



US008431360B2

(12) **United States Patent**
Glass et al.

(10) **Patent No.:** US 8,431,360 B2
(45) **Date of Patent:** Apr. 30, 2013

(54) **METHODS AND COMPOSITIONS FOR IMPROVING SUGAR TRANSPORT, MIXED SUGAR FERMENTATION, AND PRODUCTION OF BIOFUELS**

- (75) Inventors: **N. Louise Glass**, Orinda, CA (US); **Chaoguang Tian**, Beijing (CN); **William T. Beeson, IV**, Oakland, CA (US); **Huimin Zhao**, Champaign, IL (US); **Jing Du**, Champaign, IL (US); **Jin Ho Choi**, Urbana, IL (US); **James H. Doudna Cate**, Berkeley, CA (US); **Jonathan M. Galazka**, Berkeley, CA (US); **Suk-Jin Ha**, Champaign, IL (US); **Yong-Su Jin**, Champaign, IL (US); **Soo Rin Kim**, Urbana, IL (US); **Sijin Li**, Urbana, IL (US); **Xiaomin Yang**, Albany, CA (US)
- (73) Assignees: **The Regents of the University of California**, Oakland, CA (US); **The Board of Trustees of the University of Illinois**, Urbana, IL (US); **BP Corporation North America Inc.**, Warrenville, IL (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 168 days.

(21) Appl. No.: **12/843,844**

(22) Filed: **Jul. 26, 2010**

(65) **Prior Publication Data**

US 2011/0020910 A1 Jan. 27, 2011

Related U.S. Application Data

(60) Provisional application No. 61/271,833, filed on Jul. 24, 2009; provisional application No. 61/285,526, filed on Dec. 10, 2009.

(51) **Int. Cl.**

C12N 1/00 (2006.01)
C12P 21/04 (2006.01)
C12P 21/06 (2006.01)

(52) **U.S. Cl.**

USPC **435/69.1; 435/71.1; 435/243**

(58) **Field of Classification Search** None
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

5,100,791 A 3/1992 Spindler et al.
6,207,436 B1 3/2001 Bjornvad et al.
7,344,876 B2 3/2008 Levine
2008/0280325 A1 11/2008 Johansen et al.
2009/0053784 A1 2/2009 Travassos Leandro et al.
2011/0177542 A1 7/2011 Van zyl et al.

OTHER PUBLICATIONS

Galazka et al., "A New Diet for Yeast to Improve Biofuel Production", Bioengineered Bugs, vol. 2, Issue 4, Jul./Aug. 2011, pp. 1-4.

Glazka et al., "Improving the Bioconversion of Plant Biomass to Biofuels: A Multidisciplinary Approach", Energy & Environmental Science, vol. 4, 2011, pp. 3329-3333.

Ha et al., "Cofermentation of Cellobiose and Galactose by an Engineered *Saccharomyces cerevisiae* Strain", Applied and Environment Microbiology, vol. 77, No. 16, Aug. 2011, pp. 5822-5825.

Ha et al., "Engineered *Saccharomyces cerevisiae* Capable of Simultaneous Cellobiose and Xylose Fermentation", Proceeding of the National Academy of Sciences, vol. 108, No. 2 Jan. 11, 2011, pp. 504-509.

Ha, et al., "Xylitol Does Inhibit Xylose Fermentation by Engineered *Saccharomyces cerevisiae* Expressing XyIA as Severely as its Inhibits Xylose Isomerase Reaction in Vitro", Applied Microbiology and Biotechnology, 2011, 8 pages.

Hamacher et al., "Characterization of the Xylose-transporting Properties of Yeast Hexose Transporters and their Influence on Xylose Utilization", Microbiology, vol. 148, 2002, pp. 2783-2788.

Koning et al., "Cellobiose Uptake in the Hyperthermophilic Archaeon *Pyrococcus furiosus* Is Mediated by an Inducible, High-Affinity ABC Transporter", Journal of Bacteriology, vol. 183, No. 17, Sep. 2001, pp. 4979-4984.

Leandro et al., "The Expression *Saccharomyces cerevisiae* of Glucose/xylose Symporter from *Candida intermedia* is Affected by Presence of a Glucose/xylose Facilitator", Microbiology, vol. 154, 2008, pp. 1646-1655.

Li et al., "Overcoming Glucose Repression in Mixed Sugar Fermentation by Co-expressing a Cellobiose Transporter and a β -glucosidase in *Saccharomyces cerevisiae*", Molecular BioSystems, vol. 6, 2010, pp. 2129-2132.

Pao et al., "Major Facilitator Superfamily", Microbiology and Molecular Biology Reviews, vol. 62, No. 1, Mar. 1998, pp. 1-34.

Ren et al., "TransportDB: A Relational Database of Cellular Membrane Transport Systems", Nucleic Acids Research, vol. 32, Database Issue, 2004, pp. D284-D288.

Suzuki et al., "Cellotriose and Cellobetraose as Inducers of the Genes Encoding Cellobiohydrolases in the Basidiomycete Phanerochaete chrysosporium", Applied and Environmental Microbiology, vol. 76, No. 18, Sep. 2010, pp. 6164-6170.

(Continued)

Primary Examiner — Albert Navarro

(74) **Attorney, Agent, or Firm — Morrison & Foerster LLP**

(57) **ABSTRACT**

The present disclosure relates to host cells containing a recombinant polynucleotide encoding a polypeptide where the polypeptide transports celldextrin into the cell. The present disclosure further relates to methods of increasing transport of celldextrin into a cell, methods of increasing growth of a cell on a medium containing celldextrin, methods of co-fermenting cellulose-derived and hemicellulose-derived sugars, and methods of making hydrocarbons or hydrocarbon derivatives by providing a host cell containing a recombinant polynucleotide encoding a polypeptide where the polypeptide transports celldextrin into the cell. The present disclosure relates to host cells containing a recombinant polynucleotide encoding a polypeptide where the polypeptide transports a pentose into the cell, methods of increasing transport of a pentose into a cell, methods of increasing growth of a cell on a medium containing pentose sugars, and methods of making hydrocarbons or hydrocarbon derivatives by providing a host cell containing a recombinant polynucleotide encoding a polypeptide where the polypeptide transports a pentose into the cell.

OTHER PUBLICATIONS

- Shill et al., "Ionic Liquid Pretreatment of Cellulosic Biomass: Enzymatic Hydrolysis and Ionic Liquid Recycle" *Biotechnology and Bioengineering*, vol. 108, No. 3, Mar. 2011, pp. 511-520.
- Invitation to Pay Additional Fees received for PCT Patent Application No. PCT/US2010/43279, mailed on Apr. 26, 2011, 7 pages.
- International Search Report and Written Opinion received for PCT Patent Application No. PCT/US2010/043279, mailed on Jul. 7, 2011, 16 pages.
- Sadie et al., "Co-expression of a cellobiose phosphorylase and lactose permease enables intracellular cellobiose utilisation by *Saccharomyces cerevisiae*", *Applied Microbiology and Biotechnology*, Feb. 20, 2011, 8 pages.
- Jacobson et al. (2004). "Neurospora in temperate forests of western North America." *Mycologia* 96(1):66-74.
- Jeffries et al. (2007). "Genome sequence of the lignocellulose-bioconverting and xylose-fermenting yeast *Pichia stipitis*." *Nat Biotechnol* 25(3):319-326.
- Kasuga et al. (2008). "Dissecting colony development of *Neurospora crassa* using mRNA profiling and comparative genomics approaches." *Eukaryot Cell* 7(9): 1549-64.
- Kasuga et al. (2005). "Long-oligomer microarray profiling in *Neurospora crassa* reveals the transcriptional program underlying biochemical and physiological events of conidial germination." *Nucleic Acids Res* 33(20): 6469-85.
- Katahira et al. (2008) "Improvement of ethanol productivity during xylose and glucose co-fermentation by xylose-assimilating *S. cerevisiae* via expression of glucose transpoter Sut1." *Enzyme Microb Tech* 43(2):115-119.
- Katahira et al. (2006). "Ethanol fermentation from lignocellulosic hydrolysates by a recombinant xylose- and cellooligosaccharide-assimilating yeast strain." *Appl Microbiol Biotechnol* 72(6):1136-1143.
- Korkhin et al. (1998) "NADP-dependent bacterial alcohol dehydrogenases: crystal structure, cofactor-binding and cofactor specificity of the ADHs of *Clostridium beijerinckii* and *Thermoanaerobacter brockii*." *J Mol Biol* 278(5):967-981.
- Kotaka et al. (2008). "Direct ethanol production from barley beta-glucan by sake yeast displaying *Aspergillus oryzae* beta-glucosidase and endoglucanases." *J Biosci Bioeng* 105(6):622-627.
- Kötter, P. and M. Ciriacy (1993). "Xylose fermentation by *Saccharomyces cerevisiae*." *Appl Microbiol Biotechnol* 38(6):776-783.
- Kubicek et al. (1993). "Triggering of cellulase biosynthesis by cellulose in *Trichoderma reesei*. Involvement of a constitutive, sophorose-inducible, glucose inhibited beta-diglucoside permease." *J Biol Chem* 268(26):19364-19368.
- Kumar et al. (2008). "Bioconversion of lignocellulosic biomass: biochemical and molecular perspectives." *J Ind Microbiol Biotechnol* 35(5):377-391.
- Kuyer et al. (2003). "High-level functional expression of a fungal xylose isomerase: the key to efficient ethanolic fermentation of xylose by *Saccharomyces cerevisiae*?" *FEMS Yeast Res* 4(1):69-78.
- Lang, J. M. and V. P. Cirillo (1987). "Glucose transport in a kinaseless *Saccharomyces cerevisiae* mutant." *J Bacteriol* 169(7):2932-2937.
- Leandro et al. (2006). "Two glucose/xylose transporter genes from the yeast *Candida intermedia*: first molecular characterization of a yeast xylose-H⁺ symporter." *Biochem J* 395(3):543-549.
- Linder, M. and T. T. Teeri (1996). "The cellulose-binding domain of the major cellobiohydrolase of *Trichoderma reesei* exhibits true reversibility and high exchang rate on crystalline cellulose." *Proc Natl Acad Sci USA* 93(22):12251-12255.
- Lynd et al. (2008). "How biotech can transform biofuels." *Nat Biotechnol.* 26(2):169-172.
- Lynd et al. (2002). "Microbial cellulose utilization: fundamental and biotechnology." *Microbiol Mol Biol Rev* 66(3):506-77, table of contents.
- Madhavan et al. (2009). "Xylose isomerase from polycentric fungus orpinomyces: gene sequencing, cloning, and expression in *Saccharomyces cerevisiae* for bioconversion of xylose to ethanol." *Appl Microbiol Biotechnol* 82(6):1067-1078.
- Martin et al. (2010). "Périgord black truffle genome uncovers evolutionary origins and mechanisms and symbiosis." *Nature* 464(7291):1033-1038.
- Martinez et al. (2008). "Genome sequencing and analysis of the biomass-degrading fungus *Trichoderma reesei* (syn. *Hypocreopsis jecorina*)."*Nat Biotechnol* 36(5):553-60.
- Martinez et al. (2009). "Genome, transcriptome, and secretome analysis of wood decay fungus *Postia placenta* supports unique mechanisms of lignocellulose conversion." *Proc Natl Acad Sci USA* 106(6):1954-1959.
- Martinez et al. (2004). "Genome sequence of the lignocellulose degrading fungus *Phanerochaete chrysosporium* strain RP78." *Nat Biotechnol* 22(6):695-700.
- Matsushika et al. (2008) "Expression of protein engineered NADP+-dependent xylitol dehydrogenase increases ethanol production from xylose in recombinant *Saccharomyces cerevisiae*." *Appl Microbiol Biotechnol* 81(2):243-255.
- Matteucci, M. D. and M. H. Caruthers (1980). "The synthesis of oligodeoxypyrimidines on a polymer support," *Tetrahedron Lett* 21:719-722.
- McCluskey, K. (2003). "The Fungal Genetics Stock Center: from molds to molecules." *Adv Appl Microbiol* 52:245-262.
- Medve et al. (1998). "Hydrolysis of microcrystalline cellulose by cellobiohydrolase I and endoglucanase II from *Trichoderma reesei*: adsorption, sugar production pattern, and synergism of the enzymes." *Biotechnol Bioeng* 59(5):621-34.
- Miyasaka, H. (1999). "The positive relationship between codon usage bias and translation initiation AUG context in *Saccharomyces cerevisiae*." *Yeast* 15(8):633-637.
- Nair, N. and H. Zhao (2007). "Biochemical characterization of an L-Xylulose reductase from *Neurospora crassa*." *Appl Environ Microbiol* 73(6):2001-2004.
- Nakamura et al. (2008). "Effective xylose/cellobiose co-fermentation and ethanol production by xylose-assimilation *S. cerevisiae* via expression of β-glucosidase on its cell surface." *Enzyme Microb Tech* 43(3):233-236.
- Noguchi et al. (2009) "Genes regulated by AoXlnR, the xylanolytic and cellulolytic transcriptional regulator, in *Aspergillus oryzae*." *Appl Microbiol Biotechnol* 85(1):141-154.
- Notredame et al. (2000). "T-Coffee: A novel method for fast and accurate multiple sequence alignment." *J Mol Biol* 302(1):205-217.
- Pandit, A. and R. Maheshwari (1996). "Life-history of *Neurospora intermedia* in a sugar cane field." *J biosci (Bangalore)* 21(1):57-79.
- Pauly et al. (2003). "X-ray crystallographic and kinetic studies of human sorbitol dehydrogenase." *Structure* 11(9):1071-1085.
- Pedelacq et al. (2006). "Engineering and characterization of a superfolder green fluorescent protein." *Nat Biotechnol* 24(1):79-88.
- Perkins et al. (1976). "Strains of *Neurospora* collected from nature." *Evolution* 30: 281-313.
- Ramos et al. (1988). "Relationship between low- and high-affinity glucose transport systems of *Saccharomyces cerevisiae*." *J Bacteriol* 170(11):5375-5377.
- Rawat, U. and M. Rao (1997). "Site and significance of cysteine residues in xylose reductase from *Neurospora crassa* as deduced by fluorescence studies." *Biochemical and Biophysical Research Communications* 239(3): 789-93.
- Reifenberger et al. (1997). "Kinetic characterization of individual hexose transporters of *Saccharomyces cerevisiae* and their relation to the triggering mechanisms of glucose repression." *Eur J Biochem* 245(2):324-333.
- Romero et al. (1999). "Cellulase production by *Neurospora crassa* on wheat straw." *Enzyme Microb Tech* 25: 244-250.
- Roy et al. (2010). "I-TASSER: a unified platform for automated protein structure and function prediction." *Nat Protoc* 5(4):725-738.
- Rubin E. M. (2008). "Genomics of cellulosic biofuels." *Nature* 454(7206):841-845.
- Ruepp et al. (2004). "The FunCat, a functional annotation scheme for systematic classification of proteins from whole genomes." *Nucleic Acids Res* 32(18):5539-5545.
- Runquist et al. (2009). "Expression of the Gxf1 transporter from *Candida intermedia* improves fermentation performance in recombinant xylose-utilizing *Saccharomyces cerevisiae*." *Appl Microbiol Biotechnol* 82(1):123-130.
- Runquist et al. (2010). "Comparison of heterologous xylose transporter in recombinant *Saccharomyces cerevisiae*." *Biotechnol Biofuels* 3:5.

- Saha, B. C. (2003). "Hemicellulose bioconversion." *J Ind Microbiol Biotechnol* 30:279-291.
- Saloheimo et al. (2007). "Xylose transport studies with xylose-utilizing *Saccharomyces cerevisiae* strains expressing heterologous and homologous permeases." *Appl Microbiol Biotechnol* 74(5):1041-1052.
- Sarkar, G. and S. S. Sommer (1990). "The "megaprimer" method of site-directed mutagenesis." *Biotechniques* 8(4):404-407.
- Sauer, U. (2001). "Evolutionary engineering of industrially important microbial phenotypes." *Adv Biochem Eng Biotechnol* 73:129-169.
- Scarborough, G. A. (1973). "Transport in Neurospora." *Int Rev Cytol* 34:103-122.
- Seibold et al. (1997). "Role of four major cellulases in triggering of cellulase gene expression by cellulose in *Trichoderma reesei*." *J Bacteriol* 179(17):5318-5320.
- Seidl et al. (2008). "The *Hypocrea jecorina* (*Trichoderma reesei*) hypercellulolytic mutant RUT C30 lacks a 85 kb (29 gene-encoding) region of the wild-type genome." *BMC Genomics* 9:327.
- Shao et al. (2009). "DNA assembler, an in vivo genetic method for rapid construction of biochemical pathways." *Nucleic Acids Res* 37(2):e16.
- Sikorski, R. S. and P. Hieter (1989). "A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in *Saccharomyces cerevisiae*." *Genetics* 122(1):19-27.
- Skory et al. (1996). "Expression and secretion of the *Candida wickerhamii* extracellular beta-glucosidase gene, bgIB, in *Saccharomyces cerevisiae*." *Curr Genet* 30(5):417-422.
- Smith et al. (2000). "Vegetative incompatibility in the het-6 region of *Neurospora crassa* is mediated by two linked genes." *Genetics* 155(3):1095-1104.
- Spindler et al. (May 1992). "Evaluation of the cellobiose-fermenting yeast *Brettanomyces custersii* in the simultaneous saccharification and fermentation of cellulose." *Biotech Lett* 15(5):403-407.
- Spindler et al. (1989). "Simultaneous saccharification and fermentation of pretreated wheat straw to ethanol with selected yeast strains of β -glucosidase supplementation." *App Biochem Biotech* 20:529.
- Stambuk et al. (2003). "D-Xylose transport by *Candida succiphila* and *Kluyveromyces marxianus*." *Appl Biochem Biotechnol* 105-108:255-263.
- Stephanopoulos, G. (2007). "Challenges in Engineering Microbes for Biofuels Production." *Science* 315(5813):801-804.
- Stricker et al. (2008). "Regulation of transcription of cellulases- and hemicellulases-encoding genes in *Aspergillus niger* and *Hypocrea jecorina* (*Trichoderma reesei*)." *Appl Microbiol Biotechnol* 78(2):211-220.
- Sullivan, R. and H. Zhao (2007). "Cloning, characterization, and mutational analysis of a highly active and stable L-arabinitol 4-dehydrogenase from *Neurospora crassa*." *Appl Microbiol Biotechnol* 77(4):845-52.
- Sun Y. and J. Cheng (2002). "Hydrolysis of lignocellulosic materials for ethanol production: a review." *Bioresource Technol* 83(1):1-11.
- Suominen et al. (1993). "High frequency one-step gene replacement in *Trichoderma reesei*. II. Effects of deletions of individual cellulase genes." *Mol Gen Genet* 241(5-6):523-530.
- Tian et al. (2009). "Systems analysis of plant cell wall degradation by the model filamentous fungus *Neurospora crassa*." *Proc Natl Acad Sci USA* 106(52):22157-22162.
- Tian et al. (2007). "Transcriptional profiling of cross pathway control in *Neurospora crassa* and comparative analysis of the Gcn4 and CPC1 regulons." *Eukaryot Cell* 6(6): 1018-1029.
- Townsend J. P. (2004). "Resolution of large and small differences in gene expression using models for the Bayesian analysis of gene expression levels and spotted DNA microarrays." *BMC Bioinformatics* 5:54.
- Townsend J. P. and D. L. Hartl (2002). "Bayesian analysis of gene expression levels: statistical quantification of relative mRNA level across multiple strains or treatments." *Genome Biol* 3(12):research0071.1-0071.16.
- USDA, "The Economic Feasibility of Ethanol Production from Sugar in the United States," United States Department of Agriculture (2006), 78 pages.
- Van Rooyen et al. (2005). "Construction of cellobiose-growing and fermenting *Saccharomyces cerevisiae* strains." *J Biotechnol* 120(3):284-295.
- Vogel, H. J. (1956). "A convenient growth medium fro *Neurospora*." *Microbiol. Genet. Bull.* 13:42-46.
- Wagner et al. (2006). "Rationalizing membrane protein overexpression." *Trends Biotechnol* 24(8):364-371.
- Watanabe et al. (2007). "Ethanol production from xylose by recombinant *Saccharomyces cerevisiae* expressing protein-engineered NADH-preferring xylose reductase from *Pichia stipitis*." *Microbiology* 153(Pt 9):3044-3054.
- Watanabe et al. (2005). "L-Arabinose 1-dehydrogenase: a novel enzyme involving in bacterial L-arabinose metabolism." *Nucleic Acids Symp Ser (Oxf)* 48:309-310.
- Watanabe et al. (2005). "Complete reversal of coenzyme specificity of xylitol dehydrogenases and increase of thermostability by the introduction of structural zinc." *J Biol Chem* 280(11):10340-10349.
- Wiedemann, B. and E. Boles (2008). "Codon-optimized bacterial genes improve L-arabinose fermentation in recombinant *Saccharomyces cerevisiae*." *Appl Environ Microbiol* 74(7):2043-2050.
- Wisselink et al. (2007). "Engineering of *Saccharomyces cerevisiae* for efficient anaerobic alcoholic fermentation of L-arabinose." *Appl Environ Microbiol* 73(15):4881-4891.
- Wisselink et al. (2009). "Novel Evolutionary Engineering Approach for Accelerated Utilization of Glucose, Xylose, and Arabinose Mixtures by Engineered *Saccharomyces cerevisiae* Strains." *Appl Environ Microbiol* 75(4):907-914.
- Woodyer et al. (2005). "Heterologous expression, purification, and characterization of a highly active xylose reductase from *Neurospora crassa*." *Appl Environ Microbiol* 71(3): 1642-7
- Wymelenberg et al. (2009). "Transcriptome and Secretome Analyses of Phanerochaete Chrysosporium Reveal Complex Patterns of Gene Expression." *Appl Environ Microbiol* 75(12):4058-4068.
- Xin et al. (1993). "Acceleration of ethanol production from paper mill waste fiber by supplementation with β -glucosidase." *Enzyme Microb Tech* 15(1):62.
- International Preliminary Report on Patentability received for PCT Patent Application No. PCT/US2010/043279, mailed on Feb. 2, 2012, 10 pages.
- Arendt et al. (2007). "Genetic selection for a highly functional cysteine-less membrane protein using site saturation mutagenesis." *Anal Biochem* 365(2):185-193.
- Bai et al. (2008). "Ethanol fermentation technologies from sugar and starch feedstocks." *Biotechnol Adv* 26(1): 89-105.
- Bailey et al. (1992). "Interlaboratory testing of methods for assay of xylanase activity." *J Biotechnol* 23:257-270.
- Basso et al. (2008). "Yeast selection for fuel ethanol production in Brazil." *FEMS Yeast Res* 8(7):1155-1163.
- Becker, D. M. and V. Lundblad (2001). "Introduction of DNA into yeast cells." *Curr Protoc Mol Biol Chapter* 13, Unit13.7.
- Bendtsen et al. (2004). "Improved prediction of signal peptides: SignalP 3.0." *J Mol Biol* 340(4):783-95.
- Bengsston et al. (2009). "Xylose reductase from *Pichia stipitis* with altered coenzyme preference improves ethanolic xylose fermentation by recombinant *Saccharomyces cerevisiae*." *Biotechnol Biofuels* 2:9.
- Bhat, K. M. and R. Maheshwari (1987). "Sporotrichum thermophile Growth, Cellulose Degradation, and Cellulase Activity." *Appl Environ Microbiol* 53(9):2175-2182.
- Bouffard et al. (1994). "Dependence of lactose metabolism upon mutarotase encoded in the gal operon in *Escherichia coli*." *J Mol Biol* 244(3):269-278.
- Bouws et al. (2008). "Fungal secretomes—nature's toolbox for white biotechnology." *Appl Microbiol Biotechnol* 80(3):381-388.
- Brat et al. (2009). "Functional Expression of a Bacterial Xylose Isomerase in *Saccharomyces cerevisiae*." *Appl Environ Microbiol* 75(8):2304-2311.
- Canevacchini, G. (1988). "Cellobiose dehydrogenase from *Sporotrichum thermophile*." *Method Enzymol* 160:443-448.
- Cantarel et al. (2009). "The Carbohydrate-Active EnZymes database (CAZy): an expert resource for glycogenomics." *Nucleic Acids Res* 37(Database issue):D233-238.
- Chauve et al. (2010). "Comparative kinetic analysis of two fungal beta-glucosidases." *Biotechnol Biofuels* 3(1):3.

- Christakopoulos et al. (1994). "Enzymatic synthesis of trisaccharides and alkyl beta-D-glucosides by the transglycosylation reaction of beta-glucosidase from *Fusarium oxysporum*." *Int J Biol Macromol* 16(6):331-334.
- Christianson et al., (1992). "Multifunctional yeast high-copy number shuttle vectors." *Gene* 110(1):119-122.
- Chu, B. C. and H. Lee (2007). "Genetic improvement of *Saccharomyces cerevisiae* for xylose fermentation." *Biotechnol Adv* 25(5):425-441.
- Colot et al. (2006). "A high-throughput gene knockout procedure for *Neurospora* reveals functions for multiple transcription factors." *Proc. Natl Acad Sci USA* 103(27):10352-10357.
- Cullen et al. (1987). "Sequence and centromere proximal location of a transformation enhancing fragment ans1 from *Aspergillus nidulans*." *Nucleic Acids Res* 15(22):9163-9175.
- Davies et al. (2000). "Structure and function of Humicola insolens family 6 cellulases: structure of the endoglucanase, Cel6B, at 1.6 Å resolution." *Biochem J* 348 Pt 1:201-207.
- Davis, R. H. and D. D. Perkins (2002). "Timeline: *Neurospora*: a model of model microbes." *Nat Rev Genet* 3(5):397-403.
- De Groot et al. (2005). "Metabolic control analysis of *Aspergillus niger* L-arabinose catabolism." *Biotechnol Prog* 21(6):1610-1616.
- Dementhon K et al. (2006). "VIB-1 is required for expression of genes necessary for programmed cell death in *Neurospora crassa*." *Eukaryo Cell* 5(12):2161-2173.
- Doran-Peterson et al. (2009). "Simultaneous saccharification and fermentation and partial saccharification and co-fermentation of lignocellulosic biomass for ethanol production." *Methods Mol Biol* 581:263-280.
- Drissen et al. (2009). "Modelling ethanol production from cellulose separate hydrolysis and fermentation versus simultaneous saccharification and fermentation." *Biocatal Biotransfor* 27(1):27-35.
- Dunlap et al. (2007). "Enabling a Community to Dissect an Organism: Overview of the *Neurospora* Functional Genomics Project." *Adv Genet* 57:49-96.
- Du Preez et al. (1986). "The Fermentation of hexose and pentose sugars by *Candida shehatae* and *Pichia Stipitis*." *Appl Microbiol Biotechnol* 3(3):228-233.
- Eberhart et al. (1977). "Cellulase of *Neurospora crassa*." *J Bacteriol* 130(1):181-186.
- Ellis, K. J. and J. F. Morrison (1982). "Buffers of constant ionic strength for studying pH-dependent processes." *Methods Enzymol* 87:405-426.
- Espagne et al. (2008). "The genome sequence of the model ascomycete fungus *Podospora anserina*." *Genome Biol* 9(5):R77.
- Fonseca et al. (2007). "L-Arabinose transport and catabolism in yeast." *FEBS J* 274(14):3589-3600.
- Freer, S. N. (1991). "Fermentation and aerobic metabolism of cellobextrins by yeasts." *Appl Environ Microbiol* 57(3):655-659.
- Freer, S. N. and R. V. Greene (1990). "Transport of glucose and cellobiose by *Candida wickerhamii* and *Clavispora lusitaniae*." *J Biol Chem* 265(22):12864-12868.
- Galagan et al. (2003). "The genome sequence of the filamentous fungus *Neurospora crassa*." *Nature* 422(6934):859-868.
- Gems et al. (1991). "An autonomously replicating plasmid transforms *Aspergillus nidulans* at high frequency." *Gene* 98(1):61-67.
- Guindon, S. and O. Gascuel (2003). "A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood." *Syst Biol* 52(5):696-704.
- Hahn-Hägerdal et al. (2007). "Towards industrial pentose-fermenting yeast strains." *Appl Microbiol Biotechnol* 74(5):937-953.
- Hector et al. (2008). "Expression of a Heterologous xylose transporter in a *Saccharomyces cerevisiae* strain engineered to utilize xylose improves aerobic xylose consumption." *Appl Microbiol Biotechnol* 80(4):675-684.
- Himmel et al. (2007). "Biomass recalcitrance: engineering plants and enzymes for biofuels production." *Science* 315(5813):804-807.
- Hosaka et al., (1992). "A dominant mutation that alters the regulation of INO1 expression in *Saccharomyces cerevisiae*." *J Biochem* 111(3):352-358.
- Law et al., "Ins and Outs of Major Facilitator Superfamily Antiporters", NIH Public Access Author Manuscript, Annu Rev Microbiol. 2008; 62: 289-305.
- Pace et al., "A Helix Propensity Scale Based on Experiment Studie of Peptides and Proteins", Biophysical Journal, vol. 75, Jul. 1998, pp. 422-427.
- Stryer, Lubert, Biochemistry Fourth Edition, 1995, pp. 28-30.
- Vilkund et al., "Best α -helical transmembrane protein topology predictions are achieved using hidden Markov models and evolutionary information", Protein Science (2004), vol. 13, pp. 1908-1917, Cold Spring Harbor Laboratory Press.

Figure 1

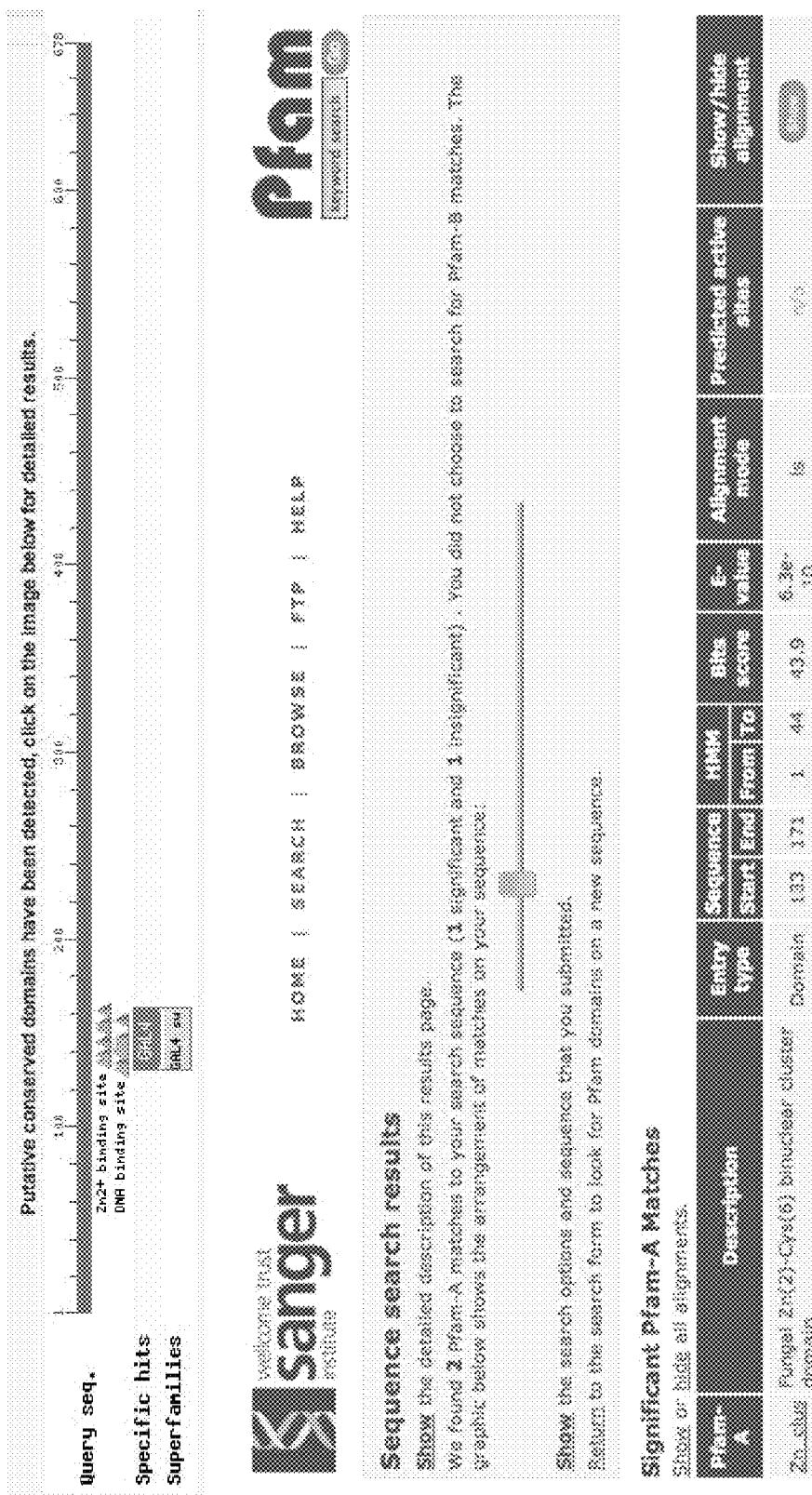


Figure 2

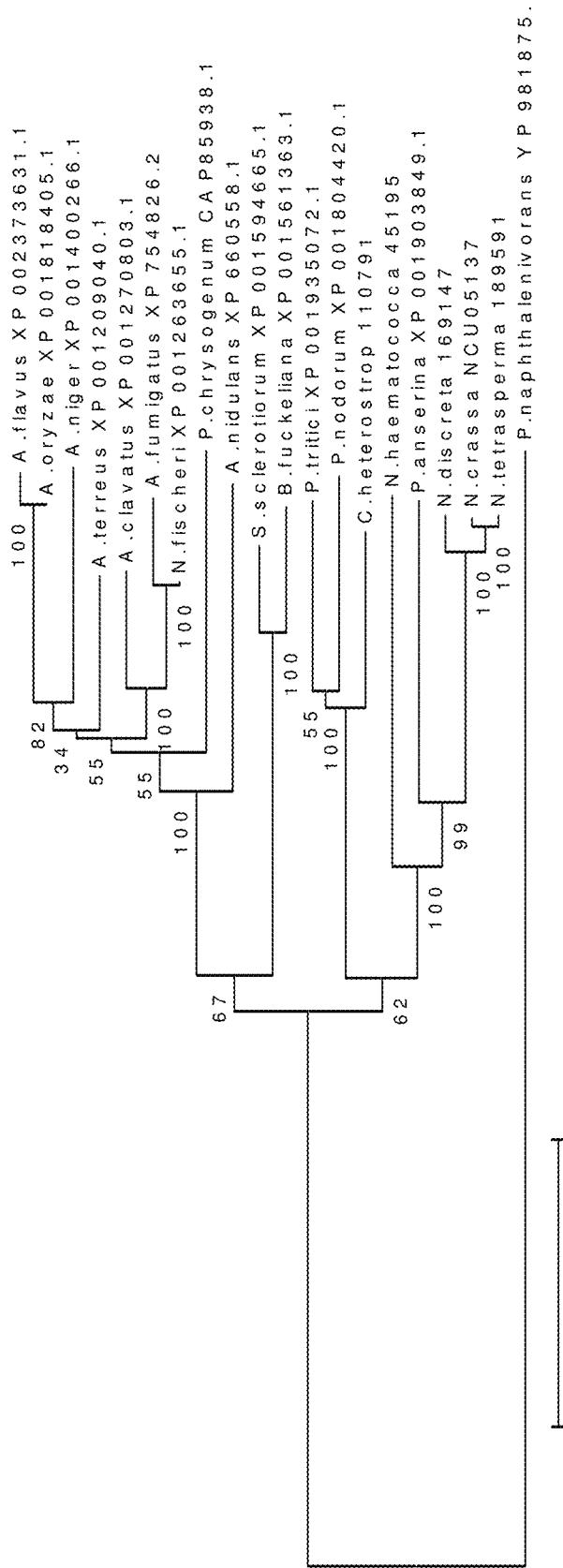


Figure 3

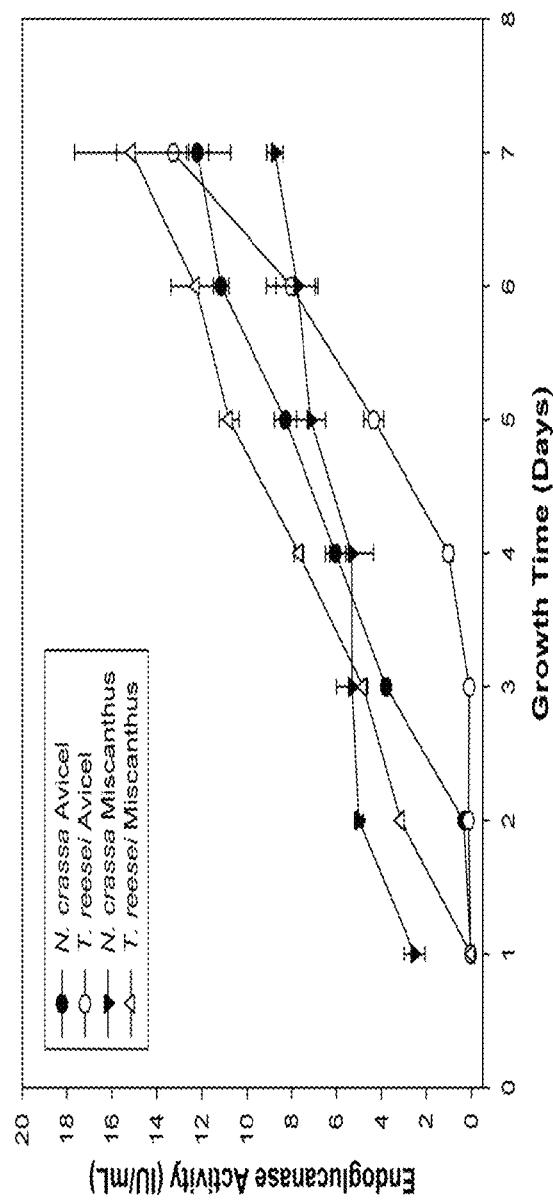


Figure 4

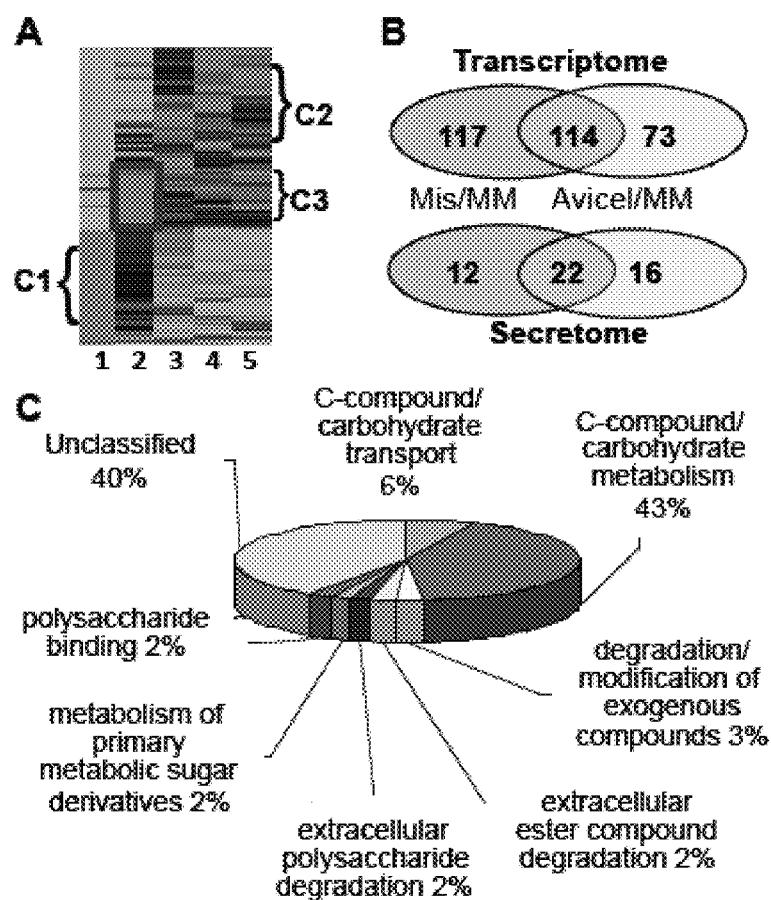


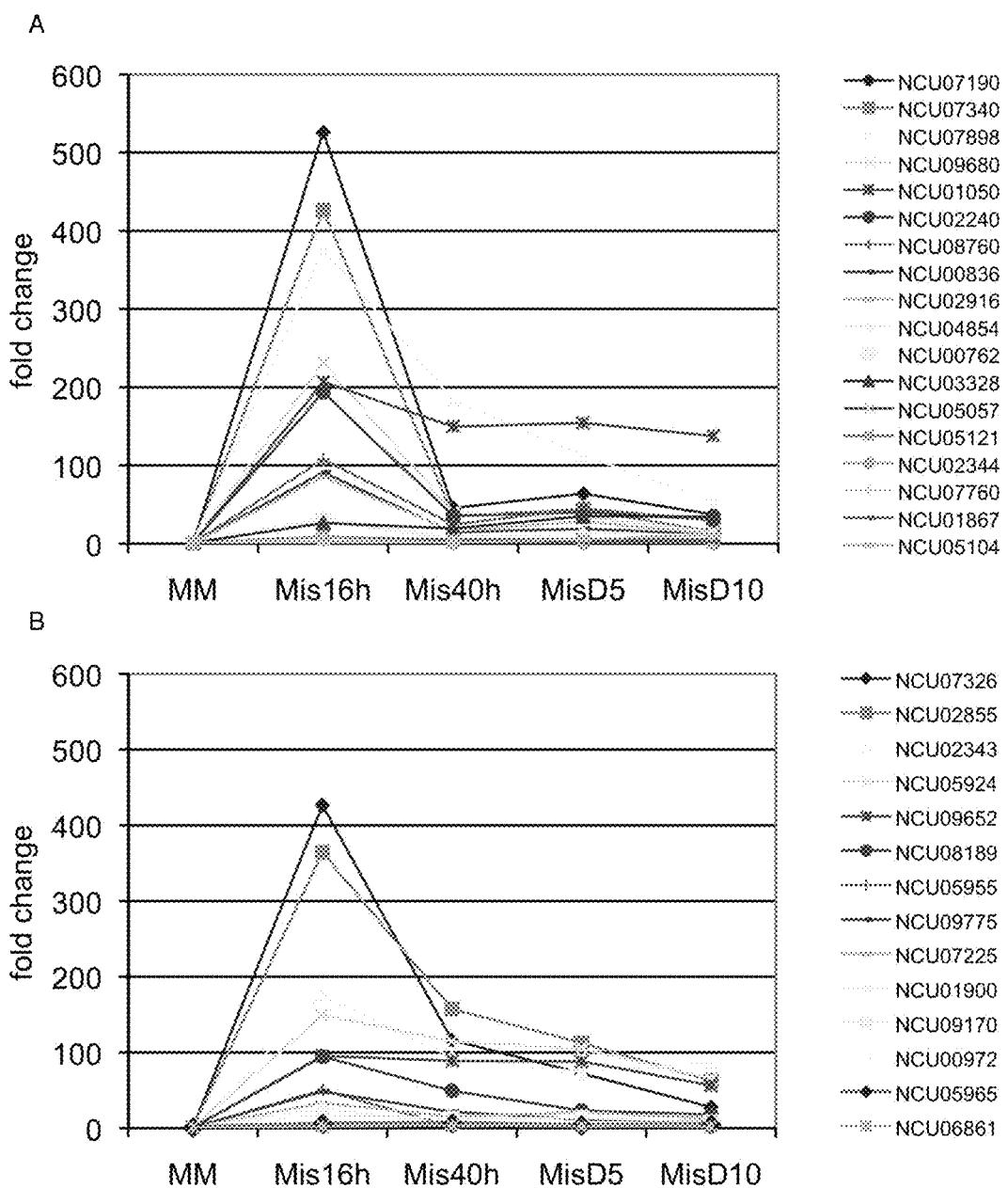
Figure 5

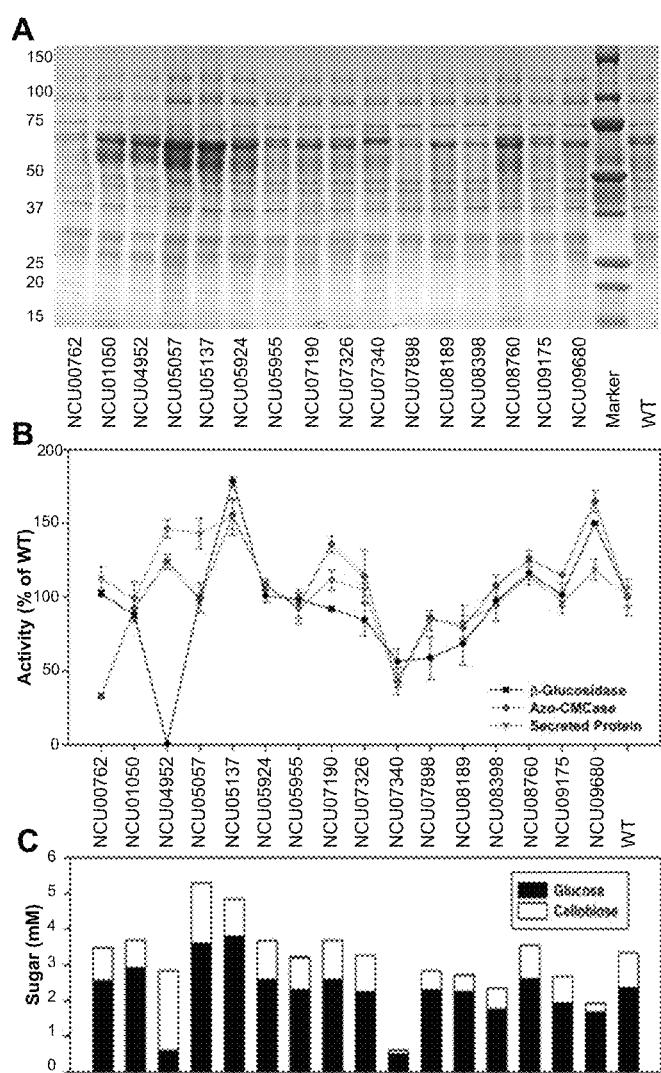
Figure 6

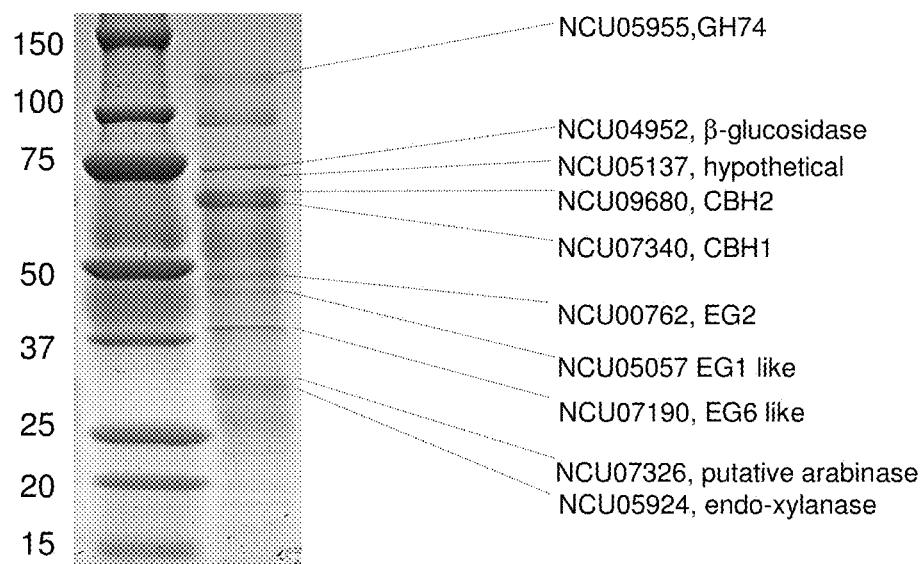
Figure 7

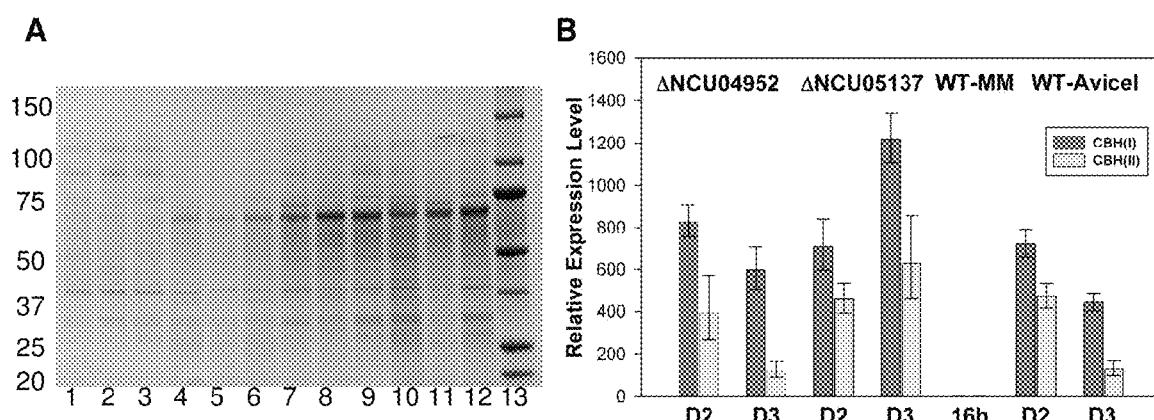
Figure 8

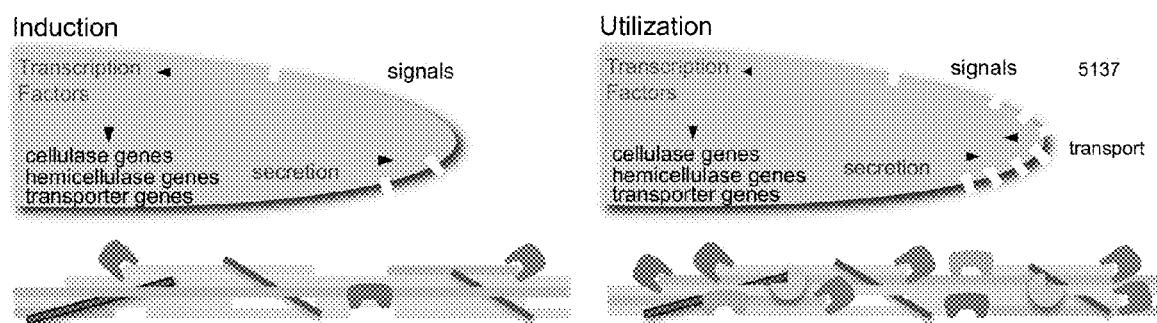
Figure 9

Figure 10-1

Query= NCU00801 | neurospora_crassa hypothetical protein similar to MFS lactose permease (580 nt)
 (579 letters)

Database: Spoth1_GeneModels_FilteredModeis1_aa.fasta
 8806 sequences; 4,205,009 total letters

Searching.....done

Sequences producing significant alignments:	Score (bits)	E Value
jgi Spoth1 43941 e_gw1.2.4209.1	361	e-100
>jgi Spoth1 43941 e_gw1.2.4209.1 Length = 537		
Score = 361 bits (926), Expect = e-100, Method: Compositional matrix adjust. Identities = 199/534 (37%), Positives = 284/534 (53%), Gaps = 29/534 (5%)		
Query: 33 KCHGQTATKPGQKEVRNAALFAAIKE----SNIKPWSKESIHLHYFAIFVAFCCACANG 88 + +T A + +R A L + + PWS I LY + +++ C C NG Sbjct: 5 RDEKETVVADHADDDALREADLAVQVAHDADGTVYSPWSLRMIRLYLVLSLSYLCGCLNG 64		
Query: 89 YDGSLMTGIIAMDKFQNQFHTGDTGPKVSVIFS LYTVGAMVGAPFAAILSDRFGRKKGMF 148 YDGSLM G+ M +Q FH G +IF++Y +G++ F ++D FGR+ GMF Sbjct: 65 YDGSLMGLNCMTSYQRYFHMSTACSTTCLIFAMYNIGSVAAVFFTGPVNDFGRRWGMF 124		
Query: 149 IGGIFIIVGSIIIVASSSKLAQFVVGRFLGLGIAIMTVAAPAYSIEIAPPHWRGRCTGFY 208 +G + +IVG+ + A + QF+ GRFLG G++ V+AP Y E+A P WRG TG Y Sbjct: 125 VGALLVIVGTCVQAPCTTRGQFLAGRFVLGFGVSFCVSAAPCYVSEMAHPKWRGTLTGLY 184		
Query: 209 NCGWFCCSIPAACITYCCYFIKSNSWRIPPLILQAFTCLIVMSSVFFLPESPRFLFANR 268 NC W+ GSI A+ + YGC +I + +WRIP+ Q T +V VF+LPESPR+L A R Sbjct: 185 NCTWYIGSIVASWVYGCYIDTLAWRIPIWCQMVTSGVCLGVFWLPESPRWLMAQDR 244		
Query: 269 DAEAVAFLVKYHGNQDPNSKLVLLETEEMRDGIRTDGVKWWWDYRPLFMTHSGRWRMAQ 328 +A L YHG G + LV L+ +EM + I T+ DK W+DY L+ THS R R+ Sbjct: 245 HDDAARVLA TYHGEGRADHPLVKLQMQEMMNQISTEASDKWYD YHELNNTHSARRRLIC 304		
Query: 329 VLMISIFGQFSGNGL-GYFNNTVIFKNIGNVTSTSQQLAYNILNSVISAIGALTAVSMTDRM 387 V+ +++FGQ SGN L Y+ + K+ G+T + LA N +N +S +GA+ MTD + Sbjct: 305 VIGMAVFGQISGNSLSSYYLVNMLKSAGITEERRVLA LNGVNPALSFLGAILGARMTDVV 364		
Query: 388 PRRAVLIIGTFMCAAALATNSGLSATLDKQTQRGTQINLNQGMNEQDAKDNDAYLHVDSNY 447 RR +L+ + A +G S R N Sbjct: 365 GRRPPLLTYTIVFASVCFAVITGTSKMATDDPTRTAAAN----- 402		
Query: 448 AKGALAAYFLFNVIIFSFTYTPLQGVIPTEALETTIRGKGLALSGFIVNAMGFINQFAGPI 507 +A F+F ++FSF +TPLQ + E L T R KG A+ F + I Q+A Sbjct: 403 --ATIAFIFIFGIVFSFGWTPLQSMYIAETLPTATRAKGTAVGNFSSSVASTILOQYASGP 460		
Query: 508 ALHNIGYKYIFVFVGVWDLIELTVAWYFFFVESQGRITLEQLEWVYDQPNPVKASLK 561 A IGY + VFV WDLIE YF+ E++ RTLE+LE V+ PNPVK SL+ Sbjct: 461 AFEGIGYFYLVFVFWDLIELGAIMYFYFPETKDRTEELEEVFSAPNPVKKSL 514		

Figure 10-2

Query= NCU04963 : neurospora_crassa hypothetical protein similar to
MFS monosaccharide transporter (528 nt)
(527 letters)

Database: Spoth1_GeneModels_FilteredModels1_aa.fasta
8806 sequences; 4,205,009 total letters

Searching.....done

Sequences producing significant alignments:	Score (bits)	E Value
jgi Spoth1 62521 estExt_Genewise1.C_21757	541	e-155
>jgi Spoth1 62521 estExt_Genewise1.C_21757 Length = 566		
Score = 541 bits (1395), Expect = e-155, Method: Compositional matrix adjust. Identities = 263/508 (51%), Positives = 366/508 (72%), Gaps = 10/508 (1%)		
Query: 6 KKPEGVPGKSWPAIVIGLEVAFGGVLFGYDTGTIGGILAMPYWQDLFSTGYRNPEHHLDV 65 +KP+ V G S PAI+-GLFVA GG+LFGYDTG I GILAM +++ F+TGY + + + Sbjct: 9 QKPDNVAGSSAPAIMVGLFWATGGLLFGYDTGAINGTLAMDTFKEDFTTGYTDKQGKPGL 68		
Query: 66 TASQSATIVSILSAGIFFGALGAAPLADWAGRRLGLILSSFVFIFGVILQTAAVSIPLFL 125 AS+ + IV++LSAGT GAL +AP+ D GRRL L1++ VF G 1+Q A ++ + + Sbjct: 69 YASEVSLIVAMLSGTATGALLSAPMGDRWGRRLSILIVAIGVFCVGAIIQVCATNVAMLV 128		
Query: 126 AGRFFAGLGVLISATIPLYQSETAPKWIRGVIVGSYQLAITIGLLLASIVNNATHNMQN 185 GR AG+GVG++S +FLYQSE APKWIRG +V +YQL+IT GLL A+ VN T+ + + Sbjct: 129 VGRTLAGIGVGVVSVLVPLYQSEMAPKWIIRGTLVCAVQLSITAGLLAAATVNILTYKLKS 188		
Query: 186 TGCYRIPIAVQFAWAIILIVGMIILPETPRFHIKRDNLPAATRSLAILRRLEQNHPAIE 245 YRIPPI +Q WA++L +G++ILPETPR+ +KR AA SL+ LRRL+ HPA+IE Sbjct: 189 AAAYRIPIGLQLTWAIVLVAIGLIVLIPETPRYLVKRGLKEAAALSLSRIRRIDITHPALIE 248		
Query: 246 ELSEIQANHEFEKSLGKATYLDCLKG--NLLKRLLTGCFIQLQSLQQLTGINFIFYGTQFF 303 EL+EI+ANHE+E +LG TY D + G +L +R LTGC IQ IQLQTG+NFI YYGT FF Sbjct: 249 ELAEIEANHEYEMALGPDTYKDIIIFGEPHLGRRTLTCGCGLQMLQQLTGVNFIIMYYGTTFF 308		
Query: 304 KNSGFSDSFLISLITNLNVVSTLPGLYAIDKWGRPVLLWGAVGMCVCQFIVAILGTTT 363 +G ++F +SLI ++N+VST PGL+ ++ WGRR +L+ G+VGM +CQ ++A T + Sbjct: 309 YGAGIGNAFTVSLIMQVINLVSTFPGLFVVESWGRKLLIVGSVGMAICQLLIASFATAS 368		
Query: 364 TSQDASGMIIVHNLAFAQAAIAFICFYIFFFFAASWGPVAWVVTGEIFPLKVRAKSLSITT 423 + + + + I F+ YIFFFAASWGPV WVVT EI+PLKVRAKS+SI+T Sbjct: 369 GNDNKP-----TQNQ1L11IFVAIYIFFFAASWGPVVVVVTSEIYPLKVRAKSMSIST 420		
Query: 424 ASNWLNNWAIAYSTPYLVNYGPGNNANLQSKIFFVWGCCCCFICIAFVYFMIYETKGLTLEQ 483 ASNW+LN+ IAY TPYLV+ G+ +L S++FFVWG C + IAFV++M+YET ++LEQ Sbjct: 421 ASNWVLNFGIAYGTPYLVTSGDSPDLGSRVFFFVGAFICLISIAFVWYMWYETSKISLEQ 480		
Query: 484 VDELYEEVSDARKSIGWVPIITFREIRE 511 +DE+YE V+ A S + P+ +F+++R+ Sbjct: 481 IDEMYERVAHAWNSRSFEPWSFQQMRD 508		

Figure 10-3

```

Query= jgi|Spoth1|48439|e_gw1.3.3367.1
      (512 letters)

Database: neurospora_crassa_7_proteins_no_asterisks.fas
      9822 sequences; 4,775,003 total letters

Searching.....done

Score      E
Sequences producing significant alignments:          (bits) Value
                                         727    0.0
NCU01132 |  neurospora_crassa hypothetical protein similar to mo... 727    0.0
>NCU01132 |  neurospora_crassa hypothetical protein similar to
              monosaccharide transporter (554 nt)
Length = 553

Score = 727 bits (1876), Expect = 0.0, Method: Compositional matrix adjust.
Identities = 382/548 (69%), Positives = 431/548 (78%), Gaps = 42/548 (7%)

Query: 1 MKKFLGLRGQALNLAVGTLAGCDFLLFGYDQGVMGGILTLKVFLDAFPMINPEAAGLSHD 60
       MK FLGLRGQ LNLAVG +AGCDFLLFGYDQGVMGGILTL FL F INP+A GL+
Sbjct: 1 MKPFLGLRGQPLNLAVGAVAGCDFLLFGYDQGVMGGILTLPEFLGYFEQINPDAPGLTPH 60

Query: 61 ESSTRSTYQGIAVASYNLGCFLGAIITIFIGNPLGRKRVIMLGTTSVMVIGAILQASSTTL 120
       ESS RSTYQGI+VASYNLGCF+GAIITIFIGNP GRK++I+LGTS+M++GAILQAS+TTL
Sbjct: 61 ESSMRSTYQGISVASYNLGCFIGAIITIFIGNPWRKKIILLGTTSIMIVGAILQASATT 120

Query: 121 PQFIVGRIITGLNGGNSTVPTWQSETSKAHKRKGKMIFF-----CALLIAF----- 167
       FI+GRIITG+GNGGNSTVPTWQSETSKAHKRKGK+           I+L++
Sbjct: 121 GHFIIGRIITGIGNGGNSTVPTWQSETSKAHKRKGKVMIEGSLVTAGIMSYWIDLGLS 180

Query: 168 -----IPFLPESPRWLILKGREDEAREVIAALEBDTDSR 202
       IPFLPESPRWL+ KGR+ EA+E++AAL D + D
Sbjct: 181 FAPGSVAWRFPPLAFQIIFCILILIFIPLPESPRWLVFKGRDAEAKEILAALNVEELDP 240

Query: 203 IVENEFIAKETVLEMSKGTFRDLFTMDKNRNLHRTLLAYFNQVFQQISGINLITYYAAV 262
       IV+ EF I +TV+EMSKG+F+DLFTMDK+RN HRTLLAY NQVFQQISGINLITYYAAV
Sbjct: 241 IVDTEFHFIHDVVEMSKGSFKDLFTMDKDRNFHRTLLAYLNQVFQQISGINLITYYAAV 300

Query: 263 IYKGLGMSDFLSRLLAALNGTEYFLASWPATLVERVGRRNMLFGAVGQAATMAILAGV 322
       IY GLGMSDFL+RLLAALNGTEYF+ASWPATLVERVGRR LMLFGA+GQAATMAILAGV
Sbjct: 301 IYSGLGMSDFLARLLAALNGTEYFIASWPATLVERGRRKLMFGAIGQAATMAILAGV 360

Query: 323 NSR--QETGFQIAGIVFLFVNTFFAVGWLGMTWLYPAEIVPLRIRAPANALSTSANWIFN 381
       NSR + I+QIAGIVFLFVFT FAVGWLGM+WLYPAEIVPLRIRAPANALSTSANWIFN
Sbjct: 361 NSRPDDKPYQIAGIVFLFVNTVFAVGWLMSWLYPAEIVPLRIRAPANALSTSANWIFN 420

Query: 382 FLVVMITPVAFNNIGYQTYIIFAVINAFMVPVCVYFFYPETAYRSLEEMDNIFHKVADGWK 441
       F+VVMITPVAFN I YQTYIIFAVINAF+VP VYFFYPETA RSLEEMD IFHKV DGWK
Sbjct: 421 FMVVVMITPVAFNPKIKYQTYIIFAVINAFIVPVVYFFYPETACRSLEEMDMIFHKV-DGWK 479

Query: 442 GVFTVVHQAKVEPRWYKGNGELL--VDYQOTEHHRRHLQQQEGAVTASEKRSVEGAGSGS 499
       G FTVVHQAKVEP+WY K+G+ + D+++T ++ H + K VE +
Sbjct: 480 GYFTVVHQAKVEPKWYDKDGQRIGGADFEKTAGYQSHSIPESSGFEKPTKAHVESPRADD 539

```

Figure 10-4

```

Query: 500 GSGDVVKQD 507
      G     D
Sbjct: 540 GITSSSSD 547

Query= jgi|Spoth1|112305|estExt_fgenesh1_kg.C_60263
      (554 letters)

Database: neurospora_crassa_7_proteins_no_asterisks.fas
            9822 sequences; 4,775,003 total letters

Searching.....done

Score      E
Sequences producing significant alignments:          (bits) Value
                                             
NCU10021 |  neurospora_crassa high affinity glucose transporter ...    937    0.0
>NCU10021 |  neurospora_crassa high affinity glucose transporter
              (554 nt)
Length = 553

Score = 937 bits (2421), Expect = 0.0, Method: Compositional matrix adjust.
Identities = 474/542 (87%), Positives = 507/542 (93%), Gaps = 7/542 (1%)

Query: 1 MSSSEKEATGPVAAHVGNLATTQDVEKIEAPVTWKAYLICAFASFGGIFFGYDSGYINGV 60
      MSS+ ++ T P AAH+G LA TQDVE+IEAPVTWKAYLICAFASFGGIFFGYDSGYINGV
Sbjct: 1 MSSAHEKETAPTAAHIG-LAHTQDVERIEAPVTWKAYLICAFASFGGIFFGYDSGYINGV 59

Query: 61 LASKLFINAVEGAG--KDAISESHSSLIVSILSCGTFFGALIAGDLADFIGRKYTIVLGC 118
      L S++FI+AVEG      +DA+SESH SL+VSILSCGTFFGALIAGDLAD IGRK+TVILGC
Sbjct: 60 LGSQLFIDAVEGTSPVRDALSEHQSLLVVSILSCGTFFGALIAGDLADMIGRKWTIVLGC 119

Query: 119 LIYIIGCVIQIITGLGNALGAIVAGRRIAGIGVGFESEAIVILYLMSEICPRKVRGALVAGY 178
      LIY+IGCVIQ+ITGLG+ALGAIVAGRRIAGIGVGFESEA+VILYLMSEICPRKVRGALVAGY
Sbjct: 120 LIYLIGCVIQMITGLGDALGAIVAGRRIAGIGVGFESEAIVILYLMSEICPRKVRGALVAGY 179

Query: 179 QFCITIGLMLASCVVYGTQNRQDTGQYRIPIGIPIQFIWALILGGGLCLPDSPRYFVKRGR 238
      QFCITIGLMLASCVVY TQ+R DTG YRIPPI IQFIWALIL GGL+CLPDSPRYFVK+C
Sbjct: 180 QFCITIGLMLASCVVYATQDRPDTGAYRIPIAIQFIWALILLAGGIMCLPDSPRYFVKGN 239

Query: 239 LADATSALSRLRGQPEDSEYIQVELAEIVANEEYERQLIPSTTWFGSWANCFKGSLFKAN 298
      LA ATS+LSRLRGQ  +SEYIQVELAEIVANEEYERQLIPSTTWFGSWANCFKGSL+KAN
Sbjct: 240 LAAATSSLRLRGQDPNSEYIQVELAEIVANEEYERQLIPSTTWFGSWANCFKGSLWGAN 299

Query: 299 SNLRKTILGTSLOQQMOWTGVNFIIFYYSTPFLKSTGAIDPFELMSMVFTIIINVFSTPISF 358
      SNLRKTILGTSLOQQMOWTGVNFIIFYYSTPFLKSTGAI + FL+SMVFTIIINVFSTPISF
Sbjct: 300 SNLRKTILGTSLOQQMOWTGVNFIIFYYSTPFLKSTGAISNTPLISMVFTIIINVFSTPISF 359

Query: 359 YTVERFGRRTILFWGALGMLICQFLVAIVGVTVGFNHTHPAPTADDPEATLANNISAVNA 418
      +TVERFGRRTILFWGALGMLICQFLVAI+GVTVGFN TH P + ++ANN+SAVNA
Sbjct: 360 WTVERFGRRTILFWGALGMLICQFLVAIIGVTVGFNTHMGPDGE----SMANNVSANNA 415

Query: 419 QIAFIAIFIFFFASTWGPAGAWIVIGEIFPLPIRSRGVGLSTASNWLWNNTIIAVITPYMVG 478
      QIAFIAIFIFFFASTWGPAGAWI+IGEIFPLPIRSRGVGLSTASNWLWNNTIIAVITPYMVG
Sbjct: 416 QIAFIAIFIFFFASTWGPAGAWILIGEIFPLPIRSRGVGLSTASNWLWNNTIIAVITPYMVG 475

```

Figure 10-5

```

Query: 479 EDRGNMKSSVFFVWGLCTCAFVYTYFLVPETKGLSLEQVDKMMEEETTPRTSAWKPTT 538
        E RGN+KSSVFFVWGLCTCAF+YTYFLVPETKGLSLEQVDKMMEEETTPRTSAWKWP TT
Sbjct: 476 EQRGNLKSSVFFVWGLCTCAFVYTYFLVPETKGLSLEQVDKMMEEETTPRTSAWKPRTT 535

Query: 539 FA 540
        FA
Sbjct: 536 FA 537

Query= jgi|Spothi|114107|estExt_fgeneshl_pm.C_20669
       (522 letters)

Database: neurospora_crassa_7_proteins_no_asterisks.fas
           9822 sequences; 4,775,003 total letters

Searching.....done

Score      E
Sequences producing significant alignments:          (bits) Value
NCU08114 |  neurospora_crassa hypothetical protein similar to MF...     832   0.0
>NCU08114 |  neurospora_crassa hypothetical protein similar to MFS
              hexose transporter (526 nt)
              Length = 525

Score = 832 bits (2148), Expect = 0.0, Method: Compositional matrix adjust,
Identities = 397/523 (75%), Positives = 452/523 (86%), Gaps = 6/523 (1%)

Query: 1  MGIFAFNKQKPNAEATAVAQ--EEAPQFERVDWKRDPLRKLYFYAFVLCIASATTGYDG 58
        MGIF      +KP A+A + Q  EEAPQFERVDWK+DPGLRKLYFYAF+LCIASATTGYDG
Sbjct: 1  MGIF-----NKKPVAQAVDLNQIQQEEAPQFERVDWKDFGLRKLYFYAFILCIA SATTGYDG 57

Query: 59  MFFNSVQNFETWENYFNHPTGSKLGVLGALYQIGSLASIPLVPIIADRVGRKIPIAIGCV 118
          MFFNSVQNFETW YF P GS+LG+LGALYQIGS+ SIP VP++ D GRK PI ICCV
Sbjct: 58  MFFNSVQNFETWIKYFGDPRGSELGLGALYQIGSIGSIPFVPLTDNFGRKTPIIIGCV 117

Query: 119 IMIVGAVLQAACRNLGTFMGGRFLLGFGNSLAQLCSPMLLTEL AHPQHRGRLLTVYNCLW 178
        IMIVGAVLQA +NL TFMGGR +LGFGNSLAQ+ SPMLLTEL AHPQHR RLTT+YNCLW
Sbjct: 118 IMIVGAVLQATAKNLDTFMGGRTMLGFGNSLAQI ASPMILTEL AHPQHRARLTTIYNCLW 177

Query: 179 NVGALVVAVSFGTDYLKSWSWRIPALIQAFPSVIQLLFIFWVPESPRYLMAKD KHERA 238
        NVGALVV+W++FGT+Y+ +DWSWRIPAL+QAFPS+IQLL I+WVPESPR+L+AKDKH+ A
Sbjct: 178 NVGALVVWSWIAGTNYINNDWSWRIPALLQAFFSIIQLLGIWVWPESPRFLIAKDKHDEA 237

Query: 239 LAILAKYHANGDANHPTVQFEYREIKETIRLEFEASKSSYLD FVRTRGNRYRLA VLI SL 298
        L IIAKYHANGD NHPTVQFE+REIKET+RLE E++K+SSYLD F ++RGNRYRLA+L+SL
Sbjct: 238 LHIIAKYHANGDPNHPTVQFEFREIKETIRLEMESTKNSSYLDFFKSRGNR YRLA ILLSL 297

Query: 299 GIFSQWSGNAIISNYSSKLYDTAGVTGSTQKLGLSAGQTGLSIIISVTM ALLVDKFGRRP 358
        G FSQWSGNAIISNYSSKLY+TAGVT ST KLGLSAGQTGL+LI+SVTM ALLVDK GRR
Sbjct: 298 GFFSQWSGNAIISNYSSKLYETAGVT DSTAKLGLSAGQTGLALIVS VTM ALLVDKLG RRL 357

Query: 359 MFLTSTAGMFCTFIFWT LTSGLYEEHNADGARYAMILFIWIHGIFY SISWSGLLVGYAIE 418
        FL ST GM TF+ WTLT+GLY EH GA AMI FIW+ GIFY S++WSGLLVGYAIE
Sbjct: 358 AFLASTGGMC GTFVIWTL TAGLYGEHRLKGADKAMIFFI WVGIFY S LAWSGLLVGYAIE 417

```

Figure 10-6

```

Query: 419 VLPYKLRALKLIMMNLTVAQALTLNTYANPVAFDAF-EGHISWKLYIYYTIWIFLELCFW 477
       +LPY+LR KGLM+MN++VQ ALTLNTYANPVAFD F HSWKLY+IYT WI E FV+
Sbjct: 418 ILPYRLRGKGLMVNMNSVQCALTNTYANPVAFDYFGPDHSWKLYLIYTCAAAEFVFV 477

Query: 478 KMYIETKGPTLEELAKIIDGDEAAVAHVVDIKQVEKETHINEEK 520
       MY+ETKGPTLEELAK+IDGDEA VAH+DI QVEKE I+E +
Sbjct: 478 FMYVETKGPTLEELAKVIDGDEADVAHIDIHQVEKEVEIHEHE 520

Query= jgi|Spoth1|70023|estExt_Genewise1.C_53218
      (504 letters)

Database: neurospora_crassa_7_proteins_no_asterisks.fas
            9822 sequences; 4,775,003 total letters

Searching.....done

Score      E
Sequences producing significant alignments:          (bits) Value
NCU06138 | neurospora_crassa hypothetical protein similar to MF...    263   6e-71
>NCU06138 | neurospora_crassa hypothetical protein similar to MFS
           monosaccharide transporter (584 nt)
           Length = 583

Score = 263 bits (673), Expect = 6e-71, Method: Compositional matrix adjust.
Identities = 159/488 (32%), Positives = 259/488 (53%), Gaps = 35/488 (7%)

Query: 10 FLGVVFASLGSLLYGYDLGVIAQVIASQSFKSRFSPSD--NEEAA-----VVSVFTGG 61
       F + FA +G +LYGY+ G+ +V+A +F+ D +E A+ + ++ GA
Sbjct: 39 FSIACFACIGGVLYGYNQGMFSGVLAMPAPQKHMGEYDPIDENASQTKKGWLTAILELGA 98

Query: 62 FFGAMAAGPMGDKLGRWRITLCLGALVFCCLGGALQTGAQALSY--LYAGRHSIAGLVGVLC 119
       + G + +G M + L R++ +L LVF LG +Q + + + + AGR I G+GVG L
Sbjct: 99 WLGTLLSGFMAEVLSRKYGVLVACLVFMLGVVIQATSISGGHETILAGRFITGMGVGSLA 158

Query: 120 MIVPMYQAEIAPSHRGRITALQQFMLGTGALAAAISYGT-YVG---FAPTNDQQWRTS 175
       MI+P+Y +E+A P +RG + ALQQ + G + + WI YGT Y+G +D W
Sbjct: 159 MTIPYINSEVAPPEVRGALVALQQLAICFGIMVSFWIDYGTNYIGGTKLETQSDAAWLVP 218

Query: 176 LGIQVIPAVFLAALILLFFESPRWLIDHGRSEEGLRTLQLHSHGDVDDAWVQAELYQQIR 235
       + +Q+ PA+ L + + P SPRWLI HGR E + L+ L D V+ E+ +I+
Sbjct: 219 VCLQLAPALILFFGMMFMPPFSPRWLHGGREAEARKILSTLRGLPQ-DHELVELEFLEIK 277

Query: 236 ESVEFVRENEAKSYAEL-----FRDRSCFRRFLACAIQGSVQMIG 276
       F + + A+ + EL FR ++ FRR+ +A Q +C
Sbjct: 278 AQSLFEKRSTAELFPELREQTAWNTFKLQFVAIEKLFRKAMFRRVIVATVTMFFQQWSG 337

Query: 277 VSIAIQYYSVTIYGLMGIEGDDT-LKYQAIISSIIALVAQALCILIDRLGRRWTLLGGNLG 335
       ++AI YY+ I+ +G+ G+ T L + I+ +A +L IDR+GR+ L G LG
Sbjct: 338 INAILYYAPQIFKQLGLSGNTTSLLATGVVGIVMFIAVPAVLWIDRVGRKPVLTIGALG 397

Query: 336 NCVTFIIATVMLARYPPGTSSNKAAAAGFIVVVTWVYNSFSATCGPLSWIIPAEIFDTKT 395
       II V++A+ + +KAA W + + W++ F + GP +WII AEI+ T
Sbjct: 398 MATCHIIIHAVIVAKNVDQWETHKAAGWAAVAMVWLFIHFGYSGPCAWIIVAEIWPLST 457

