformed by electroporation according to the protocol of Dower (Dower *et al.*, 1988) and Wirth (Wirth, 1989) using a Bio-Rad Gene Pulser<sup>TM</sup>. DNA (from *E. coli* or yeast DNA preparations) was added to the frozen competent *E. coli* cells and the sample was incubated and thawed for 10 min on ice. The cell suspension was then transferred to electroporation cuvettes and directly pulsed. The Bio-Rad Gene Pulser was set to a voltage of

2.5 kV per cm, a resistance of 200  $\Omega$  and a capacity of 25  $\mu$ F. Immediately after the pulse the cells were mixed with 1 ml of pre-warmed SOC medium and transferred to an eppendorf tube. The cells were incubated at 37°C for 45 min at 600 - 800 rpm in a Thermomixer (Eppendorf) before plating on selective LB agar plates containing kanamycin or ampicillin. In case the cells were transformed with a HXT7-coding plasmid, the incubation was performed at room temperature for 4 hours without shaking or at 20-25°C with shaking for 2 hours.

### **Transformation of *S. cerevisiae***

For transformation of *S. cerevisiae*, two different protocols of the LiAc/SS carrier DNA/PEG method from Gietz et al. (Gietz and Schiestl, 2007a, Gietz and Schiestl, 2007b) were used with small deviations. Liquid cultures were grown in suitable medium to an OD of 0.6 – 1.0. Centrifugation of the culture and for the washing steps were shortened to 2 minutes at 3000x g. The single-stranded carrier DNA was used as a 10 mg/ml solution, allowing a volume of 54  $\mu$ l or 74  $\mu$ l of DNA in the transformation mix, respectively. The duration of the heat-shock was 35 minutes. After transformation the whole cell suspension was directly plated on the selection medium or, in case of transformations with a dominant selection marker, transferred to 5 ml of appropriate liquid medium for regeneration. After regeneration cells were pelleted, resuspended in 50-100  $\mu$ l medium and plated out.

DNA amounts for transformations were approx. 500 ng for single plasmids,  $\geq 1000$  ng each for co-transformations with multiple plasmids and  $\geq 2000$  ng and more for integrative DNA-cassettes (e.g. for gene deletions).

### **Codon optimization of genes**

The ORF of some genes has been codon-optimized. The codons have been adapted to the codon-usage of *S. cerevisiae* as determined by the preferred codons of the glycolytic genes. Described in Wiedemann et al. (Wiedemann and Boles, 2008).

### **Plasmid construction by homologous recombination (recombinational cloning)**

Plasmids were constructed in vivo by homologous recombination of suitable DNA fragments (vector backbone and insert(s)) in *S. cerevisiae*. For this purpose the respective vector was linearized at the site of insertion by restriction digestion. Optionally, the resulting vector backbone was purified by agarose gel-electrophoresis and subsequent gel extraction. The inserts were designed to have flanking sequences ( $> 30$  bp) homologous to the region targeted for insertion or, in case of multiple insert fragments, to each other. Inserts were amplified by PCR

and could be provided with homologous sequences by using primers with corresponding 5' ends (homologous overhangs). *S. cerevisiae* was transformed with the DNA fragments and transformants were plated out on selective medium. Colonies were picked to inoculate selective liquid medium. DNA was isolated from these cultures and used for transformation of *E. coli* for plasmid separation and proliferation. Plasmids containing the gyrase inhibitor gene *ccdB* are toxic to most *E. coli* strains so *ccdB*-resistant strain *E. coli* DB3.1 was used for these plasmids. Plasmids were isolated from *E. coli* single-colony cultures and verified by analytic restriction digestion and DNA sequencing. Glycerol stock cultures were set up for correct clones.

### Genomic gene deletion or insertion by homologous recombination

For gene deletions in the genome of *S. cerevisiae*, marker cassettes were integrated into the respective gene by homologous recombination (Carter and Delneri, 2010, Güldener *et al.*, 1996, Sauer, 1987). The marker cassettes were amplified by PCR using primers with 5' ends homologous to the target gene to enable site-directed insertion. The cassettes are composed of a dominant marker gene (*kanMX4*/G418, *hphNT1*/Hygromycin B, *natNT2*/clonNAT), flanked by a promoter (*pTEF*) and a terminator (*tTEF*, *tCYC1* and *tADHI*, respectively) and *loxP* sites. These sites allow excision of the genome-integrated marker cassettes by the cre recombinase, which clears the marker for another round of gene deletion. After transformation of *S. cerevisiae* with a deletion cassette, cells were plated out on selective medium and replica plated on the same medium once. Single colonies were streaked out again to obtain single clones, which were then picked and grown in selective medium. The DNA was isolated from these cultures and the correct integration was confirmed by PCR with different primer combinations. Primers for confirmation are termed as seen in the Table below. Glycerol stock cultures were set up for correct clones.

### Nomenclature of primers used for confirmation of genomic integration by PCR

| name | position                                                | direction  |
|------|---------------------------------------------------------|------------|
| A1   | upstream of the integration site                        | downstream |
| A2   | within the region that gets replaced by the integration | upstream   |
| A3   | within the region that gets replaced by the integration | downstream |
| A4   | downstream of the integration site                      | upstream   |
| K2   | within the deletion cassette/gene that is integrated    | upstream   |

---

|    |                                                      |            |
|----|------------------------------------------------------|------------|
| K3 | within the deletion cassette/gene that is integrated | downstream |
|----|------------------------------------------------------|------------|

---

For recycling of the marker, cells were transformed with a plasmid encoding the cre recombinase under control of the galactose-inducible *GAL1*-promotor (pSH47 or pNatCre). After brief induction of the recombinase, cells were selected for loss of the dominant marker by replica plating. Since full expression of the recombinase is lethal in *hxt*<sup>0</sup> strains, basal expression of the recombinase under non-inducing conditions was used for these strains. Removal of the cassette was again controlled by PCR (see above). The integration of gene cassettes for overexpression of genes was done accordingly. Within these cassettes only the dominant marker is flanked by loxP sites and is excised, the rest of the cassette remains in the genome.

