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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
13 December 2007 (13.12.2007)

PCT

(10) International Publication Number
WO 2007/143247 A2

(51) International Patent Classification:
C12N 1/18 (2006.01)

(21) International Application Number:
PCT/US2007/064418

(22) International Filing Date: 20 March 2007 (20.03.2007)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/810,274 2 June 2006 (02.06.2006) US

(71) Applicant (*for all designated States except US*): **MID-WEST RESEARCH INSTITUTE** [US/US]; 425 Volker Boulevard, Kansas City, Missouri 64100 (US).

(72) Inventors; and

(75) Inventors/Applicants (*for US only*): **KNOSHAUG, Eric** [US/US]; 204 East Street, Golden, Colorado 80403 (US). **JARVIS, Eric** [US/US]; 2265 Dartmouth Avenue, Boulder, Colorado 80305 (US). **SINGH, Arjun** [US/US]; 6977 West Virginia Place, Lakewood, Colorado 80226 (US). **FRANDEN, Mary Ann** [US/US]; 3653 E. Lake Drive, Centennial, Colorado 80121 (US). **ZHANG, Min** [US/US]; 14178 W. Warren Drive, Lakewood, Colorado 80228 (US).

(74) Agents: **WHITE, Paul J.** et al.; 1617 Cole Boulevard, Golden, Colorado 80401 (US).

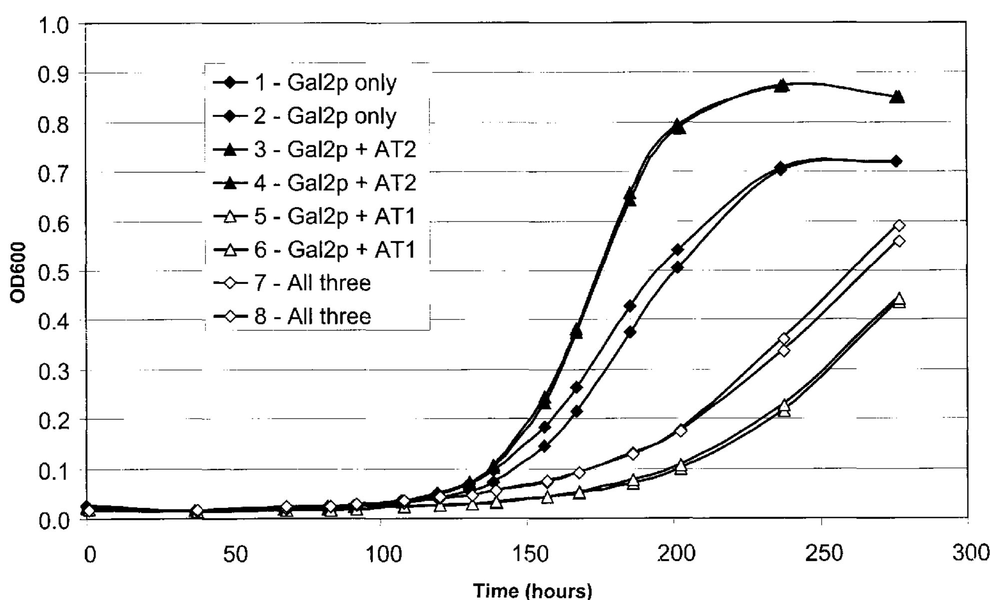
(81) Designated States (*unless otherwise indicated, for every kind of national protection available*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP,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

(84) Designated States (*unless otherwise indicated, for every kind of regional protection available*): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, MT, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Declaration under Rule 4.17:
— *of inventorship (Rule 4.17(iv))*

[Continued on next page]

(54) **Title:** CLONING AND CHARACTERIZATION OF L-ARABINOSE TRANSPORTERS FROM NON-CONVENTIONAL YEAST



(57) **Abstract:** Two genes from non-conventional yeast encode arabinose transporters. These arabinose transporters are capable of transporting arabinose across the cell membrane. These genes may be expressed heterologously in a host that is not otherwise capable of taking up significant amounts of arabinose from the environment of use. Methods are disclosed to use such genetically engineered hosts to ferment pentose such as arabinose, to produce ethanol.

WO 2007/143247 A2

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

CLONING AND CHARACTERIZATION OF L-ARABINOSE TRANSPORTERS FROM NON-CONVENTIONAL YEAST

RELATED APPLICATIONS

[0001] This application claims priority under 35 U.S.C. 119(e) to U.S. Provisional Patent Application No. 60/810,274 entitled "CLONING AND CHARACTERIZATION OF L-ARABINOSE TRANSPORTERS FROM NON-CONVENTIONAL YEAST," filed June 2, 2006, the disclosure of which is hereby incorporated by reference in its entirety.

CONTRACTUAL ORIGIN

[0002] The United States Government has rights in this invention under Contract No. DE-AC36-99GO10337 between the United States Department of Energy and the National Renewable Energy Laboratory, a Division of the Midwest Research Institute.

BACKGROUND

[0003] Fuel ethanol is a suitable alternative to fossil fuels. Ethanol may be produced from plant biomass, which is an economical and renewable resource that is available in large amounts. Examples of biomass include agricultural feedstocks, paper wastes, wood chips and so on. The sources of biomass vary from region to region based on the abundance of natural or agricultural biomass that is available in a particular region. For example, while sugar cane is the primary source of biomass used to produce ethanol in Brazil, corn-derived biomass, such as corn starch and corn fiber, is a large source of biomass used to produce ethanol in the United States. Other agricultural feedstocks include, by way of example: straw; grasses such as switchgrass; grains; and any other cellulose or starch-bearing material.

[0004] A typical biomass substrate contains from 35-45% cellulose, 25-40% hemicellulose, and 15-30% lignin, although sources may be found that deviate from these general ranges. As is known in the art, cellulose is polymer of glucose subunits, and hemicellulose contains mostly xylose. Arabinose is also a significant fermentable substrate that may be found in biomass, such as corn fiber and many herbaceous crops. Other researchers have investigated the utilization of arabinose and hemicellulose, as reported by Hespell, R. B. 1998. *Extraction and characterization of hemicellulose from the corn fiber*

produced by corn wet-milling processes. J. Agric. Food Chem. **46**:2615-2619, and McMillan, J. D., and B. L. Boynton. 1994. *Arabinose utilization by xylose-fermenting yeasts and fungi*. Appl. Biochem. Biotechnol. **45-46:**569-584. The two most abundant types of pentose that exist naturally are D-xylose and L-arabinose.

[0005] It is problematic that most of the naturally available ethanol-producing microorganisms are only capable of utilizing hexose sugar, such as glucose. This is confirmed by a review of the art, such as is reported by Barnett, J. A. 1976. The utilization of sugars by yeasts. Adv. Carbohydr. Chem. Biochem. 32:125-234. Many types of yeast, especially *Saccharomyces cerevisiae* and related species, are very effective in fermenting glucose-based feedstocks into ethanol through anaerobic fermentation. However, these glucose-fermenting yeasts are unable to ferment xylose or L-arabinose, and are unable to grow solely on these pentose sugars. Although some other yeast species, such as *Pichia stipitis* and *Candida shehatae*, may ferment xylose to ethanol, they are not as effective as *Saccharomyces* for fermentation of glucose and have a relatively low level of ethanol tolerance. Thus, the present range of available yeast are not entirely suitable for large scale industrial production of ethanol from biomass.

[0006] Most bacteria, including *E. coli* and *Bacillus subtilis*, utilize L-arabinose for aerobic growth, but they do not ferment L-arabinose to ethanol. Other microorganisms, such as *Zymononas mobilis*, have also been genetically modified to produce ethanol from hexose or pentose. This has been reported, for example, in Deanda, K., M. Zhang, C. Eddy, and S. Picataggio. 1996, Development of an arabinose-fermenting *Zymomonas mobilis* strain by metabolic pathway engineering. Appl. Environ. Microbiol. 62:4465-4470; and Zhang, M., C. Eddy, K. Deanda, M. Finkelstein, and S. Picataggio. 1995 Metabolic engineering of a pentose metabolism pathway in ethanologenic *Zymomonas mobilis*. Science 267:240-243. However, it remains the case that low alcohol tolerance of these non-yeast microorganisms limits their utility in the ethanol industry.

[0007] Much effort has been made over the last decade or so, without truly overcoming the problem of developing new yeast strains that ferment xylose to generate ethanol. Such efforts are reported, for example, in Kötter, P., R. Amore, C. P. Hollenberg, and M. Ciriacy. 1990. Isolation and characterization of the *Pichia stipitis* xylitol dehydrogenase gene, XYL2, and construction of a xylose-utilizing *Saccharomyces cerevisiae* transformant. Curr. Genet. 18:493-500; and Wahlbom, C. F., and B. Hahn-Hägerdal. 2002. Furfural, 5-hydroxymethyl furfural, and acetoin act as external electron acceptors during anaerobic fermentation of xylose in recombinant *Saccharomyces cerevisiae*, Biotechnol.

Bioeng. 78:172-178. Recent studies have been conducted on yeast strains that may ferment arabinose. Sedlak, M., and N. W. Ho. 2001. Expression of *E. coli araBAD* operon encoding enzymes for metabolizing L-arabinose in *Saccharomyces cerevisiae*, Enzyme Microb. Technol. 28:16-24 discloses the expression of an *E. coli* araBAD operon encoding enzymes for metabolizing L-arabinose in *Saccharomyces cerevisiae*. Although this strain expresses *araA*, *araB* and *araD* proteins, it is incapable of producing ethanol.

[0008] United States Patent Application 10/983,951 by Boles and Becker discloses the creation of a yeast strain that may ferment L-arabinose. However, the overall yield is relatively low, at about 60% of theoretical value. The rate of arabinose transport into *S. cerevisiae* may be a limiting factor for complete utilization of the pentose substrate. Boles and Becker attempted to enhance arabinose uptake by over expressing the *GAL2*-encoded galactose permease in *S. cerevisiae*. However, the rate of arabinose transport using galactose permease was still much lower when compared to that exhibited by non-conventional yeast such as *Kluyveromyces marxianus*. Another limitation that may have contributed to the low yield of ethanol in the modified strain of Becker and Boles is the poor activity of the L-arabinose isomerase encoded by the bacterial *araA* gene. Although Becker and Boles used an *araA* gene from *B. subtilis* instead of one from *E. coli*, the specific activity of the enzyme was still low. Other workers in the field have reported that low isomerase activity is a bottleneck in L-arabinose utilization by yeast.

[0009] There remains a need for new arabinose-fermenting strains that are capable of producing ethanol at high yield. There is further a need to identify novel arabinose transporter for introduction into *Saccharomyces cerevisiae* to boost the production of ethanol from arabinose.

[0010] The foregoing examples of the related art and limitations related therewith are intended to be illustrative and not exclusive. Other limitations of the related art will become apparent to those of skill in the art upon a reading of the specification and a study of the drawings.

SUMMARY

[0011] The following embodiments and aspects thereof are described and illustrated in conjunction with systems, tools and methods which are meant to be exemplary and illustrative, not limiting in scope. In various embodiments, one or more of the above-

described problems have been reduced or eliminated, while other embodiments are directed to other improvements.

[0012] Systems, tools and methods are provided for the identification of yeast strains having efficient, single component L-arabinose transport mechanisms. These strains were identified by sequentially screening yeast strains for L-arabinose utilization, ameability of L-arabinose utilizing yeast strains to genetic and biochemical manipulation, autotrophy, and specific L-arabinose transport rates. Yeast strains identified using the systems, tools and methods of the embodiment were then further characterized to identify novel arabinose transporter genes.

[0013] As such, the presently disclosed instrumentalities provide cloned and characterized novel arabinose transporter genes, termed *KmLAT1*and *PgLAT2*, from two non-conventional yeast species, *Kluyveromyces marxianus* and *Pichia guilliermondii* (also known as *Pichia guilliermondii*), respectively. It is disclosed herein that both *Kluyveromyces marxianus* and *Pichia guilliermondii* are efficient utilizers of L-arabinose, which renders them ideal sources for cloning L-arabinose transporter genes.

[0014] The *KmLAT1* gene may be isolated using functional complementation of an adapted *S. cerevisiae* strain that could not grow on L-arabinose because it lacked sufficient L-arabinose transport activity. KmLat1 protein has a predicted length of 556 amino acids encoded by a single ORF of 1668 bp. It is a transmembrane protein having high homology to sugar transporters of many different yeast species. When KmLat1 is expressed in *S. cerevisiae*, transport assays using labeled L-arabinose show that this transporter has the kinetic characteristics of a low affinity arabinose transporter, with K_m = 230 mM and V_{max} = 55 nmol/mg·min. Transport of L-arabinose by KmLat1 is not significantly inhibited by common uncoupling agents but is out-competed by glucose, galactose, xylose, and maltose.

[0015] The *PgLAT2* gene may be isolated using the technique of differential display from *Pichia guilliermondii*. The *PgLAT2* gene has an ORF of 1617 nucleotides encoding a protein with a predicted length of 539 amino acids. It is also predicted to be a transmembrane protein and shows high homology to sugar transporters of many different yeast species. When PgLat2 is expressed in *S. cerevisiae*, transport assays show that this transporter has almost identical L-arabinose transport kinetics as that of wildtype *Pichia guilliermondii*. The PgLat2 transporter when expressed in *S. cerevisiae* has a K_m of 0.07 mM and V_{max} of 18 nmol/mg-min for L-arabinose transport. Inhibition experiments show significant inhibition of the PgLat2 transporter by protonophores (e.g., NaN_3, DNP, and CCP) and H+-adenosine triphosphatase (ATPase) inhibitors (e.g., DESB and DCCD) similar to inhibition in wildtype

P. guilliermondii. Competition experiments show that L-arabinose uptake by the PgLat2 transporter is inhibited by glucose, galactose, xylose and to a lesser extent by maltose.

[0016] The transport kinetics of *S. cerevisiae* Gal2p have been measured and compared to those of KmLat1. The *S. cerevisiae GAL2* gene (SEQ ID NO 5) under control of a *TDH3* promoter exhibits 28 times greater (8.9 nmol/mg·min) L-arabinose transport rate as compared to *GAL2* gene under control of a *ADH1* promoter. The *GAL2*-encoded permease (SEQ ID NO 6) shows a K_m of 550 m*M* and a V_{max} of 425 nmol/mg·min for L-arabinose transport and a K_m of 25 m*M* and a V_{max} of 76 nmol/mg·min for galactose transport. Although L-arabinose transport by both *KmLAT1* and *GAL2* encoded permeases is out-competed by glucose or galactose, the inhibitory effects of glucose or galactose are greater on the *GAL2* encoded permease than on the *KmLAT1* encoded transporter.

[0017] It is further disclosed here that a *S. cerevisiae* strain may be transformed with different combinations of the *KmLAT1* and *PgLAT2* transporter genes and a plasmid carrying the *GAL2* gene native to *S. cerevisiae*. The doubling time for the PgLat2p and Gal2p co-expressing cells grown on L-arabinose is markedly shorter than that of the cells expressing only Gal2p, suggesting that L-arabinose uptake may have been enhanced in these cells. In addition, the PgLat2p and Gal2p co-expressing cells appear to grow to a higher optical density at saturation, suggesting that this strain may be able to utilize the L-arabinose in the medium more completely. This conclusion is supported by HPLC analysis which shows significantly less residual L-arabinose in the culture of cells expressing PgLat2p and Gal2p.

[0018] In addition to the exemplary aspects and embodiments described above, further aspects and embodiments will become apparent by reference to the drawings and by study of the following descriptions.

BRIEF DESCRIPTION OF THE DRAWINGS

[0019] Exemplary embodiments are illustrated in referenced figures of the drawings. It is intended that the embodiments and figures disclosed herein are to be considered illustrative rather than limiting.

[0020] Figure 1 shows phenotypes of L-arabinose-negative mutants obtained from *A. adeninivorans*. The numbers of mutants in each class are indicated in the boxes. The grey shaded box indicates the mutant class in which transport is expected to have been impacted.

[0021] Figure 2 shows phenotypes of L-arabinose-negative mutants obtained from *D. hansenii* var. *fabryii*. The numbers of mutants in each class are indicated in the boxes. The

grey shaded box indicates the mutant class in which transport is expected to have been impacted.

[0022] Figure 3 shows phenotypes of L-arabinose-negative mutants obtained from *P. guilliermondii*. The numbers of mutants in each class are indicated in the boxes. The grey shaded box indicates the mutant class in which transport is expected to have been impacted.

[0023] Figure 4 shows testing for impurities in L-(1-^{14}C)arabinose, 0.1 μl of D-(1-^{14}C)galactose (lane 1), D-(1-^{14}C)xylose (lane 2) and L-(1-^{14}C)arabinose (lane 3) were separated on Whatman K6 TLC plates. The positions of migration of other related compounds on this TLC system are indicated.

[0024] Figure 5 shows the identification of L-(1-^{14}C)arabinose. Sample incubated without galactose dehydrogenase (lane 1); Sample incubated with galactose dehydrogenase (lane 2). The TLC was performed as described in Materials and Methods of Example 1.

[0025] Figure 6 shows the Eadie-Hofstee plots of L-arabinose transport. Initial rates of labeled L-arabinose uptake (0.065-592.2 mM) by L-arabinose grown cells were determined. A: *A. adeninivorans*. B: *D. hansenii* var. *fabryii*. C: *K. marxianus*. D: *P. guilliermondii*.

[0026] Figure 7 shows the fungal pathway for L-arabinose metabolism.

[0027] Figure 8 shows the relationship between *KmLAT1* and other transporters based on the neighbor-joining method (Saitou and Nei 1987).

[0028] Figure 9 shows the DNA (SEQ ID NO. 1) sequence of *Kluyveromyces marxianus KmLAT1*, and the predicted protein sequence (SEQ ID NO. 2).

[0029] Figure 10 shows the library insert from genomic *K. marxianus* DNA complements adapted *S. cerevisiae* for growth on L-arabinose. Cloning into the library expression vector is at the indicated *BamHI* restriction sites. The black block arrow is the L-arabinose transporter ORF responsible for complementation (Km*LAT1*). The block arrow with vertical stripes is the interrupted transporter ORF. The block arrow with the horizontal stripes is an un-related ORF ligated in place gratuitously during library construction. The *Sau3AI* restriction site where the transporter ORF was interrupted is shown. The primer used for PCR based genomic walking in *K. marxianus* is shown.

[0030] Figure 11 shows the growth curve of *S. cerevisiae* expressing *KmLAT1* (Δ), *GAL2* (■) or a control vector (♦) on 2% L-arabinose.

[0031] Figure 12 (A): Eadie-Hofstee plot of L-arabinose uptake by KmLat1 (♦) or Gal2 (■) expressed in *S. cerevisiae* grown on 2% L-arabinose. (B): Comparison of Eadie-

Hofstee plots of KmLat1 expressed in *S. cerevisiae* (♦) and wild type transport activity of *K. marxianus* (Δ) both grown on 2% L-arabinose.

[0032] Figure 13 shows the DNA (SEQ ID NO. 3) sequence of *Pichia guilliermondii PgLAT2*, and the predicted protein sequence (SEQ ID NO. 4).

[0033] Figure 14 shows the induction of L-arabinose transport in *P. guilliermondii*. Uptake of 13 mM labeled sugar was assayed for cells grown in minimal media containing 2% L-arabinose, D-galactose or D-xylose. White bars indicate labeled L-arabinose transport. Black bars indicate labeled galactose transport. Bars with vertical stripes indicate labeled xylose transport.

[0034] Figure 15 shows the sugar transport competition analysis in *P. guilliermondii* grown in minimal L-arabinose medium.

[0035] Figure 16 shows the transport kinetics of L-arabinose by the PgLat2 transporter expressed in *S. cerevisiae*. Open triangles indicate transport for wild type *P. guilliermondii* grown on L-arabinose. Black diamonds indicate transport for PgLat2 expressed in *S. cerevisiae* grown on L-arabinose.

[0036] Figure 17 shows comparison of the growth curves in 0.2% L-arabinose for *S. cerevisiae* cells expressing either Gal2p alone or both Gal2p and PgLat2. The maximum growth density and growth rate are significantly enhanced in the strain expressing both Gal2p and PgLat2.

DETAILED DESCRIPTION

[0037] There will now be shown and described systems, methods and tools for identifying single component, high efficiency arabinose transport proficient yeast strains. These identified yeast stains provide the source material for identifying and characterizing single component arabinose transporter genes identified herein. In particular, several highly efficient arabinose transport genes, for example *KmLAT1* and *PgLAT2*, are provided. The identified transporter genes, for example *KmLAT1* and *PgLAT2,* may be inserted into a host for metabolizing arabinose and, under conditions described herein, produce ethanol in high yield.

[0038] In the discussion below, parenthetical mention is made to publications from the references section for a discussion of related procedures that may be found useful from a perspective of one skilled in the art. This is done to demonstrate what is disclosed by way of non-limiting example.

[0039] The following definitions are provided to facilitate understanding of certain terms used frequently herein and are not meant to limit the scope of the present disclosure:

[0040] "Amino acid" refers to any of the twenty naturally occurring amino acids as well as any modified amino acid sequences. Modifications may include natural processes such as posttranslational processing, or may include chemical modifications which are known in the art. Modifications include but are not limited to: phosphorylation, ubiquitination, acetylation, amidation, glycosylation, covalent attachment of flavin, ADP-ribosylation, cross linking, iodination, methylation, and the like.

[0041] "Antibody" refers to a generally Y-shaped molecule having a pair of antigen binding sites, a hinge region and a constant region. Fragments of antibodies, for example an antigen binding fragment (Fab), chimeric antibodies, antibodies having a human constant region coupled to a murine antigen binding region, and fragments thereof, as well as other well known recombinant antibodies are included in this definition.

[0042] "Antisense" refers to polynucleotide sequences that are complementary to target "sense" polynucleotide sequence.

[0043] "Biomass" refers collectively to organic non-fossil material. "Biomass" in the present disclosure refers particularly to plant material that is used to generate fuel, such as ethanol. Examples of biomass includes but are not limited to corn fiber, dried distiller's grain, jatropha, manure, meat and bone meal, miscanthus, peat, plate waste, landscaping waste, maize, rice hulls, silage, stover, maiden grass, switchgrass, whey, and bagasse from sugarcane.