```

Figure 10-7

```

Query: 396 RSKGVSIATMTSPALNTMIGQVTGPAMKTVGYRFYLLFVVCNFTNALFFWAFLPETAKRP 455
        R GVS+    +++ N ++GQVT    +K + Y Y++F +   + A F W F+PET +
Sbjct: 458 RPYGVSLGASSNWMNNFIVGQVTPDMLKAIPYGTYYIFGLLTYMGAAFIWFFVPETKRLT 517

Query: 456 LEEMNRLF 463
        LEEM+ +F
Sbjct: 518 LEEMDMIF 525

Query= jgi|Spoth1|102977|fgenesh1_pm.5_#_763
          (481 letters)

Database: neurospora_crassa_7_proteins_no_asterisks.fas
          9822 sequences; 4,775,003 total letters

Searching.....done

Score      E
Sequences producing significant alignments:          (bits) Value
                                             
NCU05897 |  neurospora_crassa hypothetical protein similar to 1-...    713    0.0
>NCU05897 |  neurospora_crassa hypothetical protein similar to
              1-fucose permease (472 nt)
Length = 471

Score = 713 bits (1841), Expect = 0.0, Method: Compositional matrix adjust.
Identities = 347/447 (77%), Positives = 385/447 (86%), Gaps = 2/447 (0%)

Query: 1  MLSSGFWKRRSLRVPDNQRTKAAELTLRESLYPLSLVTILFFLWGFSYGLLDTLNKHFQN 60
        M S +WKRRSLRV D++ TKAAEL+LRESL PL LVTILFFLWGFSYGLLDTLNKHFQ
Sbjct: 1  MFSREWWKRRSLRVRDDKVTKAAELSIRESLPLCLVTILFFLWGFSYGLLDTLNKHFQE 60

Query: 61  TLGITKTRSSGLQAAAYFGAYPLASLGHAAWILRHGYGRAVFIWGLFLYGLGALLAIPSIM 120
        TL ITK RS+GLQAAAYFGAYPLASLGHA+ILR + YRAVFIWGLFLYGLGALLAIP I
Sbjct: 61  TLHITKARSAGLQAAAYFGAYPLASLGHAAYILRRFSYRAVFIWGLFLYGLGALLAIPCIR 120

Query: 121 HHSFAGFCVCICIFIIGNGLGSLETAANPYIITVCGPPKFSEIRINVAQAFNGIGTVVAPVLG 180
        SFAGFCVCICIFIIGNGLGSLETAANPYIITVCGPPK+SEIRIN AQAFNGIGTVVAPVLG
Sbjct: 121 AKSFAGFCVCICIFIIGNGLGSLETAANPYIITVCGPPKYSEIRINFAQAFNGIGTVVAPVLG 180

Query: 181 SYVFFTDDQTALRNVQWVYLAIAFCVFLLAGVFFLSVIPEITDADMFAQAAETHAGADD 240
        SYVFF FDD AL+NQWVYLAIA FV++LA VFFF +PEITDADM QAAETHAG D
Sbjct: 181 SYVFFGFDDNLALQNQWVYLAIAFVYIILAVVFFLIELPEITDADMHQHQAETHAGDAD 240

Query: 241 RPFHTQYRLFHAAFAQFCYTGAQVAIAGYFINYATETRPNTDSSLGSKFLAGSQAGFAVG 300
        +PF QYRLFHA+F+QFCYTGAQ+AIAGYFINY TETR NTDS+LG++FLAG+Q FAVG
Sbjct: 241 QPFRKQYRLFHASFSQFCYTGAQIAIAGYFINYVIETRKNTDSALGAQFLAGAQGTFAVG 300

Query: 301 RFAGAAMMQFIKPRKVFALFMTMCIVFSAPAITQRGNAGLSMLYLVMFFESICFPPIAL 360
        RF GAA+M F++PRKVF LF+T CI+F AP ITQR N G+S+LY+ +FFESICFPPI+AL
Sbjct: 301 RFAGAAIMHFVRPRKVFLFLTACIIFVAPТИTQRЕНТGMSLLYVTЛFFESICFPPIAL 360

Query: 361 GMRGLGRHTKRGSGWIVAGVLGGACVPPLMGAAADARGTGFMSMLVPLCFFVAAWTYALAV 420
        GMRGLGRHTKRGSG++VAGV GGA VPPLMGA AD T SM+VPL FF AWTYALAV
Sbjct: 361 GMRGLGRHTKRGSGFLVAGVFGGAVVPPFLMGAVADMHDТАMSMVVPLAFFAVAATYALAV 420

```

Figure 10-8

Figure 10-9

```

Query: 497 TSGRTLEELAFLFEDASLNEKAAA AVEKQIHYGDEK 532
      T GRTLEEL F+FE   L +A AA EK +++ +
Sbjct: 481 TMGRTLEELTFMFEGEDLQRQANAAA EKVVNHTED 516

Query= jgi|Spoth1|116270|estExt_fgenesh1_pm.C_50266
(488 letters)

Database: neurospora_crassa_7_proteins_no_asterisks.fas
9822 sequences; 4,775,003 total letters

Searching.....done

Score      E
Sequences producing significant alignments:          (bits) Value
NCU05519 |  neurospora_crassa hypothetical protein similar to Tn...    700    0.0

>NCUC5519 |  neurospora_crassa hypothetical protein similar to Tn1
              (520 nt)
Length = 519

Score = 700 bits (1807), Expect = 0.0, Method: Compositional matrix adjust.
Identities = 338/512 (66%), Positives = 400/512 (78%), Gaps = 25/512 (4%)

Query: 1  MADEKRMGS S DKA AVQHSETLPGV SSTA ERG FAATD QNGQP IVQFDLK AEARLRRKL 60
        M EKR S           + E PG + AERG AATD +G P+V+ D AE +LRRK+
Sbjct: 9  MTSEKRRQSL SESDTKEGYFENAPC~AHYRAERGQAATDIHGNPL VELDPVAETKLRKI 67

Query: 61  DLFI VPTV SLL YLFC FIDRANIGNARIAGLE KDLNLTGYDYNALLSVFY I SYIVF EIPSN 120
        DL++VPTV++LYLFC FIDRANIGNAR+ LEKDL+L GYDYNALLSVFY+ YIVF EIP+N
Sbjct: 68  DLYVVPTV AIL YLFC FIDRANIGNARLDKLEKDLHGYDYNALLSVFY VGYIVF EIPAN 127

Query: 121 IACKWIGPGWFPIPAISLGFGV VSLATAFVDNF AQAAGVRFLLGVF EAGMMPGIAYYLSRW 180
        I CKW+GPGWF+P SLGFG++S+ AFV+NF+QA GVRFLLGVF EAGM+PGIAYYLSRW
Sbjct: 128 IMCKWMGP GWFLPLTSLGF GIM SVCMAFVN NFSQACGVRFLLGVF EAGMLPGIAYYLSRW 187

Query: 181 YRRAELTFRLS LYIVMAPMAGAFGGLLASGILSLDHVGVTGWRMIFV VEGIITIGLSVI 240
        YRR+ELT RLS LYIVM+P+AGAFGGLLASGIL LDH G + GWRMIF +EGIIT+GLS+I
Sbjct: 188 YRRSELT LRLS LYIVMSPLAGAFGGLLASGILKLDHF GS LHGWRMIFGIEGIITVGLSLI 247

Query: 241 SFITLTDR PETARWL TQEEKDLAIARVK SERV ATTEV LDRMDT KKL IQGILSPVTLATSF 300
        F+TLTD P TA+WL+QEEKDLAIARVK SER+ TE++D+MD KKL +GI +PV T F
Sbjct: 248 GFITLT DH PATAKWL S QEEKDLAIARVK SERI GQTEII DKMDA KL KRGIFNPV VFTGF 307

Query: 301 MFLNNIT-----QLFTVPPY VVGGFFT LALPLLSWYLD RR 336
        F LNNIT                   QL TVPPY+VGGFFT+ +PL+S +LD RR
Sbjct: 308 AFLLNNITVQGLAFFAPTVVATIYPTKNI IQKQLL TVPPY I VGGFFT VLMPLISR WL D RR 367

Query: 337 QIII ILLSTPLV I VGYSMFL GTTNPSARYGATFLLSSSL FAVG AL SNSQV SANV VSDTARS 396
        QIII+ PLV+VGY MFL T N RYGA FL+S+S+FA G L+NSQV SANV VSDTARS
Sbjct: 368 QIII I ACCPLV MVGY IMFLATE NAHVRYGA AF LVST SVFAAGPLT NSQV SANV VSDTARS 427

Query: 397 SAIGL NVMMGNVGGLI ATWSY LPWDGP NYKIGN GLNL AACCTVL ILSAVT LLWMKWDNRR 456
        SAI NVM+GNV GGL+ATWS+LP+D P Y IGNG+NLAA LI++ L+WMK DN +
Sbjct: 428 SAIAYNVMLGNVGGLV ATWSFLPFDAPKY HIGNGTINLA SGGT LTIALCLLI WMKRDNNK 487

```

Figure 10-10

```

Query: 457 REGRNAEEELAGMSRQEIQDLDWKHPAFRWRP 488
       RE RNAEEEL GMS+QEIQDLD-KHP FRW+P
Sbjct: 488 RELRNAEEELTGMSSQEIQDLDYKHPGFRWKP 519

Query= jgi|Spoth1|79030|estExt_Genewise1Plus.C_31624
          (519 letters)

Database: neurospora_crassa_7_proteins_no_asterisks.fas
           9822 sequences; 4,775,003 total letters

Searching.....done

Score      E
Sequences producing significant alignments:          (bits) Value
NCU01231 |  neurospora_crassa hypothetical protein similar to ca...   834   0.0
>NCU01231 |  neurospora_crassa hypothetical protein similar to
              carboxylic acid transport protein (520 nt)
Length = 519

Score = 834 bits (2154), Expect = 0.0, Method: Compositional matrix adjust.
Identities = 395/503 (78%), Positives = 441/503 (87%), Gaps = 1/503 (0%)

Query: 1 MESTHEPADPIAKGVLATAKQSWHDLFIFKQRVVVTNTELGETSTEWARPVPLRNPISSL 60
       MESTHEPADP+AKG+L TA+QSW DLF+I+KQRVVVTN GET+TEWA+PVPL+NPISLLA
Sbjct: 1 MESTHEPADPVAKGILPTARQSWKDLFIWKQRVVVTNVYGETATEWAKPVPLKNPISSL 60

Query: 61 QLSARNWLFFIVGFLAWVADAYDFHALSIQQVKLAEFYNTTKINISTAITLTLLLRSVGA 120
       QLS R+W+ F+FVG AW ADA+DFHALSIQQVKLA +Y +KT++STAITLTLLLRS+GA
Sbjct: 61 QLSGRDWICFLVGFCAWSADAFDFHALSIQQVKLAAYYGVSKTSVSTAITLELLRSIGA 120

Query: 121 AFFGLAGDKWGRKWPVMANMIVLGVLQIGTIYSVTFSDFLAVRALFGLFMGGVYGNIAIM 180
       A FGLAGD+WGRKWPVM NMIVLG+LQI TIYS T+S FL VRALFGLFMGGVYGNIAIM
Sbjct: 121 AAFGLAGDRWGRKWPVMVNMIVLGILQIATIYSSTYSQFLGVRALFGLFMGGVYGNIAIM 180

Query: 181 ALENSPPDARGLMSGILQQGYSLGYVIAACANLGVGDDNSWKTIVFWIGAGLSIGVGLR 240
       ALENSP DARGLMSGILQQGY+ GYV AACANEVGG +SWKTVFWI AGLSIGVG+ +R
Sbjct: 181 ALENSPVDARGLMSGILQQGYAFGYVCAACANLGVGDDTSWKTIVFWIAAGLSIGVGIIR 240

Query: 241 CFFPESQQFLEARAAGKGQASASAFWKEVKMMLAQEWMKCVCYCILMLTFNYYSHSQDS 300
       CFFPES+QFLEAR GK A+ S FW+ETK+ML QEWMKCVCYC ILMTWFNYYSHSQD+
Sbjct: 241 CFFPESKQFLEARKEGKAHANPSQFWRETKVMLRQEWMKCVCYCILMLTFNYYSHSQDN 300

Query: 301 YTTFMLTQKELDNDGASRASILMKVGACVGGTIIGYISQWFGRRTIIVAAALISGLIIPA 360
       YTTF+L KE+DN ASRASI+MK GACVGGTIIGY+SQ+FGRRRTIIV+ +LISG +IPA
Sbjct: 301 YTTFVLRAKEMDNAASRASIIMKAGACVGGTIIGYLSQYFGRRRTIIVSSLISGCMIPA 360

Query: 361 WILPEGERSLSVTGFFMQFFVQGAWGVIPIHLNELSPPAFRSSFPGLTYQLGNMISSPSA 420
       WILP ER+LS TGFFMQFFVQGAWGVIPIHLNEL+PPAFRSSFPG+TYQ+GNM+SSPSA
Sbjct: 361 WILPNSERALSATGFFMQFFVQGAWGVIPIHLNELAPPFRSSFPGITYQVGNMVSSPSA 420

Query: 421 QIVNAIAESHSVTSKSGKSVNAYGPTMGIATAIIATGIAVTTALGPEKRGREFEKLIPAG 480
       QIVNA++E + S +GK V AYGPTMGIATAII GI VTIA GPEKRGREFEK LPAG
Sbjct: 421 QIVNAVSEKIHIVSHTGKLVEAYGPTMGIATAIIIVMGIVVTTAFGPEKRGREFEKALPAG 480

```

Figure 10-11

```

Query: 481 MNI--IQGGKAVDDLEKGDSRDEK 502
        MN+ Q GK VDDLE EK
Sbjct: 481 MNLQKQHKGQVDDLEMETGHMEK 503

Query= jgi|Spoth1|108890|estExt_fgenesh1_pg.C_60848
      (533 letters)

Database: neurospora_crassa_7_proteins_no_asterisks.fas
          9822 sequences; 4,775,003 total letters

Searching.....done

Score      E
Sequences producing significant alignments:           (bits) Value
Score = 878 bits (2269), Expect = 0.0, Method: Compositional matrix adjust.
Identities = 429/538 (79%), Positives = 471/538 (87%), Gaps = 10/538 (1%)

Query: 1   MGLS--TKILQKIVRNEAMASDPPEIYGRVYLLACSGAMSFGWDSSVIGGVIVLPP 58
        MGLS +IL+KIV+NEAMA DPPEIYGRVYLLACSGAMSFGWDSSVIGGVI L P
Sbjct: 1   MGLSIGNRILRKIVRNEAMAEDPPEIYGRVYLLACSGAMSFGWDSSVIGGVIEEP 60

Query: 59  FIRDFN-LCDPKSQASANLSANIVSTLQAGCFLGALVASPMTDRFGRKWCLIGVSLIII 117
          F DF +G+ K A ANL ANIVSTLQAGCFLGAL+ASP+TDRFGRKWCLI VSL++II
Sbjct: 61  FKHDFFGFIGNDK--AKANLGANIVSTLQAGCFLGALIASPITDRFGRKWCLIAVSLVII 118

Query: 118 GIIMQAAASGNLGPMPYAGRFTIAGAGVGAASITINPIYVSENAPRAIRGLLTGLYQLFIVTG 177
          GIIMQAAASGNL PMY GRF+AG GVGAAS INP++VSENAPR+IRGLLTGLYQLFIVTG
Sbjct: 119 GIIMQAAASGNLAPMYIGRFVAGVGVAASCINPVFVSENAPRSIRGLLTGLYQLFIVTG 178

Query: 178 GMIAFWINYSVSIHFPEVKIMYVFPLAIQALPAALLECLCMILLCQESPRWLARRDRWEDTK 237
          GMIAFWINYSVS+HF + K MY+FPLAIQ LPA LLC+CMLLC ESPRWLARRDRWE+ K
Sbjct: 179 GMIAFWINYSVSLHF-KGKSMYIFPLAIQGLPAGLCLVCMLLCESPRWLARRDRWEECK 237

Query: 238 RVLSRIRNLPPDHPYIQQDEFQEIVAQLEHERRLIGDASFVNLLQREMWTIAGNRRVLISI 297
          VL+RIRNLPPDHPYI DEF+EI QLE ERRL GDA++W+L R+MWI+AGNR+R LISI
Sbjct: 238 SVLARIRNLPPDHPYIVDEFREIQDQLEQERRLQGDATYWLDRDMWTVAGNRKRALISI 297

Query: 298 ILMICQQMTGTNAINTYAPTIKFNLGLTGTSTSFLSTGVYGIKVVTSCIIFLLFMADSLG 357
          LMICQQMTGTNAINTYAPTIKFNLG+TGTSTSFLSTG+YGIKV KV SC+IFLLF+ADSLG
Sbjct: 298 FLMICQQMTGTNAINTYAPTIKFNLGITGTSTSFLSTGIYGIKVVKVSCVIFLLFLADSLG 357

Query: 358 RRRSLLWTSIAQGLAMFYIGLYVRAPPKEGESVPPAGYFALVCIFLFAAFFQFGWGPGAC 417
          RRRSLLWTSIAQGLAMFYIGLYVR+PP +G+ VPPAGY ALVCIFLFAAFFQFGWGPGAC
Sbjct: 358 RRRSLLWTSIAQGLAMFYIGLYVRISPPIIDGQPVPPAGYVALVCIFLFAAFFQFGWGPGAC 417

Query: 418 WIYASEIPAARLRSLNVAYAAATQWLFNFVVARTVPVMIVTMGEGGYGYTLFGSFCFSM 477
          WIYASEIPAARLRSLNV+YAAATQWLFNFVVVAR VP M+VI+G GYGYTL+FGSFC SM
Sbjct: 418 WIYASEIPAARLRSLNVSIAATQWLFNFVVARAVPTMLTVGPHGYGYTLIFGSFCLSM 477

```

Figure 10-12

Query: 478 FVFWFFVPETKGVSLAAMDKLFGVTD----ESSKSLTVDEDAKEKEKDGP^{HARQTEV} 531
FVFWFFVPETKG+SLE MD+LFGVTD E S D+ E K ++ EV
Sbjct: 478 FVFWFFVPETKGI^SLEHMDELFGVTDGPAAEKSSVHGGDDVGSEM^GKGDQSKHVEV 535

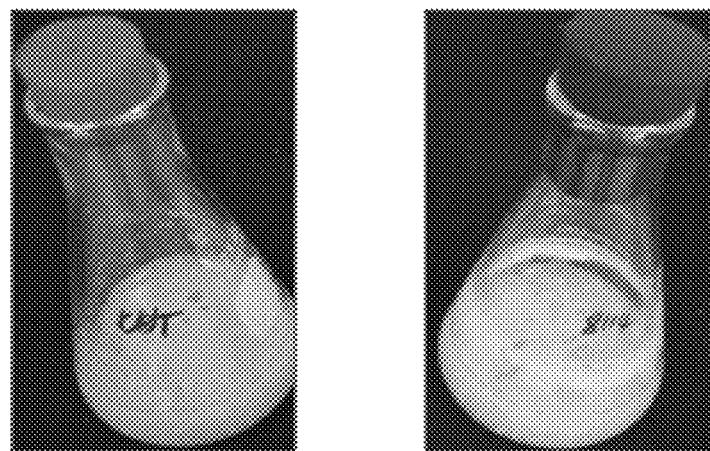
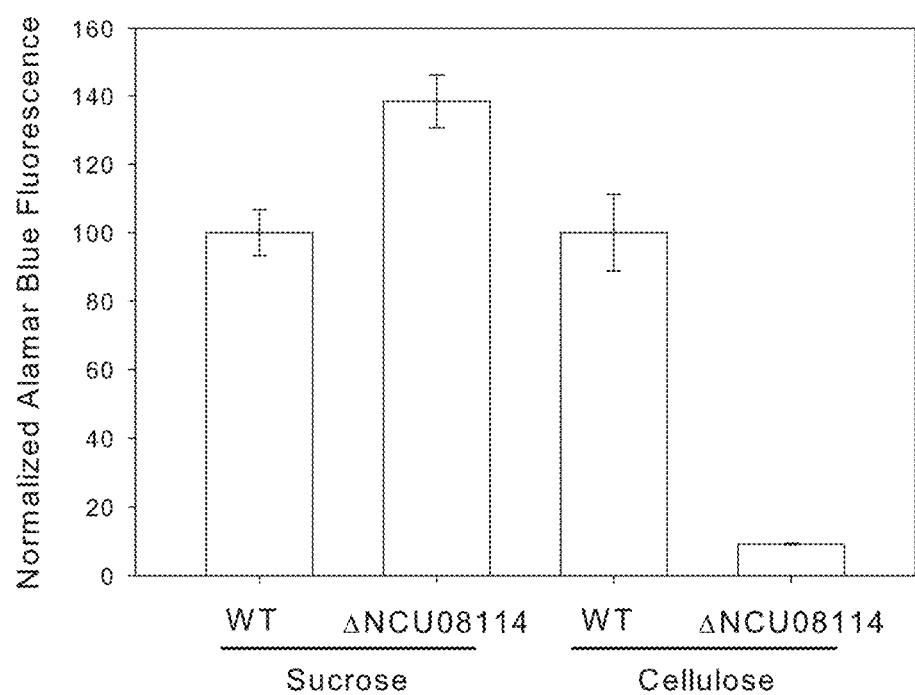
Fig. 11**A****B**

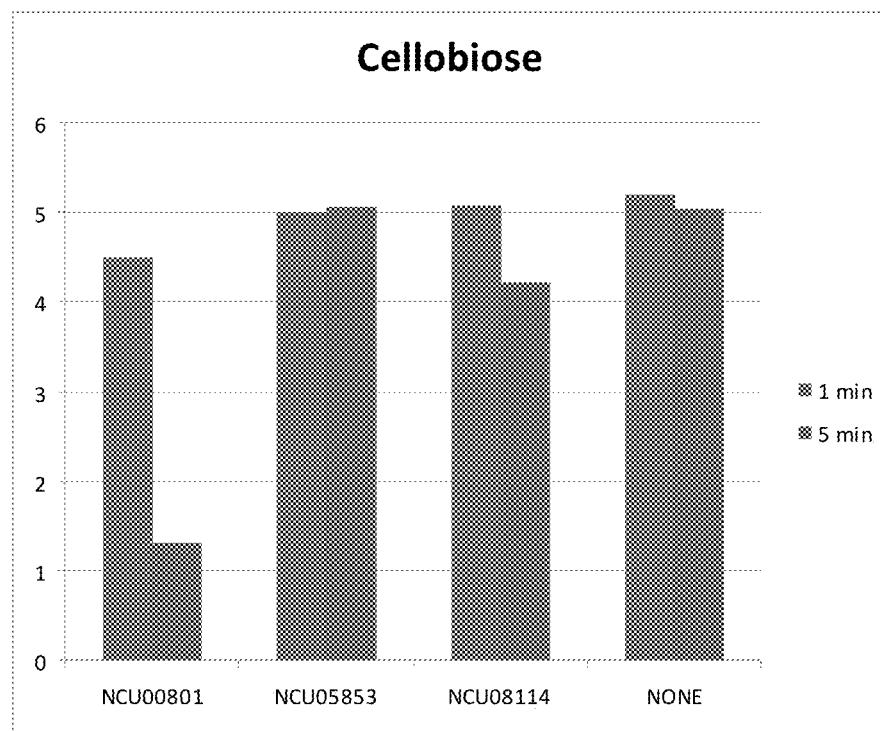
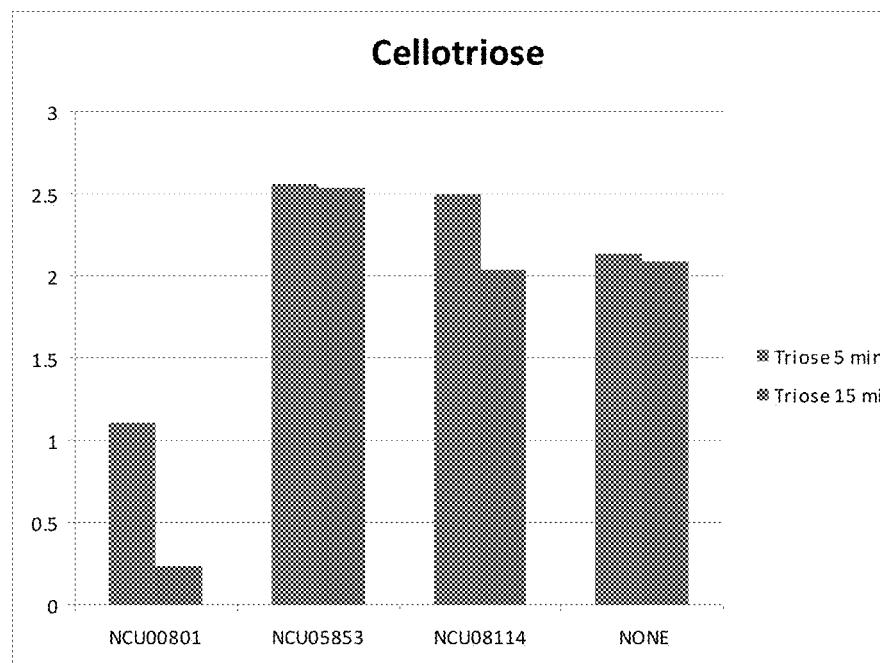
Fig. 12**A.****B.**

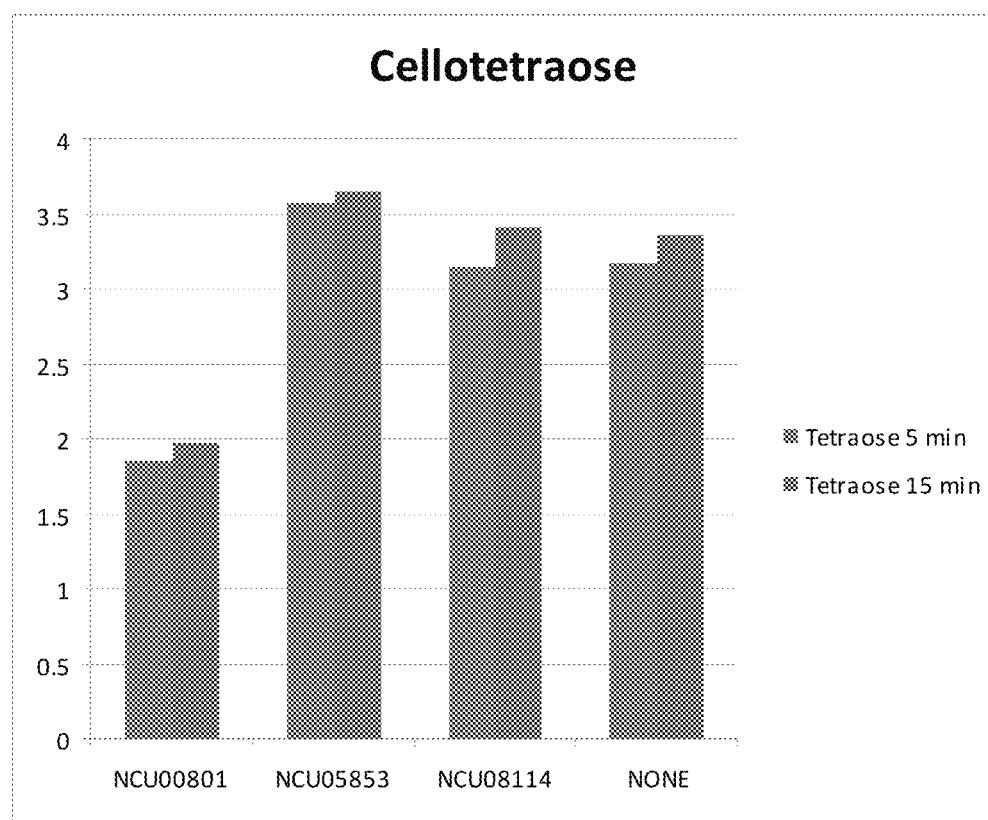
Fig. 12 (cont.)**C.**

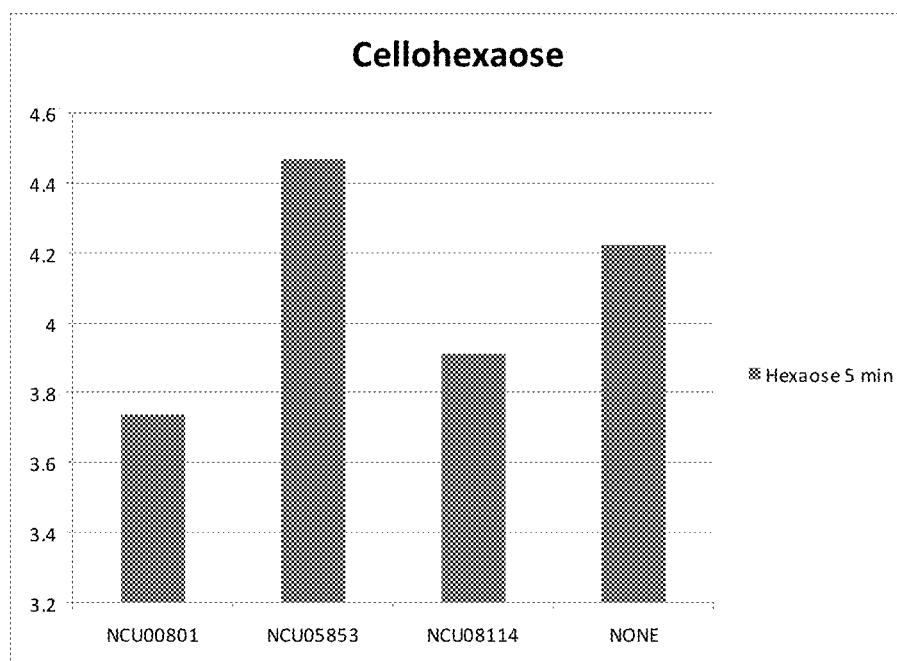
Fig. 12 (cont.)**D.**

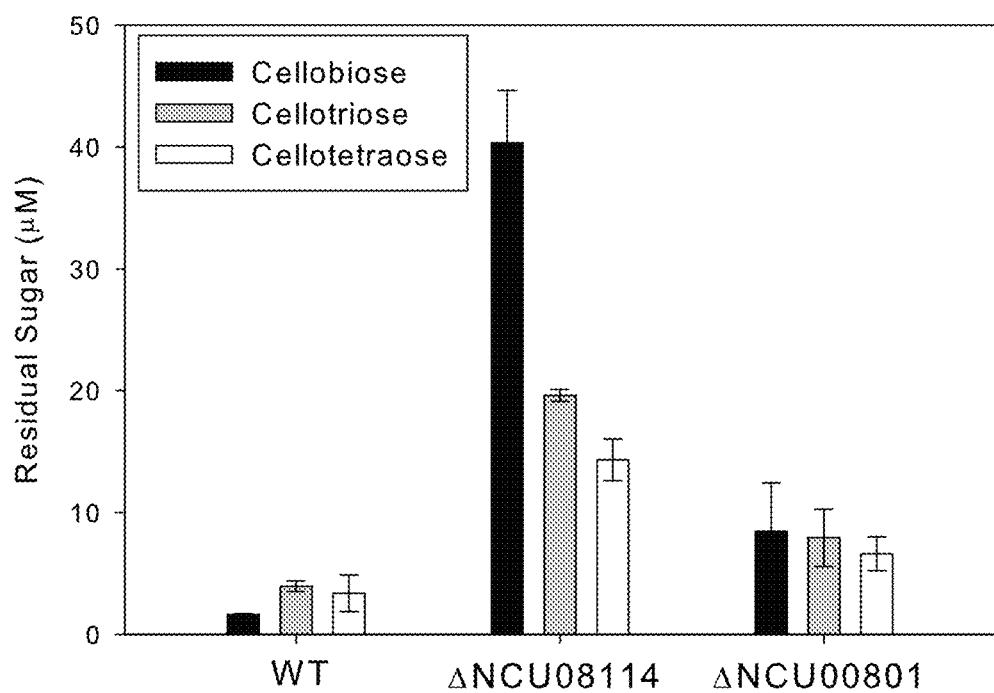
Fig. 13

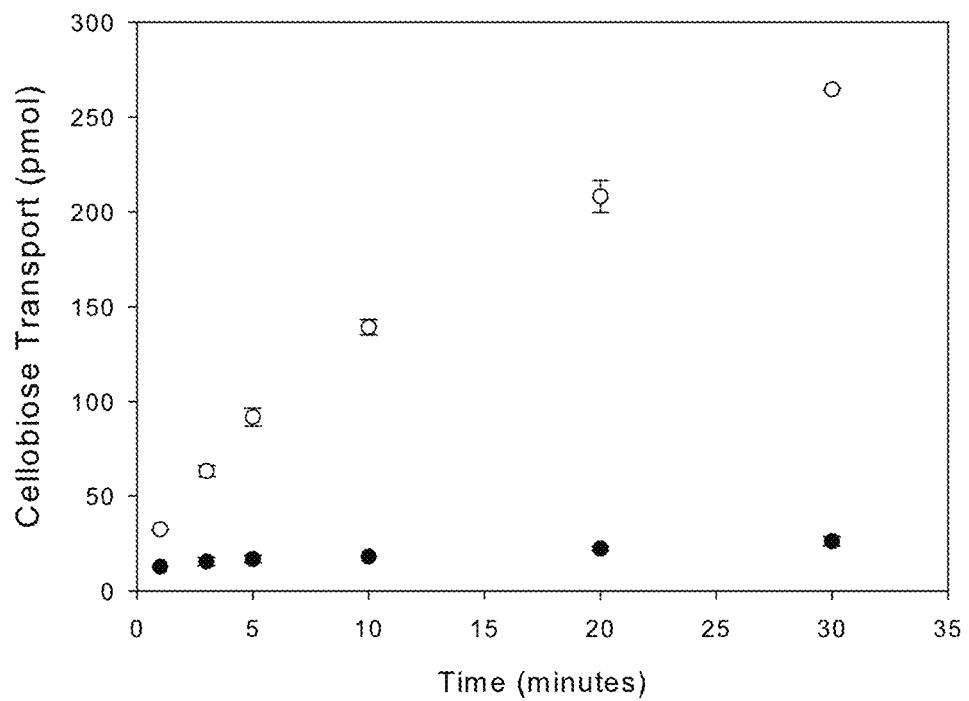
Fig. 14

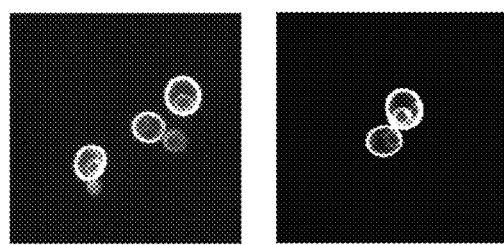
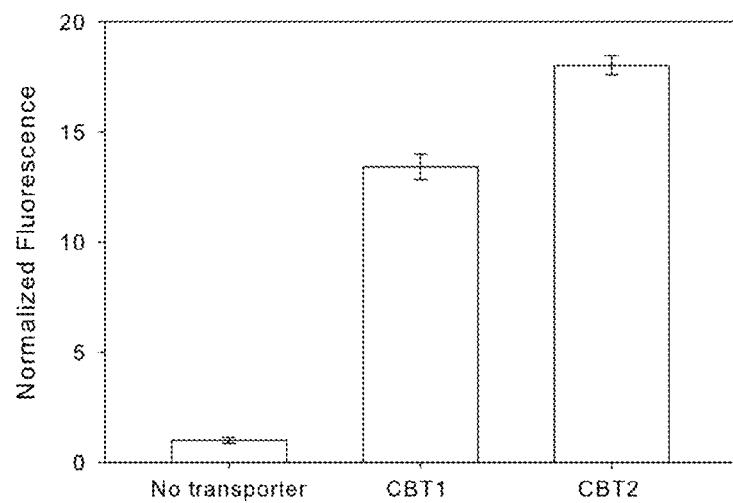
Fig. 15**A****B**

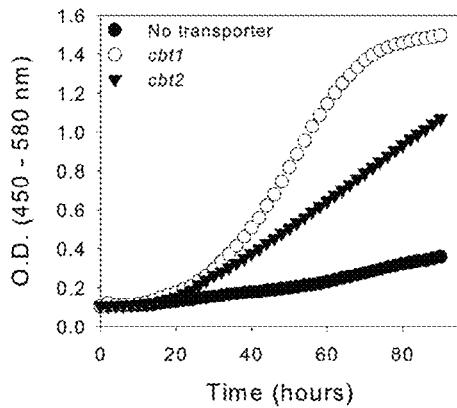
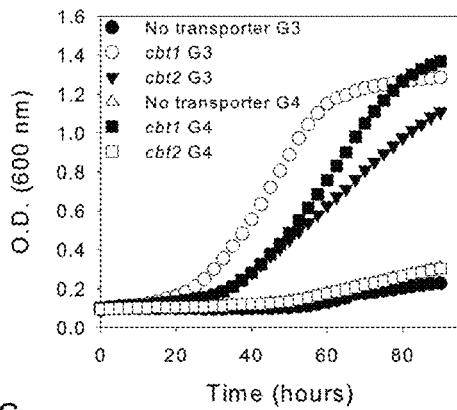
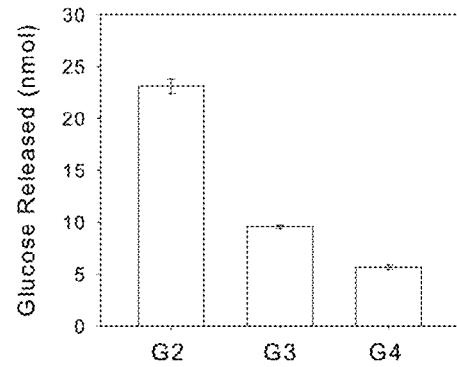
Fig. 16**A****B****C**

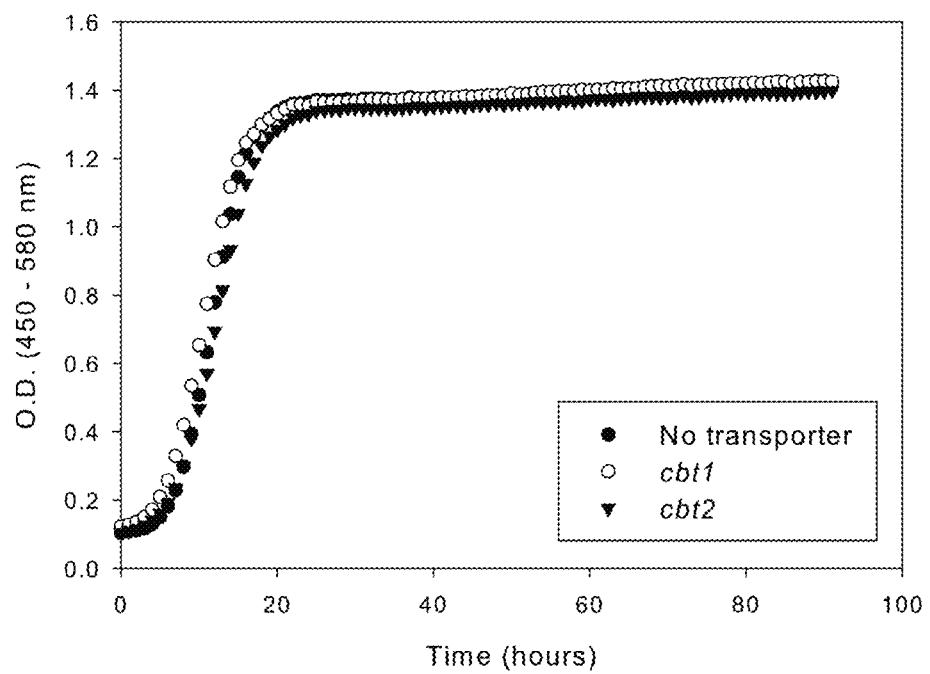
Fig. 17

Fig. 18

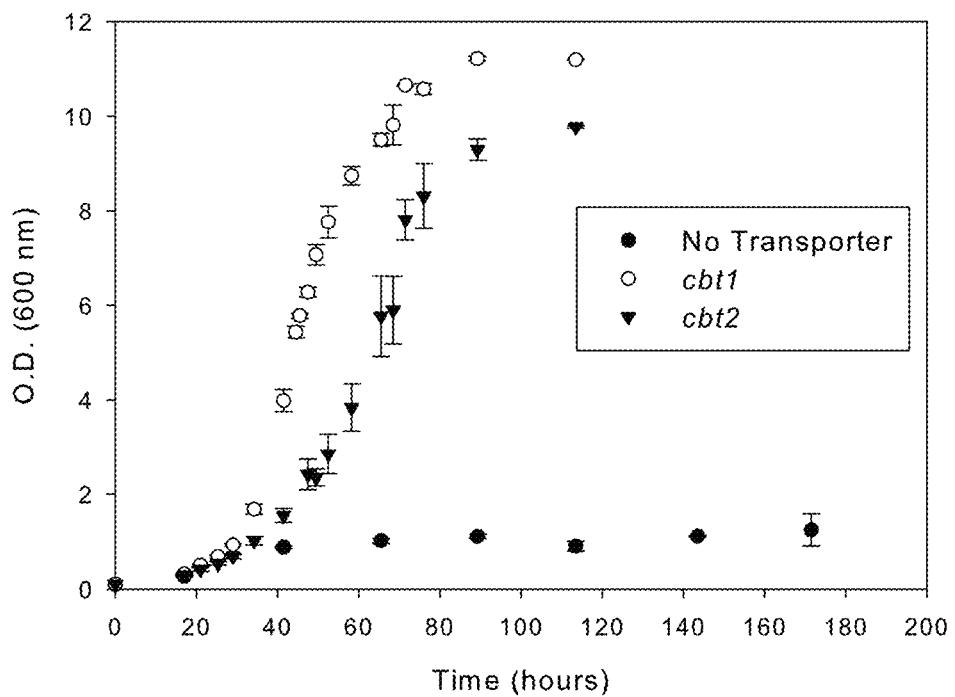


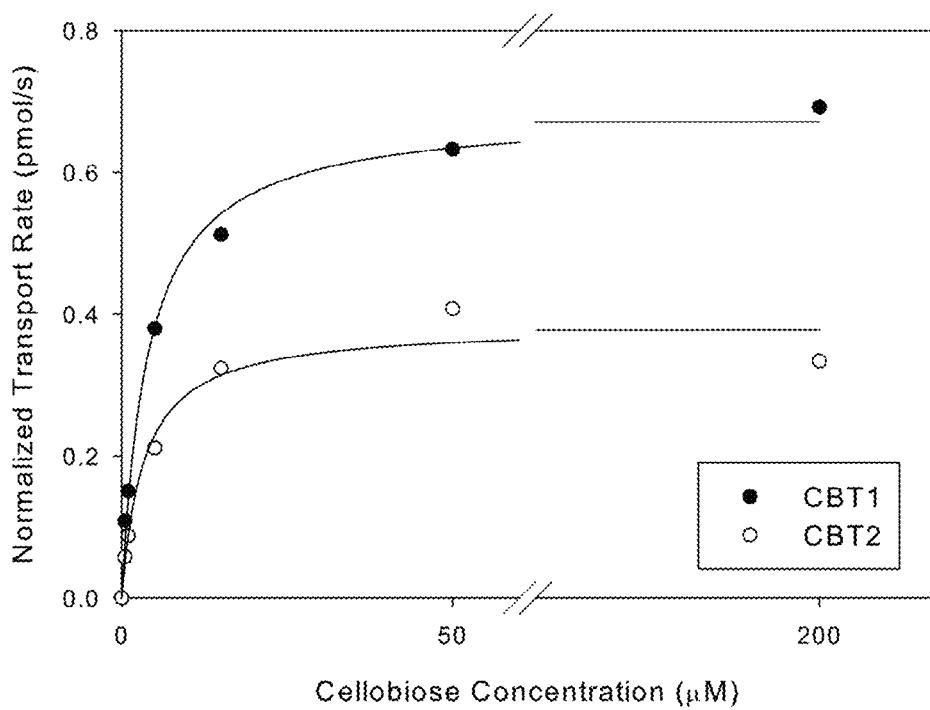
Fig. 19

Fig. 20

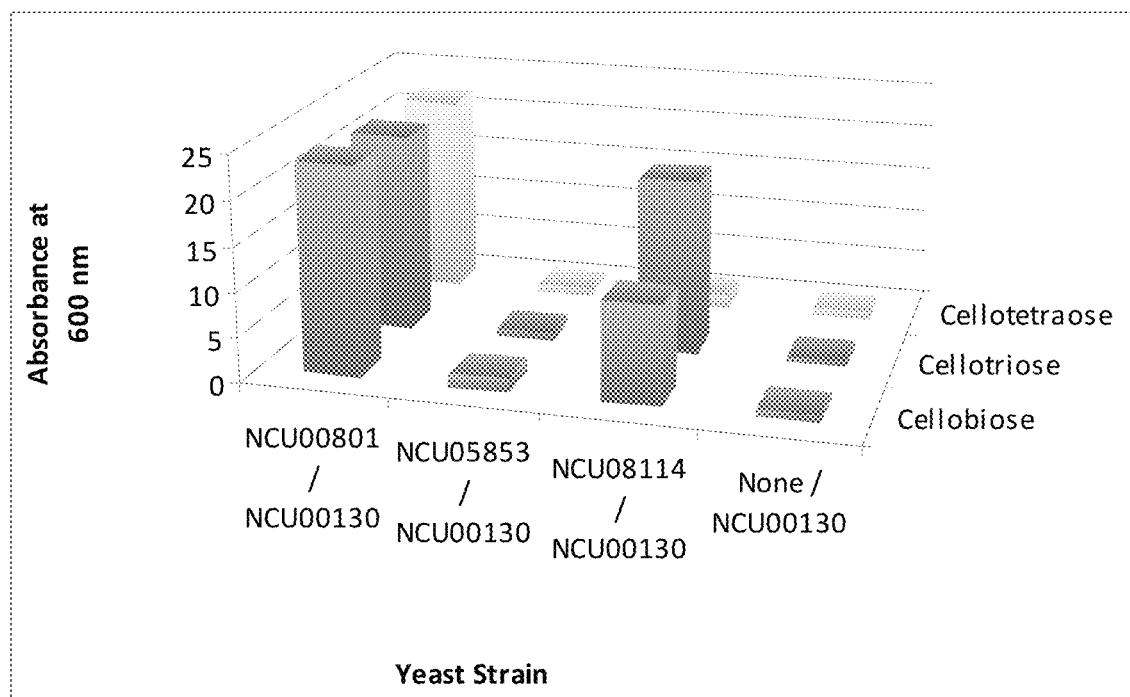


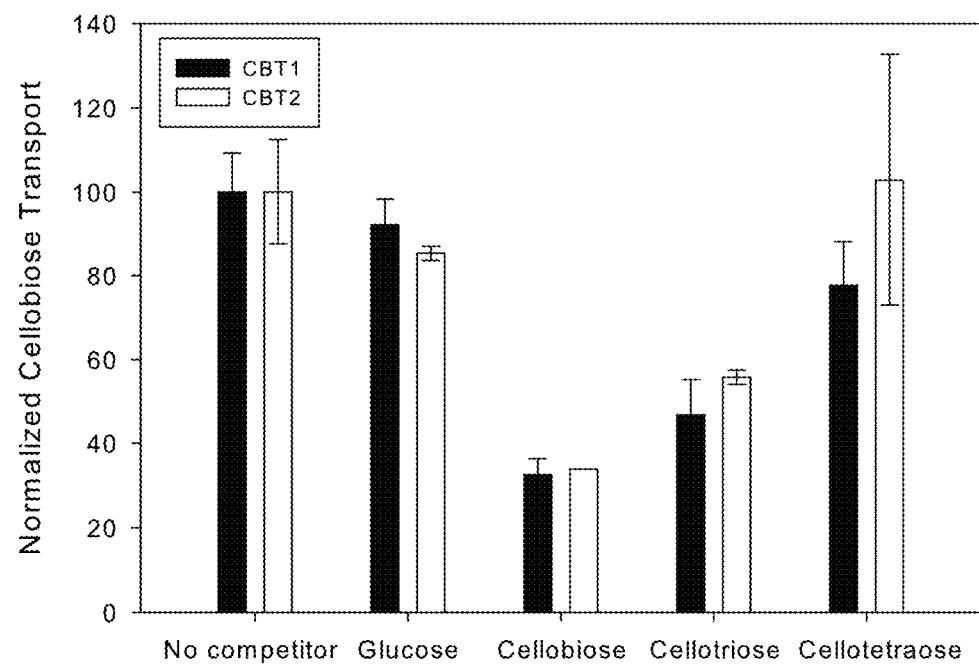
Fig. 21

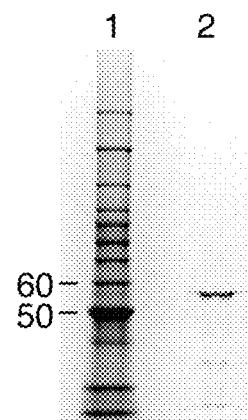
Fig. 22

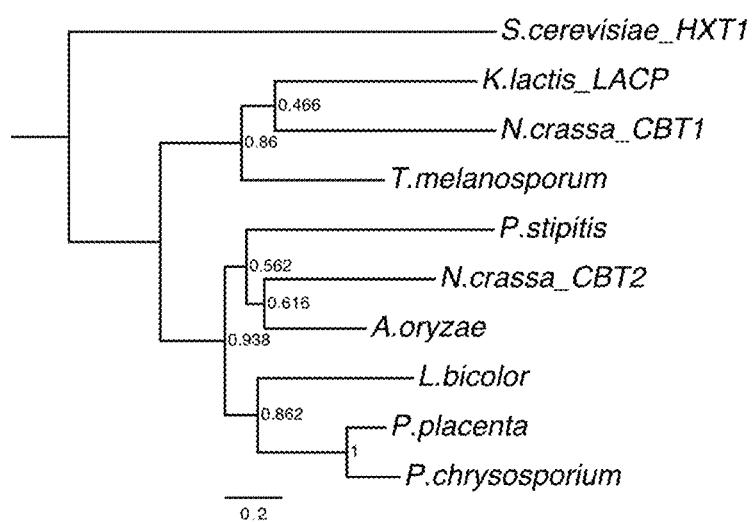
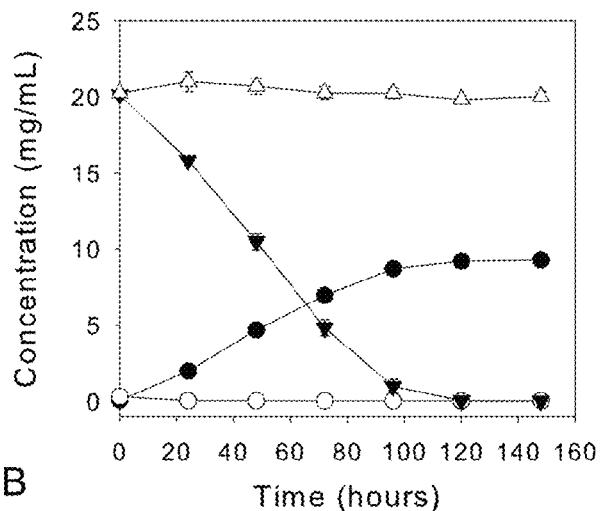
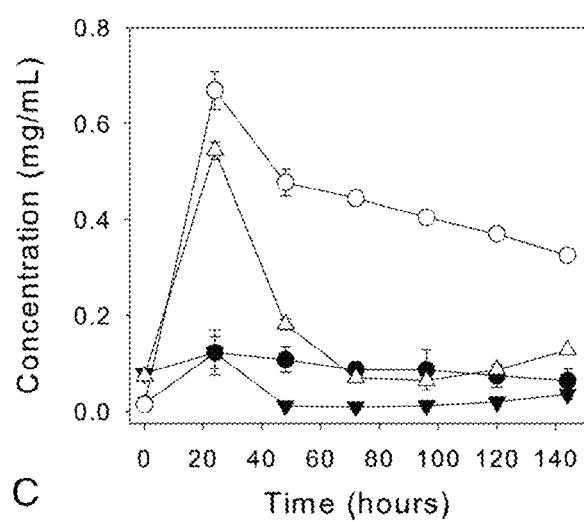
Fig. 23

Fig. 24

A



B



C

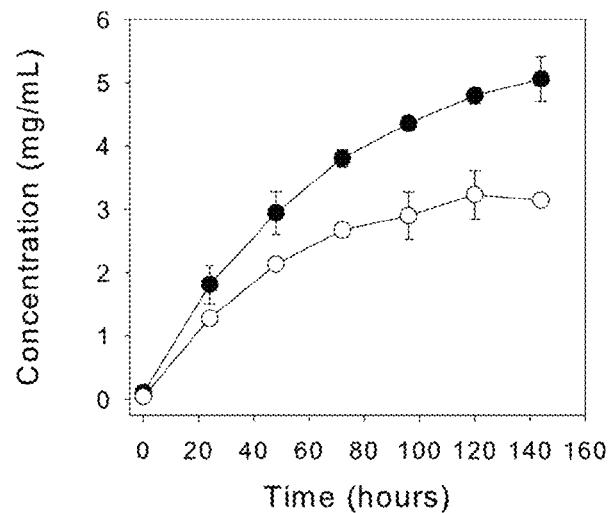


Fig. 25

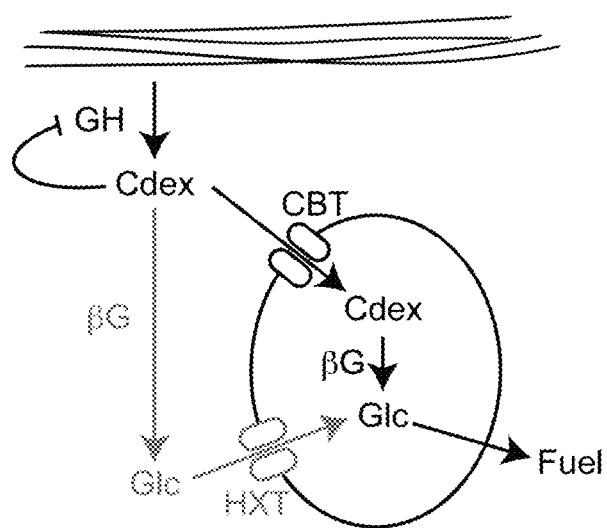


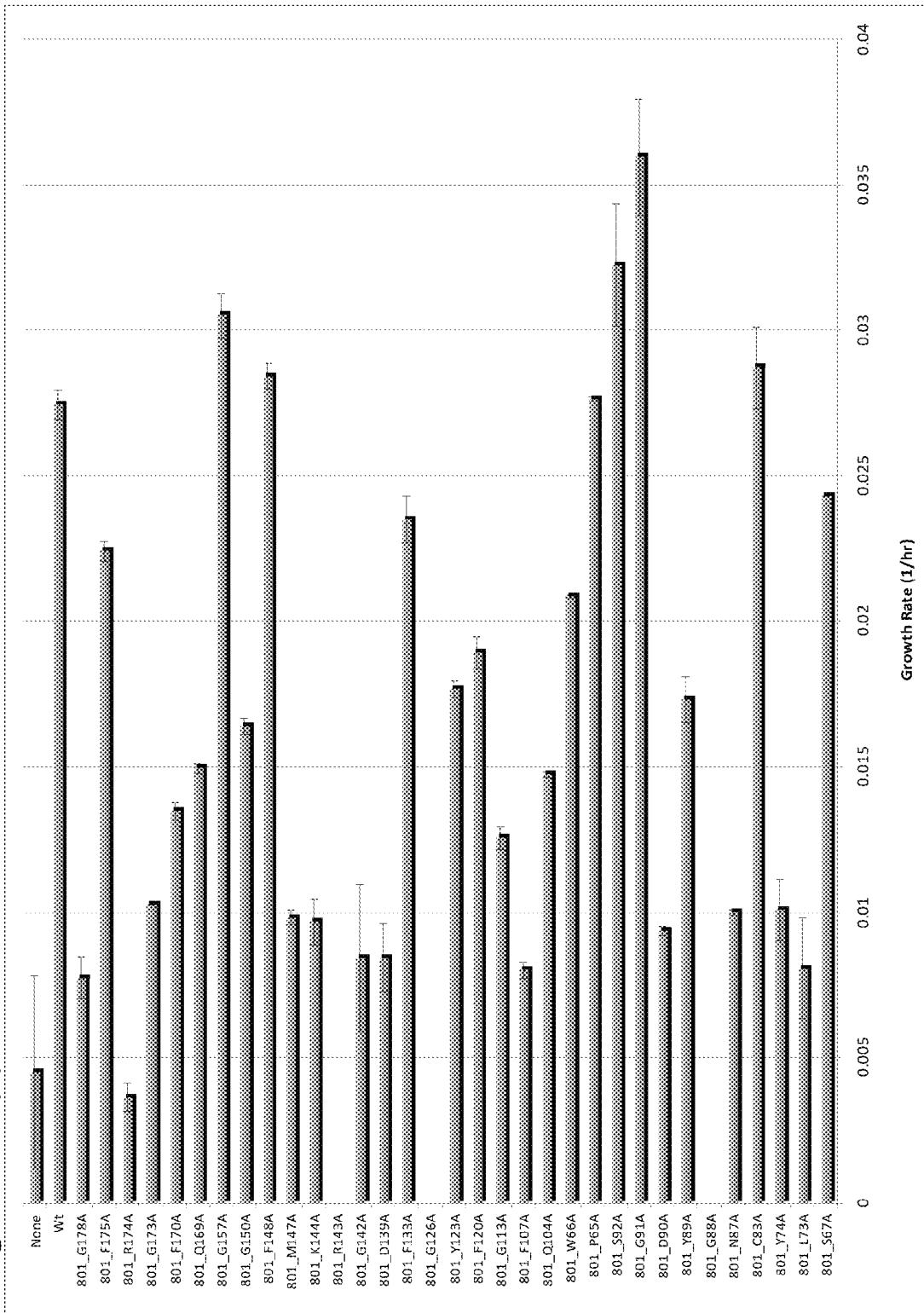
Figure 26 (a)

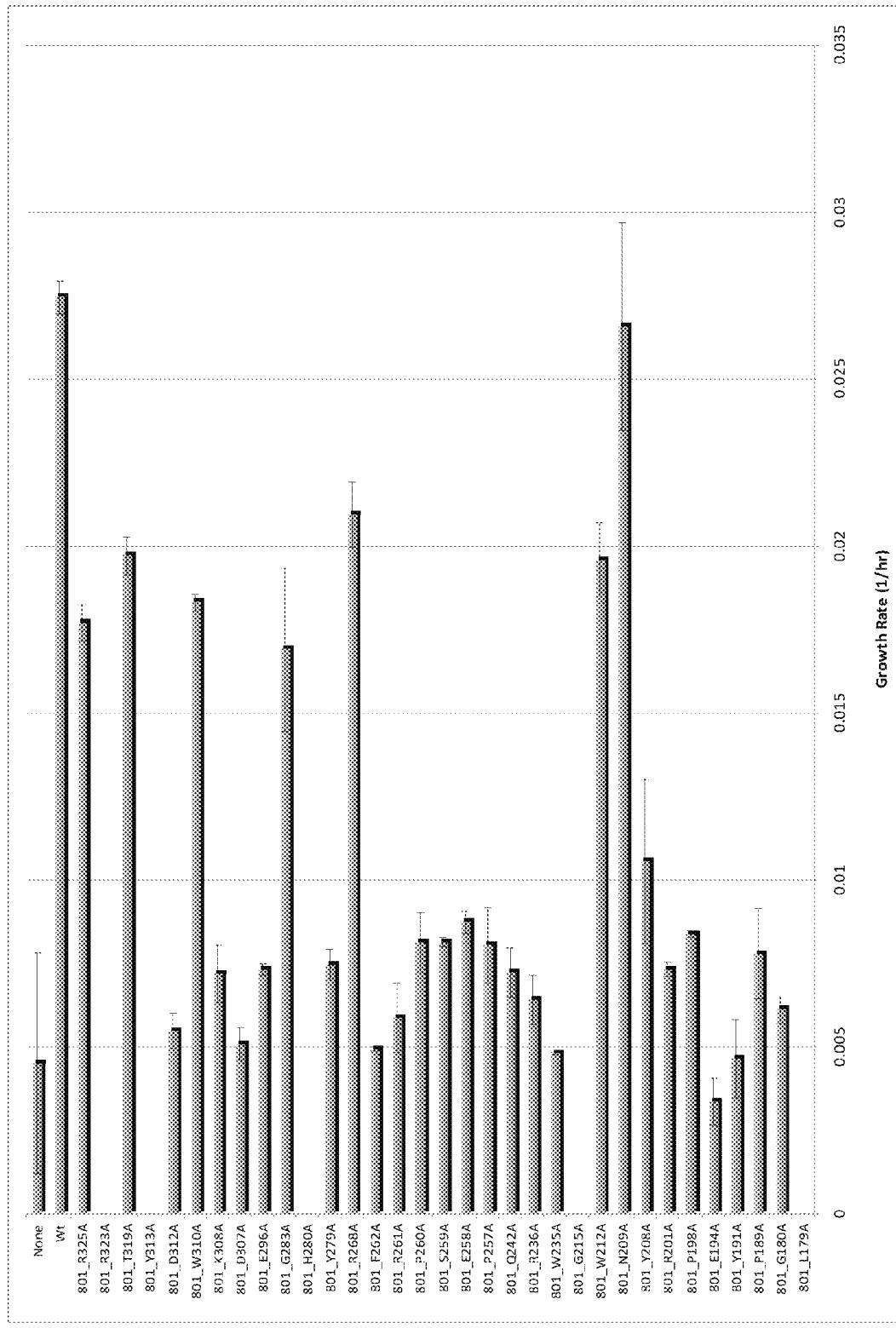
Figure 26 (a) (cont.)

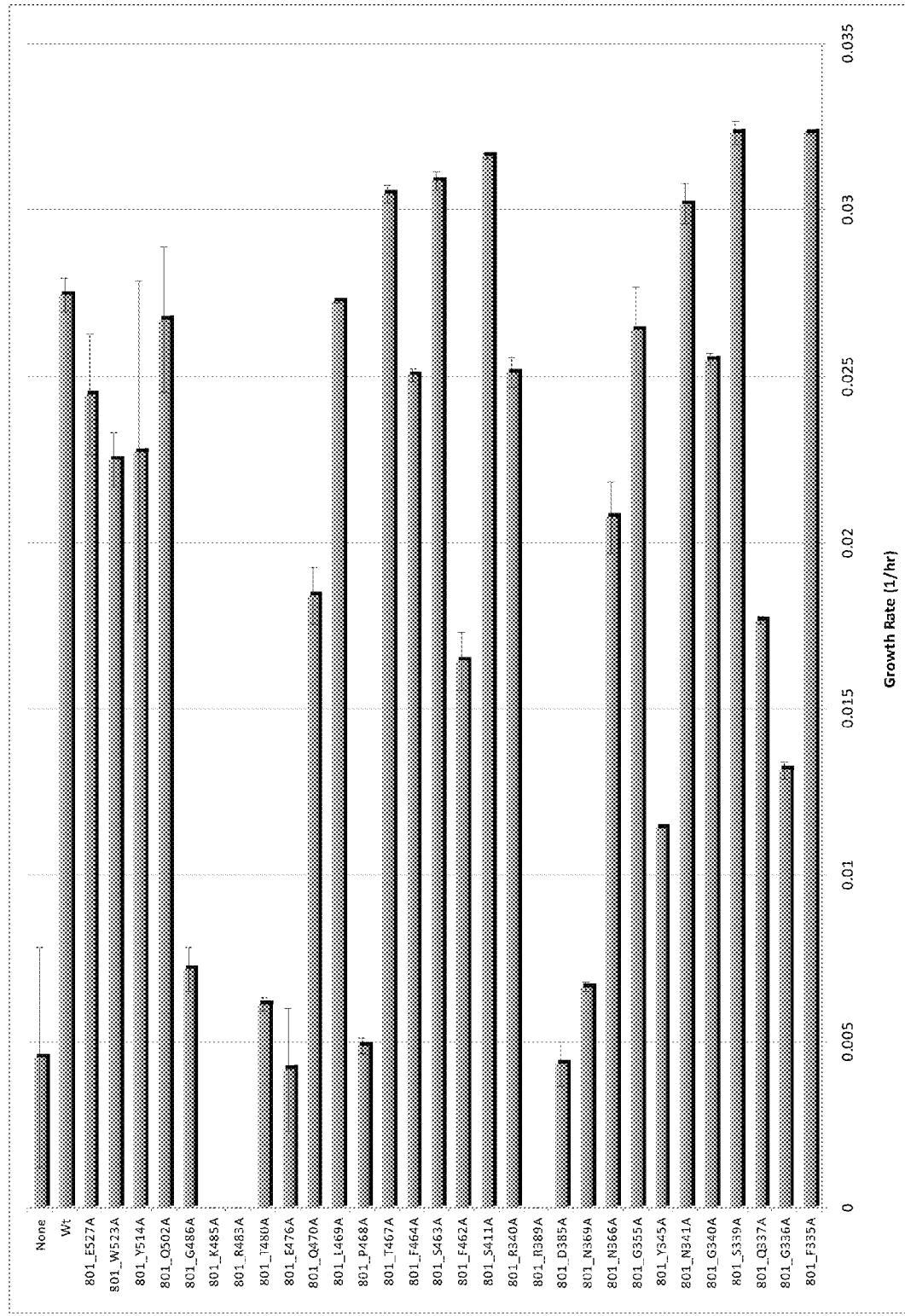
Figure 26 (a) (cont.)

Fig. 26

(b)

Polypeptide Sequence: NCU00801

MSSHGSHDGASTEKHLATHDIAPTHDAIKIVPKGHGQTATKPGQAQEKEVR
NAALFAAIKESNIKPWSKESIHLYFAIFVAFCCACANGY**D**GSLMTGI IAM
DKFQNQFHTGDTGPKVSVIFSLYTVGAMVGAPFAAILSDRF**GR**KKGMFIG
GIFIIVGSIIVASSSKLAQFVVGRFVL**GL**GIAIMTVAAPAYSIEIAPPHW
RGRCTGFY**N**CGWFGGSPAACITYGCYFIKSNWSWRIPLILQAFTCLIVM
SSVFFLPESPRFLFANGRDAEAVAFLVKYHGNGDPNSKLVLLETEEMRDG
IRTDGVDKVWWDYRPLFMTHSGRWRMAQVLMISIFG**QFS**GNGLGYFNTVI
FKNIGVTSTSQQLAYNILNSVISAIGALTAVSMTDRMPRRAVLIIGTFMC
AAALATNGLSSATLDKQTQRGTQINLQGMNEQDADNAYLHVDSNYAKG
ALAAYFLENVIESFTYTPLQGVIPTEALETIRGKGLALSGFIVNAMGF
NQFAGPIALHNIGYKYIFVFVGWDLIETVAWYFFGVESQGRTLEQLEWY
DQPNPVKASLKVEKVVVQADGHVSEAIVA

Italicized: All transporters

Double-underlined: β-linked transporters

Capped: NCU00801 clade

Bold: Essential in *S. cerevisiae* Hxt1/Hxt3

Underlined: functionally important according to alanine scanning experiment

Fig. 26 (cont.)

(c)

Polypeptide Sequence: NCU08114

MGIFNKKPVAQAVDILNQIQEEAPQFERVDWKKDPGLRKLFYAFILCIAS
ATTGYDGMFFNSVQNFETWIKYFGDPRGSELLLGALYQIGSIGSIPFVP
LLTDNFGRKTPIIIGCVIMIVGAVLQATAKNLDTFMGGRTMLGFGNSLAQ
IASPMLLTELAHPQHRARLTIYNCLWNVGALVVSWLAFGTNYINNDWS
RIPALLQAFPSIIQLLGIWWVPESPRFLIAKDKHDEALHILAKYHANGDP
NHPTVQFEFREIKETIRLEMESTKNSSYLDFFKSRGNRYRLAILLSLGFF
SQWSGNAIISNYSSKLYETAGVTDSTAKLGLSAGQTGLALIVSVTMALLV
DKLGRRLAFLASTGGMCGTFVIWTLAGLYGEHRLKGADKAMIFFIWVFG
IFYSLAWSGLLVGYYAEEILPYRLRGKGLMVMNMSVQCALTNTYANPVAF
DYFGPDHSWKLYLIYTCWIAAEFVFVFFMYVETKGPTLEELAKVIDGDEA
DVAHIDIHQVEKEVEIHEHEGKSVA

Italicized: All transporters

Double-underlined: β-linked transporters

Capped: NCU08114 clade

Bold: Essential in *S. cerevisiae* Hxt1/Hxt3

Underlined: functionally important according to alanine scanning experiment

Figure 27

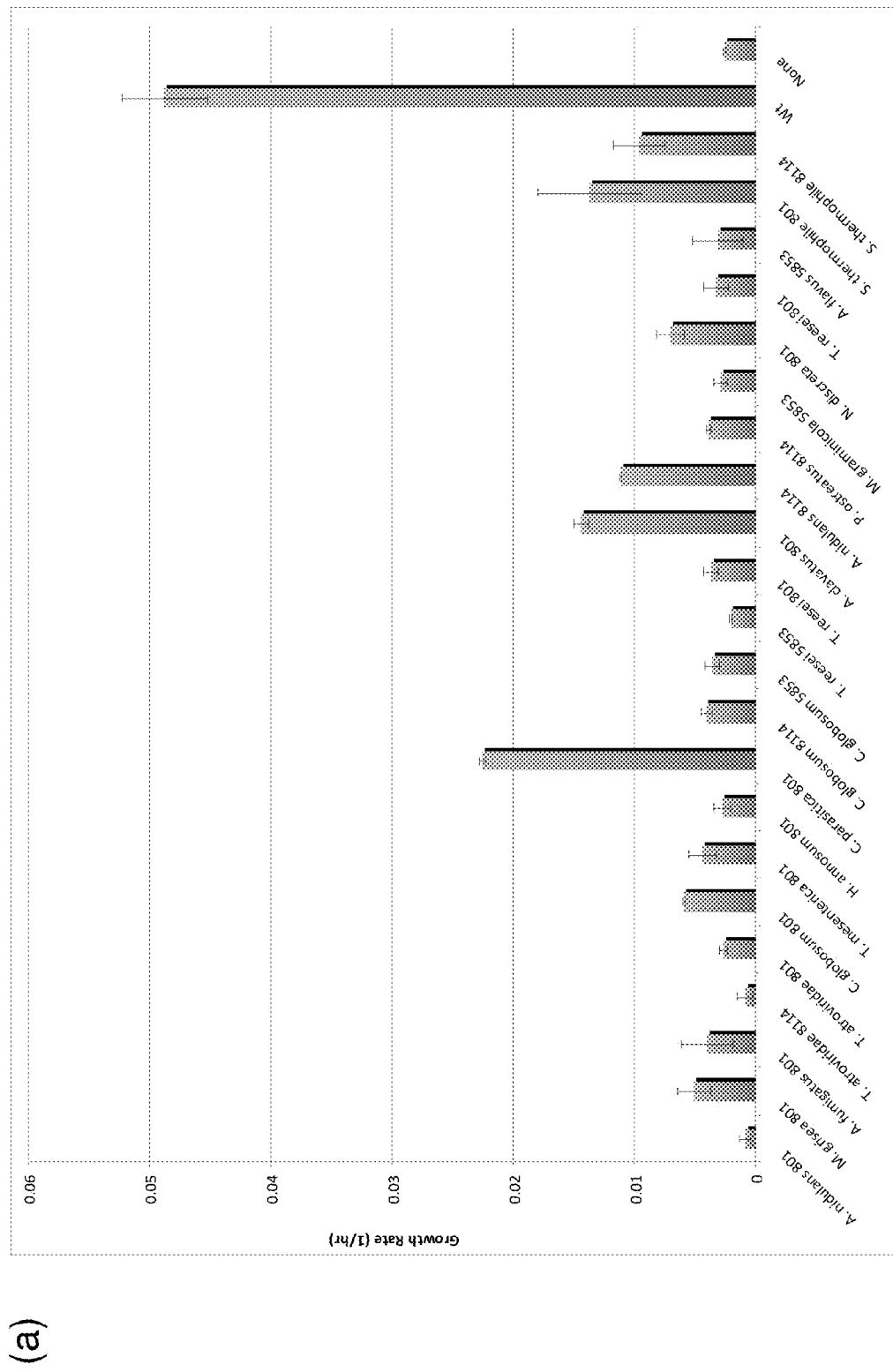


Figure 27 (cont.)