#### List of primers

| Primer name      | Sequence (5'-3') [SEQ ID NO.]                                                      | Description                                               |
|------------------|------------------------------------------------------------------------------------|-----------------------------------------------------------|
| GAL2_for         | AACACAAAAACAAAAAGTTT TTTTAATTTTA<br>ATCAAAAAATGGCAGTTGAGGAGAACAA<br>[SEQ ID NO. 2] | forward primer for<br>GAL2 amplification                  |
| GAL2_rev         | GAATGTAAGCGTGACATAACTAATTACATG<br>ACTCGAGTTATTCTAGCATGGCCTTGTACC<br>[SEQ ID NO. 3] | reverse primer for<br>GAL2 amplification                  |
| GAL2 T354A<br>fw | TTATTTTTTCTACTACGGTGCCGTTATTTTCA<br>AGTCAG<br>[SEQ ID NO. 4]                       | forward primer for<br>mutagenesis of T354<br>to A in GAL2 |
| GAL2 T354A<br>rv | GACTTGAAAATAACGGCACC GTAGTAGAAA<br>AAATAATTG<br>[SEQ ID NO. 5]                     | reverse primer for<br>mutagenesis of T354<br>to A in GAL2 |
| GAL2 V71I<br>fw  | GTCTGAATATGTTACCATTTCTTGCTTTGTT<br>TGTG<br>[SEQ ID NO. 6]                          | forward primer for<br>mutagenesis of V71 to<br>I in GAL2  |
| GAL2 V71I<br>rv  | AAACAAAGCAAGGAAATGGTAACATATTCA<br>GACATG<br>[SEQ ID NO. 7]                         | reverse primer for<br>mutagenesis of V71 to<br>I in GAL2  |

#### *Methods for cell cultivation and fermentation experiments*

##### **Spectrophotometrical determination of cell density**

The cell concentration in a liquid culture was quantified spectrophotometrically by measuring the optical density at 600 nm (OD<sub>600</sub>). Samples of the cell culture or dilutions thereof were placed in a polystyrene (PS) cuvette and analysed in a Ultrospec 2100 pro spectrophotometer (GE Healthcare, USA) at 600 nm.

**Glycerol stock cultures**

For long time storage of specific strains and plasmid-containing *E. coli* glycerol stock cultures were prepared. For this purpose stationary cultures of *S. cerevisiae* or growing cultures of *E. coli* were mixed 1:1 with 50 % (v/v) glycerol and stored at -80°C.

**Semi-solid agar cultivation**

The Semi-solid agar method of cultivation was chosen to expand the plasmid cDNA library (in *E. coli*). By this method representational biases that can occur during growth in liquid culture can be minimized. Incubation is done at 30°C helping to stabilize unstable clones (Hanahan *et al.*, 1991, Sassone-Corsi, 1991). The protocol can be found at Life technologies website. In brief, 2x concentrated LB medium is mixed with 3 g/l SeaPrep agarose while stirring, autoclaved and cooled down to 37°C. The antibiotic and  $4 \cdot 10^5$  to  $6 \cdot 10^5$  (per 450 ml medium) are added to the medium and mixed for 2 minutes. The bottles are then incubated in an ice-bath at 0°C for 1 hour and then gently transferred to 30°C for 40-45h of incubation (without disturbance). After growth the cells can be pelleted from the semi-solid agar by centrifugation at 10400x g.

**Serial dilution spot assays (drop tests)**

For easy comparison of growth of different *S. cerevisiae* strains under various growth conditions a serial dilution spot assay was performed. Cells were grown in liquid culture to exponential phase in appropriate medium, collected by centrifugation (2000x g, 2 min), washed twice with sterile water and then resuspended to an OD<sub>600</sub> of 1.0 in selective medium without carbon source. From this cell suspension a ten-fold serial dilution was prepared in selective medium (four dilution steps). 6 µl of each cell suspension were spotted on plates of the media to be examined and allowed to dry. Plates were incubated at 30°C.

**Aerobic batch fermentations**

Aerobic batch fermentations were done in shake flasks of varying sizes (volume 5-10x of the culture volume) on rotary shakers (150-180 rpm) usually at 30°C. The evolutionary engineering was done as a serial aerobic batch fermentation (details see below)

**Anaerobic batch fermentations**

For anaerobic batch fermentations, shake flasks were sealed with a rubber plug and a fermentation lock. The volume of the flasks was matching the culture volume of 100 ml. The cultures were stirred continuously with 120 rpm on a magnetic stirrer at 30°C. In this work the fermentations were done at  $OD_{600}=10$ . For this purpose grown cells were harvested and set to an  $OD_{600}$  of 20 in 50 ml fermentation medium without carbon source. To start the experiment this cell suspension was added to the prepared flasks containing 50 ml fermentation medium supplemented with 2x concentrated carbon source. Samples for determination of cell concentration and for HPLC analysis were taken through an inserted, sterile needle and syringe.

### **Anaerobic fermentations in a fermenter**

Some fermentations with industrial *S. cerevisiae* strains were conducted in an Infors Multifors fermenter (2x 1.4l) with a working volume of 800 ml and equipped with temperature, pH, O<sub>2</sub> and CO<sub>2</sub> sensors. The fermenter was filled with 530 ml of concentrated fermentation medium (without carbon source and supplements) and then autoclaved. Prior to the start of the fermentation, 250 ml concentrated carbon source solution, 100 µl/l Antifoam 204 and the other supplements (vitamins, trace elements and antibiotics) were added to the concentrated medium. The fermenter was purged with nitrogen gas to initially create anaerobic conditions. A feed of nitrogen gas to the head space (0.4 l/min) was applied during the experiment to maintain anaerobic conditions. The gas outlet was cooled to condense and return vapor and then piped through a gas washing bottle. The culture was stirred with 300 rpm, temperature kept at 35°C and pH kept at 5.0 by automated addition of 2M KOH or 2M H<sub>2</sub>PO<sub>4</sub>. The inoculum of cells (in 20 ml fermentation medium) was added when all set parameters were reached and constant. During the fermentations samples were taken for cell dry weight determination and HPLC analysis. Iris 5.2 Software (Infors) was used to operate the fermenter and monitor the experiment.