[0044] "Complementary" or "complementarity" refers to the ability of a polynucleotide in a polynucleotide molecule to form a base pair with another polynucleotide in a second polynucleotide molecule. For example, the sequence 5'-A-G-T-3' is complementary to the sequence 3'-T-C-A-5'. Complementarity may be partial, in which only some of the polynucleotides match according to base pairing, or complete, where all the polynucleotides match according to base pairing.

[0045] "Expression" refers to transcription and translation occurring within a host cell. The level of expression of a DNA molecule in a host cell may be determined on the basis of either the amount of corresponding mRNA that is present within the cell or the amount of DNA molecule encoded protein produced by the host cell (Sambrook et al., 1989, *Molecular cloning: A Laboratory Manual*, 18.1- 18.88).

[0046] "Fusion protein" refers to a first protein attached to a second, heterologous protein. Preferably, the heterologous protein is fused via recombinant DNA techniques, such

that the first and second proteins are expressed in frame. The heterologous protein may confer a desired characteristic to the fusion protein, for example, a detection signal, enhanced stability or stabilization of the protein, facilitated oligomerization of the protein, or facilitated purification of the fusion protein. Examples of heterologous proteins useful as fusion proteins include molecules having full-length or partial protein sequence of KmLat1 or PgLat2. Further examples include peptide tags such as histidine tag (6-His), leucine zipper, substrate targeting moieties, signal peptides, and the like. Fusion proteins are also meant to encompass variants and derivatives of KmLat1 or PgLat2 polypeptides that are generated by conventional site-directed mutagenesis and more modern techniques such as directed evolution, discussed infra.

[0047] "Genetically engineered" refers to any recombinant DNA or RNA method used to create a prokaryotic or eukaryotic host cell that expresses a protein at elevated levels, at lowered levels, or in a mutated form. In other words, the host cell has been transfected, transformed, or transduced with a recombinant polynucleotide molecule, and thereby been altered so as to cause the cell to alter expression of the desired protein. Methods and vectors for genetically engineering host cells are well known in the art; for example various techniques are illustrated in Current Protocols in Molecular Biology, Ausubel et al., eds. (Wiley & Sons, New York, 1988, and quarterly updates). Genetic engineering techniques include but are not limited to expression vectors, targeted homologous recombination and gene activation (see, for example, U.S. Patent No. 5,272,071 to Chappel) and trans activation by engineered transcription factors (see, for example, Segal et al., 1999, *Proc Natl Acad Sci USA* 96(6):2758-63).

[0048] "Homology" refers to a degree of similarity between polynucleotides, having significant effect on the efficiency and strength of hybridization between polynucleotide molecules. The term also refers to a degree of similarity between polypeptides. Two polypeptides having greater than or equal to about 60% similarity are presumptively homologous.

[0049] "Host," "Host cell" or "host cells" refers to cells expressing a heterologous polynucleotide molecule. The term "heterologous" means non-native. For instance, when a gene that is not normally expressed in an organism is introduced and expressed in that host organism, such an expression is heterologous. Host cells of the present disclosure express polynucleotides encoding *KmLAT1* or *PgLAT2* or a fragment thereof. Examples of suitable host cells useful in the present disclosure include, but are not limited to, prokaryotic and eukaryotic cells. Specific examples of such cells include bacteria of the genera *Escherichia,*

Bacillus, and *Salmonella,* as well as members of the genera *Pseudomonas, Streptomyces,* and *Staphylococcus*; fungi, particularly filamentous fungi such as *Trichoderma* and *Aspergillus*, *Phanerochaete chrysosporium* and other white rot fungi; also other fungi including *Fusaria*, molds, and yeast including *Saccharomyces* sp., *Pichia* sp., and *Candida* sp. and the like; plants e.g. *Arabidopsis*, cotton, barley, tobacco, potato, and aquatic plants and the like; SF9 insect cells (Summers and Smith, 1987, *Texas Agriculture Experiment Station Bulletin*, 1555), and the like. Other specific examples include mammalian cells such as human embryonic kidney cells (293 cells), Chinese hamster ovary (CHO) cells (Puck et al., 1958, *Proc. Natl. Acad. Sci. USA* 60, 1275-1281), human cervical carcinoma cells (HELA) (ATCC CCL 2), human liver cells (Hep G2) (ATCC HB8065), human breast cancer cells (MCF-7) (ATCC HTB22), human colon carcinoma cells (DLD-1) (ATCC CCL 221), Daudi cells (ATCC CRL-213), murine myeloma cells such as P3/NSI/1-Ag4-1 (ATCC TIB-18), P3X63Ag8 (ATCC TIB-9), SP2/0-Ag14 (ATCC CRL-1581) and the like. The most preferred host is *Saccharomyces cerevisiae.*

[0050] "Hybridization" refers to the pairing of complementary polynucleotides during an annealing period. The strength of hybridization between two polynucleotide molecules is impacted by the homology between the two molecules, stringency of the conditions involved, the melting temperature of the formed hybrid and the G:C ratio within the polynucleotides.

[0051] "Identity" refers to a comparison of two different DNA or protein sequences by comparing pairs of nucleic acid or amino acids within the two sequences. Methods for determining sequence identity are known. See, for example, computer programs commonly employed for this purpose, such as the Gap program (Wisconsin Sequence Analysis Package, Version 8 for Unix, Genetics Computer Group, University Research Park, Madison Wisconsin), that uses the algorithm of Smith and Waterman, 1981, *Adv. Appl. Math., 2:* 482-489.

[0052] "Isolated" refers to a polynucleotide or polypeptide that has been separated from at least one contaminant (polynucleotide or polypeptide) with which it is normally associated. For example, an isolated polynucleotide or polypeptide is in a context or in a form that is different from that in which it is found in nature.

[0053] "Nucleic acid sequence" refers to the order or sequence of deoxyribonucleotides along a strand of deoxyribonucleic acid. The order of these deoxyribonucleotides determines the order of amino acids along a polypeptide chain. The deoxyribonucleotide sequence thus codes for the amino acid sequence.

[0054] "Polynucleotide" refers to a linear sequence of nucleotides. The nucleotides may be ribonucleotides, or deoxyribonucleotides, or a mixture of both. Examples of polynucleotides in this context include single and double stranded DNA, single and double stranded RNA, and hybrid molecules having mixtures of single and double stranded DNA and RNA. The polynucleotides may contain one or more modified nucleotides.

[0055] "Protein," "peptide," and "polypeptide" are used interchangeably to denote an amino acid polymer or a set of two or more interacting or bound amino acid polymers.

[0056] "Purify," or "purified" refers to a target protein makes up for at least about 90% of a composition. In other words, it refers to a target protein that is free from at least 5-10% of contaminating proteins. Purification of a protein from contaminating proteins may be accomplished using known techniques, including ammonium sulfate or ethanol precipitation, acid precipitation, heat precipitation, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography, size-exclusion chromatography, and lectin chromatography. Various protein purification techniques are illustrated in *Current Protocols in Molecular Biology*, Ausubel et al., eds. (Wiley & Sons, New York, 1988, and quarterly updates).

[0057] "Selectable marker" refers to a marker that identifies a cell as having undergone a recombinant DNA or RNA event. Selectable markers include, for example, genes that encode antimetabolite resistance such as the DHFR protein that confers resistance to methotrexate (Wigler et al, 1980, *Proc Natl Acad Sci* USA 77:3567; O'Hare et al., 1981, *Proc Natl Acad Sci USA*, 78:1527), the GPT protein that confers resistance to mycophenolic acid (Mulligan & Berg, 1981, *PNAS USA*, 78:2072), the neomycin resistance marker that confers resistance to the aminoglycoside G-418 (Calberre-Garapin et al., 1981, *J Mol Biol*, 150:1), the Hygro protein that confers resistance to hygromycin B (Santerre et al., 1984, *Gene* 30:147), and the Zeocin™ resistance marker (Invitrogen). In addition, the herpes simplex virus thymidine kinase, hypoxanthine-guanine phosphoribosyltransferase and adenine phosphoribosyltransferase genes may be employed in tk^-, $hgprt^-$ and $aprt^-$ cells, respectively.

[0058] "Transform" means the process of introducing a gene into a host cell. The gene may be foreign in origin, but the gene may also derive from the host. A transformed host cell is termed a "transformant." The introduced gene may be integrated onto the chromosome of the host, or the gene may remain on a stand-alone vector independent of the host chromosomes.

[0059] "Variant", as used herein, means a polynucleotide or polypeptide molecule that differs from a reference molecule. Variants may include nucleotide changes that result in amino acid substitutions, deletions, fusions, or truncations in the resulting variant polypeptide when compared to the reference polypeptide.

[0060] "Vector," "extra-chromosomal vector" or "expression vector" refers to a first polynucleotide molecule, usually double-stranded, which may have inserted into it a second polynucleotide molecule, for example a foreign or heterologous polynucleotide. The heterologous polynucleotide molecule may or may not be naturally found in the host cell, and may be, for example, one or more additional copy of the heterologous polynucleotide naturally present in the host genome. The vector is adapted for transporting the foreign polynucleotide molecule into a suitable host cell. Once in the host cell, the vector may be capable of integrating into the host cell chromosomes. The vector may optionally contain additional elements for selecting cells containing the integrated polynucleotide molecule as well as elements to promote transcription of mRNA from transfected DNA. Examples of vectors useful in the methods disclosed herein include, but are not limited to, plasmids, bacteriophages, cosmids, retroviruses, and artificial chromosomes.

[0061] For purpose of this disclosure, unless otherwise stated, the techniques used may be found in any of several well-known references, such as: *Molecular Cloning: A Laboratory Manual* (Sambrook et al. (1989) Molecular cloning: A Laboratory Manual), *Gene Expression Technology* (Methods in Enzymology, Vol. 185, edited by D. Goeddel, 1991 Academic Press, San Diego, CA), "Guide to Protein Purification" in *Methods in Enzymology* (M.P. Deutshcer, 3d., (1990) Academic Press, Inc.), *PCR Protocols: A Guide to Methods and Applications* (Innis et al. (1990) Academic Press, San Diego, CA), Culture of Animal Cells: A Manual of Basic Technique, 2nd ed. (R.I. Freshney (1987) Liss, Inc., New York, NY), *and Gene Transfer and Expression Protocols*, pp 109-128, ed. E.J. Murray, The Humana Press Inc., Clifton, N.J.).

[0062] Unless otherwise indicated, the term "yeast," "yeast strain" or "yeast cell" refers to baker's yeast, *Saccharomyces cerevisiae*. Other yeast species, such as *Kluyveromyces marxianus* or *Pichia guilliermondii*, are referred to as non-conventional yeast in this disclosure. Yeast strains of *S. cerevisiae* and plasmids used for this disclosure are listed in Table 1. The yeast *Kluyveromyces marxianus* CBS-1089 is obtained from the Centraalbureau voor Schimmel cultures (CBS) collection. *Pichia guilliermondii* NRRL Y-2075 is obtained from the Agricultural Research Service Culture Collection (NRRL).

Table 1. *S. cerevisiae* Strains and Plasmids Used in The Disclosure

Strain	Genotype	Plasmids
BFY002	*MAT*a *ura3-52 trp1-Δ63 his3-Δ200 leu2-Δ1 yhr104w::LEU2*	
BFY507	*MAT*a *ura3-52 trp1-Δ63 his3-Δ200 leu2-Δ1 yhr104w::LEU2* adapted for growth on L-arabinose	p138, p42
BFY518	same as BFY507	p138
BFY566	same as BFY518	p138, p171
BFY590	same as BFY518 *gal2Δ::HIS3*	p138
BFY597	same as BFY590	p138, p42
BFY598	same as BFY590	p138, p187
BFY605	same as BFY590	p138, p244
BFY604	*MATa his3Δ1 leu2Δ0 ura3Δ0 met15Δ0 gal80Δ::G418 trp1Δ*	
BFY607	Same as BFY604	p138
BFY609	Same as BFY607	p12, p13, p138
BFY612	Same as BFY607	p12, p138, p204
BFY625	Same as BFY609, adapted for growth on L-arabinose	p12, p13, p138
BFY626	Same as BFY612, adapted for growth on L-arabinose	p12, p138, p204
BFY057	*MATa his3D1 leu2D0 ura3D0 met15D0 gal80D::G418 yhr104w::LEU2*	
BFY534	same as BFY057	p144, p165
BFY535	same as BFY057	p144, p13

Plasmid	Marker and expressed genes
p12, p13	*URA3* or *HIS3* control vectors respectively
p42	*URA3*, *GAL2* over-expression
p138	*TRP1*, *B. subtilis araA*, *E. coli araB*, *E. coli araD*
p171	*HIS3*, 8.8 kb *K. marxianus* genomic DNA fragment
p187	*URA3*, *KmLAT1* over-expression plasmid
p204	*HIS3*, *PgLAT2* over-expression plasmid
p244	*URA3*, Pg*LAT2* over-expression plasmid
p144	*E. coli araB,D*; *B. subtilis araA* in pBFY012
p165	*HIS3*, *GAL2* over-expression

[0063] Yeast strains may be grown on liquid or solid media with 2% agar for solid media. Where appropriate, some amino acids or nucleic acids are purposely left out from the media for plasmid maintenance. Growth conditions are typically 30°C unless otherwise indicated, with shaking in liquid cultures. An anaerobic condition is generally more favorable to metabolize the various sugars to ethanol.

[0064] Yeast cells may be grown in rich media YPD or minimum media conventionally used in the field. YPD medium contains about 1% yeast extract, 2% peptone and 2% dextrose. Yeast minimum media typically contains 0.67% of yeast nitrogen base ("YNB") without amino acids supplemented with appropriate amino acids or purine or pyrimidine bases. A number of sugar, typically at 2% unless otherwise indicated, may be used as carbon source, including glucose (dextrose), galactose, maltose or L-arabinose among others. Adaptation for growth on L-arabinose is performed as described in Becker and Boles (2003) with modifications as detailed in Example 3.

[0065] Over-expression plasmids are constructed by cloning the gene for over-expression downstream of the *S. cerevisiae PGK1* or *TDH3* promoter in a 2μ-based vector. Other like overexpression plasmid for expression of a gene may also be used as is known in the art. Construction of DNA library is detailed in the Examples.

[0066] *E. coli* cells may be grown in LB liquid media or on LB agar plates supplemented with ampicillin at 100 μg/ml as needed. Transformation of *E. coli* DH5α is by electrotransformation according to a protocol by Invitrogen (Invitrogen 11319-019). After transformation, the bacterial cells are plated on LB plates containing 100 μg/ml ampicillin for selection. Transformation of *S. cerevisiae* was performed using a DMSO-enhanced lithium-acetate procedure as described with the following modifications (Hill et al., 1991). Cells are harvested and initially washed in water. 600 μl of PEG4000 solution is added and 70 μl DMSO is added just prior to heat shocking. Cells are heat-shocked for 15 min at 42°C and the last wash step is skipped. Cells are resuspended in 10 mM TE solution and plated.

[0067] Yeast DNA is isolated using the Easy DNA kit according to manufacturer's protocol (Invitrogen, K1800-01). DNA manipulations and library construction are performed as described in Molecular Cloning: A Laboratory Manual (1989), except otherwise specifically indicated in this disclosure. Plasmids are cured from yeast by growing the strain in rich non-selective media overnight followed by plating on non-selective media. Isolated colonies are replica plated to screen for loss of selective markers. Plasmid rescue is performed by transforming isolated yeast DNA into *E. coli* followed by isolation and

characterization. *E. coli* plasmid isolation is accomplished using plasmid spin mini-prep kit according to the manufacturer's manual (Qiagen, 27106). PCR-based chromosomal walking is performed using the Universal GenomeWalker Kit as described (BD Biosciences, K1807-1).

[0068] For transport assays, cells may be grown in minimal media supplemented with 20 g/L of L-arabinose. Cells are collected in mid-growth and washed twice before suspension in water at 30 mg/ml. Uptake of L-(1-^{14}C)arabinose (54 mCi/mmol, Moravek Biochemicals Inc.) or D-(1-^{14}C) galactose (57 mCi/mmol, Amersham Biosciences) is measured as previously described by Stambuk et al. (2003). Assays are performed in 30 seconds to maintain initial rates after appropriate experiments to ensure uptake is linear for at least 1 minute. Transport activity is described as nano-moles of labeled sugar transported per mg cell dry weight per minute. Inhibition and competition assays are performed as previously described by Stambuk et al. (2003).

[0069] Embodiments described herein provide systems, tools and methods for the identification of yeast strains efficient in utilization of L-arabinose as a sugar source. In some aspects the identified yeast strains are amenable to genetic and biochemical manipulation. In further aspects certain identified yeast strains are shown to include a single component responsible for transporting arabinose into the yeast. These identified single components correspond to arabinose transporter genes that can then be identified and cloned (as described herein).

[0070] Identification of arabinose utilizing or fermenting yeast strains in accordance with embodiments described herein includes: (1) selecting yeast strains for screening that are not pathogenic, did not, or rarely, form hyphae and grow primarily as a single cell(s); (2) growing selected yeast strain(s) or minimal growth medium containing from 0.2 to 2% L-arabinose as a sole sugar source; (3) classifying the yeast strain for its capacity to utilize L-arabinose (typically via periodic growth measurements of optical density); (4) determining if a strain that shows growth on arabinose is amenable to genetic and biochemical manipulation, i.e., only strains that are amenable to manipulation are useful for ultimate cloning and identification of the arabinose transporter gene; and (5) perform arabinose transport assays on the selected yeast strains that grow on L-arabinose to identify single affinity arabinose transporter strains of yeast. Single component high affinity arabinose transporter systems are targeted for further identification of arabinose transporter genes (see below). As discussed in detail in Example 1, at least two strains of yeast were identified as containing high affinity, single component transporter systems, *K. marxianus* and *P. guilliermondii*. As detailed

below and in Examples 2-9, the genes identified are *KmLATl* and *PgLAT2*, other like genes are identified able using the novel methods and tools described herein, each of which is within the scope of this disclosure.

[0071] Briefly, and by way of example, the *KmLAT1* transporter gene was identified using complementation of a strain adapted for growth on L-arabinose as described above. This strain was able to utilize L-arabinose only if a suitable transporter was present. After adaptation, the plasmid carrying the *GAL2* transporter was cured (removed) from the strain rendering this strain unable to grow on L-arabinose. A genomic library was then introduced and colonies selected that regained the ability to grow on L-arabinose. The genomic fragment isolated in this manner contained the ORF for the *KmLAT1* transporter.

[0072] The *PgLAT2* transporter gene was identified using differential display (differential expression). The identified strain using methods described herein was separately grown on L-arabinose and a control sugar (D-xylose). Total RNA was isolated from cells grown on the two sugars and analyzed to detect MRNAs that were only expressed or much more highly expressed when the cells were grown on L-arabinose compared to the samples from the MRNA from cells grown on the control sugar. The gene fragments corresponding to the differentially expressed genes were sequenced and the complete gene was then isolated from the genome of *P.guilliermondii* by genome walking using oligonucleotide primers designed from the sequences of the fragments. Complete sequence of the gene was then determined and the gene was engineered for expression in *S. cerevisiae.*

[0073] Note that similar techniques can be used to identify transporter genes from other like L-arabinose fermenting yeast stems. Examples 1-10 illustrate various of the methods described herein.

[0074] Sequencing results showed that the *KmLAT1* gene contains an open reading frame ("ORF") of 1668 bp in length. The predicted amino acid sequence of KmLat1 shares homology with high-affinity glucose transporters, in particular, with *HGT1* from *K. lactis* (Table 2). KmLat1 transporter shows a much higher sequence similarity with high-affinity glucose transporters from non-conventional yeast than with transporter proteins encoded by the bacterial *araE* gene or hexose transporters from *S. cerevisiae* (See Fig. 8).

Table 2. Properties and similarities of KmLat1 to other sugar transporters.

gene	Predicted protein (no. of aa/no. of kDa)	pI of protein	Predicted transmembrane regions	Degree of identity (%) /similarity (%)	Organism	Putative function of gene product
KmLAT1	556 / 61.3	8.22	12	-	*K.marxianus*[1]	L-arabinose transporter
KlHGT1	551 / 60.8	5.76	12	77 / 89	*K. lactis*[2]	high affinity glucose transporter
AEL042Cp	547 / 59.8	8.82	12	65 / 82	*A. gossypii*[3]	putative hexose transporter
DEHA0E01738g	545 / 61.1	5.55	12	52 / 70	*D. hansenii*[4]	hexose transporter
CaHGT1	545 / 60.7	8.05	12-13	50 / 71	*C. albicans*[5]	putative hexose transporter
CaHGT2	545 / 60.4	8.48	12-14	51 / 71	*C. albicans*[6]	putative hexose transporter

Accession numbers: 1: Not yet assigned, 2: 1346290, 3: AEL042C, 4: DEHA0E01738g, 5: CAA76406, 6: orf19.3668

[0075] Transmembrane regions predicted for KmLat1 and PgLat2 by the software Tmpred show 12 transmembrane regions with a larger intercellular loop between regions 6 and 7 (Fig. 2) (See Hofmann et al, 1993), typical of Gal2 and other yeast sugar transporters having 10-12 transmembrane regions (*See e.g.*, Alves-Araujo et al., 2004; Day et al., 2002; Kruckeberg et al., 1996; Pina et al. ,2004; and Weierstall et al. 1999).

[0076] Like other members of the transporter family, and in particular sugar transporters, KmLat1 and PgLat2 polypeptides are useful in facilitating the uptake of various sugar molecules into the cells. It is envisioned that KmLat1 or PgLat2 polypeptides may be used for other purposes, for example, in analytical instruments or other processes where uptake of sugar is required. KmLat1 or PgLat2 polypeptides may be used alone or in combination with one or more other transporters to facilitate the movement of molecules across a membrane structure, which function may be modified by one skilled in the relevant art, all of which are within the scope of the present disclosure.

[0077] KmLat1 polypeptides may include isolated polypeptides having an amino acid sequence as shown below in Example 2; and in SEQ ID NO:2, as well as variants and derivatives, including fragments, having substantial sequence similarity to the amino acid sequence of SEQ ID NO:2 and that retain any of the functional activities of KmLat1. PgLat2 polypeptides may include isolated polypeptides having an amino acid sequence as shown

below in Example 5; and in SEQ ID NO:4, as well as variants and derivatives, including fragments, having substantial sequence similarity to the amino acid sequence of SEQ ID NO:4 and that retain any of the functional activities of PgLat2. The functional activities of the KmLat1 or PgLat2 polypeptides include but are not limited to transport of L-arabinose across cell membrane. Such activities may be determined, for example, by subjecting the variant, derivative, or fragment to a arabinose transport assay as detailed in Example 4.