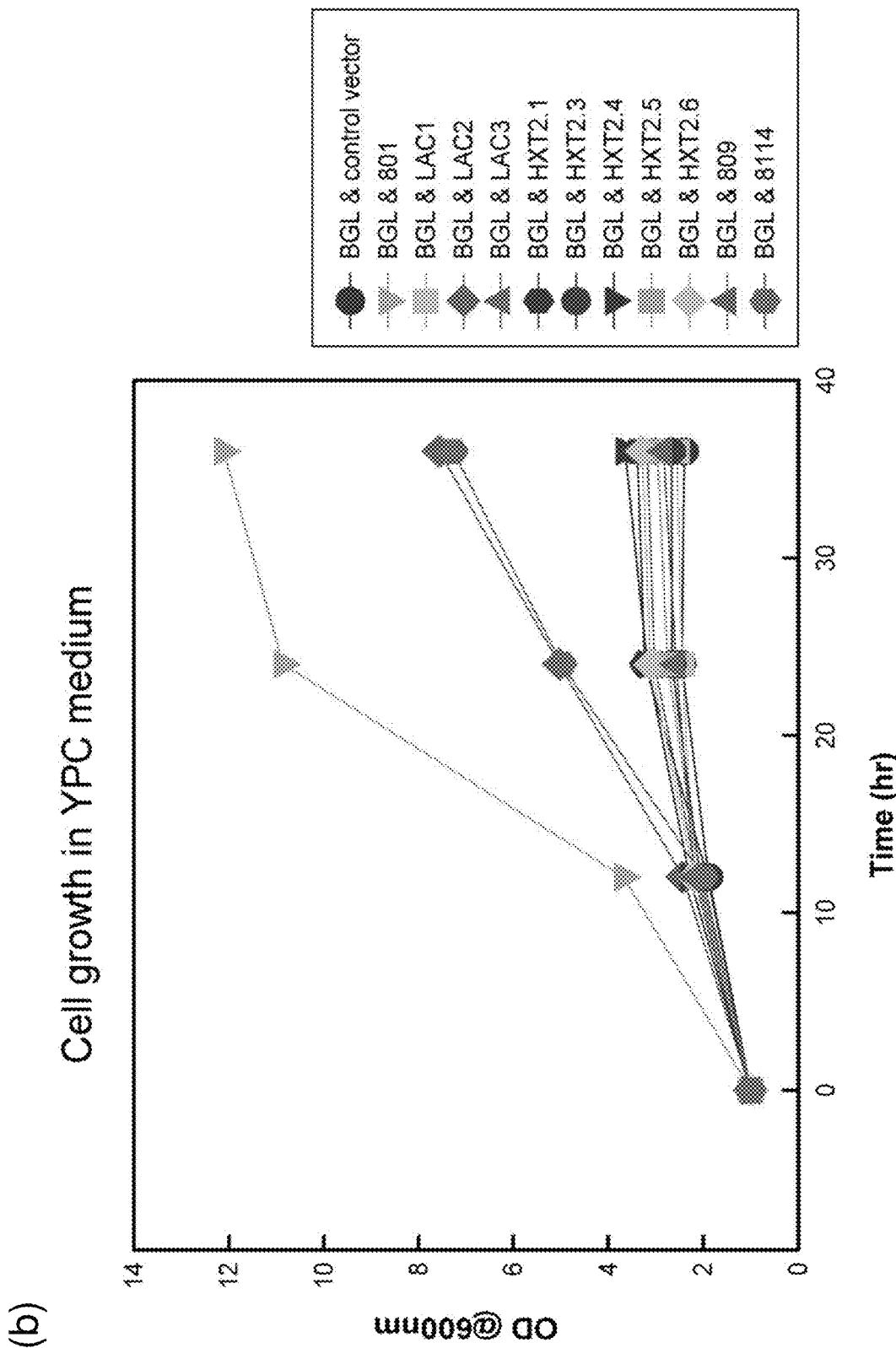


Figure 27 (cont.)

(c)

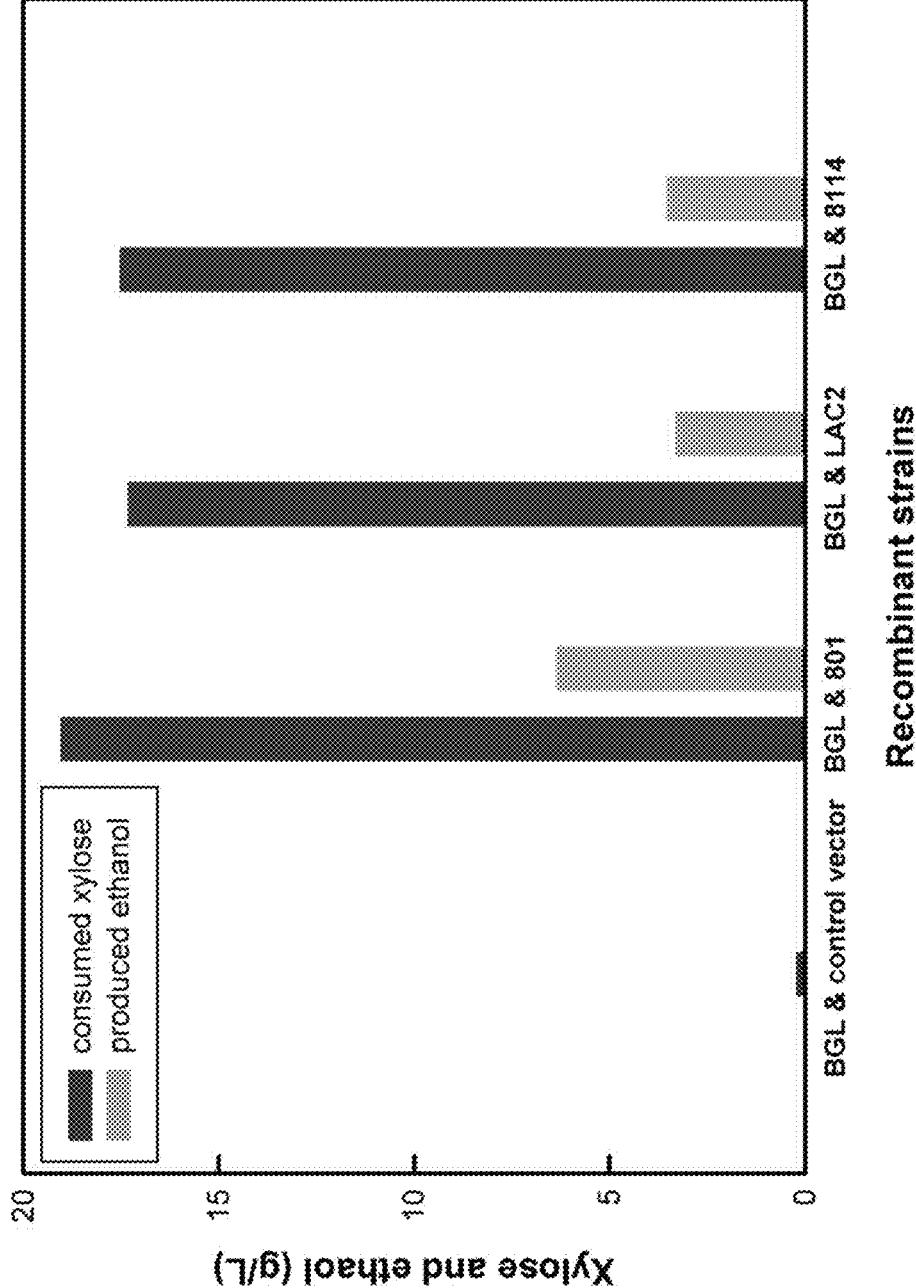
Consumed xylose and produced ethanol

Fig. 28 (a)

Underlined: LYF-PESPR-KYH motifs

Capped: strictly conserved residues, critical to function (Ala-scan results)

NCU00801	-----MSSHGSHD----GASTEKH-LATHDIAPTHDA--IKIVPKGHG
NCU008114	-----MG
LAC2	-----
C_parasiti	-----MSTNSLNDNSYNPSSTKEKD-IVVQSEALADVA--I-----
A_clavatus	MILQDVQTTVFKRIDLIMSAEKTDP----VVDESNH-GISADV----Q--YPVVAEGIG
A_nidulans	-----MSSTD-----KHGISVHHDSPGSDADFKPAVAAGQG
S_thermo_l	-----MGKD-----TFTLA--G-----
S_thermo_i	-----MSSKGS-D-----HIGDEKR-PVSPEIATAAVP--V-----S-----MG
NCU00801	QTA-TK-PGAQEKEVRNAALFA-----AIKESNIKPWSKESIH- <u>LYFAIFVAFCCACAN</u>
NCU008114	IFN-KK-PVAQAVDNLNQIQEEA-----PQFERVDWKDPGLRK- <u>LYFYAFILCIASATT</u>
LAC2	ETAFET-DG-YKK-----IFQEHEPVPRWTKLRLS- <u>IYFTCLVIYLVSTTN</u>
C_parasiti	KTL-DTSAGAKAREVHNAELFA-----AVQESNIERWSKSSIQ- <u>LYFAVFIAPCCACAN</u>
A_clavatus	ETA-KV-GGAKTRQVHNAELYA-----AITETPIEKWSKESIH- <u>MYFAIFVAFCCACAN</u>
A_nidulans	R-EWP-----KVNWWMKGMRSVYLTGAAMVTSATN
S_thermo_l	HTP-KV-QG-KEKKVHNAELYA-----AILEAKIEPWSRTSIH- <u>LYFSIFIAPCCACAN</u>
S_thermo_i	IF-----AFNKQKPNAEATAVAQEEAPQFERVDWKRDPLRK- <u>LYFYAFVLCIASATT</u> *: : : :
NCU00801	GYDGSLMTGIIAMDKFQNQFHTGDTGPKVSVIFSLYTGVAMVGAPFAAILSDREFGRKKGM
NCU008114	GYDGMFFNSVQNFEETWIKYFGDP-RGSELGLLGALYQIGSIGSIPFVPLLTDNFGRKTP
LAC2	GYDGSSLSSLITMPPEFISHLNIK-SASGTGIVFAIFQVGQMVALTFV-WLGDFIGRRNAI
C_parasiti	GYDGLIGSITAMKPFMDTFNSELTGTVSISSLSVGTIVTSPILAAMLSDRGRRNWSM
A_clavatus	GYDGLSMGAILAMDHFQNTFHTGMDGPKVSLVTSLYTVGSIAATPFSAVLSDKLGRRKCM
A_nidulans	GYDGLSMNGLEALDDWKRSYNHP-EGATLGLLAASMSIGSILAIPVVPYVADLFGRREGV
S_thermo_l	GYDGLSMTSIIAMPHFQQTFDVGKVGTGAAVVFSLYVVGAMVGSPFAATLSDKFGRRKSM
S_thermo_i	GYDGMFFNSVQNFEETWENYFNHP-TGSKLGVLGALYQIGSLASIPIINVPIIADRVGRKTP **** : : : : : * : . . : * . ** : :
NCU00801	FIGGIFIIVGSIIVASSSKLAQFVVGFRVLGLGIAIMTVAAPAYSIEIAPHWRGRCTGF
NCU008114	IIIGCVIMIVGAVLQATAKNLDTFMGGRTMLGFGNSLAQIASPMILLTELAHPQHRAARLT
LAC2	FIGSVIVCLGAIITSIANNSTFIGGRFLLSFGSGISCALSTTYLLEITSPDERSALCAI
C_parasiti	FIGGWVIIVGSVIACSSSTVAQITVGRFILGGGISMIMTVAAPAYSIEIAPHWRGRCTGF
A_clavatus	FVGAWVIIAGSIIIAIAKHLQFYVGRVVLGFGIQIMVVSAPAYAVEIAPHWRGRAVGF
A_nidulans	VLGCMIMIGGVAMVSIGYKIALFVVGRGIILGFGLGIAQECSPLLVTELVHPQHRAIYSTI
S_thermo_l	FAGGLTIIIGMILVSTAHHLPQFVVGRFVLGLGIAIMTVAAPAYAIEVAPHWRGRCTGF
S_thermo_i	AIGCVIMIVGAVLQAACRNLGTFMGGRFLLGFGNSLAQLCSPMILLTELAHPQHGRGLTTV * : * : . : * : * : . : * : . : * : . : * : . : * : . : * : .
NCU00801	YNCGWFGGSIPAACYGCY--FIK-SNWSWRIPLILOAF-TCLIVM-SSVFFLPESPRF
NCU008114	YNCLWNVNGALVVSWLAFGTN--YIN-NDWSWRIPALLQAF-PSIIQL-LGIWWV <u>PESPRF</u>
LAC2	YNSLYYIGSIIATWSSYATSISYAN-SVLSFRIPLWLQILCPALVVI G VAPESPRF

Fig. 28 (a) (cont.)

Fig. 28 (a) (cont.)

NCU00801	FINQFAGPIALHNI-----GYKYIFVFVGWDLIETVAWYFFGVESQGRTEQLEWVYDQPN
NCU008114	TLN TYANPVAFDYFGPDHSWKLYLIYT C WIAAEFVFVFFMYVETKGPTLEELAKVIDGDE
LAC2	FINTYAAPVAMQNI-----KYWYYVFFFVWDTFEVIIYLYFVETKNLTLEEIELIFESAT
C_parasiti	FINQFAGPIALANI-----NTNYVYVFVAWDVIESILWWIFGVESQGRTEQLEWVYNQPK
A_clavatus	FVSQYASPIGLRNI-----STHYFWIFVGWDLFEALCWYLFGVESQGRTEELEWVYQQPN
A_nidulans	LFNQYVNPIALLDI-----GWKYYIVYCVWLLFELFVWWKYYVETKNTPLEEIVKFFDGDQ
S_thermo_l	FINQFAGPIALERI-----GYKYIYVFVAWCIEALAWYLFGVESQGRTEQLEWVYNQPN
S_thermo_i	TLNTYANPVAFDAGE-GHSWKLYIIYTIWIFLELCFWWKMYIETKGPTLEELAKIIDGDE ... : *::: : . : * * : :*:.. .::*:: . : .
NCU00801	PVKASLK V-----EKVVVQA-----DGHVSEA-IV-----A
NCU008114	ADVAHIDI-----HQVEKEV-----EIHEHEGKSV-----A
LAC2	PVKTSMII-----SKPGHAANEEKL-RLANLKL GKN-YVA-----
C_parasiti	PVQASIKV-----DKVVVQA-----DGKVTEK-IVADAAS
A_clavatus	PVKASLQV-----DKVVVQA-----DGQVTEK-IT-D--A
A_nidulans	AVLGGAAATEKIH E LVT--VQARS AEE TIGEKGPA-VSAEAR-
S_thermo_l	PVKASLK V-----DKVILTD-----DGKVAEK-IV-----A
S_thermo_i	AAVAHVDI-----KQV---EK-----ETHINEE-KS-----V .

Fig.28 (b)

Fig. 28 (c)

BLASTP 2.2.24+

Reference: Stephen F. Altschul, Thomas L. Madden, Alejandro A. Schaffer, Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman (1997), "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs", Nucleic Acids Res. 25:3389-3402.

Reference for compositional score matrix adjustment: Stephen F. Altschul, John C. Wootton, E. Michael Gertz, Richa Agarwala, Aleksandr Morgulis, Alejandro A. Schaffer, and Yi-Kuo Yu (2005) "Protein database searches using compositionally adjusted substitution matrices", FEBS J. 272:5101-5109.

RID: 4HJKWKG311R

Query= NCU00801
Length=579

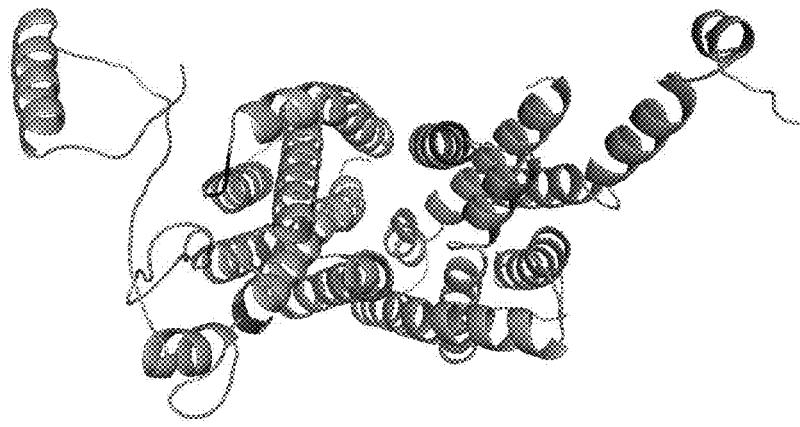
		Score
E	Sequences producing significant alignments:	(Bits)
Value		
1c1 47823	NCU08114	238
4e-67		
ALIGNMENTS		
>1c1 47823	NCU08114	
Length=525		
Score = 238 bits (608), Expect = 4e-67, Method: Compositional matrix adjust.		
Identities = 152/518 (29%), Positives = 257/518 (49%), Gaps = 35/518 (6%)		
Query 66	WSKES--IHLYFAIFVAFCCACANGYDGSLMTGIIAMDKFQNQFHTGD-TGPKVSIFSL	122
Sbjct 30	W K+ LYF F+ + GYDG + + + F GD G ++ ++ +L	87
	WKKDPGLRKLYFYAFILCIASATTGYDGMFFNSVQNFETWIKYF--GDPRGSELGLLGAL	
Query 123	YTVGAMVGAPFAILSDRGRKKGMFIGGIFIIVGSIIVASSSKLAQFVVGRFVLGLGIA	182
Sbjct 88	Y +G++ PF +L+D FGRK + IG + +IVG+++ A++ L F+ GR +LG G +	147
	YQIGSIGSIPFPVPLLTNDNFRKTPPIIGCVIMIVGAVLQATAKNLDTFMGGRTMLFGNS	
Query 183	IMTVAAPAYSIEIAPPHWGRRCTGFYNCGWFGGSIPAACYFIKSNSWWRIPLILQ	242
Sbjct 148	+ +A+P E+A P R R T YNC W G++ + + +G +I ++WSWRIP +LQ	207
	LAQIASPMLTEAHQPHARLTTIYNCLWNV GALVVSWLAFGTNYINNDWSWRIPALLQ	
Query 243	AFTCLIVMSSVFLPESPRFLFANGRDAEAVFLVKYHGNQDPNSKLVLLETEEMRDGIR	302
Sbjct 208	AF +I + +++++PESPRFL A + EA+ L KYH NGDPN V E E+++ IR	267
	AFPSIIQLLGIVWWVPESPRFLIAKDKHDEALHILAKYHANGDPNHPTVQFEFREIKETIR	
Query 303	TDGVDKVWWDYRPLFMTHSGRWRMAQVLMISIFGQFSGNG-LGYFNTVIFKNIGVTSTSQ	361
Sbjct 268	+ Y F + R+R+A +L + F Q+SGN + +++ +++ GVT ++	327
	LEMESTKNSSYLDFFKSRGNRYRLAILLSLGFSSQWSGNAIISNYSSKLYETAGVTDSTA	
Query 362	QLAYNILNSVISAIGALTAVSMTDRMPRRAVLIIGTFMCAAALATNSGLSATLDKQTQRG	421
Sbjct 328	+L + + + I ++T + D++ RR A LA+ G+ T T	374
	KLGLSAGQTGLALIVSVTMAILVDKLGR-----LAFLASTGGMCFTVIWT---	

Fig. 28 (c) (cont.)

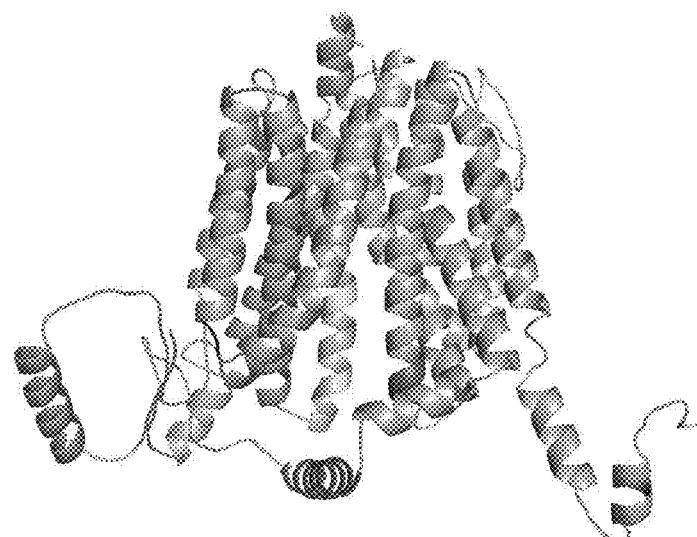
Query	422	TQINLNQGMNEQDAKDNEYLHVDSNYAKGALAAAYFLFNVI	FSFTYTPLQGVIPTEA	LETT	481	
		L G+ + H K + ++F + +S ++ L	E L			
Sbjct	375	-----LTAGLYGE-----	HRLKGADKAMIFFIWFGIFYSLAWSGLLVGYAIEILPYR		422	
Query	482	IRGKGLALSGFIVNAMGF	INQFAGPIALHNIG	---YKYIFVFVGVWDLIETVA	WYFFGVE	537
		+RGKGL + V +N +A P+A G +K ++ W E V +F VE				
Sbjct	423	LRGKGLMVNMMSVQCALT	LNTYANPVAFDYFGPDHSWKLYLIYTCWIAAEFVFVFFFMYVE		482	
Query	538	SQGR	TLEWVYDQP	NPVKASLKVEKVVVQADGHVSE	575	
		++G TLE+L V D A + + +V + + H E				
Sbjct	483	TKGPT	LEELAKVIDGDEADVAHIDIHQVEKEVEIHEHE		520	

Fig. 29

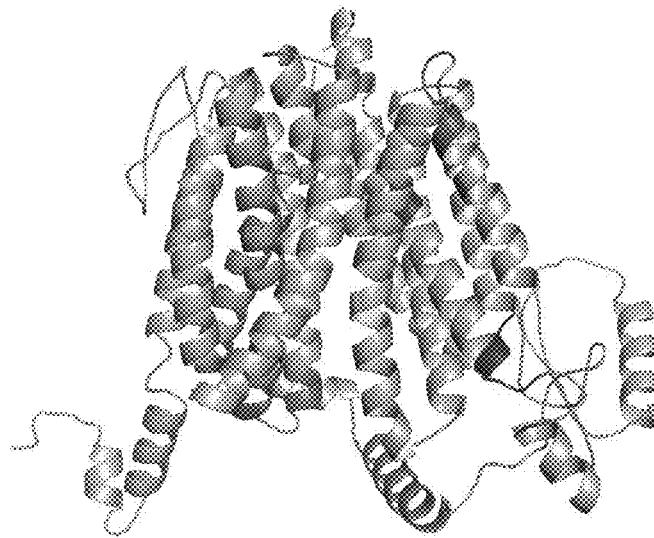
(a)



I.



II.

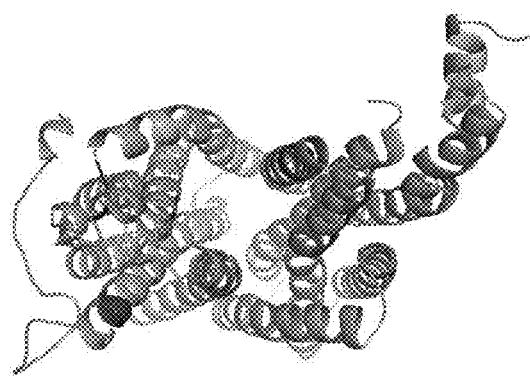


III.

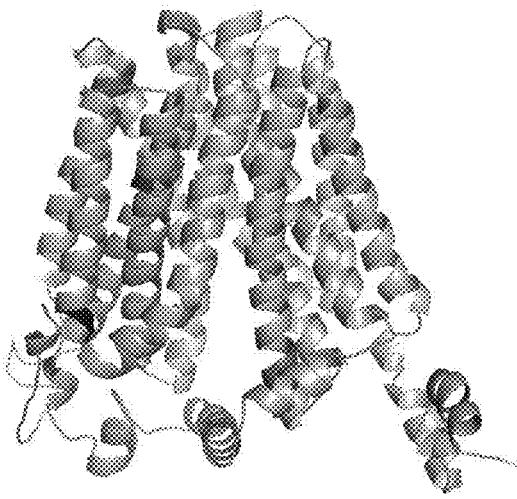
Fig. 29 (cont.)

(b)

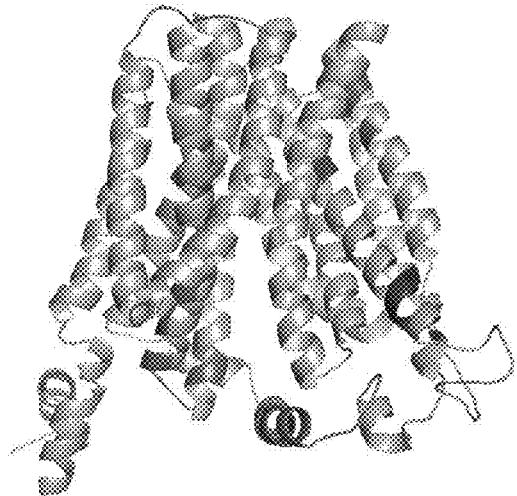
I.



II.



III.



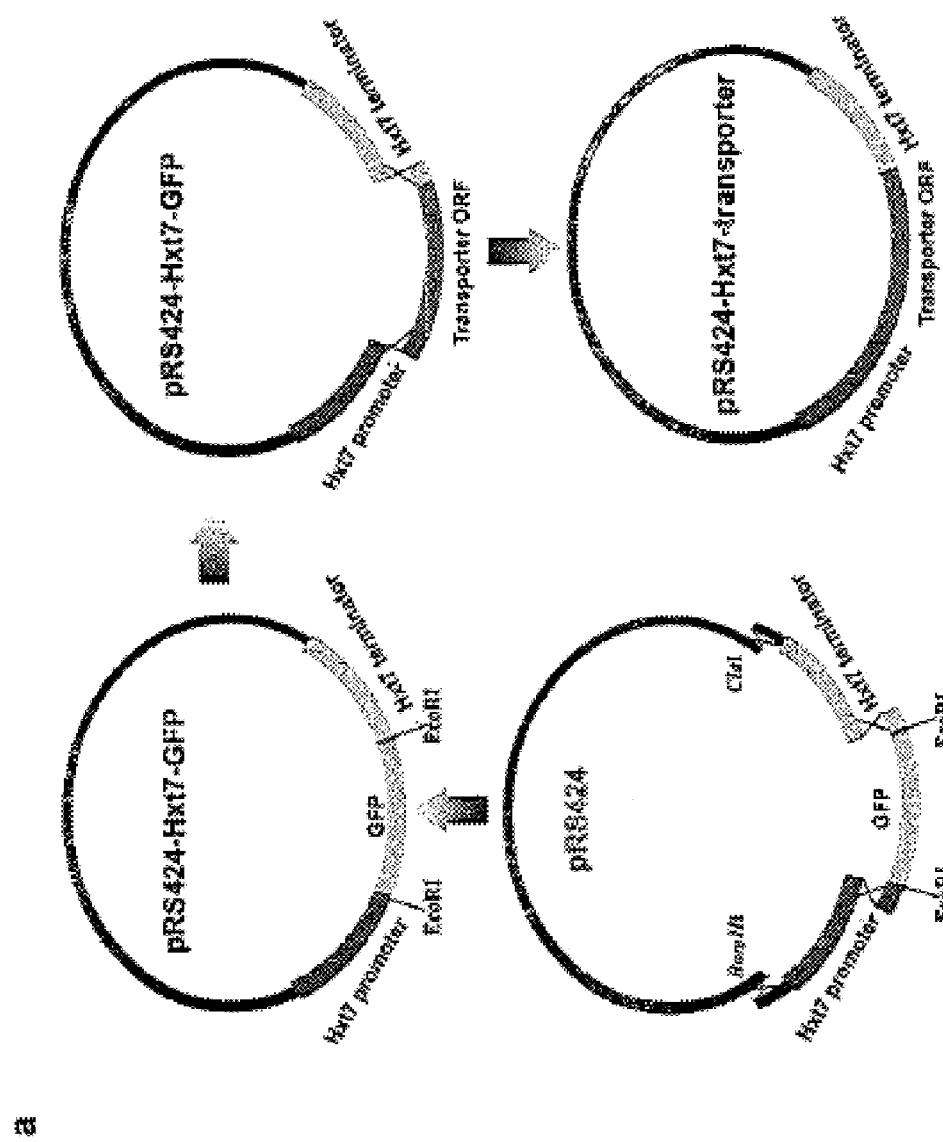


Figure 30

Figure 30 (cont.)

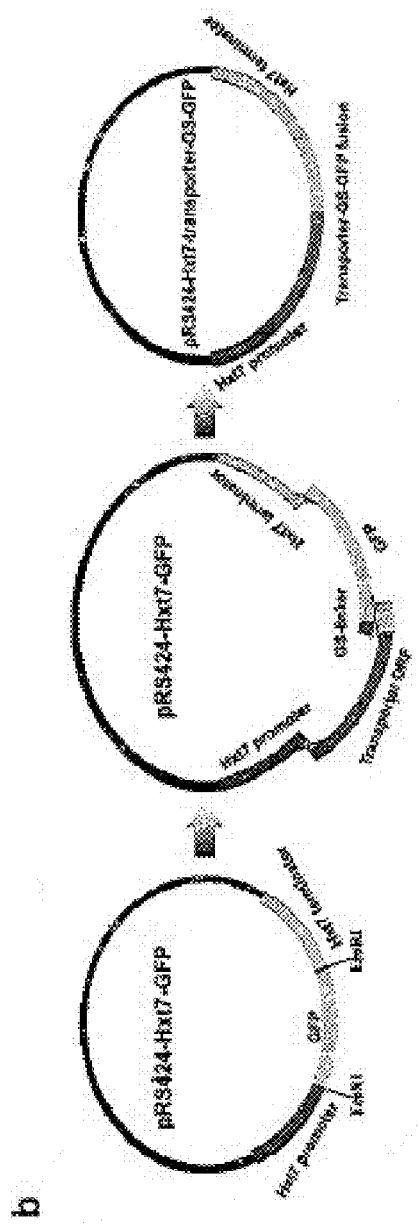


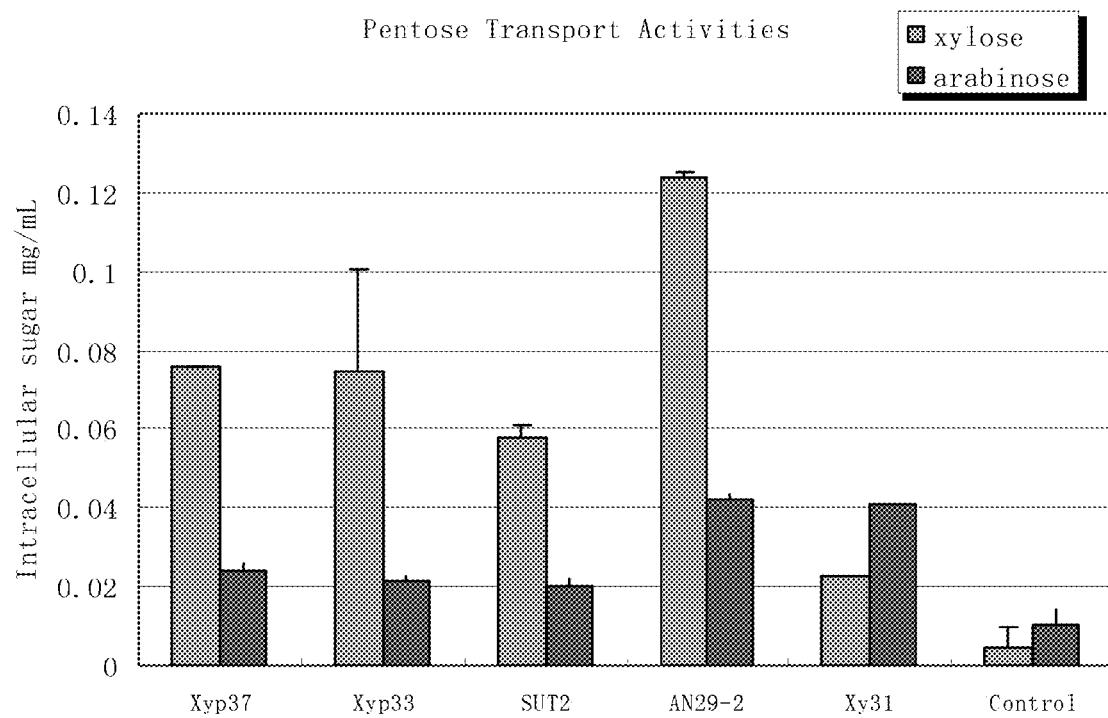
Figure 31

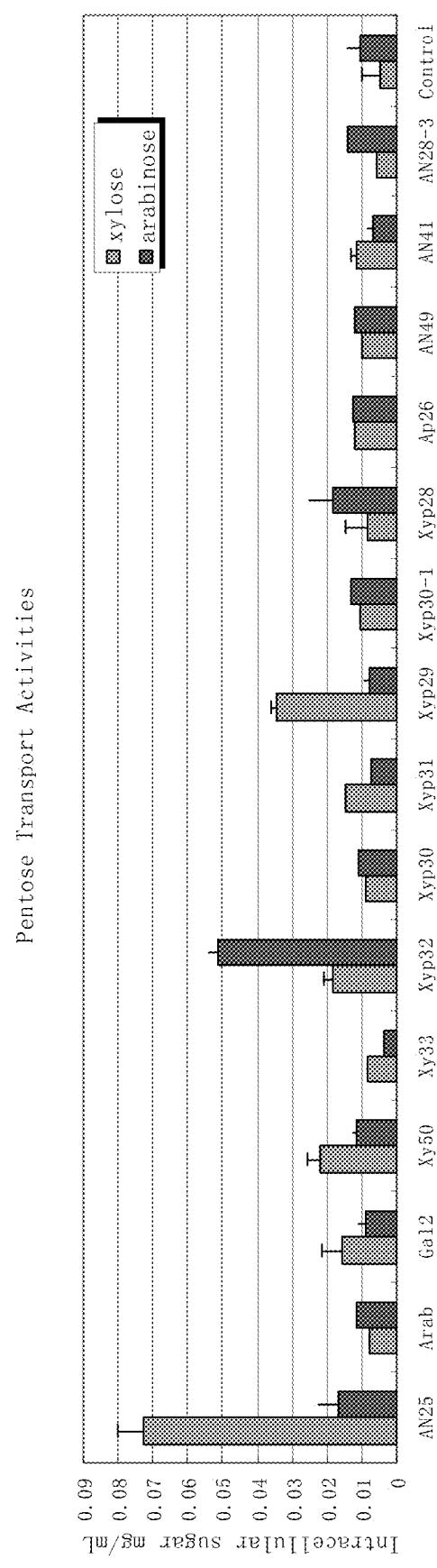
Figure 32

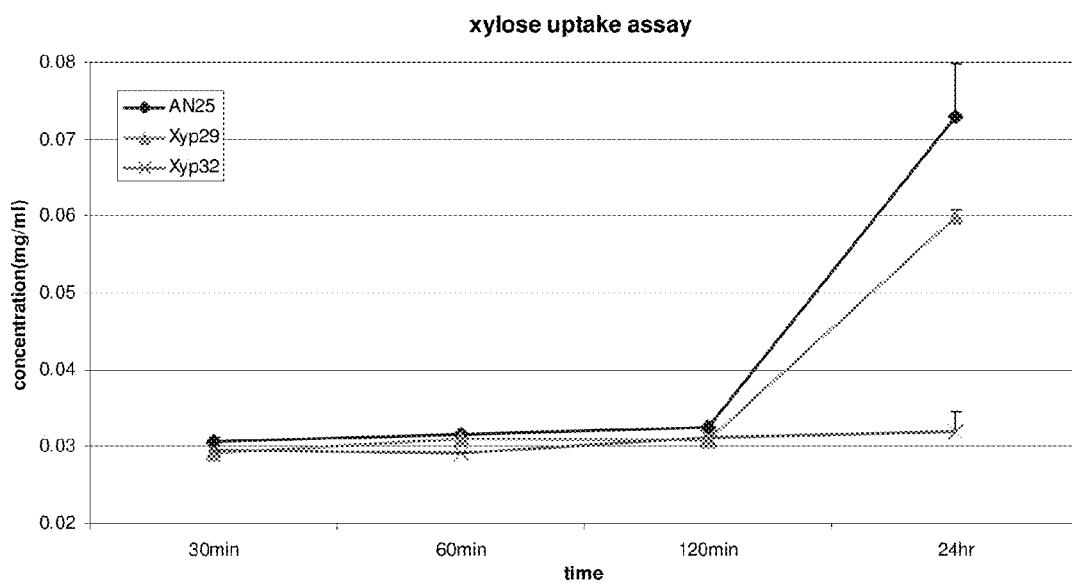
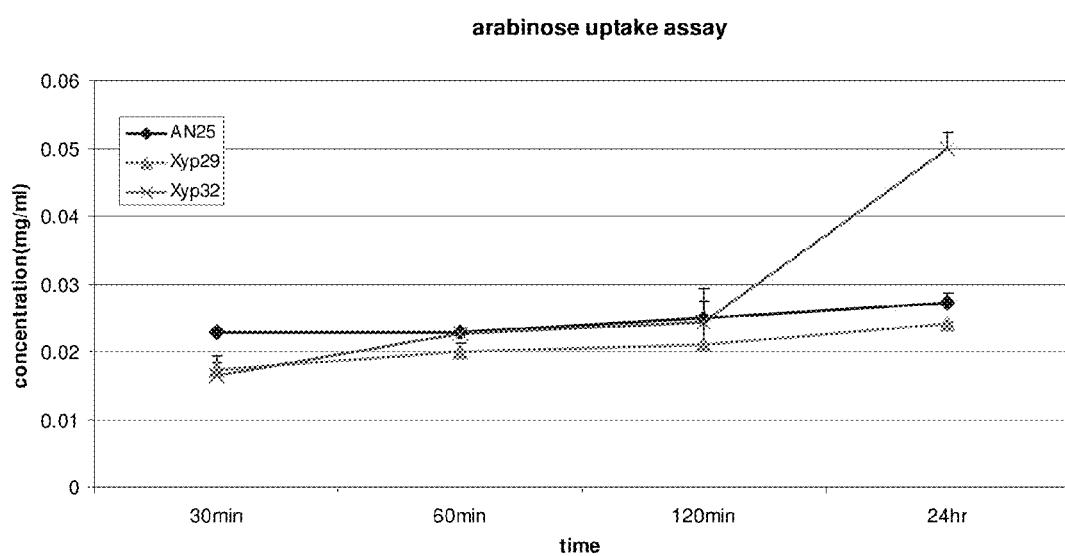
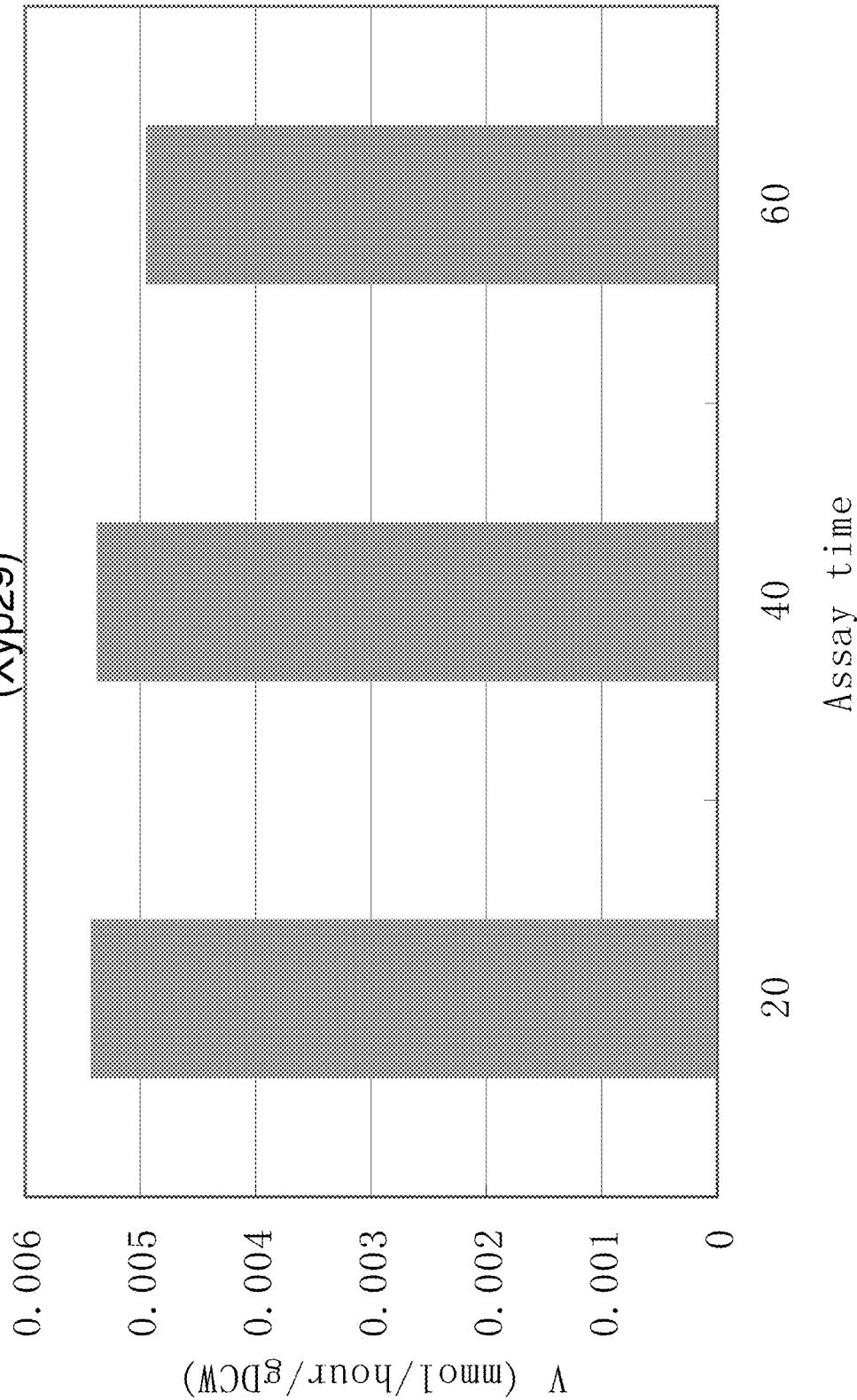
Figure 33**A****B**

Figure 34STL12/XUT6
(Xyp29)

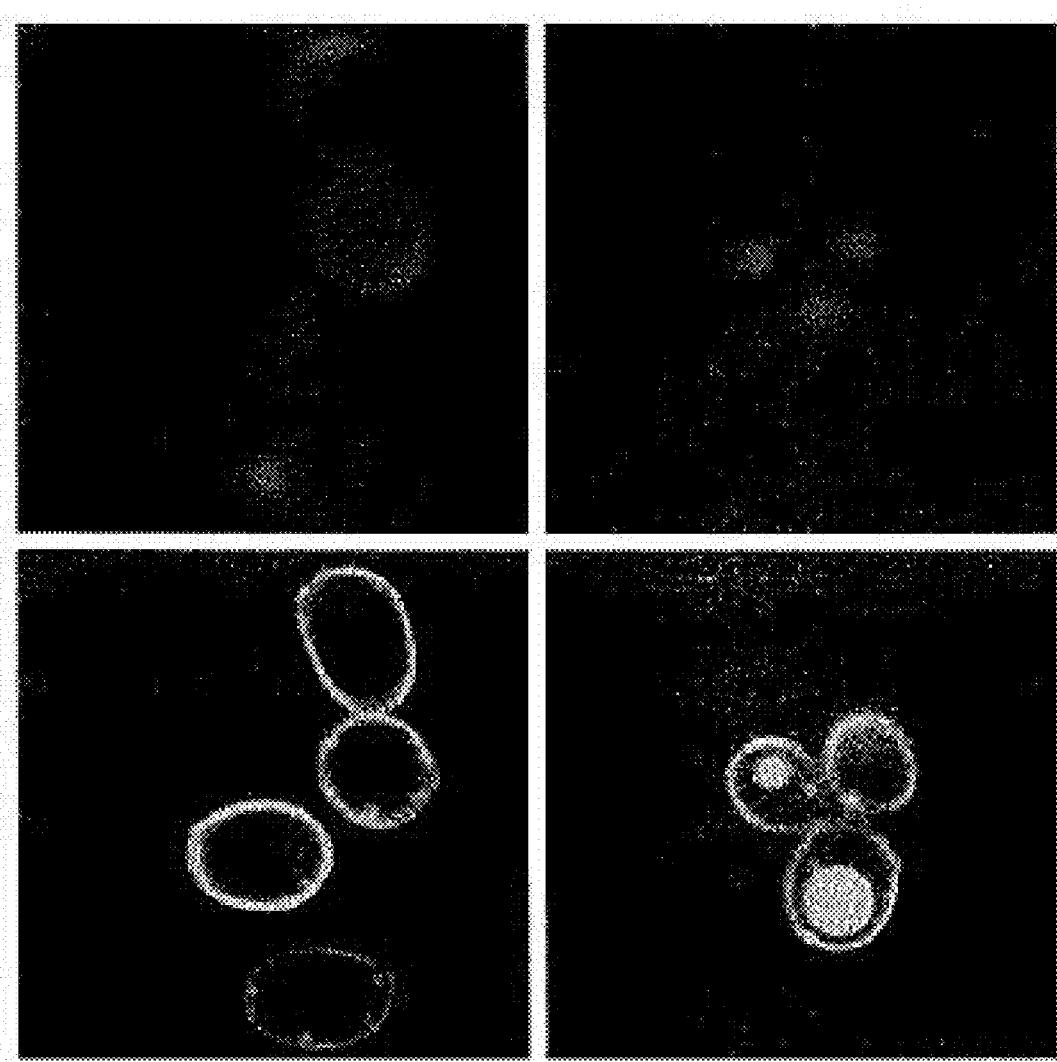


Figure 35

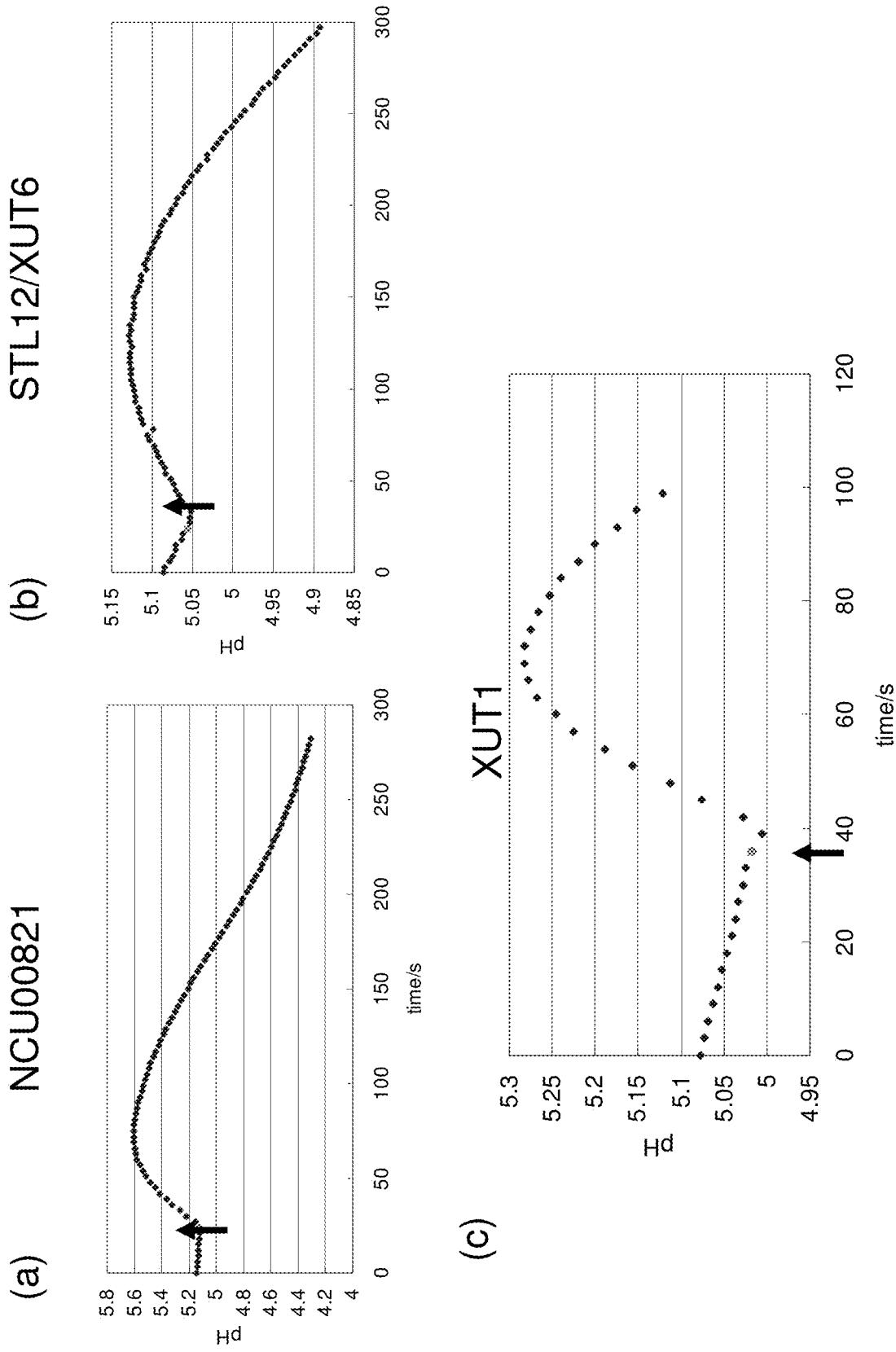
Figure 36

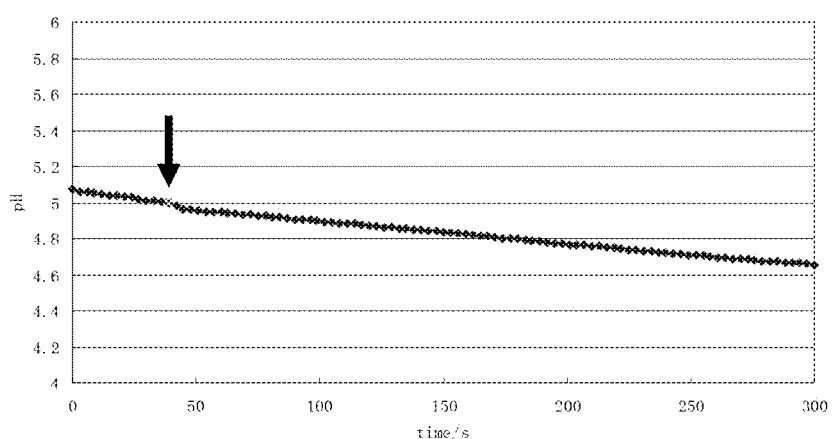
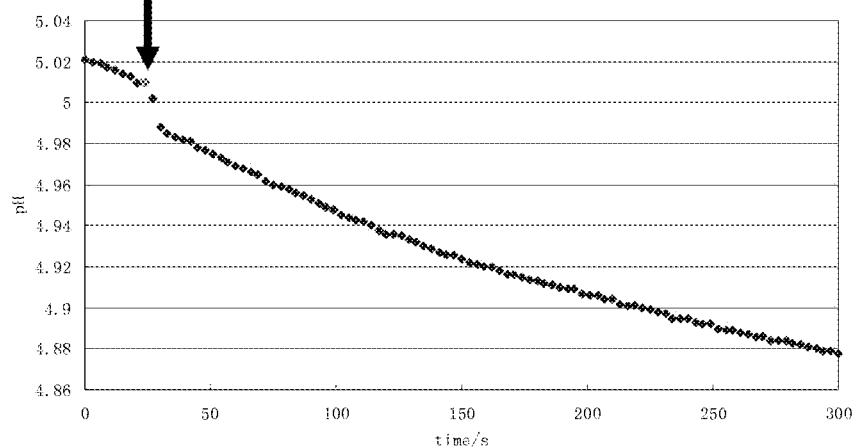
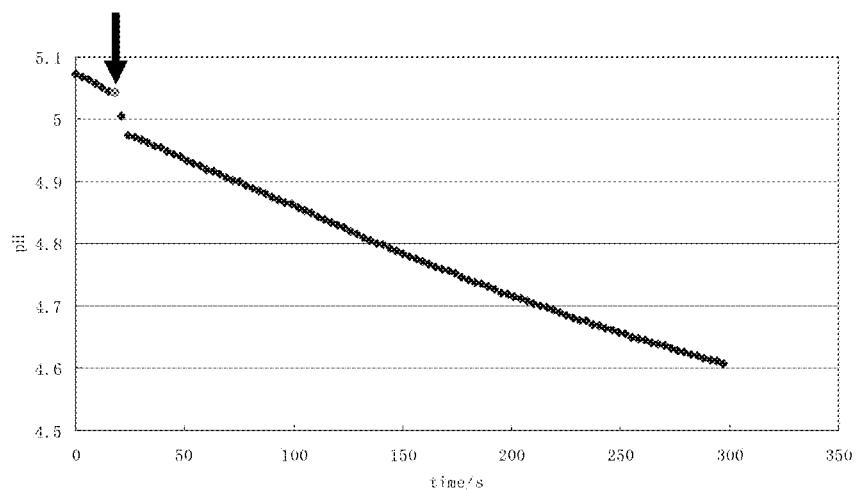
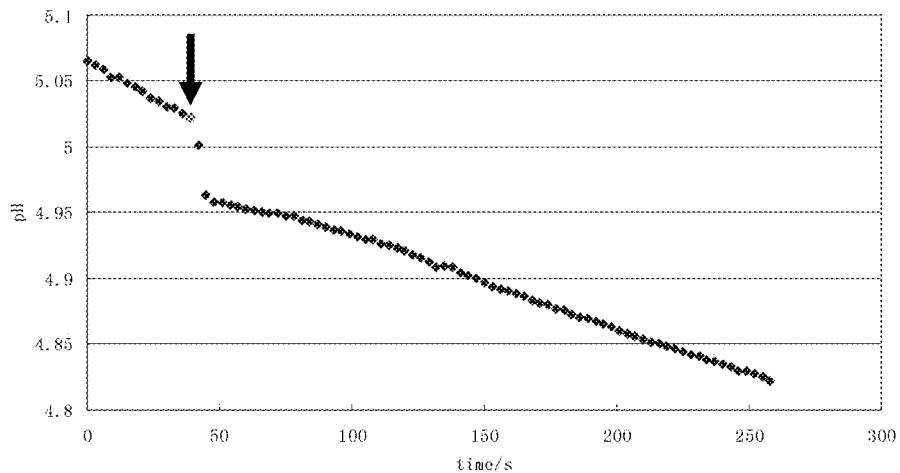
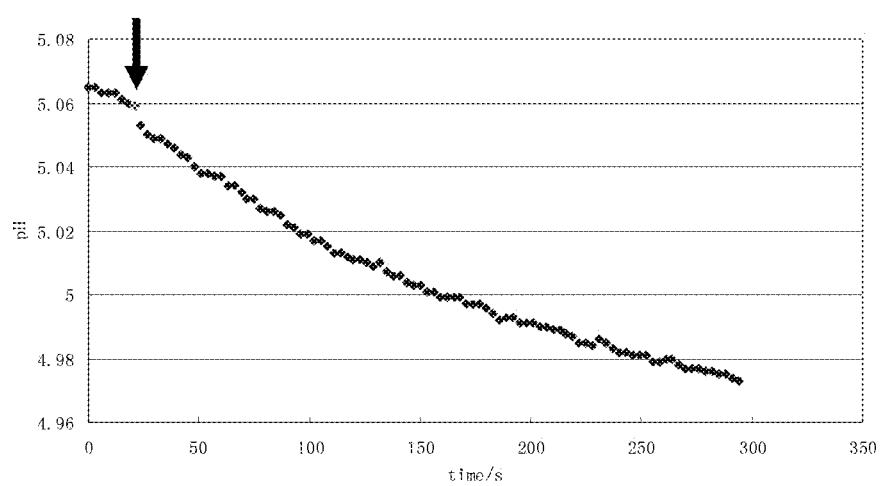
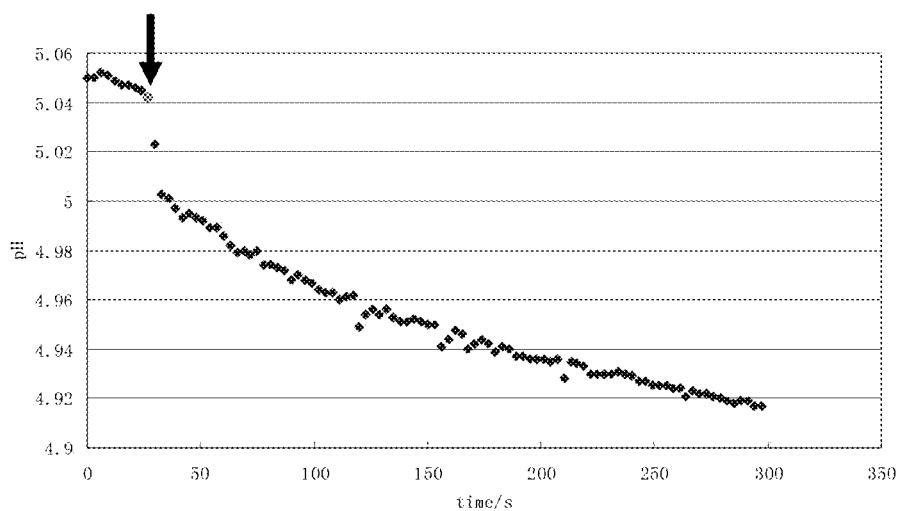
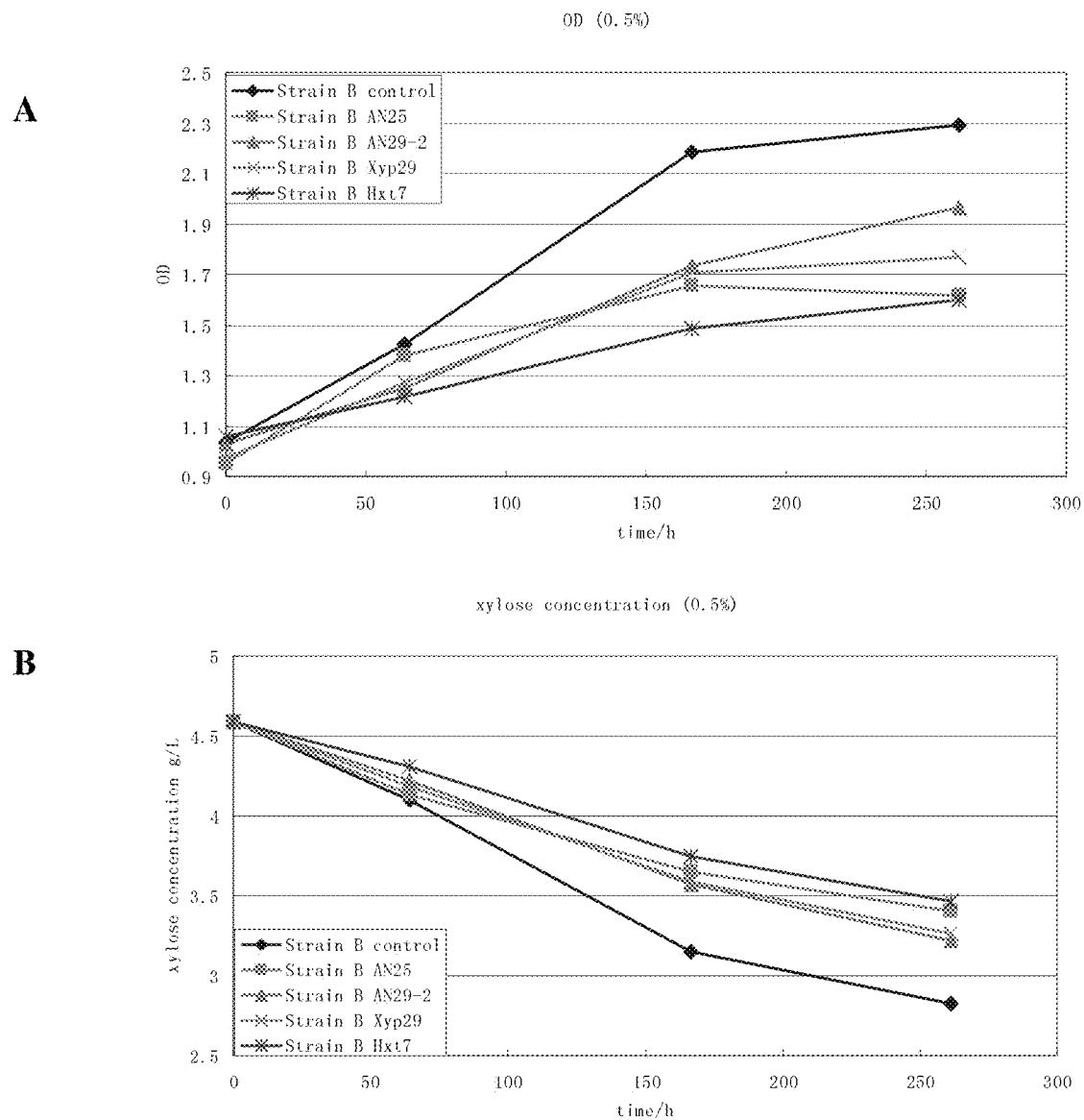
Figure 37**A****B****C**

Figure 37 (cont.)**D****E****F**

**Figure 38**

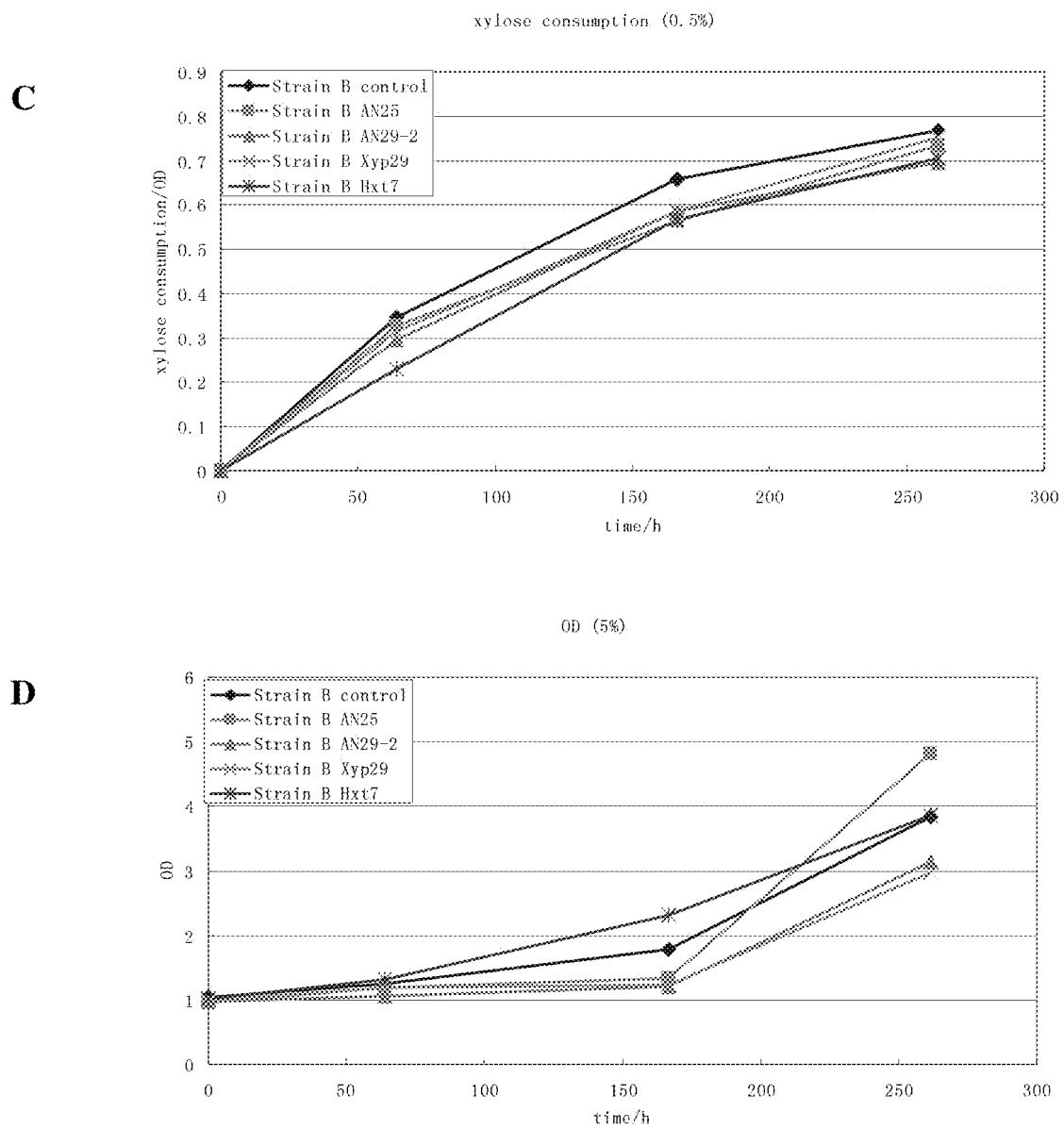


Figure 38 (cont.)

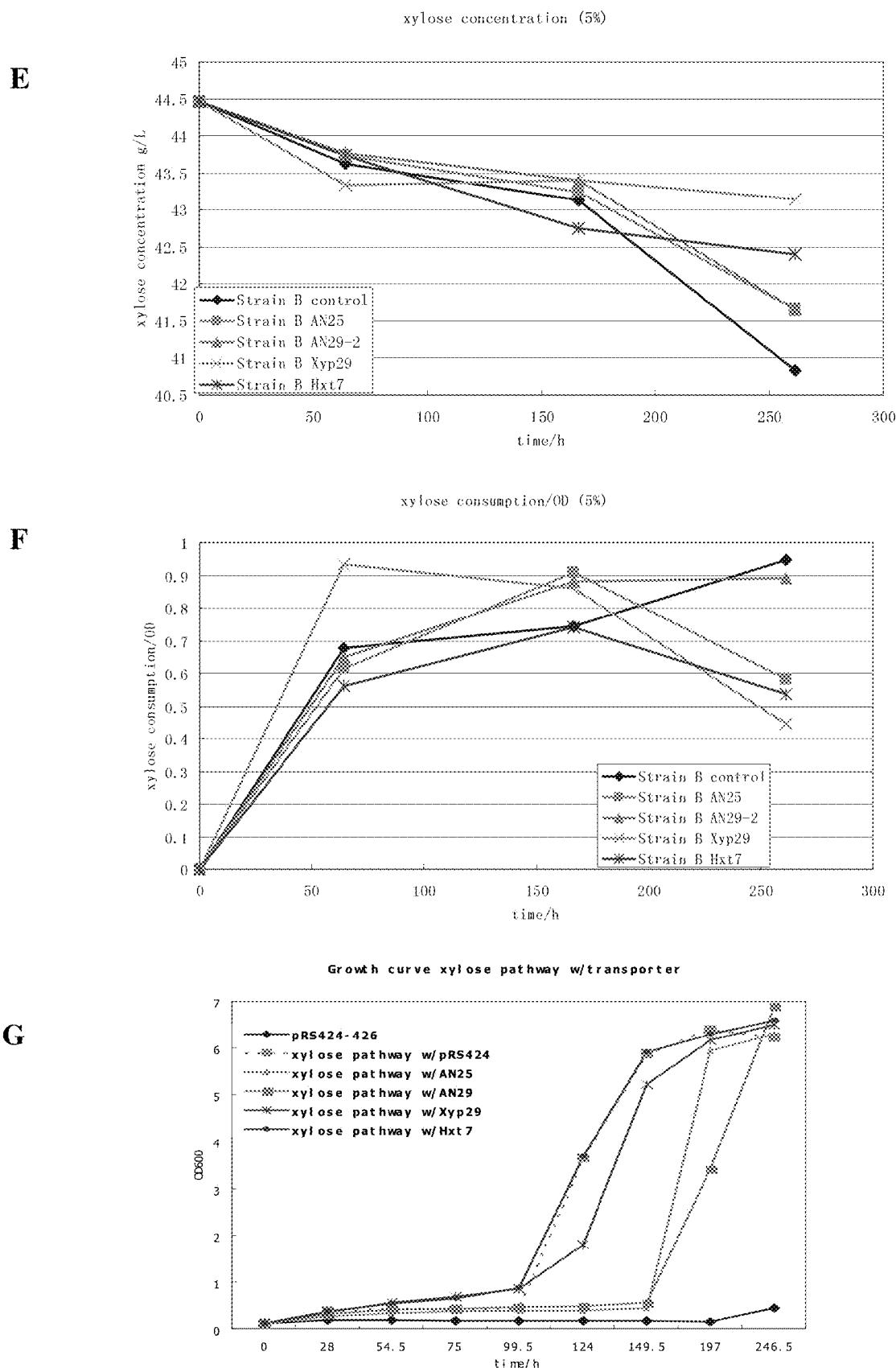


Figure 38 (cont.)

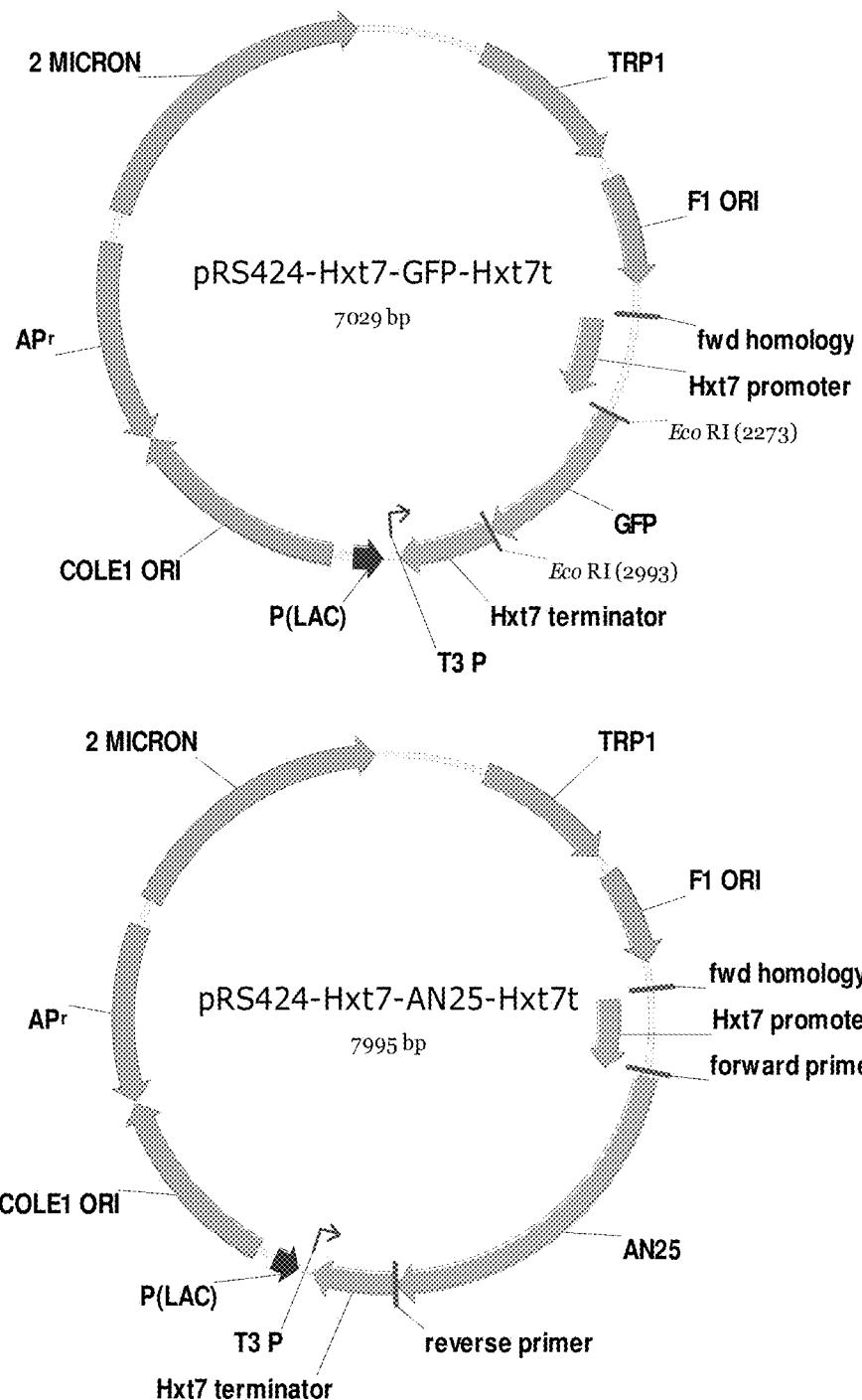
Figure 39

Figure 39 (cont.)