### **Determination of cell dry mass**

To determine cell dry mass 5-10 ml of a liquid culture were vacuum filtered through a pre-washed and dried (as described below) filter (nitrocellulose, pore size 0.45 µm) and subsequently washed twice with ddH<sub>2</sub>O. The filters were dried in a microwave oven at 120-150 W for 15 min and left to cool and dry further in a desiccator for another 15 min. The filters were weighted prior to filtering and after drying to measure the cell dry weight. The method described has been adapted from Ask et al. (Ask *et al.*, 2013).

## Evolutionary engineering

The evolutionary engineering was done as an serial aerobic batch fermentation. Transformants were first inoculated in selective SCE<sub>2</sub> to obtain biomass and then grown in selective SC and SM medium with 20 g/l xylose to adapt the strain to xylose utilization. After these initial cultures the cells were switched to medium with 10 g/l xylose and increasing concentrations of glucose to apply an evolutionary pressure. When they were reaching late exponential to early stationary phase cells were harvested by centrifugation (2000x g, 2 min) and transferred to fresh medium to an OD<sub>600</sub> of 0.2. Glucose concentrations were increased every time an adaptation could be seen or growth was not negatively influenced by the current glucose concentration. To all media, liquid and solid, G418 was added to select for the AFY10 strain background. Liquid media additionally contained 0.5 g/l 2-deoxy-D-glucose (2-DOG) to suppress formation of suppressor mutants of the *hxx*<sup>0</sup> phenotype. *hxx*<sup>0</sup> strains, but not wild type strains, are resistant to 2-DOG (Subtil and Boles, 2012). Lack of glucose-consumption has been confirmed by streaking out culture samples on glucose-media-plates and also by HPLC analysis.

## Metabolite analysis by HPLC

For analysis of metabolites cell-free samples (5-10 min, 4°C, 16000x g) were mixed with 1/9 volumes of 50 % (w/v) 5-sulfosalicylic acid and centrifuged (5-10 min, 4°C, 16000x g). The supernatant was analyzed in an UHPLC<sup>TM</sup> system by Thermo Scientific (Dionex<sup>TM</sup> UltiMate 3000) equipped with a HyperREZ XP Carbohydrate H<sup>+</sup> 8 µm column and a refractive index detector (Thermo Shodex RI-101). Separation was carried out at column temperature of 65°C with 5 mM sulfuric acid as mobile phase with a flow rate of 0.6 ml/min. Chromeleon 6.80 software was used to control the system and to analyze the data. Five standards (mixtures of D-glucose, D-xylose, xylitol, acetate, glycerol and ethanol with concentrations of 0.01 - 3 % (w/v)) were analyzed for quantification of the different compounds.

## Sugar uptake assays

Sugar uptake assays were done as described by Bisson et al. (Bisson and Fraenkel, 1983) with modifications according to Walsh et al. (Walsh *et al.*, 1994).

Transformants of strain EBY.VW4000 were grown in selective YEPE to an OD of 1.1 - 1.6, harvested by centrifugation and washed twice in ice-cold uptake-buffer (RT, 3 min, 3000x g). Cells were kept on ice from here on. The cell pellet was resuspended in ice-cold uptake-buffer to a concentration of 60 mg<sub>ww</sub>/ml and aliquoted to 110 µl. One cell suspension aliquot and one sugar solution were incubated in a water bath at 30°C for 4-5 min. 100 µl of the cell suspension

were pipetted to the sugar solution (50  $\mu$ l), mixed briefly by pipetting and incubated for 5 (D-[U- $^{14}$ C]-glucose) or 20 sec (D-[1- $^{14}$ C]-xylose). The uptake reaction was stopped by transferring 100  $\mu$ l of the mixture into 10 ml ice-cold quenching-buffer, which was immediately filtered through a Durapore membrane filter (0.22  $\mu$ m pore size, Millipore<sup>TM</sup>). The filter was washed twice with 10 ml ice-cold quenching-buffer, transferred to a scintillation vial containing 4 ml scintillation cocktail (Rotiszint eco plus, Roth) and shaken thoroughly. Additionally to this filter sample (cpm<sub>filter</sub>), 10  $\mu$ l of each reaction were transferred directly to a scintillation vial with 4 ml scintillation cocktail for determination of the total counts in the reaction (cpm<sub>total</sub>). To determine a value for sugar that is bound unspecifically to the cell surface or the filter (cpm<sub>blank</sub>) a few samples of 33,3  $\mu$ l sugar solution and 66,6  $\mu$ l cell suspension were mixed in 10 ml icecold quenching buffer and treated as described above. Radioactivity of all vials was analysed in an Wallac 1409 liquid scintillation counter.

Stocksolutions of 2M, 500 mM or 20 mM glucose or xylose (in H<sub>2</sub>O) were used to prepare the sugar solutions for the assays (threefold of the desired concentration in the uptake reaction (S, substrate concentration), 50  $\mu$ l aliquots). Uptake was measured at sugar concentrations 0.2, 1, 5, 25 and 100 mM for glucose and 1, 5, 25, 66, 100, 200 and 500 mM for xylose. Inhibition of xylose uptake by glucose was measured at 25, 66 and 100 mM xylose with additional 25 and 100 mM unlabelled glucose. Sugar solutions contained 0.135 to 0.608  $\mu$ Ci of D-[U- $^{14}$ C]-glucose (290-300 mCi/mmol) or D-[1- $^{14}$ C]-xylose (55 mCi/mmol) (American Radiolabeled Chemicals Inc., St. Louis, MO, USA).