[0078] Variants and derivatives of KmLat1 or PgLat2 include, for example, KmLat1 or PgLat2 polypeptides modified by covalent or aggregative conjugation with other chemical moieties, such as glycosyl groups, polyethylene glycol (PEG) groups, lipids, phosphate, acetyl groups, and the like.

[0079] The amino acid sequence of these KmLat1 or PgLat2 variants or derivatives is preferably at least about 60% identical, more preferably at least about 70% identical, still more preferably at least 80% identical, or in some embodiments at least about 90%, 95%, 96%, 97%, 98%, or 99% identical, to the KmLat1 and PgLat2 amino acid sequences of SEQ ID NO: 2 and SEQ ID NO: 4, respectively. The percentage sequence identity, also termed homology (see definition above) may be readily determined, for example, by comparing the two polypeptide sequences using any of the computer programs commonly employed for this purpose, such as the Gap program (Wisconsin Sequence Analysis Package, Version 8 for Unix, Genetics Computer Group, University Research Park, Madison Wisconsin), which uses the algorithm of Smith and Waterman, 1981, *Adv. Appl. Math.* 2: 482-489.

[0080] Variants and derivatives of the KmLat1 or PgLat2 polypeptides may further include, for example, fusion proteins formed of a KmLat1 or PgLat2 polypeptide and another polypeptide. Fusion protein may be formed between a fragment of the KmLat1 or PgLat2 polypeptide and another polypeptide, such that the fusion protein may retain all or only part of the activities normally performed by the full-length KmLat1 or PgLat2 polypeptide. Preferred polypeptides for constructing the fusion protein include those that facilitate purification or oligomerization, or those that enhance KmLat1 or PgLat2 stability and/or transport capacity or transport rate for sugars, especially for arabinose. Preferred polypeptides may also include those that gain enhanced transport capability when fused with KmLat1, PgLat2 or fragments thereof.

[0081] KmLat1 or PgLat2 variants and derivatives may contain conservatively substituted amino acids, meaning that one or more amino acid may be replaced by an amino acid that does not alter the secondary and/or tertiary structure of the polypeptide. Such substitutions may include the replacement of an amino acid, by a residue having similar

physicochemical properties, such as substituting one aliphatic residue (Ile, Val, Leu, or Ala) for another, or substitutions between basic residues Lys and Arg, acidic residues Glu and Asp, amide residues Gln and Asn, hydroxyl residues Ser and Tyr, or aromatic residues Phe and Tyr. Phenotypically silent amino acid exchanges are described more fully in Bowie *et al.,* 1990. In addition, functional KmLat1 or PgLat2 polypeptide variants include those having amino acid substitutions, deletions, or additions to the amino acid sequence outside functional regions of the protein.

[0082] The KmLat1 or PgLat2 polypeptides may be provided in an isolated form, or in a substantially purified form. The polypeptides may be recovered and purified from recombinant cell cultures by known methods, including, for example, ammonium sulfate or ethanol precipitation, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography, and lectin chromatography. Preferably, protein chromatography is employed for purification.

[0083] A preferred form of KmLat1 or PgLat2 polypeptides is that of recombinant polypeptides expressed by suitable hosts. In one preferred embodiment, when heterologous expression of KmLat1 or PgLat2 is desired, the coding sequences of KmLat1 or PgLat2 may be modified in accordance with the codon usage of the host. Such modification may result in increase protein expression of a foreign in the host. Furthermore, the hosts may simultaneously produce other transporters such that multiple transporters are expressed in the same cell, wherein the different transporters may form oligomers to transport the same sugar. Alternatively, the different transporters may function independently to transport different sugars. Such recombinant cells may be useful in crude fermentation processing or in other industrial processing.

[0084] KmLat1 or PgLat2 polypeptides may be fused to heterologous polypeptides to facilitate purification. Many available heterologous peptides (peptide tags) allow selective binding of the fusion protein to a binding partner. Non-limiting examples of peptide tags include 6-His, thioredoxin, hemaglutinin, GST, and the OmpA signal sequence tag. A binding partner that recognizes and binds to the heterologous peptide may be any molecule or compound, including metal ions (for example, metal affinity columns), antibodies, antibody fragments, or any protein or peptide that preferentially binds the heterologous peptide to permit purification of the fusion protein.

[0085] KmLat1 or PgLat2 polypeptides may be modified to facilitate formation of KmLat1 or PgLat2 oligomers. For example, KmLat1 polypeptides may be fused to peptide

moieties that promote oligomerization, such as leucine zippers and certain antibody fragment polypeptides, for example, Fc polypeptides. Techniques for preparing these fusion proteins are known, and are described, for example, in WO 99/31241 and in Cosman et.al., 2001. Fusion to an Fc polypeptide offers the additional advantage of facilitating purification by affinity chromatography over Protein A or Protein G columns. Fusion to a leucine-zipper (LZ), for example, a repetitive heptad repeat, often with four or five leucine residues interspersed with other amino acids, is described in Landschultz et al., 1988.

[0086] It is also envisioned that an expanded set of variants and derivatives of *KmLAT1* or *PgLAT2* polynucleotides and/or polypeptides may be generated to select for useful molecules, where such expansion is achieved not only by conventional methods such as site-directed mutagenesis but also by more modern techniques, either independently or in combination.

[0087] Site-directed-mutagenesis is considered an informational approach to protein engineering and may rely on high-resolution crystallographic structures of target proteins for specific amino acid changes (van den Burg et al. 1998). For example, modification of the amino acid sequence of KmLat1 or PgLat2 polypeptides may be accomplished as is known in the art, such as by introducing mutations at particular locations by oligonucleotide-directed mutagenesis. Site-directed-mutagenesis may also take advantage of the recent advent of computational methods for identifying site-specific changes for a variety of protein engineering objectives (Hellinga, 1998).

[0088] The more modern techniques include, but are not limited to, non-informational mutagenesis techniques (referred to generically as “directed evolution”). Directed evolution, in conjunction with high-throughput screening, allows testing of statistically meaningful variations in protein conformation (Arnold, 1998). Directed evolution technology may include diversification methods similar to that described by Crameri et al. (1998), site-saturation mutagenesis, staggered extension process (StEP) (Zhao et al., 1998), and DNA synthesis/reassembly (U.S. Patent 5,965,408).

[0089] Fragments of the KmLat1 or PgLat2 polypeptide may be used, for example, to generate specific anti-KmLat1 or PgLat2 antibodies. Using known selection techniques, specific epitopes may be selected and used to generate monoclonal or polyclonal antibodies. Such antibodies have utility in the assay of KmLat1 or PgLat2 activity as well as in purifying recombinant KmLat1 or PgLat2 polypeptides from genetically engineered host cells.

[0090] The disclosure also provides polynucleotide molecules encoding the KmLat1 or PgLat2 polypeptides discussed above. *KmLAT1* or *PgLAT2* polynucleotide molecules of

the disclosure include polynucleotide molecules having the nucleic acid sequence shown in SEQ ID NO:1 and SEQ ID NO:3, respectively; polynucleotide molecules that hybridize to the nucleic acid sequence of SEQ ID NO:1 and SEQ ID NO:3, respectively, under high stringency hybridization conditions (for example, 42°, 2.5 hr., 6X SCC, 0.1%SDS); and polynucleotide molecules having substantial nucleic acid sequence identity with the nucleic acid sequence of SEQ ID NO:1 and SEQ ID NO:3, respectively.

[0091] The *KmLAT1* or *PgLAT2* polynucleotide molecules of the disclosure are preferably isolated molecules encoding the KmLat1 or PgLat2 polypeptide having an amino acid sequence as shown in SEQ ID NO:2 and SEQ ID NO:4, respectively, as well as derivatives, variants, and useful fragments of the *KmLAT1* or *PgLAT2* polynucleotide. The *KmLAT1* or *PgLAT2* polynucleotide sequence may include deletions, substitutions, or additions to the nucleic acid sequence of SEQ ID NO:1 and SEQ ID NO:3, respectively.

[0092] The *KmLAT1* or *PgLAT2* polynucleotide molecule of the disclosure may be cDNA, chemically synthesized DNA, DNA amplified by PCR, RNA, or combinations thereof. Due to the degeneracy of the genetic code, two DNA sequences may differ and yet encode identical amino acid sequences. The present disclosure thus provides an isolated polynucleotide molecule having a *KmLAT1* or *PgLAT2* nucleic acid sequence encoding KmLat1 or PgLat2 polypeptide, wherein the nucleic acid sequence encodes a polypeptide having the complete amino acid sequences as shown in SEQ ID NO:2 and SEQ ID NO:4, respectively, or variants, derivatives, and fragments thereof.

[0093] The *KmLAT1* or *PgLAT2* polynucleotides of the disclosure have a nucleic acid sequence that is at least about 60% identical to the nucleic acid sequence shown in SEQ ID NO:1 and SEQ ID NO:3, respectively, in some embodiments at least about 70% identical to the nucleic acid sequence shown in SEQ ID NO:1 and SEQ ID NO:3, respectively, in other embodiments at least about 80% identical to the nucleic acid sequence shown in SEQ ID NO:1 and SEQ ID NO:3, respectively and in other embodiments at least about 90%, 95%, 96%, 97%, 98%, 99%, identical to the nucleic and sequence shown in SEQ ID NO: 1 and SEQ ID NO: 3, respectively. Nucleic acid sequence identity is determined by known methods, for example by aligning two sequences in a software program such as the BLAST program (Altschul, S.F et al. (1990) J. Mol. Biol. 215:403-410, from the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/BLAST/).

[0094] The *KmLAT1* or *PgLAT2* polynucleotide molecules of the disclosure also include isolated polynucleotide molecules having a nucleic acid sequence that hybridizes under high stringency conditions (as defined above) to a the nucleic acid sequence shown in

SEQ ID NO:1 and SEQ ID NO:3, respectively. Hybridization of the polynucleotide is to at least about 15 contiguous nucleotides, or at least about 20 contiguous nucleotides, and in other embodiments at least about 30 contiguous nucleotides, and in still other embodiments at least about 100 contiguous nucleotides of the nucleic acid sequence shown in SEQ ID NO:1 and SEQ ID NO:3, respectively.

[0095] Useful fragments of the *KmLAT1* or *PgLAT2* polynucleotide molecules described herein, include probes and primers. Such probes and primers may be used, for example, in PCR methods to amplify and detect the presence of *KmLAT1* or *PgLAT2* polynucleotides *in vitro*, as well as in Southern and Northern blots for analysis of *KmLAT1* or *PgLAT2*. Cells expressing the *KmLAT1* or *PgLAT2* polynucleotide molecules may also be identified by the use of such probes. Methods for the production and use of such primers and probes are known. For PCR, 5' and 3' primers corresponding to a region at the termini of the *KmLAT1* or *PgLAT2* polynucleotide molecule may be employed to isolate and amplify the *KmLAT1* or *PgLAT2* polynucleotide using conventional techniques.

[0096] Other useful fragments of the *KmLAT1* or *PgLAT2* polynucleotides include antisense or sense oligonucleotides comprising a single-stranded nucleic acid sequence capable of binding to a target *KmLAT1* or *PgLAT2* mRNA (using a sense strand), or DNA (using an antisense strand) sequence.

[0097] The present disclosure also provides vectors containing the polynucleotide molecules, as well as host cells transformed with such vectors. Any of the polynucleotide molecules of the disclosure may be contained in a vector, which generally includes a selectable marker and an origin of replication, for propagation in a host. The vectors may further include suitable transcriptional or translational regulatory sequences, such as those derived from a mammalian, fungal, bacterial, viral, or insect genes, operably linked to the *KmLAT1* or *PgLAT2* polynucleotide molecule. Examples of such regulatory sequences include transcriptional promoters, operators, or enhancers, mRNA ribosomal binding sites, and appropriate sequences which control transcription and translation. Nucleotide sequences are operably linked when the regulatory sequence functionally relates to the DNA encoding the target protein. Thus, a promoter nucleotide sequence is operably linked to a *KmLAT1* or *PgLAT2* DNA sequence if the promoter nucleotide sequence directs the transcription of the *KmLAT1* or *PgLAT2* sequence.

[0098] Selection of suitable vectors for the cloning of *KmLAT1* or *PgLAT2* polynucleotide molecules encoding the target KmLat1 or PgLat2 polypeptides of this disclosure will depend upon the host cell in which the vector will be transformed, and, where

applicable, the host cell from which the target polypeptide is to be expressed. Suitable host cells for expression of KmLat1 or PgLat2 polypeptides include prokaryotes, yeast, and higher eukaryotic cells, each of which is discussed below.

[0099] The KmLat1 or PgLat2 polypeptides to be expressed in such host cells may also be fusion proteins that include sequences from other proteins. As discussed above, such regions may be included to allow, for example, enhanced functionality, improved stability, or facilitated purification of the KmLat1 or PgLat2 polypeptide. For example, a nucleic acid sequence encoding a peptide that binds strongly to arabinose may be fused in-frame to the transmembrane sequence of the KmLat1 or PgLat2 polypeptides so that the resulting fusion protein binds arabinose and transports the sugar across the cell membrane at a higher rate than the KmLat1 or PgLat2 transporter.

[00100] Suitable host cells for expression of target polypeptides include prokaryotes, yeast, and higher eukaryotic cells. Suitable prokaryotic hosts to be used for the expression of these polypeptides include bacteria of the genera *Escherichia, Bacillus,* and *Salmonella,* as well as members of the genera *Pseudomonas, Streptomyces,* and *Staphylococcus.*

[00101] Expression vectors for use in prokaryotic hosts generally comprise one or more phenotypic selectable marker genes. Such genes encode, for example, a protein that confers antibiotic resistance or that supplies an auxotrophic requirement. A wide variety of such vectors are readily available from commercial sources. Examples include pSPORT vectors, pGEM vectors (Promega, Madison, WI), pPROEX vectors (LTI, Bethesda, MD), Bluescript vectors (Stratagene), and pQE vectors (Qiagen).

[00102] *KmLAT1* or *PgLAT2* may also be expressed in yeast host cells from genera including *Saccharomyces, Pichia,* and *Kluveromyces.* Preferred yeast host is *S. cerevisiae.* Yeast vectors will often contain an origin of replication sequence from a 2μ yeast plasmid for high copy vectors and a CEN sequence for a low copy number vector. Other sequences on a yeast vector may include an autonomously replicating sequence (ARS), a promoter region, sequences for polyadenylation, sequences for transcription termination, and a selectable marker gene. Vectors replicable in both yeast and *E. coli* (termed shuttle vectors) are preferred. In addition to the above-mentioned features of yeast vectors, a shuttle vector will also include sequences for replication and selection in *E. coli.*

[00103] Insect host cell culture systems may also be used for the expression of KmLat1 or PgLat2 polypeptides. The target polypeptides are preferably expressed using a baculovirus expression system, as described, for example, in the review by Luckow and Summers, 1988.

[00104] The choice of a suitable expression vector for expression of KmLat1 or PgLat2 polypeptides will depend upon the host cell to be used. Examples of suitable expression vectors for *E. coli* include pET, pUC, and similar vectors as is known in the art. Preferred vectors for expression of the KmLat1 or PgLat2 polypeptides include the shuttle plasmid pIJ702 for *Streptomyces lividans*, pGAPZalpha-A, B, C and pPICZalpha-A, B, C (Invitrogen) for *Pichia pastoris*, and pFE-1 and pFE-2 for filamentous fungi and similar vectors as is known in the art. The vectors preferred by expression in *S. cerevisiae* are listed in Table 1.

[00105] Modification of a *KmLAT1* or *PgLAT2* polynucleotide molecule to facilitate insertion into a particular vector (for example, by modifying restriction sites), ease of use in a particular expression system or host (for example, using preferred host codons), and the like, are known and are contemplated for use as descibed herein. Genetic engineering methods for the production of KmLat1 or PgLat2 polypeptides include the expression of the polynucleotide molecules in cell free expression systems, in host cells, in tissues, and in animal models, according to known methods.

[00106] This disclosure also provides reagents, compositions, and methods that are useful for analysis of KmLat1 or PgLat2 activity and for assessing the amount and rate of arabinose transport.

[00107] The KmLat1 or PgLat2 polypeptides of the present disclosure, in whole or in part, may be used to raise polyclonal and monoclonal antibodies that are useful in purifying KmLat1 or PgLat2, or detecting KmLat1 or PgLat2 polypeptide expression, as well as a reagent tool for characterizing the molecular actions of the KmLat1 or PgLat2 polypeptide. Preferably, a peptide containing a unique epitope of the KmLat1 or PgLat2 polypeptide is used in preparation of antibodies, using conventional techniques. Methods for the selection of peptide epitopes and production of antibodies are known. See, for example, *Antibodies: A Laboratory Manual*, Harlow and Land (eds.), 1988 Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.; *Monoclonal Antibodies, Hybridomas: A New Dimension in Biological Analyses*, Kennet *et al.* (eds.), 1980 Plenum Press, New York.

[00108] Agents that modify, for example, increase or decrease, KmLat1 or PgLat2 transport of arabinose or other sugars may be identified, for example, by the transport assay described in Example 5. Performing the transport assay in the presence or absence of a test agent permits screening of such agents.

[00109] The KmLat1 or PgLat2 transport activity is determined in the presence or absence of a test agent and then compared. For instance, a lower KmLat1 transport activity

in the presence of the test agent, than in the absence of the test agent, indicates that the test agent has decreased the activity of the KmLat1. Stimulators and inhibitors of KmLat1 or PgLat2 may be used to augment, inhibit, or modify KmLat1 or PgLat2 transport activity, and therefore may have potential industrial uses as well as potential use in further elucidation of the molecular actions of KmLat1 or PgLat2.

[00110] The KmLat1 or PgLat2 polypeptide of the disclosure is an effective arabinose transporter. In the methods of the disclosure, the sugar transporting effects of KmLat1 or PgLat2 are achieved by mixing cells expressing KmLat1 or PgLat2 with pure sugar or sugar-containing biomass. KmLat1 or PgLat2 may also be used in a cell-free system. KmLat1 or PgLat2 may be used under other conditions, for example, at elevated temperatures or under acidic pH. Other methods of using KmLat1 or PgLat2 to transport sugar, especially arabinose, for fermentation, are envisioned to be within the scope of the present disclosure. KmLat1 or PgLat2 polypeptides may be used in any known application currently utilizing a sugar transporter, all of which are within the scope of the disclosure. It should be noted that the KmLat1 and PgLat2 polypeptides are also capable of transporting other sugars, including D-xylose, and thus may have utility for transport of other biomass-derived sugars.

[00111] It is also shown in this disclosure that Gal2p is an effective L-arabinose transporter at high concentrations of arabinose whereas KmLat1 or PgLat2 may be more effective at different concentrations of L-arabinose. Combination of the Gal2p and the two new transporters from non-conventional yeast may be employed to provide complementary transport into *S. cerevisiae* of L-arabinose down to very low residual concentration of arabinose.

[00112] It is shown that combinatorial expression of Gal2p and PgLat2 may enhance the overall rate and extent of arabinose utilization by recombinant *S. cerevisiae* cells expressing these transporters. As shown in Example 9, the doubling time for *S. cerevisiae* strain expressing both PgLat2 and Gal2p is shorter than *S. cerevisiae* cells expressing Gal2p alone (15 hours *vs*. 19 hours), suggesting that L-arabinose uptake may be enhanced by the synergistic effect of PgLat2 and Gal2p in these cells. Moreover, the PgLat2 expressing strain appears to grow to a higher overall optical density at saturation, suggesting that this strain was able to utilize the carbon source (L-arabinose) in the medium more completely. This hypothesis is supported by HPLC analysis of the final culture media (Table 3) which indicates that there is significantly less residual L-arabinose in the culture of cells expressing Gal2p and PgLat2 than in the culture of those expressing Gal2p alone. Thus, heterologous expression of either or both KmLat1 and PgLat2 in *S. cerevisiae* may enhance arabinose

utilization by facilitating arabinose transport when the concentration of arabinose is relatively low.

Table 3. Doubling times and HPLC Measurement of Residual Arabinose Concentration

Flask	Transporters Expressed	Doubling Time (hours)	*Final* OD_{600}	L-arabinose (g/L) by HPLC
1	Gal2p only	19.2	0.72	0.68
2		18.6	0.72	0.67
3	Gal2p + PgLat2	15.0	0.85	0.49
4		14.8	0.85	0.48

*starting L-arabinose concentration 1.89 g/L and media without L-arabinose had an undetectable level (<0.1 g/L).

[00113] Note that yeast strains BFY013, BFY534, BFY598 and BFY626 were deposited by the Inventors at American Type Culture Collection, 10801 University Boulevard, Manassas, VA, on March 16, 2007. Strain BFY013 has accession number ____ strain BFY504 has accession number ____; strain BFY598 has accession number ____ and strain BFY626 has accession number ____. All strains were deposited in accordance with the provisions of the Budapest Treaty on the International Recognition of the Deposit of Microorganisms for the Purposes of Patent Protection.

[00114] The examples herein illustrate the disclosure and are not meant as limiting in nature. The chemicals, biological agents and other ingredients are presented as typical components or reactants, and the procedures described herein may represent but one of the typical ways to accomplish the goal of the particular experiment. It is understood that various modification may be derived in view of the foregoing disclosure without departing from the spirit of the present disclosure.

EXAMPLES

Example 1-Identification of Yeast Strains Capable of Efficiently Fermenting Arabinose

[00115] To better understand L-arabinose transport and use in yeasts and to identify a source for efficient L-arabinose transporters, 165 non-*Saccharomyces* yeast strains were

studied. These yeast strains were arranged into 6 groups based on the minimum time required to utilize 20 g/L of L-arabinose. Transport rates of L-arabinose were determined for several strains and a more comprehensive transport studies was done in four selected strains. Detailed transport kinetics in *Arxula adeninivorans* showed a transport system consisting of low and high affinity components while that of *Debaryomyces hansenii* var. *fabryii*, *Kluyveromyces marxianus* and *Pichia guilliermondii* showed that these strains have single component, high affinity active transport systems.