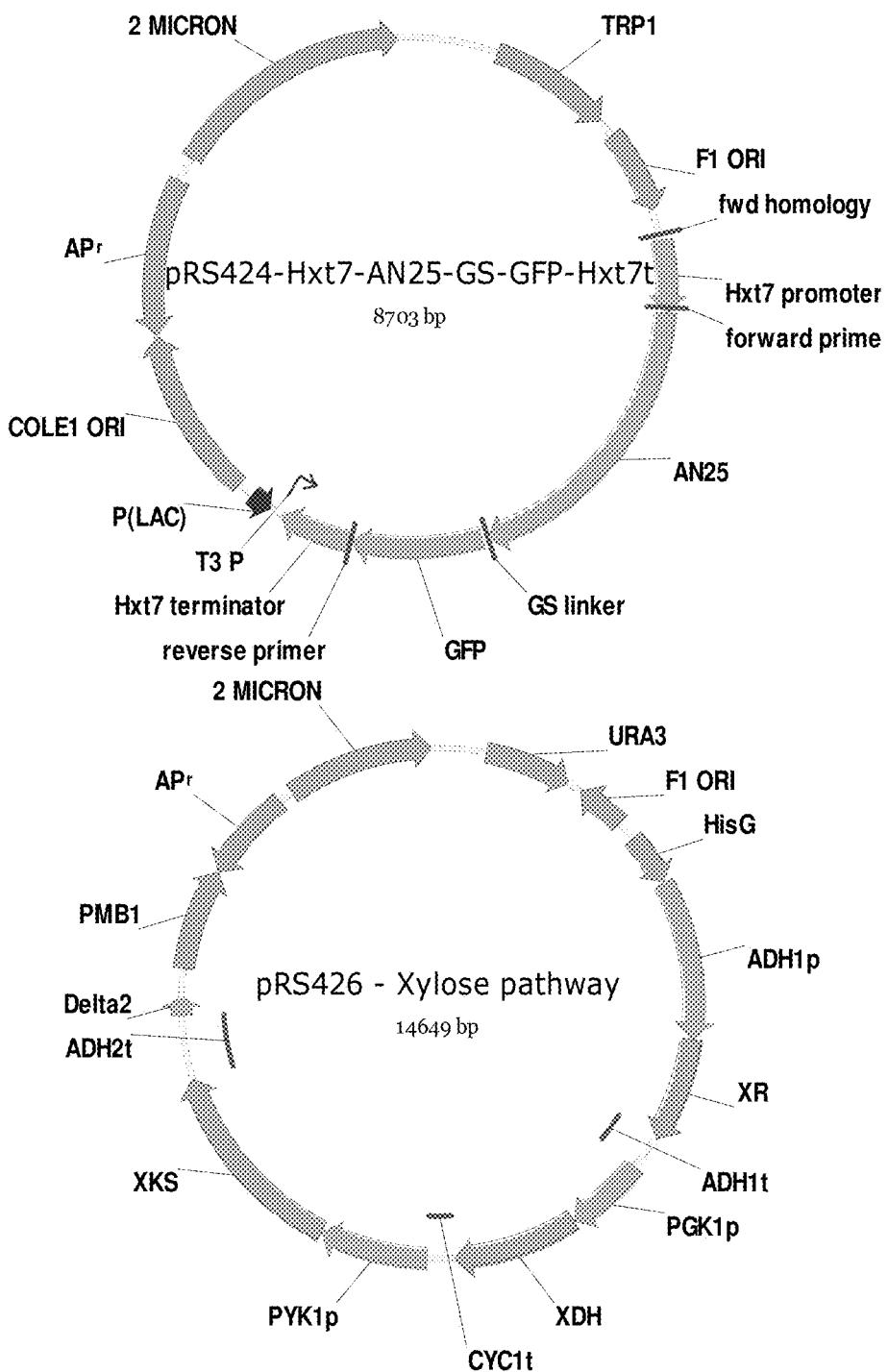


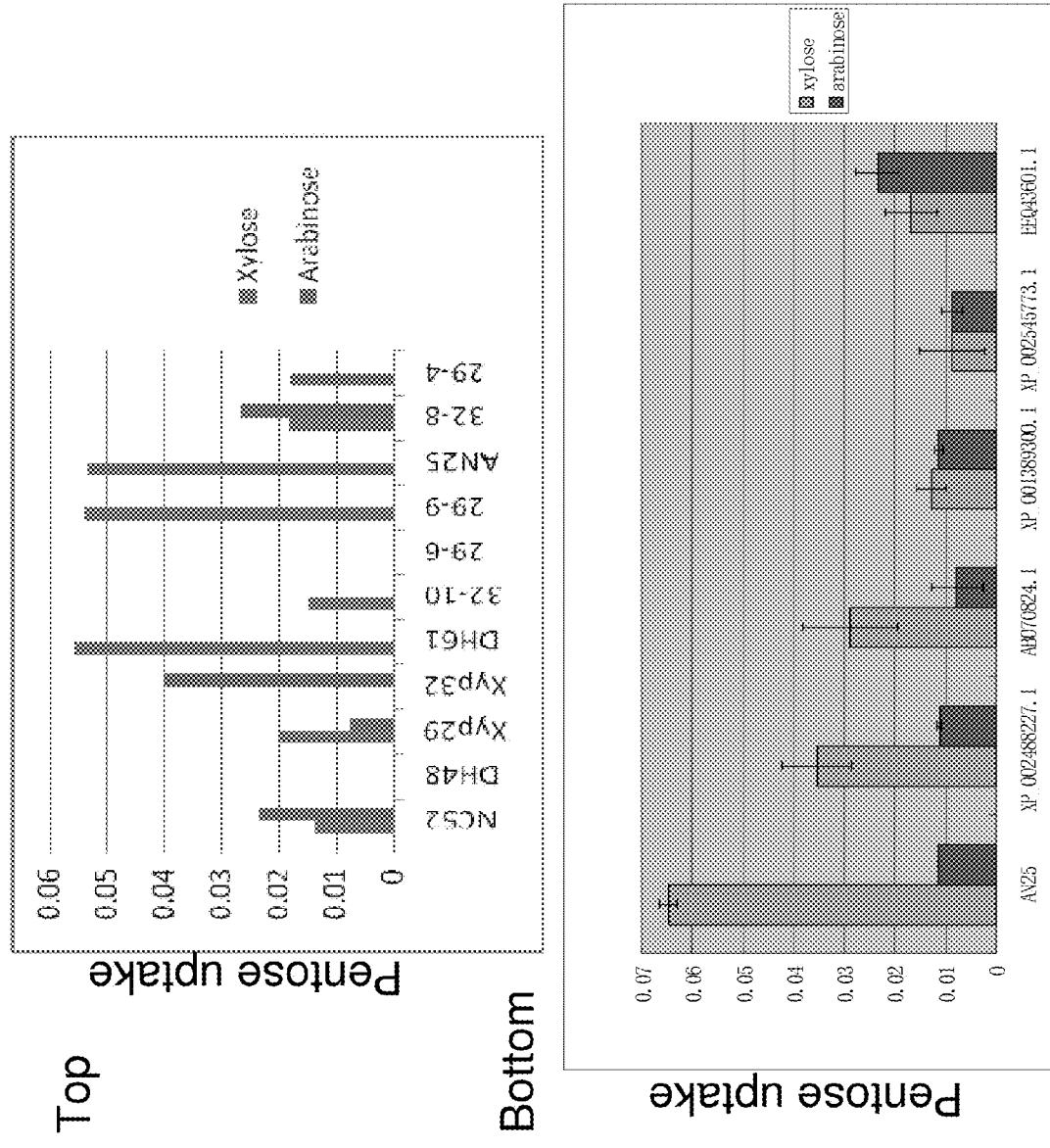
Figure 40

Figure 41

Consensus key

- * - single, fully conserved residue
 : - conservation of strong groups
 . - conservation of weak groups
 - no consensus

CLUSTAL W (1.81) multiple sequence alignment

(a)

NCU00821	MAPPKFLGLSGRPLSLAVSTVATTGFLFGYDQGVMM
XP_002488227.1	MAVPQFAGMSGRKLSWSISTIATLGFLFGYDQGVMM
XP_002382573.1	MAPTFAGLSGRPLSLAVST-----
XUT6	MEFSSVEKSAETASYTSQVSASGSAKTNSYLGLRGHKLNFAVSCFAGVGFLFGYDQGVMM

; *; *; *; ::*

NCU00821	SGIITAPAFNNFTPTKD--NSTMQGLITAIYEIGCLIGAMFVLWTGDLLGRRRNIMVGA
XP_002488227.1	SGIISAKPFNTVFKATED--DSTMQGTVTAIYEIGCLFGAMFILWFGDWLGRRKSVMLGA
XP_002382573.1	-----
XUT6	GSLLTLPSENTFPAMKASNNATLQGAVIALYEIGCMSSSLATIYLGDRRLGRLKIMFIGC

NCU00821	FIMALGVI IQVTCQAGSNPPAQLFVGRVVMGIGNGMTSTIPTYQAEC SKTSNRGLLICI
XP_002488227.1	AIMILGVVI QVTAYTGHVPLAQFIVGRVVTGVGNINTSTIPTYQAEC SRITNRGLLICI
XP_002382573.1	-----AECSKTSNRGLLICI
XUT6	VIVCIGAALQASAFT----IAHLTVA RIITGLGTGFITSTVPVYQSECSPAKKRGQL IMM

;*** ; ; *** ***

NCU00821	EGGVIAFGTLIAYWIDYGASYGPDD-----LVWRFPIAFQLLFAIPICVPMFYLPESPRW
XP_002488227.1	EGGVIAFGTLIAYWIDFGASYGPDD-----LTWRFPIAFQIIVPGLGIIAGMFELPESPRW
XP_002382573.1	EGGVIAIGTAAIAYWIDFGAHYGPDD-----LVWRFPIAFQIVPGVIIIVGMFFLPDSPRY
XUT6	EGSLIALGIAISYWIDFGFYFLRNDGLHSSASWRAPIALQCVPAVLLISTVFFFFPESPRW

Figure 41 (cont.)

Figure 41 (cont.)**(b)**

EEQ43601. 1	MSKGDLLELDIQLIVEKRLEQSNGSGFATMKRNKRALGTCLFVSLGGILYGYNQGMFGQ
XUT1	-MHGGGDGNDITEIIAARRLQIAGKSGVAGLVANSRSFFIAVFASLGLVYGYNQGMFGQ ;*: ; ** ; : *; : ***; : , **; * ; *; *; : . ; *; ***; ;*****;
EEQ43601. 1	VSSMHSFGETVGIGKIQDNPTLQGLLTSILELGAWVGVLMGYVADALGRRAVVIGCIL
XUT1	ISGMYSFSKAIGVEKIQDNPTLQGLLTSILELGAWVGVLMGYIADRLGRKKSVVVGVFF ;*; *; **; ; ; *; *****; *****; *****; ***; ***; ; ***; * ; :
EEQ43601. 1	FNLGVLIQAVARDADYGYLGGRFVIGLGVGVLSMVVPLYNSEISRAEIRGANTAIYQLS
XUT1	FFIGVIVQAVARGGNYDYILGGRFVVGIGVGILSMVVPLYNAEIISPPEIRGSLVALQQLA * ***; ***; , ; *; ***; ***; ; *; ***; ***; ; ***; , ***; , *; ; **;
EEQ43601. 1	ITFGIMISYWITYGTNFIGGTGDNQSQASWLVPMCIQAAPAIIIAVFIYSFPESPRWLN
XUT1	ITFGIMISYWITYGTNYIGGTGSGQSKASWLVPICIQLVPAALLGVGIFMPESPRWLMN *****; *****; ; ***; ; ***; ; ***; ; ***; ; ***; ; ***; ; ***; ; ***; ; ***;
EEQ43601. 1	VGQEDKALEVLAWLRETEQENVGLQIEFLEMKAQKIEPQTLETAEAYPHLQDGTKMSKFKI
XUT1	EDREDECLSVLSNLRSLSKEDTLVQMEFLEMKAQKLERELSAKYFPHLQDGSAKSNFLI . ; ***; , *; *; ; **; . ; *; . ; *; ; ***; ; ***; ; ***; ; ***; ; ***; ; ***;
EEQ43601. 1	NLNQYKSMVTHLPTFKRVSVACLTMVFQQWTGAYNFIYYAPPFIFASLGLSGNTTSLLAS
XUT1	GPNQYKSMITHYPTFKRVAACLIMTRQQWTG-VNFILYYAPFIFSSLGLSGNTISLLAS . ; ***; ; ***; ; ***; ; ***; ; ***; ; ***; ; ***; ; ***; ; ***; ; ***; ; ***;
EEQ43601. 1	GVVGIVMFLCTIPAVMWVDKVRKPPLLISGALVMGLCHFVVAGLLGGYSDNIGSHKAAGW
XUT1	GVVGIVMFLATIPAVLWDRILGRKPVLISGAIIIMGICHFVVAAILGQPGGNFVNHSAGGW *****; ; ***; ; ***; ; ***; ; ***; ; ***; ; ***; ; ***; ; ***; ; ***; ; ***;
EEQ43601. 1	VAVVIWIFAGAFGYSWGPCAWVIVAEVPLGMRAKGVSLGSSPNWLNNPSVA1STPKFV
XUT1	VAVVFVWIFAIAGFGYSWGPCAWVLVAEVPLGLRAKGVSIGASSNWLNPAVAMSTPDFV *****; ; ***; ; ***; ; ***; ; ***; ; ***; ; ***; ; ***; ; ***; ; ***; ; ***;
EEQ43601. 1	ANAKYGAYIFLGLMCVIGSMYVYFMVPETKNKTLDELDEVFGDFTGTSKKESELREKILK
XUT1	AKAKFGAYIFLGLMCIFGAAYVQFFCPETKGRTLEEIDELFGDTSGTSKMEKEIHEQKLK *; ; ***; ; ***; ; ; *; ; **; ; ; ***; ; ; ***; ; ; ***; ; ; ***; ; ; ***; ; ; ***;

Figure 41 (cont.)

EEQ43601. 1 QVGLVDLLVGSDKELDSFRSKPEVEYKEKEAHSE
XUT1 EVGLLQLLG-EENASESENSKADYHVEK-----
;***; ;** . :; ;* , **. ;* ; **

Fig. 41 (cont.)**(c)**

NCU00821	-----MAPPKFLGLSGRPLSLAVSTVATIGFLFGYDQGV
XP_002488227. 1	-----MAVPQFAGMSGRKLSWSISTIATLGFLFGYDQGV
XP_002382573. 1	-----MAPTFAGLSGRPLSLAVST-----
XUT6	-----MEFSSVEKSAETASYTSQVSASGSAKTNSYLGLRGHKLNFAVSCFAGVGFLFGYDQGV
EEQ43601. 1	---MSKGDLLELDIQKLIVEKRLEQSNGSGFATMKRNKRALGTCFLVSLGGILYGYNQGMF
XUT1	---MHGGGDNDITEIIAARRLQIAGKSGVAGLVANSRSFFIAVFASLGGLVYGYNQGMF
	: . .

NCU00821	SGIITAPAFNNFTPTKD--NSTMQGLITAIYEIGCLIGAMFVLWTGDLLGRRRNIMVGA
XP_002488227. 1	SGIISAKPFNTVFKATED--DSTMQGTVTAIYEIGCLFGAMFILWFGDWLGRRKSVMLGA
XP_002382573. 1	-----
XUT6	GSLLTLPSFENTFPAMKASNNTLQGAVIALYEIGCMSSSLATIYLGDRRLGRKIMFIGC
EEQ43601. 1	GQVSSMHSFGETVGIGKIQDNPTLQGLLTSILELGAWVGVLMLNGYVADALGRRASVVIGC
XUT1	GQISGMYSFSKAIGVEKIQDNPTLQGLLTSILELGAWVGVLMLNGYIADRLGRKKSVVGV

NCU00821	FIMALGVIIQVTCQAGSNPFAQLFVGRVVMGIGNGMNTSTIPTYQAECSKTSNRGLLICI
XP_002488227. 1	AIMILGVVIQVTAYTGHVPLAQFIVGRVVTGVGNINTSTIPTYQAECSSRTTNRGLLICI
XP_002382573. 1	-----AECSKTSNRGLLICI
XUT6	VIVCIGAALQASAFT---IAHLTVAARIITGLGTGFITSTPVYQSECSPAKKRGQLIMM
EEQ43601. 1	ILFNIGVIIQAVARDAD--YGYILGGRFVIGLGVGVLSMVPLYNSEISRAEIRGANTAI
XUT1	FFFFFIGVIVQAVARGGN--YDYILGGRFVVGIGVGILSMMVPLYNAEISPPEIRGSLVAL
	: * * . ** :

NCU00821	EGGVIAFGTLIAYWIDYGASYGPDD-----LVWRPIAFQLLFAIFICVPMFYLPESPR
XP_002488227. 1	EGGVIAFGTLIAYWIDFGASYGPDD-----LTWRPIAFQIVFGLGIAGMFELPESPR
XP_002382573. 1	EGGVIAITGTAIAYWIDFGAHYGPDD-----LVWRPIAFQIVFGVIIIVGMFFLPDSPR
XUT6	EGSLIALGIAISYWIDFGFYFLRNDGL-HSSASWRAPIALQCVFAVLLISTVFFFPESPR
EEQ43601. 1	YQLAITFGIMISYWITYGTNFIGGTGDNQSQASWLPVPMCIQAAPATIILAVFIYSPESPR
XUT1	QQLAITFGIMISYWITYGTNYIGGTGSGQSKASWLPFICIQLVPALLGVGIFFMPESPR
	* : * * ; *** : * : . * * : ; * : : : : ; * ; ***

NCU00821	WLLSHGRTQEADKVIALLRGYEIDGPETIQERNLIVDSLRA-----S
XP_002488227. 1	WLFMRERYQEGERAVIAALLNEETNSHHVQLQKTLVLDLSIRA-----S
XP_002382573. 1	YLIISKDRIQEGEYVLAALGGYEVHDQETQTQKNLVIDSIRAYVCCLEVLPNKSVTNLDS
XUT6	WLINKGRTEEAREVPSALYDLPADSEKISIQIEEQAAIDL-----E
EEQ43601. 1	WLINVGQEDKALEVLAWLRETEQENVGLQIEFLEMKAQKIFEQTLETEAYP-----H
XUT1	WLNNEDREDECLSVLSNLRSLSKEDTLVQMFLEMKAQKLFERELSAYFP-----H
	* : : : : * : * : .. : : :

Figure 41 (cont.)

Figure 41 (cont.)

NCU00821	VEIDVVDEHGVESGFGDGINTKETR
XP_002488227. 1	AETTGTPKQGEEH-----VSKMV-
XP_002382573. 1	SE-----
XUT6	-----
EEQ43601. 1	VEYKEKEAHSE-----
XUT1	VYHVEK-----

Figure 42

Strain backgrounds vs. xylose fermentation

- Construction of engineered strains using an identical construct and different laboratory strains

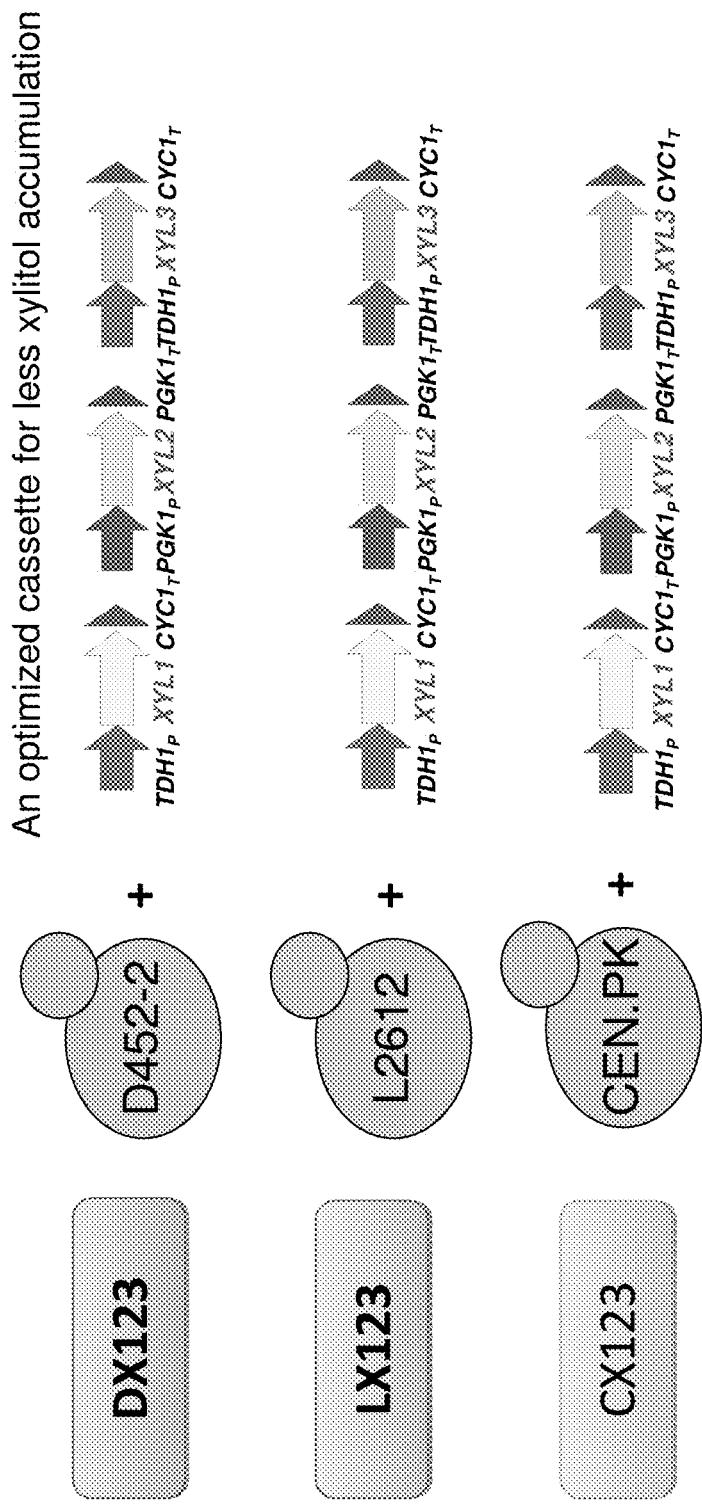


Figure 43
Identical expression cassette under different strain backgrounds

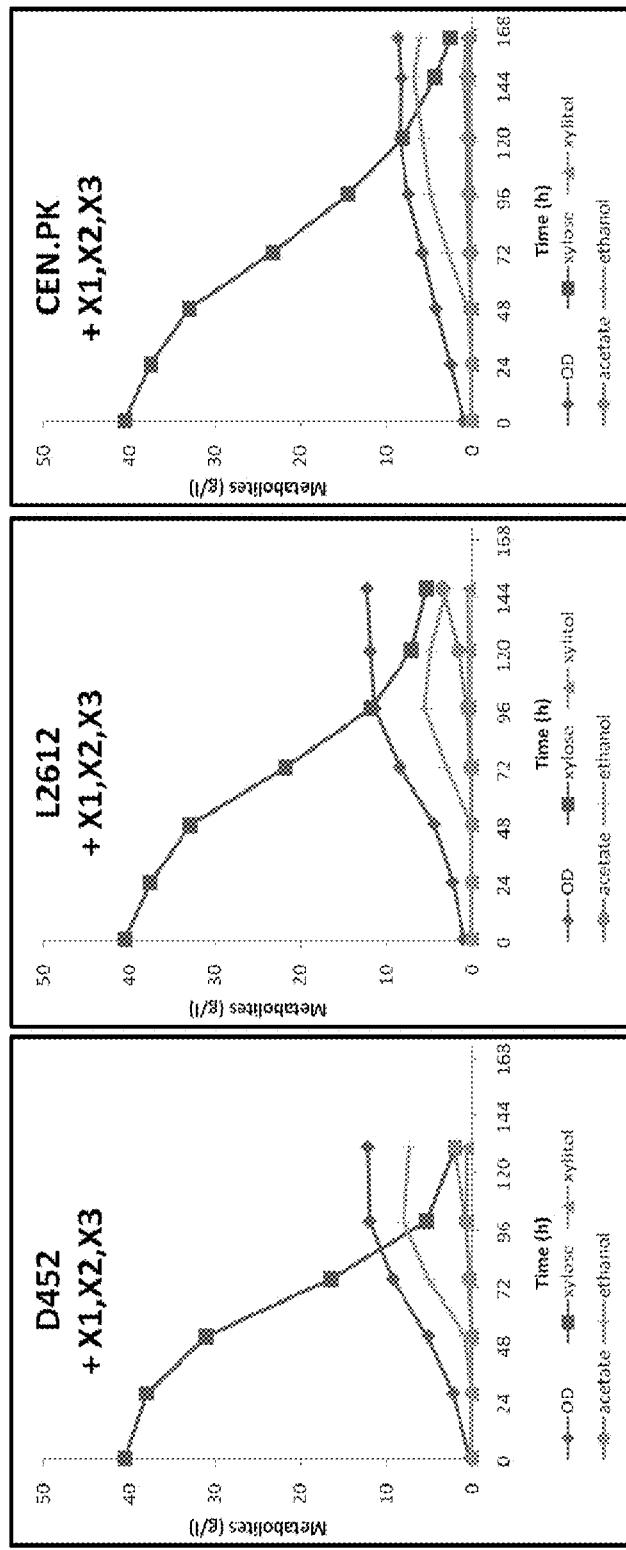
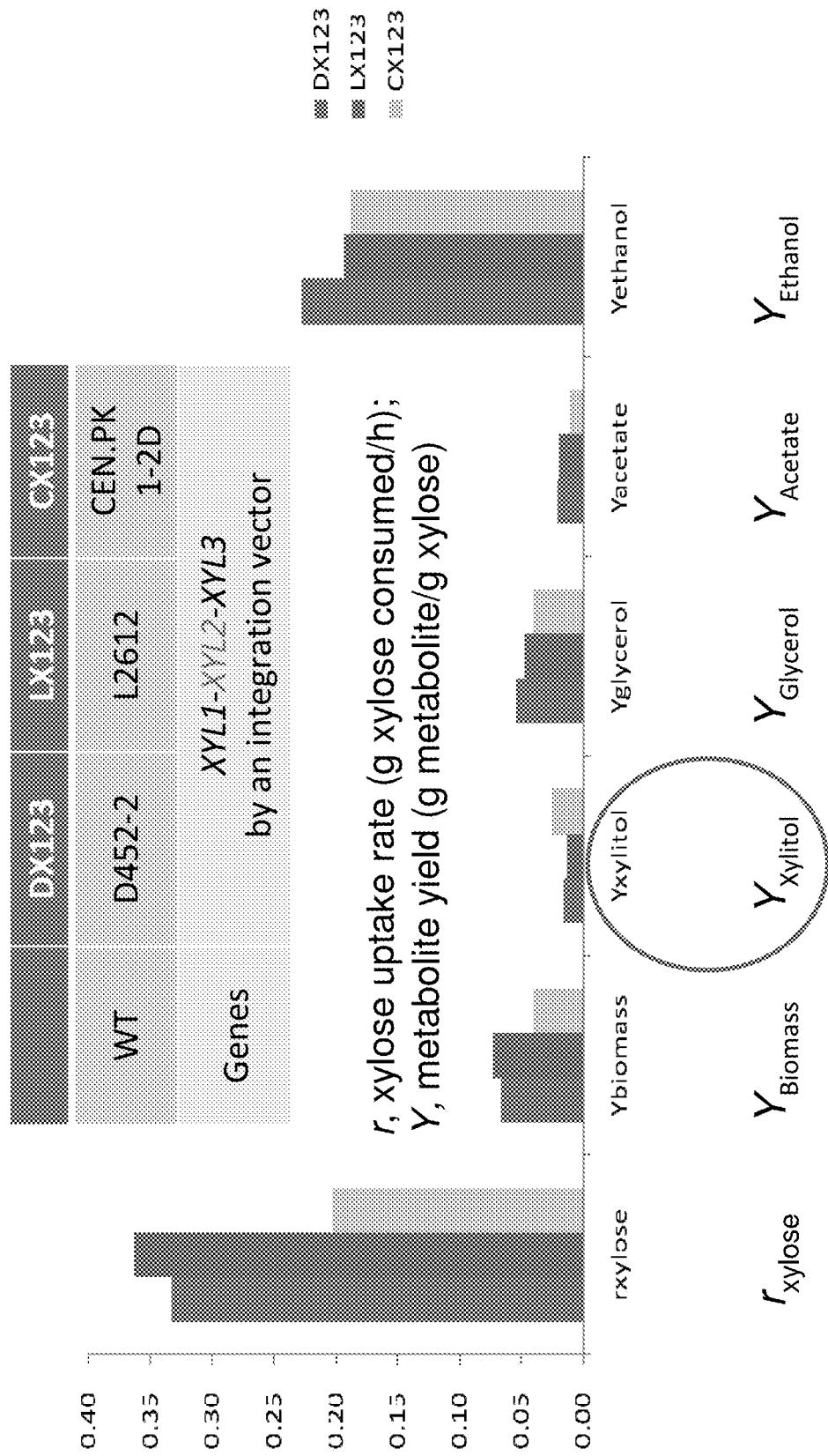


Figure 44 Three different laboratory strains

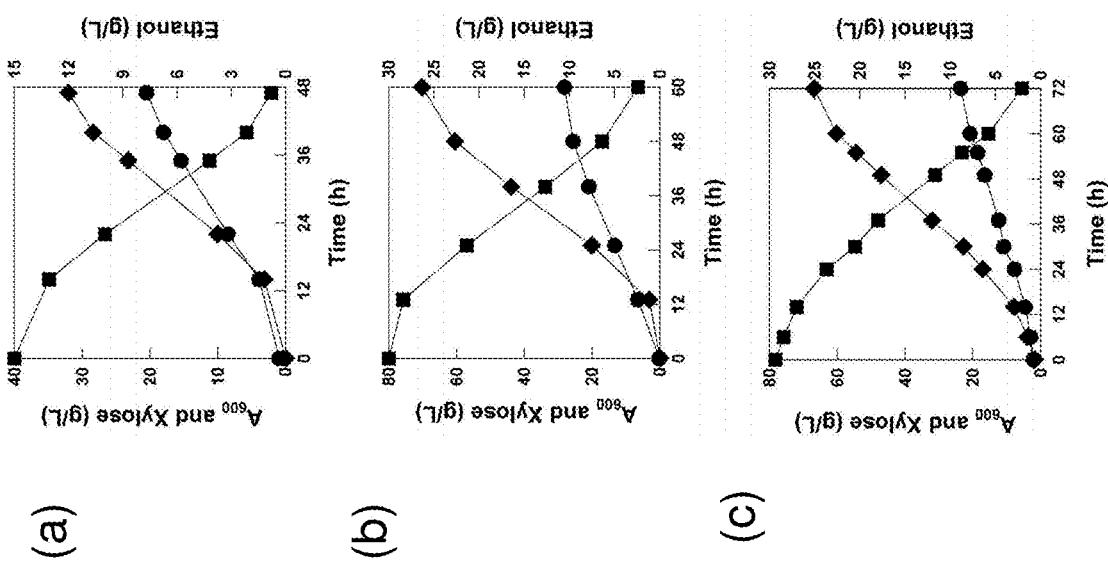


Figure 45 (a)

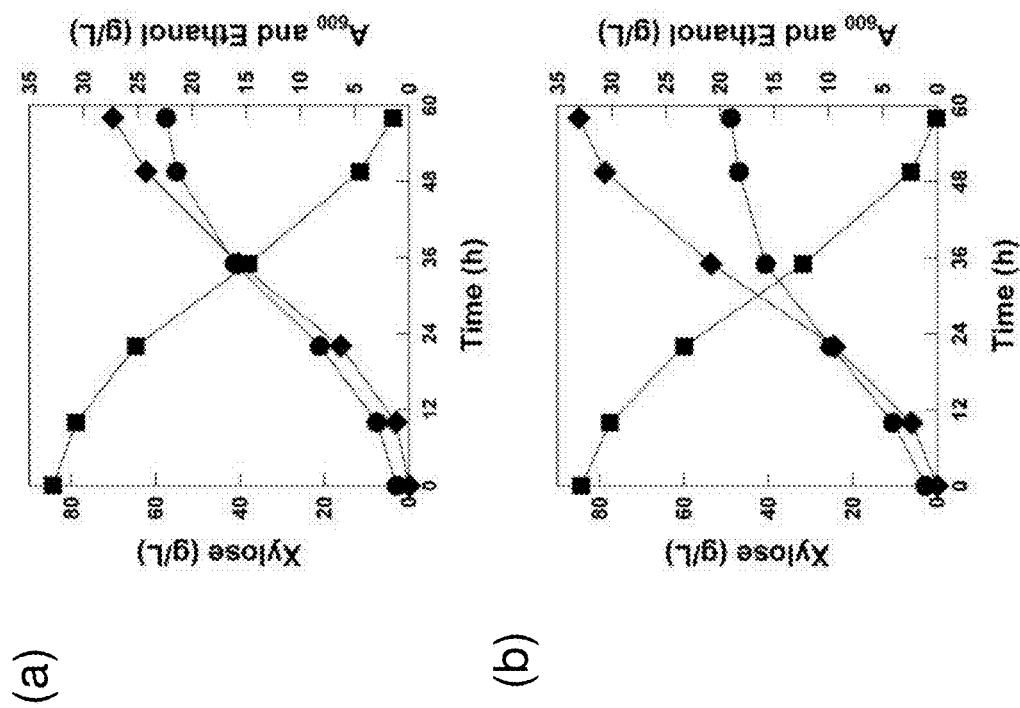
Figure 46

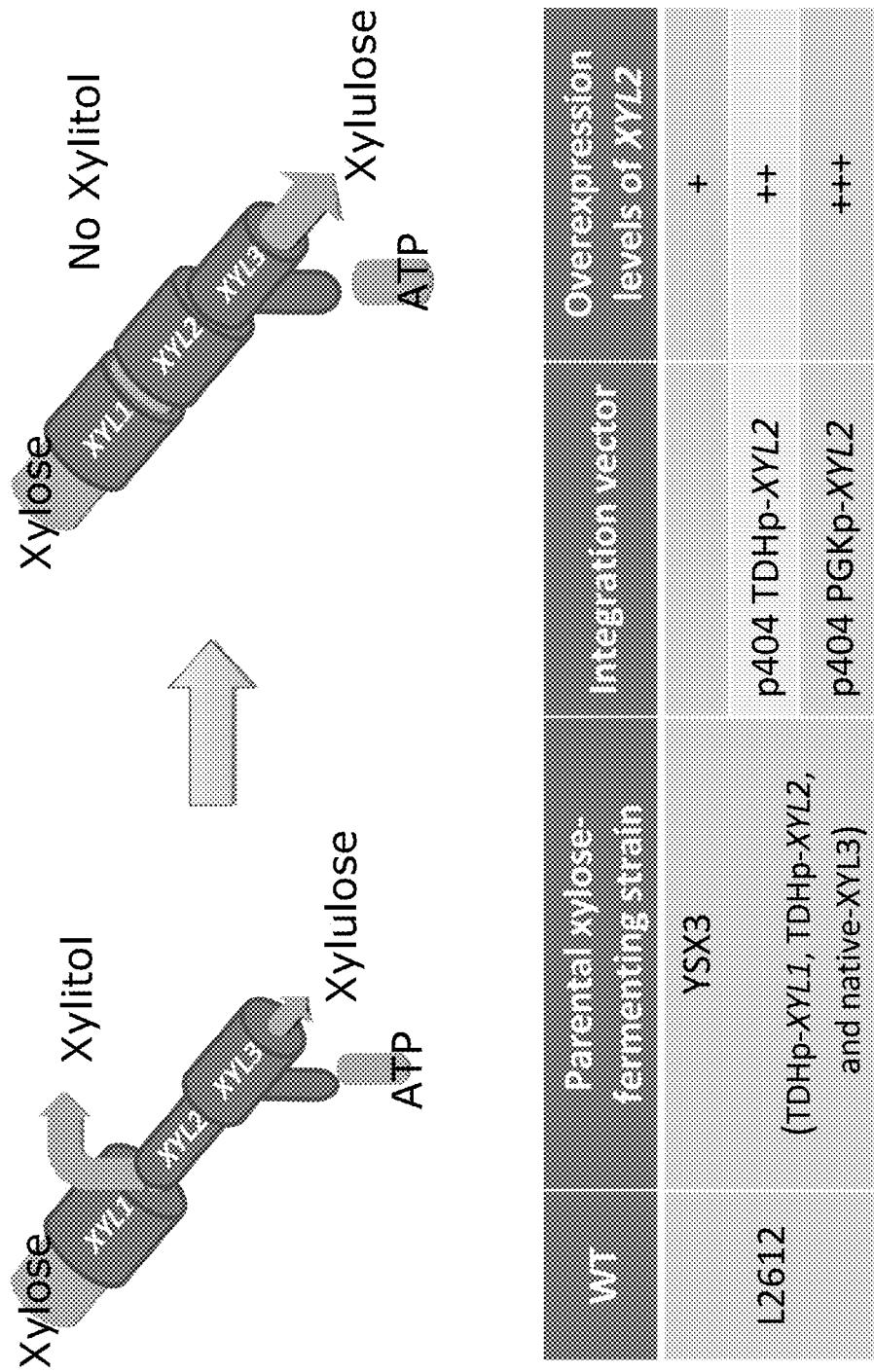
Figure 47**Hypothesis and experimental design**

Figure 48
Additional *XYL2* integration improves xylose fermenting ability

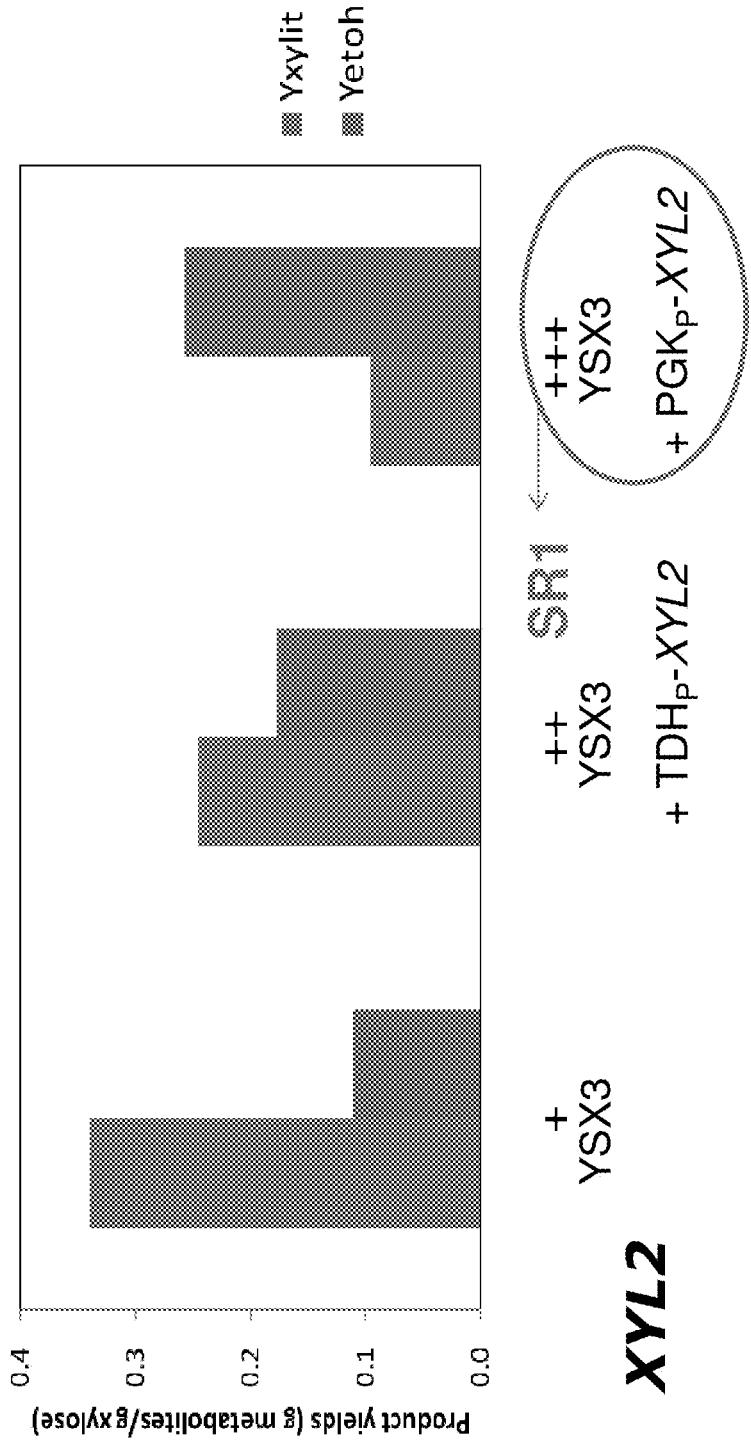


Figure 49
**Negligible amounts of xylitol were
accumulated after additional overexpression
of *XYL2-XYL3***

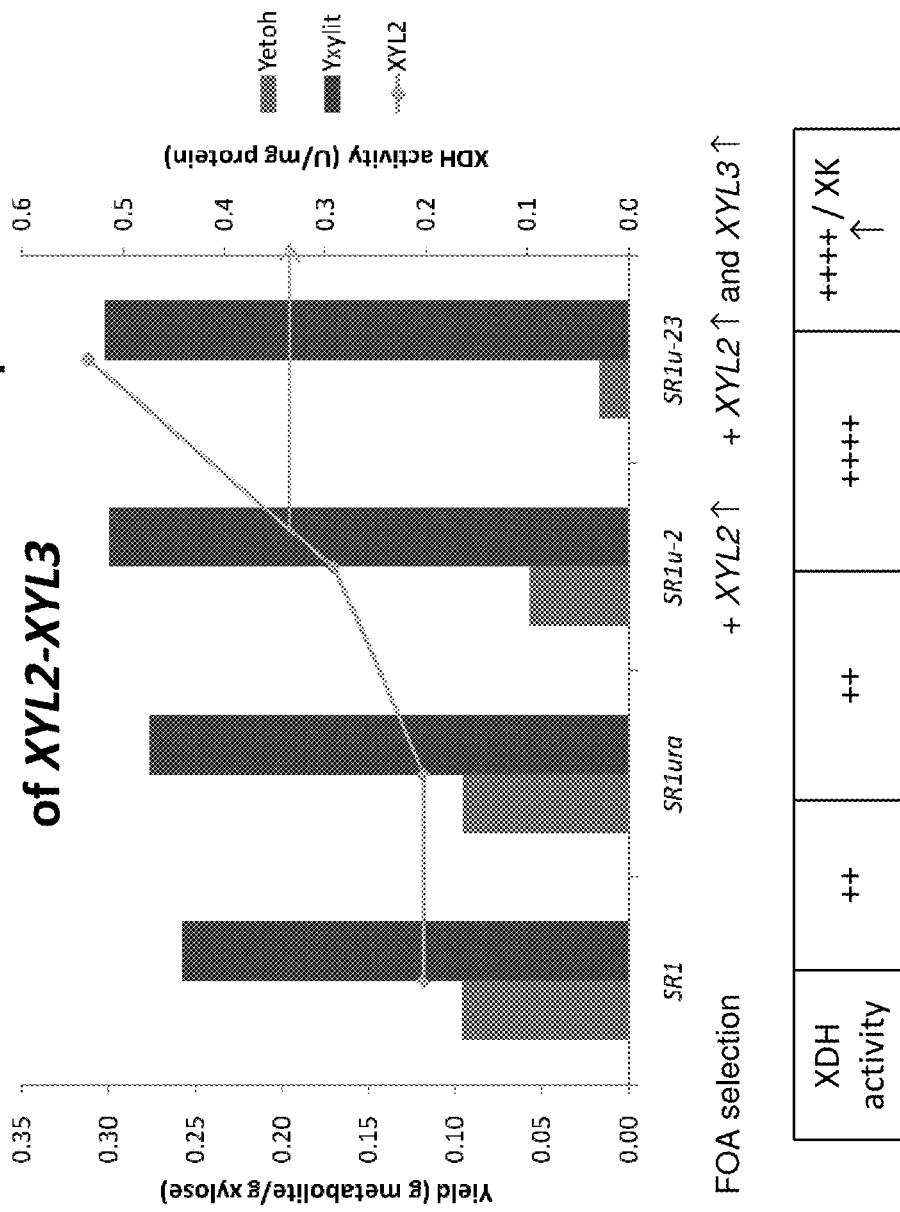


Figure 50
Construction of engineered strains exhibiting
different XR activities

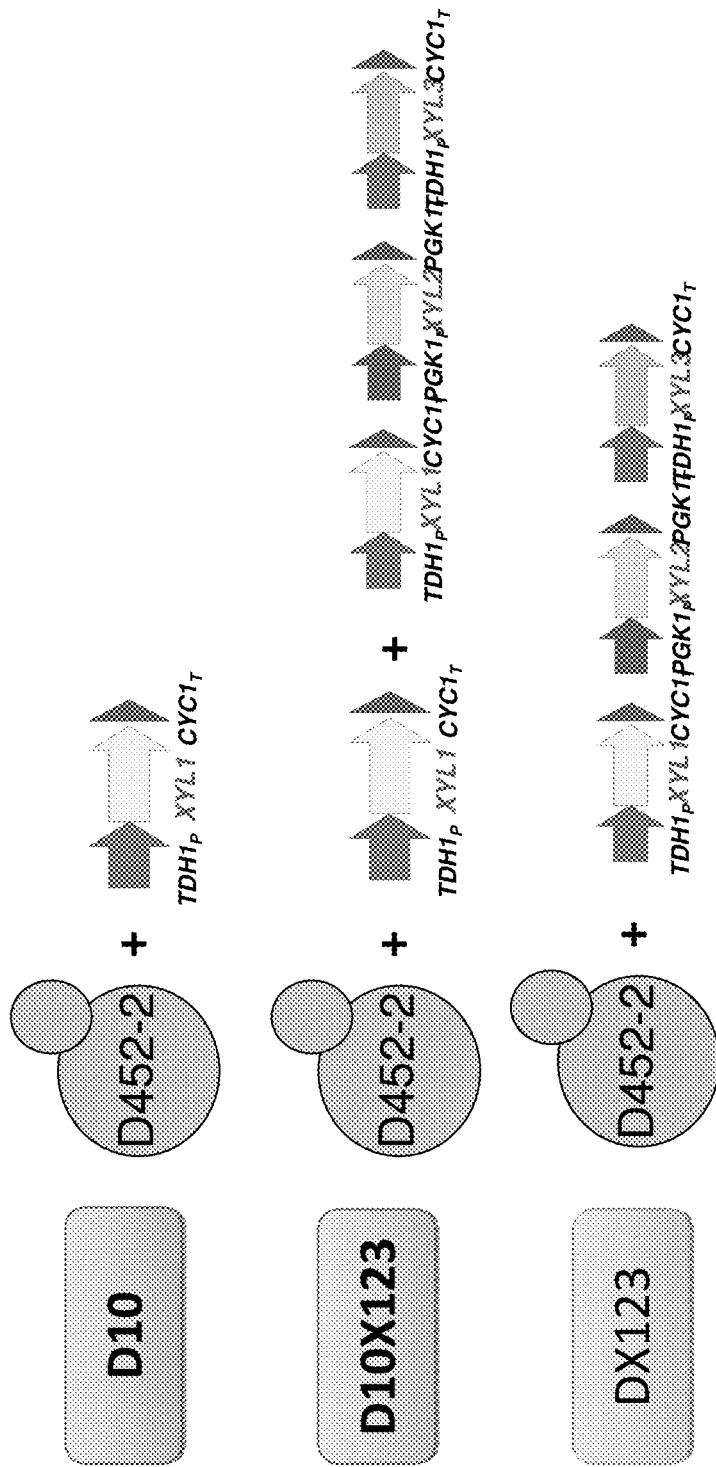


Figure 51
Strong *XYL1* with *XYL2/XYL3* caused xylitol accumulation

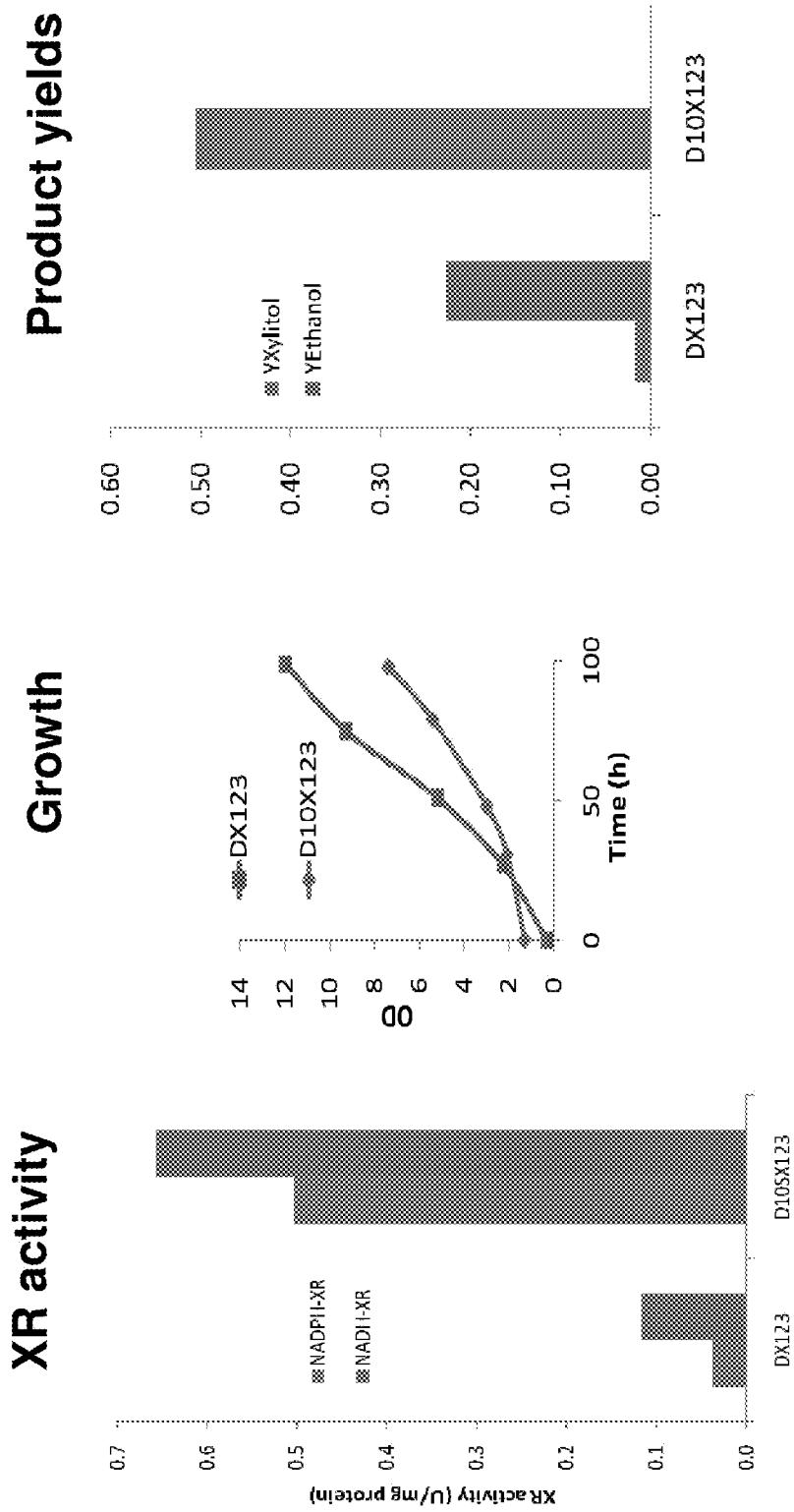


Figure 52
Two equivalent engineered strains expressing
XYL1 and *GRE3* were constructed

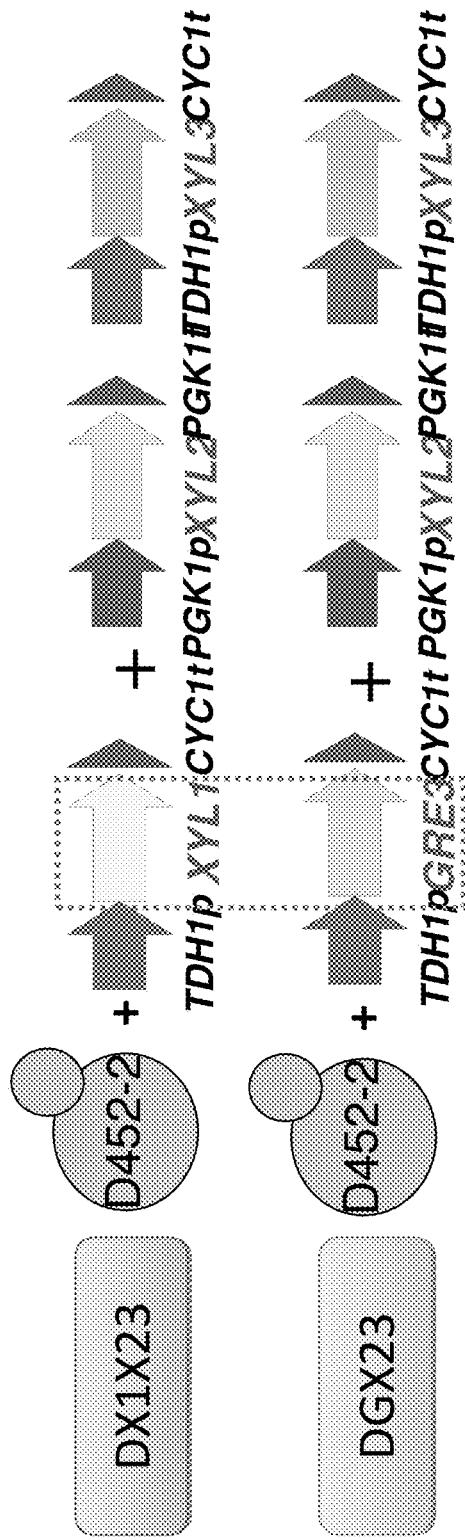


Figure 53
**Xylose fermentation with high OD
(OD=10) inoculations**

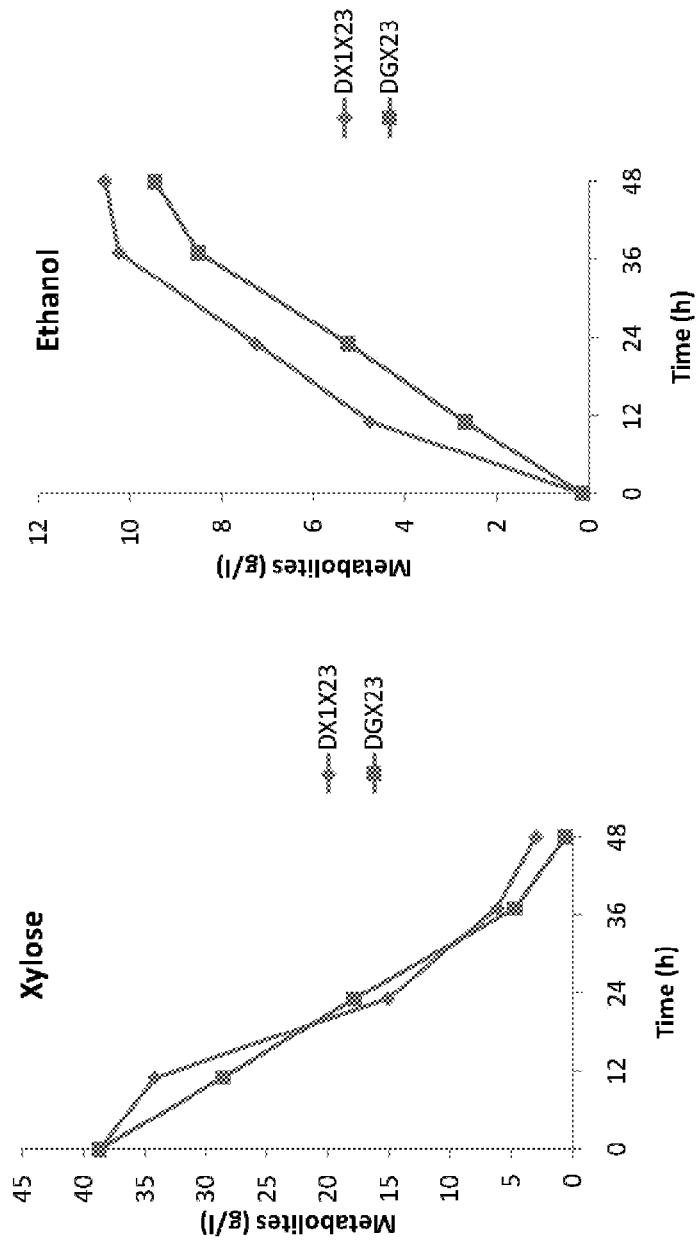


Figure 54
**80 g/L of xylose fermentation with
high OD (OD=15) inoculations**

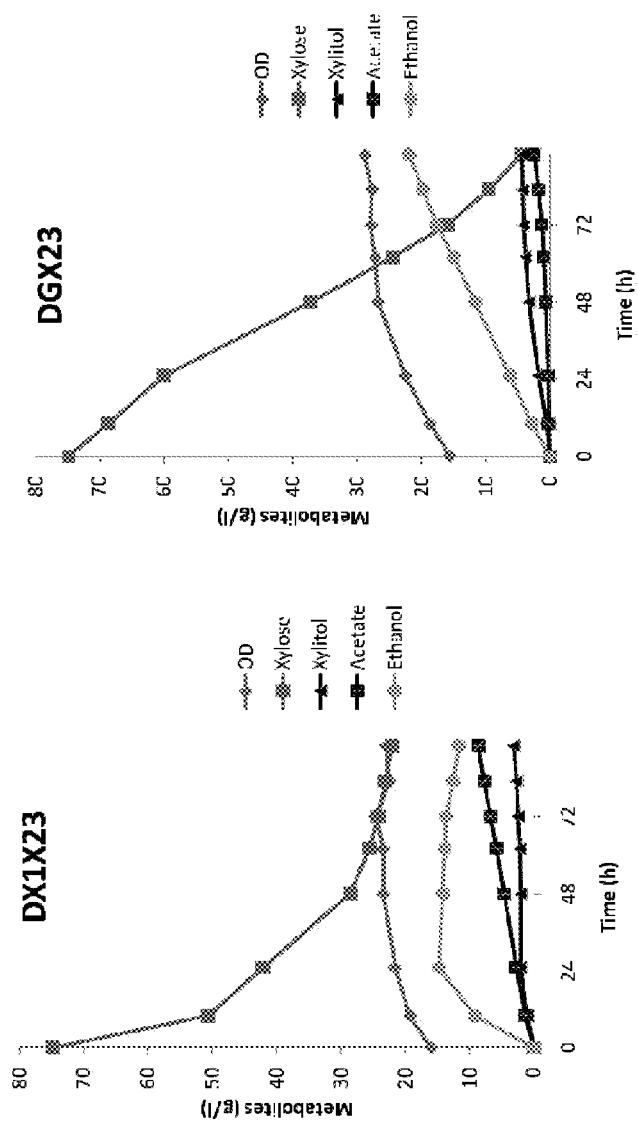


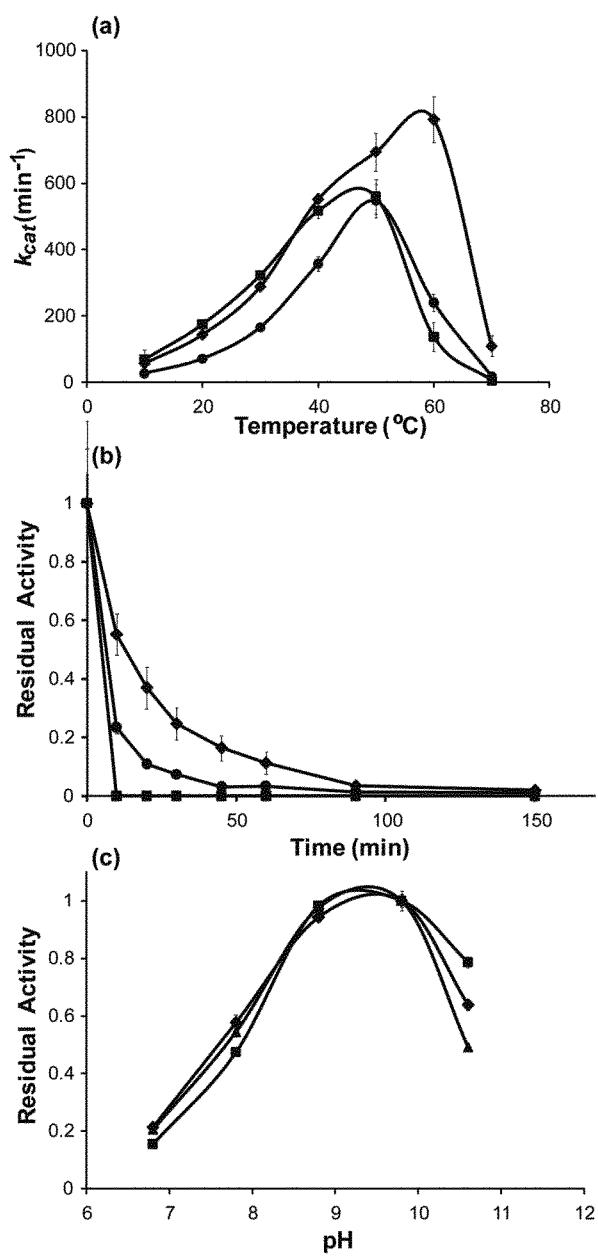
Figure 55

Figure 56

Score = 299 bits (765), Expect = 3e-79, Method: Compositional matrix adjust.

Identities = 169/365 (46%), Positives = 222/365 (60%), Gaps = 22/365 (6%)

psXDH ANPSLVLNKIDDISFETYDAPEISEPTDVLVQVKKTGICGSIDIHFYAHGRIGNFVLTKPM 62
+N S VLNK D+ F+ P+I+ P DVLV V TGICGSD+H++ HG IG+FV+ PM

ncXDH SNLSFVLNKPLDVCFQDKPVPKINSPHDVLVAVNYTGICGSDVHYWLHGAIGHFVVKDPM 66

psXDH VLGHESAGTVVQVGKGVTSLKVDNVIAIEPGIPSRESDEYKSGHYNLCPHMAFAATPNSK 122
VLGHESAGT+V VG V +L VGD VA+EPG P R SGHYNLCP M FAATP

ncXDH VLGHESAGTIVAVGDAVKTLSVGDRVALEPGYPCRRCVHCLSGHYNLCPEMRFAATPPYD 126

psXDH EGEPNPPGTLCKYFKSPEDFLVKLPDHVSLELGALVEPLSVGvhASKLGSVAFGDYVAVF 182
GTL ++ +P DF KLP+ VSL+ GAL+EPL+V VH +K + G V V

ncXDH -----GTLTGFWTAPADFCYKLPETVSLQEGALIEPLAVAVHITKQAKIQPGQTVVVM 179

psXDH GAGPVGLLAAAVAKTFGAKGVIVVDIFDNKLKMAKDIGAATHTFNSKTGGSEE----LI 237
GAGPVGLL AAVAK +GA V+ VDI +KL+ AK AATHT+ S+ EE +

ncXDH GAGPVGLLCAAVAKAYGASKVVSVDIVPSKLEFAKSF-AATHTYLSQRVSPEENARNIIA 238

psXDH KAFGNVPNVVLECTGAEPCIKGVDIAPIPGGRFVQVGNAAGPVSFPITVFAMKELTFLG 297
A G +V++ +GAEP I+ + + GG +VQ G ++FPI +KE+T G

ncXDH AADLGEGADAVIDASGAEPSIQAALHVVRQGGHYVQGGMGKDNITFPIMALCIKEVTASG 298

psXDH SFRYGFNDYKTAVGIFDTNYQNGRENAPIDFEQLITHRYKFKAIEAYDLVRAGKGAVKC 357
SFRYG DY+ A+ + E +D ++L+ FK+A EA+ V+ G+ +K

ncXDH SFRYGSGYRLAIQLV-----EQGKVDVKLVLNGVVPFKNAEEAFKKVKEGE-VIKI 349

psXDH LIDGP 362

LI GP

ncXDH LIAGP 354

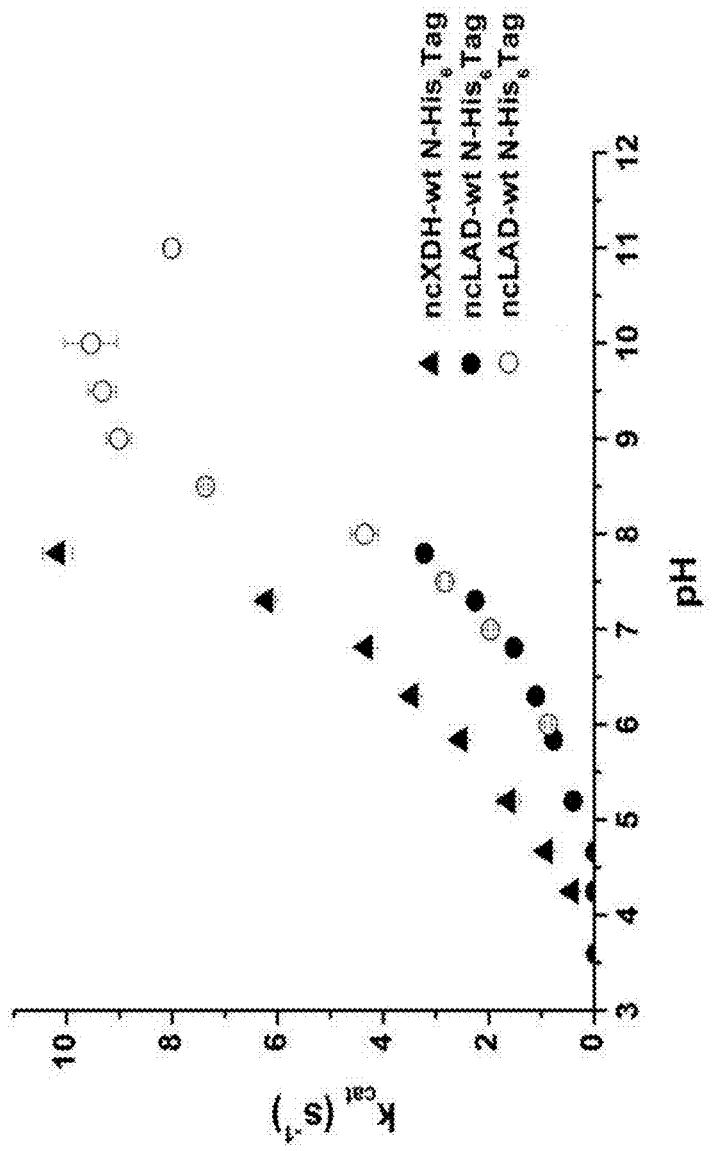
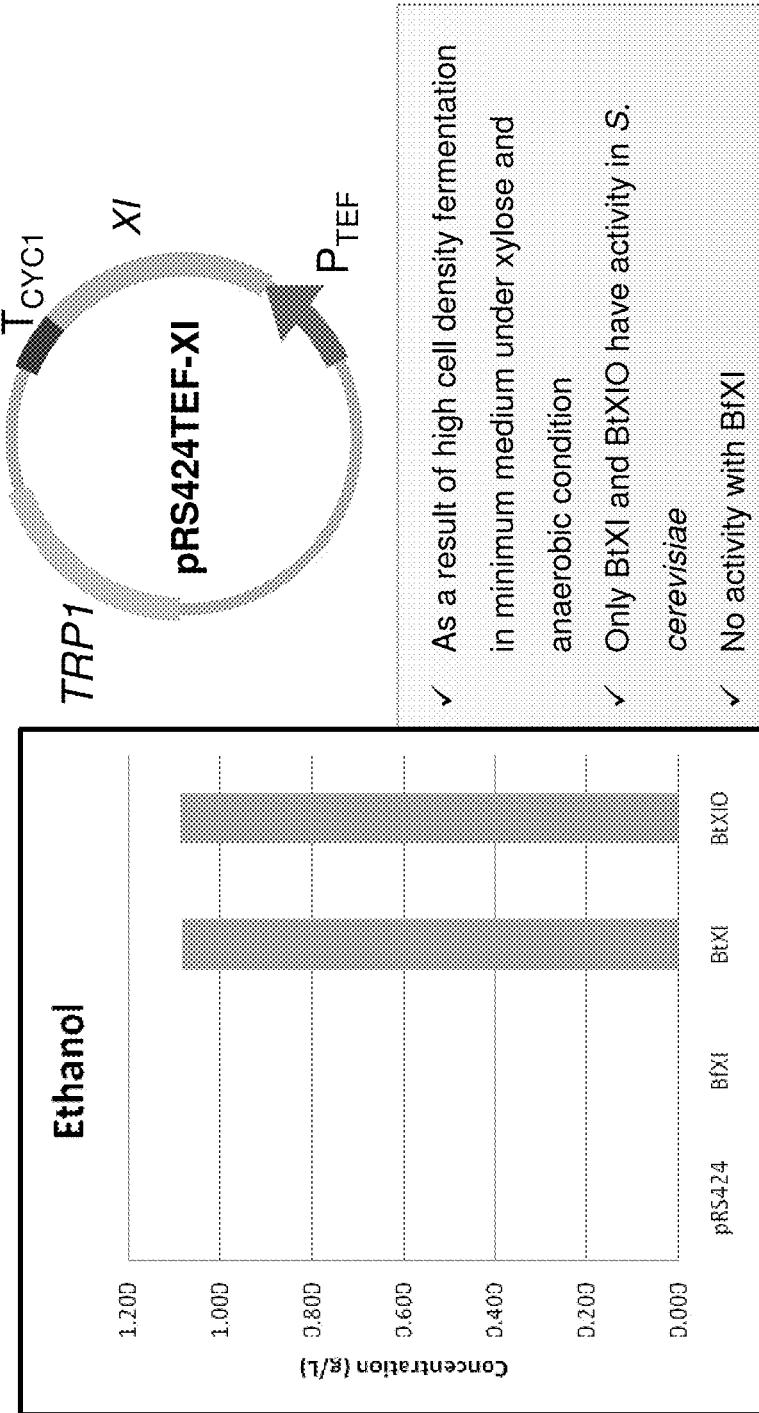
FIG. 57

Figure 58

	Plasmid (L2612)	Integration (D452-2)		
	OD 48	OD 96	OD 51	OD 96
Ethanol (g/L)	4.92	5.15	7.67	7.88
Yield (EtoH)	0.35	0.34	0.30	0.30
Productivity (g/L-hr)	0.04	0.04	0.05	0.06
Xylytol (g/L)	3.62	3.61	2.80	2.74
Yield (Xt)	0.28	0.24	0.15	0.14

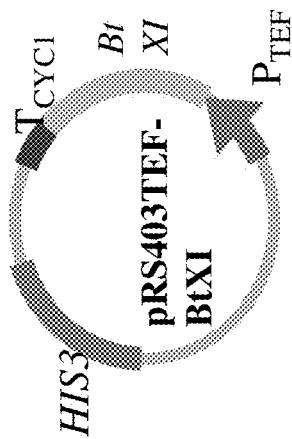
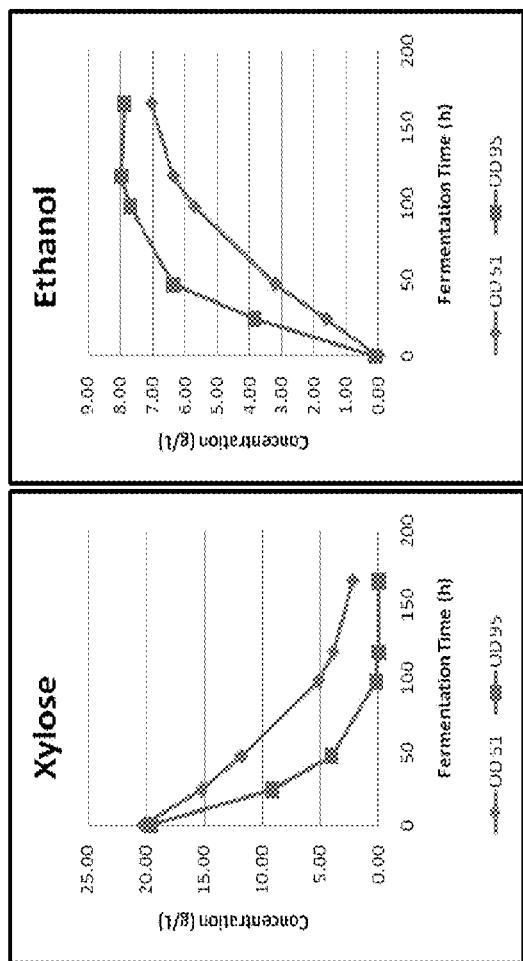
**Figure 59**