Data of the uptake assays were used for following calculations:

The amount of sugar ( $A_{\text{sugar}}$ , in nmol) taken up during the incubation time ( $t$ , in seconds) at a certain sugar concentration ( $S$ , in mM):

$$A_{\text{sugar}} = ((\text{cpm}_{\text{sample}} - \text{cpm}_{\text{blank}}) / (\text{cpm}_{\text{total}} \cdot 10)) \cdot S \cdot 100 \mu\text{l}$$

Transport velocity (in  $\text{nmol} \cdot \text{min}^{-1} \cdot \text{mg}_{\text{ww}}^{-1}$ ) calculated per milligrams of cell ( $m$ , in  $\text{mg}_{\text{ww}}$ ):

$$V = (A_{\text{sugar}} \cdot 60 \text{ s}) / (t \cdot m)$$

Calculation of  $K_m$  (Michaelis constant),  $V_{\text{max}}$  (maximal initial uptake velocity) and  $K_i$  (inhibitor constant for competitive inhibition) was done by nonlinear regression analysis and global curve fitting in Prism 5 (GraphPad Software, Inc.) with values of three independent measurements.

### ***Bioinformatic methods***

DNA sequences were obtained from the Saccharomyces Genome Database (SGD, (Cherry *et al.*, 2012)). Sequence alignments for transporter proteins were conducted using the PRALINE multiple alignment server ((Simossis and Heringa, 2005)) with standard settings plus PHOBIUS transmembrane structure prediction (Käll *et al.*, 2004). Phylogenetic trees were calculated from PRALINE alignments with ClustalW2 phylogeny (Larkin *et al.*, 2007) and visualized with Phylodendron software .

Similarities and identities between protein sequences were calculated from PRALINE alignments using SIAS .

Figures for sequence alignments were created with ALINE software (Bond and Schüttelkopf, 2009).

## **EXAMPLE 1            T354A**

### **1.1 Investigation of Gal2\_T354A**

Ethanol Red is an industrial strain, which is a promising candidate for fermentations of lignocellulosic hydrolysates. Genes encoding enzymes for the xylose and arabinose metabolic pathway could be integrated into the genome, resulting in strain HDY.GUF5 (Demeke *et al.*, 2013). HDY.GUF5 was further evolved on xylose and engineered by genetic engineering, finally resulting in strain HDY.GUF9 (Dietz 2013). HDY.GUF9 was further evolved by evolutionary engineering on arabinose, resulting in strain HDY.GUF10. This strain also had an improved growth behavior on xylose. The xylose consumption rate was improved by about 80%, the arabinose consumption rate by about 25%. Determination of the xylose uptake rate with radioactive sugar uptake assays revealed that HDY.GUF10 had a xylose uptake rate 35% higher than HDY.GUF9. As Gal2 is the only transporter in *S. cerevisiae* which can transport xylose and arabinose in significant amounts, the *GAL2* gene was isolated from both strains and sequenced. One amino acid substitution was found in Gal2\_HDY.GUF10 compared to Gal2\_HDY.GUF9, which is T354A, probably located within the transport channel in transmembrane helix 7 at the extracellular side. This position might play a role for the alteration of conformation of the transporter. The desired sequence was amplified from chromosomal DNA of HDY.GUF10 and cloned into p426 in order to investigate the modified Gal2p. As a control Gal2p of Ethanol Red was used. The received vectors were transformed into the screening strains to test the growth on several media.

### **1.2    Test for functionality of Gal2\_T354A**

First a growth test was performed with the vectors in VW4000 on glucose and maltose media. The galactose transporter of the wild type CEN.PK2, as well as Gal2\_ep3.1 were used as controls and as a comparison. The Gal2p wild type of the industrial strain Ethanol Red showed the same growth like the wild type of CEN.PK2, as expected. The growth of Gal2\_GUF10 was like the growth of both wild types on galactose, whereas the growth on glucose looked like the growth of the error prone mutant (Figure 1).

Furthermore this was also done with AFY10 in order to investigate the xylose specificity and the glucose affinity of the transformants. The wild type Gal2\_EtOH Red was not able to grow on a low xylose concentration, like 0.2%. However growth of Gal2\_GUF10 was observed on this media. Nevertheless the growth of Gal2\_GUF10 was inferior to Gal2\_ep3.1 on 2% xylose. Therefore Gal2\_GUF10 seems to have a higher xylose affinity. On the media with additional glucose only the error prone mutant showed growth (Figure 2).

## **EXAMPLE 2      T354A/V71I**

### **2.1      Effect of T354A in combination with other mutations**

The amino acid exchange T354A in Gal2 was found in strain Ethanol Red HDY.GUF10<sup>TM</sup> which was evolved on arabinose originating from strain HDY.GUF9. This indicates that the transport properties of Gal2 for arabinose were improved. It was also shown that the T354A mutation in Gal2 of HDY.GUF10 could restore growth on low xylose concentrations. The sequence of Ethanol Red Gal2 and Gal2\_HDY.GUF9 differs in two amino acids from Gal2 of strain CEN.PK (L280R and V71I). To determine the effects of these two differences, plasmid expression constructs were made with Gal2 of CEN.PK which have V71I and L280R alone or in combination with T354A. To determine their properties, growth drop tests were performed after transformation of the constructs with the screening strains EBY.VW4000 und AFY10. As controls CEN.PK Gal2 wild type, p426\_empty vector and T354A alone in CEN.PK Gal2 were used.

### **2.2      Growth of cells with Gal2\_T354A in combination with L280R and V71I on hexoses**

The following variants of Gal2 were constructed in expression vectors: T354A, T354A+V71I, T354A+L280R, L280R and I71V. They were transformed into competent EBY.VW4000 cells. Then drop tests were performed on different carbon sources.

As can be seen (Figure 3) V71I and L280R alone have no effect on uptake of glucose or galactose. T354A alone strongly impairs growth on glucose. The combination of V71I and T354A does not mediate growth on glucose. The combination of T354A and L280R however can mediate growth on glucose. Growth is however slower than with Gal2 wild type.