[00116] The rationale for selecting yeast species and strains was based on reasoning that an organism that grows well on L-arabinose must have an efficient mechanism for the uptake of this sugar. In order to facilitate future experiments, we limited our search for species that were not reported to be pathogenic, did not, or rarely, formed hyphae, and grew primarily as single cells. In order to choose strains based on these criteria, we relied, in part, on the publication on yeasts by Barnett, Payne and Yarrow (Barnett, et al., 2000), previously described L-arabinose fermenting strains (Dien, et al., 1996; Kurtzman and Dien, 1998), and descriptions in the publications by various culture collections. We obtained 165 strains from 123 different species from the American Type Culture Collection (ATCC), the National Center for Agricultural Utilization Research (NRRL), or the Centraalbureau voor Schimmelcultures (CBS) (Table 1).

[00117] The routine growth medium for the growth and maintenance of the yeast strains was YPD (1% Bacto-yeast extract (Difco), 2% Bacto-peptone (Difco), 2% dextrose, and when needed for solidification, 2 % Bacto agar (Difco)). For the testing of growth on L-arabinose, similar medium (YPA) was used in which glucose was replaced with L-arabinose. Minimal medium contained 0.67% yeast nitrogen base without amino acids and 2% glucose or 2% L-arabinose. In order to determine auxotrophic requirements of various mutants, “drop-out” media (Sigma Aldrich or Clontech) were used. These media contain various supplements, but are missing adenine, uracil, or one of the amino acids.

[00118] Cells, grown in YDP medium to early stationary phase, were collected by centrifugation, washed with water and suspended in 25 ml YPA medium at an initial density of OD_{600} of 0.2 /ml. Cultures were incubated at 30°C with shaking at 220 rpm. Growth was monitored by periodic measurement of the optical density of the cultures. Utilization of L-arabinose was determined by measuring the sugar remaining in the medium after various periods of growth by the analysis of the filtered media by high pressure liquid chromatography (HPLC) on Hewlett-Packard (HP) 1090 instrument using a Bio-Rad HPX-

87H hydrogen ion resin column and an HP 1047A external refractive index detector. The mobile phase, 0.001 N H_2SO_4, was run at 55°C with a flow rate of 0.6 ml/min.

[00119] Cells were grown on L-arabinose to mid-log phase and collected by centrifugation, washed twice in water, and then suspended in water. The yeast suspension was adjusted to 30-60 mg dry weight/ml. Uptake rates and inhibition of L-arabinose transport were measured as previously described (Stambuk, et al., 2003). For our initial transport studies, we had purchased and used L-(1-^{14}C)arabinose from American Radiolabeled Chemicals, Inc. (St Louis, MO. USA). However, we discovered that the 1-^{14}C-labeled L-arabinose from this source was heavily contaminated with other radioactive compounds (see below), transport experiments reported here were done using L-(1-^{14}C)arabinose (54 mCi/mmol) that was custom-synthesized by Moravek Biochemicals Inc. (Brea, CA. USA). The analysis of the radioactive substrate was performed by thin layer chromatography (TLC) as described (Han and Robyt, 1998) using Whatman Silica Gel 60Å TLC plates (cat# 4410-221) for the separation of the compounds to visualize the products of this reaction. We also established that the compound migrating to the position expected for L-arabinose is, in fact, L-arabinose. For these determinations we used a previously described method (Sturgeon, 1986). Briefly, the preparations were dried down and resuspended to a final concentration of 180 μM in a 50 μL volume containing 80 mM Tris HCl, pH 8.6 and 500 μM NAD. This suspension was then treated with β-galactose dehydrogenase (Roche) for one hour at room temperature. Disappearance of L-arabinose, which is oxidized to form L-arabinono-1,5-lactone, was confirmed by analyzing the reaction mixture by the TLC method described above.

[00120] **Mutagenesis**. Cells were grown in YPD to stationary phase, washed, and then treated with 3% ethyl methanesulfonate (EMS) in 0.1 M potassium phosphate buffer (pH 7.2) for 1.0 h or 1.5 h at 30°C. At the end of the treatment, EMS was inactivated by diluting the cells in 5% sodium thiosulfate solution. The treated cells were washed and plated on YPD medium. After 2 days of growth, the yeast colonies were tested for auxotrophy by replica plating onto the synthetic minimal medium. For testing of growth of mutants on other sugars, the glucose from the minimal medium was replaced with appropriate sugars.

RESULTS

[00121] **Utilization of L-arabinose.** The first step in the selection of strains for transport studies was to determine how well various strains utilized L-arabinose for growth.

The efficiency with which the selected strains utilized the sugar was determined by measuring their growth on L-arabinose and by measuring the amount of L-arabinose that was consumed by these strains during various periods of growth. The amount of L-arabinose utilized was determined by HPLC analysis of the growth medium. Based on the rate of L-arabinose utilization, the 165 strains were classified into one of six groups (Tables 4 and 5). The group classification and final optical density of the cultures when grown on L-arabinose are indicated in Table 4 and a summary is provided in Table 5. Further studies of L-arabinose transport focused on the strains in groups 1-3.

Table 4: Yeast strains

Species	Strain	Group	OD_{600}
Ambrosiozyma monospora	CBS-2554	2	18.1
*Ambrosiozyma monospora**	NRRL Y-1484	2	25.5
Arxula adeninivorans	CBS-8244	1	37.2
Bullera coprosmaensis	CBS-8284	3	20.9
Bullera dendrophila	ATCC-24608	4	4.2
Bullera globispora	CBS-6981	4	23.5
Bullera mrakii	CBS-8288	3	32.5
Bullera penniseticola	CBS-8623	4	16.3
Bullera pseudoalba	CBS-7227	3	26.8
Bullera sinensis	CBS-7345	5	7.1
Bullera varaiabilis	CBS-7354	3	21.6
Bulleromyces albus	ATCC-18568	6	0.3
Bulleromyces albus	CBS-6302	4	21.2
*Candida arabinofermentans**	NRRL YB-1299	2	34.6
*Candida arabinofermentans**	NRRL YB-1984	2	31.3
Candida arabinofermentans	NRRL YB-2248	3	26.2
Candida atlantica	CBS-5263	3	28.8
Candida auringiensis	CBS-6913	3	32.4
Candida auringiensis	NRRL Y-11848	3	34.6
Candida auringiensis	NRRL Y-11849	3	32.5
Candida bertae	ATCC-58889	4	32.3
Candida bertae var. *bertae*	CBS-4605	4	32.2
Candida blankii	ATCC-18735	2	37.7

Candida blankii	CBS-6734	6	9.7
Candida cellulolytica	CBS-7920	2	32.4
Candida chilensis	CBS-5719	6	0.2
Candida conglobata	CBS-5808	4	24.5
Candida conglobata	NRRL Y-1504	3	26.4
Candida diddensiae	CBS-6032	3	29.5
*Candida diddensiae**	NRRL Y-7589	2	26.5
*Candida entomaea**	NRRL Y-7785	2	33.5
Candida insectorum	ATCC-22940	2	35.3
Candida ishiwadae	ATCC-22018	6	5.0
Candida ishiwadae	CBS-7348	6	8.2
Candida membranifaciens	CBS-1952	2	35.5
*Candida membranifaciens**	NRRL Y-2089	2	27.1
Candida methanosorbosa	CBS-7029	1	35.6
Candida mogii	CBS-2032	3	23.6
Candida nanaspora	CBS-7200	3	20.8
Candida nitratophila	CBS-2027	3	25.6
*Candida ovalis**	NRRL Y-17662	2	14.8
Candida peltata	CBS-5576	1	32.1
Candida pignaliae	CBS-6071	3	23.4
Candida populi	CBS-7351	4	26.0
Candida psychrophila	CBS-5956	6	0.2
*Candida rhagii***	ATCC-22983	5	12.3
Candida rhagii	CBS-4432	6	5.1
Candida santjacobensis	CBS-8183	6	3.8
Candida sequanensis	CBS-8118	6	3.9
Candida shehatae var. *shehatae*	CBS-2779	6	3.4
Candida shehatae var. *shehatae*	NRRL Y-17029	6	4.8
Candida silvicultrix	CBS-6269	5	8.0
Candida sonorensis	CBS-6793	2	31.1
Candida succiphila	CBS-7297	2	32.0
*Candida succiphila**	NRRL Y-11997	2	14.4
Candida succiphila	NRRL Y-11998	3	15.4

Candida tenuis	CBS-2885	3	25.3
*Candida tenuis***	ATCC-10573	5	10.6
Candida vanderwaltii	CBS-5524	1	34.8
Cryptococcus albidus	ATCC-10666	5	13.1
Cryptococcus albidus var. *albidus*	CBS-8395	3	32.8
Cryptococcus aerius	CBS-155	3	23.5
Cryptococcus cellulolyticus	CBS-8294	3	27.2
Cryptococcus curvatus	ATCC-10567	6	10.3
Cryptococcus curvatus	CBS-5324	6	10.4
Cryptococcus luteolus	ATCC-32044	6	2.0
Cryptococcus luteolus	CBS-8014	2	33.8
Cryptococcus terreus	ATCC-11799	4	16.5
Cryptococcus terreus	CBS-7528	3	21.5
Cystofilobasidium capitatum	CBS-7420	6	0.2
*Debaryomyces hansenii***	ATCC-2357	5	10.0
Debaryomyces hansenii	ATCC-36239	4	21.1
Debaryomyces hansenii	CBS-941	3	32.2
Debaryomyces hansenii var. *fabryii*	CBS-2753	1	45.2
Debaryomyces nepalensis	NRRL Y-7108	1	25.8
Debaryomyces polymorphus	NRRL Y-2022	6	6.4
Debaryomyces robertsiae	CBS-2934	5	10
Debaryomyces yamadae	NRRL Y-11714	6	5.6
Fellomyces borneensis	CBS-8282	5	19.7
Fellomyces chinensis	CBS-8278	6	2.8
Fellomyces distylii	CBS-8545	4	32.2
Fellomyces fuzhouensis	CBS-6133	4	21.8
Fellomyces horovitziae	CBS-7515	6	0.2
Fellomyces lichenicola	CBS-8315	6	1.7
Fellomyces ogasawarensis	CBS-8544	2	37.0
Fellomyces penicillatus	ATCC-32128	5	12.4
Fellomyces penicillatus	CBS-5491	5	20.0
Fellomyces polyborus	ATCC-32821	6	11.5
Fellomyces polyborus	CBS-8333	5	18.3

Fellomyces sichuanensis	CBS-8277	6	6.4
Fibulobasidium inconspicuum	CBS-6963	3	27.7
Filobasidium floriforme	ATCC-22367	3	15.2
Filobasidium floriforme	CBS-6241	4	14.1
Hansenula glucozyma	ATCC-18938	6	5.4
Kluveromyces marxianus	CBS-712	3	20.9
Kluveromyces marxianus	CBS-1089	2	23
Kluveromyces marxianus	CBS-1555	2	24.9
Kluveromyces marxianus	CBS-1557	3	26.8
Kluveromyces marxianus	CBS-2173	3	19.9
Kluveromyces marxianus	NRRL Y-8281	3	13.7
Kockovaella machilophila	CBS-8607	5	12.7
Kockovaella sacchari	CBS-8624	4	21.4
Myxozyma geophila	CBS-7219	4	24.9
Myxozyma kluyveri	CBS-7332	6	2.5
Myxozyma lipomycoides	CBS-7038	3	16.3
Myxozyma melibiosi	CBS-2102	3	27.4
Myxozyma monticola	CBS-7806	4	13.9
Myxozyma mucilagina	CBS-7071	3	17.0
Myxozyma neglecta	CBS-7058	3	20.2
Myxozyma vanderwaltii	CBS-7517	3	26.4
Pachysolens tannophilus	ATCC-32691	3	20.1
Pichia angophorae	CBS-5823	6	3.0
Pichia bovis	NRRL YB-4184	2	36.6
Pichia capsulate	CBS-136	4	30.7
Pichia capsulate	NRRL Y-1842	3	26.2
Pichia ciferrii	NRRL Y-1031	6	3.4
*Pichia guilliermondii**	NRRL Y-2075	2	40.0
Pichia haplophila	NRRL Y-7860	6	5.1
Pichia heimii	CBS-6139	5	5.6
*Pichia holstii***	ATCC-13689	5	11.5
Pichia holstii	CBS-2026	6	5.4
Pichia kodamae	CBS-7081	5	21.9

Pichia kodamae	NRRL Y-17234	4	25.8
Pichia methanolica	ATCC-46071	2	26.9
Pichia methanolica	CBS-6515	1	27.5
Pichia mississippiensis	NRRL YB-1294	2	30.0
Pichia naganishii	ATCC-32148	2	35.5
Pichia naganishii	CBS-7259	2	38.6
Pichia nakazawae var. *nakazawae*	NRRL Y-7903	5	12.8
Pichia philogaea	NRRL Y-7813	3	21.7
Pichia rabaulensis	CBS-6797	1	38.7
Pichia scolyti	CBS-4802	2	29.1
Pichia scolyti	NRRL Y-5512	1	45.8
Pichia silvicola	NRRL Y-1679	3	20.8
Pichia stipitis	ATCC-62970	6	9.0
Pichia stipitis	CBS-5773	6	8.0
Pichia tannicola	NRRL Y-7499	6	7.3
Pichia trehalophila	CBS-5361	2	33.1
Pichia triangularis	CBS-4094	3	31.4
Pseudozyma flocculosa	CBS-167.88	5	6.6
Psuedozyma fusiformata	CBS-6951	6	1.3
Pseudozyma rugulosa	CBS-170.88	3	21.8
Rhodosporidium sphaerocarpum	CBS-5939	3	27.7
Rhodotorula acuta	ATCC-42713	3	32.4
Rhodotorula fragaria	CBS-6254	4	16.0
Rhodotorula pustula	ATCC-32034	6	6.2
Sirobasidium intermedium	CBS-7805	4	15.9
Smithiozyma japonica	CBS-7319	4	13.2
Sporidiobolus ruineniae	CBS-5001	6	2.5
Stephanoascus smithiae	CBS-5657	3	37.6
Sterigmatomyces elviae	CBS-5922	3	29.3
Sympodiomycopsis paphiopedili	CBS-7429	3	27.5
Tremella aurantia	CBS-6965	4	15.1
Tremella cinnabarina	CBS-8234	6	2.5
Tremella enchephala	CBS-6968	3	20.7

Tremella foliacea	CBS-8228	6	1.0
Tremella fuciformis	CBS-8225	6	3.3
Tremella indecorata	CBS-6976	4	15.5
Tremella nivalis	CBS-8487	3	25.3
Trichosporon laibachii	ATCC-90037	3	19.7
Trichosporon laibachii	CBS-2495	3	16.7
Trichosporon loubieri	ATCC-56048	3	24.5
Trichosporon loubieri	CBS-7719	4	20.7
Trichosporon moniliiforme	ATCC-22164	3	22.9
Trichosporon moniliiforme	CBS-2820	3	29.1

*These strains were not analyzed at 18 hours; some of them could be in group 1.

**These strains were not analyzed at 144 hours; some of them could be in group 4.

Table 5. Utilization of L-arabinose by yeast strains

Group	Criteria	No. of strains
1	Used > 90% of available L-arabinose within 18 hours	9
2	Used all L-arabinose within 24 hours	27
3	Used all L-arabinose within 48 hours	51
4	Used all L-arabinose within 144 hours	24
5	Used > 10% of available L-arabinose in 144 hours	17
6	Did not use any or < 10% of available L-arabinose in 144 hours	37

[00122] **Mutagenesis of selected strains.** Our ultimate objective for these studies was not only to characterize L-arabinose utilization and uptake by various yeast species, but to identify the strains that contain efficient L-arabinose transporter(s) that could be targeted for isolation and engineering for expression in *S. cerevisiae* (See Examples 2-9). Since use of appropriate mutants is one of the approaches for identifying and isolating genes of interest, we examined the colony and cell morphology of the strains collected for these studies and chose only the strains that would be expected to be amenable to the application of genetic and biochemical methods. For example, we excluded strains such as *Bullera penniseticola*, *Pichia*

capsulate, *Pichia kodamae*, *Smithiozyma japonica*, *Sterigmatomyces elviae*, and the *Myxozyma* and *Tremella* species that formed slimy colonies and would be difficult to replica plate. We also excluded strains such as *Ambrosiozyma monospora*, *Trichosposon laibachii*, and the *Psuedozyma* species that tended to display significant mycelial form of growth from which it would be difficult to obtain single colony mutants. Since very limited or no genetic information is available for most of the yeast species collected for this investigation, we chose several strains, mainly from groups 1 and 2, for mutagenesis to determine if they would yield auxotophic mutants at a reasonable frequency. It is clear that a variety of mutants can be isolated from each of the selected strains with reasonable frequency (Table 6). We therefore concluded that it would be possible to isolate appropriate mutants from one or more of the organisms that were determined to harbor suitable L-arabinose transporter(s). Screening through approximately 155,000 colonies for *Arxula adeninivorans*, 132,000 for *Debaryomyces hansenii* var. *fabryii*, and 76,400 for *Pichia guilliermondii*, we also isolated mutants that were unable to grow on L-arabinose (See Fig. 1, Fig. 2, and Fig. 3). These mutants were further tested for growth on arabitol, xylose, and xylitol in an attempt to identify class of mutants expected to contain L-arabinose transport deficient mutants for use in future studies.

Table 6. EMS-induced mutagenesis of selected strains

Species	Strain	No. of colonies tested	No. of auxotrophs obtained	Mutant frequency (%)	Types of auxotrophs obtained
A. adeninivorans	CBS-8244	14440	5	0.03	Ade^-, Arg^-, Met^-
C. arabinofermentans	NRRL YB-1299	19750	13	0.07	Ade^-, Arg^-, Lys^-
D. hansenii var. *fabryii*	CBS-2753	24400	26	0.11	Ade^-, Arg^-, His^-, Ile^-, Leu^-, Lys^-, Met^-, Thr^-, Trp^-, Tyr^-
D. nepalensis	NRRL Y-7108	2760	5	0.18	Ade^-, Arg^-
K. marxianus	CBS-712	32650	4	0.01	Ade^-, Val^-
P. guilliermondii	NRRL Y-2075	6920	10	0.14	Ade^-, Arg^-, Lys^-, Thr^-, Tyr^-
P. methanolica	CBS-6515	27200	13	0.05	Ade^-, Arg^-, Lys^-, Met^-

[00123] **Transport of L-arabinose.** For our initial transport studies, 1-^{14}C-labeled L-arabinose from American Radiolabeled Chemicals Inc. was used as this was the only supplier from where L-(1-^{14}C)arabinose was available. However, our TLC analysis of the chemical obtained from this source showed that it contained a mixture of L-arabinose and other unknown compounds and that L-arabinose constituted only a small fraction of these chemicals (Fig. 4). We did not attempt to identify the unexpected compounds present in the L-arabinose preparations. In the chromatogram of the L-(1-^{14}C)arabinose sample, we have indicated the positions to which other related compounds are expected to migrate in this TLC system (Fig. 4). It should be noted that the D-(1-^{14}C)xylose and D-(1-^{14}C)galactose obtained from the same source are essentially pure (Fig. 4). Additionally, we used the method described by Sturgeon to establish that the compound migrating to the position expected for L-arabinose was , in fact, L-arabinose. This method uses galactose dehydrogenase to convert L-arabinose to L-arabino-lactone which is then converted to L-arabinonic acid in aqueous solution (Sturgeon, 1986). The spot expected to be L-arabinose disappears when the preparation is treated with galactose dehydrogenase while the other spots remain unaffected (Fig. 5). Subsequently, we obtained L-(1-^{14}C)arabinose from custom synthesis by Moravek Biochemicals Inc. This preparation was also analyzed by the methods described above and was found to be free of contaminating compounds (data not show).

[00124] In order to select the eventual strain(s) for isolation of an L-arabinose transporter, we evaluated L-arabinose transport rates in several strains deemed suitable for further investigation (Table 7). We chose these strains based on their L-arabinose utilization efficiency (group 1-3) and suitability for genetic analysis and mutagenesis. We also tested whether L-arabinose transport in these strains is active by measuring transport in the presence of 2,4-dinitrophenol (DNP), an inhibitor of active transport. Transport of L-arabinose was inhibited by DNP in all of the strains tested indicating that these strains harbor active transport mechanisms for L-arabinose (Table 7).

Table 7 Rates of L-arabinose transport (nmole $mg^{-1}min^{-1}$)

Group	Species	Source	L-arabinose concentration (mM)			Inhibition by 1.25 mM DNP
			133	13.3	1.3	(%)[a]
1	*A.adeninivorans*	CBS-8244	32.4±0.5	4.4±0.3	5.2±0.2	96

2	*C.arabinofermentans*	NRRL YB-1299	10.8±0.0	12.6±0.4	11.5±1.9	99[b]
2	*C.blankii*	ATCC-18735	2.6±0.3	1.6±0.4	0.4±0.2	96
3	*Dhansenii*	CBS-941	2.7±0.1	2.3±0.1	1.7±0.1	89
1	*D.hansenii var. fabryii*	CBS-2753	14.2±2.5	15.2±0.7	13.3±1.5	98
1	*D.nepalensis*	NRRL Y-7108	5.0±2.0	3.8±0.8	3.3±0.1	99
3	*K.marxianus*	CBS-712	4.8±0.2	5.7±0.6	1.2±0.1	88
2	*K.marxianus*	CBS-1089	28.5±0.1	20.5±1.3	20.2±0.6	96[c]
2	*K.marxianus*	CBS-1555	4.4±0.3	4.5±0.9	1.0±0.0	94
2	*P.bovis*	NRRL YB-4184	10.0±2.2	7.7±0.5	7.9±0.7	91
2	*P.guilliermondii*	NRRL Y-2075	8.2±8.0	16.8±3.0	22.0±1.8	100[d]
1	*P.methanolica*	CBS-6515	14.2±1.3	10.6±0.7	9.2±0.1	99[b]
1	*P.scolyti*	NRRL Y-5512	4.6±0.8	4.1±0.1	2.9±0.2	94

[a] Transport inhibition assayed at 10.0 mM unless indicated.
[b] Transport inhibition assayed at 1.3 mM.
[c] Transport inhibition assayed at 3.0 mM.
[d] Transport inhibition assayed at 0.3 mM.