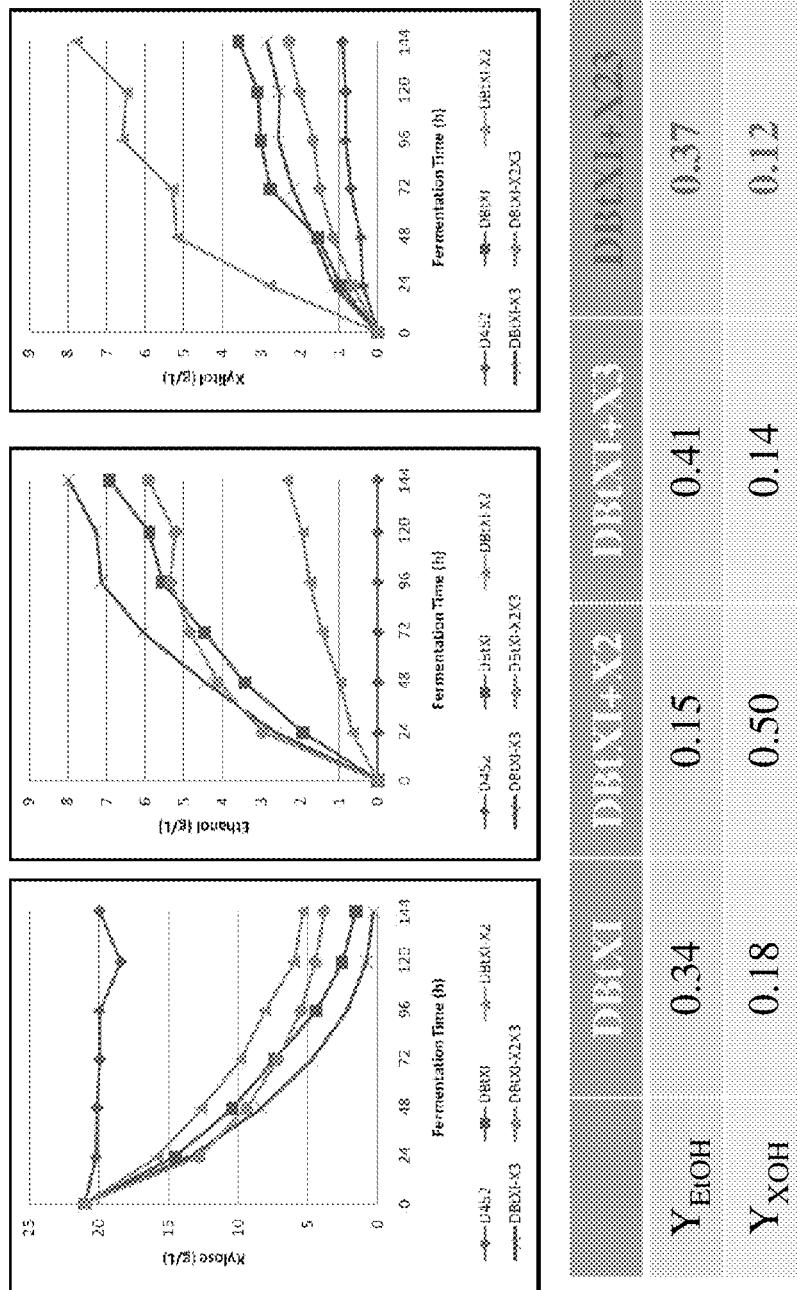
Figure 60

Figure 61
Cell growth on xylulose

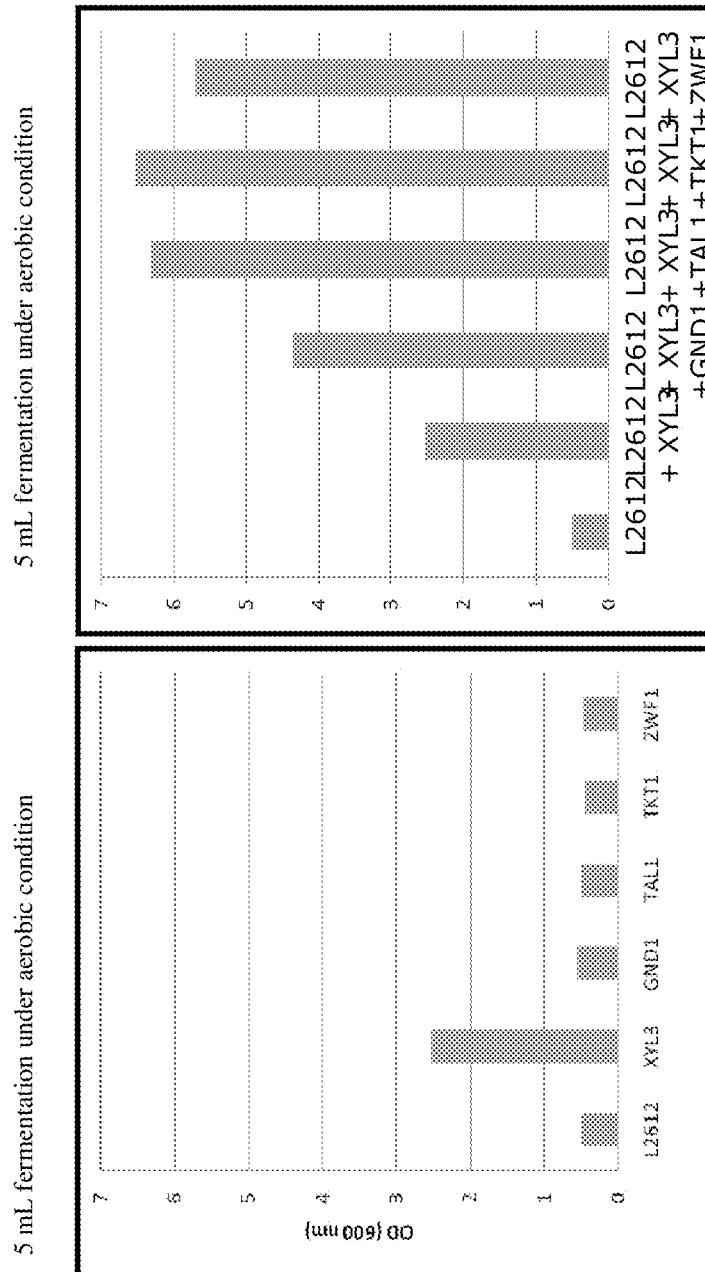


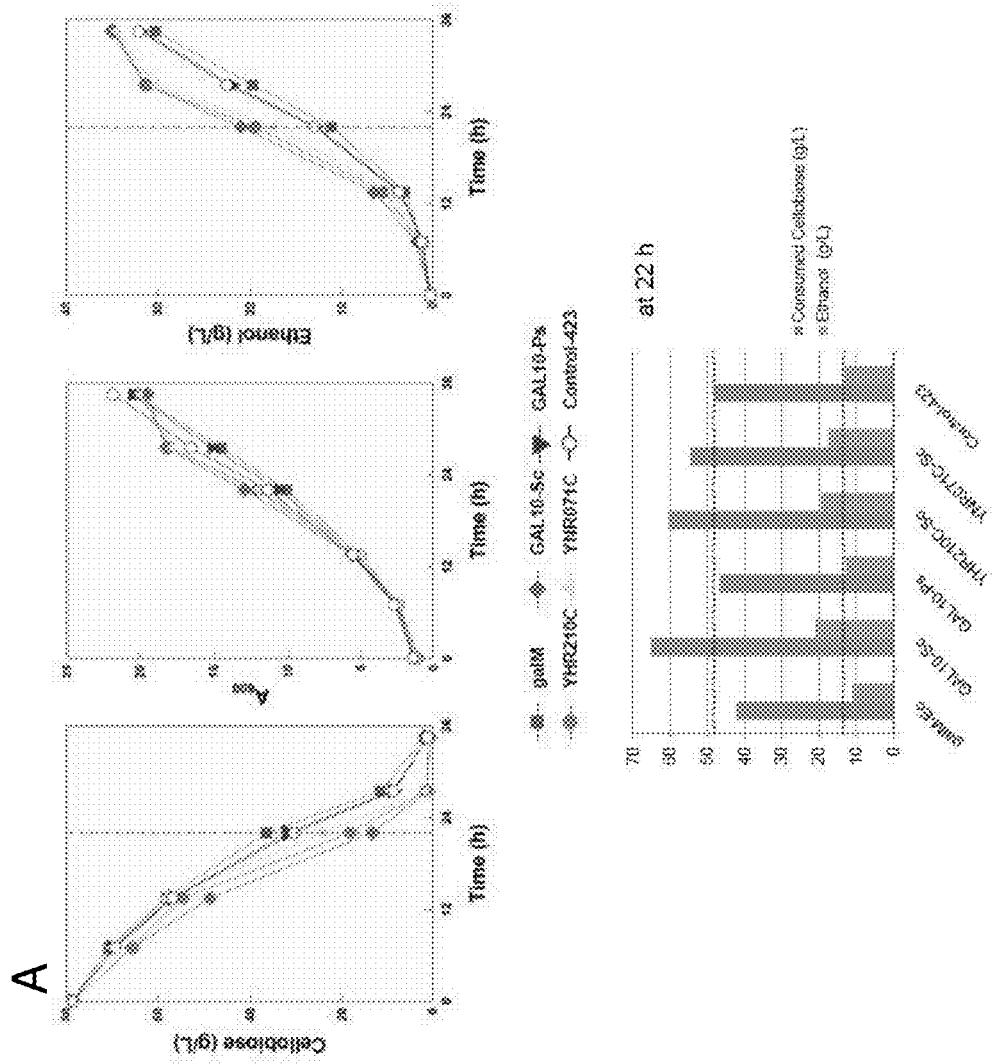
Figure 62

Figure 62 (cont.)**B**

	0.8 (%)	2.0 (%)	4.0 (%)	7.0 (%)	10.0 (%)
galM-Ec	20.4	30.3	0.40	0.88	
GAL10-Sc	19.5	34.9	0.45	1.01	
GAL10-Ps	20.5	31.7	0.41	0.92	
YHR210C-Sc	20.0	34.7	0.45	1.01	
YNR071C-Sc	20.0	34.6	0.45	1.00	
Control-423	21.8	32.0	0.42	0.93	

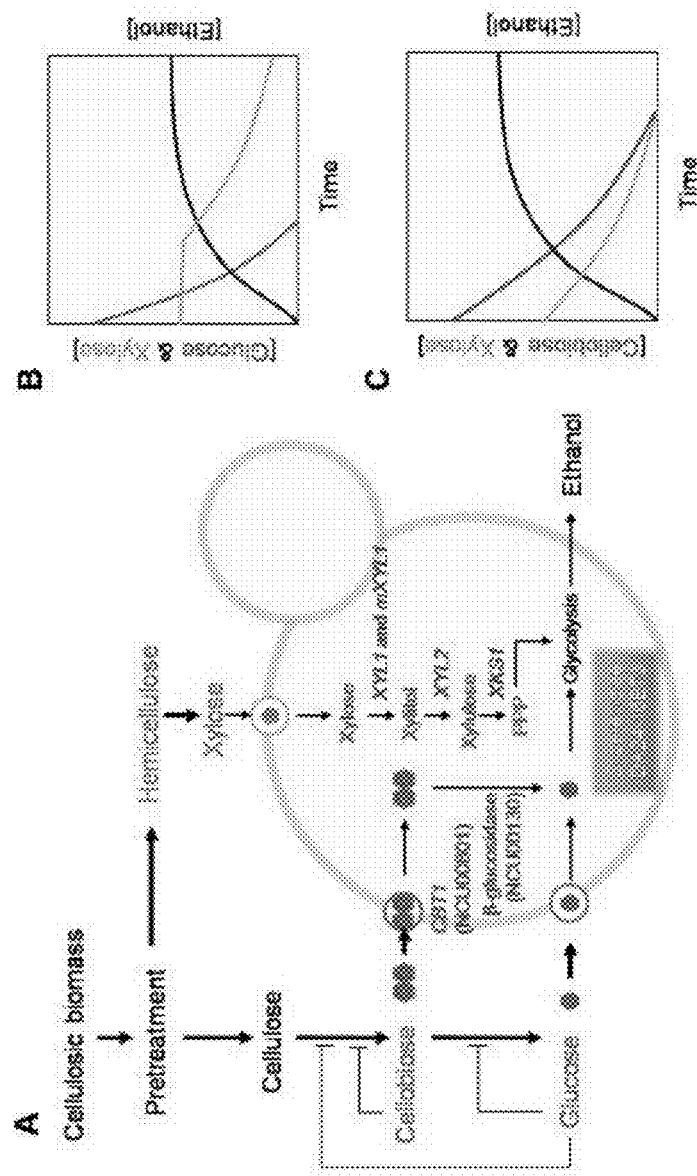
Figure 63

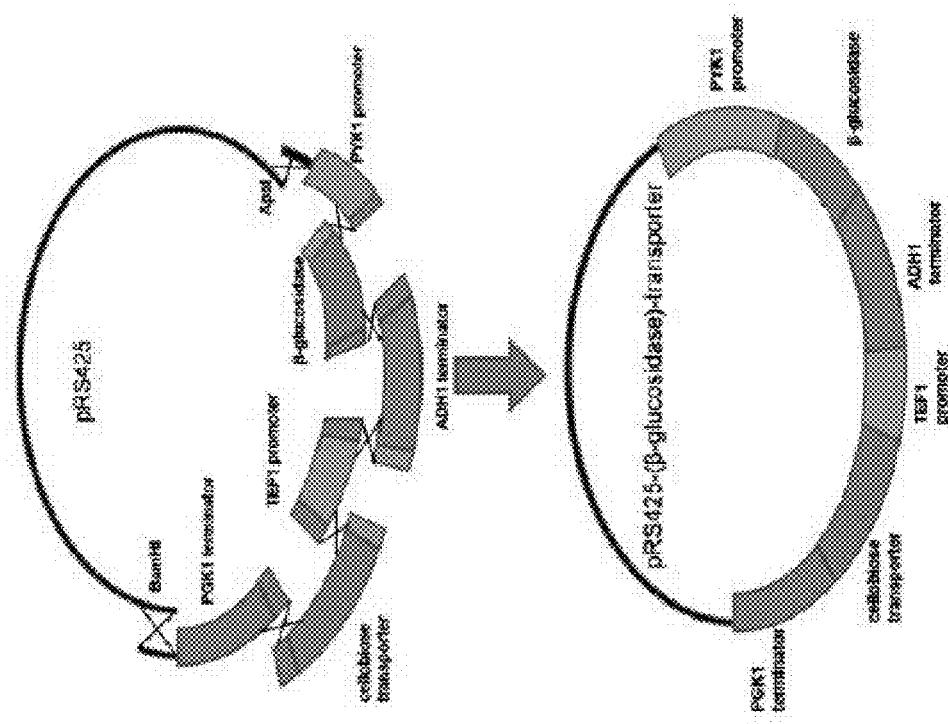
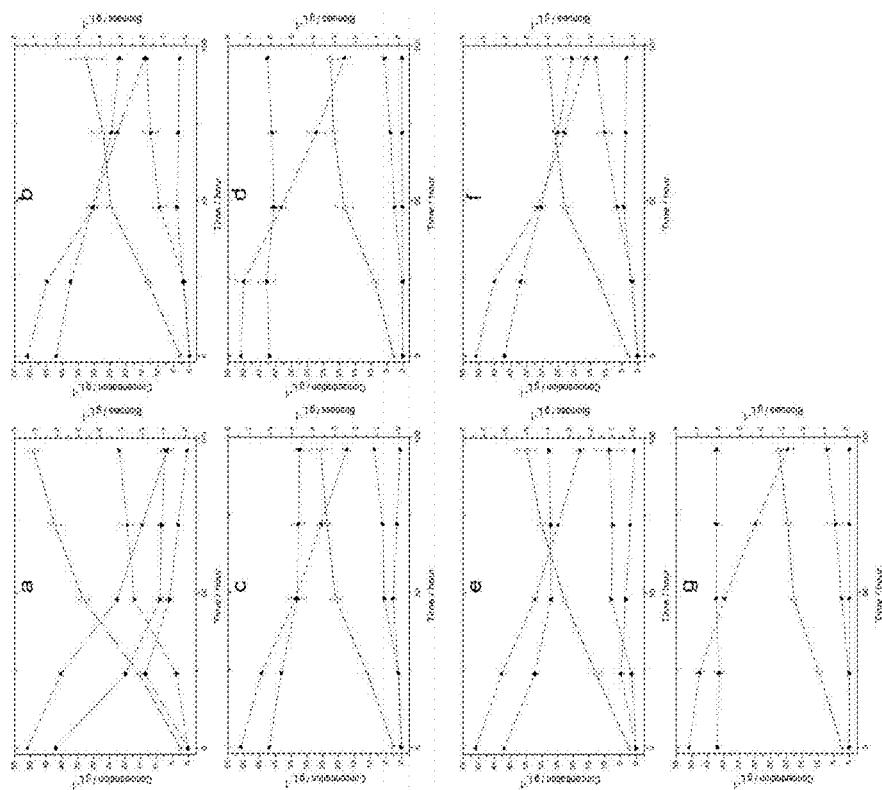
Figure 64

Figure 65

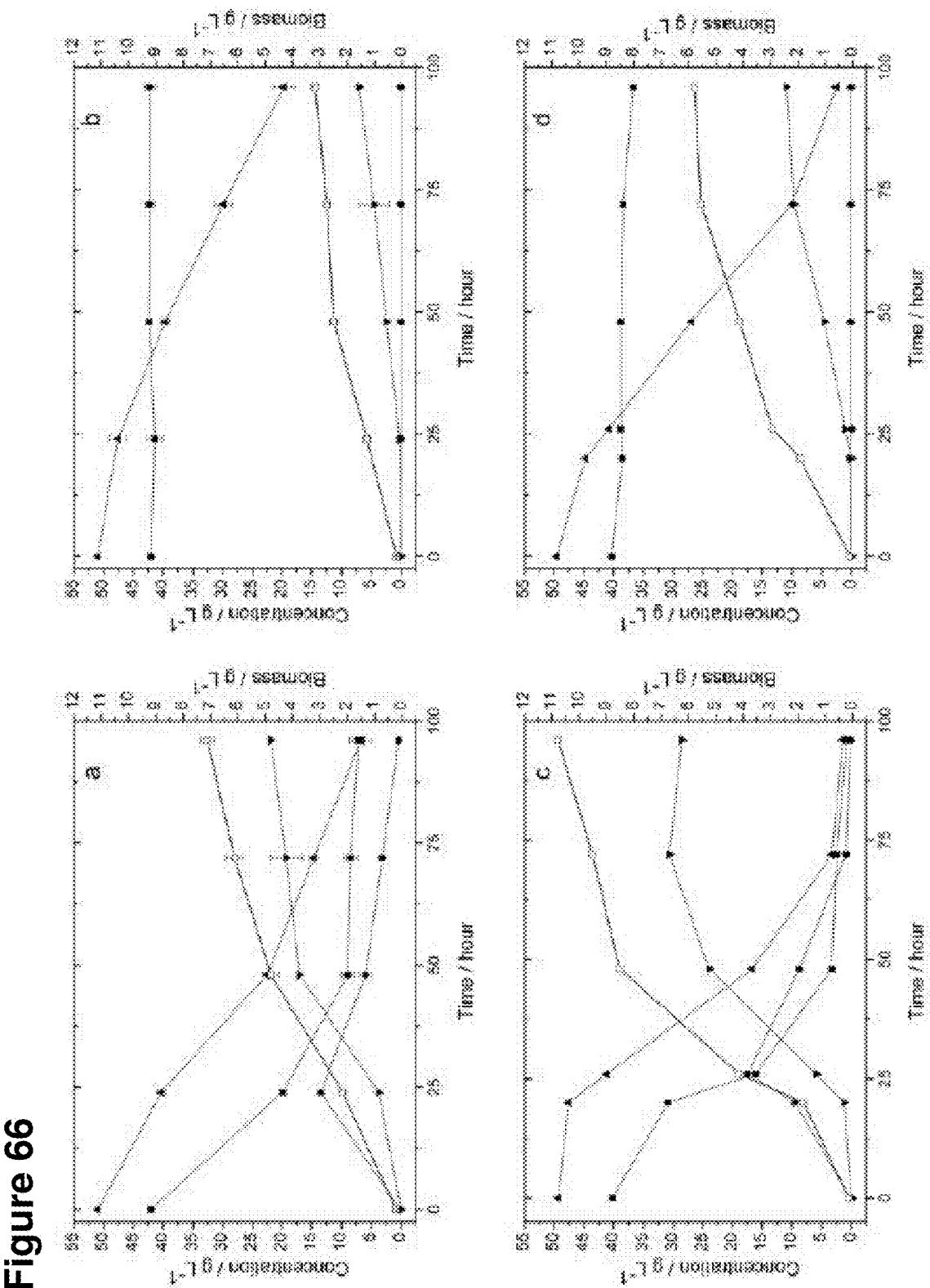


Figure 66

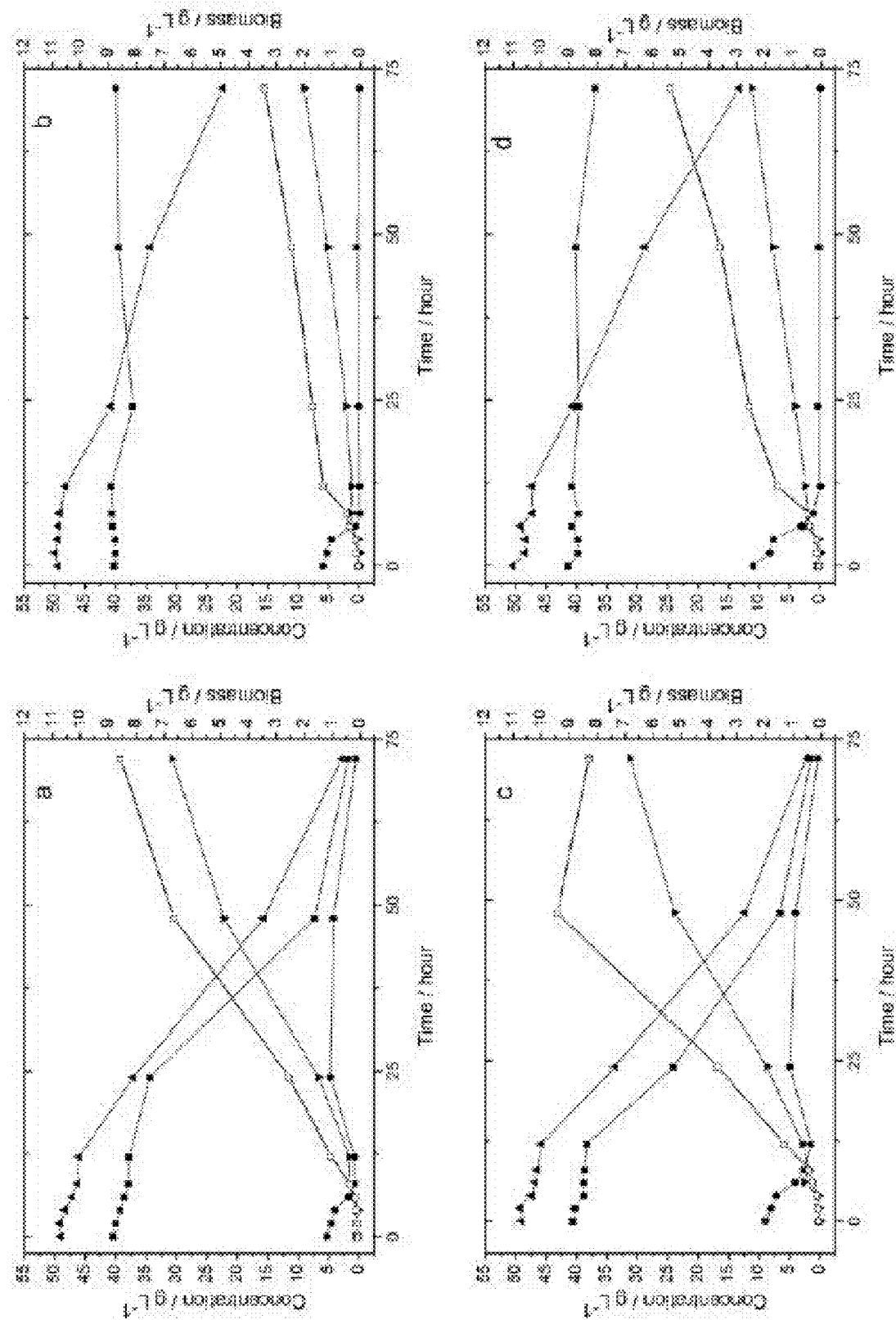


Figure 67

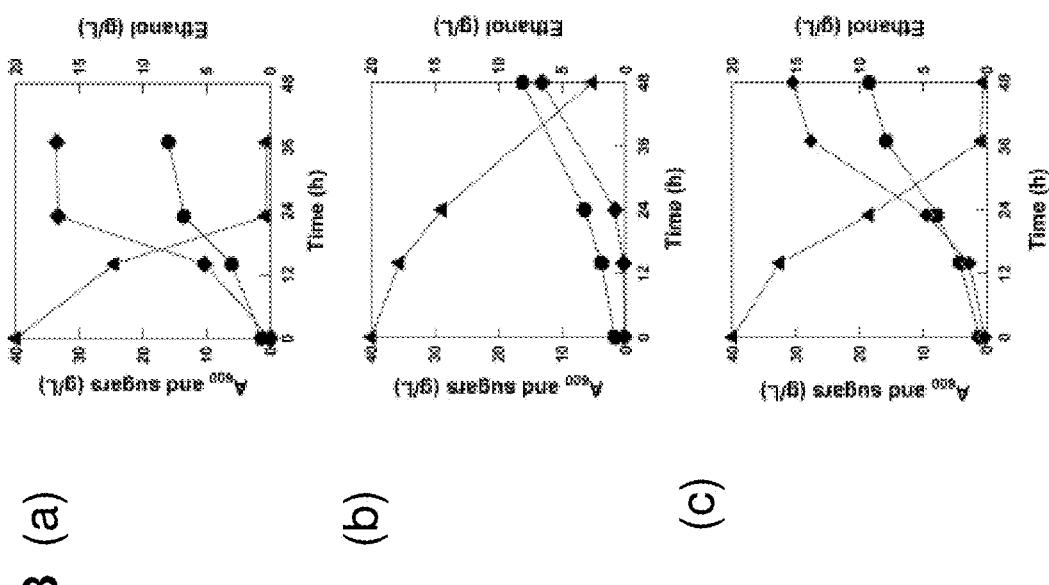


Figure 68 (a)

(b)

(c)

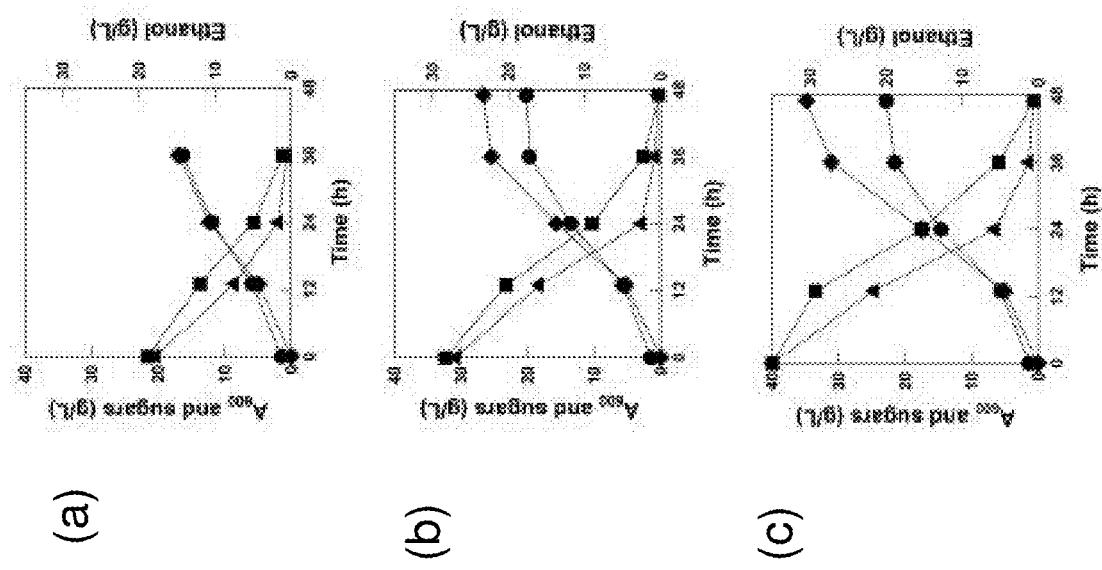


Figure 69 (a)

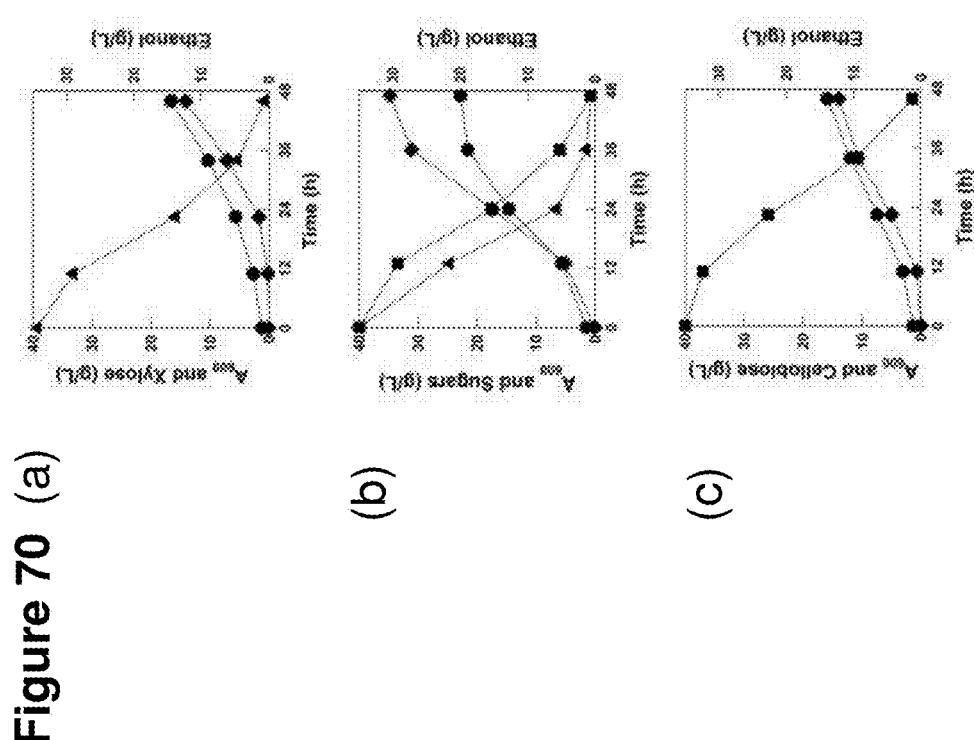


Figure 70 (a)

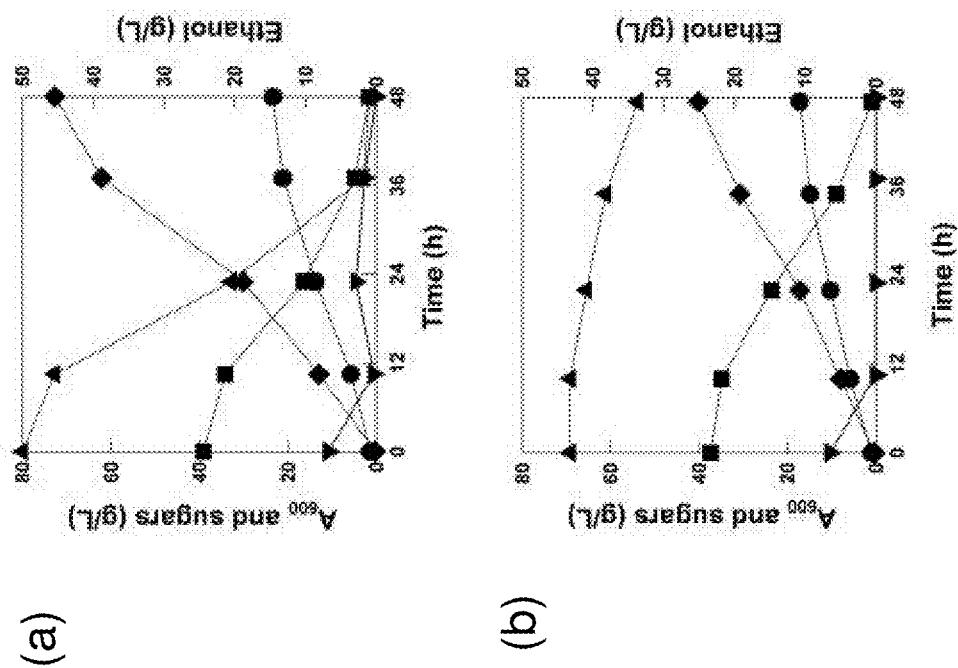
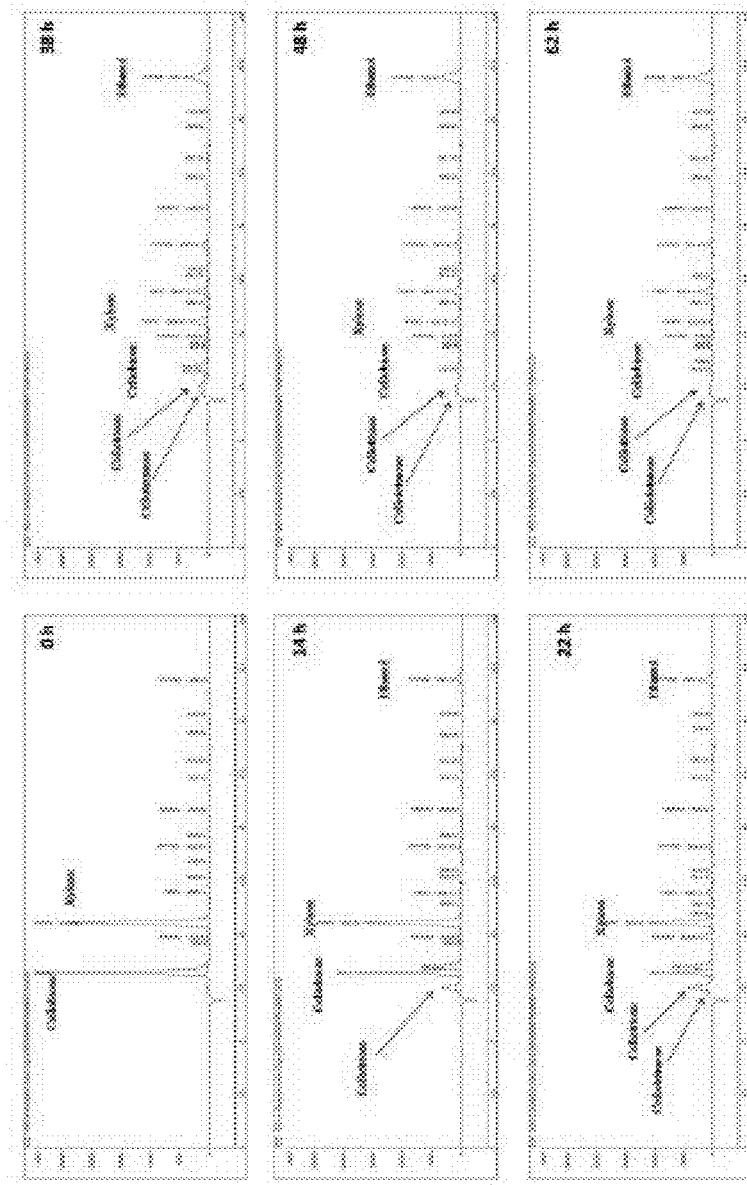
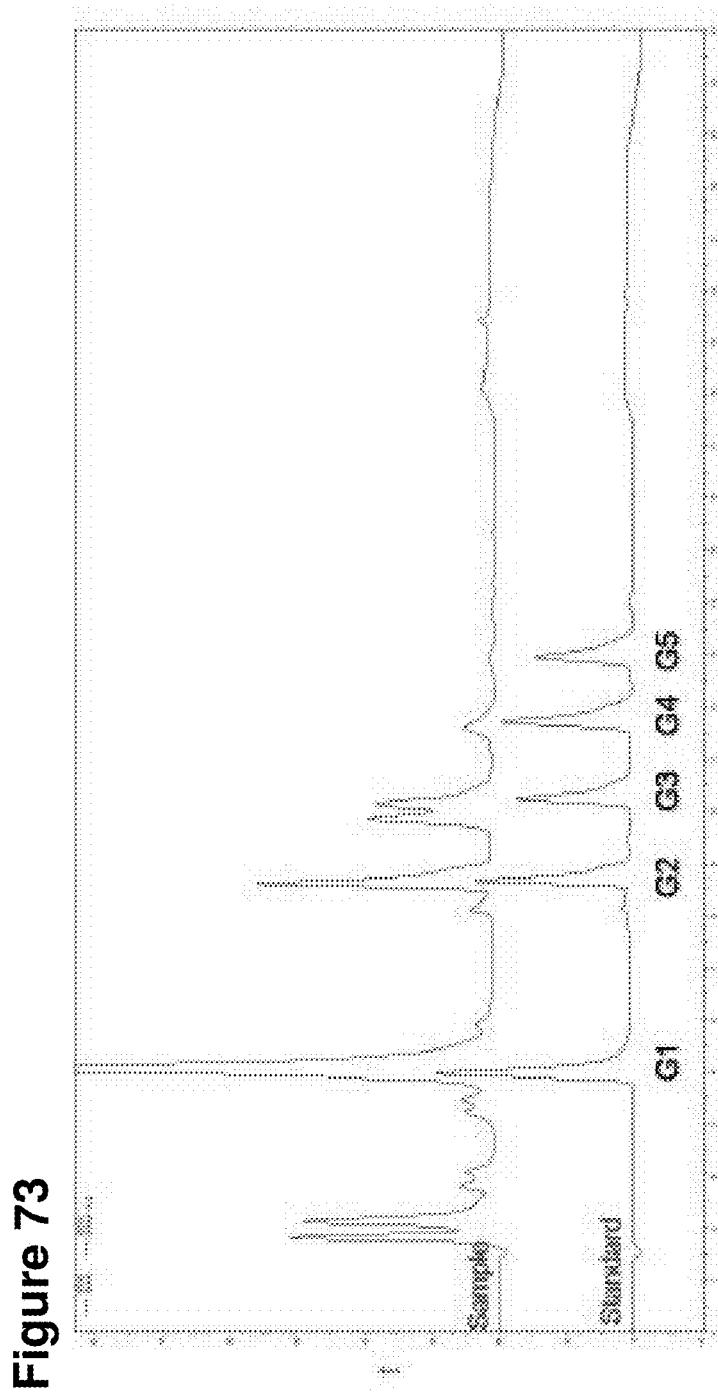


Figure 71 (a)

Figure 72



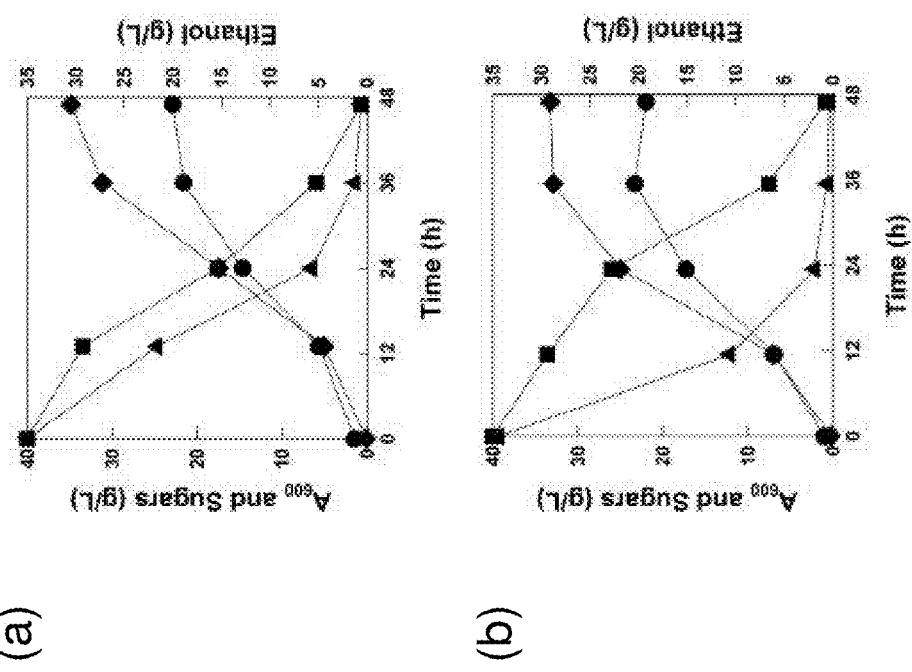


Figure 74 (a)

(b)

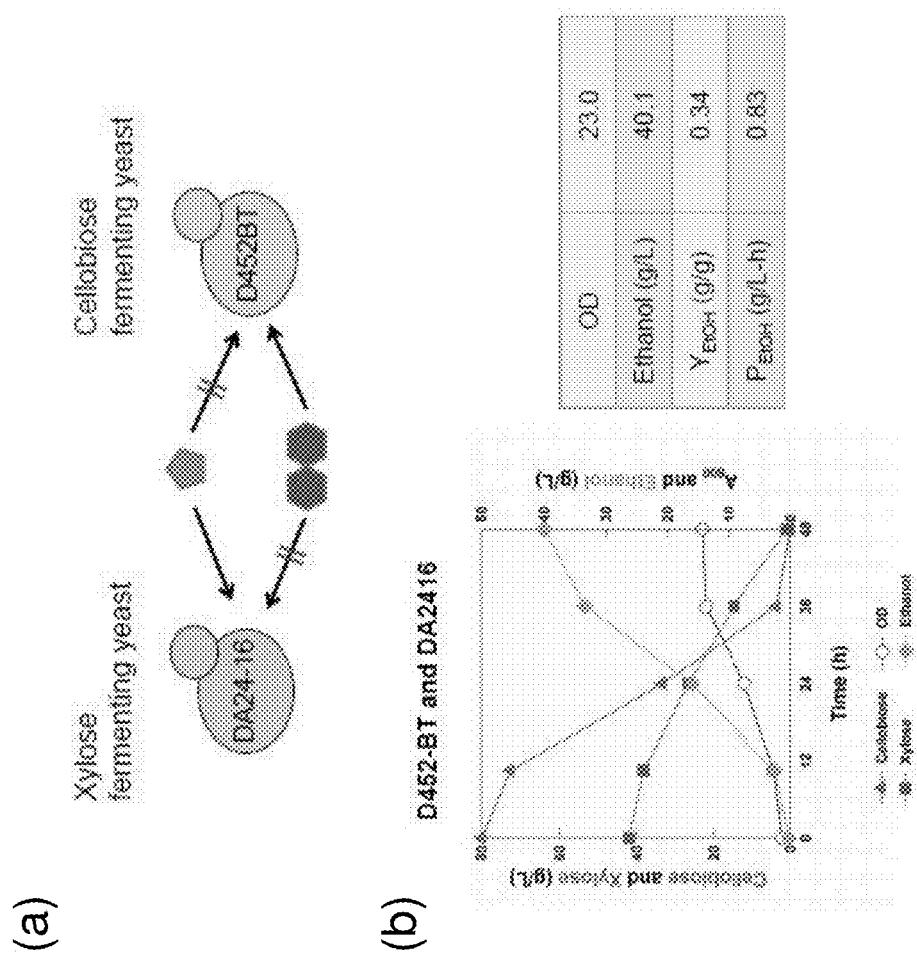
Figure 75

Fig. 76

NCU00010.2	NCU01906.2	NCU04373.2	NCU06656.2	NCU09051.2b
NCU00028.2	NCU01940.2	NCU04388.2	NCU06681.2	NCU09115.2
NCU00060.2	NCU01970.2	NCU04394.2	NCU06707.2	NCU09133.2
NCU00073.2	NCU01989.2	NCU04400.2	NCU06752.2	NCU09138.2
NCU00102.2	NCU02018.2	NCU04401.2	NCU06803.2	NCU09169.2
NCU00111.2	NCU02027.2	NCU04415.2	NCU06977.2	NCU09175.2
NCU00122.2	NCU02097.2	NCU04417.2	NCU06999.2	NCU09209.2
NCU00124.2	NCU02124.2	NCU04460.2	NCU07027.2	NCU09210.2
NCU00130.2	NCU02136.2	NCU04475.2	NCU07063.2	NCU09266.2
NCU00173.2	NCU02179.2	NCU04491.2b	NCU07064.2b	NCU09267.2
NCU00292.2	NCU02188.2	NCU04510.2	NCU07133.2	NCU09285.2
NCU00299.2	NCU02238.2	NCU04521.2	NCU07143.2	NCU09316.2
NCU00304.2	NCU02307.2	NCU04605.2	NCU07158.2	NCU09427.2
NCU00305.2	NCU02316.2	NCU04623.2	NCU07225.2	NCU09491.2
NCU00378.2	NCU02342.2	NCU04674.2	NCU07267.2	NCU09532.2
NCU00379.2	NCU02343.2	NCU04675.2	NCU07273.2	NCU09533.2
NCU00591.2	NCU02361.2	NCU04676.2	NCU07286.2	NCU09652.2
NCU00608.2	NCU02397.2	NCU04781.2	NCU07287.2b	NCU09682.2
NCU00611.2	NCU02453.1	NCU04801.2	NCU07310.2	NCU09705.2
NCU00642.2	NCU02455.2	NCU04815.2	NCU07311.2	NCU09763.2b
NCU00643.2	NCU02512.2	NCU04905.2	NCU07313.2b	NCU09783.2
NCU00663.2	NCU02582.2	NCU04906.2	NCU07392.2	NCU09821.2
NCU00673.2	NCU02583.2	NCU04908.2a	NCU07453.2	NCU09886.2
NCU00695.2	NCU02596.2	NCU04908.2b	NCU07624.2	NCU09923.2
NCU00709.2	NCU02606.2	NCU04909.2	NCU07705.2	NCU09924.2
NCU00809.2	NCU02637.2	NCU04910.2	NCU07706.2	NCU10020.2
NCU00810.2	NCU02653.2	NCU04918.2	NCU07723.2	NCU10021.2
NCU00821.2	NCU02654.2	NCU04930.2	NCU07771.2	NCU10040.2
NCU00864.2	NCU02701.2	NCU04963.2	NCU07788.2	NCU10045.2
NCU00870.2	NCU02730.2	NCU05134.2	NCU07883.2	NCU10107.2a
NCU00884.2	NCU02887.2	NCU05137.2	NCU07890.2a	NCU10398.2a
NCU00888.2b	NCU03013.2	NCU05143.2	NCU07941.2	NCU10398.2b
NCU00890.2	NCU03043.2	NCU05148.2	NCU07997.2	NCU10547.2a
NCU00891.2	NCU03086.2	NCU05159.2	NCU08017.2	NCU10656.2c
NCU00892.2	NCU03188.2	NCU05164.2	NCU08076.2	NCU10733.2a
NCU00937.2	NCU03216.2	NCU05170.2	NCU08092.2	NCU10966.2a
NCU00972.1	NCU03222.2	NCU05315.2	NCU08114.2	NCU10966.2b
NCU00988.2	NCU03253.2	NCU05395.2	NCU08130.2	NCU10997.2a
NCU00992.2	NCU03263.2	NCU05493.2	NCU08131.2	xnc066_080
NCU00999.1	NCU03281.2	NCU05498.2	NCU08189.2	NCU00132.2
NCU01045.2	NCU03322.2	NCU05627.2	NCU08190.2	NCU00157.2
NCU01068.2	NCU03323.2	NCU05751.2	NCU08224.2	NCU00442.2
NCU01107.2	NCU03550.2	NCU05755.2	NCU08273.2	NCU00987.2
NCU01132.2	NCU03558.2	NCU05770.2	NCU08282.2	NCU01517.2
NCU01140.2	NCU03588.2	NCU05832.2	NCU08330.2	NCU02042.2
NCU01145.2	NCU03598.2	NCU05837.2	NCU08331.2	NCU02128.2
NCU01181.2	NCU03415.2	NCU05850.2	NCU08333.2	NCU03156.2
NCU01230.2	NCU03523.2	NCU05853.2	NCU08351.2	NCU03284.2
NCU01231.2	NCU03605.2	NCU05897.2	NCU08356.2	NCU03518.2b
NCU01233.2	NCU03637.2	NCU05924.2	NCU08384.2	NCU03636.2
NCU01240.2	NCU03639.2	NCU05965.2	NCU08428.2	NCU04053.2
NCU01258.2	NCU03731.2	NCU05977.2	NCU08457.2	NCU04197.2
NCU01275.2	NCU03749.2	NCU05994.2	NCU08516.2	NCU04525.2
NCU01320.2	NCU03753.2	NCU06005.2	NCU08541.2	NCU04626.2
NCU01328.2	NCU03779.2	NCU06043.2	NCU08549.2	NCU05303.2
NCU01417.2	NCU03803.2	NCU06123.2	NCU08554.2	NCU06081.1
NCU01419.2	NCU03813.2	NCU06125.2	NCU08677.2	NCU06961.2
NCU01420.2	NCU03893.2	NCU06138.2	NCU08687.2	NCU07190.2
NCU01424.2	NCU03949.2	NCU06143.2	NCU08750.2	NCU07215.2
NCU01430.2	NCU03965.2	NCU06255.2	NCU08752.2	NCU07325.2
NCU01436.2	NCU04109.2	NCU06261.2	NCU08755.2	NCU09161.2
NCU01555.2	NCU04133.2	NCU06265.2	NCU08771.2	NCU09689.2
NCU01701.2	NCU04264.2	NCU06277.2	NCU08779.2	NCU09698.2
NCU01704.2	NCU04265.2	NCU06305.2	NCU08943.2	NCU09777.2
NCU01740.2	NCU04266.2	NCU06358.2	NCU08949.2	NCU09792.2
NCU01759.2	NCU04272.2	NCU06364.2	NCU08977.2	NCU09883.2
NCU01815.2	NCU04287.2	NCU06380.2	NCU09010.2	NCU09906.2
NCU01847.1	NCU04295.2	NCU06490.2	NCU09013.2	
NCU01866.2	NCU04298.2	NCU06597.2	NCU09027.2	
NCU01900.2	NCU04349.2a	NCU06603.2	NCU09034.2	
NCU01904.2	NCU04369.2	NCU06650.2	NCU09041.2	
NCU01905.2	NCU04371.2	NCU06652.2	NCU09043.2	

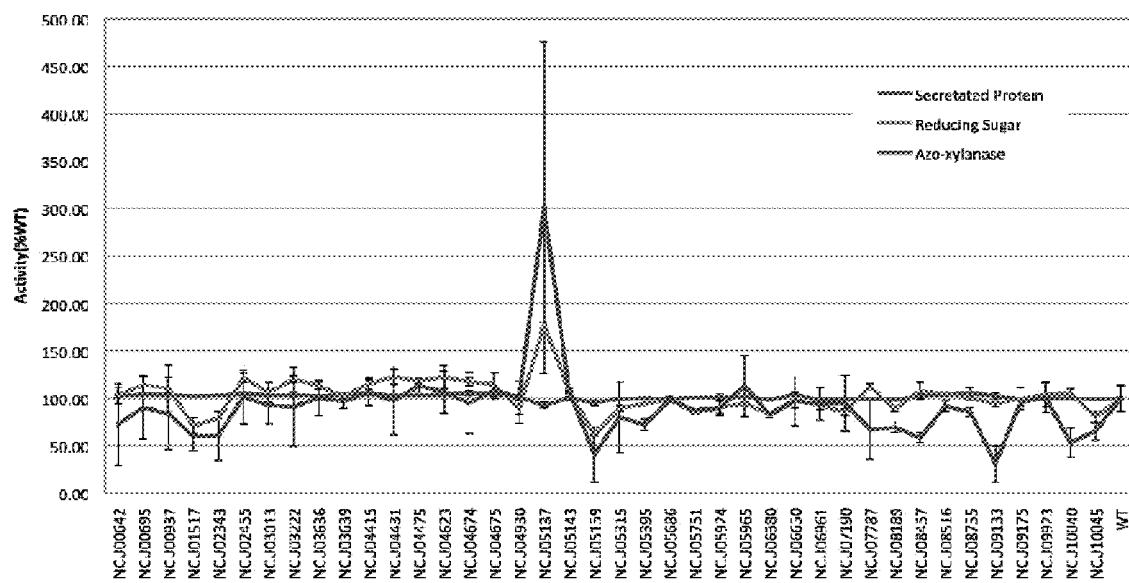
Figure 77

Figure 78**A**

	Bradford	CMC ^a activity
Wt-1	0.57	0.50
Wt-2	0.58	0.53
Ko-1	0.89	0.76
Ko-2	0.89	0.65
GFP	0.56	0.54
GFP	0.60	0.58

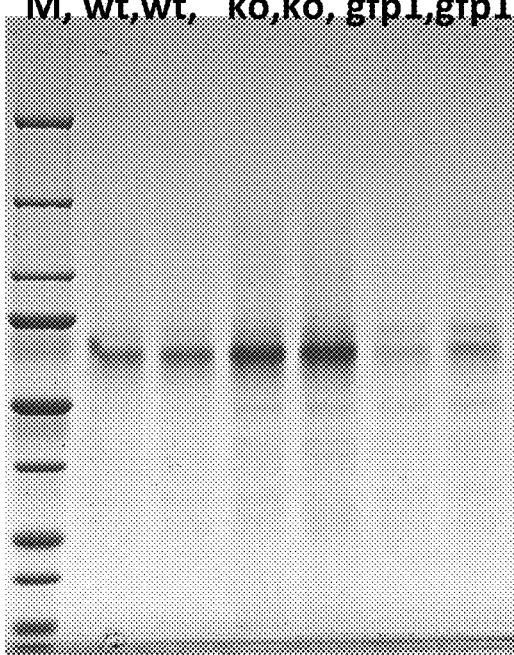
B**M, wt,wt, ko,ko, gfp1,gfp1**

Figure 79

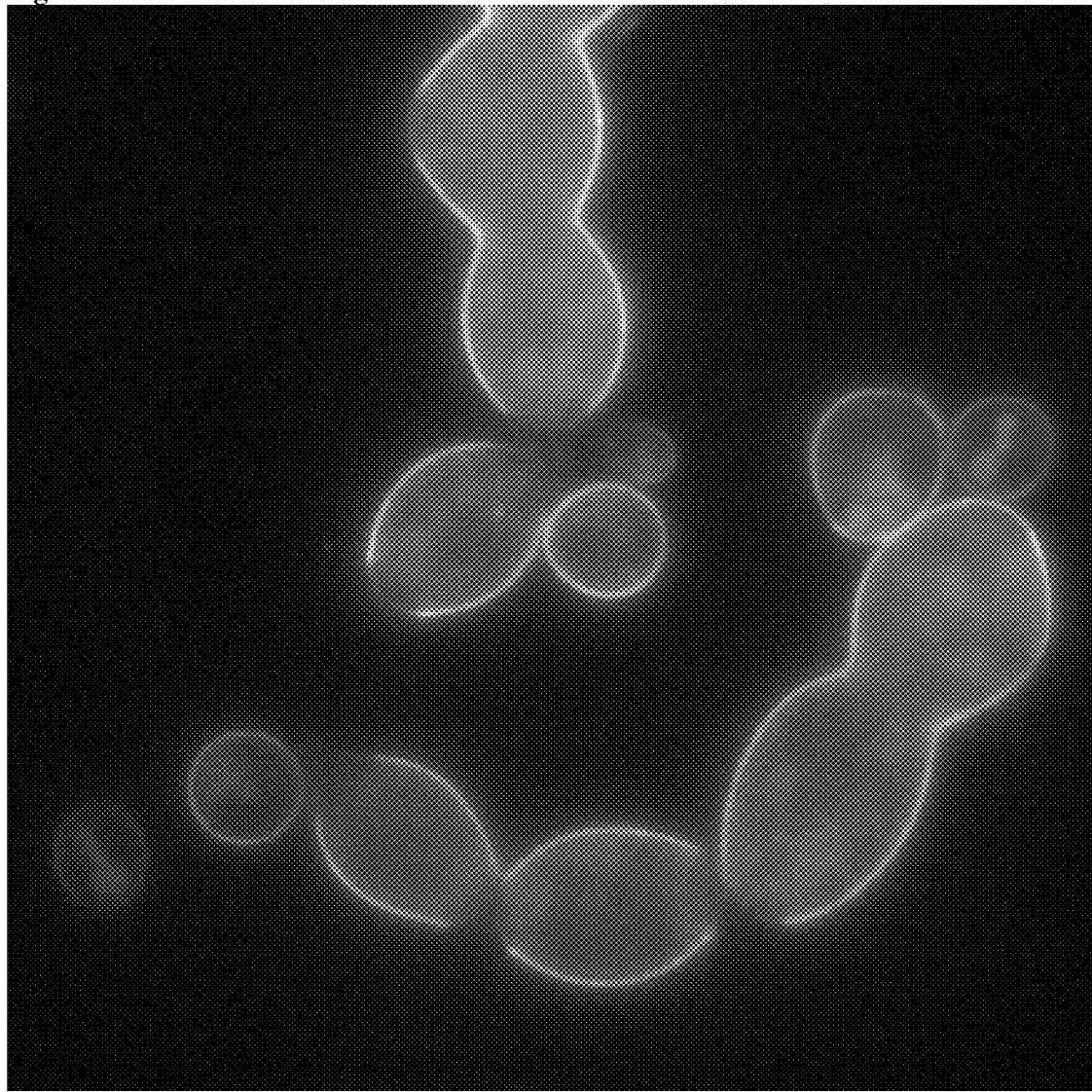
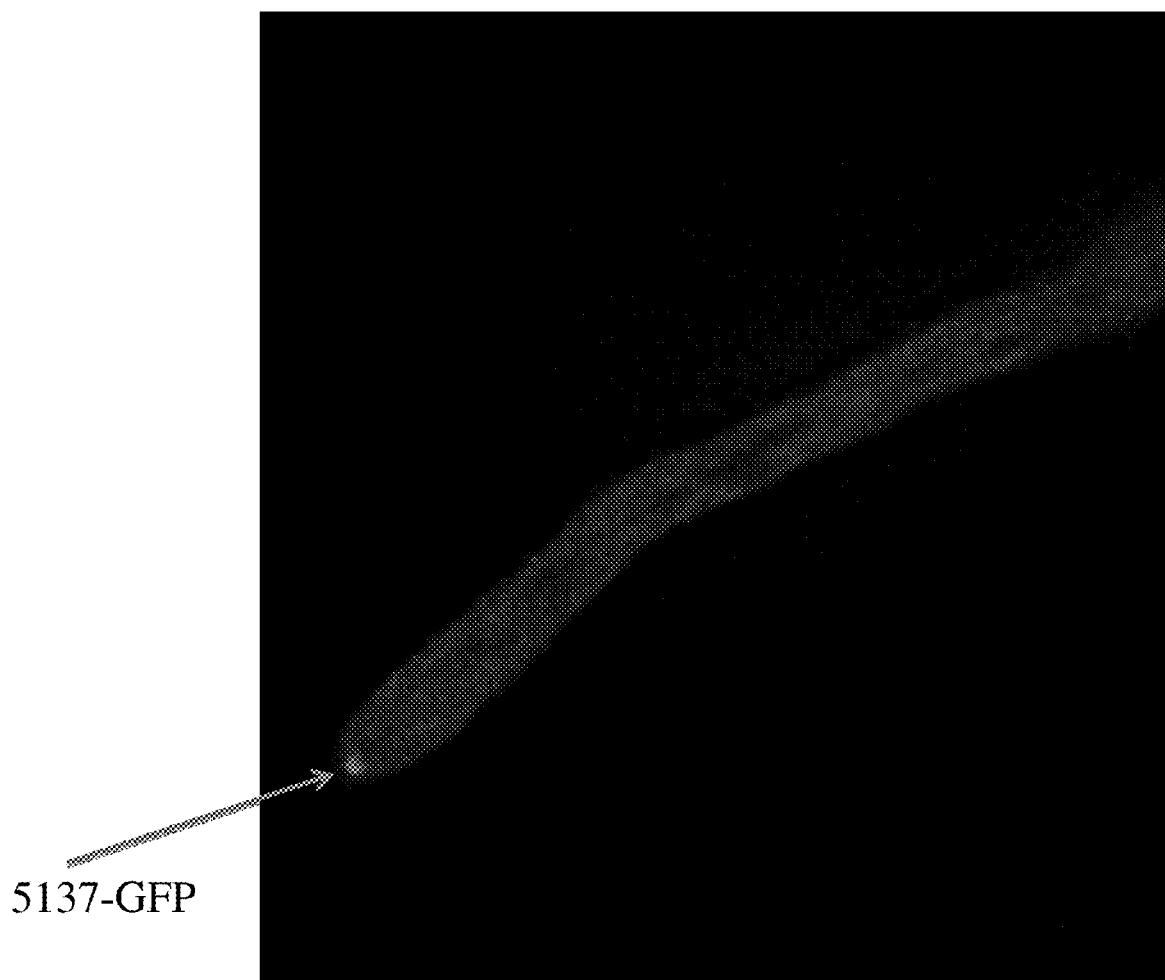


Figure 80



1

**METHODS AND COMPOSITIONS FOR
IMPROVING SUGAR TRANSPORT, MIXED
SUGAR FERMENTATION, AND
PRODUCTION OF BIOFUELS**

**CROSS-REFERENCE TO RELATED
APPLICATIONS**

This application claims the benefit of U.S. Provisional Application No. 61/285,526, filed Dec. 10, 2009, and U.S. Provisional Application No. 61/271,833, filed Jul. 24, 2009, both of which are hereby incorporated by reference in their entirety.

**SUBMISSION OF SEQUENCE LISTING ON
ASCII TEXT FILE**

The content of the following submission on ASCII text file is incorporated herein by reference in its entirety: a computer readable form (CRF) of the Sequence Listing (file name: 677792000100SeqList.txt, date recorded: Jul. 26, 2010, size: 104 KB).

FIELD OF THE INVENTION

The present disclosure relates to methods and compositions for increasing the transport of sugars into cells, for increasing growth of cells, for increasing synthesis of hydrocarbons and hydrocarbon derivatives, and for co-fermenting cellulose-derived and hemicellulose-derived sugars.

BACKGROUND OF THE INVENTION

Biofuels are under intensive investigation due to the increasing concerns about energy security, sustainability, and global climate change (Lynd et al., 1991). Bioconversion of plant-derived lignocellulosic materials into biofuels has been regarded as an attractive alternative to chemical production of fossil fuels (Lynd et al. 2008; Hahn-Hagerdal et al. 2006). Lignocellulosic biomass is composed of cellulose, hemicellulose, and lignin.

The engineering of microorganisms to perform the conversion of lignocellulosic biomass to ethanol efficiently remains a major goal of the biofuels field. Much research has been focused on genetically manipulating microorganisms that naturally ferment simple sugars to alcohol to express cellulases and other enzymes that would allow them to degrade lignocellulosic biomass polymers and generate ethanol within one cell. However, an area that has been less well studied is that of sugar transporters. An understanding of the regulation of sugar transport and the genetic engineering of microorganisms to have improved sugar-uptake ability will greatly improve efficiency (Stephanopoulos 2007). Furthermore, other types of proteins involved in the regulation of cellulase expression and activity remain to be fully explored.

Saccharomyces cerevisiae, also known as baker's yeast, has been used for bioconversion of hexose sugars into ethanol for thousands of years. It is also the most widely used microorganism for large scale industrial fermentation of D-glucose into ethanol. *S. cerevisiae* is a very suitable candidate for bioconversion of lignocellulosic biomass into biofuels (van Maris et al., 2006). It has a well-studied genetic and physiological background, ample genetic tools, and high tolerance to high ethanol concentration and inhibitors presented in lignocellulosic hydrolysates (Jeffries 2006). The low fermentation pH of *S. cerevisiae* can also prevent bacterial contamination during fermentation.

2

Unfortunately, wild type *S. cerevisiae* cannot utilize pentose sugars (Hector et al., 2008). To overcome this limitation, pentose utilization pathways from pentose-assimilating organisms have been introduced into *S. cerevisiae*, allowing 5 fermentation of D-xylose and L-arabinose (Hahn-Hagerdal et al., 2007; Brat et al., 2009; Wisselink et al., 2007, 2009; Wiedemann and Boles 2008; Karhumma et al., 2006). However, efficient conversion of pentose sugars into biofuels is limited by multiple issues including cellular redox imbalance, low influx of pentose phosphate pathway, and lack of efficient pentose transport into the cell (Hector et al., 2008).

In addition, both natural and engineered microorganisms show reduced ethanol tolerance during xylose fermentation as compared to glucose fermentation (Jeffries and Jin 2000).

15 Combined with the lower fermentation rate, the reduced ethanol tolerance during xylose fermentation poses a significant problem in fermentation of sugar mixtures containing the high concentrations of glucose (~70-100 g/L) and xylose (~40-60 g/L) present in cellulosic hydrolysates. Since micro- 20 organisms utilize glucose preferentially, at the time of glucose depletion (when cells begin to use xylose), the ethanol concentration is already high enough (~35-45 g/L) to further reduce the xylose fermentation rate. As a result, sequential utilization of xylose after glucose depletion because of "glucose 25 repression" is a significant challenge to be overcome in order to successfully utilize mixed sugars in cellulosic hydrolysates.

Thus, a need exists for the identification of additional genes that are critical for the degradation of lignocellulose and for 30 their use in the engineering of microorganisms for improved growth on lignocellulose and uptake of compounds resulting from lignocellulose degradation. A further need exists for improved methods of efficient conversion of pentose sugars into biofuels and of mixed sugar fermentation for the production of biofuels.

BRIEF SUMMARY OF THE INVENTION

In order to meet these needs, the invention described herein 40 provides methods of increasing transport of cellobextrin into a cell, methods of increasing growth of a cell on a medium containing cellobextrin, methods of co-fermenting cellulose-derived and hemicellulose-derived sugars, and methods of making hydrocarbons or hydrocarbon derivatives by providing 45 a host cell containing a recombinant polynucleotide encoding a polypeptide where the polypeptide transports cellobextrin into the cell. Further described are host cells containing a recombinant polynucleotide encoding a polypeptide where the polypeptide transports cellobextrin into the cell. 50 Further described herein are host cells containing a recombinant polynucleotide encoding a polypeptide where the polypeptide transports a pentose into the cell, methods of increasing transport of a pentose into a cell, methods of increasing growth of a cell on a medium containing pentose sugars, and methods of making hydrocarbons or hydrocarbon derivatives by providing a host cell containing a recombinant polynucleotide encoding a polypeptide where the polypeptide transports a pentose into the cell.

As used herein, cellobextrin refers to glucose polymers of 60 varying length and includes, without limitation, cellobiose (2 glucose monomers), cellotriose (3 glucose monomers), cellotetraose (4 glucose monomers), cellopentaose (5 glucose monomers), and cellohexaose (6 glucose monomers).

Thus one aspect includes methods of increasing transport 65 of cellobextrin into a cell, including providing a host cell, where the host cell contains a recombinant polynucleotide encoding a polypeptide containing transmembrane α -helix 1,

α -helix 2, α -helix 3, α -helix 4, α -helix 5, α -helix 6, α -helix 7, α -helix 8, α -helix 9, α -helix 10, α -helix 11, α -helix 12, and transmembrane α -helix 1 contains SEQ ID NO: 1, and culturing the cell in a medium such that the recombinant polynucleotide is expressed, where expression of the recombinant polynucleotide results in increased transport of cellobextrin into the cell compared with a cell that does not contain the recombinant polynucleotide.

Another aspect includes methods of increasing transport of cellobextrin into a cell, including providing a host cell, where the host cell contains a recombinant polynucleotide encoding a polypeptide containing transmembrane α -helix 1, α -helix 2, α -helix 3, α -helix 4, α -helix 5, α -helix 6, α -helix 7, α -helix 8, α -helix 9, α -helix 10, α -helix 11, α -helix 12, and transmembrane α -helix 2 contains SEQ ID NO: 2, and culturing the cell in a medium such that the recombinant polynucleotide is expressed, where expression of the recombinant polynucleotide results in increased transport of cellobextrin into the cell compared with a cell that does not contain the recombinant polynucleotide.

Another aspect includes methods of increasing transport of cellobextrin into a cell, including providing a host cell, where the host cell contains a recombinant polynucleotide encoding a polypeptide containing transmembrane α -helix 1, α -helix 2, α -helix 3, α -helix 4, α -helix 5, α -helix 6, α -helix 7, α -helix 8, α -helix 9, α -helix 10, α -helix 11, α -helix 12, and a loop connecting transmembrane α -helix 2 and transmembrane α -helix 3 contains SEQ ID NO: 3, and culturing the cell in a medium such that the recombinant polynucleotide is expressed, where expression of the recombinant polynucleotide results in increased transport of cellobextrin into the cell compared with a cell that does not contain the recombinant polynucleotide.

Another aspect includes methods of increasing transport of cellobextrin into a cell, including providing a host cell, where the host cell contains a recombinant polynucleotide encoding a polypeptide containing transmembrane α -helix 1, α -helix 2, α -helix 3, α -helix 4, α -helix 5, α -helix 6, α -helix 7, α -helix 8, α -helix 9, α -helix 10, α -helix 11, α -helix 12, and transmembrane α -helix 5 contains SEQ ID NO: 4, and culturing the cell in a medium such that the recombinant polynucleotide is expressed, where expression of the recombinant polynucleotide results in increased transport of cellobextrin into the cell compared with a cell that does not contain the recombinant polynucleotide.

Another aspect includes methods of increasing transport of cellobextrin into a cell, including providing a host cell, where the host cell contains a recombinant polynucleotide encoding a polypeptide containing transmembrane α -helix 1, α -helix 2, α -helix 3, α -helix 4, α -helix 5, α -helix 6, α -helix 7, α -helix 8, α -helix 9, α -helix 10, α -helix 11, α -helix 12, and transmembrane α -helix 6 contains SEQ ID NO: 5, and culturing the cell in a medium such that the recombinant polynucleotide is expressed, where expression of the recombinant polynucleotide results in increased transport of cellobextrin into the cell compared with a cell that does not contain the recombinant polynucleotide.

Another aspect includes methods of increasing transport of cellobextrin into a cell, including providing a host cell, where the host cell contains a recombinant polynucleotide encoding a polypeptide containing transmembrane α -helix 1, α -helix 2, α -helix 3, α -helix 4, α -helix 5, α -helix 6, α -helix 7, α -helix 8, α -helix 9, α -helix 10, α -helix 11, α -helix 12, and sequence between transmembrane α -helix 6 and transmembrane α -helix 7 contains SEQ ID NO: 6, and culturing the cell in a medium such that the recombinant polynucleotide is expressed, where expression of the recombinant polynucle-

otide results in increased transport of cellobextrin into the cell compared with a cell that does not contain the recombinant polynucleotide.

Another aspect includes methods of increasing transport of cellobextrin into a cell, including providing a host cell, where the host cell contains a recombinant polynucleotide encoding a polypeptide containing transmembrane α -helix 1, α -helix 2, α -helix 3, α -helix 4, α -helix 5, α -helix 6, α -helix 7, α -helix 8, α -helix 9, α -helix 10, α -helix 11, α -helix 12, and transmembrane α -helix 7 contains SEQ ID NO: 7, and culturing the cell in a medium such that the recombinant polynucleotide is expressed, where expression of the recombinant polynucleotide results in increased transport of cellobextrin into the cell compared with a cell that does not contain the recombinant polynucleotide.

Another aspect includes methods of increasing transport of cellobextrin into a cell, including providing a host cell, where the host cell contains a recombinant polynucleotide encoding a polypeptide containing transmembrane α -helix 1, α -helix 2, α -helix 3, α -helix 4, α -helix 5, α -helix 6, α -helix 7, α -helix 8, α -helix 9, α -helix 10, α -helix 11, α -helix 12, and transmembrane α -helix 10 and transmembrane α -helix 11 and the sequence between them contains SEQ ID NO: 8, and culturing the cell in a medium such that the recombinant polynucleotide is expressed, where expression of the recombinant polynucleotide results in increased transport of cellobextrin into the cell compared with a cell that does not contain the recombinant polynucleotide.

In certain embodiments that may be combined with any of the preceding aspects, the polypeptide has at least 29%, at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 99%, or at least 100% amino acid identity to NCU00801 or NCU08114. In certain embodiments that may be combined with any of the preceding embodiments, the host cell contains a second recombinant polynucleotide encoding at least a catalytic domain of a β -glucosidase. In certain embodiments that may be combined with the preceding embodiments having a host cell containing a second recombinant polynucleotide encoding at least a catalytic domain of a β -glucosidase, the β -glucosidase is from *Neurospora crassa*. In certain embodiments that may be combined with the preceding embodiments having a host cell containing a second recombinant polynucleotide encoding at least a catalytic domain of a β -glucosidase from *Neurospora crassa*, the β -glucosidase is encoded by NCU00130. In certain embodiments that may be combined with any of the preceding embodiments, the host cell further contains one or more recombinant polynucleotides where the one or more polynucleotides encode one or more enzymes involved in pentose utilization. In certain embodiments that may be combined with the preceding embodiments having a host cell further containing one or more recombinant polynucleotides where the one or more polynucleotides encode one or more enzymes involved in pentose utilization, the one or more enzymes are selected from one or more of the group consisting of L-arabinose isomerase, L-ribulokinase, L-ribulose-5-P 4 epimerase, xylose isomerase, xylulokinase, aldose reductase, L-arabinitol 4-dehydrogenase, L-xylulose reductase, and xylitol dehydrogenase. In certain embodiments that may be combined with any of the preceding embodiments, the host cell further contains a third recombinant polynucleotide where the third recombinant polynucleotide encodes a pentose transporter. In certain embodiments that may be combined with the preceding embodiments having the host cell further containing a third recombinant polynucleotide where

the third recombinant polynucleotide encodes a pentose transporter, the pentose transporter is selected from the group consisting of NCU00821, NCU04963, NCU06138, STL12/XUT6, SUT2, SUT3, XUT1, and XUT3.

Another aspect includes methods of increasing growth of a cell, including providing a host cell, where the host cell contains a recombinant polynucleotide encoding a polypeptide containing transmembrane α -helix 1, α -helix 2, α -helix 3, α -helix 4, α -helix 5, α -helix 6, α -helix 7, α -helix 8, α -helix 9, α -helix 10, α -helix 11, α -helix 12, and transmembrane α -helix 1 contains SEQ ID NO: 1, and the polypeptide is a celldextrin transporter, and culturing the host cell in a medium containing celldextrin, where the host cell grows at a faster rate in the medium than a cell that does not contain the recombinant polynucleotide.

Another aspect includes methods of increasing growth of a cell, including providing a host cell, where the host cell contains a recombinant polynucleotide encoding a polypeptide containing transmembrane α -helix 1, α -helix 2, α -helix 3, α -helix 4, α -helix 5, α -helix 6, α -helix 7, α -helix 8, α -helix 9, α -helix 10, α -helix 11, α -helix 12, and transmembrane α -helix 2 contains SEQ ID NO: 2, and the polypeptide is a celldextrin transporter, and culturing the host cell in a medium containing celldextrin, where the host cell grows at a faster rate in the medium than a cell that does not contain the recombinant polynucleotide.

Another aspect includes methods of increasing growth of a cell, including providing a host cell, where the host cell contains a recombinant polynucleotide encoding a polypeptide containing transmembrane α -helix 1, α -helix 2, α -helix 3, α -helix 4, α -helix 5, α -helix 6, α -helix 7, α -helix 8, α -helix 9, α -helix 10, α -helix 11, α -helix 12, and a loop connecting transmembrane α -helix 2 and transmembrane α -helix 3 contains SEQ ID NO: 3, and the polypeptide is a celldextrin transporter, and culturing the host cell in a medium containing celldextrin, where the host cell grows at a faster rate in the medium than a cell that does not contain the recombinant polynucleotide.

Another aspect includes methods of increasing growth of a cell, including providing a host cell, where the host cell contains a recombinant polynucleotide encoding a polypeptide containing transmembrane α -helix 1, α -helix 2, α -helix 3, α -helix 4, α -helix 5, α -helix 6, α -helix 7, α -helix 8, α -helix 9, α -helix 10, α -helix 11, α -helix 12, and transmembrane α -helix 5 contains SEQ ID NO: 4, and the polypeptide is a celldextrin transporter, and culturing the host cell in a medium containing celldextrin, where the host cell grows at a faster rate in the medium than a cell that does not contain the recombinant polynucleotide.

Another aspect includes methods of increasing growth of a cell, including providing a host cell, where the host cell contains a recombinant polynucleotide encoding a polypeptide containing transmembrane α -helix 1, α -helix 2, α -helix 3, α -helix 4, α -helix 5, α -helix 6, α -helix 7, α -helix 8, α -helix 9, α -helix 10, α -helix 11, α -helix 12, and transmembrane α -helix 6 contains SEQ ID NO: 5, and the polypeptide is a celldextrin transporter, and culturing the host cell in a medium containing celldextrin, where the host cell grows at a faster rate in the medium than a cell that does not contain the recombinant polynucleotide.

Another aspect includes methods of increasing growth of a cell, including providing a host cell, where the host cell contains a recombinant polynucleotide encoding a polypeptide containing transmembrane α -helix 1, α -helix 2, α -helix 3, α -helix 4, α -helix 5, α -helix 6, α -helix 7, α -helix 8, α -helix 9, α -helix 10, α -helix 11, α -helix 12, and sequence between transmembrane α -helix 6 and transmembrane α -helix 7 con-

tains SEQ ID NO: 6, and the polypeptide is a celldextrin transporter, and culturing the host cell in a medium containing celldextrin, where the host cell grows at a faster rate in the medium than a cell that does not contain the recombinant polynucleotide.

Another aspect includes methods of increasing growth of a cell, including providing a host cell, where the host cell contains a recombinant polynucleotide encoding a polypeptide containing transmembrane α -helix 1, α -helix 2, α -helix 3, α -helix 4, α -helix 5, α -helix 6, α -helix 7, α -helix 8, α -helix 9, α -helix 10, α -helix 11, α -helix 12, and transmembrane α -helix 7 contains SEQ ID NO: 7, and the polypeptide is a celldextrin transporter, and culturing the host cell in a medium containing celldextrin, where the host cell grows at a faster rate in the medium than a cell that does not contain the recombinant polynucleotide.

Another aspect includes methods of increasing growth of a cell, including providing a host cell, where the host cell contains a recombinant polynucleotide encoding a polypeptide containing transmembrane α -helix 1, α -helix 2, α -helix 3, α -helix 4, α -helix 5, α -helix 6, α -helix 7, α -helix 8, α -helix 9, α -helix 10, α -helix 11, α -helix 12, and transmembrane α -helix 10 and transmembrane α -helix 11 and the sequence between them contain SEQ ID NO: 8, and the polypeptide is a celldextrin transporter, and culturing the host cell in a medium containing celldextrin, where the host cell grows at a faster rate in the medium than a cell that does not contain the recombinant polynucleotide.