### **2.3 Growth of cells with Gal2\_T354A in combination with L280R and V71I on xylose and sugar mixtures**

The various constructs were transformed into AFY10 cells together with vector YEp181\_pHXT7-optXI\_Clos, and serial dilution growth drop tests were performed.

No variant can mediate growth on xylose-glucose mixture plates indicating that the mutant transporters are all inhibited by glucose. Growth with high xylose concentrations (20 g/l) is not much different between the various constructs. Growth on low xylose concentrations (2 g/l) however shows significant differences: L280R mediates growth like Gal2 wild type; growth of the combination of L280R and T354A looks like T354A alone or the empty vector control. V71I mediates very slow growth on 2 g/l xylose. The combination of V71I and T354A however mediates growth like the Gal2 wild type (Figure 4) whereas, in contrast, T354A and V71I alone mediate poor growth on low xylose concentrations. This demonstrates that amino acid exchange V71I in Gal2 of Ethanol Red is responsible for the low xylose uptake activity especially at low xylose concentrations. This defect is suppressed by the additional exchange of T354A. This explains the improved growth behavior of GUF-10 compared to GUF-9. As GUF-10 was evolved on arabinose medium it can be concluded that also uptake of arabinose of Gal2 from Ethanol Red is improved by the T354A mutation by suppressing the V71I exchange.

The features disclosed in the foregoing description, in the claims and/or in the accompanying drawings may, both separately and in any combination thereof, be material for realizing the invention in diverse forms thereof.

### **REFERENCES**

Becker, J. and Boles, E. (2003) A modified *Saccharomyces cerevisiae* strain that consumes L-Arabinose and produces ethanol. *Appl Environ Microbiol* 69(7),4144-50.

Brat D, Boles E, Wiedemann B (2009) Functional expression of a bacterial xylose isomerase in *Saccharomyces cerevisiae*. *Appl Environ Microbiol*. 75:2304-11.

Bruder S (2011) Analyse eines unbekannten Glukose Transporters in *Saccharomyces cerevisiae* und Entwicklung eines Xylose-Transporter Screening Systems. Diploma thesis (Goethe Universität Frankfurt).

Demeke MM, Dietz H, Li Y, Foulquié-Moreno MR, Mutturi S, Deprez S, Den Abt T, Bonini BM, Liden G, Dumortier F, Verplaetse A, Boles E, Thevelein JM (2013) Development of a D-xylose fermenting and inhibitor tolerant industrial *Saccharomyces cerevisiae* strain with high performance in lignocellulose hydrolysates using metabolic and evolutionary engineering. *Biotechnol Biofuels*. 6:89.

Dien, B.S., Kurtzman, C.P., Saha, B.C. and Bothast, R.J. (1996) Screening for L-arabinose fermenting yeasts. *Appl Biochem Biotechnol* 57-58, 233-42.

Dietz H (2013) Konstruktion und Charakterisierung von pentosefermentierenden Industriehefestämmen. PhD thesis. Goethe University Frankfurt, Germany

Farwick A, Bruder S, Schadeweg V, Oreb M, Boles E. Engineering of yeast hexose transporters to transport D-xylose without inhibition by D-glucose. *Proc Natl Acad Sci U S A*. 2014 Apr 8;111(14):5159-64.

Hahn-Hagerdal, B., Wahlbom, C.F., Gardonyi, M., van Zyl, W.H., Cordero Otero, R.R. and Jonsson, L.J. (2001) Metabolic engineering of *Saccharomyces cerevisiae* for xylose utilization. *Adv Biochem Eng Biotechnol* 73,53-84.

Horak J and Wolf DH (1997) Catabolite inactivation of the galactose transporter in the yeast *Saccharomyces cerevisiae*: ubiquitination, endocytosis, and degradation in the vacuole. *J Bacteriol* 179(5):1541-1549.

Jeppson M, Bengtsson O, Franke K, Lee H, Hahn-Hagerdal B, Gorwa-Grauslund MF. (2006) The expression of a *Pichia stipitis* xylose reductase mutant with higher K(M) for NADPH increases ethanol production from xylose in recombinant *Saccharomyces cerevisiae*.

*Biotechnol Bioeng.* 93(4):665-73.

Jin YS, Jeffries TW. (2004) Stoichiometric network constraints on xylose metabolism by recombinant *Saccharomyces cerevisiae*. *Metab Eng.* 6(3):229-38.

Katahira S, Mizuike A, Fukuda H, Kondo A. (2006) Ethanol fermentation from lignocellulosic hydrolysate by a recombinant xylose- and cellooligosaccharide-assimilating yeast strain. *Appl Microbiol Biotechnol.* 72(6):1136-43.

Kotter,P. and Ciriacy, M. (1993) Xylose fermentation by *Sacharomyces cerevisiae*. *Appl Microbiol Biotechnol* 38, 776-783.

Kou, S.C., Christensen, M.S. and Cirillo, V.P. (1970) Galactose transport in *Saccharomyces cerevisiae*. II. Characteristics of galactose uptake and exchange in galactokinaseless cells. *J Bacteriol* 103(3), 671-8.

Kuyper, M., Toirkens, M.J., Diderich, J.A., Winkler, A.A., van Dijken, J.P. and Pronk, J.T. (2005b) Evolutionary engineering of mixed-sugar utilization by a xylose-fermenting *Saccharomyces cerevisiae* strain. *FEMS Yeast Res* 5(10), 925-34.

Kuyper, M., Winkler, A.A., van Dijken, J.P. and Pronk, J.T. (2004) Minimal metabolic engineering of *Saccharomyces cerevisiae* for efficient anaerobic xylose fermentation: a proof of principle. *FEMS Yeast Res* 4(6), 655-64.

Lucas, C. and Uden, N.v. (1986) Transport of hemicellulose monomers in the xylose-fermenting yeast *Candida shehatae*. *Appl Microbiol Biotechnol* 23,491-495.