[00125] **Detailed transport studies with *A. adeninivorans*, *D. hansenii* var. *fabryii*, *Kluyveromyces marxianus*, and *P. guilliermondii*.** Based on the relatively high L-arabinose transport velocities found in our initial transport studies, we selected *A. adeninivorans*, *D. hansenii* var. *fabryii*, *K. marxianus* (CBS-1089), and *P. guilliermondii* for detailed L-arabinose transport studies. In the case of *A. adeninivorans*, both low and high affinity transport systems are present as indicated by the non-linear Eadie-Hofstee plot (Fig. 6, panel A). The low affinity transport system has a K_m of 250 mM and a V_{max} of 20.0 nmol/mg·min while the high affinity system has a K_m of 0.3 mM and a V_{max} of 6.7 nmol/mg·min. In our initial screen of transport activity and inhibition of transport in *A. adeninivorans*, an L-arabinose concentration of 10 mM was used (Table 7). At this low level of substrate, only the high affinity transport system would be impacted showing that this component is an active transport system.

[00126] The linear Eadie-Hofstee plots for *D. hansenii* var. *fabryii*, *K. marxianus*, and *P. guilliermondii* indicate single component, high affinity transport systems were responsible for L-arabinose uptake (Fig. 6, panels B, C, and D respectively). These transport systems had K_m values of 0.10 mM, 0.14 mM and 0.07 mM and V_{max} values of 15.0 nmol/mg·min, 24.0 nmol/mg·min, and 22.5 nmol/mg·min for *D. hansenii* var. *fabryii*, *K. marxianus*, and *P. guilliermondii*, respectively. These high capacity transport systems allowed these strains to effectively metabolize 20 g/L of L-arabinose within 18 h for *A. adeninivorans* and *D. hansenii* var. *fabryii* and within 24 h for *K. marxianus* and *P. guilliermondii*.

DISCUSSION

[00127] There is very high degree of variability in the extent and the rate of L-arabinose utilization among various species of yeast. This example provides screening of a variety of strains for the purpose of isolating an efficient L-arabinose transporter by reasoning that strains that grew well on L-arabinose would have an efficient transport mechanism for this sugar. It was determined that of the strains may also have efficient L-arabinose transporter(s), but don't grow well on the sugar due to other defects in the L-arabinose utilization pathway. An examination of Table 4 reveals that there is significant variability in the efficiency of L-arabinose utilization among the various strains of the same species. For example, various strains of *D. hansenii*, which has been shown to be polyphyletic and highly variable at the genomic level (Corredor, et al., 2003; Kurtzman and Robnett, 1997), can be classified in any one of the groups 1, 3, 4, or 5. On the other hand, some genera and species showed little or no diversity. The 3 strains of *C. arabinofermentans*, 2 strains of *C. diddensiae*, 3 strains of *C. succiphila*, and 5 strains of *K. marxianus* are in similar groups (group 2 or 3). Similarly, 5 of the 6 strains of the *Trichosporon* sp. are in group 3 and the remaining one is in group 4. All strains of *A. monospora*, *C. auringiensis*, *C. bertae*, *C. membranifaciens*, *C. shehetae*, *C. curvatus, F. penicillatus*, *P. holstii*, *P. naganishii*, *P. stipitis, T. laibachii*, and *T. moniliiforme* were assigned to the same group as the other strains of the same species. Generally a strain from one collection had similar growth properties as the equivalent strain from a different collection. For example, *A. monospora* (NRRL Y1484 and CBS-2554), *C. membranifaciens* (NRRL Y-2089 and CBS-195), and *K. marxianus* (NRRL Y-8281 and CBS-712) behaved identically from both collections. However, there were instances where the equivalent strains from different collections had somewhat different growth properties as is the case for *P. kodamae* (NRRL Y-17234 and CBS-7081) and *P. scolyti* (NRRL Y-5512 and CBS-4802). The strains that were previously found to ferment L-arabinose grew well on L-arabinose in our screening (Dien, et al., 1996; Kurtzmann and Dien, 1998). In our study, three of these strains utilized 20 g/L of L-arabinose in 24 h (group 2) while *C. auringiensis* consumed all L-arabinose in 48 h (group 3).

[00128] Initial uptake rates varied widely for strains in groups 1-3. No clear relationship could be established between the initial uptake rates and the time it took to completely exhaust the L-arabinose from the culture media. Strains in group 1 did not always have higher transport rates than those in groups 2 or 3. The transport of the sugar is, of course, just one (first) step in its metabolism and the velocities of other enzymes in the

pathway, the flux rates of the intermediates of the pathway, and various metabolic regulatory mechanisms contribute to the rate at which L-arabinose would be utilized by the organisms.

[00129] Three strains were selected to determine if mutants unable to utilize L-arabinose for growth could be isolated. A number of such mutants were isolated from each of these strains (Fig. 1-3). Except for the mutants that do not grow on xylitol, all the mutants can be explained by a single mutation in one of the steps in the L-arabinose utilization pathway in yeasts (Chiang and Knight, 1960) (Fig. 7). This pathway has been confirmed to be functional by demonstrating that *S. cerevisiae* expressing genes for each of the enzymatic steps can utilize L-arabinose for growth (Richard, et al., 2003). The mutants that do not grow on xylitol may have defects at more than one step or, alternatively, there is yet an unknown pathway for xylose utilization.

[00130] The only known yeast transporter for L-arabinose, galactose permease (Gal2p) of *S. cerevisiae*, is a facilitated diffusion permease and transports L-arabinose as long as the external concentration of L-arabinose is higher than the internal cellular concentration (Cirillo, 1968; Kou, et al., 1970) with a transport velocity of 0.32 nmol/mg·min at 10 mM L-arabinose (Becker and Boles, 2003). It was shown previously that Gal2p is a very high capacity L-arabinose transporter, but only at high concentrations of L-arabinose (Knoshaug et. al., L-arabinose and D-galactose transport by the galactose permease of *Saccharomyces cerevisiae*, submitted to Applied and Environmental Microbiology). Suprisingly, the L-arabinose transport rates in the non-conventional yeast strains were higher than that reported for Gal2p in *S. cerevisiae* at low concentrations of L-arabinose. In particular, the active, high affinity transport systems of *K. marxianus* and *P. guilliermondii* had transport velocities that were, respectively, 64 and 53 times greater than reported velocities for Gal2p of *S. cerevisiae*.

[00131] Interestingly, *A. adeninivorans* has a combination of transporters. The combination of low and high affinity, high capacity L-arabinose transport systems endogenously present in *A. adeninivorans* allows for the complete assimilation of 20 g/L of L-arabinose within 18 h demonstrating that, through the combination of transport systems with different affinities for high and low concentrations of L-arabinose, a strain can be developed that can utilize L-arabinose quickly.

[00132] Two particular strains *K. marxianus* and *P. guilliermondii* of yeast were selected for further characterization and identification of single component arabinose transport.

Example 2—Cloning of the new transporter gene *KmLAT1*

[00133] A *K. marxianus* genomic library was constructed in our yeast vector pBFY13 which contains the yeast 2μ origin of replication, a *URA3* selection cassette, and a *BamHI* site located between the *PGK1* promoter and *GAL10* terminator. After partial digestion of 200 μg of genomic DNA with *Sau3AI* restriction enzyme, fragments of 2-8 kb in length were gel-isolated and ligated into the *BamHI* site of pBFY013. This ligation reaction was then transformed into *E. coli* and plated for recovery. Plate counts produced ~3000 cfu's/10 μl of transformed cells and the plasmid DNA from 24 colonies was screened for presence of insert revealing 22 of 24 transformants had an insert ranging from 1 kb to ~8 kb giving an average insert size of 3.2 kb. The transformed cells were scraped from the plates, DNA recovered, and 5 μl was transformed into competent BFY518 cells. The strain, BFY518, was cured of the *GAL2* over-expression plasmid negating its ability to form colonies on agar plates containing L-arabinose as the sole carbon source enabling restoration of colony formation by complementation with a heterologous L-arabinose transporter. To count the number of transformed yeast cells, 10 μl of the yeast library transformation were plated onto minimal glucose media yet the colonies were so dense that only an estimate of ~5000 colonies was possible. The rest of the transformation mix (~140 μl) was plated onto minimal media containing 2% L-arabinose for selection from which a small amount of background growth was noticed. The plates were then replica plated to fresh L-arabinose minimal media. The total number of cells plated for selection represented ~280,000 transformants representing ~8 fold coverage of the 10.7 mb *K. marxianus* genome (See Dujon et al., 2004). Two colonies grew on the replica plates and the plasmid DNA was rescued and re-transformed into BFY518 allowing growth once again on L-arabinose confirming that the *K. marxianus* genomic insert carried on these plasmids was responsible for growth. Restriction analysis suggested both plasmids harbored the same insert of approximately 8.8 kb in size.

Example 3—Sequence Analysis of the *KmLAT1* gene

[00134] Sequencing results showed that both plasmids had identical inserts of 8838 kb containing two ORFs on the 5' end of the insert. Both of these ORFs showed strong homology to yeast sugar transporters. One transporter ORF was interrupted by a fragment of an unrelated ORF suggesting that recombination of fragments during ligation into the vector occurred in library construction (Fig. 10). Recombination of library fragments during ligation into the vector was shown by PCR walking experiments performed on *K. marxianus* genomic DNA. Walking was performed out of the transporter in a 5' direction and additional transporter sequence including the start codon was recovered rather than the additional

sequence from the unrelated ORF. The uninterrupted transporter ORF, termed Km*LAT1*, was recovered twice more in another subsequent library screening. This ORF was 1668 bp in length and shared homology with high-affinity glucose transporters in particular, *HGT1* from *K. lactis* (Table 2) and showed a much closer association with high-affinity glucose transporters from non-conventional yeasts than the bacterial araE genes or *S. cerevisiae* hexose transporters (Fig. 8).

[00135] Transmembrane region prediction by the software Tmpred shows 12 transmembrane regions with a larger intercellular loop between regions 6 and 7 in KmLat1 (Fig. 9) (See Hofmann et al, 1993), typical of Gal2p and other yeast sugar transporters having 10-12 transmembrane regions (See e.g., Alves-Araujo et al., 2004; Day et al., 2002; Kruckeberg et al., 1996; Pina et al. ,2004; and Weierstall et al. 1999).

Example 4--*KmLAT1* Expressed in *S. cerevisiae* Enables Growth on Arabinose

[00136] The coding sequence of *KmLAT1* was isolated by PCR from genomic DNA of *K. marxianus* and cloned into a yeast 2μ plasmid under control of the *PGK1* promoter of *S. cerevisiae*. This construct was transformed into a *GAL2* deleted strain of *S. cerevisiae* adapted to L-arabinose, BFY590. Briefly, cells are grown in appropriate selective glucose minimal media until saturation then washed and diluted to a starting OD_{600} of 0.2 in minimal media supplemented with 2% L-arabinose. Cultures are incubated until exponential growth is observed then the cultures are diluted twice into the same media for continued growth to establish the final L-arabinose utilizing adapted strain which is purified on streak plates. Control plasmids carrying the yeast *GAL2* gene and an empty vector were also used to transform yeast cells.

[00137] Yeast cells with a 2μ plasmid carrying the *KmLAT1* or *GAL2* gene or cells with an empty 2μ plasmid were grown with shaking in liquid minimum media containing 2% L-arabinose as the sole carbon source. The OD_{600} of each culture was measured and monitored by 140 hours. Growth curve results show that *KmLAT1* is sufficient to support growth on L-arabinose when compared to cells harboring the empty vector which does not show any signs of growth (Fig. 11). This result confirms that the *KmLAT1* gene encodes an arabinose transporter that enables yeast cells to grow on L-arabinose.

Example 5—Comparison of the Arabinose Transport Kinetics between Gal2p and KmLat1 expressed in *S. cerevisiae*

[00138] The transport characteristics of the KmLat1 and the Gal2p transporters expressed in *S. cerevisiae* were compared. Both transporters were expressed in a host, BFY590, adapted for growth on L-arabinose in which the endogenous copy of *GAL2* had been entirely replaced with a *HIS3* selection marker. The KmLat1 transporter showed a low-affinity transporter having a Km = 230 mM and a Vmax = 55 nmol/mg·min (Fig. 12A). This is in contrast to the high-affinity active transport activity induced in the wild type *K. marxianus* when grown on 2% L-arabinose (Fig. 12B). These results suggest there are at least 2 transporters in *K. marxianus* that may transport L-arabinose but only the high-affinity activity is induced in the wild type when grown on 2% L-arabinose. Inhibition experiments showed that KmLat1 expressed in *S. cerevisiae* is not significantly inhibited by protonophores such as NaN3, DNP, and CCP. Neither is KmLat1 inhibited by H+-adenosine triphosphatase (ATPase) inhibitors such as DESB and DCCD (Table 8). This is in contrast to the transport activity in wild type *K. marxianus*, suggesting that KmLat1 is a facilitated diffusion permease similar to the Gal2 permease. Competition experiments showed that KmLat1 is out-competed by glucose, galactose, xylose, and maltose when expressed in *S. cerevisiae* (Table 8).

Table 8. Effect of Inhibitors or Competing Sugars on the Rate of L-Arabinose Transport in L-Arabinose-Grown *S. cerevisiae* Expressing *GAL2* or *KmLAT1*

Inhibitor or Competing Sugar	Concentration(m*M*)	Relative L-arabinose transport (%)	
		Gal2	KmLat1
None	NA	100[a]	100[c]
NaN_3	10	66	11
CCCP	5	46	61
DCCD	5	69	55
DNP	5	72	75
DESB	5	81	100
None	NA	100[b]	100[d]
Glucose	900	10	17
Galactose	900	3	23
Xylose	900	25	25
Maltose	450	ND	38

[a] Uptake rate was 66.0 nmol mg^{-1} min^{-1} determined with 118 m*M* labeled L-arabinose.
[b] Uptake rate was 18.9 nmol mg^{-1} min^{-1} determined with 30 m*M* labeled L-arabinose.
[c] Uptake rate was 7.7 nmol mg^{-1} min^{-1} determined with 118 m*M* labeled L-arabinose.

[d] Uptake rate was 3.6 nmol mg^{-1} min^{-1} determined with 30 m*M* labeled L-arabinose.
ND, Not Done.

[00139] Transport kinetics of *S. cerevisiae* BFY597 over-expressing the Gal2 permease grown on 2% L-arabinose showed a K_m of 550 mM and a V_{max} of 425 nmol/mg·min for L-arabinose transport (Fig. 12A). Inhibition assays showed a reduction but not a complete inhibition of transport suggestive of facilitated diffusion transport as previously reported (Kuo et al. 1970, see also Table 4). Competition studies showed that glucose, galactose, and xylose significantly reduced L-arabinose transport indicating that these sugars are preferentially transported over L-arabinose (Table 8). The kinetics of galactose transport were also measured in this strain and indicate that Gal2p has a K_m of 25 m*M* and a V_{max} of 76 nmol/mg·min for galactose transport, demonstrating a higher affinity for galactose that would out-compete L-arabinose for transport.

Example 6—Cloning of the New Transporter Gene *PgLAT2*

[00140] *Pichia guilliermondii* cells were grown in minimal media supplemented with 2% L-arabinose, galactose, or xylose. Cells were collected in mid-growth and washed twice in water before suspension in water at about 30 mg/ml. RNA was extracted from the cells using the acid phenol method (Ausubel, et al., Short Protocols in Molecular Biology, John Wiley and Sons, 1999). Briefly, approximately 15 mL of fresh culture was added to ~25 mL of crushed ice and centrifuged at 4° C for 5 min at 3840 x g. Cells were washed twice with cold DEPC-treated water, and the pellets were frozen at -80° C. After the pellets were resuspended in 400 ul TES (10 mM Tris HCl, pH 7.5, 5 mM EDTA, 0.5% SDS), 400 ul of acid phenol was added. The samples were vortexed vigorously for 10 sec, followed by incubation for 30-60 min at 65° C with occasional vortexing. The tubes with the samples were then chilled on ice and spun for 5 min at 4° C. The aqueous phase was removed and re-extracted with chloroform. The aqueous phase was then ethanol precipitated using 0.1 volume of 3 M sodium acetate (pH 5.3) and two volumes of 100% ethanol. The pellet was washed using 80% ethanol, dried, and resuspended in 50 ul DEPC H_2O. Total RNA concentration was quantitated by measuring the OD_{260} and visualized on agarose gels.

[00141] RNA purification, synthesis of cDNA, and differential display were performed at GenHunter Corporation according to standard techniques. DNA Bands showing higher levels of expression from arabinose-grown cells relative to xylose- or galactose-grown cells were reamplified using the differential display amplification primers. Direct sequencing was

performed on the PCR products using the GenHunter arbitrary primers. In cases that did not yield clean sequence, the amplification products were cloned in the TOPO-TA vector pCR2.1 (Invitrogen) and individual clones were sequenced. Sequences were then compared to the databases using BLASTX analysis and those that showed similarity to known transporters or transporter-like proteins were examined further. One of these sequences led to the identification of a novel transporter gene, *PgLAT2* from *Pichia guilliermondii.*

[00142] The full-length *PgLAT2* gene was isolated by genome walking in *P. guilliermondii.* PCR-based walking was done in both the 5' and 3' directions from the sequence isolated by differential display. The entire gene was then isolated from genomic DNA using PCR primers based on the flanking DNA sequences. *PgLAT2* gene has an ORF of 1617 nucleotides encoding a protein with a predicted length of 539 amino acids (Fig. 13). Sequence similarity was observed between *PgLAT2* and other sugar transporter genes, including high affinity glucose transporters from *Candida albicans* and *Kluyveromyces lactis*. Similar to KmLat1, the predicted PgLat2 polypeptide showed 12 transmembrane regions with a larger intercellular loop between regions 6 and 7, typical of yeast sugar transporters.

Example 7—Characteristics of Sugar Transport by *Pichia guilliermondii*

[00143] The induction of L-arabinose transport in wild type *P. guilliermondii* was examined. Wildtype *Pichia guilliermondii* cells were grown in minimal media supplemented with 2% L-arabinose, galactose, or xylose while BFY605 cells were grown in the same media supplemented with 0.2% L-arabinose. Cells were collected in mid-growth and washed twice in water before suspension in water at about 30 mg/ml. Uptake of L-(1-^{14}C)arabinose (54 mCi/mmol, Moravek Biochemicals Inc.), D-(1-^{14}C)galactose (57 mCi/mmol, Amersham Biosciences), or D-(1-^{14}C)xylose (53 mCi/mmol, Moravek Biochemicals Inc.) was measured as previously described (Stambuk, Franden et al. 2003). Assays were performed in 5, 10, or 30 second periods to maintain initial rates. Appropriate experiments ensured uptake was linear for at least 1 minute. Transport activity was described as nmoles of labeled sugar transported per mg cell dry weight per minute. Inhibition and competition assays were performed as previously described (Stambuk, Franden et al. 2003).

[00144] Cells grown on L-arabinose were able to transport L-arabinose whereas cells grown on galactose or xylose were not able to transport L-arabinose. Additionally, xylose transport was about double in cells grown in L-arabinose media compared to cells grown in xylose media. Galactose was transported at the same rate independent of growth substrate (Fig. 14). Transport competition between L-arabinose and xylose was also examined.

Uptake of labeled L-arabinose was reduced by 96% when 100x un-labeled xylose was included in the transport assay whereas uptake of labeled xylose was only reduced by 16% when 100x un-labeled L-arabinose was included in the assay (Fig. 15). These data suggest that in *P. guilliermondii*, growth on L-arabinose induces expression of a specific transport system capable of transporting L-arabinose and xylose. Additionally, this system preferentially transports xylose at the expense of L-arabinose if both sugars are present and has a higher transport velocity for xylose than the transport system induced when grown on xylose. By contrast, transport activity for L-arabinose is not induced when grown on xylose.

Example 8—Arabinose Transport Kinetics of PgLat2 expressed in *S. cerevisiae*

[00145] The L-arabinose transport characteristics of the PgLat2 transporter expressed from a 2μ plasmid under control of the *PGK1* promoter of *S. cerevisiae* in *S. cerevisiae* grown on 0.2% L-arabinose medium showed the same L-arabinose transport characteristics as wildtype *P. guilliermondii* (Fig. 16). The PgLat2 transporter when expressed in *S. cerevisiae* has a K_m = 0.07 mM and V_{max} = 18 nmol/mg-min. Inhibition experiments showed significant inhibition of transport by protonophores (NaN_3, DNP, and CCP) and H+-adenosine triphosphatase (ATPase) inhibitors (DESB and DCCD) similar to the inhibition observed in wildtype *P. guilliermondii* (Table 9). Competition experiments showed that L-arabinose uptake by the PgLat2 transporter was inhibited by glucose, galactose, xylose and to a lesser extent by maltose (Table 9).

Table 9
Effect of Inhibitors or Competing Sugars on the Rate of L-Arabinose Transport in L-Arabinose-Grown *P. guilliermondii* Y-2075 and *S. cerevisiae* BFY605

Inhibitor or Competing Sugar	Concentration (mM)	Relative L-arabinose transport	
		P. guilliermondii	***S. cerevisiae* (PgLat2 transporter)**
None [a]	-	100	100
NaN_3	10	1	16
DNP	5	0	4
CCCP	5	0	2
DCCD	5	22	36
DESB	5	8	1
None [b]	-	100	100
Glucose	120	ND	17
Galactose	120	ND	20
Xylose	120	4	0
Maltose	120	ND	30

[a] Rate of L-arabinose transport was 11.2 nmol mg^{-1} min^{-1} for *P. guilliermondii* and 10.4 nmol mg^{-1} min^{-1} for *S. cerevisiae* (PgLat2 transporter) determined with 0.33 mM labeled L-arabinose.

[b] Rate of L-arabinose transport was 14.2 nmol mg^{-1} min^{-1} for *P. guilliermondii* and 14.4 nmol mg^{-1} min^{-1} for *S. cerevisiae* (PgLat2 transporter) determined with 1.2 mM labeled L-arabinose.

[00146] The transport activities, inhibition profiles, and competition rates with respect to xylose of wildtype *P. guilliermondii* and of the PgLat2 transporter expressed in *S. cerevisiae* are very similar, suggesting that *P. guilliermondii* has a single, high affinity, active transporter charged with uptake of L-arabinose. There are no L-arabinose transport activities that are unaccounted for which suggests the presence of a single L-arabinose transporter in *P. guilliermondii*.