In certain embodiments that may be combined with any of the preceding aspects of increasing growth of cells, the polypeptide has at least 29%, at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 99%, or at least 100% amino acid identity to NCU00801 or NCU08114. In certain embodiments that may be combined with any of the preceding embodiments, the host cell contains an endogenous or a second recombinant polynucleotide where the polynucleotide encodes at least a catalytic domain of a β -glucosidase. In certain embodiments that may be combined with the preceding embodiments having a host cell containing an endogenous or a second recombinant polynucleotide where the polynucleotide encodes at least a catalytic domain of a β -glucosidase, the β -glucosidase is from *Neurospora crassa*.

In certain embodiments that may be combined with the preceding embodiments having a host cell containing an endogenous or a second recombinant polynucleotide where the polynucleotide encodes at least a catalytic domain of a β -glucosidase from *Neurospora crassa*, the β -glucosidase is encoded by NCU00130.

Another aspect includes methods of co-fermenting cellulose-derived and hemicellulose-derived sugars, containing providing a host cell, where the host cell contains a first recombinant polynucleotide encoding a celldextrin transporter and a second recombinant polynucleotide encoding a catalytic domain of a β -glucosidase, and culturing the host cell in a medium containing a cellulose-derived sugar and a hemicellulose-derived sugar, where expression of the recombinant polynucleotides enables co-fermentation of the cellulose-derived sugar and the hemicellulose-derived sugar. In certain embodiments, the first recombinant polynucleotide encodes a polypeptide containing transmembrane α -helix 1, α -helix 2, α -helix 3, α -helix 4, α -helix 5, α -helix 6, α -helix 7, α -helix 8, α -helix 9, α -helix 10, α -helix 11, α -helix 12, and transmembrane α -helix 1 contains SEQ ID NO: 1. In certain embodiments, the first recombinant polynucleotide encodes a polypeptide containing transmembrane α -helix 1, α -helix 2,

α -helix 3, α -helix 4, α -helix 5, α -helix 6, α -helix 7, α -helix 8, α -helix 9, α -helix 10, α -helix 11, α -helix 12, and transmembrane α -helix 2 contains SEQ ID NO: 2. In certain embodiments, the first recombinant polynucleotide encodes a polypeptide containing transmembrane α -helix 1, α -helix 2, α -helix 3, α -helix 4, α -helix 5, α -helix 6, α -helix 7, α -helix 8, α -helix 9, α -helix 10, α -helix 11, α -helix 12, and a loop connecting transmembrane α -helix 2 and transmembrane α -helix 3 contains SEQ ID NO: 3. In certain embodiments, the first recombinant polynucleotide encodes a polypeptide containing transmembrane α -helix 1, α -helix 2, α -helix 3, α -helix 4, α -helix 5, α -helix 6, α -helix 7, α -helix 8, α -helix 9, α -helix 10, α -helix 11, α -helix 12, and transmembrane α -helix 5 contains SEQ ID NO: 4. In certain embodiments, the first recombinant polynucleotide encodes a polypeptide containing transmembrane α -helix 1, α -helix 2, α -helix 3, α -helix 4, α -helix 5, α -helix 6, α -helix 7, α -helix 8, α -helix 9, α -helix 10, α -helix 11, α -helix 12, and transmembrane α -helix 6 contains SEQ ID NO: 5. In certain embodiments, the first recombinant polynucleotide encodes a polypeptide containing transmembrane α -helix 1, α -helix 2, α -helix 3, α -helix 4, α -helix 5, α -helix 6, α -helix 7, α -helix 8, α -helix 9, α -helix 10, α -helix 11, α -helix 12, and sequence between transmembrane α -helix 6 and transmembrane α -helix 7 contains SEQ ID NO: 6. In certain embodiments, the first recombinant polynucleotide encodes a polypeptide containing transmembrane α -helix 1, α -helix 2, α -helix 3, α -helix 4, α -helix 5, α -helix 6, α -helix 7, α -helix 8, α -helix 9, α -helix 10, α -helix 11, α -helix 12, and transmembrane α -helix 7 contains SEQ ID NO: 7. In certain embodiments, the first recombinant polynucleotide encodes a polypeptide containing transmembrane α -helix 1, α -helix 2, α -helix 3, α -helix 4, α -helix 5, α -helix 6, α -helix 7, α -helix 8, α -helix 9, α -helix 10, α -helix 11, α -helix 12, and transmembrane α -helix 10 and transmembrane α -helix 11 and the sequence between them contain SEQ ID NO: 8. In certain embodiments that may be combined with any of the preceding embodiments, the polypeptide has at least 29%, at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 99%, or at least 100% amino acid identity to NCU00801 or NCU08114. In certain embodiments that may be combined with any of the preceding embodiments, the β -glucosidase is from *Neurospora crassa*. In certain embodiments that may be combined with the preceding embodiments having a host cell containing a second recombinant polynucleotide encoding a catalytic domain of a β -glucosidase from *Neurospora crassa*, the β -glucosidase is encoded by NCU00130. In certain embodiments that may be combined with any of the preceding embodiments, the host cell further contains one or more recombinant polynucleotides where the one or more polynucleotides encode one or more enzymes involved in pentose utilization. In certain embodiments that may be combined with the preceding embodiments having a host cell further containing one or more recombinant polynucleotides where the one or more polynucleotides encode one or more enzymes involved in pentose utilization, the one or more enzymes are selected from one or more of the group consisting of L-arabinose isomerase, L-ribulokinase, L-ribulose-5-P 4 epimerase, xylose isomerase, xylulokinase, aldose reductase, L-arabinitol 4-dehydrogenase, L-xylulose reductase, and xylitol dehydrogenase. In certain embodiments that may be combined with any of the preceding embodiments, the host cell further contains a third recombinant polynucleotide where the third recombinant polynucleotide encodes a pentose transporter. In certain embodiments that may be com-

bined with the preceding embodiments having the host cell further containing a third recombinant polynucleotide where the third recombinant polynucleotide encodes a pentose transporter, the pentose transporter is selected from the group consisting of NCU00821, NCU04963, NCU06138, STL12/XUT6, SUT2, SUT3, XUT1, and XUT3. In certain embodiments that may be combined with any of the preceding embodiments, the cellulose-derived sugar is selected from the group consisting of cellobiose, cellotriose, and cellotetraose, and the hemicellulose-derived sugar is xylose.

Another aspect includes methods of increasing the synthesis of hydrocarbons or hydrocarbon derivatives by a host cell, including providing a host cell, where the host cell contains a recombinant polynucleotide encoding a polypeptide containing transmembrane α -helix 1, α -helix 2, α -helix 3, α -helix 4, α -helix 5, α -helix 6, α -helix 7, α -helix 8, α -helix 9, α -helix 10, α -helix 11, α -helix 12, and transmembrane α -helix 1 contains SEQ ID NO: 1, and the polypeptide transports cellobextrin into the host cell for the synthesis of hydrocarbons or hydrocarbon derivatives, and culturing the host cell in a medium containing cellobextrin or a source of cellobextrin to increase the synthesis of hydrocarbons or hydrocarbon derivatives by the host cell, where transport of cellobextrin into the cell is increased upon expression of the recombinant polynucleotide.

Another aspect includes methods of increasing the synthesis of hydrocarbons or hydrocarbon derivatives by a host cell, containing providing a host cell, where the host cell contains a recombinant polynucleotide encoding a polypeptide containing transmembrane α -helix 1, α -helix 2, α -helix 3, α -helix 4, α -helix 5, α -helix 6, α -helix 7, α -helix 8, α -helix 9, α -helix 10, α -helix 11, α -helix 12, and transmembrane α -helix 2 contains SEQ ID NO: 2, and the polypeptide transports cellobextrin into the host cell for the synthesis of hydrocarbons or hydrocarbon derivatives, and culturing the host cell in a medium containing cellobextrin or a source of cellobextrin to increase the synthesis of hydrocarbons or hydrocarbon derivatives by the host cell, where transport of cellobextrin into the cell is increased upon expression of the recombinant polynucleotide.

Another aspect includes methods of increasing the synthesis of hydrocarbons or hydrocarbon derivatives by a host cell, including providing a host cell, where the host cell contains a recombinant polynucleotide encoding a polypeptide containing transmembrane α -helix 1, α -helix 2, α -helix 3, α -helix 4, α -helix 5, α -helix 6, α -helix 7, α -helix 8, α -helix 9, α -helix 10, α -helix 11, α -helix 12, and a loop connecting transmembrane α -helix 2 and transmembrane α -helix 3 contains SEQ ID NO: 3, and the polypeptide transports cellobextrin into the host cell for the synthesis of hydrocarbons or hydrocarbon derivatives, and culturing the host cell in a medium containing cellobextrin or a source of cellobextrin to increase the synthesis of hydrocarbons or hydrocarbon derivatives by the host cell, where transport of cellobextrin into the cell is increased upon expression of the recombinant polynucleotide.

Another aspect includes methods of increasing the synthesis of hydrocarbons or hydrocarbon derivatives by a host cell, including providing a host cell, where the host cell contains a recombinant polynucleotide encoding a polypeptide containing transmembrane α -helix 1, α -helix 2, α -helix 3, α -helix 4, α -helix 5, α -helix 6, α -helix 7, α -helix 8, α -helix 9, α -helix 10, α -helix 11, α -helix 12, and transmembrane α -helix 5 contains SEQ ID NO: 4, and the polypeptide transports cellobextrin into the host cell for the synthesis of hydrocarbons or hydrocarbon derivatives, and culturing the host cell in a medium containing cellobextrin or a source of cellobextrin to

increase the synthesis of hydrocarbons or hydrocarbon derivatives by the host cell, where transport of celldextrin into the cell is increased upon expression of the recombinant polynucleotide.

Another aspect includes methods of increasing the synthesis of hydrocarbons or hydrocarbon derivatives by a host cell, including providing a host cell, where the host cell contains a recombinant polynucleotide encoding a polypeptide containing transmembrane α -helix 1, α -helix 2, α -helix 3, α -helix 4, α -helix 5, α -helix 6, α -helix 7, α -helix 8, α -helix 9, α -helix 10, α -helix 11, α -helix 12, and transmembrane α -helix 6 contains SEQ ID NO: 5, and the polypeptide transports celldextrin into the host cell for the synthesis of hydrocarbons or hydrocarbon derivatives, and culturing the host cell in a medium containing celldextrin or a source of celldextrin to increase the synthesis of hydrocarbons or hydrocarbon derivatives by the host cell, where transport of celldextrin into the cell is increased upon expression of the recombinant polynucleotide.

Another aspect includes methods of increasing the synthesis of hydrocarbons or hydrocarbon derivatives by a host cell, including providing a host cell, where the host cell contains a recombinant polynucleotide encoding a polypeptide containing transmembrane α -helix 1, α -helix 2, α -helix 3, α -helix 4, α -helix 5, α -helix 6, α -helix 7, α -helix 8, α -helix 9, α -helix 10, α -helix 11, α -helix 12, and sequence between transmembrane α -helix 6 and transmembrane α -helix 7 contains SEQ ID NO: 6, and the polypeptide transports celldextrin into the host cell for the synthesis of hydrocarbons or hydrocarbon derivatives, and culturing the host cell in a medium containing celldextrin or a source of celldextrin to increase the synthesis of hydrocarbons or hydrocarbon derivatives by the host cell, where transport of celldextrin into the cell is increased upon expression of the recombinant polynucleotide.

Another aspect includes methods of increasing the synthesis of hydrocarbons or hydrocarbon derivatives by a host cell, including providing a host cell, where the host cell contains a recombinant polynucleotide encoding a polypeptide containing transmembrane α -helix 1, α -helix 2, α -helix 3, α -helix 4, α -helix 5, α -helix 6, α -helix 7, α -helix 8, α -helix 9, α -helix 10, α -helix 11, α -helix 12, and transmembrane α -helix 7 contains SEQ ID NO: 7, and the polypeptide transports celldextrin into the host cell for the synthesis of hydrocarbons or hydrocarbon derivatives, and culturing the host cell in a medium containing celldextrin or a source of celldextrin to increase the synthesis of hydrocarbons or hydrocarbon derivatives by the host cell, where transport of celldextrin into the cell is increased upon expression of the recombinant polynucleotide.

Another aspect includes methods of increasing the synthesis of hydrocarbons or hydrocarbon derivatives by a host cell, including providing a host cell, where the host cell contains a recombinant polynucleotide encoding a polypeptide containing transmembrane α -helix 1, α -helix 2, α -helix 3, α -helix 4, α -helix 5, α -helix 6, α -helix 7, α -helix 8, α -helix 9, α -helix 10, α -helix 11, α -helix 12, and transmembrane α -helix 10 and transmembrane α -helix 11 and the sequence between them contain SEQ ID NO: 8, and the polypeptide transports celldextrin into the host cell for the synthesis of hydrocarbons or hydrocarbon derivatives, and culturing the host cell in a medium containing celldextrin or a source of celldextrin to increase the synthesis of hydrocarbons or hydrocarbon derivatives by the host cell, where transport of celldextrin into the cell is increased upon expression of the recombinant polynucleotide.

In certain embodiments that may be combined with any of the preceding aspects increasing the synthesis of hydrocarbons or hydrocarbon derivatives, the polypeptide has at least 29%, at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 99%, or at least 100% amino acid identity to NCU00801 or NCU08114. In certain embodiments that may be combined with any of the preceding embodiments, 5 the host cell further contains a second recombinant polynucleotide where the polynucleotide encodes at least a catalytic domain of a β -glucosidase. In certain embodiments that may be combined with preceding embodiments having the host cell further containing a second recombinant polynucleotide where the 10 polynucleotide encodes at least a catalytic domain of a β -glucosidase from *Neurospora crassa*, the β -glucosidase is encoded by NCU00130. In certain embodiments that may be combined with any of the preceding embodiments, the source of the celldextrin contains cellulose. In certain embodiments that may be combined with any of the preceding embodiments, the hydrocarbons or hydrocarbon derivatives can be used as fuel. In certain embodiments that may be combined with the preceding embodiments having the hydrocarbons or hydrocarbon derivatives used as fuel, the hydrocarbons or hydrocarbon derivatives contain ethanol. In certain embodiments that may be combined with the preceding embodiments having the hydrocarbons or hydrocarbon derivatives used as fuel, the hydrocarbons or hydrocarbon derivatives contain butanol.

15 In certain embodiments that may be combined with any of the preceding aspects, the medium contains a cellulase-containing enzyme mixture from an altered organism, where the cellulase-containing mixture has reduced β -glucosidase activity compared to a cellulase-containing mixture from an unaltered organism. In certain embodiments that may be combined with any of the preceding aspects, the host cell is selected from the group consisting of *Saccharomyces* sp., *Saccharomyces cerevisiae*, *Saccharomyces monacensis*, *Saccharomyces bayanus*, *Saccharomyces pastorianus*, *Saccharomyces carlsbergensis*, *Saccharomyces pombe*, *Kluyveromyces* sp., *Kluyveromyces marxianus*, *Kluyveromyces lactis*, *Kluyveromyces fragilis*, *Pichia stipitis*, *Sporotrichum thermophile*, *Candida shehatae*, *Candida tropicalis*, *Neurospora crassa*, *Zymomonas mobilis*, *Clostridium* sp., *Clostridium phytofermentans*, *Clostridium thermocellum*, *Clostridium beijerinckii*, *Clostridium acetobutylicum*, *Moorella thermoacetica*, *Escherichia coli*, *Klebsiella oxytoca*, *Thermoanaerobacter saccharolyticum*, and *Bacillus subtilis*. In certain embodiments that may be combined with any of the preceding aspects, celldextrin is selected from one or more of the group consisting of cellobiose, celotriose, and cellotetraose.

Another aspect includes host cells containing a recombinant polynucleotide encoding a polypeptide having transmembrane α -helix 1, α -helix 2, α -helix 3, α -helix 4, α -helix 5, α -helix 6, α -helix 7, α -helix 8, α -helix 9, α -helix 10, α -helix 11, α -helix 12, an intracellular N-terminus, an intracellular C-terminus, and a sequence selected from the group consisting of SEQ ID NO: 1 in transmembrane α -helix 1, 60 SEQ ID NO: 2 in transmembrane α -helix 2, SEQ ID NO: 3 in a loop connecting transmembrane α -helix 2 and transmembrane α -helix 3, SEQ ID NO: 4 in transmembrane α -helix 5,

11

SEQ ID NO: 5 in transmembrane α -helix 6, SEQ ID NO: 6 in the sequence between transmembrane α -helix 6 and transmembrane α -helix 7, SEQ ID NO: 7 in transmembrane α -helix 7, and SEQ ID NO: 8 in transmembrane α -helix 10 and transmembrane α -helix 11 and the sequence between them, where the polypeptide is a cellobextrin transporter. In certain embodiments, the polypeptide has at least 29%, at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 99%, or at least 100% amino acid identity to NCU00801 or NCU08114. In certain embodiments that may be combined with either of the preceding embodiments, the host cell further contains a second recombinant polynucleotide where the second recombinant polynucleotide encodes a catalytic domain of a β -glucosidase. In certain embodiments that may be combined with preceding embodiments having the host cell further containing a second recombinant polynucleotide where the second recombinant polynucleotide encodes a catalytic domain of a β -glucosidase, the β -glucosidase is from *Neurospora crassa*. In certain embodiments that may be combined with the preceding embodiments having the host cell further containing a second recombinant polynucleotide where the second recombinant polynucleotide encodes a catalytic domain of a β -glucosidase from *Neurospora crassa*, the β -glucosidase is encoded by NCU00130. In certain embodiments that may be combined with any of the preceding embodiments, the host cell further contains one or more recombinant polynucleotides where the one or more polynucleotides encode one or more enzymes involved in pentose utilization. In certain embodiments that may be combined with the preceding embodiments having the host cell further containing one or more recombinant polynucleotides where the one or more polynucleotides encode one or more enzymes involved in pentose utilization, the one or more enzymes are selected from one or more of the group consisting of L-arabinose isomerase, L-ribulokinase, L-ribulose-5-P 4 epimerase, xylose isomerase, xylulokinase, aldose reductase, L-arabinitol 4-dehydrogenase, L-xylulose reductase, and xylitol dehydrogenase. In certain embodiments that may be combined with any of the preceding embodiments, the host cell further contains a third recombinant polynucleotide where the third recombinant polynucleotide encodes a pentose transporter. In certain embodiments that may be combined with the preceding embodiment having the host cell further containing a third recombinant polynucleotide where the third recombinant polynucleotide encodes a pentose transporter, the pentose transporter is selected from the group consisting of NCU00821, NCU04963, NCU06138, STL12/XUT6, SUT2, SUT3, XUT1, and XUT3.

In certain embodiments that may be combined with any of the preceding aspects, the host cell further contains one or more inducible promoters operably linked to the one or more recombinant polynucleotides.

Another aspect includes a host cell containing a recombinant polynucleotide encoding a polypeptide selected from the group consisting of NCU00821 and STL12/XUT6, where the polypeptide transports xylose into the cell.

Another aspect includes a host cell containing a recombinant polynucleotide encoding a XUT1 polypeptide, where the polypeptide transports arabinose into the cell.

Another aspect includes a host cell containing a recombinant polynucleotide encoding an NCU06138 polypeptide, where the polypeptide transports arabinose and glucose into the cell.

Another aspect includes a host cell containing a recombinant polynucleotide encoding a polypeptide selected from the

12

group consisting of SUT2, SUT3, and XUT3, where the polypeptide transports xylose and glucose into the cell.

Another aspect includes a host cell containing a recombinant polynucleotide encoding an NCU04963 polypeptide, where the polypeptide transports xylose, arabinose, and glucose into the cell.

In certain embodiments that may be combined with any of the preceding aspects having a host cell containing a recombinant polynucleotide encoding a pentose transporter, the host cell further contains one or more recombinant polynucleotides where the one or more polynucleotides encode one or more enzymes involved in pentose utilization. In certain embodiments that may be combined with the preceding embodiment having the host cell further containing one or more recombinant polynucleotides where the one or more polynucleotides encode one or more enzymes involved in pentose utilization, the one or more enzymes are selected from one or more of the group consisting of L-arabinose isomerase, L-ribulokinase, L-ribulose-5-P 4 epimerase, xylose isomerase, xylulokinase, aldose reductase, L-arabinitol 4-dehydrogenase, L-xylulose reductase, and xylitol dehydrogenase.

Another aspect includes methods of increasing transport of xylose into a cell, including providing a host cell, where the host cell contains a recombinant polynucleotide encoding a polypeptide selected from the group consisting of NCU00821 and STL12/XUT6, and culturing the cell such that the recombinant polynucleotide is expressed, where expression of the recombinant polynucleotide results in increased transport of xylose into the cell compared with a cell that does not contain the recombinant polynucleotide.

Another aspect includes methods of increasing transport of arabinose into a cell, including providing a host cell, where the host cell contains a recombinant polynucleotide encoding a XUT1 polypeptide, and culturing the cell such that the recombinant polynucleotide is expressed, where expression of the recombinant polynucleotide results in increased transport of arabinose into the cell compared with a cell that does not contain the recombinant polynucleotide.

Another aspect includes methods of increasing transport of arabinose or glucose into a cell, including providing a host cell, where the host cell contains a recombinant polynucleotide encoding a NCU06138 polypeptide, and culturing the cell such that the recombinant polynucleotide is expressed, where expression of the recombinant polynucleotide results in increased transport of arabinose or glucose into the cell compared with a cell that does not contain the recombinant polynucleotide.

Another aspect includes methods of increasing transport of xylose or glucose into a cell, including providing a host cell, where the host cell contains a recombinant polynucleotide encoding a polypeptide selected from the group consisting of SUT2, SUT3, and XUT3, and culturing the cell such that the recombinant polynucleotide is expressed, where expression of the recombinant polynucleotide results in increased transport of xylose or glucose into the cell compared with a cell that does not contain the recombinant polynucleotide.

Another aspect includes methods of increasing transport of xylose, arabinose, or glucose into a cell, including providing a host cell, where the host cell contains a recombinant polynucleotide encoding a NCU04963 polypeptide, and culturing the cell such that the recombinant polynucleotide is expressed, where expression of the recombinant polynucleotide results in increased transport of xylose, arabinose, or glucose into the cell compared with a cell that does not contain the recombinant polynucleotide.

In certain embodiments that may be combined with any of the preceding aspects of increasing transport of xylose, arabinose, or glucose into cells, the method further includes one or more recombinant polynucleotides where the one or more polynucleotides encode one or more enzymes involved in pentose utilization. In certain embodiments that may be combined with the preceding embodiments having the method further including one or more recombinant polynucleotides where the one or more polynucleotides encode one or more enzymes involved in pentose utilization, the one or more enzymes are selected from one or more of the group consisting of L-arabinose isomerase, L-ribulokinase, L-ribulose-5-P 4 epimerase, xylose isomerase, xylulokinase, aldose reductase, L-arabinitol 4-dehydrogenase, L-xylulose reductase, and xylitol dehydrogenase.

Another aspect includes methods of increasing growth of a cell, including providing a host cell, where the host cell contains a recombinant polynucleotide where the polynucleotide encodes a polypeptide selected from the group consisting of NCU00821 and STL12/XUT6, and the polypeptide transports xylose, and culturing the host cell in a medium containing xylose, where the host cell grows at a faster rate in the medium than a cell that does not contain the recombinant polynucleotide.

Another aspect includes methods of increasing growth of a cell, including providing a host cell, where the host cell contains a recombinant polynucleotide where the polynucleotide encodes a XUT1 polypeptide, and the polypeptide transports arabinose, and culturing the host cell in a medium containing arabinose, where the host cell grows at a faster rate in the medium than a cell that does not contain the recombinant polynucleotide.

Another aspect includes method of increasing growth of a cell, including providing a host cell, where the host cell contains a recombinant polynucleotide where the polynucleotide encodes an NCU06138 polypeptide, and the polypeptide transports arabinose and glucose, and culturing the host cell in a medium containing arabinose or glucose, where the host cell grows at a faster rate in the medium than a cell that does not contain the recombinant polynucleotide.

Another aspect includes methods of increasing growth of a cell, including providing a host cell, where the host cell contains a recombinant polynucleotide where the polynucleotide encodes a polypeptide selected from the group consisting of SUT2, SUT3, and XUT3, and the polypeptide transports xylose and glucose, and culturing the host cell in a medium including xylose or glucose, where the host cell grows at a faster rate in the medium than a cell that does not contain the recombinant polynucleotide.

Another aspect includes methods of increasing growth of a cell, including providing a host cell, where the host cell contains a recombinant polynucleotide where the polynucleotide encodes a NCU04963 polypeptide, and the polypeptide transports xylose, arabinose, and glucose, and culturing the host cell in a medium containing xylose, arabinose, or glucose, where the host cell grows at a faster rate in the medium than a cell that does not contain the recombinant polynucleotide.

In certain embodiments that may be combined with the preceding aspects of increasing growth of cells by culturing a host cell containing a recombinant polynucleotide encoding a polypeptide that transports xylose and/or arabinose and/or glucose, the host cell further contains one or more endogenous or recombinant polynucleotides encoding one or more enzymes involved in pentose utilization. In certain embodiments that may be combined with the preceding embodiments having the host cell further containing one or more endogenous or recombinant polynucleotides encoding one or

more enzymes involved in pentose utilization, the one or more enzymes are selected from one or more of the group consisting of L-arabinose isomerase, L-ribulokinase, L-ribulose-5-P 4 epimerase, xylose isomerase, xylulokinase, aldose reductase, L-arabinitol 4-dehydrogenase, L-xylulose reductase, and xylitol dehydrogenase.

Another aspect includes methods of increasing the synthesis of hydrocarbons or hydrocarbon derivatives by a host cell, including providing a host cell, where the host cell contains a recombinant polynucleotide encoding a polypeptide selected from the group consisting of NCU00821 and STL12/XUT6, where the polypeptide transports xylose into the host cell for the synthesis of hydrocarbons or hydrocarbon derivatives, and culturing the host cell in a medium containing xylose or a source of xylose to increase the synthesis of hydrocarbons or hydrocarbon derivatives by the host cell, where transport of xylose into the cell is increased upon expression of the recombinant polynucleotide.

Another aspect includes methods of increasing the synthesis of hydrocarbons or hydrocarbon derivatives by a host cell, including providing a host cell, where the host cell contains a recombinant polynucleotide encoding a XUT1 polypeptide, where the polypeptide transports arabinose into the host cell for the synthesis of hydrocarbons or hydrocarbon derivatives, and culturing the host cell in a medium containing arabinose or a source of arabinose to increase the synthesis of hydrocarbons or hydrocarbon derivatives by the host cell, where transport of arabinose into the cell is increased upon expression of the recombinant polynucleotide.

Another aspect includes methods of increasing the synthesis of hydrocarbons or hydrocarbon derivatives by a host cell, including providing a host cell, where the host cell contains a recombinant polynucleotide encoding an NCU06138 polypeptide, where the polypeptide transports arabinose or glucose into the host cell for the synthesis of hydrocarbons or hydrocarbon derivatives, and culturing the host cell in a medium containing arabinose or glucose or a source of arabinose or glucose to increase the synthesis of hydrocarbons or hydrocarbon derivatives by the host cell, where transport of arabinose or glucose into the cell is increased upon expression of the recombinant polynucleotide.

Another aspect includes method of increasing the synthesis of hydrocarbons or hydrocarbon derivatives by a host cell, including providing a host cell, where the host cell contains a recombinant polynucleotide encoding a polypeptide selected from the group consisting of SUT2, SUT3, and XUT3, where the polypeptide transports xylose or glucose into the host cell for the synthesis of hydrocarbons or hydrocarbon derivatives, and culturing the host cell in a medium containing xylose or glucose or a source of xylose or glucose to increase the synthesis of hydrocarbons or hydrocarbon derivatives by the host cell, where transport of xylose or glucose into the cell is increased upon expression of the recombinant polynucleotide.

Another aspect includes methods of increasing the synthesis of hydrocarbons or hydrocarbon derivatives by a host cell, including providing a host cell, where the host cell contains a recombinant polynucleotide encoding an NCU04963 polypeptide, where the polypeptide transports xylose, arabinose, or glucose into the host cell for the synthesis of hydrocarbons or hydrocarbon derivatives, and culturing the host cell in a medium containing xylose, arabinose, or glucose or a source of xylose, arabinose, or glucose to increase the synthesis of hydrocarbons or hydrocarbon derivatives by the host cell, where transport of xylose, arabinose, or glucose into the cell is increased upon expression of the recombinant polynucleotide.

15

In certain embodiments that may combine any of the preceding aspects of increasing the synthesis of hydrocarbons or hydrocarbon derivatives by culturing a host cell containing a recombinant polynucleotide encoding a polypeptide that transports glucose, the source of glucose contains cellulose. In certain embodiments that may combine any of the preceding embodiments, the source of xylose or arabinose contains hemicellulose. In certain embodiments that may combine any of the preceding embodiments, the hydrocarbons or hydrocarbon derivatives can be used as fuel. In certain embodiments that may combine the preceding embodiment having the hydrocarbons or hydrocarbon derivatives used as fuel, the hydrocarbons or hydrocarbon derivatives contain ethanol. In certain embodiments that may combine the preceding embodiment having the hydrocarbons or hydrocarbon derivatives used as fuel, the hydrocarbons or hydrocarbon derivatives contain butanol.

In certain embodiments that may combine any of the preceding embodiments, the host cell is selected from the group consisting of *Saccharomyces* sp., *Saccharomyces cerevisiae*, *Saccharomyces monacensis*, *Saccharomyces bayanus*, *Saccharomyces pastorianus*, *Saccharomyces carlsbergensis*, *Saccharomyces pombe*, *Kluyveromyces* sp., *Kluyveromyces marxianus*, *Kluyveromyces lactis*, *Kluyveromyces fragilis*, *Pichia stipitis*, *Sporotrichum thermophile*, *Candida shehatae*, *Candida tropicalis*, *Neurospora crassa*, *Zymomonas mobilis*, *Clostridium* sp., *Clostridium phytofermentans*, *Clostridium thermocellum*, *Clostridium beijerinckii*, *Clostridium acetobutylicum*, *Moarella thermoacetica*, *Escherichia coli*, *Klebsiella oxytoca*, *Thermoanaerobacterium saccharolyticum*, and *Bacillus subtilis*.

Another aspect includes methods of increasing growth of a cell, including providing a host cell, where the host cell contains a recombinant polynucleotide where the polynucleotide encodes a NCU07705 polypeptide, and culturing the cell in a medium containing cellulose, where the host cell grows at a faster rate in the medium than a cell that does not contain the recombinant polynucleotide. In certain embodiments, the host cell is selected from the group consisting of *Saccharomyces* sp., *Saccharomyces cerevisiae*, *Saccharomyces monacensis*, *Saccharomyces bayanus*, *Saccharomyces pastorianus*, *Saccharomyces carlsbergensis*, *Saccharomyces pombe*, *Kluyveromyces* sp., *Kluyveromyces marxianus*, *Kluyveromyces lactis*, *Kluyveromyces fragilis*, *Pichia stipitis*, *Sporotrichum thermophile*, *Candida shehatae*, *Candida tropicalis*, *Neurospora crassa*, *Zymomonas mobilis*, *Clostridium* sp., *Clostridium phytofermentans*, *Clostridium thermocellum*, *Clostridium beijerinckii*, *Clostridium acetobutylicum*, *Moarella thermoacetica*, *Escherichia coli*, *Klebsiella oxytoca*, *Thermoanaerobacterium saccharolyticum*, and *Bacillus subtilis*. In certain embodiments, the host cell further contains an inducible promoter operably linked to the recombinant polynucleotide. In certain embodiments, expression of cellulases is increased in the host cell upon expression of the recombinant polynucleotide.

Another aspect includes methods of increasing growth of a cell on a biomass polymer, including providing a host cell, where the host cell contains an endogenous polynucleotide where the polynucleotide encodes an NCU05137 polypeptide, inhibiting expression of the endogenous polynucleotide, and culturing the cell in a medium containing the biomass polymer, where the host cell grows at a faster rate in the medium than a cell in which expression of the endogenous polynucleotide is not inhibited. In certain embodiments, the host cell is selected from the group consisting of *Saccharomyces* sp., *Saccharomyces cerevisiae*, *Saccharomyces monacensis*, *Saccharomyces bayanus*, *Saccharomyces pasto-*

16

rianus, *Saccharomyces carlsbergensis*, *Saccharomyces pombe*, *Kluyveromyces* sp., *Kluyveromyces marxianus*, *Kluyveromyces lactis*, *Kluyveromyces fragilis*, *Pichia stipitis*, *Sporotrichum thermophile*, *Candida shehatae*, *Candida tropicalis*, *Neurospora crassa*, *Zymomonas mobilis*, *Clostridium* sp., *Clostridium phytofermentans*, *Clostridium thermocellum*, *Clostridium beijerinckii*, *Clostridium acetobutylicum*, *Moarella thermoacetica*, *Escherichia coli*, *Klebsiella oxytoca*, *Thermoanaerobacterium saccharolyticum*, and *Bacillus subtilis*. In certain embodiments, cellulase activity of the host cell is increased upon inhibiting expression of the endogenous polynucleotide. In certain embodiments, hemicellulase activity of the host cell is increased upon inhibiting expression of the endogenous polynucleotide. In certain embodiments, inhibiting expression of the endogenous polynucleotide contains mutating or deleting a gene containing the endogenous polynucleotide. In certain embodiments, the biomass polymer is cellulose. In certain embodiments, the biomass polymer is hemicellulose.

Another aspect includes methods of increasing growth of a cell, including providing a host cell, where the host cell contains a recombinant polynucleotide where the polynucleotide encodes a polypeptide selected from the group consisting of NCU01517, NCU09133, and NCU10040, and culturing the cell in a medium containing hemicellulose, where the host cell grows at a faster rate in the medium than a cell that does not contain the recombinant polynucleotide. In certain embodiments, the host cell is selected from the group consisting of *Saccharomyces* sp., *Saccharomyces cerevisiae*, *Saccharomyces monacensis*, *Saccharomyces bayanus*, *Saccharomyces pastorianus*, *Saccharomyces carlsbergensis*, *Saccharomyces pombe*, *Kluyveromyces* sp., *Kluyveromyces marxianus*, *Kluyveromyces lactis*, *Kluyveromyces fragilis*, *Pichia stipitis*, *Sporotrichum thermophile*, *Candida shehatae*, *Candida tropicalis*, *Neurospora crassa*, *Zymomonas mobilis*, *Clostridium* sp., *Clostridium phytofermentans*, *Clostridium thermocellum*, *Clostridium beijerinckii*, *Clostridium acetobutylicum*, *Moarella thermoacetica*, *Escherichia coli*, *Klebsiella oxytoca*, *Thermoanaerobacterium saccharolyticum*, and *Bacillus subtilis*. In certain embodiments, the host cell further contains an inducible promoter operably linked to the recombinant polynucleotide. In certain embodiments, hemicellulase activity of the host cell is increased upon expression of the recombinant polynucleotide.

Another aspect includes methods of degrading cellulose, including providing a composition containing cellulose, and contacting the composition with a cellulase-containing enzyme mixture from an altered organism, where the cellulase-containing mixture has reduced β -glucosidase activity compared to a cellulase-containing mixture from an unaltered organism, and where the cellulose is degraded by the cellulase-containing mixture. In certain embodiments, the organism is altered by mutation of a gene encoding a β -glucosidase.

In certain embodiments, the organism is altered by reducing the expression of a β -glucosidase. In certain embodiments that may be combined with any of the preceding embodiments, the organism is selected from the group consisting of a fungus and a bacterium. In certain embodiments that may be combined with any of the preceding embodiments having the organism selected from the group consisting of a fungus and a bacterium, the organism is a filamentous fungus. In certain embodiments that may be combined with any of the preceding embodiments, the cellulose is from plant material. In certain embodiments that may be combined with the preceding embodiments having the cellulose from plant material, the plant material is selected from the group consisting of

switchgrass, *Miscanthus*, rice hulls, bagasse, flax, bamboo, sisal, abaca, straw, leaves, grass clippings, corn stover, corn cobs, distillers grains, legume plants, sorghum, sugar cane, sugar beet pulp, wood chips, sawdust, and biomass crops.

Yet another aspect includes methods of increasing the synthesis of hydrocarbons or hydrocarbon derivatives by a host cell comprising providing a host cell, wherein the host cell comprises a recombinant polynucleotide wherein the polynucleotide encodes a polypeptide encoded by a sequence selected from the group consisting of NCU00801, NCU00988, NCU01231, NCU04963, NCU05519, NCU05853, NCU05897, NCU06138, NCU00809, NCU08114, NCU10021, and any of the genes listed in Table 15 and culturing the host cell in a medium comprising a source of a compound to increase the synthesis of hydrocarbons or hydrocarbon derivatives by the host cell, wherein the compound is a substrate for the synthesis of the hydrocarbons or hydrocarbon derivatives, and wherein transport of the compound into the cell is increased upon expression of the recombinant polynucleotide. In certain embodiments, the host cell is selected from the group consisting of *Saccharomyces cerevisiae*, *Escherichia coli*, *Zymomonas mobilis*, *Neurospora crassa*, *Candida shehatae*, *Clostridium* sp., *Clostridium phytofermentans*, *Clostridium thermocellum*, *Moorella thermocetica*, *Thermoanaerobacterium saccharolyticum*, *Klebsiella oxytoca*, and *Pichia stipitis*. In certain embodiments, the host cell further comprises an inducible promoter operably linked to the recombinant polynucleotide. In certain embodiments, the recombinant polynucleotide encodes a polypeptide having at least 50% amino acid identity to the polypeptide encoded by a sequence selected from the group consisting of NCU00801, NCU00988, NCU01231, NCU04963, NCU05519, NCU05853, NCU05897, NCU06138, NCU00809, NCU08114, NCU10021, and any of the genes listed in Table 15. In some embodiments, the hydrocarbons or hydrocarbon derivatives can be used as fuel. In certain embodiments, the medium comprises cellulose. In other embodiments, the medium comprises hemicellulose. In certain embodiments, the compound is a sugar. In certain embodiments that may be combined with the preceding embodiments, the sugar is a pentose. In certain embodiments that may be combined with the preceding embodiments, the sugar is a hexose. In certain embodiments that may be combined with the preceding embodiments, the sugar is a disaccharide. In certain embodiments that may be combined with the preceding embodiments, the sugar is an oligosaccharide. In other embodiments, the compound is a plant phenol. In certain embodiments that may be combined with the preceding embodiments, the plant phenol is quinic acid. In certain embodiments that may be combined with the preceding embodiments, the plant phenol is nicotinamide. In other embodiments, the compound is pyruvate or lactate.

Another aspect includes methods of increasing growth of a cell on a biomass polymer comprising providing a host cell, wherein the host cell comprises a recombinant polynucleotide wherein the polynucleotide encodes a polypeptide encoded by any of the *Neurospora* or *Pichia stipitis* genes listed in Table 10, in Supplemental Data, Dataset S1, page 3 in Tian et al., *PNAS*, 2009, vol. 106, no. 52, 22157-22162, the disclosure of which is hereby incorporated by reference, in Table 15, or NCU01517, NCU09133, or NCU10040 and culturing the cell in a medium comprising the biomass polymer, wherein the host cell grows at a faster rate in the medium than a cell that does not comprise the recombinant polynucleotide. In certain embodiments, the polynucleotide encodes a polypeptide encoded by any of the sequences NCU00130.2, NCU00248.2, NCU00326.2, NCU00762.2, NCU00810.2,

NCU00890.2, NCU03328.2, NCU03415.2, NCU03731.2, NCU03753.2, NCU04197.2, NCU04249.2, NCU04287.2, NCU04349.2, NCU04475.2, NCU04997.2, NCU05057.2, NCU05159.2, NCU05493.2, NCU05751.2, NCU05770.2, 5 NCU05932.2, NCU06009.2, NCU06490.2, NCU07340.2, NCU07853.2, NCU07997.2, NCU08744.2, NCU08746.2, NCU08760.2, NCU09108.2, NCU09495.2, NCU09680.2, or NCU10045.2. In certain embodiments, the polynucleotide encodes a polypeptide encoded by NCU07705. In certain 10 embodiments, the recombinant polynucleotide encodes a polypeptide having at least 50% amino acid identity to the polypeptide encoded by any of the *Neurospora* or *Pichia stipitis* genes listed in Table 10, in Supplemental Data, Dataset S1, page 3 in Tian et al., 2009, or in Table 15. In certain 15 embodiments, the polynucleotide encodes a polypeptide having at least 50% amino acid identity to the polypeptide encoded by any of the sequences NCU00130.2, NCU00248.2, NCU00326.2, NCU00762.2, NCU00810.2, NCU00890.2, NCU03328.2, NCU03415.2, NCU03731.2, 20 NCU03753.2, NCU04197.2, NCU04249.2, NCU04287.2, NCU04349.2, NCU04475.2, NCU04997.2, NCU05057.2, NCU05159.2, NCU05493.2, NCU05751.2, NCU05770.2, NCU05932.2, NCU06009.2, NCU06490.2, NCU07340.2, NCU07853.2, NCU07997.2, NCU08744.2, NCU08746.2, 25 NCU08760.2, NCU09108.2, NCU09495.2, NCU09680.2, or NCU10045.2. In certain embodiments, the recombinant polynucleotide encodes a polypeptide having at least 50% amino acid identity to the polypeptide encoded by NCU07705. In certain 30 embodiments, the biomass polymer is cellulose. In other embodiments, the biomass polymer is hemicellulose. In certain embodiments, the host cell is selected from the group consisting of *Saccharomyces cerevisiae*, *Escherichia coli*, *Zymomonas mobilis*, *Neurospora crassa*, *Candida shehatae*, *Clostridium* sp., *Clostridium phytofermentans*, *Clostridium thermocellum*, *Moorella thermocetica*, *Thermoanaerobacterium saccharolyticum*, *Klebsiella oxytoca*, and *Pichia stipitis*. In certain 35 embodiments, the host cell further comprises an inducible promoter operably linked to the recombinant polynucleotide. In certain 40 embodiments, expression of cellulases is increased in the host cell upon expression of the recombinant polynucleotide. In other embodiments, expression of hemicellulases is increased in the host cell upon expression of the recombinant polynucleotide.

45 Yet another aspect includes methods of increasing growth of a cell on a biomass polymer comprising providing a host cell, wherein the host cell comprises an endogenous polynucleotide wherein the polynucleotide encodes a polypeptide encoded by any of the *Neurospora* or *Pichia stipitis* genes listed in Table 10, in Supplemental Data, Dataset S1, page 3 in Tian et al., 2009, or in Table 15, or, inhibiting expression of the endogenous polynucleotide, and culturing the cell in a medium comprising the biomass polymer, wherein the host cell grows at a faster rate in the medium than a cell in which expression of the endogenous polynucleotide is not inhibited. 50 In certain embodiments, the endogenous polynucleotide encodes a polypeptide encoded by any of the sequences NCU00130.2, NCU00248.2, NCU00326.2, NCU00762.2, NCU00810.2, NCU00890.2, NCU03328.2, NCU03415.2, NCU03731.2, NCU03753.2, NCU04197.2, NCU04249.2, NCU04287.2, NCU04349.2, NCU04475.2, NCU04997.2, NCU05057.2, NCU05159.2, NCU05493.2, NCU05751.2, NCU05770.2, NCU05932.2, NCU06009.2, NCU06490.2, NCU07340.2, NCU07853.2, NCU07997.2, NCU08744.2, NCU08746.2, NCU08760.2, NCU09108.2, NCU09495.2, NCU09680.2, or NCU10045.2. In certain 55 embodiments, the endogenous polynucleotide encodes a polypeptide encoded

by NCU05137. In certain embodiments, the endogenous polynucleotide encodes a polypeptide having at least 50% amino acid identity to the polypeptide encoded by any of the *Neurospora* or *Pichia stipitis* genes listed in Table 10, in Supplemental Data, Dataset S1, page 3 in Tian et al., 2009, or in Table 15. In certain embodiments, the endogenous polynucleotide encodes a polypeptide having at least 50% amino acid identity to the polypeptide encoded by any of the sequences NCU00130.2, NCU00248.2, NCU00326.2, NCU00762.2, NCU00810.2, NCU00890.2, NCU03328.2, NCU03415.2, NCU03731.2, NCU03753.2, NCU04197.2, NCU04249.2, NCU04287.2, NCU04349.2, NCU04475.2, NCU04997.2, NCU05057.2, NCU05159.2, NCU05493.2, NCU05751.2, NCU05770.2, NCU05932.2, NCU06009.2, NCU06490.2, NCU07340.2, NCU07853.2, NCU07997.2, NCU08744.2, NCU08746.2, NCU08760.2, NCU09108.2, NCU09495.2, NCU09680.2, or NCU10045.2. In certain embodiments, the endogenous polynucleotide encodes a polypeptide having at least 50% amino acid identity to the polypeptide encoded by NCU05137. In certain embodiments, the host cell is selected from the group consisting of *Saccharomyces cerevisiae*, *Escherichia coli*, *Zymomonas mobilis*, *Neurospora crassa*, *Candida shehatae*, *Clostridium* sp., *Clostridium phytofermentans*, *Clostridium thermocellum*, *Moorella thermocetica*, *Thermoanaerobacterium saccharolyticum*, *Klebsiella oxytoca*, and *Pichia stipitis*. In certain embodiments, the biomass polymer is cellulose. In other embodiments, the biomass polymer is hemicellulose. In certain embodiments, cellulase activity of the host cell is increased upon inhibiting expression of the endogenous polynucleotide. In other embodiments, hemicellulase activity of the host cell is increased upon inhibiting expression of the endogenous polynucleotide. In certain embodiments, inhibiting expression of the endogenous polynucleotide comprises mutating or deleting a gene comprising the endogenous polynucleotide.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 shows the domain structure of the polypeptide encoded by NCU07705.

FIG. 2 shows the phylogenetic analysis of NCU05137. The predicted orthologs of *N. crassa* NCU05137 were retrieved from NCBI and JGI based on amino acid sequences showing significant similarity by BLAST. All identified filamentous fungal orthologs are shown; NCBI E values were 0.0 except for *B. fuckeliana*, which was 9e-175. Homologs of NCU05137 were also identified in a number of bacteria (E value~e-30). YP_981875 from *Polaromonas naphthalenivorans* (a beta-proteobacterium) was used as an outgroup. A=*Aspergillus*; N=*Neosartorya*; P=*P. chrysogenum*=*Penicillium*; S=*Sclerotinia*; B=*Botryotinia*; P=*Pyrenopora*; C=*Cochliobolus*; N=*haematococca*=*Nectria*; P=*anserina*=*Podospora*; N=*Neurospora*. The tree was made by MEGA3, NJ. Bar=0.2 substitutions per amino acid site.

FIG. 3 shows an analysis of *N. crassa* FGSC2489 and *T. reesei* QM9414 endoglucanase activity when grown on *Miscanthus* and Avicel as a sole carbon source. Endoglucanase activity in culture filtrates of *N. crassa* WT strain FGSC2489 and *T. reesei* QM9414. *N. crassa* was grown on Vogel's minimal medium containing 2% of either Avicel or *Miscanthus* powder as a sole carbon source at 25° C. *T. reesei* strain was inoculated in MA medium with either 1% Avicel or *Miscanthus* powder as sole carbon source at 25° C. Both strains were inoculated with the same amount of conidia (1×10^6 /mL in 100 mL culture). The endoglucanase activity in

the cultures at different time points were measured at pH 4.5 using Azo-CM-cellulose as a substrate according to the manufacturer's instructions (Megazyme, Ireland).

FIG. 4 shows transcriptional profiling of *N. crassa* grown on *Miscanthus* and Avicel. (A) Hierarchical clustering analysis of 769 genes showing expression differences in *Miscanthus* culture. Dark shading indicates higher relative expression and light shading indicates lower relative expression. Lane 1: Expression profile of a 16 hr Vogel's minimal medium *N. crassa* culture (Vogel 1956). Lane 2: Expression profile of a culture grown on *Miscanthus* as a sole carbon source for 16 hrs. Lanes 3, 4, 5: Expression profiles from cultures grown on *Miscanthus* for 40 hrs, 5 days, and 10 days. The three clusters are shown as C1, C2, and C3. The cluster that showed increased expression levels of most of the cellulase and hemicellulase genes is boxed (C3 cluster). (B) Analysis of the overlap in expression profiles between the *N. crassa* *Miscanthus* versus Avicel grown cultures (Top). Analysis and overlap of proteins detected in the culture filtrates of *N. crassa* grown on *Miscanthus* and Avicel by tandem mass spectrometry (Bottom). (C) Functional category (FunCat) enrichment analysis (Ruepp 2004) of the 231 genes that showed an increase in relative expression levels in *Miscanthus* cultures. Functional categories that showed significant enrichment ($p < 0.001$), including the unclassified group are shown.

FIG. 5 shows the relative expression levels of *N. crassa* genes encoding cellulases (A) and hemicellulases (B) during growth on minimal medium (MM) and during growth on *Miscanthus* for 16 hr, 40 hr, 5 days and 10 days.

FIG. 6 shows the protein profile and enzymatic activity of culture supernatants from strains containing deletions of genes encoding secreted proteins identified by MS. (A) SDS-PAGE of proteins present in the culture filtrates of 16 deletion strains as compared to wild type when grown on Avicel for 7 days. Deletion strains were chosen based on identification of the protein by mass spectrometry in both *Miscanthus* and Avicel culture filtrates. Strains are ordered based on gene NCU number, the wild-type strain is FGSC 2489. Missing protein bands that correspond to the deleted genes are marked with boxes. (B) Total secreted protein, azo-CMCase, and β -glucosidase activity assays (see Example 5) performed on 16 deletion strains and the wild-type parental strain (FGSC 2489) using the same sample from (A). Activities and protein concentrations were normalized compared to wild type levels and represent the average of triplicate biological measurements. (C) Cellulase activity of the culture filtrates from the 16 deletion strains using the same samples as in (A). Culture filtrates were diluted 10 fold and mixed with 5 mg/mL Avicel (see Example 5) to assess Avicelase activity. Glucose (black) and cellobiose (white) were measured after 8 hours of incubation at 40° C.

FIG. 7 shows the identity of *N. crassa* secreted proteins based on mutant analysis from a culture grown on Avicel as a sole carbon source. SDS-PAGE of secreted proteins from WT *N. crassa* (FGSC 2489) grown on 2% Avicel in 100 mL shake flasks for 7 days at 25° C. 15 μ L of unconcentrated culture filtrate was loaded onto Criterion 4-15% 26-well gel. Proto Blue Safe (Coomassie) from National Diagnostics was used to stain the gel. The protein bands were identified in this study as shown in FIG. 6A based on analysis of secreted proteins in deletion strains.

FIG. 8 shows the profile of secreted proteins and expression of cbh-1 (NCU07340) and gh6-2 (NCU09680; CBHII) in Δ NCU04952 and Δ NCU05137. (A) SDS-PAGE of total secreted proteins in WT, Δ NCU04952, and Δ NCU05137. Cultures were grown on Avicel from conidia, and harvested at 30 hrs, two days (48 hrs) and three days (72 hrs) (see Example

5). Lanes 1-3, 20 \times concentrated culture filtrates after 30 hrs of growth on Avicel from WT, Δ NCU04952, and Δ NCU05137 strains, respectively. Lanes 4-6, unconcentrated culture filtrates after two days of growth from WT, Δ NCU04952, and Δ NCU05137 strains, respectively. Lane 7-9, unconcentrated culture filtrates after three days of growth from WT and Δ NCU04952 and Δ NCU05137 strains, respectively. (B) RT-PCR of cbh-1 (NCU07340; CBHI) and gh6-2 (NCU09680; CBHII) in the WT, Δ NCU04952, and Δ NCU05137 strains during growth on Avicel. The WT and deletion strains were grown on Avicel from conidia, and harvested at 48 hrs and 72 hrs (see Example 5). The minimal medium (MM) culture, with sucrose as a sole carbon source (Vogel 1956), was grown for 16 hrs (similar developmental time point). The fold induction of cbh-1 and gh6-2 were relative to the expression of these genes under MM conditions, with actin gene expression used as the control in all samples.

FIG. 9 shows a model of plant cell wall deconstruction in *N. crassa*. Induction: Extracellular enzymes expressed at low levels generate secondary metabolites that signal *N. crassa* to dramatically increase the expression level of genes encoding plant cell wall degrading enzymes, most of which are secreted. Utilization: Extracellular enzymes and transporters specific for translocation of cell wall degradation products enable *N. crassa* to utilize plant cell material for growth. Some extracellular proteins (NCU05137, NCU05057, and NCU04952) may generate metabolites that modulate gene expression of cellulases and hemicellulase during the utilization phase; double hexagon (cellobiose), double pentagon (xylobiose), hexagon (glucose), and pentagon (xylose). The depicted plant cell wall-degrading enzymes include CBH(I), CBH(II), EG2, EG1, EG6, and xylanase. Additional cellulolytic enzymes are not shown. Thickness of arrows indicates relative strength of response.

FIG. 10 shows BLAST results from searching the sequences of *N. crassa* putative transporters against a database of *S. thermophile* protein sequences or from searching the sequences of *S. thermophile* putative transporters against a database of *N. crassa* protein sequences.

FIG. 11 shows the growth phenotype of a *N. crassa* strain lacking NCU08114. (A) Shaker flasks of WT (left) and Δ NCU08114 (right) *N. crassa* strains after 3 days of growth with crystalline cellulose as a carbon source. (B) Mean Alamar Blue® fluorescence from *N. crassa* cultures grown with either sucrose or crystalline cellulose as a carbon source for 16 or 28 hours, respectively. Fluorescence was normalized by setting WT to 100%. Error bars were the standard deviation between measurements from three biological replicates. *N. crassa* lacking NCU0801 did not have an obvious phenotype. *N. crassa* secreted β -glucosidases (Tian et al., 2009) that hydrolyzed cellobextrins to glucose, which was subsequently taken up by monosaccharide transporters (Scarborough 1973). This alternate route of consumption led to an underestimate of the cellobextrin transport defect in these deletion lines.

FIG. 12 shows (A) cellobiose consumption for *S. cerevisiae* strains expressing NCU00801, NCU05853, or NCU08114 along with NCU00130; (B) cellotriose consumption for *S. cerevisiae* strains expressing NCU00801, NCU05853, or NCU08114 along with NCU00130; (C) celotetraose consumption for *S. cerevisiae* strains expressing NCU00801, NCU05853, or NCU08114 along with NCU00130; and (D) cellohexaose consumption for *S. cerevisiae* strains expressing NCU00801, NCU05853, or NCU08114 along with NCU00130.

FIG. 13 shows cellobextrin consumption by *N. crassa* strains lacking NCU008114 or NCU00801. The indicated *N.*

crassa strains were incubated with 90 μ M of the respective sugars for 15 minutes. Bars represent the mean concentration of sugars remaining in the supernatant following the incubation from two independent experiments. Error bars were the standard deviation between these experiments.

FIG. 14 shows cellobiose transport by a *S. cerevisiae* strain expressing NCU00801/cbt1. Shown is cellobiose transport by yeast with (○) or without (●) CBT1. Both strains expressed the intracellular β -glucosidase, NCU00130. The initial concentration of cellobiose was 50 μ M. All values were the mean between two measurements, with error bars representing the standard deviation between these measurements.

FIG. 15 shows localization and quantification of GFP fused to CBT1 and CBT2. (A) Images of *S. cerevisiae* strains expressing cbt1 (left), or cbt2 (right), fused to GFP at their C-terminus. (B) GFP fluorescence of yeast strains without a cellobiose transporter, or expressing cbt1 or cbt2 fused to GFP at their C-terminus. Values were the mean from three biological replicates, and error bars represent the standard deviation between these replicates.

FIG. 16 shows cellobextrin transport by *N. crassa* transport systems expressed in *S. cerevisiae*. (A) Cellobiose-mediated growth of yeast strains expressing the gene NCU00801 (named cbt1, ○), NCU08114 (named cbt2, ▼), or no transporter (●). All strains also expressed the intracellular β -glucosidase, NCU00130. A representative experiment is shown. Growth rates from three independent experiments were as follows: cbt1, $0.0341 \pm 0.0010 \text{ hr}^{-1}$; cbt2, $0.0131 \pm 0.0008 \text{ hr}^{-1}$; no transporter, $0.0026 \pm 0.0001 \text{ hr}^{-1}$. (B) Growth of yeast strains on cellotriose and cellotetraose. Strains expressing the intracellular β -glucosidase, NCU00130, as well as the transporters listed in the legend, were grown with 0.5% (w/v) of cellotriose (G3) or cellotetraose (G4) serving as the sole carbon source. A representative experiment is shown. Growth rates from three independent experiments were as follows: cbt1 cellotriose, $0.0332 \pm 0.0004 \text{ hr}^{-1}$; cbt1 cellotetraose $0.0263 \pm 0.0020 \text{ hr}^{-1}$; no transporter cellotriose, $0.0043 \pm 0.0015 \text{ hr}^{-1}$; cbt2 cellotriose, $0.0178 \pm 0.0005 \text{ hr}^{-1}$; cbt2 cellotetraose $0.0041 \pm 0.0003 \text{ hr}^{-1}$; no transporter cellotetraose, $0.0031 \pm 0.0008 \text{ hr}^{-1}$. (C) Glucose produced from cellobiose (G2), cellotriose (G3), and cellotetraose (G4) hydrolysis by purified NCU00130. The mean and standard deviation of three independent measurements are shown. Residual glucose in incubations without enzyme (2 nmol) was subtracted from the values shown.

FIG. 17 shows growth of *S. cerevisiae* strains expressing cbt1 (○), cbt2 (▼), or no transporter (●) on glucose. All strains expressed the β -glucosidase, NCU00130. A representative experiment is shown.

FIG. 18 shows cellobiose-mediated growth of *S. cerevisiae* strains in 250 mL flasks. Values represent the mean OD between two replicate cultures of yeast strains expressing the β -glucosidase, NCU00130, cbt1 or cbt2, or a strain expression NCU00130, but lacking any transporters. Error bars represent the standard deviation between replicates.

FIG. 19 shows kinetics of cellobiose transport by CBT1 and CBT2. The rate of cellobiose transport was determined as a function of cellobiose concentration by yeast strains expressing either cbt1 or cbt2. The transport rate was normalized for transporter abundance.

FIG. 20 shows the ability of *S. cerevisiae* expressing the combinations of *Neurospora* genes shown on the x-axis to grow on cellobiose, cellotriose, or cellotetraose.

FIG. 21 shows competition by cellobextrins for cellobiose transport in strains carrying cbt1 or cbt2. A 5-fold excess of the respective unlabeled sugar was included during assays of [³H]-cellobiose transport. Substrates of CBT1 or CBT2

would decrease the [³H]-cellobiose transport rate by competing for binding. Bars represent the mean from three replicates. Error bars represent the standard deviation between these replicates. Values were normalized by setting the rate of [³H]-cellobiose transport without a competing sugar to 100.

FIG. 22 shows the SDS-PAGE gel of purified NCU00130. Lane 1, Protein molecular weight standards, in kDa. Lane 2, NCU00130 after purification over nickel-NTA resin. Molecular weights in kDa are shown to the left.

FIG. 23 shows maximum likelihood phylogenetic analysis of the cellobiose transporters NCU08114 and NCU00801. With the exception of *S. cerevisiae* HXT1 and *K. lactis* LACP, all genes encoding proteins shown are reported to increases in expression level when the fungus comes into contact with plant cell wall material or cellobiose (Tian et al., 2009; Noguchi et al., 2009; Wymelenberg et al., 2010; Martin et al., 2010). *S. cerevisiae* HXT1, a low affinity glucose transporter (Reifenberger et al., 1997), was used as an outgroup.

FIG. 24 shows cellobiose fermentation, and simultaneous saccharification and fermentation of cellulose, by *S. cerevisiae* expressing the cellobiose transport system from *N. crassa*. (A) Cellobiose fermentation to ethanol. Ethanol produced by yeast strains with CBT1 (●), or without CBT1 (○). Cellobiose concentration during the fermentation reaction using yeast strains with CBT1 (▼), or without CBT1 (Δ). (B) SSF using yeast strains with and without CBT1. Cellobiose (●) and glucose (▼) concentrations in the presence of a strain with CBT1, and cellobiose (○) and glucose (Δ) concentrations in the presence of a strain lacking CBT1. Note, 0.1 mg/mL cellobiose=292 μM. (C) Ethanol produced during SSF using a strain with CBT1 (●), or without CBT1 (○). In all panels, values are the mean of 3 biological replicates. Error bars were the standard deviation between these replicates. All strains also expressed the intracellular β-glucosidase, NCU00130.

FIG. 25 shows use of cellodextrin transport pathways from filamentous fungi during simultaneous saccharification and fermentation of cellulose by yeast. The cellodextrin (Cdex) transport pathway (black) includes a cellodextrin transporter (CBT) and intracellular β-glucosidase (βG). The sugar catabolism pathway presented in standard yeast includes hexose transporters (HXT). In SSF, both cellulases (GH) and extracellular β-glucosidase (βG) could be used.

FIG. 26 shows residues in NCU00801 and NCU08114 that are critical for function. (A) Ala-scan of cbt1/NCU00801. (B) Polypeptide sequence (important residues marked) of cbt1/NCU00801. (C) Polypeptide sequence (important residues marked) of cbt2/NCU08114.

FIG. 27 shows a comparison of *S. cerevisiae* strains expressing cellobiose transporters from *P. stipitis*. (A) Cell growth of *S. cerevisiae* strains expressing β-glucosidase and orthologs of cellobiose transporters NCU00801, NCU08114, and NCU05853. (B) Comparison of cellobiose transporters from *P. stipitis*: cell growth of *S. cerevisiae* strains expressing β-glucosidase and cellobiose transporters. (C) Comparison of cellobiose transporters from *P. stipitis*: xylose consumption and ethanol production by *S. cerevisiae* strains expressing β-glucosidase and cellobiose transporters.

FIG. 28 shows alignments of cellobiose transporter orthologs. (A) Alignment of cellobiose transporter orthologs including ones that did not appear to have transporter function under the conditions tested. (B) Alignment of cellobiose transporter orthologs that had transport function. (C) Alignment of NCU00801 and NCU08114.

FIG. 29 shows functionally important motifs marked in homology models of NCU00801 and NCU08114. (A) Location of cellobiose transporters motifs on NCU00801 homol-

ogy model. Motif [LIVM]-Y-[FL]-x(13)-[YF]-D (SEQ ID NO: 1) is shown in red. Motif [YF]-x(2)-G-x(5)-[PVF]-x(6)-[DQ] (SEQ ID NO: 2) is shown in light green. Motif G-R-[RK] (SEQ ID NO: 3) is shown in dark blue. Motif R-x(6)-[YF]-N (SEQ ID NO: 4) is shown in yellow. Motif WR-[IVLA]-P-x(3)-Q (SEQ ID NO: 5) is shown in magenta. Motif P-E-S-P-R-x-L-x(8)-A-x(3)-L-x(2)-Y-H (SEQ ID NO: 6) is shown in cyan. Motif F-[GST]-Q-x-S-G-N-x-[LIV] (SEQ ID NO: 7) is shown in orange. Motif L-x(3)-[YIV]-x(2)-E-x-L-x(4)-R-[GA]-K-G (SEQ ID NO: 8) is shown in dark green. I. View of NCU00801 from the cytoplasmic side looking into the putative cellobiose binding pore. Note that in this image, some of the residues connecting transmembrane helices 6 and 7 have been removed for clarity as they occlude the pore. II. View of one side of NCU00801. III. View of the side opposite to that shown in II. (B) Location of cellobiose transporters motifs on NCU08114 homology model. Motif [LIVM]-Y-[FL]-x(13)-[YF]-D (SEQ ID NO: 1) is shown in red. Motif [YF]-x(2)-G-x(5)-[PVF]-x(6)-[DQ] (SEQ ID NO: 2) is shown in light green. Motif G-R-[RK] (SEQ ID NO: 3) is shown in dark blue. Motif R-x(6)-[YF]-N (SEQ ID NO: 4) is shown in yellow. Motif WR-[IVLA]-P-x(3)-Q (SEQ ID NO: 5) is shown in magenta. Motif P-E-S-P-R-x-L-x(8)-A-x(3)-L-x(2)-Y-H (SEQ ID NO: 6) is shown in cyan. Motif F-[GST]-Q-x-S-G-N-x-[LIV] (SEQ ID NO: 7) is shown in orange. Motif L-x(3)-[YIV]-x(2)-E-x-L-x(4)-R-[GA]-K-G (SEQ ID NO: 8) is shown in dark green. I. View of NCU08114 from the cytoplasmic side looking into the putative cellobiose binding pore. Note that in this image, some of the residues connecting transmembrane helices 6 and 7 have been removed for clarity as they occlude the pore. II. View of one side of NCU08114. III. View of the side opposite to that shown in II. (c) Predicted secondary structures in NCU00801 and NCU08114.

FIG. 30 shows the cloning process used in the construction of plasmid expressing: (A) putative transporters and (B) transporter-GFP fusion proteins.

FIG. 31 shows pentose transport activity of putative transporters identified to have glucose-uptake activity.

FIG. 32 shows pentose transport activity of putative transporters identified to not have glucose-uptake activity.

FIG. 33 shows pentose uptake of NCU00821 (AN25), STL12/XUT6 (Xyp29), and XUT1 (Xyp32). Part (A) shows xylose uptake and part (B) shows arabinose uptake.

FIG. 34 shows ¹⁴C-labeled sugar uptake by *S. cerevisiae* expressing STL12/XUT6 (Xyp29).

FIG. 35 shows localizations of transporters expressed in *S. cerevisiae* cells as monitored by GFP fluorescence. First row from left to right: NCU00821-GFP fluorescence, NCU00821 nuclei; second row from left to right: STL12/XUT6-GFP fluorescence, STL12/XUT6 nuclei.

FIG. 36 shows the effect on pH upon addition of maltose to un-buffered cell suspension expressing: (a) NCU00821 (AN25), (b) STL12/XUT6 (Xyp29), and (c) XUT1 (Xyp32).

The black arrows indicate the time points when maltose was added.

FIG. 37 shows results of a symporter assay of NCU00821, STL12/XUT6, and XUT1. Part (A) shows NCU00821 for xylose, part (B) shows NCU00821 for arabinose, part (C) shows XUT1 for arabinose, part (D) shows XUT1 for xylose, part E shows STL12/XUT6 for arabinose, and part F shows STL12/XUT6 for xylose. The black arrows the time points when maltose was added.

FIG. 38 shows phenotypic analyses of transporter overexpression. Part (A) shows OD, part (B) shows xylose concentration, and part (C) shows xylose consumption in 0.5% xylose-containing media. Part (D) shows OD, part (E) shows

xylose concentration, and part (F) shows xylose consumption in 5% xylose-containing media. Part (G) shows the growth curve of *S. cerevisiae* containing pentose transporters introduced on pRS424, a multicopy plasmid.

FIG. 39 shows maps of the plasmids used for cloning of heterologous transporters.

FIG. 40 shows results of the sugar-uptake assay by *S. cerevisiae* strains expressing pentose transporter orthologs.

FIG. 41 shows sequence alignments of the pentose transporter orthologs by Clustal W (1.81). (a) Alignment of the xylose transporter orthologs. (b) Alignment of the arabinose transporters. (c) Alignment of xylose and arabinose transporters. Consensus key: *—single, fully conserved residue; :—conservation of strong groups; .—conservation of weak groups.

FIG. 42 describes the different *S. cerevisiae* strains engineered to express xylose-utilizing enzymes.

FIG. 43 shows xylose metabolism (as monitored by xylose consumption, ethanol production, etc.) of three *S. cerevisiae* strains of different backgrounds expressing identical cassettes containing xylose utilization pathway enzymes.

FIG. 44 shows xylose-uptake rates and metabolite yields of three *S. cerevisiae* strains of different backgrounds expressing identical cassettes containing xylose utilization pathway enzymes.

FIG. 45 shows xylose fermentation by the *S. cerevisiae* strain DA24 under various conditions. (a) 40 g/L xylose in a shaker flask, (b) 80 g/L xylose in a shaker flask, and (c) 80 g/L xylose in a bioreactor. Symbols: xylose (■), ethanol (◆), and OD₆₀₀ (●).

FIG. 46 shows a comparison of xylose consumption and ethanol production between (a) *S. cerevisiae* DA24 and (b) *P. stipitis*. Symbols: xylose (■), ethanol (◆), and OD₆₀₀ (●).

FIG. 47 describes the experimental design used to test the effect of XYL2 over-expression levels on xylose metabolism in engineered *S. cerevisiae*.

FIG. 48 shows the effect of additional XYL2 integration (i.e. increased XYL2 expression level) into the genome of engineered xylose-fermenting *S. cerevisiae*.

FIG. 49 shows the effect of additional simultaneous over-expression of XYL2 and XYL3 on xylose fermentation by engineered *S. cerevisiae*.

FIG. 50 describes *S. cerevisiae* strains expressing different levels of xylose-fermenting enzymes.

FIG. 51 shows the effect of differential XYL1 expression of fermentation by engineered *S. cerevisiae*.

FIG. 52 describes *S. cerevisiae* strains engineered to over-express identical XYL2 and XYL3 but different reductases (XYL1 vs. GRE3).

FIG. 53 shows the effect of over-expressing XYL1 versus GRE3 on xylose fermentation by engineered *S. cerevisiae* grown in 40 g/L xylose.

FIG. 54 shows the effect of over-expressing XYL1 versus GRE3 on xylose fermentation by engineered *S. cerevisiae* grown in 80 g/L xylose.

FIG. 55 shows the thermal and pH-dependent properties of different wild-type LAD enzymes: anLAD (■), tLAD (◆), and pcLAD (●). (a) Temperature-dependent catalytic activities, (b) Thermal inactivation at 50° C. over time, and (c) pH-dependent catalytic activities. Error bars indicate standard error of the mean (n=3).

FIG. 56 shows an alignment of XDH from *N. crassa* (ncXDH) and *P. stipitis* (psXDH).

FIG. 57 show a comparison of pH rate profiles of *N. crassa* LAD and XDH. Data taken from the characterization of LAD was performed in universal buffer MES/Tris/glycine, and overlapped with data for ncXDH (closed triangles) and

ncLAD (closed circles) performed in universal buffer acetic acid/MES/Tris for lower pH values.

FIG. 58 shows ethanol production by *S. cerevisiae* strain L2612 transformed with xylose isomerase enzyme from *Bacteroids stercoris* (BtXI), *Bifidobacterium longum* (BtXI), and BtXIO coding for codon-optimized BtXI. The XI gene was cloned into the pRS424TEF vector.

FIG. 59 shows xylose consumption and ethanol production by *S. cerevisiae* strain D452-2, which had BtXI integrated into its genome by the vector pRS403TEF. Comparison is also made to xylose-fermentation by *S. cerevisiae* strain L2612, which expresses BtXI from a plasmid.

FIG. 60 shows xylose fermentation by *S. cerevisiae* strain, containing integrated BtXI and expressing XYL2 or XYL3 or XYL2 and XYL3.

FIG. 61 shows the necessity of XYL3 expression in *S. cerevisiae* engineered to over-express enzymes, such as GND1, involved in the pentose phosphate pathway in order to efficiently metabolize xylose.

FIG. 62 shows the effect of over-expression of NCU09705 homologs in *E. coli*, *S. cerevisiae*, and *P. stipitis* on fermentation parameters. Over-expression of galM, GAL10-Sc, GAL10-Ps, YHR210C, and YNR071c on (A) cellobiose consumption, growth, and ethanol production; and on (B) ethanol yield and productivity.

FIG. 63 shows the experimental design enabling simultaneous co-fermentation of cellobiose and xylose without glucose repression through integration of a cellobextrin assimilation pathway from filamentous fungi (*N. crassa*) and modified xylose metabolic pathway from the xylose-fermenting yeast *P. stipitis* into *S. cerevisiae*. (a) A strain improvement strategy to engineer yeast strain capable of fermenting two non-metabolizable sugars (cellobiose and xylose). The cellobextrin assimilation pathway consists of a cellobextrin transporter (NCU00801) and an intracellular β-glucosidase (NCU00130) from *N. crassa*. The modified xylose metabolic pathway utilizes xylose reductase isozymes (wild-type XR and mutant XR^{R276H}), xylitol dehydrogenase (XYL2), and xylulokinase (XKS1). (b) Fermentation profile of a sugar mixture containing glucose and xylose by the engineered *S. cerevisiae* developed in this study. Glucose fermentation repressed xylose fermentation completely so that xylose fermentation begins only after glucose depletion. (c) Fermentation profile of a sugar mixture containing cellobiose and xylose by the engineered *S. cerevisiae* developed in this study. Cellobiose and xylose are simultaneously utilized, as neither carbon source repressed consumption of the other.

FIG. 64 shows the scheme for plasmid construction. The pRS425 shuttle vector was linearized followed by assembly of the cellobiose transporter and β-glucosidase genes using the DNA assembler method (Shao et al., 2009).

FIG. 65 shows the change in concentrations of cellobiose (■), glucose (●), D-xylose (▲), ethanol (▼), and biomass (□) during co-fermentation of 4% cellobiose and 5% D-xylose by *S. cerevisiae* strains (a) SL01, (b) SL04, (c) SL02, (d) SL05, (e) SL03, (f) SL06, and (g) SL00 as a function of time.

FIG. 66 shows the change in concentrations of cellobiose (■), glucose (●), D-xylose (▲), ethanol (▼), and biomass (□) in *S. cerevisiae* strains SL01 (a, c) and SL00 (b, d) grown in cellobiose-xylose mixtures in shake-flasks (a, b) or bioreactors (c, d) plotted as a function of time.

FIG. 67 shows the change in concentrations of cellobiose (■), glucose (●), D-xylose (▲), ethanol (▼), and biomass (□) in *S. cerevisiae* strains SL01 (a, c) and SL00 (b, d) grown in media containing 5 g/L glucose-40 g/L cellobiose-50 g/L

xylose mixture (a, b) or 10 g/L glucose-40 g/L cellobiose-50 g/L xylose mixture (c, d) in bioreactors, plotted as a function of time.

FIG. 68 shows a comparison of cellobiose utilizations by β -glucosidase (NCU00130)-containing *S. cerevisiae* strain expressing (a) NCU00801, (b) NCU00809, and (c) NCU08114. Symbols: cellobiose (■), ethanol (◆), and OD₆₀₀ (●).

FIG. 69 shows co-fermentation of cellobiose and xylose by the *S. cerevisiae* strain DA24-16BT3 grown in mixtures containing various concentrations of the two sugars: (a) 20 g/L (each) of cellobiose and xylose, (b) 30 g/L (each) of cellobiose and xylose, and (c) 40 g/L (each) of cellobiose and xylose. Symbols: cellobiose (▲), xylose (■), ethanol (◆), and OD₆₀₀ (●).

FIG. 70 shows the synergistic effects of co-fermentation of cellobiose and xylose by the *S. cerevisiae* strain DA24-16BT3. Symbols: cellobiose (▲), xylose (■), ethanol (◆), and OD₆₀₀ (●): (a) 40 g/L cellobiose, (b) 40 g/L (each) of cellobiose and xylose, and (c) 40 g/L xylose.