Pitkanen JP, Rintala E, Aristidou A, Ruohonen L, Penttilä M. (2005) Xylose chemostat isolates of *Saccharomyces cerevisiae* show altered metabolite and enzyme levels compared with xylose, glucose, and ethanol metabolism of the original strain. *Appl Microbiol Biotechnol.* 67(6):827-37.

Reifenberger E, Boles E and Ciriacy M (1997) Kinetic characterization of individual hexose transporters of *Saccharomyces cerevisiae* and their relation to the triggering mechanisms of glucose repression. *Eur J Biochem* 245(2):324-333.

Richard, P., Putkonen, M., Vaananen, R., Londesborough, J. and Penttila, M. (2002) The missing link in the fungal L-arabinose catabolic pathway, identification of the L-xylulose reductase gene. *Biochemistry* 41(20), 6432-7.

Sambrook J. and Russell D.W. (2001) *Molecular cloning. A laboratory manual*. New York: Cold Spring Harbor.

Subtil T and Boles E (2012) Competition between pentoses and glucose during uptake and catabolism in recombinant *Saccharomyces cerevisiae*. *Biotechnol Biofuels* 5(1):14.

Wieczorke, R., Krampe, S., Weierstall, T., Freidel, K., Hollenberg, C.P. and Boles, E. (1999) Concurrent knock-out of at least 20 transporter genes is required to block uptake of hexoses in *Saccharomyces cerevisiae*. *FEBS Lett* 464(3), 123-8.

Wiedemann B., Boles E. (2008) Codon-optimized bacterial genes improve L-Arabinose fermentation in recombinant *Saccharomyces cerevisiae*. *Appl Environ Microbiol.* 74 (7):2043-50.

### **Claims**

1. Polypeptide, comprising at least one amino acid substitution, wherein the at least one amino acid substitution includes an amino acid substitution at a position corresponding to T354 of the amino acid sequence of SEQ ID NO: 1,  
wherein the polypeptide has at least 90% or 95% sequence identity with the amino acid sequence of SEQ ID NO: 1, and wherein the polypeptide has an *in vitro* and/or *in vivo* pentose transport function.
2. The polypeptide according to claim 1, wherein the polypeptide is Gal2 of *Saccharomyces cerevisiae*.
3. The polypeptide according to claim 1 or 2, wherein the amino acid substitution at a position corresponding to T354 of the amino acid sequence of SEQ ID NO: 1 is T354A.
4. The polypeptide according to any one of claims 1 to 3, comprising a further amino acid substitution at a position corresponding to V71 of the amino acid sequence of SEQ ID NO: 1.
5. The polypeptide according to claim 4, wherein the amino acid substitution at a position corresponding to V71 of the amino acid sequence of SEQ ID NO: 1 is V71I.
6. The polypeptide according to any one of claims 1 to 5, wherein the amino acid substitution at a position corresponding to T354 increases the activity of the *in vitro* and/or *in vivo* pentose transport function compared to the polypeptide without such amino acid substitution.
7. The polypeptide according to any of claims 1 to 6, wherein the amino acid substitution at a position corresponding to T354 increases the affinity of the polypeptide for pentoses compared to the polypeptide without such amino acid substitution.
8. The polypeptide of any one of claims 1 to 7, wherein the pentose is D-xylose and/or L-arabinose.

9. A nucleic acid molecule, coding for a polypeptide according to any one of claims 1 to 8.
10. The nucleic acid molecule of claim 9, further comprising vector nucleic acid sequences, promoter nucleic acid sequences and terminator nucleic acid sequences, other regulatory nucleic acid sequences, or a combination thereof,  
and/or wherein the nucleic acid molecule comprises dsDNA, ssDNA, PNA, CNA, RNA, mRNA or combinations thereof.
11. The nucleic acid molecule of claim 10, wherein the vector nucleic acid sequences are expression vector sequences.
12. Host cell, containing a nucleic acid molecule according to any one of claims 9 to 11.
13. The host cell of claim 12, which expresses the nucleic acid molecule according to any one of claims 9 to 11.
14. The host cell according to any of claims 12 or 13, which is a fungus cell.
15. The host cell according to claim 14, which is a yeast cell.
16. The host cell according to claim 15, which is a cell of *Saccharomyces* species, *Kluyveromyces* sp., *Hansenula* sp., *Pichia* sp. or *Yarrowia* sp.
17. The host cell according to claim 16, which belongs to the species *Saccharomyces cerevisiae*.
18. The host cell according to any one of claims 12 to 17, which has an increased uptake rate for D-xylose and/or L-arabinose compared to a cell not containing a nucleic acid molecule according to any one of claims 9 to 11.
19. The host cell according to any one of claims 12 to 18, further containing  
- nucleic acid molecules which code for proteins of a xylose metabolic pathway,

and/or

- nucleic acid molecules which code for proteins of an arabinose metabolic pathway.

20. The host cell according to claim 19, wherein the proteins of a xylose metabolic pathway are xylose isomerase and xylulokinase, and/or wherein the proteins of an arabinose metabolic pathway are arabinose isomerase, ribulokinase, and ribulose-5-P 4-epimerase.

21. The host cell according to claim 20, which has an increased D-xylose and/or L-arabinose consumption rate, a faster growth rate with D-xylose and/or L-arabinose, or both, as compared to a cell not containing a nucleic acid molecule according to any one of claims 9 to 11.

22. Method for the production of bioethanol and/or other bio-based compounds, comprising expressing a nucleic acid molecule according to any one of claims 9 to 11, wherein the other bio-based compounds are selected from alcohols, organic acids, amino acids, alkanes, terpenes, isoprenoids, solvents, pharmaceutical compounds, and vitamins.

23. The method of claim 22, wherein the alcohols are selected from 1-butanol, isobutanol, and 2-butanol.

24. The method of claim 22, wherein the organic acids are selected from lactic acid, acetic acid, succinic acid and malic acid.