Example 9—Synergistic Effect on Growth Rate and Sugar Utilization by *S. cerevisiae* Expressing Gal2p and the New Transporter Protein-PgLat2

[00147] To determine the complementary effects on arabinose transport by the transporters, namely, Gal2p and PgLat2, yeast strains were constructed with appropriate selection markers to allow different pathway and transporter combinations to be expressed. Transporter combinations were generated by introducing a transporter expression plasmid for *PgLAT2* (or an empty vector) into *S. cerevisiae* strain BFY607 expressing the bacterial genes

araA, *araB* and *araD* (See e.g., Becker and Boles, for examples of yeast strain expressing these bacterial proteins for arabinose metabolism). The resulting strains, expressed Gal2p due to the *gal80-* genotype that de-represses *GAL2* expression. Strains BFY609 and BFY612 containing a control vector or a PgLat2 expression vector, respectively, were able to grow on 2% or 0.2% L-arabinose after extensive lag times (A process termed "adaptation."). A relatively low concentration of L-arabinose (0.2%) was used in this experiment as strain differences are more pronounced at this concentration. Once "adapted" to growth on 0.2% L-arabinose, the strains were able to grow more quickly and growth curves for the two transporter combination were generated as shown in Fig. 17 (Also see Table 3). A significant lag time was observed due to their inoculation from stationary cultures. However, once growth initiated, the growth rate was relatively rapid. The doubling time for each culture in the exponential phase of the curve is shown in Table 3. The doubling time for the PgLat2 and Gal2p co-expressing cells was markedly shorter than in the cells expressing only Gal2p (15 hours *vs*. 19 hours). A second observation relates to the overall extent of growth. The PgLat2 expressing strain appeared to grow to a higher overall optical density at saturation, suggesting that this strain was able to utilize the carbon source (L-arabinose) in the medium more completely (Fig. 17).

Example 10—Co-expression of Gal2p with PgLat2 Enables more Complete Utilization of Arabinose by Recombinant *S. cerevisiae*

[00148] Doubling times for the cultures described above in Example 9 were measured in early exponential phase for each culture. Doubling time was measured by the period of time taken for the number of cells to double in a given cell culture (See generally, Guthrie and Fink, 1991). The concentration of remaining L-arabinose at the 276 hour time point was determined by HPLC (for saturated cultures only). The concentration of L-arabinose in the starting media was about 1.89 g/L and the concentration of L-arabinose in media without L-arabinose had an undetectable level (<0.1 g/L). As shown in Table 3, significantly less residual L-arabinose remained in the culture of cells expressing both Gal2p and PgLat2 than in the culture of cells expressing Gal2p alone.

[00149] While a number of exemplary aspects and embodiments have been discussed above, those of skill in the art will recognize certain modifications, permutations, additions and sub-combinations thereof. It is therefore intended that the following appended claims and claims hereafter introduced are interpreted to include all such modifications, permutations, additions and sub-combinations as are within their true spirit and scope.

[00150] This specification contains numerous citations to references such as patents, patent applications, and scientific publications. Each is hereby incorporated by reference for all purposes.

References Cited

Alves-Araujo, C., M. Hernandez-Lopez, M. Sousa, J. Prieto, and F. Randez-Gil. 2004. Cloning and characterization of the MAL11 gene encoding a high-affinity maltose transporter from *Torulaspora delbrueckii*. FEMS Yeast Research 4:467-476.

Arnold, F.H. Nature Biotechnol. 1998, 16, 617-618.

Barnett, J. A. 1976. The utilization of sugars by yeasts. Adv. Carbohydr. Chem. Biochem. **32:**125-234.

Bowie *et al.,* 1990, *Science 247*:1306-10.

Cosman et.al., 2001 *Immunity* 14:123-33.

Crameri A. et al., 1998, Nature 391: 288-91.

Day, R., V. Higgins, P. Rogers, and I. Dawes. 2002. Characterization of the putative maltose transporters encoded by YDL247w and YJR160c. Yeast **19:**1015-1027.

Deanda, K., M. Zhang, C. Eddy, and S. Picataggio. 1996. Development of an arabinose-fermenting *Zymomonas mobilis* strain by metabolic pathway engineering. Appl. Environ. Microbiol. **62:**4465-4470.

Dujon B, S. D., Fischer G, Durrens P, Casaregola S, Lafontaine I, De Montigny J, Marck C, Neuveglise C, Talla E, Goffard N, Frangeul L, Aigle M, Anthouard V, Babour A, Barbe V, Barnay S, Blanchin S, Beckerich JM, Beyne E, Bleykasten C, Boisrame A, Boyer J, Cattolico L, Confanioleri F, De Daruvar A, Despons L, Fabre E, Fairhead C, Ferry-Dumazet H, Groppi A, Hantraye F, Hennequin C, Jauniaux N, Joyet P, Kachouri R, Kerrest A, Koszul R, Lemaire M, Lesur I, Ma L, Muller H, Nicaud JM, Nikolski M, Oztas S, Ozier-Kalogeropoulos O, Pellenz S, Potier S, Richard GF, Straub ML, Suleau A, Swennen D, Tekaia F, Wesolowski-Louvel M, Westhof E, Wirth B, Zeniou-Meyer M, Zivanovic I, Bolotin-Fukuhara M, Thierry A, Bouchier C, Caudron B, Scarpelli C, Gaillardin C, Weissenbach J, Wincker P, Souciet JL. 2004. Genome evolution in yeasts. Nature **430:**35-44.

Guthrie, C., and G. R. Fink, eds. 1991. Guide to Yeast Genetics and Molecular Biology. Methods in Enzymology, Vol.194, Academic Press.

Hellinga, H.W. Nature Structural. Biol. 1998, 5, 525-7.

Hespell, R. B. 1998. Extraction and characterization of hemicellulose from the corn fiber produced by corn wet-milling processes. J. Agric. Food Chem. **46:**2615-2619.

Hill, J., K. Ian, G. Donald, and D. Griffiths. 1991. DMSO-enhanced whole cell yeast transformation. Nucleic Acids Research **19:**5791.

Hofmann, K., and S. W. 1993. A database of membrane spanning proteins segments. Biol. Chem. 374:166.

Kötter, P., R. Amore, C. P. Hollenberg, and M. Ciriacy. 1990. Isolation and characterization of the *Pichia stipitis* xylitol dehydrogenase gene, *XYL2*, and construction of a xylose-utilizing *Saccharomyces cerevisiae* transformant. Curr. Genet. **18:**493-500.

Kruckeberg, A. 1996. The hexose trasnporter family of Saccharomyces cerevisiae. Arch. Microbiol. **166:**283-292.

Kou, S., M. Christensen, and V. Cirillo. 1970. Galactose transport in *Saccharomyces cerevisiae* II. Characteristics of galactose uptake and exchange in galactokinaseless cells. J Bacteriol **103:**671-678.

Landschultz et al., 1988, *Science*, 240:1759.

Luckow and Summers, 1988 *Bio/Technology* 6:47.

McMillan, J. D., and B. L. Boynton. 1994. Arabinose utilization by xylose-fermenting yeasts and fungi. Appl. Biochem. Biotechnol. **45-46:**569-584.

Pina, C., P. Goncalves, C. Prista, and M. Loureiro-Dias. 2004. Ffz1, a new transporter specific for fructose from *Zygosaccharomyces bailii*. Microbiol **150:**2429-2433.

Saitou, N., and M. Nei. 1987 The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol. Jul;4(4):406-25.

Sambrook, J., E. Fritsch, and T. Maniatis. 1989. Molecular Cloning: a Laboratory Manual, 2nd. ed. Cold Spring Harbor Laboratory Press, NY.

Sedlak, M., and N. W. Ho. 2001. Expression of *E. coli araBAD* operon encoding enzymes for metabolizing L-arabinose in *Saccharomyces cerevisiae*. Enzyme Microb. Technol. **28:**16-24.

Stambuk, B., M. Franden, A. Singh, and M. Zhang. 2003. D-Xylose transport by Candida succiphila and Kluyveromyces marxianus. Appl Biochem Biotechnol **105-108:**255-263.

Van den Burg, B.; Vriend, G.; Veltman, O.R.; Venema, G.; Eijsink, V.G.H. 1998. Engineering an enzyme to resist boiling. Proc. Nat. Acad. Sci. U.S., 95:2056-60.

Wahlbom, C. F., and B. Hahn-Hägerdal. 2002. Furfural, 5-hydroxymethyl furfural, and acetoin act as external electron acceptors during anaerobic fermentation of xylose in recombinant *Saccharomyces cerevisiae*. Biotechnol. Bioeng. **78:**172-178.

Weierstall, T., C. Hollenberg, and E. Boles. 1999. Cloning and characterization of three genes (SUT1-3) encoding glucose transporters of the yeast *Pichia stipitis*. Mol Microbiol **31:**871-883.

Zhang, M., C. Eddy, K. Deanda, M. Finkelstein, and S. Picataggio. 1995. Metabolic engineering of a pentose metabolism pathway in ethanologenic *Zymomonas mobilis*. Science **267:**240-243.

Zhao, H.; Giver, L.; Shao, Z.; Affholter, J.A.; Arnold, F.H. Nature Biotechnol. 1998, 16, 258-62.

CLAIMS

What is claimed is:

1. A isolated non-conventional yeast arabinose transporter comprising the capability of adapting a conventional yeast for growth on arabinose when the arabinose transporter is included in the conventional yeast.

2. The arabinose transporter of claim 1 wherein the conventional yeast is *S. cerevisiae*.

3. The arabinose transporter of claim 1 wherein the non-conventional yeast is *A. adeninivorans*, *K. marxianus* or *P. guilliermondii*.

4. The arabinose transporter of claim 3 wherein the transporter has at least 95% identity to a sequence of SEQ ID NO: 2

5. The arabinose transporter of claim 4 wherein the transporter has a sequence of SEQ ID NO: 2.

6. The arabinose transporter of claim 3 wherein the transporter has at least 95% identity to a sequence of SEQ ID NO: 4.

7. The arabinose transporter of claim 6 wherein the transporter has a sequence of SEQ ID NO: 4.

8. The arabinose transporter of claim 1 further comprising high affinity arabinose transport into the conventional yeast.

9. A vector comprising the polynucleotide of claim 1.

10. A method of identifying a yeast arabinose transporter comprising:
obtaining a yeast strain for screening for presence of an arabinose transporter;
classifying the yeast strain based on ability to utilize L-arabinose as a sole source of fermentation;

determine if yeast strain is amenable to genetic and biochemical testing procedures; and

determine if yeast strain has single component high affinity arabinose transport;

wherein yeast strains having the capacity to grow on L-arabinose, are amenable to genetic and biochemical manipulation and have a single component high affinity arabinose transport are identified as including an arabinose transporter.

11. The method of claim 10 wherein the yeast strain is a non-conventional yeast strain.

12. The method of claim 11 wherein the non-conventional yeast strain is selected from a group consisting of *A. adeninivorans*, *K. marxianus* or *P. guilliermondii*.

13. The method of claim 10 wherein high affinity transport is arabinose transport less than a Km of 1mm and a Vmax of at least 15mmol/mg-minute.

14. The method of claim 10 wherein the genetic and biochemical testing procedures include replica plating.

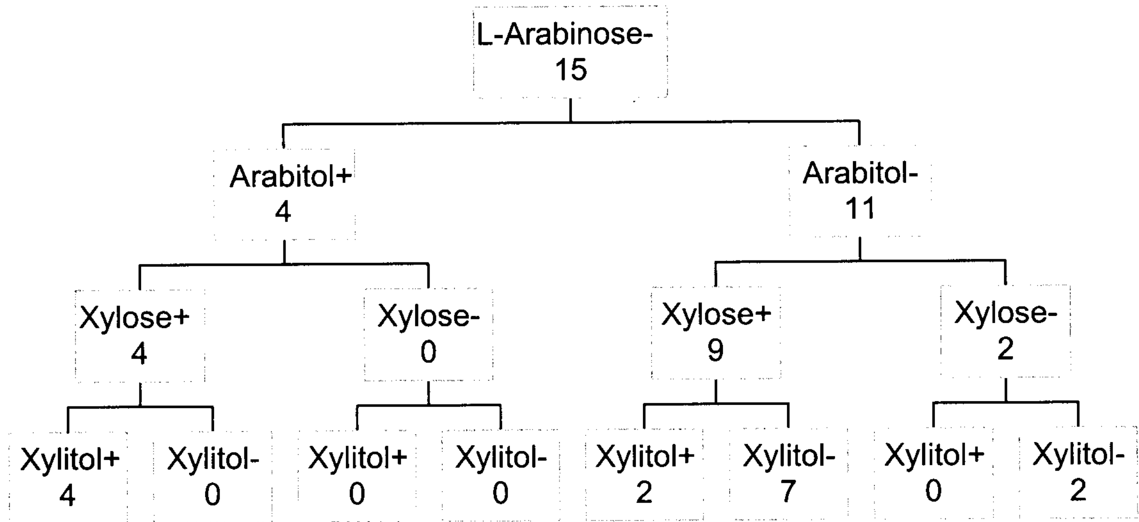


FIG. 1.

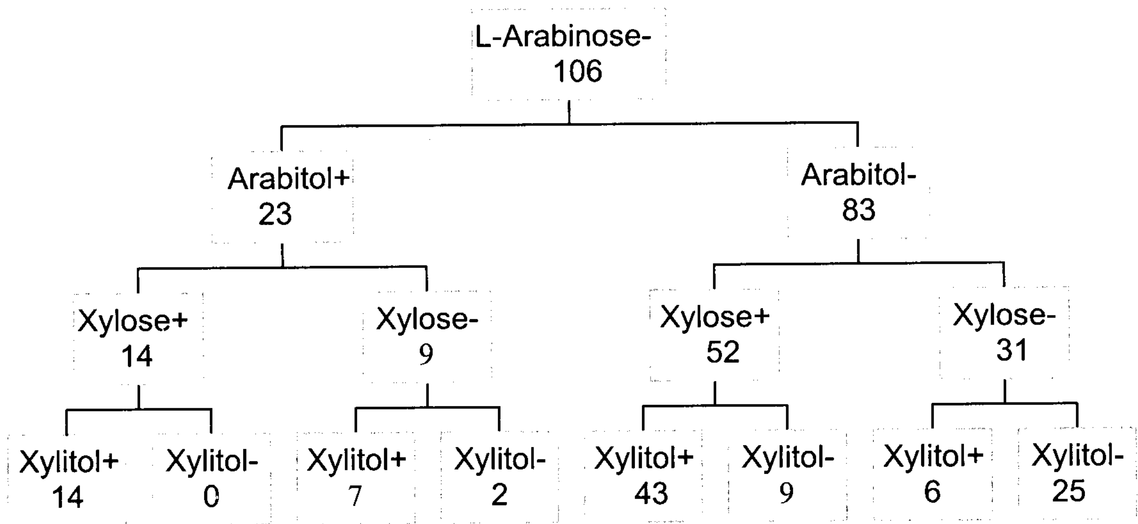


FIG. 2

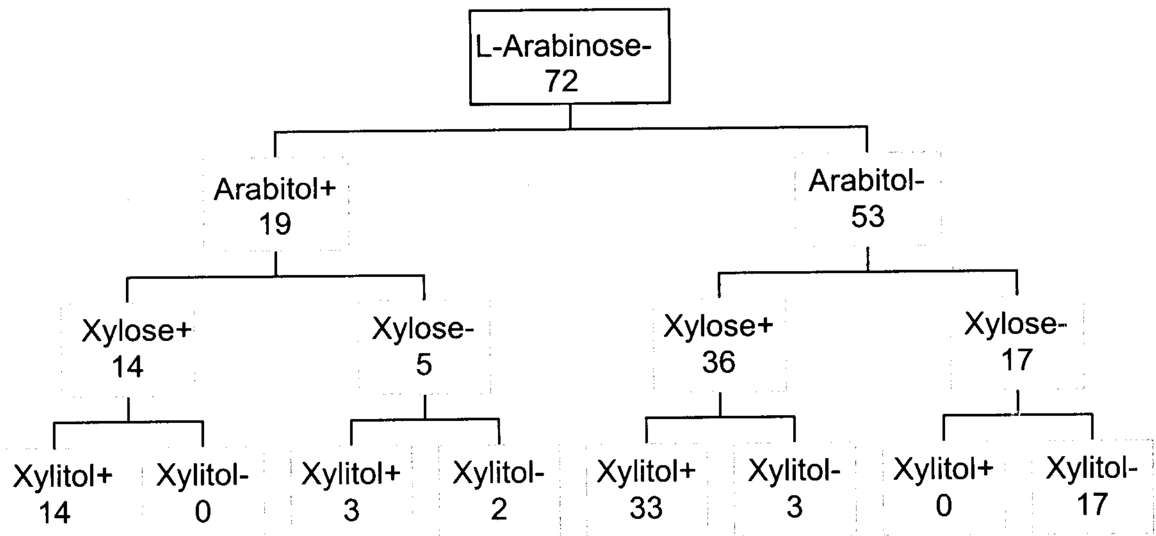


FIG. 3.

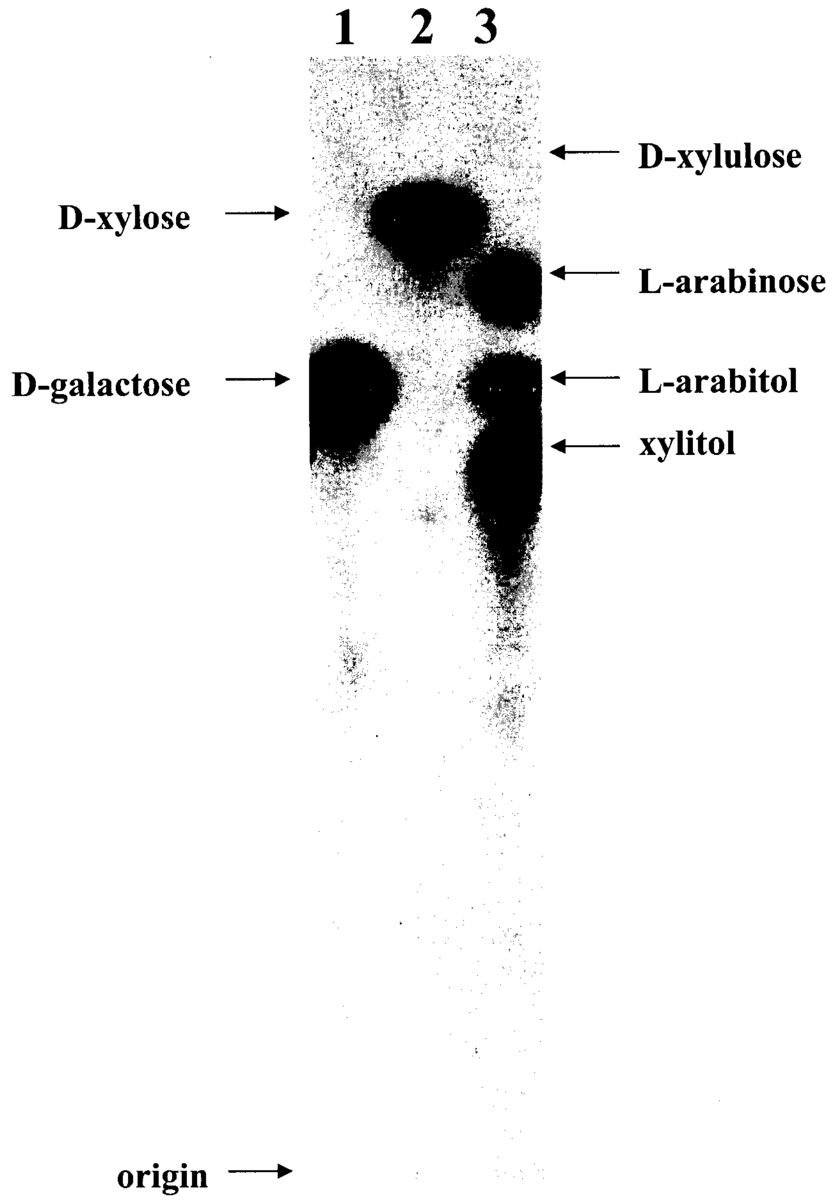
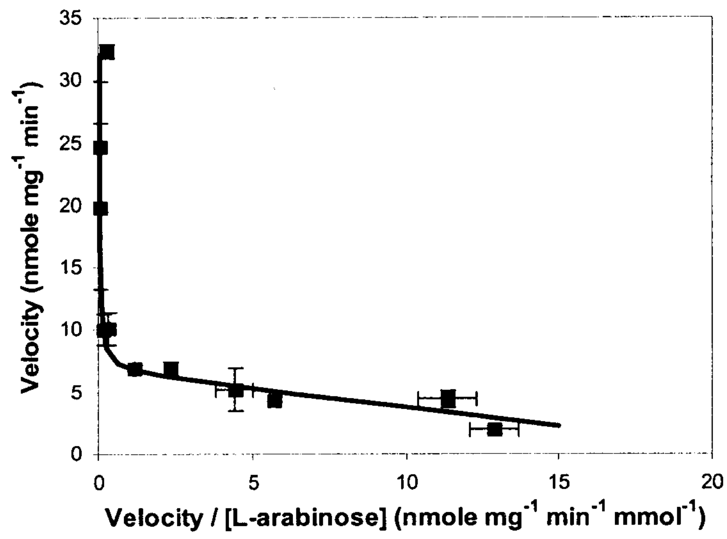


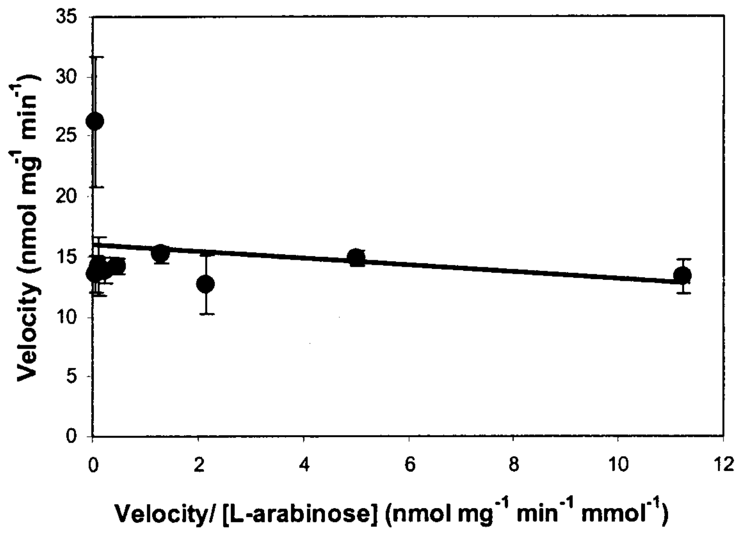
FIG. 4.