FIG. 71 shows co-fermentation of glucose, cellobiose, and xylose by the *S. cerevisiae* strain DA24-16BT3 and the wild-type *P. stipitis* strain. Symbols: cellobiose (▲), xylose (■), ethanol (◆), OD₆₀₀ (●), and glucose (▼). (a) DA24-16BT3 and (b) *P. stipitis*.

FIG. 72 shows HPLC chromatograms from each time point, suggesting celotriose and celotetraose accumulation during c-fermentation of cellobiose and xylose by the *S. cerevisiae* strain DA24-16BT3.

FIG. 73 shows HPAEC analysis demonstrating cellodextrin accumulation in fermentation medium after 22 hours fermentation by the *S. cerevisiae* strain DA24-16BT3 during co-fermentation of cellobiose and xylose. (G1: glucose, G2: cellobiose, G3: celotriose, G4: celotetraose, and G5: cellopentaose).

FIG. 74 shows a comparison of sugar utilization by *S. cerevisiae* transformants expressing (a) an integrated copy of NCU00801 and (b) NCU00801 on a multi-copy plasmid, during co-fermentation of 40 g/L (each) of cellobiose and xylose. Symbols: cellobiose (▲), xylose (■), ethanol (◆), and OD₆₀₀ (●).

FIG. 75 shows ethanol production by cultivation of two different yeast strains. (a) The two different *S. cerevisiae* strains used in study: DA24-16 and D452BT. A xylose molecule is shown as a pentagon and a cellobiose molecule is shown as two hexagons; and (b) Mixed cultures of xylose-fermenting strain and cellobiose-fermenting strain.

FIG. 76 shows a listing of 354 xylan-induced genes in *N. crassa*.

FIG. 77 shows secreted protein levels, reducing sugar, and azo-xylanase activity for various *N. crassa* knock-out strains. Secreted protein levels were relatively constant for all strains.

FIG. 78 shows (A) total secreted protein and CMC-activity for wild type, Δ NCU05137, and Δ NCU05137/ Δ NCU05137-GFP *Neurospora* strains, and (B) a Coomassie stain of total protein in supernatants from cultures of the three different strains.

FIG. 79 shows localization of NCU05137-GFP in conidia.

FIG. 80 shows localization of NCU05137-GFP in the hypha tip.

DETAILED DESCRIPTION OF THE INVENTION

The present disclosure relates to host cells containing a recombinant polynucleotide encoding a polypeptide containing transmembrane α -helix 1, α -helix 2, α -helix 3, α -helix 4, α -helix 5, α -helix 6, α -helix 7, α -helix 8, α -helix 9, α -helix

10, α -helix 11, and α -helix 12, where one or more of the following is true: transmembrane α -helix 1 comprises SEQ ID NO: 1, transmembrane α -helix 2 comprises SEQ ID NO: 2, the loop connecting transmembrane α -helix 2 and transmembrane α -helix 3 comprises SEQ ID NO: 3, transmembrane α -helix 5 comprises SEQ ID NO: 4, transmembrane α -helix 6 comprises SEQ ID NO: 5, sequence between transmembrane α -helix 6 and transmembrane α -helix 7 comprises SEQ ID NO: 6, transmembrane α -helix 7 comprises SEQ ID NO: 7, and transmembrane α -helix 10 and transmembrane α -helix 11 and the sequence between them comprise SEQ ID NO: 8, and where the polypeptide transports celldextrin into the cell. Further described herein are methods of increasing transport of celldextrin into a cell, methods of increasing growth of a cell on a medium containing celldextrin, methods of co-fermenting cellulose-derived and hemicellulose-derived sugars, and methods of making hydrocarbons or hydrocarbon derivatives using the host cells. Further described herein are host cells containing a recombinant polynucleotide encoding a polypeptide where the polypeptide transports a pentose into the cell, methods of increasing transport of a pentose into a cell, methods of increasing growth of a cell on a medium containing pentose sugars, and methods of making hydrocarbons or hydrocarbon derivatives by providing a host cell containing a recombinant polynucleotide encoding a polypeptide where the polypeptide transports a pentose into the cell.

As used herein, celldextrin refers to glucose polymers of varying length and includes, without limitation, cellobiose (2 glucose monomers), celotriose (3 glucose monomers), celotetraose (4 glucose monomers), cellopentaose (5 glucose monomers), and cellohexaose (6 glucose monomers).

As used herein, sugar refers to monosaccharides (e.g., glucose, fructose, galactose, xylose, arabinose), disaccharides (e.g., cellobiose, sucrose, lactose, maltose), and oligosaccharides (typically containing 3 to 10 component monosaccharides).

Polynucleotides of the Invention

The invention herein relates to host cells and methods of using such host cells where the host cells comprise recombinant polynucleotides encoding polypeptides capable of transporting various sugars.

As used herein, the terms "polynucleotide," "nucleic acid sequence," "sequence of nucleic acids," and variations thereof shall be generic to polydeoxyribonucleotides (containing 2-deoxy-D-ribose), to polyribonucleotides (containing D-ribose), to any other type of polynucleotide that is an N-glycoside of a purine or pyrimidine base, and to other polymers containing non-nucleotidic backbones, provided that the polymers contain nucleobases in a configuration that allows for base pairing and base stacking, as found in DNA and RNA. Thus, these terms include known types of nucleic acid sequence modifications, for example, substitution of one or more of the naturally occurring nucleotides with an analog; inter-nucleotide modifications, such as, for example, those with uncharged linkages (e.g., methyl phosphonates, phosphotriesters, phosphoramidates, carbamates, etc.), with negatively charged linkages (e.g., phosphorothioates, phosphorodithioates, etc.), and with positively charged linkages (e.g., aminoalkylphosphoramidates, aminoalkylphosphotriesters); those containing pendant moieties, such as, for example, proteins (including nucleases, toxins, antibodies, signal peptides, poly-L-lysine, etc.); those with intercalators (e.g., acridine, psoralen, etc.); and those containing chelators (e.g., metals, radioactive metals, boron, oxidative metals, etc.). As used herein, the symbols for nucleotides and polynucleotides

are those recommended by the IUPAC-IUB Commission of Biochemical Nomenclature (Biochem. 9:4022, 1970).

As used herein, a “polypeptide” is an amino acid sequence comprising a plurality of consecutive polymerized amino acid residues (e.g., at least about 15 consecutive polymerized amino acid residues, optionally at least about 30 consecutive polymerized amino acid residues, at least about 50 consecutive polymerized amino acid residues). In many instances, a polypeptide comprises a polymerized amino acid residue sequence that is a transporter, a transcription factor, a predicted protein of unknown function, or a domain or portion or fragment thereof. A transporter is involved in the movement of ions, small molecules, or macromolecules, such as a carbohydrate, across a biological membrane. A transcription factor can regulate gene expression and may increase or decrease gene expression in a host cell. The polypeptide optionally comprises modified amino acid residues, naturally occurring amino acid residues not encoded by a codon, and non-naturally occurring amino acid residues.

As used herein, “protein” refers to an amino acid sequence, oligopeptide, peptide, polypeptide, or portions thereof whether naturally occurring or synthetic.

Recombinant polynucleotides of the invention include any polynucleotides that encode a polypeptide encoded by any of the genes listed in Table 10, in Supplemental Data, Dataset S1, page 3 in Tian et al., 2009; in Tables 14, 15, 16, 29; or in FIG. 76. In preferred embodiments, polynucleotides of the invention include any polynucleotides that encode a polypeptide encoded by any of the sequences NCU00801, NCU00809, NCU08114, NCU00130, NCU00821, NCU04963, NCU06138, STL12/XUT6, SUT2, SUT3, XUT1, XUT3, NCU07705, NCU05137, NCU01517, NCU09133, or NCU10040.

In certain embodiments, the recombinant polynucleotides of the invention encode polypeptides having at least about 20%, or at least about 29%, or at least about 30%, or at least about 40%, or at least about 50%, or at least about 55%, or at least about 60%, or at least about 65%, or at least about 70%, or at least about 75%, or at least about 80%, or at least about 85%, or at least about 90%, or at least about 92%, or at least about 94%, or at least about 96%, or at least about 98%, or at least about 99%, or at least about 100% amino acid residue sequence identity to a polypeptide encoded by any of the genes listed in genes listed in Table 10, in Supplemental Data, Dataset S1, page 3 in Tian et al., 2009; in Tables 14, 15, 16, 29; or in FIG. 76. In preferred embodiments, the polynucleotides of the invention encode polypeptides having at least about 20%, or at least about 29%, or at least about 30%, or at least about 40%, or at least about 50%, or at least about 55%, or at least about 60%, or at least about 65%, or at least about 70%, or at least about 75%, or at least about 80%, or at least about 85%, or at least about 90%, or at least about 92%, or at least about 94%, or at least about 96%, or at least about 98%, or at least about 99%, or at least about 100% amino acid residue sequence identity to a polypeptide encoded by any of the sequences NCU00801, NCU00809, NCU08114, NCU00130, NCU00821, NCU04963, NCU06138, STL12/XUT6, SUT2, SUT3, XUT1, XUT3, NCU07705, NCU05137, NCU01517, NCU09133, or NCU10040.

Polynucleotides of the invention further include polynucleotides that encode conservatively modified variants of polypeptides encoded by the genes listed above. “Conservatively modified variants” as used herein include individual substitutions, deletions or additions to a polypeptide sequence which result in the substitution of an amino acid with a chemically similar amino acid. Conservative substitution tables providing functionally similar amino acids are well known in the art. Such conservatively modified variants are in addition to and do not exclude polymorphic variants, interspecies homologs, and alleles of the disclosure. The following eight groups contain amino acids that are conservative substitutions for one another: 1) Alanine (A), Glycine (G); 2) Aspartic acid (D), Glutamic acid (E); 3) Asparagine (N), Glutamine (Q); 4) Arginine (R), Lysine (K); 5) Isoleucine (I), Leucine (L), Methionine (M), Valine (V); 6) Phenylalanine (F), Tyrosine (Y), Tryptophan (W); 7) Serine (S), Threonine (T); and 8) Cysteine (C), Methionine (M) (see, e.g., Creighton, Protein (1984)).

Polynucleotides of the invention further include polynucleotides that encode homologs or orthologs of polypeptides encoded by any of the genes listed in Table 10, in Supplemental Data, Dataset S1, page 3 in Tian et al., 2009; in Tables 14, 15, 16, 29; or in FIG. 76. “Homology” as used herein refers to sequence similarity between a reference sequence and at least a fragment of a second sequence. Homologs may be identified by any method known in the art, preferably, by using the BLAST tool to compare a reference sequence to a single second sequence or fragment of a sequence or to a database of sequences. As described below, BLAST will compare sequences based upon percent identity and similarity. “Orthology” as used herein refers to genes in different species that derive from a common ancestor gene.

The terms “identical” or percent “identity,” in the context of two or more nucleic acids or polypeptide sequences, refer to two or more sequences or subsequences that are the same. Two sequences are “substantially identical” if two sequences have a specified percentage of amino acid residues or nucleotides that are the same (i.e., 29% identity, optionally 30%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 99% or 100% identity over a specified region, or,

Gene Name/Locus	Alter- nate Name	NCBI Reference Sequence/GenBank Accession Number	Organism
NCU00801	cbt1	XP_963801.1/EAA34565	<i>N. crassa</i>
NCU00809		XP_964302.1/EAA35116.1	<i>N. crassa</i>
NCU00821	AN25	XP_964364.2/EAA35128.2	<i>N. crassa</i>
NCU00988	Xy33	XP_963898.1/EAA34662.1	<i>N. crassa</i>
NCU01231		XP_961597.2/EAA32361.2	<i>N. crassa</i>
NCU01494	AN49	XP_955927.2/EAA26691.2	<i>N. crassa</i>
NCU02188	AN28-3	XP_959582.2/EAA30346.2	<i>N. crassa</i>
NCU04537	Xy50	XP_955977.1/EAA26741.1	<i>N. crassa</i>
NCU04963	AN29-2	XP_959411.2/EAA30175.2	<i>N. crassa</i>
NCU05519		XP_960481.1/EAA31245.1	<i>N. crassa</i>
NCU05853		XP_959844.1/EAA30608.1	<i>N. crassa</i>
NCU05897		XP_959888.1/EAA30652.1	<i>N. crassa</i>
NCU06138	Xy31	XP_960000.1/EAA30764.1	<i>N. crassa</i>
NCU08114	cbt2	XP_963873.1/EAA34637.1	<i>N. crassa</i>
NCU09287	AN41	XP_958139.1/EAA28903.1	<i>N. crassa</i>
NCU10021		XP_958069.2/EAA28833.2	<i>N. crassa</i>
XP_001387242	Ap26	XP_001387242	<i>P. stipitis</i>
HGT3	Xyp30-	XP_001386715.1/ABN68686.1	<i>P. stipitis</i>
	1		
STL1	Xyp30	XP_001383774.1/ABN65745.1	<i>P. stipitis</i>
STL12/XUT6	Xyp29	XP_001386589.1/ABN68560.1	<i>P. stipitis</i>
SUT2	Ap31	XP_001384295.2/ABN66266.2	<i>P. stipitis</i>
SUT3	Xyp37	XP_001386019.2/ABN67990.2	<i>P. stipitis</i>
XUT1	Xyp32	XP_001385583.1/ABN67554.1	<i>P. stipitis</i>
XUT2	Xyp31	XP_001387242.1/EAZ63219.2	<i>P. stipitis</i>
XUT3	Xyp33	XP_001387138.1/EAZ63115.1	<i>P. stipitis</i>
XUT7	Xyp28	XP_001387067.1/EAZ63044.1	<i>P. stipitis</i>
NCU07705	cdr-1	XP_962291.1/EAA33055	<i>N. crassa</i>
NCU05137		XP_956635.1/EAA27399	<i>N. crassa</i>
NCU01517		XP_956966.1/EAA27730	<i>N. crassa</i>
NCU09133		XP_958905.1/EAA29669	<i>N. crassa</i>
NCU10040			

when not specified, over the entire sequence), when compared and aligned for maximum correspondence over a comparison window, or designated region as measured using one of the following sequence comparison algorithms or by manual alignment and visual inspection. Optionally, the identity exists over a region that is at least about 50 nucleotides (or 10 amino acids) in length, or more preferably over a region that is 100 to 500 or 1000 or more nucleotides (or 20, 50, 200, or more amino acids) in length.

For sequence comparison, typically one sequence acts as a reference sequence, to which test sequences are compared. When using a sequence comparison algorithm, test and reference sequences are entered into a computer, subsequence coordinates are designated, if necessary, and sequence algorithm program parameters are designated. Default program parameters can be used, or alternative parameters can be designated. The sequence comparison algorithm then calculates the percent sequence identities for the test sequences relative to the reference sequence, based on the program parameters. When comparing two sequences for identity, it is not necessary that the sequences be contiguous, but any gap would carry with it a penalty that would reduce the overall percent identity. For blastn, the default parameters are Gap opening penalty=5 and Gap extension penalty=2. For blastp, the default parameters are Gap opening penalty=11 and Gap extension penalty=1.

A "comparison window," as used herein, includes reference to a segment of any one of the number of contiguous positions including, but not limited to from 20 to 600, usually about 50 to about 200, more usually about 100 to about 150 in which a sequence may be compared to a reference sequence of the same number of contiguous positions after the two sequences are optimally aligned. Methods of alignment of sequences for comparison are well known in the art. Optimal alignment of sequences for comparison can be conducted, e.g., by the local homology algorithm of Smith and Waterman (1981), by the homology alignment algorithm of Needleman and Wunsch (1970) J Mol Biol 48(3):443-453, by the search for similarity method of Pearson and Lipman (1988) Proc Natl Acad Sci USA 85(8):2444-2448, by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, Wis.), or by manual alignment and visual inspection [see, e.g., Brent et al., (2003) Current Protocols in Molecular Biology, John Wiley & Sons, Inc. (Ringbou Ed.)].

Two examples of algorithms that are suitable for determining percent sequence identity and sequence similarity are the BLAST and BLAST 2.0 algorithms, which are described in Altschul et al. (1997) Nucleic Acids Res 25(17):3389-3402 and Altschul et al. (1990) J. Mol Biol 215(3)-403-410, respectively. Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information. This algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence, which either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as the neighborhood word score threshold (Altschul et al., supra). These initial neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Cumulative scores are calculated using, for nucleotide sequences, the parameters M (reward score for a pair of matching residues; always >0) and N (penalty score for mismatching residues; always <0). For amino acid sequences, a

scoring matrix is used to calculate the cumulative score. Extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W, T, and X determine the sensitivity and speed of the alignment. The BLASTN program (for nucleotide sequences) uses as defaults a wordlength (W) of 11, an expectation (E) or 10, M=5, N=-4, and a comparison of both strands. For amino acid sequences, the BLASTP program uses as defaults a wordlength of 3, and expectation (E) of 10, and the BLOSUM62 scoring matrix [see Henikoff and Henikoff, (1992) Proc Natl Acad Sci USA 89(22):10915-10919] alignments (B) of 50, expectation (E) of 10, M=5, N=-4, and a comparison of both strands.

The BLAST algorithm also performs a statistical analysis of the similarity between two sequences (see, e.g., Karlin and Altschul, (1993) Proc Natl Acad Sci USA 90(12):5873-5877). One measure of similarity provided by the BLAST algorithm is the smallest sum probability (P(N)), which provides an indication of the probability by which a match between two nucleotide or amino acid sequences would occur by chance. For example, a nucleic acid is considered similar to a reference sequence if the smallest sum probability in a comparison of the test nucleic acid to the reference nucleic acid is less than about 0.2, more preferably less than about 0.01, and most preferably less than about 0.001.

Other than percentage of sequence identity noted above, another indication that two nucleic acid sequences or polypeptides are substantially identical is that the polypeptide encoded by the first nucleic acid is immunologically cross-reactive with the antibodies raised against the polypeptide encoded by the second nucleic acid, as described below. Thus, a polypeptide is typically substantially identical to a second polypeptide, for example, where the two peptides differ only by conservative substitutions. Another indication that two nucleic acid sequences are substantially identical is that the two molecules or their complements hybridize to each other under stringent conditions, as described below. Yet another indication that two nucleic acid sequences are substantially identical is that the same primers can be used to amplify the sequence.

As described herein, polynucleotides of the invention include members of the Major Facilitator Superfamily sugar transporter family, including NCU00988, NCU10021, NCU04963, NCU06138, NCU00801, NCU08114, and NCU05853. Members of the Major Facilitator Superfamily (MFS) (Transporter Classification #2.A.1) of transporters almost always consist of 12 transmembrane α -helices, with an intracellular N- and C-terminus (S. S. Pao, I. T. Paulsen, M. H. Saier, Jr., *Microbiol Mol Biol Rev* 62, 1 (March 1998)). While the primary sequence of MFS transporters varies widely, all are thought to share the tertiary structure of the *E. coli* lactose permease (LacY) (J. Abramson et al., *Science* 301, 610 (Aug. 1, 2003)), and the *E. coli* Pi/glycerol-3-phosphate (GlpT) (Y. Huang, M. J. Lemieux, J. Song, M. Auer, D. N. Wang, *Science* 301, 616 (Aug. 1, 2003)). In these examples the six N- and C-terminal helices form two distinct domains connected by a long cytoplasmic loop between helices 6 and 7. This symmetry corresponds to a duplication event thought to have given rise to the MFS. Substrate binds within a hydrophilic cavity formed by helices 1, 2, 4, and 5 of the N-terminal domain, and helices 7, 8, 10, and 11 of the C-terminal domain. This cavity is stabilized by helices 3, 6, 9, and 12.

The Sugar Transporter family of the MFS (Transporter Classification #2.A.1.1) is defined by motifs found in transmembrane helices 6 and 12 (PESPR (SEQ ID NO: 9)/PETK (SEQ ID NO: 10)), and loops 2 and 8 (GRR/GRK) (M. C. Maiden, E. O. Davis, S. A. Baldwin, D. C. Moore, P. J. Henderson, *Nature* 325, 641 (Feb. 12-18, 1987)). The entire Hidden Markov Model (HMM) for this family can be viewed at pfam.janelia.org/family/PF00083#tabview=tab3. PROSITE (N. Hulo et al., *Nucleic Acids Res* 34, D227 (Jan. 1, 2006)) uses two motifs to identify members of this family. The first is [LIVMSTAG]-[LIVMFSAG]-{SH}-{RDE}-[LIVMSA]-[DE]-{TD}-[LIVMFYWA]-G-R-[RK]-x(4,6)-[GSTA] (SEQ ID NO: 11). The second is [LIVMF]-x-G-[LIVMFA]-{V}-x-G-{KP}-x(7)-[LIFY]-x(2)-[EQ]-x(6)-[RK] (SEQ ID NO: 12). As an example of how to read a PROSITE motif, the following motif, [AC]-x-V-x(4)-{ED}, is translated as: [Ala or Cys]-any-Val-any-any-any-any-{any but Glu or Asp} (SEQ ID NO: 13).

As described herein, NCU00801, NCU00809, NCU08114, XP_001268541.1, and LAC2 were discovered to encode polypeptides that transport cellobextrins. Further, alanine scanning experiments and sequence analyses were used to determine that a recombinant polypeptide containing 12 transmembrane α -helices, and one or more of the sequences selected from the group consisting of SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, and SEQ ID NO: 8 encodes a polypeptide that transports cellobextrin.

Thus, in one aspect, the invention provides a polynucleotide encoding a polypeptide containing transmembrane α -helix 1, α -helix 2, α -helix 3, α -helix 4, α -helix 5, α -helix 6, α -helix 7, α -helix 8, α -helix 9, α -helix 10, α -helix 11, α -helix 12, and transmembrane α -helix 1 comprises SEQ ID NO: 1. In another aspect, the invention provides a polynucleotide encoding a polypeptide containing transmembrane α -helix 1, α -helix 2, α -helix 3, α -helix 4, α -helix 5, α -helix 6, α -helix 7, α -helix 8, α -helix 9, α -helix 10, α -helix 11, α -helix 12, and transmembrane α -helix 2 comprises SEQ ID NO: 2. In another aspect, the invention provides a polynucleotide encoding a polypeptide containing transmembrane α -helix 1, α -helix 2, α -helix 3, α -helix 4, α -helix 5, α -helix 6, α -helix 7, α -helix 8, α -helix 9, α -helix 10, α -helix 11, α -helix 12, and a loop connecting transmembrane α -helix 2 and transmembrane α -helix 3 comprises SEQ ID NO: 3. In another aspect, the invention provides a polynucleotide encoding a polypeptide containing transmembrane α -helix 1, α -helix 2, α -helix 3, α -helix 4, α -helix 5, α -helix 6, α -helix 7, α -helix 8, α -helix 9, α -helix 10, α -helix 11, α -helix 12, and transmembrane α -helix 5 comprises SEQ ID NO: 4. In another aspect, the invention provides a polynucleotide encoding a polypeptide containing transmembrane α -helix 1, α -helix 2, α -helix 3, α -helix 4, α -helix 5, α -helix 6, α -helix 7, α -helix 8, α -helix 9, α -helix 10, α -helix 11, α -helix 12, and transmembrane α -helix 6 comprises SEQ ID NO: 5. In another aspect, the invention provides a polynucleotide encoding a polypeptide containing transmembrane α -helix 1, α -helix 2, α -helix 3, α -helix 4, α -helix 5, α -helix 6, α -helix 7, α -helix 8, α -helix 9, α -helix 10, α -helix 11, α -helix 12, and sequence between transmembrane α -helix 6 and transmembrane α -helix 7 comprises SEQ ID NO: 6. In another aspect, the invention provides a polynucleotide encoding a polypeptide containing transmembrane α -helix 1, α -helix 2, α -helix 3, α -helix 4, α -helix 5, α -helix 6, α -helix 7, α -helix 8, α -helix 9, α -helix 10, α -helix 11, α -helix 12, and transmembrane α -helix 7 comprises SEQ ID NO: 7. In another aspect, the invention provides a polynucleotide encoding a polypeptide containing transmembrane α -helix 1, α -helix 2, α -helix 3,

α -helix 4, α -helix 5, α -helix 6, α -helix 7, α -helix 8, α -helix 9, α -helix 10, α -helix 11, α -helix 12, and transmembrane α -helix 10 and transmembrane α -helix 11 and the sequence between them comprise SEQ ID NO: 8.

Each of the above described aspects may be combined in any number of combinations. A polynucleotide according to any of these aspects may encode a polypeptide containing 1, 2, 3, 4, 5, 6, or 7 of any of SEQ ID NOs: 1-8, or the polypeptide may contain all of SEQ ID NOs: 1-8. For example, a polynucleotide may encode a polypeptide containing transmembrane α -helix 1, α -helix 2, α -helix 3, α -helix 4, α -helix 5, α -helix 6, α -helix 7, α -helix 8, α -helix 9, α -helix 10, α -helix 11, α -helix 12, where transmembrane α -helix 1 comprises SEQ ID NO: 1, a loop connecting transmembrane α -helix 2 and transmembrane α -helix 3 comprises SEQ ID NO: 3, and transmembrane α -helix 7 comprises SEQ ID NO: 7. Or, in another example, a polynucleotide may encode a polypeptide containing transmembrane α -helix 1, α -helix 2, α -helix 3, α -helix 4, α -helix 5, α -helix 6, α -helix 7, α -helix 8, α -helix 9, α -helix 10, α -helix 11, α -helix 12, where transmembrane α -helix 2 comprises SEQ ID NO: 2, transmembrane α -helix 3 comprises SEQ ID NO: 3, transmembrane α -helix 6 comprises SEQ ID NO: 5, and transmembrane α -helix 10 and transmembrane α -helix 11 and the sequence between them comprise SEQ ID NO: 8.

In certain embodiments of the above described aspects, the polypeptide has at least 29%, at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 99%, or at least 100% amino acid identity to NCU00801 or NCU08114.

As further described herein, NCU08221 and STL12/XUT6 were discovered to encode polypeptides that transport xylose. XUT1 was discovered to encode a polypeptide that transports arabinose. NCU06138 was discovered to encode a polypeptide that transports arabinose or glucose. SUT2, SUT3, and XUT3 were discovered to encode polypeptides that transport xylose or glucose. NCU04963 was discovered to encode a polypeptide that transports xylose, arabinose, or glucose. In preferred embodiments, polynucleotides of the invention include recombinant polynucleotides encoding a NCU08221 or STL12/XUT6 polypeptide, where the polypeptide transports xylose. In other preferred embodiments, polynucleotides of the invention include recombinant polynucleotides encoding a XUT1 polypeptide, where the polypeptide transports arabinose. In other preferred embodiments, polynucleotides of the invention include recombinant polynucleotides encoding a NCU06138 polypeptide, where the polypeptide transports arabinose or glucose. In other preferred embodiments, polynucleotides of the invention include recombinant polynucleotides encoding a SUT2, SUT3, or XUT3 polypeptide, where the polypeptide transports xylose or glucose. In other preferred embodiments, polynucleotides of the invention include recombinant polynucleotides encoding a NCU04963 polypeptide, where the polypeptide transports xylose, arabinose, or glucose.

The polynucleotides of the invention that encode polypeptides encoded by NCU07705 are predicted by FunCat (Ruepp, 2004; webpage broad.mit.edu/annotation/genome/neurospora/Home.html) to encode an unclassified protein. However, BLAST analysis of the polypeptide encoded by NCU07705 revealed that the polypeptide has high similarity to many C6 zinc finger domain containing transcription factors (see FIG. 1; a list of exemplary homologs can be found in FIG. 23 of related U.S. Appl. No. 61/271,833). Polynucleotides of the invention include polynucleotides that encode

35

these homologs of the polypeptide encoded by NCU07705 or any other homologs identified with any methods known in the art.

In another aspect of the invention, polynucleotides of the invention include those polynucleotides that encode polypeptides encoded by NCU05137. FunCat classifies the polypeptide encoded by NCU05137 to be an unclassified protein. However, NCU05137 is highly conserved in the genomes of a number of filamentous ascomycete fungi (see FIG. 2). Polynucleotides of the invention include polynucleotides that encode these homologs of the polypeptide encoded by NCU05137 or any other homologs identified with any methods known in the art.

In another aspect of the invention, polynucleotides of the invention include those polynucleotides that encode polypeptides encoded by NCU01517, NCU09133, or NCU10040. FunCat classifies the polypeptide encoded by NCU01517 to be a glucoamylase precursor. FunCat classifies the polypeptides encoded by NCU09133 and NCU10040 to be unclassified proteins. Polynucleotides of the invention include polynucleotides that encode these homologs of the polypeptide encoded by NCU01517, NCU09133, or NCU10040 or any other homologs identified with any methods known in the art.

Predicted functions of these polypeptides can be confirmed by performing functional analyses of the polynucleotide and its encoded protein. These analyses may include, for example, phenotypic analysis of strains containing deletions of the polynucleotide, genetic complementation experiments, phenotypic analysis of strains over expressing a wild-type copy of the polynucleotide, expression and purification of a recombinant form of the polypeptide, and subsequent characterization of the biochemical properties and activity of the recombinant polypeptide.

Sequences of the polynucleotides of the invention are prepared by any suitable method known to those of ordinary skill in the art, including, for example, direct chemical synthesis or cloning. For direct chemical synthesis, formation of a polymer of nucleic acids typically involves sequential addition of 3'-blocked and 5'-blocked nucleotide monomers to the terminal 5'-hydroxyl group of a growing nucleotide chain, wherein each addition is effected by nucleophilic attack of the terminal 5'-hydroxyl group of the growing chain on the 3'-position of the added monomer, which is typically a phosphorus derivative, such as a phosphotriester, phosphoramidite, or the like. Such methodology is known to those of ordinary skill in the art and is described in the pertinent texts and literature [e.g., in Matteucci et al., (1980) *Tetrahedron Lett* 21:719-722; U.S. Pat. Nos. 4,500,707; 5,436,327; and 5,700,637]. In addition, the desired sequences may be isolated from natural sources by splitting DNA using appropriate restriction enzymes, separating the fragments using gel electrophoresis, and thereafter, recovering the desired nucleic acid sequence from the gel via techniques known to those of ordinary skill in the art, such as utilization of polymerase chain reactions (PCR; e.g., U.S. Pat. No. 4,683,195).

Each polynucleotide of the invention can be incorporated into an expression vector. "Expression vector" or "vector" refers to a compound and/or composition that transduces, transforms, or infects a host cell, thereby causing the cell to express nucleic acids and/or proteins other than those native to the cell, or in a manner not native to the cell. An "expression vector" contains a sequence of nucleic acids (ordinarily RNA or DNA) to be expressed by the host cell. Optionally, the expression vector also comprises materials to aid in achieving entry of the nucleic acid into the host cell, such as a virus, liposome, protein coating, or the like. The expression vectors contemplated for use in the present invention include those

36

into which a nucleic acid sequence can be inserted, along with any preferred or required operational elements. Further, the expression vector must be one that can be transferred into a host cell and replicated therein. Preferred expression vectors are plasmids, particularly those with restriction sites that have been well documented and that contain the operational elements preferred or required for transcription of the nucleic acid sequence. Such plasmids, as well as other expression vectors, are well known to those of ordinary skill in the art.

Incorporation of the individual polynucleotides may be accomplished through known methods that include, for example, the use of restriction enzymes (such as BamHI, EcoRI, HhaI, XbaI, XmaI, and so forth) to cleave specific sites in the expression vector, e.g., plasmid. The restriction enzyme produces single stranded ends that may be annealed to a polynucleotide having, or synthesized to have, a terminus with a sequence complementary to the ends of the cleaved expression vector. Annealing is performed using an appropriate enzyme, e.g., DNA ligase. As will be appreciated by those of ordinary skill in the art, both the expression vector and the desired polynucleotide are often cleaved with the same restriction enzyme, thereby assuring that the ends of the expression vector and the ends of the polynucleotide are complementary to each other. In addition, DNA linkers maybe used to facilitate linking of nucleic acids sequences into an expression vector.

A series of individual polynucleotides can also be combined by utilizing methods that are known to those having ordinary skill in the art (e.g., U.S. Pat. No. 4,683,195).

For example, each of the desired polynucleotides can be initially generated in a separate PCR. Thereafter, specific primers are designed such that the ends of the PCR products contain complementary sequences. When the PCR products are mixed, denatured, and reannealed, the strands having the matching sequences at their 3' ends overlap and can act as primers for each other. Extension of this overlap by DNA polymerase produces a molecule in which the original sequences are "spliced" together. In this way, a series of individual polynucleotides may be "spliced" together and subsequently transduced into a host cell simultaneously. Thus, expression of each of the plurality of polynucleotides is affected.

Individual polynucleotides, or "spliced" polynucleotides, are then incorporated into an expression vector. The invention is not limited with respect to the process by which the polynucleotide is incorporated into the expression vector. Those of ordinary skill in the art are familiar with the necessary steps for incorporating a polynucleotide into an expression vector. A typical expression vector contains the desired polynucleotide preceded by one or more regulatory regions, along with a ribosome binding site, e.g., a nucleotide sequence that is 3-9 nucleotides in length and located 3-11 nucleotides upstream of the initiation codon in *E. coli*. See Shine and Dalgarno (1975) *Nature* 254(5495):34-38 and Steitz (1979) *Biological Regulation and Development* (ed. Goldberger, R. F.), 1:349-399 (Plenum, N.Y.).

The term "operably linked" as used herein refers to a configuration in which a control sequence is placed at an appropriate position relative to the coding sequence of the DNA sequence or polynucleotide such that the control sequence directs the expression of a polypeptide.

Regulatory regions include, for example, those regions that contain a promoter and an operator. A promoter is operably linked to the desired polynucleotide, thereby initiating transcription of the polynucleotide via an RNA polymerase enzyme. An operator is a sequence of nucleic acids adjacent to the promoter, which contains a protein-binding domain

where a repressor protein can bind. In the absence of a repressor protein, transcription initiates through the promoter. When present, the repressor protein specific to the protein-binding domain of the operator binds to the operator, thereby inhibiting transcription. In this way, control of transcription is accomplished, based upon the particular regulatory regions used and the presence or absence of the corresponding repressor protein. Examples include lactose promoters (Lad repressor protein changes conformation when contacted with lactose, thereby preventing the Lad repressor protein from binding to the operator) and tryptophan promoters (when complexed with tryptophan, TrpR repressor protein has a conformation that binds the operator; in the absence of tryptophan, the TrpR repressor protein has a conformation that does not bind to the operator). Another example is the tac promoter (see de Boer et al., (1983) Proc Natl Acad Sci USA 80(1):21-25). As will be appreciated by those of ordinary skill in the art, these and other expression vectors may be used in the present invention, and the invention is not limited in this respect.

Although any suitable expression vector may be used to incorporate the desired sequences, readily available expression vectors include, without limitation: plasmids, such as pSC101, pBR322, pBBRIMCS-3, pUR, pEX, pMRIOO, pCR4, pBAD24, pUC19; bacteriophages, such as M1 3 phage and λ phage. Of course, such expression vectors may only be suitable for particular host cells. One of ordinary skill in the art, however, can readily determine through routine experimentation whether any particular expression vector is suited for any given host cell. For example, the expression vector can be introduced into the host cell, which is then monitored for viability and expression of the sequences contained in the vector. In addition, reference may be made to the relevant texts and literature, which describe expression vectors and their suitability to any particular host cell.

Host Cells of the Invention

The invention herein relates to host cells containing recombinant polynucleotides encoding polypeptides where the polypeptides transport celldextrin or a pentose into the cell. Further described herein are methods of increasing transport of celldextrin into a host cell, methods of increasing growth of a host cell on a medium containing celldextrin, methods of co-fermenting cellulose-derived and hemicellulose-derived sugars, and methods of making hydrocarbons or hydrocarbon derivatives by providing a host cell containing a recombinant polynucleotide encoding a polypeptide where the polypeptide transports celldextrin into the cell. Further described herein are methods of increasing transport of a pentose into a host cell, methods of increasing growth of a host cell on a medium containing pentose sugars, and methods of making hydrocarbons or hydrocarbon derivatives by providing a host cell containing a recombinant polynucleotide encoding a polypeptide where the polypeptide transports a pentose into the cell.

"Host cell" and "host microorganism" are used interchangeably herein to refer to a living biological cell that can be transformed via insertion of recombinant DNA or RNA. Such recombinant DNA or RNA can be in an expression vector. Thus, a host organism or cell as described herein may be a prokaryotic organism (e.g., an organism of the kingdom Eubacteria) or a eukaryotic cell. As will be appreciated by one of ordinary skill in the art, a prokaryotic cell lacks a membrane-bound nucleus, while a eukaryotic cell has a membrane-bound nucleus.

Any prokaryotic or eukaryotic host cell may be used in the present invention so long as it remains viable after being transformed with a sequence of nucleic acids. Preferably, the

host cell is not adversely affected by the transduction of the necessary nucleic acid sequences, the subsequent expression of the proteins (e.g., transporters), or the resulting intermediates. Suitable eukaryotic cells include, but are not limited to, 5 fungal, plant, insect or mammalian cells.

In preferred embodiments, the host is a fungal strain. "Fungi" as used herein includes the phyla Ascomycota, Basidiomycota, Chytridiomycota, and Zygomycota (as defined by Hawksworth et al., In, Ainsworth and Bisby's 10 Dictionary of The Fungi, 8th edition, 1995, CAB International, University Press, Cambridge, UK) as well as the Oomycota (as cited in Hawksworth et al., 1995, supra, page 171) and all mitosporic fungi (Hawksworth et al., 1995, supra).

15 In particular embodiments, the fungal host is a yeast strain. "Yeast" as used herein includes ascosporogenous yeast (Endomycetales), basidiosporogenous yeast, and yeast belonging to the Fungi Imperfetti (Blastomycetes). Since the classification of yeast may change in the future, for the purposes 20 of this invention, yeast shall be defined as described in Biology and Activities of Yeast (Skinner, F. A., Passmore, S. M., and Davenport, R. R., eds, Soc. App. Bacteriol. Symposium Series No. 9, 1980).

25 In a more preferred embodiment, the yeast host is a *Candida*, *Hansenula*, *Kluyveromyces*, *Pichia*, *Saccharomyces*, *Schizosaccharomyces*, or *Yarrowia* strain.

In certain embodiments, the yeast host is a *Saccharomyces carlsbergensis* (Todkar, 2010), *Saccharomyces cerevisiae* (Duarte et al., 2009), *Saccharomyces diastaticus*, *Saccharomyces douglasii*, *Saccharomyces kluyveri*, *Saccharomyces norbensis*, *Saccharomyces monocensis* (GB-Analysts Reports, 2008), *Saccharomyces bayanus* (Kristen Publicover, 2010), *Saccharomyces pastorianus* (Nakao et al., 2007), *Saccharomyces pombe* (Mousdale, 2008), or *Saccharomyces oviformis* strain. In other preferred embodiments, the yeast host is *Kluyveromyces lactis* (O. W. Merten, 2001), *Kluyveromyces fragilis* (Pestal et al., 2006; Siso, 1996), *Kluyveromyces marxianus* (K. Kourkoutas et al., 2008), *Pichia stipitis* (Almeida et al., 2008), *Candida shehatae* (Ayhan Demirbas, 40 2003), or *Candida tropicalis* (Jamai et al., 2006). In other embodiments, the yeast host may be *Yarrowia lipolytica* (Biryukova E. N., 2009), *Brettanomyces custersii* (Spindler D. D. et al., 1992), or *Zygosaccharomyces roux* (Chaabane et al., 2006).

45 In another particular embodiment, the fungal host is a filamentous fungal strain. "Filamentous fungi" include all filamentous forms of the subdivision Eumycota and Oomycota (as defined by Hawksworth et al., 1995, supra). The filamentous fungi are generally characterized by a mycelial wall composed of chitin, cellulose, glucan, chitosan, mannan, and other complex polysaccharides. Vegetative growth is by hyphal elongation and carbon catabolism is obligately aerobic. In contrast, vegetative growth by yeasts such as *Saccharomyces cerevisiae* is by budding of a unicellular thallus and 50 carbon catabolism may be fermentative.

In preferred embodiments, the filamentous fungal host is, but not limited to, an *Acremonium*, *Aspergillus*, *Fusarium*, *Humicola*, *Mucor*, *Myceliophthora*, *Neurospora*, *Penicillium*, *Scytalidium*, *Thielavia*, *Tolypocladium*, or *Trichoderma* 55 strain.

In certain embodiments, the filamentous fungal host is an *Aspergillus awamori*, *Aspergillus foetidus*, *Aspergillus japonicus*, *Aspergillus nidulans*, *Aspergillus niger*, or *Aspergillus oryzae* strain. In other embodiments, the filamentous fungal host is a *Fusarium bactrioides*, *Fusarium cerealis*, *Fusarium crookwellense*, *Fusarium culmorum*, *Fusarium graminearum*, *Fusarium gramininum*, *Fusarium het-*

erosporum, *Fusarium negundi*, *Fusarium oxysporum*, *Fusarium reticulatum*, *Fusarium roseum*, *Fusarium sambucinum*, *Fusarium sارcochroum*, *Fusarium sporotrichoides*, *Fusarium sulphureum*, *Fusarium torulosum*, *Fusarium trichothecoides*, or *Fusarium venenatum* strain. In yet other preferred embodiments, the filamentous fungal host is a *Humicola insolens*, *Humicola lanuginosa*, *Mucor miehei*, *Myceliophthora thermophila*, *Neurospora crassa*, *Penicillium purpurogenum*, *Scyphidium thermophilum*, *Sporotrichum thermophile* (Topakas et al., 2003), or *Thielavia terrestris* strain. In a further embodiment, the filamentous fungal host is a *Trichoderma harzianum*, *Trichoderma koningii*, *Trichoderma longibrachiatum*, *Trichoderma reesei*, or *Trichoderma viride* strain.

In other preferred embodiments, the host cell is prokaryotic, and in certain embodiments, the prokaryotes are *E. coli* (Dien, B. S. et al., 2003; Yomano, L. P. et al., 1998; Moniruzzaman et al., 1996), *Bacillus subtilis* (Susana Romero et al., 2007), *Zymomonas mobilis* (B. S. Dien et al., 2003; Weuster Botz, 1993; Alterthum and Ingram, 1989), *Clostridium* sp. (Zeikus, 1980; Lynd et al., 2002; Demain et al., 2005), *Clostridium phytofermentans* (Leschine S., 2010), *Clostridium thermocellum* (Lynd et al., 2002), *Clostridium beijerinckii* (Giles Clark, 2008), *Clostridium acetobutylicum* (*Moorella thermoacetica*) (Huang W. C. et al., 2004; Dominik et al., 2007), *Thermoanaerobacterium saccharolyticum* (Marietta Smith, 2009), or *Klebsiella oxytoca* (Dien, B. S. et al., 2003; Zhou et al., 2001; Brooks and Ingram, 1995). In other embodiments, the prokaryotic host cells are *Carboxydocella* sp. (Dominik et al., 2007), *Corynebacterium glutamicum* (Masayuki Inui, et al., 2004), *Enterobacteriaceae* (Ingram et al., 1995), *Erwinia chrysanthemi* (Zhou and Ingram, 2000; Zhou et al., 2001), *Lactobacillus* sp. (McCaskey, T. A., et al., 1994), *Pediococcus acidilactici* (Zhou, S. et al., 2003), *Rhodopseudomonas capsulata* (X. Y. Shi et al., 2004), *Streptococcus lactis* (J. C. Tang et al., 1988), *Vibrio furnissii* (L. P. Wackett, 2010), *Vibrio furnissii* M1 (Park et al., 2001), *Caldicellulosiruptor saccharolyticus* (Z. Kadar et al., 2004), or *Xanthomonas campestris* (S. T. Yang et al., 1987). In other embodiments, the host cells are cyanobacteria. Additional examples of bacterial host cells include, without limitation, those species assigned to the *Escherichia*, *Enterobacter*, *Azotobacter*, *Erwinia*, *Bacillus*, *Pseudomonas*, *Klebsiella*, *Proteus*, *Salmonella*, *Serratia*, *Shigella*, *Rhizobia*, *Vitreoscilla*, *Synechococcus*, *Synechocystis*, and *Paracoccus* taxonomical classes.

In especially preferred embodiments of the invention, the host cell is *Saccharomyces* sp., *Saccharomyces cerevisiae*, *Saccharomyces monacensis*, *Saccharomyces bayanus*, *Saccharomyces pastorianus*, *Saccharomyces carlsbergensis*, *Saccharomyces pombe*, *Kluyveromyces* sp., *Kluyveromyces marxianus*, *Kluyveromyces lactis*, *Kluyveromyces fragilis*, *Pichia stipitis*, *Sporotrichum thermophile*, *Candida shehatae*, *Candida tropicalis*, *Neurospora crassa*, *Zymomonas mobilis*, *Clostridium* sp., *Clostridium phytofermentans*, *Clostridium thermocellum*, *Clostridium beijerinckii*, *Clostridium acetobutylicum*, *Moorella thermoacetica*, *Escherichia coli*, *Klebsiella oxytoca*, *Thermoanaerobacterium saccharolyticum*, or *Bacillus subtilis*. *Saccharomyces* sp. may include Industrial *Saccharomyces* strains. Argueso et al. discuss the genome structure of an Industrial *Saccharomyces* strain commonly used in bioethanol production as well as specific gene polymorphisms that are important for bioethanol production (*Genome Research*, 19: 2258-2270, 2009).

The host cells of the present invention may be genetically modified in that recombinant nucleic acids have been introduced into the host cells, and as such the genetically modified

host cells do not occur in nature. The suitable host cell is one capable of expressing one or more nucleic acid constructs encoding one or more proteins for different functions.

“Recombinant nucleic acid” or “heterologous nucleic acid” or “recombinant polynucleotide” as used herein refers to a polymer of nucleic acids wherein at least one of the following is true: (a) the sequence of nucleic acids is foreign to (i.e., not naturally found in) a given host cell; (b) the sequence may be naturally found in a given host cell, but in an unnatural (e.g., greater than expected) amount; or (c) the sequence of nucleic acids comprises two or more subsequences that are not found in the same relationship to each other in nature. For example, regarding instance (c), a recombinant nucleic acid sequence will have two or more sequences from unrelated genes arranged to make a new functional nucleic acid. Specifically, the present invention describes the introduction of an expression vector into a host cell, wherein the expression vector contains a nucleic acid sequence coding for a protein that is not normally found in a host cell or contains a nucleic acid coding for a protein that is normally found in a cell but is under the control of different regulatory sequences. With reference to the host cell’s genome, then, the nucleic acid sequence that codes for the protein is recombinant.

In some embodiments, the host cell naturally produces any of the proteins encoded by the polynucleotides of the invention. The genes encoding the desired proteins may be heterologous to the host cell or these genes may be endogenous to the host cell but are operatively linked to heterologous promoters and/or control regions which result in the higher expression of the gene(s) in the host cell. In other embodiments, the host cell does not naturally produce the desired proteins, and comprises heterologous nucleic acid constructs capable of expressing one or more genes necessary for producing those molecules.

“Endogenous” as used herein with reference to a nucleic acid molecule or polypeptide and a particular cell or microorganism refers to a nucleic acid sequence or peptide that is in the cell and was not introduced into the cell using recombinant engineering techniques; for example, a gene that was present in the cell when the cell was originally isolated from nature.

“Genetically engineered” or “genetically modified” refer 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 a 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). Genetically engineering techniques include but are not limited to expression vectors, targeted homologous recombination and gene activation (see, for example, U.S. Pat. No. 5,272,071 to Chappel) and transactivation by engineered transcription factors (see, for example, Segal et al., (1999) *Proc Natl Acad Sci USA* 96(6): 2758-2763).

Genetic modifications that result in an increase in gene expression or function can be referred to as amplification, overproduction, overexpression, activation, enhancement, addition, or up-regulation of a gene. More specifically, reference to increasing the action (or activity) of enzymes or other proteins discussed herein generally refers to any genetic modification of the host cell in question which results in

41

increased expression and/or functionality (biological activity) of the enzymes or proteins and includes higher activity or action of the proteins (e.g., specific activity or *in vivo* enzymatic activity), reduced inhibition or degradation of the proteins, and overexpression of the proteins. For example, gene copy number can be increased, expression levels can be increased by use of a promoter that gives higher levels of expression than that of the native promoter, or a gene can be altered by genetic engineering or classical mutagenesis to increase the biological activity of an enzyme or action of a protein. Combinations of some of these modifications are also possible.

Genetic modifications which result in a decrease in gene expression, in the function of the gene, or in the function of the gene product (i.e., the protein encoded by the gene) can be referred to as inactivation (complete or partial), deletion, interruption, blockage, silencing, or down-regulation, or attenuation of expression of a gene. For example, a genetic modification in a gene which results in a decrease in the function of the protein encoded by such gene, can be the result of a complete deletion of the gene (i.e., the gene does not exist, and therefore the protein does not exist), a mutation in the gene which results in incomplete or no translation of the protein (e.g., the protein is not expressed), or a mutation in the gene which decreases or abolishes the natural function of the protein (e.g., a protein is expressed which has decreased or no enzymatic activity or action). More specifically, reference to decreasing the action of proteins discussed herein generally refers to any genetic modification in the host cell in question, which results in decreased expression and/or functionality (biological activity) of the proteins and includes decreased activity of the proteins (e.g., decreased transport), increased inhibition or degradation of the proteins as well as a reduction or elimination of expression of the proteins. For example, the action or activity of a protein of the present invention can be decreased by blocking or reducing the production of the protein, reducing protein action, or inhibiting the action of the protein. Combinations of some of these modifications are also possible. Blocking or reducing the production of a protein can include placing the gene encoding the protein under the control of a promoter that requires the presence of an inducing compound in the growth medium. By establishing conditions such that the inducer becomes depleted from the medium, the expression of the gene encoding the protein (and therefore, of protein synthesis) could be turned off. Blocking or reducing the action of a protein could also include using an excision technology approach similar to that described in U.S. Pat. No. 4,743,546, incorporated herein by reference. To use this approach, the gene encoding the protein of interest is cloned between specific genetic sequences that allow specific, controlled excision of the gene from the genome. Excision could be prompted by, for example, a shift in the cultivation temperature of the culture, as in U.S. Pat. No. 4,743,546, or by some other physical or nutritional signal.

In general, according to the present invention, an increase or a decrease in a given characteristic of a mutant or modified protein (e.g., enzyme activity, ability to transport compounds) is made with reference to the same characteristic of a wild-type (i.e., normal, not modified) protein that is derived from the same organism (from the same source or parent sequence), which is measured or established under the same or equivalent conditions. Similarly, an increase or decrease in a characteristic of a genetically modified host cell (e.g., expression and/or biological activity of a protein, or production of a product) is made with reference to the same characteristic of a wild-type host cell of the same species, and preferably the same strain, under the same or equivalent con-

42

ditions. Such conditions include the assay or culture conditions (e.g., medium components, temperature, pH, etc.) under which the activity of the protein (e.g., expression or biological activity) or other characteristic of the host cell is measured, as well as the type of assay used, the host cell that is evaluated, etc. As discussed above, equivalent conditions are conditions (e.g., culture conditions) which are similar, but not necessarily identical (e.g., some conservative changes in conditions can be tolerated), and which do not substantially change the effect on cell growth or enzyme expression or biological activity as compared to a comparison made under the same conditions.

Preferably, a genetically modified host cell that has a genetic modification that increases or decreases the activity of a given protein (e.g., a transporter, an enzyme) has an increase or decrease, respectively, in the activity or action (e.g., expression, production and/or biological activity) of the protein, as compared to the activity of the wild-type protein in a wild-type host cell, of at least about 5%, and more preferably at least about 10%, and more preferably at least about 15%, and more preferably at least about 20%, and more preferably at least about 25%, and more preferably at least about 30%, and more preferably at least about 35%, and more preferably at least about 40%, and more preferably at least about 45%, and more preferably at least about 50%, and more preferably at least about 55%, and more preferably at least about 60%, and more preferably at least about 65%, and more preferably at least about 70%, and more preferably at least about 75%, and more preferably at least about 80%, and more preferably at least about 85%, and more preferably at least about 90%, and more preferably at least about 95%, or any percentage, in whole integers between 5% and 100% (e.g., 6%, 7%, 8%, etc.). The same differences are preferred when comparing an isolated modified nucleic acid molecule or protein directly to the isolated wild-type nucleic acid molecule or protein (e.g., if the comparison is done *in vitro* as compared to *in vivo*).

In another aspect of the invention, a genetically modified host cell that has a genetic modification that increases or decreases the activity of a given protein (e.g., a transporter, an enzyme) has an increase or decrease, respectively, in the activity or action (e.g., expression, production and/or biological activity) of the protein, as compared to the activity of the wild-type protein in a wild-type host cell, of at least about 2-fold, and more preferably at least about 5-fold, and more preferably at least about 10-fold, and more preferably about 20-fold, and more preferably at least about 30-fold, and more preferably at least about 40-fold, and more preferably at least about 50-fold, and more preferably at least about 75-fold, and more preferably at least about 100-fold, and more preferably at least about 125-fold, and more preferably at least about 150-fold, or any whole integer increment starting from at least about 2-fold (e.g., 3-fold, 4-fold, 5-fold, 6-fold, etc.).

Host Cell Components

In one aspect, host cells of the invention contain a polynucleotide encoding a polypeptide containing transmembrane α -helix 1, α -helix 2, α -helix 3, α -helix 4, α -helix 5, α -helix 6, α -helix 7, α -helix 8, α -helix 9, α -helix 10, α -helix 11, α -helix 12, where transmembrane α -helix 1 comprises SEQ ID NO: 1. In another aspect, host cells of the invention contain a polynucleotide encoding a polypeptide containing transmembrane α -helix 1, α -helix 2, α -helix 3, α -helix 4, α -helix 5, α -helix 6, α -helix 7, α -helix 8, α -helix 9, α -helix 10, α -helix 11, α -helix 12, where transmembrane α -helix 2 comprises SEQ ID NO: 2. In another aspect, host cells of the invention contain a polynucleotide encoding a polypeptide containing transmembrane α -helix 1, α -helix 2, α -helix 3, α -helix 4, α -helix 5, α -helix 6, α -helix 7, α -helix 8, α -helix 9, α -helix 10, α -helix 11, α -helix 12, where transmembrane α -helix 2 comprises SEQ ID NO: 3. In another aspect, host cells of the invention contain a polynucleotide encoding a polypeptide containing transmembrane α -helix 1, α -helix 2, α -helix 3, α -helix 4, α -helix 5, α -helix 6, α -helix 7, α -helix 8, α -helix 9, α -helix 10, α -helix 11, α -helix 12, where transmembrane α -helix 2 comprises SEQ ID NO: 4.

9, α -helix 10, α -helix 11, α -helix 12, where the loop connecting transmembrane α -helix 2 and transmembrane α -helix 3 comprises SEQ ID NO: 3. In another aspect, host cells of the invention contain a polynucleotide encoding a polypeptide containing transmembrane α -helix 1, α -helix 2, α -helix 3, α -helix 4, α -helix 5, α -helix 6, α -helix 7, α -helix 8, α -helix 9, α -helix 10, α -helix 11, α -helix 12, where transmembrane α -helix 5 comprises SEQ ID NO: 4. In another aspect, host cells of the invention contain a polynucleotide encoding a polypeptide containing transmembrane α -helix 1, α -helix 2, α -helix 3, α -helix 4, α -helix 5, α -helix 6, α -helix 7, α -helix 8, α -helix 9, α -helix 10, α -helix 11, α -helix 12, where transmembrane α -helix 6 comprises SEQ ID NO: 5. In another aspect, host cells of the invention contain a polynucleotide encoding a polypeptide containing transmembrane α -helix 1, α -helix 2, α -helix 3, α -helix 4, α -helix 5, α -helix 6, α -helix 7, α -helix 8, α -helix 9, α -helix 10, α -helix 11, α -helix 12, where sequence between transmembrane α -helix 6 and transmembrane α -helix 7 comprises SEQ ID NO: 6. In another aspect, host cells of the invention contain a polynucleotide encoding a polypeptide containing transmembrane α -helix 1, α -helix 2, α -helix 3, α -helix 4, α -helix 5, α -helix 6, α -helix 7, α -helix 8, α -helix 9, α -helix 10, α -helix 11, α -helix 12, where transmembrane α -helix 7 comprises SEQ ID NO: 7. In another aspect, host cells of the invention contain a polynucleotide encoding a polypeptide containing transmembrane α -helix 1, α -helix 2, α -helix 3, α -helix 4, α -helix 5, α -helix 6, α -helix 7, α -helix 8, α -helix 9, α -helix 10, α -helix 11, α -helix 12, where transmembrane α -helix 10 and transmembrane α -helix 11 and the sequence between them comprise SEQ ID NO: 8.

Each of the above described aspects may be combined in any number of combinations. A host cell may contain a polynucleotide encoding a polypeptide containing 1, 2, 3, 4, 5, 6, or 7 of any of SEQ ID NOs: 1-8, or the polypeptide may contain all of SEQ ID NOs: 1-8. For example, a host cell may contain a polynucleotide encoding a polypeptide containing transmembrane α -helix 1, α -helix 2, α -helix 3, α -helix 4, α -helix 5, α -helix 6, α -helix 7, α -helix 8, α -helix 9, α -helix 10, α -helix 11, α -helix 12, where transmembrane α -helix 1 comprises SEQ ID NO: 1, a loop connecting transmembrane α -helix 2 and transmembrane α -helix 3 comprises SEQ ID NO: 3, and transmembrane α -helix 7 comprises SEQ ID NO: 7. Or, in another example, a host cell may contain a polynucleotide encoding a polypeptide containing transmembrane α -helix 1, α -helix 2, α -helix 3, α -helix 4, α -helix 5, α -helix 6, α -helix 7, α -helix 8, α -helix 9, α -helix 10, α -helix 11, α -helix 12, where transmembrane α -helix 2 comprises SEQ ID NO: 2, transmembrane α -helix 3 comprises SEQ ID NO: 3, transmembrane α -helix 6 comprises SEQ ID NO: 5, and transmembrane α -helix 10 and transmembrane α -helix 11 and the sequence between them comprise SEQ ID NO: 8.

In certain embodiments of the above described aspects, the polypeptide has at least 29%, at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 99%, or at least 100% amino acid identity to NCU00801 or NCU08114.

In preferred embodiments, the host cells further contain a polynucleotide, where the polynucleotide encodes a catalytic domain of a β -glucosidase. As used herein, β -glucosidase refers to a β D-glucoside glucohydrolase (E.C. 3.2.1.21), which catalyzes the hydrolysis of terminal non-reducing β D-glucose residues with the release of β D-glucose. A catalytic domain of β -glucosidase has β -glucosidase activity as determined, for example, according to the basic procedure described by Venturi et al., 2002. A catalytic domain of a

β -glucosidase is any domain that catalyzes the hydrolysis of terminal non-reducing residues in β -D-glucosides with release of glucose. In preferred embodiments, the β -glucosidase is a glycosyl hydrolase family 1 member. Members of this group can be identified by the motif, [LIVMFSTC]-[LIVFYS]-[LIV]-[LIVMST]-E-N-G-[LIVMFAR]-[CSAGN] (SEQ ID NO: 14). Here, E is the catalytic glutamate (webpage expasy.org/cgi-bin/prosite-search-ac?PD00000495). In certain embodiments, the polynucleotide encoding a catalytic domain of β -glucosidase is heterologous to the host cell. In preferred embodiments, the catalytic domain of β -glucosidase is located intracellularly in the host cell. In preferred embodiments, the β -glucosidase is from *N. crassa*, and in particularly preferred embodiments, the β -glucosidase is NCU00130. In certain embodiments, the β -glucosidase may be an ortholog of NCU00130. Examples of orthologs of NCU00130 include, without limitation, *T. melanosporum*, CAZ82985.1; *A. oryzae*, BAE57671.1; *P. placenta*, EED81359.1; *P. chrysosporium*, BAE87009.1; *Kluyveromyces lactis*, CAG99696.1; *Laccaria bicolor*, EDR09330; *Clavispora lusitaniae*, EEQ37997.1; and *Pichia stipitis*, ABN67130.1. Other β -glucosidases could be used include those from the glycosyl hydrolase family 3. These β -glucosidases can be identified by the following motif according to PROSITE: [LIVM](2)-[KR]-x-[EQKRD]-x(4)-G-[LIVMFTC]-[LVT]-[LIVMF]-[ST]-D-x(2)-[SGAD-NIT] (SEQ ID NO: 15). Here D is the catalytic aspartate. Typically, any β -glucosidase may be used that contains the conserved domain of β -glucosidase/6-phospho- β -glucosidase/ β -galactosidase found in NCBI sequence COG2723. Catalytic domains from specific β -glucosidases may be preferred depending on the cellobextrin transporter contained in the host cell.

In certain embodiments, the host cell contains one or more polynucleotides, where the one or more polynucleotides encode one or more enzymes involved in pentose utilization. The one or more polynucleotides may be endogenous or heterologous to the host cell. Pentose, as used herein, refers to any monosaccharide with five carbon atoms. Examples of pentoses include, without limitation, xylose, arabinose, mannose, galactose, and rhamnose. The one or more enzymes involved in pentose utilization may include, for example, L-arabinose isomerase, L-ribulokinase, L-ribulose-5-P 4 epimerase, xylose isomerase, xylulokinase, aldose reductase, L-arabitol 4-dehydrogenase, L-xylulose reductase, and xylitol dehydrogenase in any combination. These enzymes may come from any organism that naturally metabolizes pentose sugars. Examples of such organisms include, for example, *Kluyveromyces* sp., *Zymomonas* sp., *E. coli*, *Clostridium* sp., and *Pichia* sp.

Examples 12-15 describe ways in which the pentose utilization pathway in the host cell may be improved or made to be more efficient. Strain background of a host cell can affect the efficiency of its pentose utilization pathway. In embodiments of the invention where the host cell is a *Saccharomyces* sp., preferred pentose utilizing strains include DA24-16 (see Example 13) and L2612 (see Example 16). Other host cells containing polynucleotides encoding enzymes involved in pentose utilization include a DuPont *Zymomonas* strain (WO 2009/058927) and a *Saccharomyces* strain (U.S. Pat. No. 5,789,210).

In certain embodiments of the invention, the host cell contains a recombinant polynucleotide encoding a pentose transporter. In certain embodiments, pentose transporters include those transporters discovered and described herein, including NCU00821, NCU04963, NCU06138, STL12/XUT6, SUT2, SUT3, XUT1, and XUT3 (see Example 11). In other embodi-

ments, pentose transporters may include Gxs1 from *C. intermedia*, Aut1 from *P. stipitis*, Xylhp from *D. hansenii* (Nobre et al., 1999), xylose transporter from *K. marxianus* (Stambuk et al., 2003), LAT1 and LAT2 from *Ambrosiozyma monospora* (EMBL AY923868 and AY923869, respectively, R. Verho et al.), ART1 from *C. arabinofermentans* (Fonseca et al., 2007), KmLAT1 from *K. marxianus* (Knoshaug et al., 2007), PgLAT2 from *P. guilliermondii* (Knoshaug et al., 2007), and araT from *P. stipitis* (Boles & Keller, 2008).

Methods of Producing and Culturing Host Cells of the Invention

The invention herein relates to host cells containing recombinant polynucleotides encoding polypeptides where the polypeptide transports cellobextrin or a pentose into the cell. Further described herein are methods of increasing transport of cellobextrin into a host cell, methods of increasing growth of a host cell on a medium containing cellobextrin, methods of co-fermenting cellulose-derived and hemicellulose-derived sugars, and methods of making hydrocarbons or hydrocarbon derivatives by providing a host cell containing a recombinant polynucleotide encoding a polypeptide where the polypeptide transports cellobextrin into the cell. Further described herein are methods of increasing transport of a pentose into a host cell, methods of increasing growth of a host cell on a medium containing pentose sugars, and methods of making hydrocarbons or hydrocarbon derivatives by providing a host cell containing a recombinant polynucleotide encoding a polypeptide where the polypeptide transports a pentose into the cell.

Methods of producing and culturing host cells of the invention may include the introduction or transfer of expression vectors containing the recombinant polynucleotides of the invention into the host cell. Such methods for transferring expression vectors into host cells are well known to those of ordinary skill in the art. For example, one method for transforming *E. coli* with an expression vector involves a calcium chloride treatment wherein the expression vector is introduced via a calcium precipitate. Other salts, e.g., calcium phosphate, may also be used following a similar procedure. In addition, electroporation (i.e., the application of current to increase the permeability of cells to nucleic acid sequences) may be used to transfet the host cell. Also, microinjection of the nucleic acid sequences provides the ability to transfet host cells. Other means, such as lipid complexes, liposomes, and dendrimers, may also be employed. Those of ordinary skill in the art can transfet a host cell with a desired sequence using these or other methods.

The vector may be an autonomously replicating vector, i.e., a vector which exists as an extrachromosomal entity, the replication of which is independent of chromosomal replication, e.g., a plasmid, an extrachromosomal element, a minichromosome, or an artificial chromosome. The vector may contain any means for assuring self-replication. Alternatively, the vector may be one which, when introduced into the host, is integrated into the genome and replicated together with the chromosome(s) into which it has been integrated. Furthermore, a single vector or plasmid or two or more vectors or plasmids which together contain the total DNA to be introduced into the genome of the host, or a transposon may be used.

The vectors preferably contain one or more selectable markers which permit easy selection of transformed hosts. A selectable marker is a gene the product of which provides, for example, biocide or viral resistance, resistance to heavy metals, prototrophy to auxotrophs, and the like. Selection of

bacterial cells may be based upon antimicrobial resistance that has been conferred by genes such as the amp, gpt, neo, and hyg genes.

Suitable markers for yeast hosts are, for example, ADE2, HIS3, LEU2, LYS2, MET3, TRP1, and URA3. Selectable markers for use in a filamentous fungal host include, but are not limited to, amdS (acetamidase), argB (ornithine carbamoyltransferase), bar (phosphinothricin acetyltransferase), hph (hygromycin phosphotransferase), niaD (nitrate reductase), 10 pyrG (orotidine-5'-phosphate decarboxylase), sC (sulfate adenyltransferase), and trpC (anthranilate synthase), as well as equivalents thereof. Preferred for use in *Aspergillus* are the amdS and pyrG genes of *Aspergillus nidulans* or *Aspergillus oryzae* and the bar gene of *Streptomyces hygroscopicus*. Preferred for use in *Trichoderma* are bar and amdS.

The vectors preferably contain an element(s) that permits integration of the vector into the host's genome or autonomous replication of the vector in the cell independent of the genome.

For integration into the host genome, the vector may rely on the gene's sequence or any other element of the vector for integration of the vector into the genome by homologous or nonhomologous recombination. Alternatively, the vector may contain additional nucleotide sequences for directing 20 integration by homologous recombination into the genome of the host. The additional nucleotide sequences enable the vector to be integrated into the host genome at a precise location(s) in the chromosome(s). To increase the likelihood of integration at a precise location, the integrational elements 25 should preferably contain a sufficient number of nucleic acids, such as 100 to 10,000 base pairs, preferably 400 to 10,000 base pairs, and most preferably 800 to 10,000 base pairs, which are highly homologous with the corresponding target sequence to enhance the probability of homologous 30 recombination. The integrational elements may be any sequence that is homologous with the target sequence in the genome of the host. Furthermore, the integrational elements 35 may be non-encoding or encoding nucleotide sequences. On the other hand, the vector may be integrated into the genome 40 of the host by non-homologous recombination.

For autonomous replication, the vector may further comprise an origin of replication enabling the vector to replicate autonomously in the host in question. The origin of replication 45 may be any plasmid replicator mediating autonomous replication which functions in a cell. The term "origin of replication" or "plasmid replicator" is defined herein as a sequence that enables a plasmid or vector to replicate in vivo. Examples of origins of replication for use in a yeast host are the 2 micron origin of replication, ARS1, ARS4, the combination of ARS1 and CEN3, and the combination of ARS4 and CEN6. Examples of origins of replication useful in a filamentous 50 fungal cell are AMA1 and ANS1 (Gems et al., 1991; Cullen et al., 1987; WO 00/24883). Isolation of the AMA1 gene and construction of plasmids or vectors comprising the 55 gene can be accomplished according to the methods disclosed in WO 00/24883.

For other hosts, transformation procedures may be found, for example, in Jeremiah D. Read, et al., *Applied and Environmental Microbiology*, August 2007, p. 5088-5096, for 60 *Kluyveromyces*, in Osvaldo Delgado, et al., *FEMS Microbiology Letters* 132, 1995, 23-26, for *Zymomonas*, in U.S. Pat. No. 7,501,275 for *Pichia stipitis*, and in WO 2008/040387 for *Clostridium*.

More than one copy of a gene may be inserted into the host 65 to increase production of the gene product. An increase in the copy number of the gene can be obtained by integrating at least one additional copy of the gene into the host genome or

by including an amplifiable selectable marker gene with the nucleotide sequence where cells containing amplified copies of the selectable marker gene, and thereby additional copies of the gene, can be selected for by cultivating the cells in the presence of the appropriate selectable agent.

The procedures used to ligate the elements described above to construct the recombinant expression vectors of the present invention are well known to one skilled in the art (see, e.g., Sambrook et al., 1989, supra).

The host cell is transformed with at least one expression vector. When only a single expression vector is used (without the addition of an intermediate), the vector will contain all of the nucleic acid sequences necessary.

Once the host cell has been transformed with the expression vector, the host cell is allowed to grow. Methods of the invention may include culturing the host cell such that recombinant nucleic acids in the cell are expressed. For microbial hosts, this process entails culturing the cells in a suitable medium. Typically cells are grown at 35° C. in appropriate media. Preferred growth media in the present invention include, for example, common commercially prepared media such as Luria Bertani (LB) broth, Sabouraud Dextrose (SD) broth or Yeast medium (YM) broth. Other defined or synthetic growth media may also be used and the appropriate medium for growth of the particular host cell will be known by someone skilled in the art of microbiology or fermentation science. Temperature ranges and other conditions suitable for growth are known in the art (see, e.g., Bailey and Ollis 1986).

According to some aspects of the invention, the culture media contains a carbon source for the host cell. Such a "carbon source" generally refers to a substrate or compound suitable to be used as a source of carbon for prokaryotic or simple eukaryotic cell growth. Carbon sources can be in various forms, including, but not limited to polymers, carbohydrates, acids, alcohols, aldehydes, ketones, amino acids, peptides, etc. These include, for example, various monosaccharides such as glucose, oligosaccharides, polysaccharides, a biomass polymer such as cellulose or hemicellulose, xylose, arabinose, disaccharides, such as sucrose, saturated or unsaturated fatty acids, succinate, lactate, acetate, ethanol, etc., or mixtures thereof. The carbon source can additionally be a product of photosynthesis, including, but not limited to glucose.

In preferred embodiments, the carbon source is a biomass polymer such as cellulose or hemicellulose. "A biomass polymer" as described herein is any polymer contained in biological material. The biological material may be living or dead. A biomass polymer includes, for example, cellulose, xylan, xylose, hemicellulose, lignin, mannan, and other materials commonly found in biomass. Non-limiting examples of sources of a biomass polymer include grasses (e.g., switchgrass, *Miscanthus*), rice hulls, bagasse, cotton, jute, hemp, flax, bamboo, sisal, abaca, straw, leaves, grass clippings, corn stover, corn cobs, distillers grains, legume plants, sorghum, sugar cane, sugar beet pulp, wood chips, sawdust, and biomass crops (e.g., *Crambe*).

In addition to an appropriate carbon source, media must contain suitable minerals, salts, cofactors, buffers and other components, known to those skilled in the art, suitable for the growth of the cultures and promotion of the enzymatic pathways necessary for the fermentation of various sugars and the production of hydrocarbons and hydrocarbon derivatives. Reactions may be performed under aerobic or anaerobic conditions where aerobic, anoxic, or anaerobic conditions are preferred based on the requirements of the microorganism. As the host cell grows and/or multiplies, expression of the enzymes, transporters, or other proteins necessary for growth

on various sugars or biomass polymers, sugar fermentation, or synthesis of hydrocarbons or hydrocarbon derivatives is affected.

Methods of Increasing Transport of a Sugar into a Cell

The present invention provides methods of increasing transport of a sugar into a cell. In one aspect, the invention provides a method of transporting cellobiose into a cell, including a first step of providing a host cell, where the host cell contains a recombinant polynucleotide encoding a polypeptide containing transmembrane α -helix 1, α -helix 2, α -helix 3, α -helix 4, α -helix 5, α -helix 6, α -helix 7, α -helix 8, α -helix 9, α -helix 10, α -helix 11, and α -helix 12, where one or more of the following is true: transmembrane α -helix 1 comprises SEQ ID NO: 1, transmembrane α -helix 2 comprises SEQ ID NO: 2, the loop connecting transmembrane α -helix 2 and transmembrane α -helix 3 comprises SEQ ID NO: 3, transmembrane α -helix 5 comprises SEQ ID NO: 4, transmembrane α -helix 6 comprises SEQ ID NO: 5, sequence between transmembrane α -helix 6 and transmembrane α -helix 7 comprises SEQ ID NO: 6, transmembrane α -helix 7 comprises SEQ ID NO: 7, and transmembrane α -helix 10 and transmembrane α -helix 11 and the sequence between them comprise SEQ ID NO: 8. The method includes a second step of culturing the cell such that the recombinant polynucleotide is expressed, where expression of the recombinant polynucleotide results in increased transport of cellobiose into the cell compared with a cell that does not contain the recombinant polynucleotide. Transport of cellobiose into a cell may be measured by any method known to one of skill in the art, including those methods described in Example 9 such as measuring uptake of [³H]-cellobiose into cells or measuring the ability of an *S. cerevisiae* host cell to grow when cellobiose is the sole carbon source. Typically, the host cell containing the recombinant polynucleotide and the host cell that does not contain the recombinant polynucleotide will otherwise be identical in genetic background.

In certain embodiments, the polypeptide has at least 29%, at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 99%, or at least 100% amino acid identity to NCU00801 or NCU08114. In certain embodiments, the host cell also contains a recombinant polynucleotide encoding a catalytic domain of a β -glucosidase. Such embodiments are useful for host cells lacking the endogenous ability to utilize cellobextrins. Preferably, the catalytic domain of the β -glucosidase is intracellular. In preferred embodiments, the β -glucosidase is from *Neurospora crassa*. In particularly preferred embodiments, the β -glucosidase is encoded by NCU00130.

In methods of increasing transport of cellobiose into a cell, the cell may be cultured in a medium containing a cellulase-containing enzyme mixture from an altered organism, where the mixture has reduced β -glucosidase activity compared to a cellulase-containing mixture from an unaltered organism. The organism may be altered to reduce the expression of β -glucosidase, such as by mutation of a gene encoding β -glucosidase or by targeted RNA interference or the like.