25. The method of any one of claims 22 to 24, comprising the expression of a nucleic acid molecule according to any one of claims 9 to 11 in a host cell according to any one of claims 12 to 21.

26. Use of a polypeptide according to any one of claims 1 to 8, a nucleic acid molecule according to any one of claims 9 to 11, or a host cell according to any one of claims 12 to 21 for the production of bioethanol and/or other bio-based compounds,

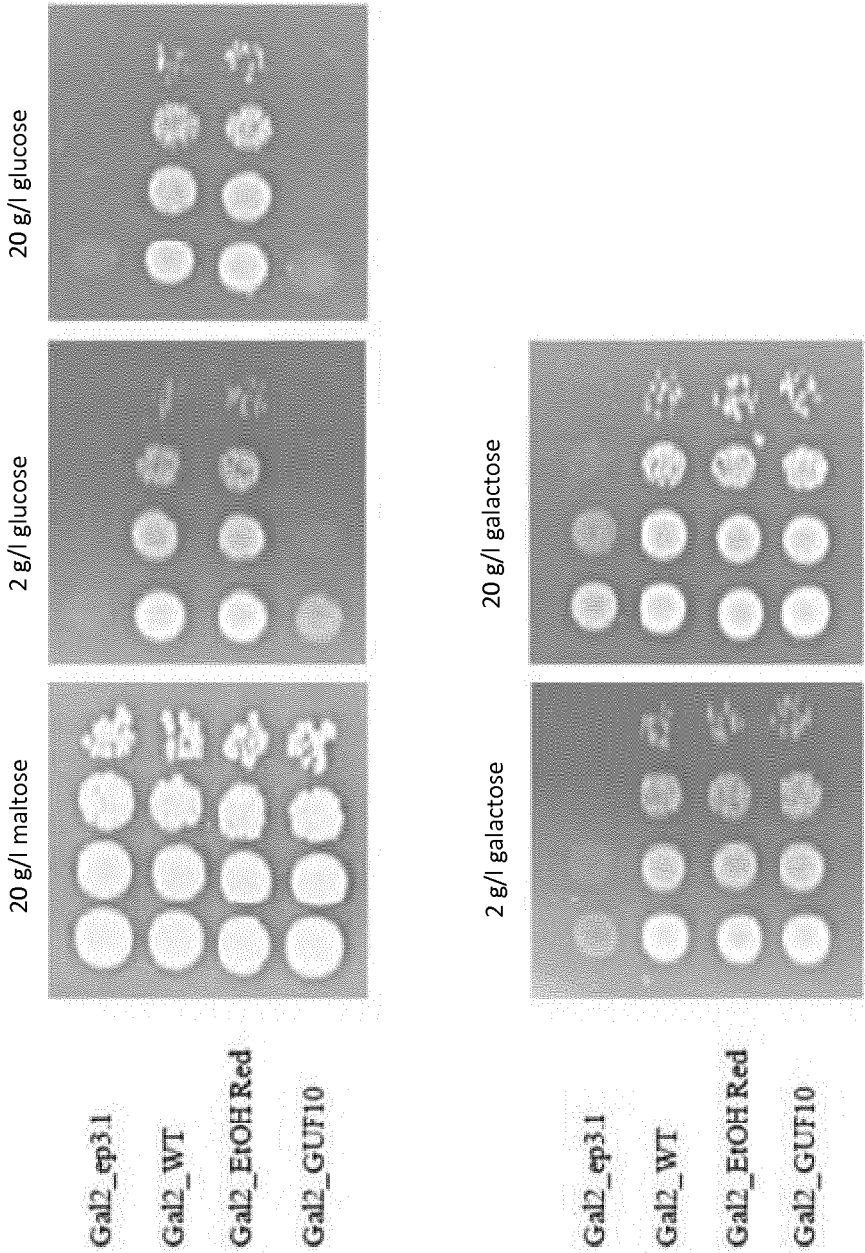
for the recombinant fermentation of biomaterial containing pentoses, or both, wherein the other bio-based compounds are selected from alcohols, organic acids, amino acids, alkanes, terpenes, isoprenoids, solvents, pharmaceutical compounds, and vitamins.

27. The use of claim 26, wherein the alcohols are selected from 1-butanol, isobutanol, and 2-butanol.

28. The use of claim 26, wherein the organic acids are selected from lactic acid, acetic acid, succinic acid and malic acid.

29. The use of any one of claims 26 to 28 for the recombinant fermentation of biomaterial containing D-xylose and/or L-arabinose.