FIG. 5

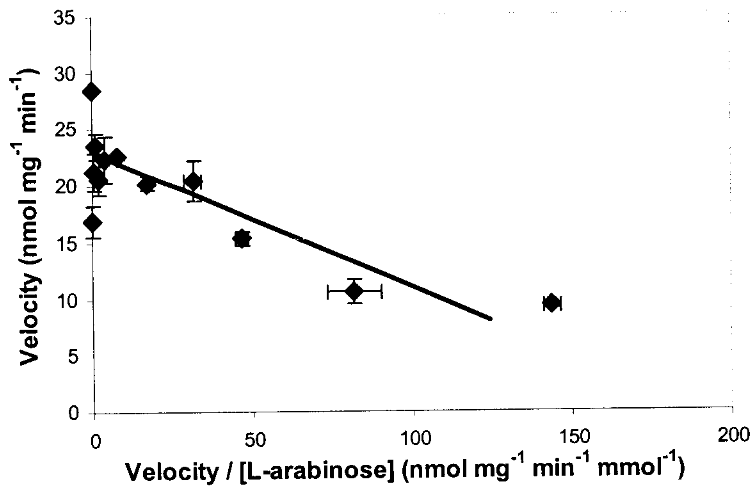
A:



B:



C:



D:

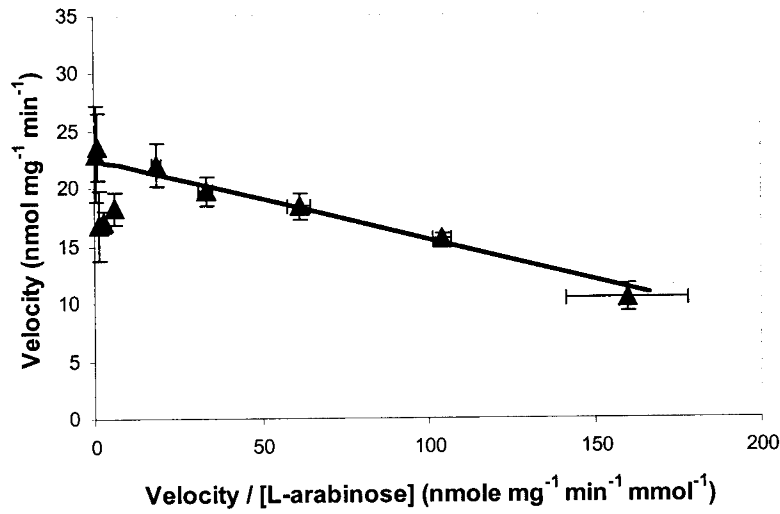


FIG. 6

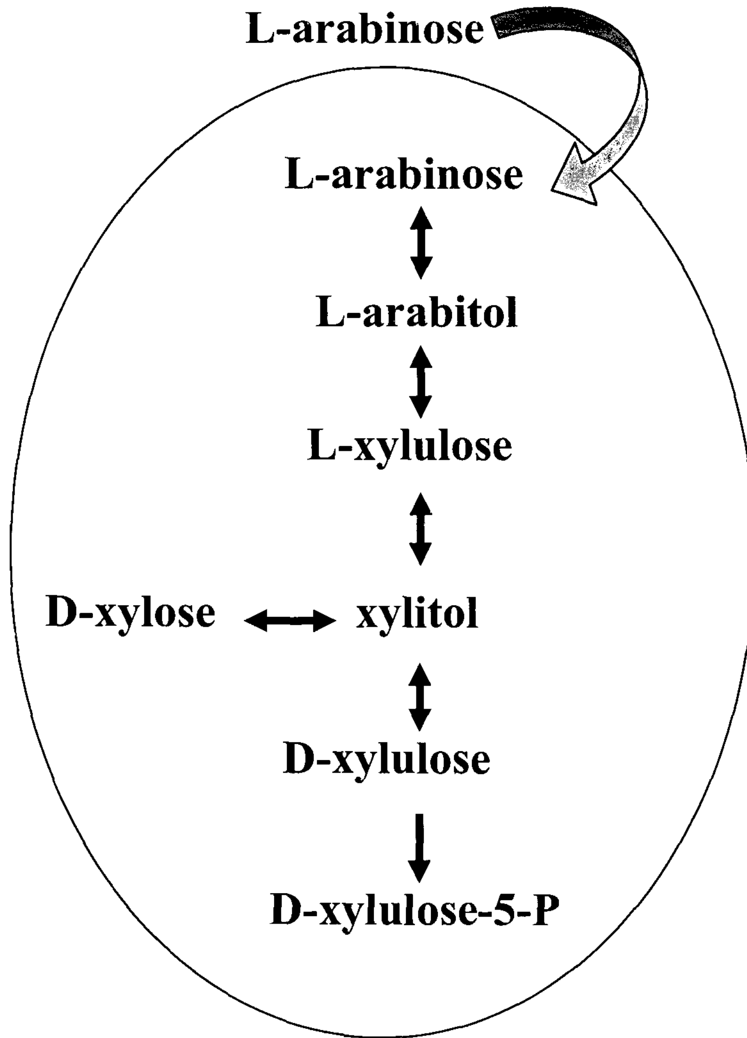


FIG. 7

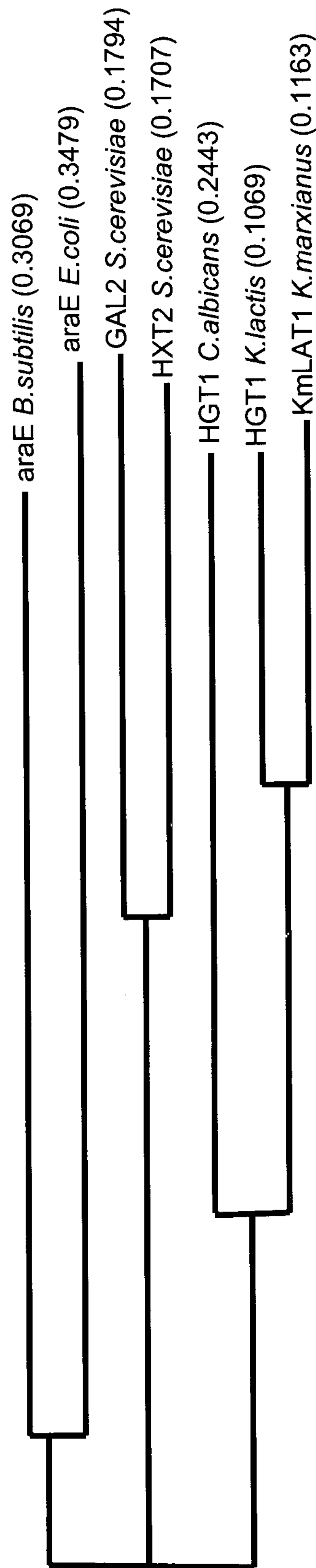


FIG. 8

```
     M T L K  D K L  L L R  N I E F  K G T  F Y A  K F P Q  I H N  I Y A  I G V   33 SEQ ID NO 2
   1 ATGACTTTAA AAGATAAACT ATTGCTCCGC AATATCGAAT TCAAGGGAAC TTTCTATGCG AAGTTCCCGC AAATTCACAA CATTTACGCA ATCGGTGTGA    SEQ ID NO 1

     I S C I  S G L  M F G F  D I S  S M S  S M I G  T E T  Y K K  Y F D H  P K S   67
 101 TTTCGTGTAT ATCTGGTCTC ATGTTTGGTT TCGATATCTC TTCAATGTCT TCCATGATCG GTACTGAAAC TTACAAAAAA TATTTTGACC ATCCAAAATC

     I T Q  G G I T  A S M  S G G  S F L G  S L L  S P A  I S D T  F G R  K V S  100
 201 CATTACCCAA GGTGGTATCA CCGCGTCAAT GTCCGGTGGT TCCTTCTTAG GCTCTTTACT CTCTCCTGCT ATTTCCGATA CCTTTGGCAG AAAAGTGTCG

     L H I C  A V L  W I V  G C I L  Q S A  A Q D  Q P M L  I A G  R V I  A G L  133
 301 TTGCACATTT GTGCCGTCTT GTGGATCGTC GGATGCATTT TGCAAAGTGC TGCCCAAGAC CAACCAATGC TAATCGCTGG CCGTGTTATC GCAGGGTTGG

     G I G F  G S G  S A P I  Y C S  E I S  P P K V  R G L  I T G  L F Q F  S I T  167
 401 GTATCGGGTT CGGCTCTGGT TCTGCTCCAA TTTACTGTTC TGAAATCTCC CCACCAAAGG TTAGAGGCTT GATCACCGGT CTTTTCCAGT TCTCTATCAC

     V G I  M I L F  Y V G  Y G C  H F L S  G N L  S F R  L T W G  L Q V  I P G  200
 501 TGTTGGTATT ATGATTCTCT TCTACGTTGG TTACGGGTGC CACTTCCTCA GTGGTAATCT TTCATTCAGA TTGACTTGGG GTTTGCAAGT TATCCCAGGA

     F V L L  V G V  L F L  P E S P  R W L  A N H  D R W E  E T E  S I V  A K V  233
 601 TTTGTGTTGC TGGTCGGTGT CCTATTCTTG CCAGAATCCC CACGTTGGTT GGCTAACCAC GACCGTTGGG AAGAAACTGA GTCAATCGTC GCCAAGGTCG

     V A K G  N V D  D E E V  K F Q  L E E  I K E Q  V I L  D A A  A K N F  S F K  267
 701 TCGCCAAGGG TAACGTAGAC GATGAAGAAG TCAAGTTCCA ATTGGAAGAA ATTAAAGAGC AGGTGATTCT TGATGCTGCC GCCAAGAACT TCTCCTTCAA

     D L L  R P K T  R K K  L F V  G V C A  Q M W  Q Q L  C G M N  V M M  Y Y I  300
 801 GGATTTGCTA AGACCAAAGA CCAGAAAGAA GCTCTTTGTT GGTGTGTGTG CTCAAATGTG GCAACAATTG TGTGGTATGA ACGTTATGAT GTACTACATT

     V Y V F  N M A  G Y T  G N T N  L V A  S S I  Q Y V L  N V L  M T F  P A L  333
 901 GTGTACGTCT TTAACATGGC TGGTTACACT GGTAACACCA ACTTGGTTGC ATCTTCCATT CAATACGTCT TGAACGTGCT AATGACTTTC CCTGCACTAT

     F L I D  K V G  R R P V  L I V  G G I  F M F T  W L F  A V A  G L L A  S Y S  367
1001 TCTTAATCGA TAAAGTCGGT AGAAGACCTG TCTTGATCGT TGGTGGTATT TTCATGTTCA CATGGTTGTT CGCTGTCGCT GGTTTGTTGG CATCATATTC

     V P A  P N G V  N G D  D T V  T I R I  P D K  H K S  A A K G  V I A  C S Y  400
1101 CGTCCCAGCT CCAAATGGTG TTAACGGTGA TGATACTGTC ACAATCAGAA TCCCAGACAA GCACAAGTCC GCCGCTAAGG GTGTCATTGC ATGTTCATAC

     L F V C  S F A  P T W  G I G I  W I Y  C S E  I F N N  M E R  A K G  S S V  433
1201 TTGTTCGTCT GCTCTTTCGC TCCAACCTGG GGTATTGGTA TCTGGATTTA CTGTTCCGAA ATTTCAACA ACATGGAAAG AGCCAAGGGT TCCTCTGTGG

     A A A T  N W A  F N F  A L A M  F V P  S A F  K N I S  W K T  Y I V  F G V F  467
1301 CTGCTGCTAC CAACTGGGCA TTCAACTTCG CTTTGGCGAT GTTCGTCCCA TCTGCATTCA AGAACATCTC ATGGAAAACA TACATCGTCT TTGGTGTCTT

     S V A  L T V Q  T Y F  M F P  E T R G  K T L  E E I  D Q M W  V D N  I P A  500
1401 TTCAGTTGCA TTGACTGTCC AAACCTACTT CATGTTCCCA GAAACTAGAG GTAAGACCTT GGAAGAAATC GACCAAATGT GGGTCGACAA CATCCCAGCC

     W K T S  S Y I  P Q L  P I I E  D E F  G N K  L G L L  G N P  Q H L  E H V  533
1501 TGGAAGACTA GCAGCTACAT CCCACAATTG CCTATCATCG AAGATGAATT TGGTAACAAG TTGGGTTTGT TGGGTAACCC ACAACATCTC GAGCATGTTA

     K S V E  K D T  V V E  K L E S  S E A  N S S  S S V *                          556
1601 AATCCGTCGA AAAGGATACT GTAGTGGAAA AATTAGAATC GTCAGAGGCT AATAGCAGCA GCTCGGTCTA G
```

FIG. 9

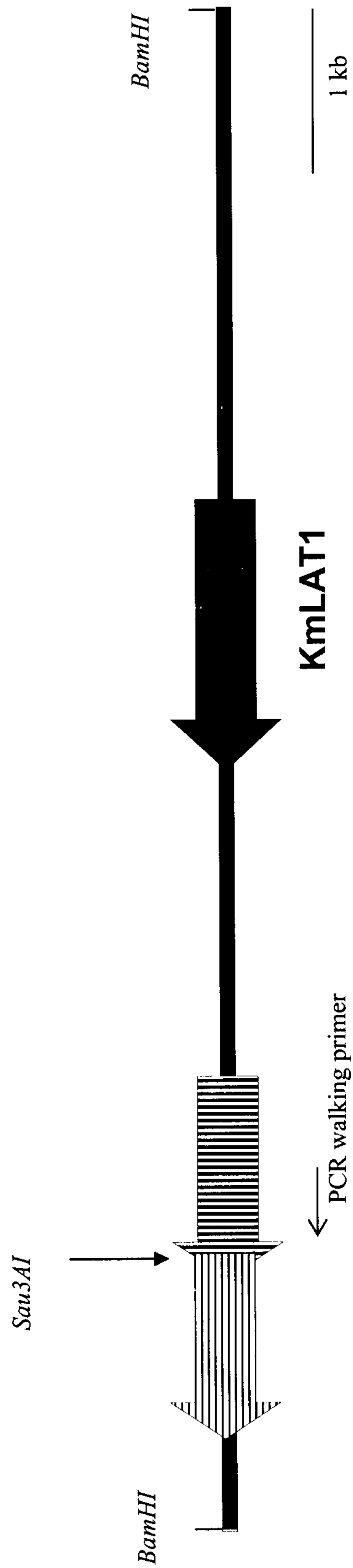


FIG. 10

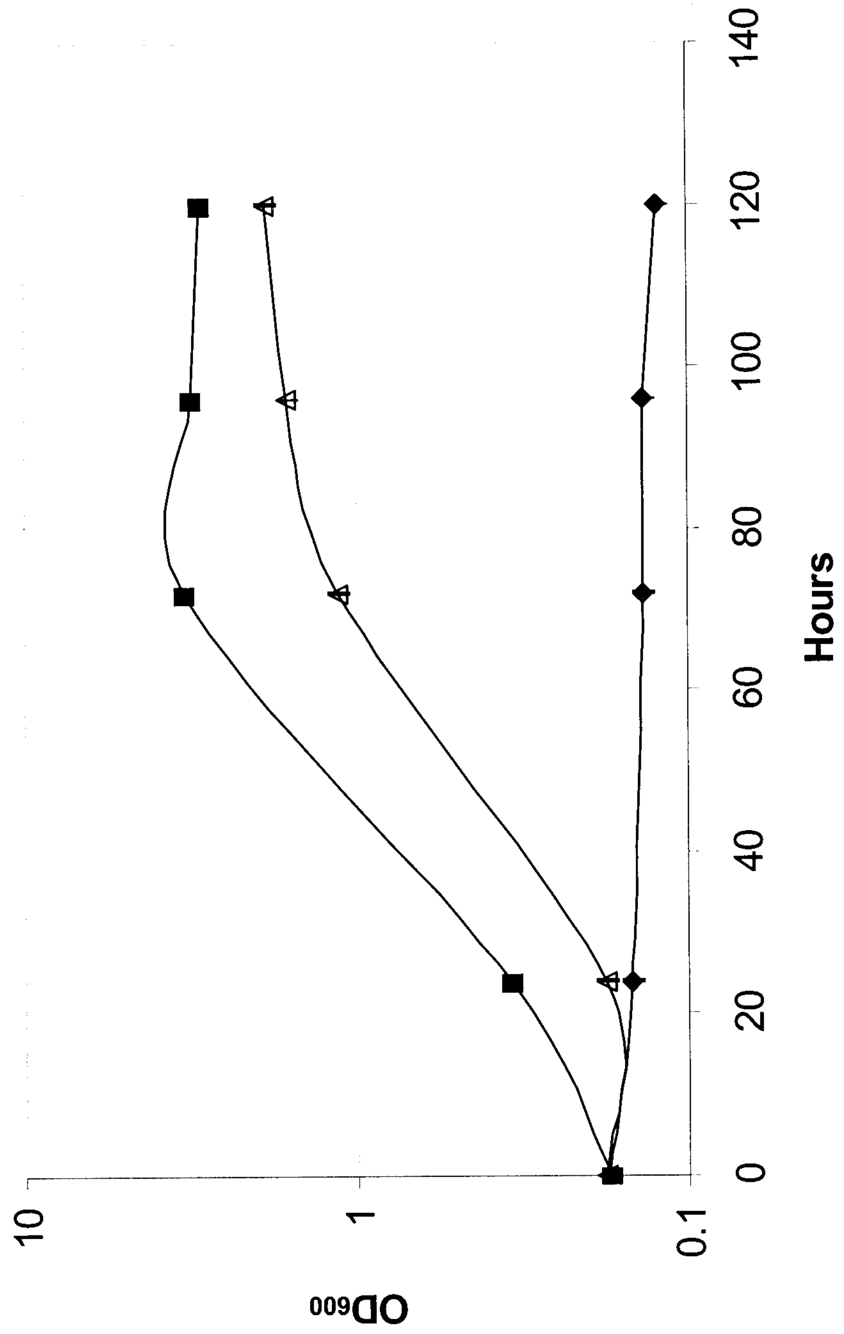


FIG. 11

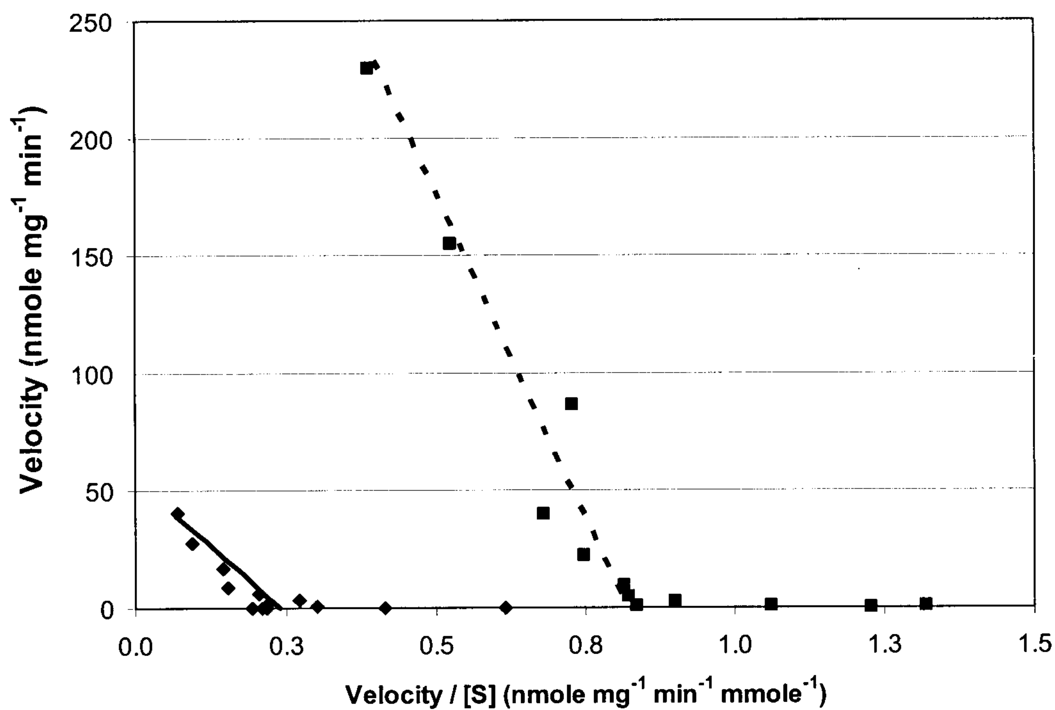


FIG. 12A

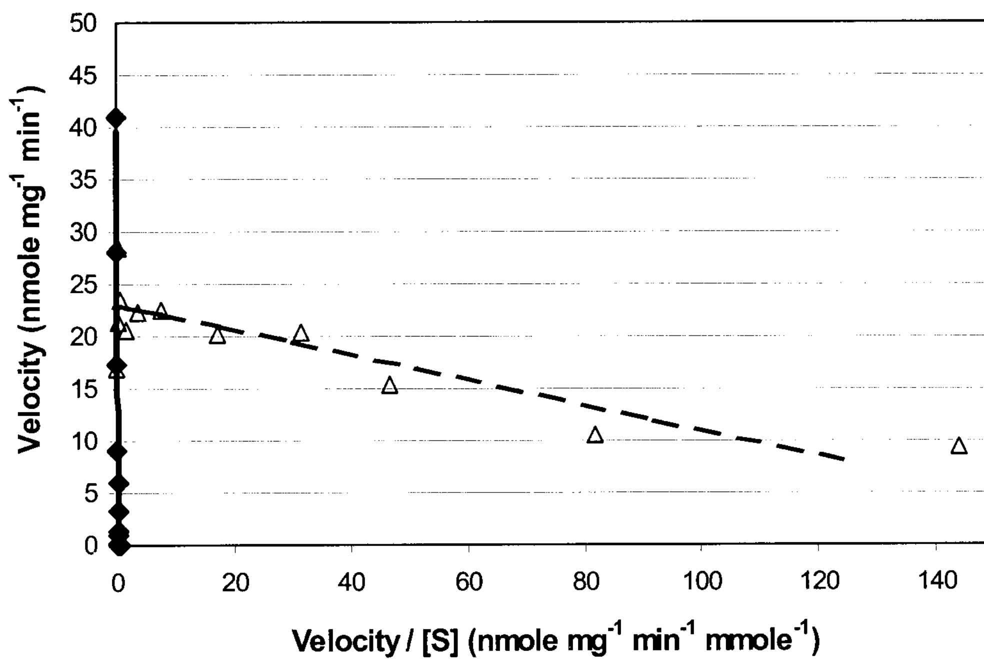


FIG. 12B

```
SEQ ID No4: 1  M  A  Y  E  D  K  L  V  A  P  A  L  K  F  R  N  F  L  D  K
SEQ ID No3: 1 ATGGCTTACGAGGACAAACTAGTGGCTCCGGCCTTGAAGTTTAGAAACTTTCTTGACAAA
        21  T  P  N  I  Y  N  P  Y  I  I  S  I  I  S  C  I  A  G  M  M
        61  ACTCCCAATATCTACAATCCATATATCATTTCTATAATCTCGTGCATTGCGGGTATGATG
        41  F  G  F  D  I  S  S  M  S  A  F  V  S  L  P  A  Y  V  N  Y
       121  TTCGGTTTTGATATTTCTTCAATGTCAGCGTTTGTCAGTTTACCAGCATACGTGAATTAT
        61  F  D  T  P  S  A  V  I  Q  G  F  I  T  S  A  M  A  L  G  S
       181  TTCGATACACCTTCAGCAGTGATTCAAGGATTTATCACATCTGCCATGGCTTTGGGTTCA
        81  F  F  G  S  I  A  S  A  F  V  S  E  P  F  G  R  R  A  S  L
       241  TTTTTCGGGTCAATTGCTTCTGCGTTTGTGTCTGAGCCATTTGGAAGACGAGCTTCCTTA
       101  L  T  C  S  W  F  W  M  I  G  A  A  I  Q  A  S  S  Q  N  R
       301  CTAACTTGTTCGTGGTTTTGGATGATAGGAGCAGCCATCCAAGCGTCTTCGCAGAACCGA
       121  A  Q  L  I  I  G  R  I  I  S  G  F  G  V  G  F  G  S  S  V
       361  GCTCAATTGATTATTGGTCGGATTATATCTGGATTTGGGGTTGGTTTCGGGTCGTCTGTG
       141  A  P  V  Y  G  S  E  M  A  P  R  K  I  R  G  R  I  G  G  I
       421  GCTCCCGTATATGGCTCCGAGATGGCACCTAGAAAAATTAGAGGAAGAATTGGTGGAATT
       161  F  Q  L  S  V  T  L  G  I  M  I  M  F  F  I  S  Y  G  T  S
       481  TTTCAATTATCTGTCACCCTCGGTATCATGATTATGTTCTTCATAAGTTACGGAACTTCT
       181  H  I  K  T  A  A  A  F  R  L  A  W  A  L  Q  I  I  P  G  L
       541  CATATTAAGACTGCGGCAGCTTTCAGGTTAGCCTGGGCACTCCAGATCATTCCTGGACTC
       201  L  M  C  I  G  V  F  F  I  P  E  S  P  R  W  L  A  K  Q  G
       601  CTCATGTGTATTGGTGTCTTCTTTATTCCAGAATCTCCTAGATGGTTGGCCAAACAAGGT
       221  H  W  D  E  A  E  I  I  V  A  K  I  Q  A  K  G  D  R  E  N
       661  CACTGGGACGAAGCCGAAATCATTGTAGCCAAAATTCAAGCCAAAGGAGATCGAGAAAAT
       241  P  D  V  L  I  E  I  S  E  I  K  D  Q  L  M  V  D  E  N  A
       721  CCCGATGTTTTGATTGAAATTTCGGAAATAAAAGACCAATTGATGGTTGACGAGAATGCC
       261  K  A  F  T  Y  A  D  L  F  S  K  K  Y  L  P  R  T  I  T  A
       781  AAAGCCTTTACCTATGCTGACTTGTTTTCGAAAAAATATCTTCCCAGAACCATCACAGCC
       281  M  F  A  Q  I  W  Q  Q  L  T  G  M  N  V  M  M  Y  Y  I  V
       841  ATGTTCGCTCAAATCTGGCAACAATTGACAGGAATGAATGTCATGATGTACTATATCGTT
       301  Y  I  F  E  M  A  G  Y  G  G  N  G  V  L  V  S  S  T  I  Q
       901  TACATTTTCGAAATGGCTGGCTACGGTGGAAATGGAGTGTTGGTATCATCGACAATTCAG
       321  Y  V  I  F  V  V  V  T  F  V  S  L  F  F  L  D  K  F  G  R
       961  TACGTTATCTTTGTCGTTGTTACATTTGTCTCATTATTCTTTTTGGACAAATTTGGAAGA
       341  R  K  I  L  L  V  G  A  A  S  M  M  T  W  Q  F  A  V  A  G
      1021  AGAAAAATTTTACTTGTCGGAGCAGCTTCCATGATGACCTGGCAGTTTGCAGTGGCAGGG
       361  I  L  A  R  Y  S  V  P  Y  D  L  S  D  T  V  K  I  K  I  P
      1081  ATCTTGGCCAGGTACTCGGTCCCGTACGATCTCAGCGATACTGTCAAAATTAAAATTCCT
       381  D  N  H  K  S  A  A  K  G  V  I  A  C  C  Y  L  F  V  A  S
      1141  GACAATCACAAATCGGCTGCAAAAGGTGTCATTGCATGCTGCTATCTTTTCGTAGCATCG
       401  F  G  F  S  W  G  V  G  I  W  L  Y  C  S  E  V  W  G  D  S
      1201  TTCGGATTTTCCTGGGGAGTTGGTATCTGGTTATACTGCTCTGAAGTCTGGGGAGACTCA
       421  Q  S  R  Q  R  G  A  A  V  S  T  A  S  N  W  I  F  N  F  A
      1261  CAATCGAGACAGAGAGGAGCCGCTGTGTCAACTGCTTCAAATTGGATTTTCAATTTTGCG
       441  L  A  M  F  T  P  S  S  F  K  N  I  T  W  K  T  Y  C  I  Y
      1321  CTCGCCATGTTCACACCATCTTCGTTTAAAAATATCACCTGGAAGACATACTGTATTTAT
       461  A  T  F  C  A  C  M  F  I  H  V  F  F  F  F  P  E  T  K  G
      1381  GCCACTTTCTGCGCATGTATGTTCATCCATGTGTTCTTCTTCTTCCCAGAAACCAAGGGG
       481  K  R  L  E  E  I  A  Q  I  W  E  E  K  I  P  A  W  K  T  T
      1441  AAGCGCTTGGAAGAAATTGCTCAAATTTGGGAAGAAAAAATTCCAGCTTGGAAAACCACC
       501  N  W  Q  P  H  V  P  L  L  S  D  H  E  L  A  E  K  I  N  A
      1501  AACTGGCAACCTCATGTTCCTTTGTTGTCGGACCACGAACTTGCGGAAAAGATCAATGCC
       521  E  H  V  E  N  V  N  S  R  E  Q  S  D  D  E  K  S  Q  V  *
      1561  GAACATGTGGAGAACGTGAATTCTAGGGAACAATCGGATGACGAGAAGTCGCAGGTATAA
```

FIG. 13

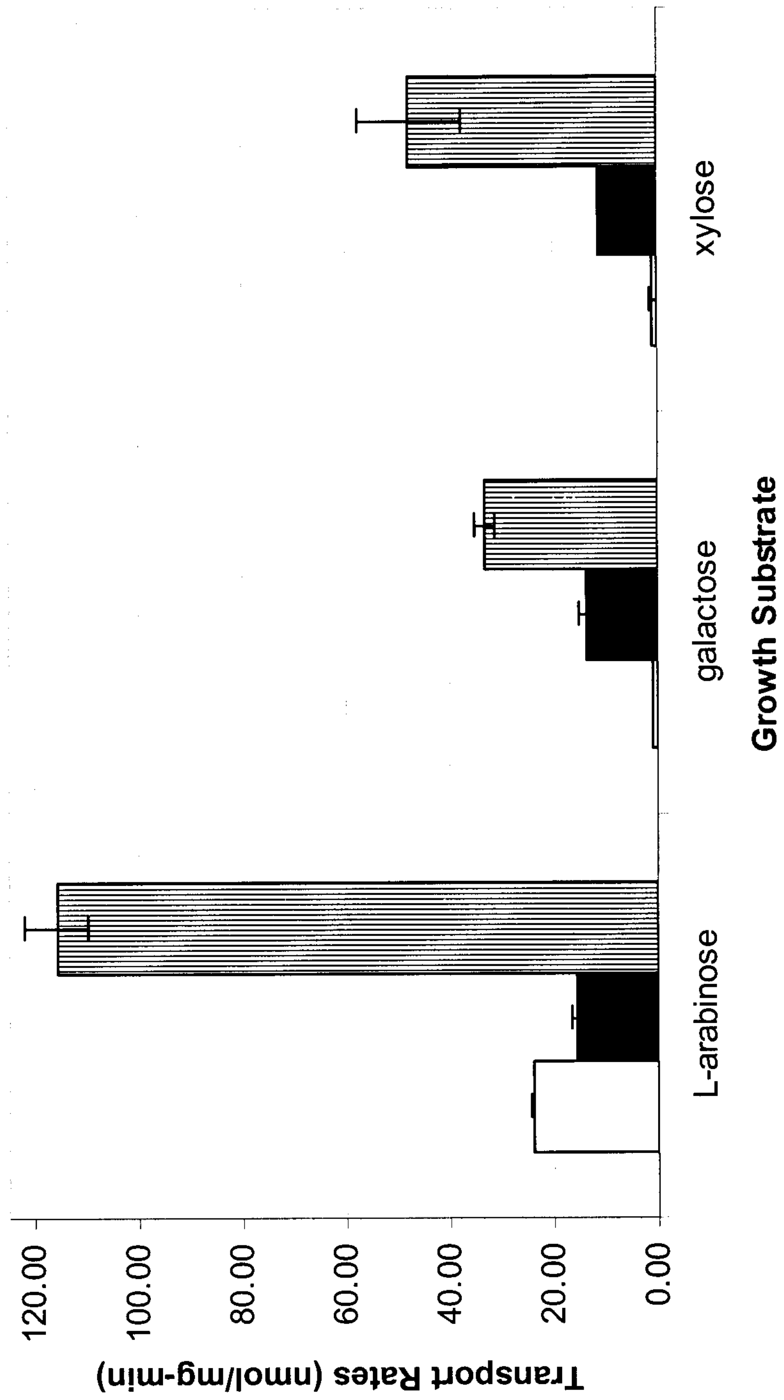


FIG. 14

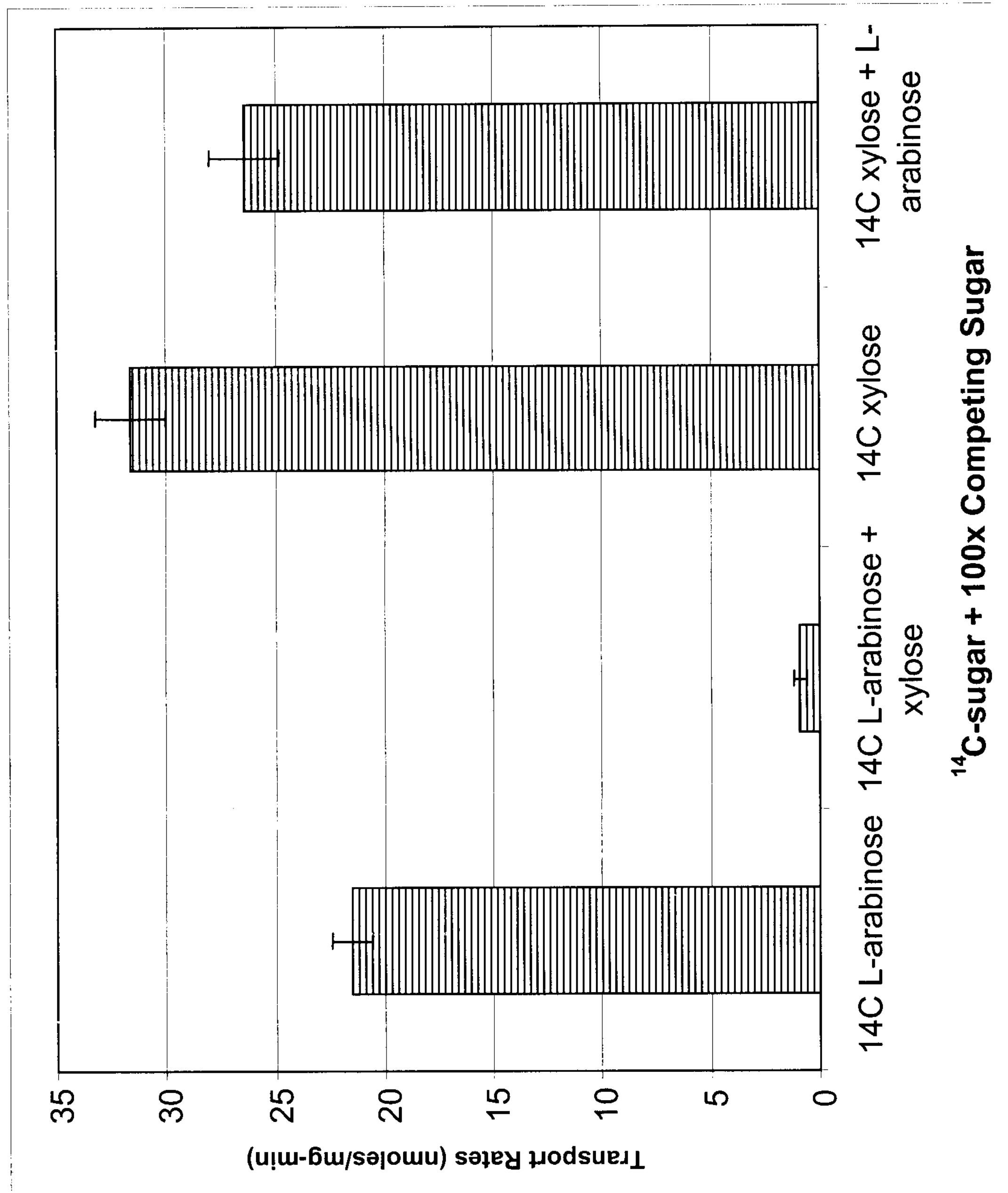


FIG. 15

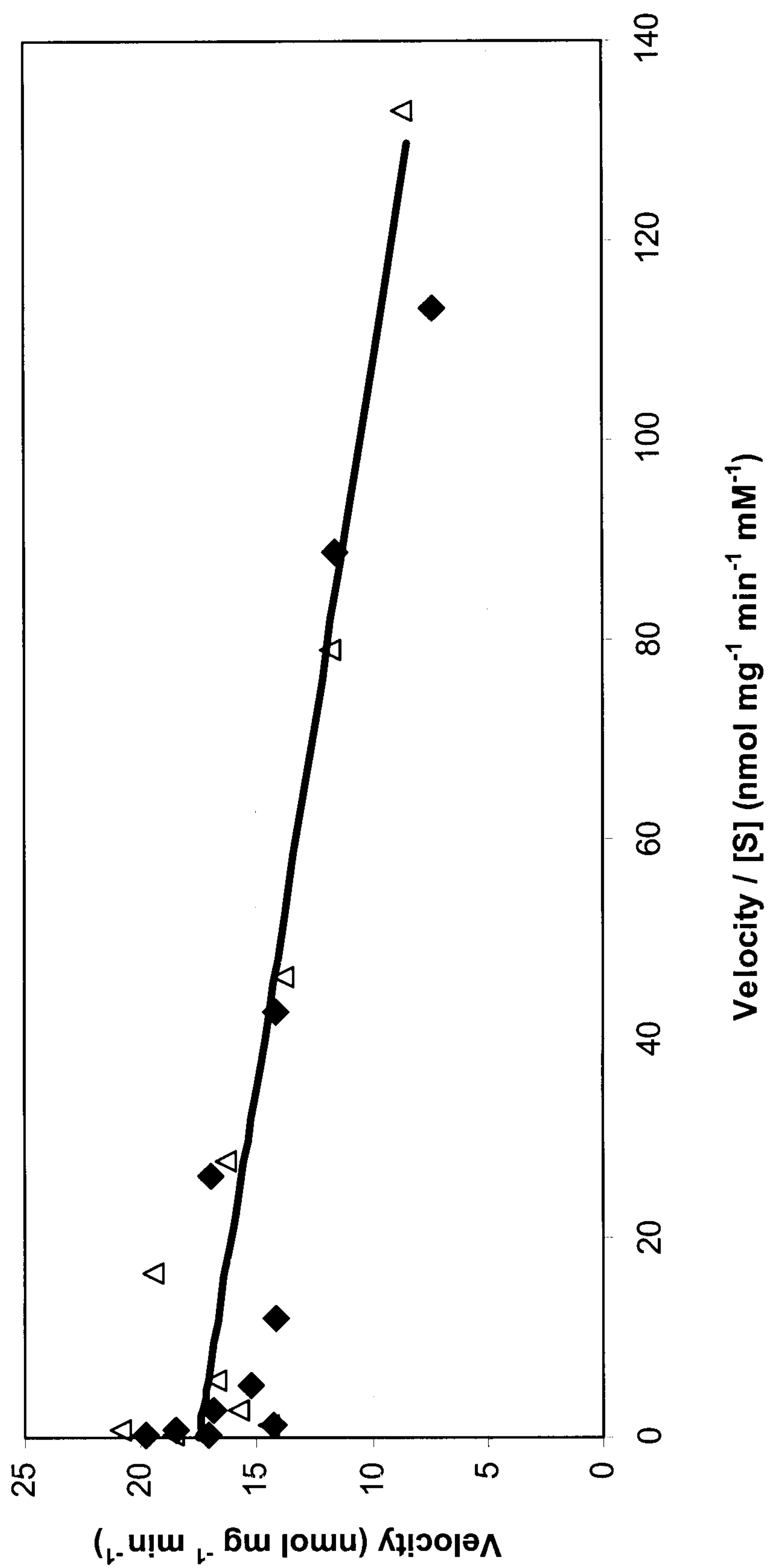


FIG. 16

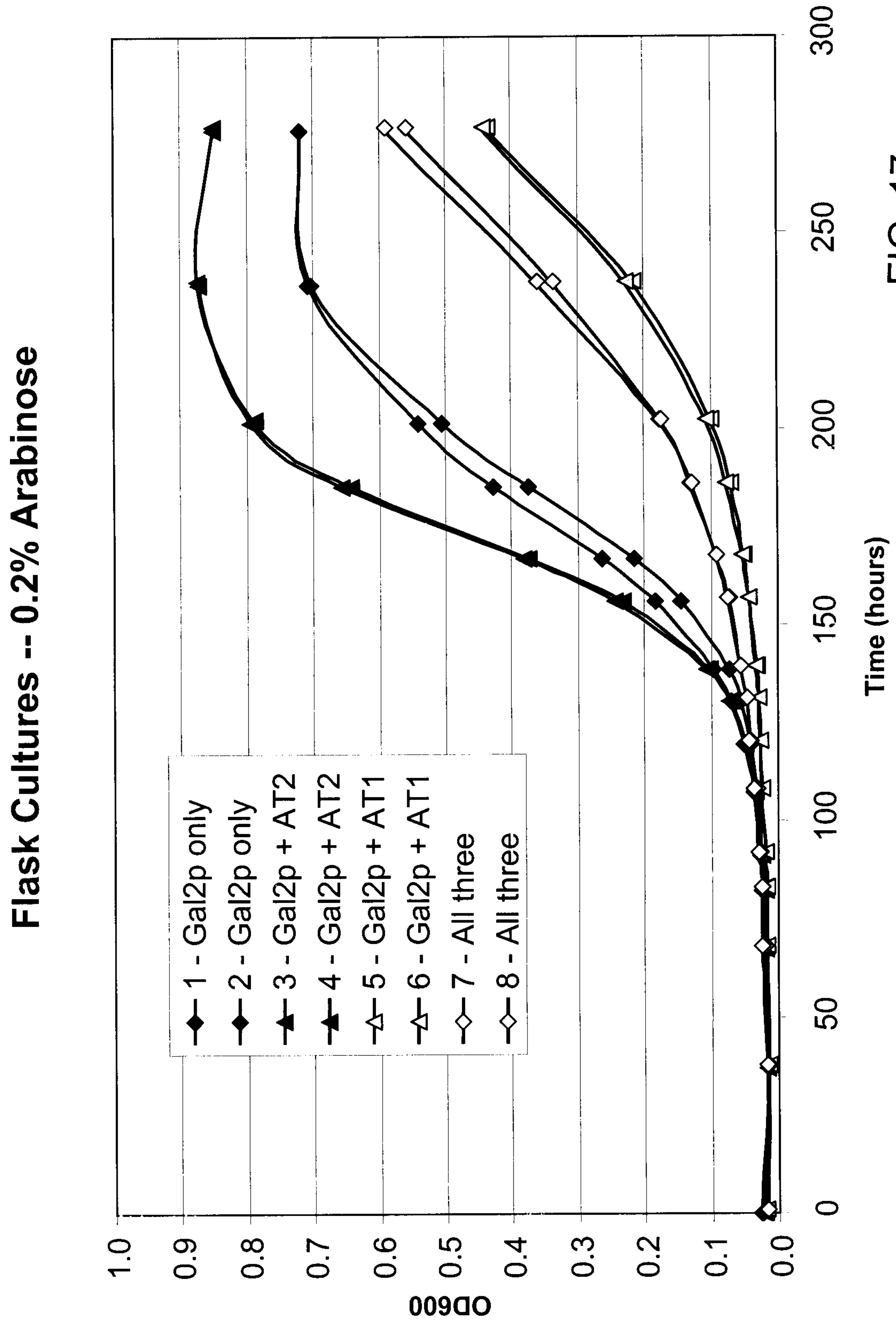


FIG. 17