Figure 1



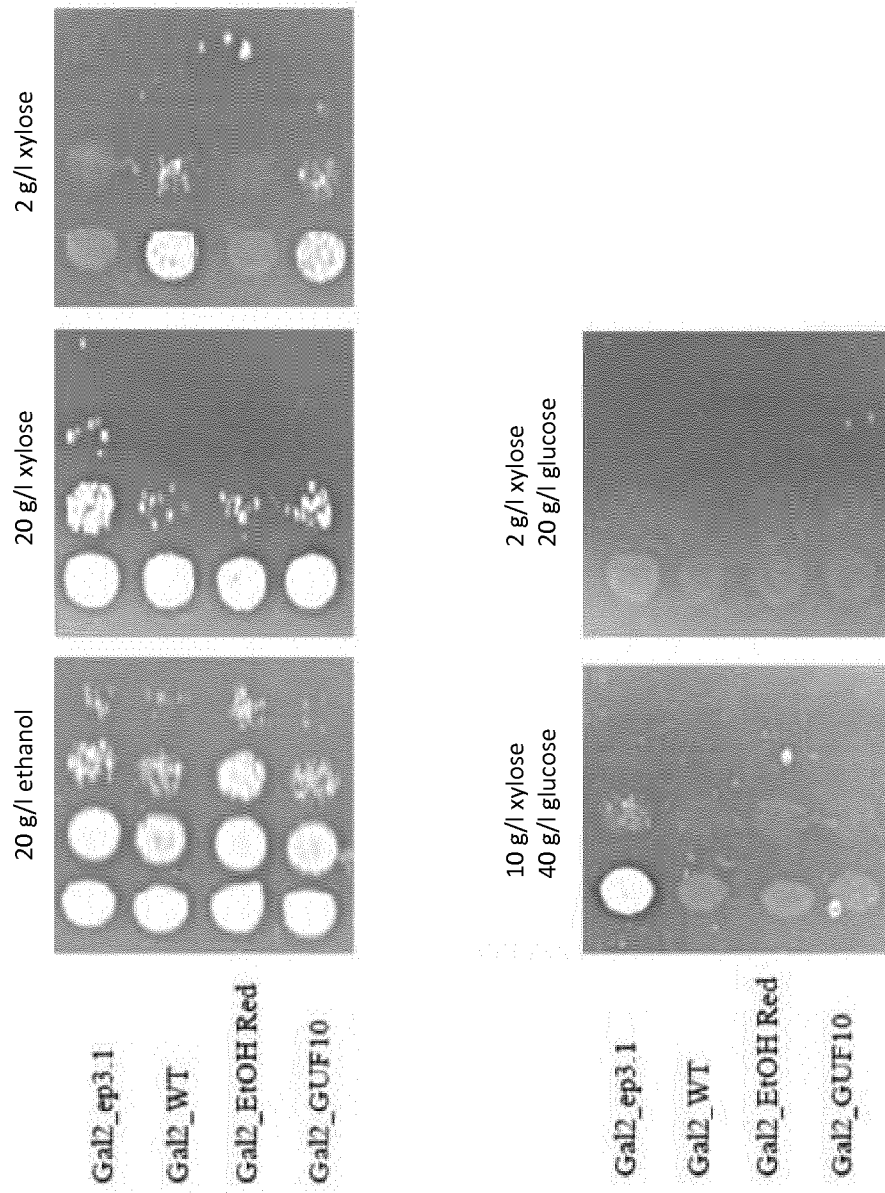
**Figure 2**

Figure 3

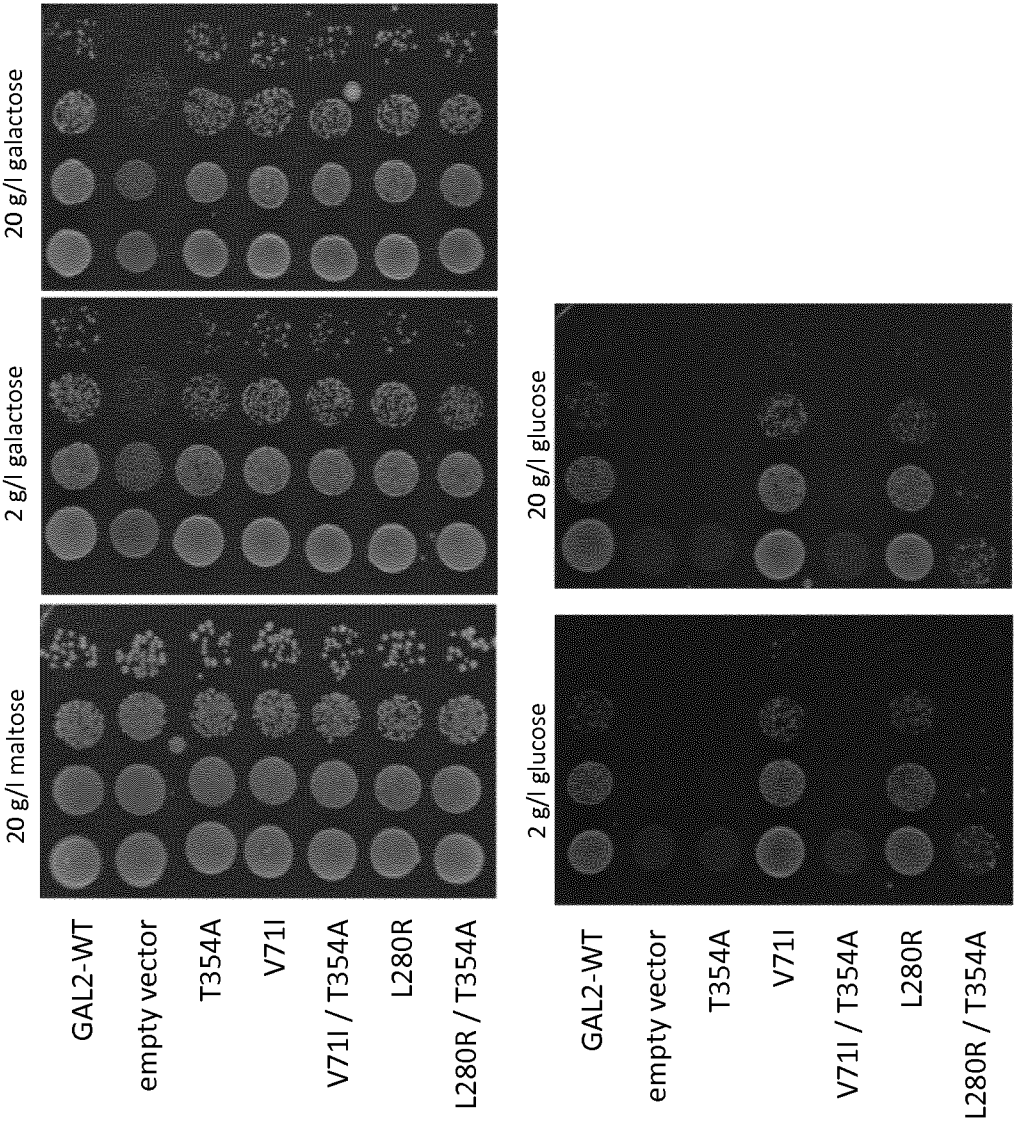


Figure 4