In another aspect, the invention provides a method of increasing transport of xylose into a cell, including the steps of providing a host cell, where the host cell contains a recombinant polynucleotide encoding a NCU00821 or STL12/XUT6 polypeptide, and culturing the cell such that the recombinant polynucleotide is expressed, where expression of the recombinant polynucleotide results in increased transport of xylose into the cell compared with a cell that does not contain the recombinant polynucleotide. In another aspect, the invention provides a method of increasing transport of arabinose

into a cell, including the steps of providing a host cell, where the host cell contains a recombinant polynucleotide encoding a XUT1 polypeptide, and culturing the cell such that the recombinant polynucleotide is expressed, where expression of the recombinant polynucleotide results in increased transport of arabinose into the cell compared with a cell that does not contain the recombinant polynucleotide. In yet another aspect, the invention provides a method of increasing transport of arabinose or glucose into a cell, including the steps of providing a host cell, where the host cell contains a recombinant polynucleotide encoding a NCU06138 polypeptide, and culturing the cell such that the recombinant polynucleotide is expressed, where expression of the recombinant polynucleotide results in increased transport of arabinose or glucose into the cell compared with a cell that does not contain the recombinant polynucleotide. In yet another aspect the invention provides a method of increasing transport of xylose or glucose into a cell, including the steps of providing a host cell, where the host cell comprises a recombinant polynucleotide encoding a SUT2, SUT3, or XUT3 polypeptide, and culturing the cell such that the recombinant polynucleotide is expressed, where expression of the recombinant polynucleotide results in increased transport of xylose or glucose into the cell compared with a cell that does not contain the recombinant polynucleotide. In another aspect, the invention provides a method of increasing transport of xylose, arabinose, or glucose into a cell, including the steps of providing a host cell, where the host cell contains a recombinant polynucleotide encoding a NCU04963 polypeptide, and culturing the cell such that the recombinant polynucleotide is expressed, where expression of the recombinant polynucleotide results in increased transport of xylose, arabinose, or glucose into the cell compared with a cell that does not contain the recombinant polynucleotide.

Transport of xylose, arabinose, or glucose into a cell may be measured by any method known to one of skill in the art, including those methods described in Example 11. These methods include, for example, measuring D-xylose or L-arabinose transport by extracting accumulated D-xylose and xylitol or L-arabinose and arabinitol from the host cell using osmosis and analyzing it using high performance liquid chromatography and measuring glucose transport by using host cells lacking the ability to grow on glucose as a sole carbon source. Typically, the host cell containing the recombinant polynucleotide and the host cell that does not contain the recombinant polynucleotide will otherwise be identical in genetic background.

In certain embodiments, the host cell also contains one or more recombinant polynucleotides where the one or more polynucleotides encode one or more enzymes involved in pentose utilization. The one or more enzymes may be, for example, L-arabinose isomerase, L-ribulokinase, L-ribulose-5-P 4 epimerase, xylose isomerase, xylulokinase, aldose reductase, L-arabitol 4-dehydrogenase, L-xylulose reductase, xylitol dehydrogenase, or any other pentose utilization enzymes known in the art.

Methods of Increasing Growth of a Cell

The present invention further provides methods of increasing the growth of a cell. In one aspect the invention provides methods of increasing growth of a cell, including a first step of providing a host cell, where the host cell contains a recombinant polynucleotide encoding a polypeptide containing transmembrane α -helix 1, α -helix 2, α -helix 3, α -helix 4, α -helix 5, α -helix 6, α -helix 7, α -helix 8, α -helix 9, α -helix 10, α -helix 11, and α -helix 12, where one or more of the following is true: transmembrane α -helix 1 comprises SEQ ID NO: 1, transmembrane α -helix 2 comprises SEQ ID NO: 2, the

loop connecting transmembrane α -helix 2 and transmembrane α -helix 3 comprises SEQ ID NO: 3, transmembrane α -helix 5 comprises SEQ ID NO: 4, transmembrane α -helix 6 comprises SEQ ID NO: 5, sequence between transmembrane α -helix 6 and transmembrane α -helix 7 comprises SEQ ID NO: 6, transmembrane α -helix 7 comprises SEQ ID NO: 7, and transmembrane α -helix 10 and transmembrane α -helix 11 and the sequence between them comprise SEQ ID NO: 8, and the polypeptide transports cellobextrin. The method includes a second step of culturing the host cell in a medium containing cellobextrin, where the host cell grows at a faster rate in the medium than a cell that does not contain the recombinant polynucleotide. The growth rate of a host cell may be measured by any method known to one of skill in the art. Typically, growth rate of a cell will be measured by evaluating cell concentration in suspension by optical density. Preferably, the host cell containing the recombinant polynucleotide and the host cell that does not contain the recombinant polynucleotide will otherwise be identical in genetic background. Media containing cellobextrins may have resulted from enzymatic treatment of biomass polymers such as cellulose.

In certain embodiments, the polypeptide has at least 29%, at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 99%, or at least 100% amino acid identity to NCU00801 or NCU08114. In certain embodiments, the host cell also contains a recombinant polynucleotide encoding a catalytic domain of a β -glucosidase. Such embodiments are useful for host cells lacking the endogenous ability to utilize cellobextrins. Preferably, the catalytic domain of the β -glucosidase is intracellular. In preferred embodiments, the β -glucosidase is from *Neurospora crassa*. In particularly preferred embodiments, the β -glucosidase is encoded by NCU00130.

In methods of increasing growth of a cell, the culturing medium may contain a cellulase-containing enzyme mixture from an altered organism, where the mixture has reduced β -glucosidase activity compared to a cellulase-containing mixture from an unaltered organism. The organism may be altered to reduce the expression of β -glucosidase, such as by mutation of a gene encoding β -glucosidase or by targeted RNA interference or the like.

In another aspect, the invention provides a method of increasing growth of a cell, including the steps of providing a host cell, where the host cell contains a recombinant polynucleotide where the polynucleotide encodes a NCU00821 or STL12/XUT6 polypeptide, and the polypeptide transports xylose, and culturing the host cell in a medium containing xylose, where the host cell grows at a faster rate in the medium than a cell that does not contain the recombinant polynucleotide. In another aspect the invention provides a method of increasing growth of a cell, including the steps of providing a host cell, where the host cell contains a recombinant polynucleotide where the polynucleotide encodes a XUT1 polypeptide, and the polypeptide transports arabinose, and culturing the host cell in a medium containing arabinose, where the host cell grows at a faster rate in the medium than a cell that does not contain the recombinant polynucleotide. In yet another aspect, the invention provides a method of increasing growth of a cell, including the steps of providing a host cell, where the host cell contains a recombinant polynucleotide where the polynucleotide encodes a NCU06138 polypeptide, and the polypeptide transports arabinose and glucose, and culturing the host cell in a medium containing arabinose or glucose, where the host cell grows at a faster rate in the medium than a cell that does not contain the recombi-

nant polynucleotide. In another aspect, the invention provides a method of increasing growth of a cell, including the steps of providing a host cell, where the host cell contains a recombinant polynucleotide where the polynucleotide encodes a SUT2, SUT3, or XUT3 polypeptide, and the polypeptide transports xylose and glucose, and culturing the host cell in a medium containing xylose or glucose, where the host cell grows at a faster rate in the medium than a cell that does not contain the recombinant polynucleotide. In yet another aspect, the invention provides a method of increasing growth of a cell, including the steps of providing a host cell, where the host cell contains a recombinant polynucleotide where the polynucleotide encodes a NCU04963 polypeptide, and the polypeptide transports xylose, arabinose, and glucose, and culturing the host cell in a medium containing xylose, arabinose, or glucose, where the host cell grows at a faster rate in the medium than a cell that does not contain the recombinant polynucleotide.

The growth rate of a host cell may be measured by any method known to one of skill in the art. Typically, growth rate of a cell will be measured by evaluating cell concentration in suspension by optical density. Preferably, the host cell containing the recombinant polynucleotide and the host cell that does not contain the recombinant polynucleotide will otherwise be identical in genetic background. Media containing xylose or arabinose may have resulted from acid treatment of biomass polymers such as hemicellulose. Media containing glucose may have resulted from enzymatic treatment of biomass polymers such as cellulose.

In certain embodiments, the host cell also contains one or more recombinant polynucleotides where the one or more polynucleotides encode one or more enzymes involved in pentose utilization. The one or more enzymes may be, for example, L-arabinose isomerase, L-ribulokinase, L-ribulose-5-P 4 epimerase, xylose isomerase, xylulokinase, aldose reductase, L-arabitol 4-dehydrogenase, L-xylulose reductase, xylitol dehydrogenase, or any other pentose utilization enzymes known in the art.

In one aspect, the invention provides methods of increasing growth of a cell on a biomass polymer. In preferred embodiments, the biomass polymer is cellulose. In other preferred embodiments, the biomass polymer is hemicellulose. According to one aspect of the invention, the method includes providing a host cell comprising a recombinant polynucleotide that encodes a NCU07705 polypeptide. According to another aspect of the invention, the method includes culturing the cell in a medium comprising the biomass polymer wherein the host cell grows at a faster rate in the medium than a cell that does not comprise the recombinant polynucleotide.

In another aspect of the invention, the invention provides a method of increasing growth of a cell, including the steps of providing a host cell, where the host cell contains a recombinant polynucleotide where the polynucleotide encodes a NCU01517, NCU09133, or NCU10040 polypeptide, and culturing the cell in a medium containing hemicellulose, where the host cell grows at a faster rate in the medium than a cell that does not contain the recombinant polynucleotide.

According to another aspect of the invention, the method includes providing a host cell comprising an endogenous polynucleotide that encodes a NCU05137 polypeptide. According to another aspect of the invention, the method includes inhibiting expression of the endogenous polynucleotide and culturing the cell in a medium comprising a biomass polymer wherein the host cell grows at a faster rate in the medium than a cell in which expression of the endogenous polynucleotide is not inhibited.

Methods of the invention may include culturing the host cell such that recombinant nucleic acids in the cell are expressed. For microbial hosts, this process entails culturing the cells in a suitable medium. Typically cells are grown at 35° C. in appropriate media. Preferred growth media in the present invention include, for example, common commercially prepared media such as Luria Bertani (LB) broth, Sabouraud Dextrose (SD) broth, or Yeast medium (YM) broth. Other defined or synthetic growth media may also be used and the appropriate medium for growth of the particular host cell will be known by someone skilled in the art of microbiology or fermentation science. Temperature ranges and other conditions suitable for growth are known in the art (see, e.g., Bailey and Ollis 1986).

The source of the biomass polymer in the medium may include, for example, grasses (e.g., switchgrass, *Miscanthus*), rice hulls, bagasse, cotton, jute, hemp, flax, bamboo, sisal, abaca, straw, leaves, grass clippings, corn stover, corn cobs, distillers grains, legume plants, sorghum, sugar cane, sugar beet pulp, wood chips, sawdust, and biomass crops (e.g., *Crambe*). In addition to a biomass polymer, the medium must contain suitable minerals, salts, cofactors, buffers and other components, known to those skilled in the art, suitable for the growth of the cultures. The rate of growth of the host cell may be measured by any methods known to one of skill in the art.

In certain embodiments of the invention, the expression of cellulases is increased in the host cell upon expression of a recombinant polynucleotide. "Cellulase" as used herein refers to a category of enzymes capable of hydrolyzing cel lulose polymers to shorter cello-oligosaccharide oligomers, cellobiose, and/or glucose. Cellulases include, without limitation, exoglucanases, exocellulobiohydrolases, endoglucanases, and glucosidases. Expression of cellulases may be measured by RT-PCR or other methods known in the art.

In certain embodiments of the invention, the expression of hemicellulases is increased in the host cell upon expression of a recombinant polynucleotide. "Hemicellulase" as used herein refers to a category of enzymes capable of hydrolyzing hemicellulose polymers. Hemicellulases include, without limitation, xylanases, mannanases, arabinases (both endo and exo kinds) and their corresponding glycosidases. Expression of hemicellulases may be measured by RT-PCR or other methods known in the art.

Inhibition of expression of the endogenous polynucleotide may be achieved, for example, by genetic modifications which result in a decrease in gene expression, in the function of the gene, or in the function of the gene product (i.e., the protein encoded by the gene) and can be referred to as inactivation (complete or partial), deletion, interruption, blockage, silencing, or down-regulation, or attenuation of expression of a gene. For example, a genetic modification in a gene which results in a decrease in the function of the protein encoded by such a gene can be the result of a complete deletion of the gene (i.e., the gene does not exist, and therefore the protein does not exist), a mutation in the gene which results in incomplete or no translation of the protein (e.g., the protein is not expressed), or a mutation in the gene which decreases or abolishes the natural function of the protein (e.g., a protein is expressed which has decreased or no enzymatic activity or action). More specifically, reference to decreasing the action of proteins discussed herein generally refers to any genetic modification in the host cell in question which results in decreased expression and/or functionality (biological activity) of the proteins and includes decreased activity of the proteins (e.g., decreased transport), increased inhibition or degradation of the proteins, as well as a reduction or elimination of expression of the proteins. For example, the action

or activity of a protein of the present invention can be decreased by blocking or reducing the production of the protein, reducing protein action, or inhibiting the action of the protein. Combinations of some of these modifications are also possible. Blocking or reducing the production of a protein can include placing the gene encoding the protein under the control of a promoter that requires the presence of an inducing compound in the growth medium. By establishing conditions such that the inducer becomes depleted from the medium, the expression of the gene encoding the protein (and therefore, of protein synthesis) could be turned off. Blocking or reducing the action of a protein could also include using an excision technology approach similar to that described in U.S. Pat. No. 4,743,546. To use this approach, the gene encoding the protein of interest is cloned between specific genetic sequences that allow specific, controlled excision of the gene from the genome. Excision could be prompted by, for example, a shift in the cultivation temperature of the culture, as in U.S. Pat. No. 4,743,546, or by some other physical or nutritional signal.

In certain embodiments of the invention, cellulase activity of the host cell is increased upon inhibiting expression of an endogenous polynucleotide. Cellulase activity may be measured as described in Example 5 and by any other methods known in the art.

In certain embodiments of the invention, hemicellulase activity of the host cell is increased upon inhibiting expression of an endogenous polynucleotide. Hemicellulase activity may be measured as described in Example 17 and by any other methods known in the art.

Methods of Co-Fermentation

One aspect of the present invention provides methods of co-fermenting cellulose-derived and hemicellulose-derived sugars. As used herein, co-fermentation refers to simultaneous utilization by a host cell of more than one sugar in the same vessel. The method includes the steps of providing a host cell, where the host cell contains a first recombinant polynucleotide encoding a cellobextrin transporter and a second recombinant polynucleotide encoding a catalytic domain of a β -glucosidase, and culturing the host cell in a medium containing a cellulose-derived sugar and a hemicellulose-derived sugar, where expression of the recombinant polynucleotides enables co-fermentation of the cellulose-derived sugar and the hemicellulose-derived sugar.

The first recombinant polynucleotide may encode any polypeptide that is capable of transporting cellobextrin into a cell. Cellobextrin transport may be measured by any method known to one of skill in the art, including the methods discussed in Example 9. In preferred embodiments, the first recombinant polynucleotide encodes a polypeptide containing transmembrane α -helix 1, α -helix 2, α -helix 3, α -helix 4, α -helix 5, α -helix 6, α -helix 7, α -helix 8, α -helix 9, α -helix 10, α -helix 11, and α -helix 12, where one or more of the following is true: transmembrane α -helix 1 comprises SEQ ID NO: 1, transmembrane α -helix 2 comprises SEQ ID NO: 2, the loop connecting transmembrane α -helix 2 and transmembrane α -helix 3 comprises SEQ ID NO: 3, transmembrane α -helix 5 comprises SEQ ID NO: 4, transmembrane α -helix 6 comprises SEQ ID NO: 5, sequence between transmembrane α -helix 6 and transmembrane α -helix 7 comprises SEQ ID NO: 6, transmembrane α -helix 7 comprises SEQ ID NO: 7, and transmembrane α -helix 10 and transmembrane α -helix 11 and the sequence between them comprise SEQ ID NO: 8. Examples of such polypeptides include NCU00801, NCU00809, NCU08114, XP_001268541.1, and LAC2. In preferred embodiments, the first recombinant polypeptide encodes NCU00801.

The second recombinant polynucleotide may encode any catalytic domain capable of catalyzing the hydrolysis of terminal non-reducing residues in β -D-glucosides with release of glucose. Preferably, the β -glucosidase catalytic domain is located intracellularly in the host cell. In certain embodiments the source of the β -glucosidase domain is a *N. crassa* β -glucosidase. In preferred embodiments the source of the β -glucosidase domain is NCU00130. Catalytic domains from different sources may work best with different cellobextrin transporters.

In certain embodiments, the host cell also contains one or more recombinant polynucleotides where the one or more polynucleotides encode one or more enzymes involved in pentose utilization. Alternatively, one or more polynucleotides encoding one or more enzymes involved in pentose utilization may be endogenous to the host cell. The one or more enzymes may include, for example, L-arabinose isomerase, L-ribulokinase, L-ribulose-5-P 4 epimerase, xylose isomerase, xylulokinase, aldose reductase, L-arabinitol 4-dehydrogenase, L-xylulose reductase, xylitol dehydrogenase, or any other pentose-utilizing enzymes known to one of skill in the art.

In certain embodiments, the host cell contains a third recombinant polynucleotide where the polynucleotide encodes a pentose transporter. Alternatively, the host cell may contain an endogenous polynucleotide encoding a pentose transporter. In preferred embodiments, the pentose transporter transports xylose and/or arabinose into the cell. In certain embodiments, the third recombinant polynucleotide encodes a polypeptide such as NCU00821, NCU04963, NCU06138, STL12/XUT6, SUT2, SUT3, XUT1, or XUT3. The expression of a pentose transporter in the host cell may enhance the efficiency of co-fermentation if glucose is present along with a pentose sugar in the growth medium.

In methods of co-fermentation as described herein, cellulose-derived sugars preferably include cellobiose, cellotriose, and cellotetraose, and hemicellulose-derived sugars preferably include xylose and arabinose. Typically, in order to prepare the cellulose-derived sugars and hemicellulose-derived sugars for co-fermentation by a host cell, lignocellulosic biomass is first pretreated to alter its structure and allow for better enzymatic hydrolysis of cellulose. Pretreatment may include physical or chemical methods, including, for example, ammonia fiber/freeze explosion, the lime method based on calcium or sodium hydroxide, and steam explosion with or without an acid catalyst. Acid treatment will release xylose and arabinose from the hemicellulose component of the lignocellulosic biomass. Next, preferably, the cellulose component of the pretreated biomass is hydrolyzed by a mixture of cellulases. Examples of commercially available cellulase mixtures include Celluclast 1.5L® (Novozymes), Spezyme CP® (Genencor) (Scott W. Pryor, 2010, *Appl Biochem Biotechnol*), and Cellulyve 50L (Lyven).

Cellulase mixtures typically contain endoglucanases, exoglucanases, and β -glucosidases. In methods of co-fermentation as described herein, the amount of β -glucosidase activity in the cellulase mixture should be minimized as much as possible. For example, the culturing medium may contain a cellulase-containing enzyme mixture from an altered organism, where the mixture has reduced β -glucosidase activity compared to a cellulase-containing mixture from an unaltered organism. The organism may be altered to reduce the expression of β -glucosidase, such as by mutation of a gene encoding β -glucosidase or by targeted RNA interference or the like.

Surprisingly, as described in Example 17, co-fermentation of cellobiose and xylose by the methods of the invention

resulted in a synergistic effect on sugar consumption and ethanol production by the host cell.

Methods of Synthesis of Hydrocarbons or Hydrocarbon Derivatives

One aspect of the present invention provides methods for increasing the synthesis of hydrocarbons or hydrocarbon derivatives by a host cell.

"Hydrocarbons" as used herein are organic compounds consisting entirely of hydrogen and carbon. Hydrocarbons include, without limitation, methane, ethane, ethene, ethyne, propane, propene, propyne, cyclopropane, allene, butane, isobutene, butene, butyne, cyclobutane, methylcyclopropane, butadiene, pentane, isopentane, neopentane, pentene, pentyne, cyclopentane, methylcyclobutane, ethylcyclopropane, pentadiene, isoprene, hexane, hexene, hexyne, cyclohexane, methylcyclopentane, ethylcyclobutane, propylcyclopropane, hexadiene, heptane, heptene, heptyne, cycloheptane, methylcyclohexane, heptadiene, octane, octene, octyne, cyclooctane, octadiene, nonane, nonene, nonyne, cyclononane, nonadiene, decane, decene, decyne, cyclodecane, and decadiene.

"Hydrocarbon derivatives" as used herein are organic compounds of carbon and at least one other element that is not hydrogen. Hydrocarbon derivatives include, without limitation, alcohols (e.g., arabinitol, butanol, ethanol, glycerol, methanol, 1,3-propanediol, sorbitol, and xylitol); organic acids (e.g., acetic acid, adipic acid, ascorbic acid, citric acid, 2,5-diketo-D-gluconic acid, formic acid, fumaric acid, glutaric acid, gluconic acid, glucuronic acid, glutaric acid, 3-hydroxypropionic acid, itaconic acid, lactic acid, malic acid, malonic acid, oxalic acid, propionic acid, succinic acid, and xylonic acid); esters; ketones (e.g., acetone); aldehydes (e.g., furfural); amino acids (e.g., aspartic acid, glutamic acid, glycine, lysine, serine, and threonine); and gases (e.g., carbon dioxide and carbon monoxide).

In preferred embodiments, the hydrocarbon or hydrocarbon derivative can be used as fuel. In particularly preferred embodiments, the hydrocarbon or hydrocarbon derivative is ethanol or butanol.

According to one aspect of the invention, a method of increasing the synthesis of hydrocarbons or hydrocarbon derivatives by a host cell includes a first step of providing a host cell, where the host cell contains a recombinant polynucleotide encoding a polypeptide containing transmembrane α -helix 1, α -helix 2, α -helix 3, α -helix 4, α -helix 5, α -helix 6, α -helix 7, α -helix 8, α -helix 9, α -helix 10, α -helix 11, and α -helix 12, where one or more of the following is true: transmembrane α -helix 1 comprises SEQ ID NO: 1, transmembrane α -helix 2 comprises SEQ ID NO: 2, the loop connecting transmembrane α -helix 2 and transmembrane α -helix 3 comprises SEQ ID NO: 3, transmembrane α -helix 5 comprises SEQ ID NO: 4, transmembrane α -helix 6 comprises SEQ ID NO: 5, sequence between transmembrane α -helix 6 and transmembrane α -helix 7 comprises SEQ ID NO: 6, transmembrane α -helix 7 comprises SEQ ID NO: 7, and transmembrane α -helix 10 and transmembrane α -helix 11 and the sequence between them comprise SEQ ID NO: 8, and where the polypeptide transports cellobextrin into the host cell for the synthesis of hydrocarbons or hydrocarbon derivatives. The method includes a second step of culturing the host cell in a medium containing cellobextrin or a source of cellobextrin to increase the synthesis of hydrocarbons or hydrocarbon derivatives by the host cell, where transport of cellobextrin into the cell is increased upon expression of the recombinant polynucleotide. In certain embodiments, the polypeptide has at least 29%, at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at

least 85%, at least 90%, at least 95%, at least 99%, or at least 100% amino acid identity to NCU00801 or NCU08114. In certain embodiments, the host cell also contains a recombinant polynucleotide encoding a catalytic domain of a β -glucosidase. Such embodiments are useful for host cells lacking the endogenous ability to utilize cellobextrins. Preferably, the catalytic domain of the β -glucosidase is intracellular. In preferred embodiments, the β -glucosidase is from *Neurospora crassa*. In particularly preferred embodiments, the β -glucosidase is encoded by NCU00130. Transport of cellobextrin into the cell may be measured by any methods known to one of skill in the art, including the methods described in Example 9. Typically, the source of the cellobextrin is cellulose.

The culturing medium may contain a cellulase-containing enzyme mixture from an altered organism, where the mixture has reduced β -glucosidase activity compared to a cellulase-containing mixture from an unaltered organism. The organism may be altered to reduce the expression of β -glucosidase, such as by mutation of a gene encoding β -glucosidase or by targeted RNA interference or the like.

According to another aspect of the invention, a method of increasing the synthesis of hydrocarbons or hydrocarbon derivatives by a host cell includes the steps of providing a host cell, where the host cell contains a recombinant polynucleotide encoding a NCU00821 or STL12/XUT6 polypeptide, where the polypeptide transports xylose into the host cell for the synthesis of hydrocarbons or hydrocarbon derivatives, and culturing the host cell in a medium containing xylose or a source of xylose to increase the synthesis of hydrocarbons or hydrocarbon derivatives by the host cell, where transport of xylose into the cell is increased upon expression of the recombinant polynucleotide.

According to another aspect, a method of increasing the synthesis of hydrocarbons or hydrocarbon derivatives by a host cell includes the steps of providing a host cell, where the host cell contains a recombinant polynucleotide encoding a XUT1 polypeptide, where the polypeptide transports arabinose into the host cell for the synthesis of hydrocarbons or hydrocarbon derivatives, and culturing the host cell in a medium containing arabinose or a source of arabinose to increase the synthesis of hydrocarbons or hydrocarbon derivatives by the host cell, where transport of arabinose into the cell is increased upon expression of the recombinant polynucleotide.

According to yet another aspect, a method of increasing the synthesis of hydrocarbons or hydrocarbon derivatives by a host cell includes the steps of providing a host cell, where the host cell contains a recombinant polynucleotide encoding a NCU06138 polypeptide, where the polypeptide transports arabinose or glucose into the host cell for the synthesis of hydrocarbons or hydrocarbon derivatives, and culturing the host cell in a medium comprising arabinose or glucose or a source of arabinose or glucose to increase the synthesis of hydrocarbons or hydrocarbon derivatives by the host cell, where transport of arabinose or glucose into the cell is increased upon expression of the recombinant polynucleotide.

According to yet another aspect, a method of increasing the synthesis of hydrocarbons or hydrocarbon derivatives by a host cell includes the steps of providing a host cell, where the host cell contains a recombinant polynucleotide encoding a SUT2, SUT3, or XUT3 polypeptide, where the polypeptide transports xylose or glucose into the host cell for the synthesis of hydrocarbons or hydrocarbon derivatives, and culturing the host cell in a medium containing xylose or glucose or a source of xylose or glucose to increase the synthesis of hydrocarbons or hydrocarbon derivatives by the host cell, where

57

transport of xylose or glucose into the cell is increased upon expression of the recombinant polynucleotide.

According to another aspect, a method of increasing the synthesis of hydrocarbons or hydrocarbon derivatives by a host cell includes the steps of providing a host cell, where the host cell contains a recombinant polynucleotide encoding a NCU04963 polypeptide, where the polypeptide transports xylose, arabinose, or glucose into the host cell for the synthesis of hydrocarbons or hydrocarbon derivatives, and culturing the host cell in a medium comprising xylose, arabinose, or glucose or a source of xylose, arabinose, or glucose to increase the synthesis of hydrocarbons or hydrocarbon derivatives by the host cell, where transport of xylose, arabinose, or glucose into the cell is increased upon expression of the recombinant polynucleotide.

Transport of xylose, arabinose, or glucose into the cell may be measured by any methods known to one of skill in the art, including the methods described in Example 11. Typically, the source of glucose is cellulose, and the source of xylose and arabinose is hemicellulose.

Methods of Degrading Cellulose

One aspect of the present invention provides methods of degrading cellulose. The methods include a first step of providing a composition comprising cellulose. The cellulose is preferably from plant material, such as switchgrass, *Miscanthus*, rice hulls, bagasse, flax, bamboo, sisal, abaca, straw, leaves, grass clippings, corn stover, corn cobs, distillers grains, legume plants, sorghum, sugar cane, sugar beet pulp, wood chips, sawdust, and biomass crops.

The methods include a second step of contacting the composition with a cellulase-containing enzyme mixture from an altered organism, where the cellulase-containing mixture has reduced β -glucosidase activity compared to a cellulase-containing mixture from an unaltered organism. The cellulose is degraded by the cellulase-containing mixture. The organism may be altered by mutation of a gene encoding a β -glucosidase or by reducing the expression of a β -glucosidase with a technique such as RNA interference. The organism may be a fungus or a bacterium. In preferred embodiments, the organism is a filamentous fungus such as *T. reesei*.

Alternatively, the methods include a second step of contacting the composition with a cellulase-containing enzyme mixture that has been altered to reduce its β -glucosidase activity. For example, the cellulase-containing enzyme mixture may be altered by affinity chromatography where β -glucosidase enzymes are captured during the chromatography, and thus removed from the mixture. In another example, the cellulase-containing enzyme mixture is altered by inactivation of β -glucosidase enzymes in the mixtures with an inhibitor. Examples of commercially available cellulase mixtures include Celluclast 1.5L® (Novozymes), Spezyme CP® (Genencor) (Scott W. Pryor, 2010, *Appl Biochem Biotechnol*), and Cellulite 50L (Lyven).

It is to be understood that, while the invention has been described in conjunction with the preferred specific embodiments thereof, the foregoing description is intended to illustrate and not limit the scope of the invention. Other aspects, advantages, and modifications within the scope of the invention will be apparent to those skilled in the art to which the invention pertains.

The invention having been described, the following examples are offered to illustrate the subject invention by way of illustration, not by way of limitation.

58

EXAMPLES

Example 1

Transcriptome Analysis of *N. crassa* grown on *Miscanthus* and Avicel

In this example, the expression profile of the *N. crassa* genome was examined during growth on *Miscanthus* or Avicel. Growth and cellulase activity of *N. crassa* (FGSC 2489) cultured on Vogel's minimal media with crystalline cellulose (Avicel) as the sole carbon source was similar to that of *T. reesei* (QM9414) (FIG. 3); *N. crassa* completely degraded Avicel in approximately 4 days. *N. crassa* also grew rapidly on ground *Miscanthus* stems, suggesting functional cellulase and hemicellulase degradative capacity. To determine the transcriptome associated with plant cell wall deconstruction in *N. crassa*, we used full genome microarrays (Kasuga and Glass 2008; Tian et al., 2007; Kasuga et al., 2005) to monitor gene expression profiles during growth of *N. crassa* on ground *Miscanthus* stems. RNA was sampled after 16 hrs of growth on sucrose and compared to RNA isolated from *N. crassa* grown on *Miscanthus* medium at 16 and 40 hrs, 5 days and 10 days (FIG. 4; also see Supplemental Data, Dataset S1, page 1 in Tian et al., 2009; data can also be found at bioinfo.townsend.yale.edu/browse.jsp, Experiment IDs 52 and 53).

A total of 769 *N. crassa* genes showed a statistically significant difference in relative expression level among the four *Miscanthus* samples as compared to the sucrose sample (see Supplemental Data, Dataset S1, page 3 in Tian et al., 2009). Hierarchical clustering showed that these genes fell into three main clusters (FIG. 4A). The first cluster of genes (C1; 300 genes) showed the highest expression levels in minimal medium with sucrose as a carbon source. Functional category (FunCat) analysis (Ruepp 2004) of these genes showed an enrichment for ribosomal proteins and other functional categories associated with primary metabolism, such as respiration, electron transport and amino acid metabolism (see Supplemental Data, Dataset S1, page 4 in Tian et al., 2009). The second cluster (C2) included 327 genes that showed the highest expression levels in *Miscanthus* cultures at later time points (40 hrs to 10 days; FIG. 4A). Within this group were 89 genes that showed a high relative expression level in *Miscanthus* cultures at all time points. For further analyses, these 89 genes were added to the cluster of genes that showed the highest expression levels from the 16 hr *Miscanthus* cultures (C3 cluster, see below). FunCat analysis (Ruepp 2004) of the remaining 238 genes showed one functional category (C-compound and carbohydrate metabolism) was slightly enriched (see Supplemental Data, Dataset S1, page 5 in Tian et al., 2009).

A third cluster of 142 genes showed the highest relative expression level after 16 hrs of growth of *N. crassa* on *Miscanthus* (C3, FIG. 4A). FunCat analysis (Ruepp 2004) of these 142 genes plus the 89 genes that showed high expression levels in *Miscanthus* cultures at all time points (C3+ cluster; total 231 genes) showed an enrichment for proteins involved with carbon metabolism, including predicted cellulases and hemicellulases (FIG. 4C; also see Supplemental Data, Dataset S1, page 6 in Tian et al., 2009). Of the 23 predicted cellulase genes in the *N. crassa* genome, 18 showed significant increases in expression levels during growth on *Miscanthus* (see Table 1 in Tian et al., 2009), particularly at the 16 hr time point (FIG. 5). Five genes showed an increase in expression level over 200-fold (cbh-1 (CBH(I);

NCU07340, gh6-2 (CBH(II)-like gene; NCU09680), gh6-3 (NCU07190) and two GH61 genes (gh61-4; NCU01050 and NCU07898)).

Plant cell walls are complex structures composed of cellulose microfibrils, hemicellulose, lignin, pectin, cutin, and protein. Thus, we compared expression profiles of *N. crassa* grown on *Miscanthus* to expression profiles of *N. crassa* grown on Avicel, a pure form of crystalline cellulose (see Tian et al., 2009, Supplemental Data, Dataset S1, page 2; data can also be found at bioinfo.townsend.yale.edu/browse.jsp, Experiment IDs 52 and 53). Over 187 genes showed a significant increase in relative expression level during growth of *N. crassa* on Avicel. Of these genes, 114 overlapped with the 231 genes in the C3+ cluster (FIG. 4B). FunCat analysis of the 114-overlap gene set showed a clear enrichment for genes predicted to be involved in carbon metabolism (see Supplemental Data, Dataset S1, page 6 in Tian et al., PNAS, 2009). Within this gene set, there was a further enrichment for secreted proteins; 53 of the 114 gene products were predicted to be secreted. Of these 53 genes, 32 encode predicted proteins that have annotation suggesting a role in plant cell wall degradation, while 16 encode putative or hypothetical proteins that lack any functional prediction. The remaining 61 genes encode predicted intracellular proteins, including ten predicted major facilitator superfamily transporters (NCU00801, NCU00988, NCU01231, NCU04963, NCU05519, NCU05853, NCU05897, NCU06138, NCU08114 and NCU10021) and 23 putative or hypothetical proteins.

Of the 117 genes within the *Miscanthus*-specific cluster (FIG. 4B), 37 encode proteins predicted to be secreted. Nine predicted hemicellulases or enzymes related to the degradation of hemicellulose were identified (NCU00710, NCU04265, NCU04870, NCU05751, NCU05965, NCU09170, NCU09775, NCU09923 and NCU09976) (Tian et al., 2009-Table 2). The remaining 80 *Miscanthus*-specific genes encode predicted intracellular proteins, including genes involved in the metabolism of pentose sugars (for example, NCU00891, xylitol dehydrogenase and NCU00643, a predicted arabinitol dehydrogenase), a predicted sugar transporter (NCU01132), and 48 proteins of unknown function.

Example 2

Secretome Analysis of *N. crassa* Grown on *Miscanthus* and Avicel

Lignocellulose degradation by fungi takes place extracellularly and requires the secretion of proteins associated with depolymerization of cell wall constituents (Lynd et al., 2002). To compare with transcriptional profiling data, which showed that genes encoding predicted cellulases, hemicellulases, and other secreted proteins increased in expression levels when *N. crassa* was grown on *Miscanthus* or Avicel, we analyzed the secretome of *N. crassa* using a shotgun proteomics approach (FIG. 4B). Supernatants from seven day old *Miscanthus* and Avicel cultures were digested with trypsin and analyzed by liquid chromatography nano-electrospray ionization tandem mass spectrometry (MS; see Example 5). Secreted proteins that bound to phosphoric acid swollen cellulose (PASC) were enriched and also analyzed by MS.

A total of 50 proteins were identified with confidence by tandem MS (Tables 2 and 3). There were 34 proteins detected in the *Miscanthus* grown *N. crassa* cultures, while 38 proteins were identified from Avicel grown cultures; twenty-two proteins were detected in both samples. Of these 22 proteins, 21 were predicted to be secreted based on computational analyses and 19 showed increased expression levels in both the *Miscanthus* and Avicel grown cultures (Table 2). The overlap dataset included eight of the 23 predicted cellulases in *N. crassa* (Table 3). There were also five predicted hemicellulases, a predicted β-glucosidase (gh3-4; NCU04952), five proteins with predicted activity on carbohydrates, and two proteins with unknown function (NCU07143 and NCU05137) (Table 4-5).

For Table 2, the annotation was generated by the Broad Institute at webpage broad.mit.edu/annotation/genome/neurospora/Home.html. The “sample detected” was the sample in which peptides were detected for a particular protein. Peptides were validated by manual inspection. A protein was determined to be present if at least 1 peptide was detected in each biological repeat. “TOTAL” refers to peptides detected from a tryptic digest of all extracellular proteins. “PASC BOUND” refers to peptides detected after enrichment for proteins that bind to phosphoric acid swollen cellulose. “UNBOUND” refers to proteins remaining in solution after removal of PASC bound proteins.

TABLE 2

Proteins identified by LC-MS/MS In the culture filtrates of Avicel grown <i>Neurospora crassa</i>		
GENE ID	ANNOTATION	SAMPLE DETECTED
NCU00206	<i>Neurospora crassa</i> hypothetical protein similar to cellobiose dehydrogenase 830 nt	TOTAL
NCU00762	<i>Neurospora crassa</i> endoglucanase 3 precursor 391 nt	TOTAL
NCU01050	<i>Neurospora crassa</i> hypothetical protein similar to endoglucanase II 239 nt	TOTAL
NCU02343	<i>Neurospora crassa</i> hypothetical protein similar to alpha L arabinofuranosidase A 668 nt	TOTAL
NCU04870	<i>Neurospora crassa</i> hypothetical protein similar to acetyl xylan esterase 313 nt	TOTAL
NCU04952	<i>Neurospora crassa</i> hypothetical protein similar to beta D glucoside glucohydrolase 736 nt	TOTAL
NCU05137	<i>Neurospora crassa</i> conserved hypothetical protein 692 nt	TOTAL
NCU05159	<i>Neurospora crassa</i> acetyl xylan esterase precursor 301 nt	TOTAL
NCU05924	<i>Neurospora crassa</i> endo 1 4 beta xylanase 330 nt	TOTAL
NCU07143	<i>Neurospora crassa</i> predicted protein 391 nt	TOTAL
NCU07190	<i>Neurospora crassa</i> conserved hypothetical protein 384 nt	TOTAL
NCU07225	<i>Neurospora crassa</i> endo 1 4 beta xylanase 2 precursor 255 nt	TOTAL
NCU07326	<i>Neurospora crassa</i> predicted protein 327 nt	TOTAL
NCU07340	<i>Neurospora crassa</i> exoglucanase 1 precursor 522 nt	TOTAL
NCU07898	<i>Neurospora crassa</i> predicted protein 242 nt	TOTAL
NCU08189	<i>Neurospora crassa</i> hypothetical protein similar to endo 1 4 beta xylanase 385 nt	TOTAL
NCU08398	<i>Neurospora crassa</i> conserved hypothetical protein 391 nt	TOTAL
NCU08760	<i>Neurospora crassa</i> predicted protein 343 nt	TOTAL

TABLE 2-continued

Proteins identified by LC-MS/MS In the culture filtrates of Avicel grown <i>Neurospora crassa</i>		
GENE ID	ANNOTATION	SAMPLE DETECTED
NCU08785	<i>Neurospora crassa</i> conserved hypothetical protein 291 nt	TOTAL
NCU09491	<i>Neurospora crassa</i> feruloyl esterase B precursor 293 nt	TOTAL
NCU09680	<i>Neurospora crassa</i> exoglucanase 2 precursor 485 nt	TOTAL
NCU09923	<i>Neurospora crassa</i> hypothetical protein similar to beta xylosidase 775 nt	TOTAL
NCU00206	<i>Neurospora crassa</i> hypothetical protein similar to cellobiose dehydrogenase 830 nt	PASC BOUND
NCU00762	<i>Neurospora crassa</i> endoglucanase 3 precursor 391 nt	PASC BOUND
NCU05159	<i>Neurospora crassa</i> acetylxyran esterase precursor 301 nt	PASC BOUND
NCU05955	<i>Neurospora crassa</i> hypothetical protein similar to Cel74a 862 nt	PASC BOUND
NCU07225	<i>Neurospora crassa</i> endo 1 4 beta xylanase 2 precursor 255 nt	PASC BOUND
NCU07340	<i>Neurospora crassa</i> exoglucanase 1 precursor 522 nt	PASC BOUND
NCU08760	<i>Neurospora crassa</i> predicted protein 343 nt	PASC BOUND
NCU09680	<i>Neurospora crassa</i> exoglucanase 2 precursor 485 nt	PASC BOUND
NCU09708	<i>Neurospora crassa</i> conserved hypothetical protein 465 nt	PASC BOUND
NCU00762	<i>Neurospora crassa</i> endoglucanase 3 precursor 391 nt	UNBOUND
NCU01651	<i>Neurospora crassa</i> conserved hypothetical protein 783 nt	UNBOUND
NCU02343	<i>Neurospora crassa</i> hypothetical protein similar to alpha L arabinofuranosidase A 668 nt	UNBOUND
NCU04202	<i>Neurospora crassa</i> nucleoside diphosphate kinase 1 153 nt	UNBOUND
NCU04870	<i>Neurospora crassa</i> hypothetical protein similar to acetyl xylan esterase 313 nt	UNBOUND
NCU04952	<i>Neurospora crassa</i> hypothetical protein similar to beta D glucoside glucohydrolase 736 nt	UNBOUND
NCU05057	<i>Neurospora crassa</i> endoglucanase EG 1 precursor 439 nt	UNBOUND
NCU05137	<i>Neurospora crassa</i> conserved hypothetical protein 692 nt	UNBOUND
NCU05751	<i>Neurospora crassa</i> conserved hypothetical protein 242 nt	UNBOUND
NCU05924	<i>Neurospora crassa</i> endo 1 4 beta xylanase 330 nt	UNBOUND
NCU06239	<i>Neurospora crassa</i> conserved hypothetical protein 514 nt	UNBOUND
NCU07143	<i>Neurospora crassa</i> predicted protein 391 nt	UNBOUND
NCU07190	<i>Neurospora crassa</i> conserved hypothetical protein 384 nt	UNBOUND
NCU07225	<i>Neurospora crassa</i> endo 1 4 beta xylanase 2 precursor 255 nt	UNBOUND
NCU07326	<i>Neurospora crassa</i> predicted protein 327 nt	UNBOUND
NCU07898	<i>Neurospora crassa</i> predicted protein 242 nt	UNBOUND
NCU08189	<i>Neurospora crassa</i> hypothetical protein similar to endo 1 4 beta xylanase 385 nt	UNBOUND
NCU08398	<i>Neurospora crassa</i> conserved hypothetical protein 391 nt	UNBOUND
NCU08412	<i>Neurospora crassa</i> conserved hypothetical protein 401 nt	UNBOUND
NCU08760	<i>Neurospora crassa</i> predicted protein 343 nt	UNBOUND
NCU08785	<i>Neurospora crassa</i> conserved hypothetical protein 291 nt	UNBOUND
NCU09024	<i>Neurospora crassa</i> conserved hypothetical protein 625 nt	UNBOUND
NCU09175	<i>Neurospora crassa</i> conserved hypothetical protein 411 nt	UNBOUND
NCU09267	<i>Neurospora crassa</i> conserved hypothetical protein 1048 nt	UNBOUND
NCU09491	<i>Neurospora crassa</i> feruloyl esterase B precursor 293 nt	UNBOUND
NCU09775	<i>Neurospora crassa</i> hypothetical protein similar to alpha L arabinofuranosidase 343 nt	UNBOUND
NCU09923	<i>Neurospora crassa</i> hypothetical protein similar to beta xylosidase 775 nt	UNBOUND

TABLE 3

22 secreted proteins detected in both <i>Miscanthus</i> and Avicel cultures					
Gene name	Gene annotation	Profiling	kos	CBM1	Signal P
NCU00206.2	CBDH	both	heter	yes	yes
NCU00762.2	probable cellulase precursor	both	16747	yes	yes
NCU01050.2	related to cel1 protein precursor	both	16543	no	yes
NCU04952.2	probable beta-D-glucoside glucohydrolase	both	13732	no	yes
NCU05057.2	probable endo-1,4-beta-glucanase	both	13343	no	yes
NCU05137.2	conserved hypothetical protein	both	11682	no	yes
NCU05924.2	probable endo-beta-1,4-D-xylanase	both	15626	no	yes
NCU05955.2	probable endoglucanase C	both	13535	yes	yes
NCU07143.2	hypothetical	both	No	no	yes
NCU07190.2	CBHII homolog	both	19315	no	yes
NCU07225.2	probable endo-1,4-beta-xylanase B precursor	both	heter	yes	yes
NCU07326.2	related to putative arabinase	both	19534	no	yes
NCU07340.2	CBHI	both	15630	yes	yes
NCU07898.2	related to cel1 protein precursor	both	19600	no	yes
NCU08189.2	similar to endo-1,4-beta xylanase	both	19861	no	yes
NCU08398.2	related to aldose 1-epimerase	both	20310	no	yes
NCU08412.2	hypothetical protein 401 nt	none	No	no	no
NCU08760.2	related to family 61 endoglucanase	both	15664	yes	yes
NCU09024.2	hypothetical protein 625 nt	none	No	no	yes
NCU09175.2	glucan 1,3-beta-glucosidase precursor	both	11750	no	yes

TABLE 3-continued

22 secreted proteins detected in both <i>Miscanthus</i> and Avicel cultures						
Gene name	Gene annotation	Profiling	kos	CBM1	Signal P	
NCU09491.2	feruloyl esterase B precursor	mis	No	no	yes	
NCU09680.2	CBHII	both	15633	yes	yes	

10

TABLE 5-continued

Gene	GH ¹ Family	CBM1 ²	SP ³	MS ⁴	EL ⁵ <i>Miscanthus</i>	EL ⁵ Avicel
NCU00762	5	yes	yes	both	29.6	31.5
NCU03996	6	no	no	ND6	ND	ND
NCU07190	6	no	yes	both	526.0	119
NCU09680	6	yes	yes	both	230.9	251.3
NCU04854	7	no	yes	ND	32.9	10.8
NCU05057	7	no	yes	both	8.7	7.4
NCU05104	7	no	yes	ND	11.6	NC7
NCU07340	7	yes	yes	both	426.4	382.2
NCU05121	45	yes	yes	avi	8.6	17.2
NCU00836	61	yes	yes	ND	91.2	31
NCU01050	61	no	yes	both	206.7	382.1
NCU01867	61	yes	yes	ND	2.2	NC
NCU02240	61	yes	yes	avi	193.5	84
NCU02344	61	no	yes	ND	8.1	4.1
NCU02916	61	yes	yes	ND	85.2	17.7
NCU03000	61	no	yes	ND	NC	ND
NCU03328	61	no	yes	ND	26.4	23.8
NCU05969	61	no	yes	ND	ND	12.7
NCU07520	61	no	yes	ND	ND	ND
NCU07760	61	yes	yes	ND	3.7	NC
NCU07898	61	no	yes	both	376.3	230
NCU07974	61	no	yes	ND	NC	NC
NCU08760	61	yes	yes	both	107.5	44.7

¹Glucoside Hydrolase;²CBM1, carbohydrate binding module;³Signal peptide prediction (signalP = webpage.cbs.dtu.dk/services/SignalP/);⁴MS, mass spectrometry analysis;⁵EL, relative expression level;

ND, not detected;

NC, no change in expression level versus minimal media.

15

Cellulases and Hemicellulases identified by LC-MS/MS						
Predicted hemicellulases in the genome of <i>Neurospora crassa</i>						
Gene ID	GH Family	AV MS	MIS MS	AV Array	MIS Array	
NCU00852	43	-	-	1	1	
NCU00972	53	-	-	9.03	15.6	
NCU01900	43	-	-	10.03	26	
NCU02343	51	-	+	6.63	174.6	
NCU02855	11	+	-	10.2	364	
NCU04997	10	-	-	1	25.6	
NCU05924	10	+	+	55.9	149.3	
NCU05955	74	+	+	19.9	50.5	
NCU05965	43	-	-	1	5.4	
NCU06861	43	-	-	1	1	
NCU07130	10	-	-	1	1	
NCU07225	11	+	+	11.43	33.5	
NCU07326	43	+	+	104.5	426.6	
NCU07351	67	-	-	1	1	
NCU08087	26	-	-	1	1	
NCU08189	10	+	+	39.8	94.4	
NCU09170	43	-	-	1	16.7	
NCU09652	43	-	-	12.2	95.4	
NCU09775	54	-	+	1	48.3	

35 GH Family—Glycosyl Hydrolase Family;

AV MS - Protein detected by LC-MS/MS in the culture filtrates of Avicel grown *Neurospora crassa*. (+) detected, (-) not detected;MIS MS - Protein detected by LC-MS/MS in the culture filtrates of *Miscanthus* grown *Neurospora crassa*. (+) detected, (-) not detected;

AV ARRAY - Fold upregulation after 30 hours of growth on Avicel relative to 16 hours of growth on sucrose from profiling data;

MIS ARRAY - Fold upregulation after 16 hours of growth on *Miscanthus* relative to 16 hours of growth on sucrose from profiling data, peptides detected only in *Miscanthus* culture filtrates.

40 There were 16 proteins identified with confidence only in the Avicel culture and 14 of these were predicted to be secreted (Table 6) including two predicted cellulases (gh61-1; NCU02240 and gh45-1; NCU05121), one xylanase (gh11-1; NCU02855), one predicted protease (NCU04205), three other proteins with predicted activity on carbohydrates (NCU08909, NCU05974 and gh30-1 (NCU04395)), three *Neurospora*-specific proteins of unknown function, and four conserved hypothetical proteins, including one protein with a cellulose binding domain (NCU09764). Twelve proteins were specific for culture filtrates of *Miscanthus* cultures and seven of these were predicted to be secreted (Table 3). Three of the five predicted intracellular proteins were conserved hypothetical proteins. The remaining two included a predicted glyoxal oxidase (NCU09267, identified from the *N. crassa* *Miscanthus* transcriptome) and a nucleoside diphosphate kinase (ndk-1; NCU04202, not identified in the *N. crassa* transcriptome). The seven proteins predicted to be secreted included three predicted esterases (NCU04870, NCU05159, and NCU08785), two predicted xylanases (GH51; NCU02343 and GH54; NCU09775), a predicted β-xilosidase (gh3-7; NCU09923) and a conserved hypothetical protein (NCU05751).

TABLE 5

Cellulases and Hemicellulases identified by LC-MS/MS					
Predicted cellulases in the genome of <i>Neurospora crassa</i>					
Gene ID	GH Family	AV MS	MIS MS	AV ARRAY	MIS ARRAY
NCU00762	5	+	+	31.5	29.6
NCU00836	61	-	-	31	91.2
NCU01050	61	+	+	382.1	206.7
NCU01867	61	-	-	1	1
NCU02240	61	+	-	84	193.5
NCU02344	61	-	-	4.1	8.1
NCU02916	61	-	-	17.7	85.2
NCU03000	61	-	-	1	1
NCU03328	61	-	-	23.8	26.4
NCU03996	6	-	-	2.5	6.3
NCU04854	7	-	-	10.8	32.9
NCU05057	7	+	+	7.4	8.7
NCU05104	7	-	-	1	1
NCU05121	45	+	-	17.2	8.6
NCU05969	61	-	-	12.7	12.3
NCU07190	6	+	+	119	526
NCU07340	7	+	+	382.2	426.4
NCU07520	61	-	-	1	1
NCU07760	61	-	-	1	3.4
NCU07898	61	+	+	230.5	376.3
NCU07974	61	-	-	1	1
NCU08760	61	+	+	44.7	107.5
NCU09680	6	+	+	251.3	230.9

TABLE 6

Proteins identified by LC-MS/MS In the culture filtrates of Avicel grown <i>Neurospora crassa</i>		
GENE ID	ANNOTATION	SAMPLE DETECTED
NCU00206	<i>Neurospora crassa</i> hypothetical protein similar to cellobiose dehydrogenase 830 nt	TOTAL
NCU00762	<i>Neurospora crassa</i> endoglucanase 3 precursor 391 nt	TOTAL
NCU00798	<i>Neurospora crassa</i> predicted protein 313 nt	TOTAL
NCU01050	<i>Neurospora crassa</i> hypothetical protein similar to endoglucanase II 239 nt	TOTAL
NCU01595	<i>Neurospora crassa</i> protein SOF1 446 nt	TOTAL
NCU02240	<i>Neurospora crassa</i> hypothetical protein similar to endoglucanase II 323 nt	TOTAL
NCU02696	<i>Neurospora crassa</i> hypothetical protein similar to DEAD DEAH box RNA helicase 1195 nt	TOTAL
NCU02855	<i>Neurospora crassa</i> endo 1 4 beta xylanase A precursor 221 nt	TOTAL
NCU04952	<i>Neurospora crassa</i> hypothetical protein similar to beta D glucoside glucohydrolase 736 nt	TOTAL
NCU05057	<i>Neurospora crassa</i> endoglucanase EG 1 precursor 439 nt	TOTAL
NCU05137	<i>Neurospora crassa</i> conserved hypothetical protein 692 nt	TOTAL
NCU05924	<i>Neurospora crassa</i> endo 1 4 beta xylanase 330 nt	TOTAL
NCU05955	<i>Neurospora crassa</i> hypothetical protein similar to Cel74a 862 nt	TOTAL
NCU07143	<i>Neurospora crassa</i> predicted protein 391 nt	TOTAL
NCU07190	<i>Neurospora crassa</i> conserved hypothetical protein 384 nt	TOTAL
NCU07225	<i>Neurospora crassa</i> endo 1 4 beta xylanase 2 precursor 255 nt	TOTAL
NCU07326	<i>Neurospora crassa</i> predicted protein 327 nt	TOTAL
NCU07340	<i>Neurospora crassa</i> exoglucanase 1 precursor 522 nt	TOTAL
NCU07898	<i>Neurospora crassa</i> predicted protein 242 nt	TOTAL
NCU08171	<i>Neurospora crassa</i> predicted protein 382 nt	TOTAL
NCU08412	<i>Neurospora crassa</i> conserved hypothetical protein 401 nt	TOTAL
NCU08760	<i>Neurospora crassa</i> predicted protein 343 nt	TOTAL
NCU09491	<i>Neurospora crassa</i> feruloyl esterase B precursor 293 nt	TOTAL
NCU09680	<i>Neurospora crassa</i> exoglucanase 2 precursor 485 nt	TOTAL
NCU09764	<i>Neurospora crassa</i> conserved hypothetical protein 406 nt	TOTAL
NCU00206	<i>Neurospora crassa</i> hypothetical protein similar to cellobiose dehydrogenase 830 nt	PASC BOUND
NCU00762	<i>Neurospora crassa</i> endoglucanase 3 precursor 391 nt	PASC BOUND
NCU05121	<i>Neurospora crassa</i> endoglucanase V 294 nt	PASC BOUND
NCU05955	<i>Neurospora crassa</i> hypothetical protein similar to Cel74a 862 nt	PASC BOUND
NCU07225	<i>Neurospora crassa</i> endo 1 4 beta xylanase 2 precursor 255 nt	PASC BOUND
NCU07340	<i>Neurospora crassa</i> exoglucanase 1 precursor 522 nt	PASC BOUND
NCU08760	<i>Neurospora crassa</i> predicted protein 343 nt	PASC BOUND
NCU09680	<i>Neurospora crassa</i> exoglucanase 2 precursor 485 nt	PASC BOUND
NCU00206	<i>Neurospora crassa</i> hypothetical protein similar to cellobiose dehydrogenase 830 nt	UNBOUND
NCU00762	<i>Neurospora crassa</i> endoglucanase 3 precursor 391 nt	UNBOUND
NCU00798	<i>Neurospora crassa</i> predicted protein 313 nt	UNBOUND
NCU01050	<i>Neurospora crassa</i> hypothetical protein similar to endoglucanase II 239 nt	UNBOUND
NCU04205	<i>Neurospora crassa</i> predicted protein 346 nt	UNBOUND
NCU04395	<i>Neurospora crassa</i> endo 1 6 beta D glucanase precursor 481 nt	UNBOUND
NCU04952	<i>Neurospora crassa</i> hypothetical protein similar to beta D glucoside glucohydrolase 736 nt	UNBOUND
NCU05057	<i>Neurospora crassa</i> endoglucanase EG 1 precursor 439 nt	UNBOUND
NCU05134	<i>Neurospora crassa</i> hypothetical protein 124 nt	UNBOUND
NCU05137	<i>Neurospora crassa</i> conserved hypothetical protein 692 nt	UNBOUND
NCU05852	<i>Neurospora crassa</i> conserved hypothetical protein 254 nt	UNBOUND
NCU05924	<i>Neurospora crassa</i> endo 1 4 beta xylanase 330 nt	UNBOUND
NCU05974	<i>Neurospora crassa</i> hypothetical protein similar to cell wall glucanosyltransferase Mwg1 365 nt	UNBOUND
NCU07143	<i>Neurospora crassa</i> predicted protein 391 nt	UNBOUND
NCU07190	<i>Neurospora crassa</i> conserved hypothetical protein 384 nt	UNBOUND
NCU07225	<i>Neurospora crassa</i> endo 1 4 beta xylanase 2 precursor 255 nt	UNBOUND
NCU07326	<i>Neurospora crassa</i> predicted protein 327 nt	UNBOUND
NCU07340	<i>Neurospora crassa</i> exoglucanase 1 precursor 522 nt	UNBOUND
NCU07898	<i>Neurospora crassa</i> predicted protein 242 nt	UNBOUND
NCU08171	<i>Neurospora crassa</i> predicted protein 382 nt	UNBOUND
NCU08189	<i>Neurospora crassa</i> hypothetical protein similar to endo 1 4 beta xylanase 385 nt	UNBOUND
NCU08398	<i>Neurospora crassa</i> conserved hypothetical protein 391 nt	UNBOUND
NCU08412	<i>Neurospora crassa</i> conserved hypothetical protein 401 nt	UNBOUND
NCU08760	<i>Neurospora crassa</i> predicted protein 343 nt	UNBOUND
NCU08909	<i>Neurospora crassa</i> hypothetical protein similar to beta 1 3 glucanosyltransferase 543 nt	UNBOUND
NCU08936	<i>Neurospora crassa</i> clock controlled gene 15 412 nt	UNBOUND
NCU09024	<i>Neurospora crassa</i> conserved hypothetical protein 625 nt	UNBOUND
NCU09046	<i>Neurospora crassa</i> predicted protein 187 nt	UNBOUND
NCU09175	<i>Neurospora crassa</i> conserved hypothetical protein 411 nt	UNBOUND
NCU09491	<i>Neurospora crassa</i> feruloyl esterase B precursor 293 nt	UNBOUND

ANNOTATION - Generated by the Broad Institute (webpage at broad.mit.edu/annotation/genome/neurospora/Home.html);

SAMPLE DETECTED - Sample in which peptides were detected for a particular protein. Peptides were validated by manual inspection. A protein was determined to be present if at least 1 peptide was detected in each biological repeat.

TOTAL, peptides detected from a tryptic digest of all extracellular proteins;

PASC BOUND, peptides detected after enrichment for proteins that bind to phosphoric acid swollen cellulose;

UNBOUND, proteins remaining in solution after removal of PASC bound proteins.

Many plant cell wall degrading enzymes contain a cellulose-binding module (CBM), which aids in attachment of the enzyme to the substrate (Linder and Teeri 1996). Within the *N. crassa* genome, proteins encoded by 19 genes are predicted to contain a CBM1 domain (Cantarel et al., 2009). Of 5 these 19 genes, 16 showed an increase in relative gene expression in *Miscanthus*-grown cultures (Table 7).

TABLE 7

Effect of <i>Miscanthus</i> and Avicel on <i>N. crassa</i> gene expression					
Gene name	CBM prediction	Annotation	Mis Array	Avicel array	MS
NCU00206	cazy and mips	probable cellobiose dehydrogenase	164	12	both
NCU00710	cazy and mips	acetylxyilan esterase	30	no detect	none
NCU00762	cazy and mips	EG2	29	31	both
NCU00836	cazy and mips	EG, GH61	91	30	none
NCU01867	cazy and mips	EG, GH61	2.2-d10	no	none
				difference	
NCU02240	cazy and mips	EG, GH61	193	84	avi
NCU02916	cazy and mips	EG, GH61	85	17	none
NCU04500	cazy and mips	similar to chitinase 4	no detect	no detect	none
NCU04997	cazy and mips	similar to xylanase	no detect	no detect	none
NCU05121	cazy and mips	EG, GH45	8.5	17	avi
NCU05159	cazy and mips	acetylxyilan esterase precursor	34	10	mis
NCU05955	cazy and mips	GH74	50	19	both
NCU07225	cazy and mips	xylanase	33	11	both
NCU07340	cazy and mips	CBH1	426	382	both
NCU07760	cazy and mips	EG, GH61	3.7	no	none
				difference	
NCU08760	cazy and mips	EG, GH61	107	44	both
NCU09416	cazy and mips	hypothetical	no detect	27	none
NCU09680	cazy and mips	CBH2	230	251	both
NCU09764	cazy and mips	hypothetical	18	16.6	avi

From the 50 proteins identified by MS, 11 contained a CBM1 domain. PASC was used to enrich for proteins that bind to cellulose (see Example 4 for methods). Nine proteins were identified by MS that bound to PASC from the supernatant of *Miscanthus*-grown *N. crassa* cultures, while eight proteins from the Avicel supernatants were identified; seven cellulose binding proteins were identified in both (Tables 2, 3,

40) 8). These included NCU00206, a predicted cellobiose dehydrogenase; gh5-1 (NCU00762), a predicted endoglucanase; NCU05955, a predicted GH74 xyloglucanase; gh11-2 (NCU07225), a predicted endoxylanase; cbh1-1 (NCU07340); gh61-5 (NCU08760), a predicted endoglucanase; and gh6-2 (NCU09680), a predicted cellobiohydrolase 2 precursor.

TABLE 8

Proteins identified by LC-MS/MS in the culture filtrates of Avicel-grown <i>Neurospora crassa</i>		
GENE ID	ANNOTATION	CULTURE
NCU00206	<i>Neurospora crassa</i> hypothetical protein similar to cellobiose dehydrogenase 830 nt	BOTH
NCU00762	<i>Neurospora crassa</i> endoglucanase 3 precursor 391 nt	BOTH
NCU01050	<i>Neurospora crassa</i> hypothetical protein similar to endoglucanase II 239 nt	BOTH
NCU04952	<i>Neurospora crassa</i> hypothetical protein similar to beta D glucoside glucohydrolase 736 nt	BOTH
NCU05057	<i>Neurospora crassa</i> endoglucanase EG 1 precursor 439 nt	BOTH
NCU05137	<i>Neurospora crassa</i> conserved hypothetical protein 692 nt	BOTH
NCU05924	<i>Neurospora crassa</i> endo 1 4 beta xylanase 330 nt	BOTH
NCU05955	<i>Neurospora crassa</i> hypothetical protein similar to Cel74a 862 nt	BOTH
NCU07143	<i>Neurospora crassa</i> predicted protein 391 nt	BOTH
NCU07190	<i>Neurospora crassa</i> conserved hypothetical protein 384 nt	BOTH
NCU07225	<i>Neurospora crassa</i> endo 1 4 beta xylanase 2 precursor 255 nt	BOTH
NCU07326	<i>Neurospora crassa</i> predicted protein 327 nt	BOTH
NCU07340	<i>Neurospora crassa</i> exoglucanase 1 precursor 522 nt	BOTH
NCU07898	<i>Neurospora crassa</i> predicted protein 242 nt	BOTH
NCU08189	<i>Neurospora crassa</i> hypothetical protein similar to endo 1 4 beta xylanase 385 nt	BOTH
NCU08398	<i>Neurospora crassa</i> conserved hypothetical protein 391 nt	BOTH

TABLE 8-continued

Proteins identified by LC-MS/MS in the culture filtrates of Avicel-grown <i>Neurospora crassa</i>		
GENE ID	ANNOTATION	CULTURE
NCU08412	<i>Neurospora crassa</i> conserved hypothetical protein 401 nt	BOTH
NCU08760	<i>Neurospora crassa</i> predicted protein 343 nt	BOTH
NCU09024	<i>Neurospora crassa</i> conserved hypothetical protein 625 nt	BOTH
NCU09175	<i>Neurospora crassa</i> conserved hypothetical protein 411 nt	BOTH
NCU09491	<i>Neurospora crassa</i> feruloyl esterase B precursor 293 nt	BOTH
NCU09680	<i>Neurospora crassa</i> exoglucanase 2 precursor 485 nt	BOTH
NCU00798	<i>Neurospora crassa</i> predicted protein 313 nt	AV
NCU01595	<i>Neurospora crassa</i> protein SOF1 446 nt	AV
NCU02240	<i>Neurospora crassa</i> hypothetical protein similar to endoglucanase II 323 nt	AV
NCU02696	<i>Neurospora crassa</i> hypothetical protein similar to DEAD DEAH box RNA helicase 1195 nt	AV
NCU02855	<i>Neurospora crassa</i> endo 1 4 beta xylanase A precursor 221 nt	AV
NCU04205	<i>Neurospora crassa</i> predicted protein 346 nt	AV
NCU04395	<i>Neurospora crassa</i> endo 1 6 beta D glucanase precursor 481 nt	AV
NCU05121	<i>Neurospora crassa</i> endoglucanase V 294 nt	AV
NCU05134	<i>Neurospora crassa</i> hypothetical protein 124 nt	AV
NCU05852	<i>Neurospora crassa</i> conserved hypothetical protein 254 nt	AV
NCU05974	<i>Neurospora crassa</i> hypothetical protein similar to cell wall glucanosyltransferase Mwg1 365 nt	AV
NCU08171	<i>Neurospora crassa</i> predicted protein 382 nt	AV
NCU08909	<i>Neurospora crassa</i> hypothetical protein similar to beta 1 3 glucanosyltransferase 543 nt	AV
NCU08936	<i>Neurospora crassa</i> clock controlled gene 15 412 nt	AV
NCU09046	<i>Neurospora crassa</i> predicted protein 187 nt	AV
NCU09764	<i>Neurospora crassa</i> conserved hypothetical protein 406 nt	AV
NCU01651	<i>Neurospora crassa</i> conserved hypothetical protein 783 nt	MIS
NCU02343	<i>Neurospora crassa</i> hypothetical protein similar to alpha L arabinofuranosidase A 668 nt	MIS
NCU04202	<i>Neurospora crassa</i> nucleoside diphosphate kinase 1 153 nt	MIS
NCU04870	<i>Neurospora crassa</i> hypothetical protein similar to acetyl xylan esterase 313 nt	MIS
NCU05159	<i>Neurospora crassa</i> acetylxyran esterase precursor 301 nt	MIS
NCU05751	<i>Neurospora crassa</i> conserved hypothetical protein 242 nt	MIS
NCU06239	<i>Neurospora crassa</i> conserved hypothetical protein 514 nt	MIS
NCU08785	<i>Neurospora crassa</i> conserved hypothetical protein 291 nt	MIS
NCU09267	<i>Neurospora crassa</i> conserved hypothetical protein 1048 nt	MIS
NCU09708	<i>Neurospora crassa</i> conserved hypothetical protein 465 nt	MIS
NCU09775	<i>Neurospora crassa</i> hypothetical protein similar to alpha L arabinofuranosidase 343 nt	MIS
NCU09923	<i>Neurospora crassa</i> hypothetical protein similar to beta xylosidase 775 nt	MIS

ANNOTATION - Generated by the Broad Institute (webpage broad.mit.edu/annotation/genome/neurospora/Home.html);

CULTURE - Culture in which peptides were detected for a particular protein.

BOTH, peptides detected in both Avicel and *Misanthus* culture filtrates;

AV, peptides detected only in Avicel culture filtrates;

MIS, peptides detected only in *Misanthus* culture filtrates.

Example 3

Characterization of Extracellular Proteins and Cellulase Activity in Strains Containing Deletions in Genes Identified in the Overlap of the Transcriptome/Secretome Datasets

Of the 22 extracellular proteins detected in both the *Misanthus* and Avicel grown cultures, homokaryotic strains containing deletions in genes encoding 16 of these extracellular proteins were available to the public (Dunlap et al., 2007). None of these 16 deletion strains had been previously characterized with respect to their influence on plant cell wall or cellulose degradation in *N. crassa*. The 16 deletion strains were grown both on media containing sucrose or Avicel as a preferred carbon source. All strains showed a wild type growth phenotype on sucrose. On medium containing Avicel, the bulk growth of the 16 deletion strains was monitored for a 7-day period. After seven days, the total secreted protein, endoglucanase activity, β -glucosidase activity, and aggregate Avicelase activity of the culture filtrates was measured and compared with the wild-type strain from which all the

mutants were derived (FIG. 6). SDS-PAGE was also done on unconcentrated culture supernatants to investigate the relative abundance of secreted proteins.

There were growth deficiencies on Avicel for strains containing deletions of two predicted exoglucanases (cbh-1; NCU07340 and gh6-2; NCU09680) and a predicted β -glucosidase (gh3-4; NCU04952). The cbh-1 mutant was the most severe; after seven days much of the Avicel remained, while in the wild-type strain all of the Avicel was degraded by this time. For 10 of the 16 deletion strains, SDS-PAGE analysis of the secreted proteins showed an altered extracellular protein profile where a single band disappeared, thus allowing assignment of a particular protein band to a predicted gene (FIG. 6A, boxes; FIG. 7). These included NCU00762 (gh5-1), NCU04952 (gh3-4), NCU05057 (gh7-1), NCU05137, NCU05924 (gh10-1), NCU05955, NCU07190 (gh6-3), NCU07326, NCU07340 (cbh-1), and NCU09680 (gh6-2).

For the majority of the deletion strains, the total secreted protein, endoglucanase, β -glucosidase, and Avicelase activities of the culture supernatants were similar to wild type (FIG. 6B, C and Table 9).

TABLE 9

Enzyme Activity of Deletion Strains						
Gene Name	Growth on Avicel	[Secreted Protein] (% of WT)	Azo-CMCCase (% of WT)	Bgl (% of WT)	[CB] (mM)	[GLC] (mM)
NCU00762	***	113 ± 8	33 ± 2	102 ± 2	0.9 ± 0.0	2.6 ± 0.1
NCU01050	***	98 ± 12	92 ± 8	88 ± 5	0.8 ± 0.2	2.9 ± 0.3
NCU04952	***	146 ± 6	124 ± 5	1 ± 0.3	2.24 ± 0.2	0.6 ± 0.0
NCU05057	***	143 ± 10	98 ± 3	100 ± 10	1.7 ± 0.1	3.6 ± 0.1
NCU05137	***	154 ± 12	156 ± 10	178 ± 3	1.0 ± 0.0	3.8 ± 0.1
NCU05924	***	108 ± 3	108 ± 5	101 ± 4	1.1 ± 0.1	2.6 ± 0.2
NCU05955	***	92 ± 10	94 ± 8	98 ± 7	0.9 ± 0.1	2.3 ± 0.1
NCU07190	***	111 ± 7	136 ± 6	92 ± 1	1.1 ± 0.0	2.6 ± 0.0
NCU07326	***	105 ± 4	114 ± 17	85 ± 11	1.0 ± 0.0	2.3 ± 0.0
NCU07340	*	41 ± 2.2	43 ± 9	56 ± 9	0.1 ± 0.0	0.5 ± 0.1
NCU07898	***	84 ± 7	86 ± 1.5	59 ± 15	0.5 ± 0.3	2.3 ± 0.5
NCU08189	***	83 ± 12	80 ± 8	69 ± 15	0.5 ± 0.1	2.3 ± 0.4
NCU08398	***	95 ± 11	107 ± 7	97 ± 3	0.6 ± 0.1	1.8 ± 0.0
NCU08760	***	115 ± 3	126 ± 6	115 ± 8	0.9 ± 0.1	2.6 ± 0.1
NCU09175	***	96 ± 7	115 ± 0	101 ± 8	0.7 ± 0.0	1.9 ± 0.1
NCU09680	**	118 ± 7	165 ± 7	150 ± 1	0.23 ± 0.1	1.7 ± 0.1
WT	***	100 ± 7	100 ± 12	100 ± 6	0.97 ± 0.0	2.4 ± 0.1

Deviations from this trend were seen with the Δ gh5-1 (NCU00762), Δ gh3-4 (NCU04952), Δ NCU05137, Δ cbh-1 (NCU07340), and Δ gh6-2 (NCU09680) mutants. In Δ gh5-1 (NCU00762), Δ gh3-4 (NCU04952), and Δ cbh-1 (NCU07340), Avicelase, endoglucanase or Δ -glucosidase activities were lower than the corresponding wild-type activities. In particular, the deletion of NCU04952 eliminated all β -glucosidase activity from the culture supernatant, as evidenced by PNPGase activity and by higher levels of cellobiose and lower levels of glucose in the Avicelase enzyme assays (FIG. 6B, C). Despite lowering endoglucanase activity, the culture filtrate from Δ gh5-1 (NCU00762) showed no significant deficiency in Avicelase activity relative to the wild-type strain (FIG. 6C). As expected, mutations in cbh-1 (NCU07340) resulted in lower endoglucanase and Avicelase activity, due to poor growth. A strain containing a deletion of NCU09680, encoding a CBH(II)-like protein (gh6-2), also showed reduced cellobiose accumulation, as observed with Δ cbh-1 mutant (FIG. 6C).

Mutations in three strains resulted in an increased level of secreted proteins, especially CBH(I) (FIG. 6A); gh3-4 (NCU04952), gh7-1 (NCU05057) and a hypothetical protein gene (NCU05137). In addition to increased levels of secreted proteins, the Δ NCU05137 mutant showed increased endoglucanase, β -glucosidase, and Avicelase activity (FIG. 6B, C). NCU05137 is highly conserved in the genomes of a number of filamentous ascomycete fungi, including other cellulolytic fungi, but notably does not have an ortholog in *T. reesei* (FIG. 2). It is possible that the increase in CBH(I) levels observed in Δ gh3-4, Δ gh7-1, and Δ NCU05137 could be due to either increased secretion, protein stability or, alternatively, feedback that results in an increase in expression of cbh-1. To differentiate these possibilities, the profile of extracellular proteins produced by Δ NCU05137 and Δ gh3-4 (NCU04952) was compared with gene expression levels of cbh-1 (NCU07340) and gh6-2 (CBH(II); NCU09680) as assayed by quantitative RT-PCR (FIG. 8). The strains Δ NCU05137 and Δ gh3-4 showed a higher level of CBH(I) protein as early as two days in an Avicel-grown culture. Quantitative RT-PCR of cbh-1 and gh6-2 from Avicel-grown cultures showed that both genes exhibited high expression levels in wild type and the Δ NCU05137 and Δ gh3-4 mutants after two days of growth. However, although expression of both of these genes decreased significantly on day three in the wild-type strain,

both cbh-1 and gh6-2 expression levels increased in the Δ NCU05137 mutant, and decreased less than wild type in Δ gh3-4 (FIG. 8). Sustained expression of cbh-1 and gh6-2 genes in the Δ NCU05137 and Δ gh3-4 mutants could be responsible for the observed increase in CBH(I) and CBH(II) protein levels.

Example 4

Materials and Methods for Transcriptome and Secretome Studies

Strains

All *Neurospora crassa* strains were obtained from the Fungal Genetics Stock Center (FGSC; webpage fgsc.net) (Supplemental Data, Dataset S1, page 1 in Tian et al., 2009). Gene deletion strains were from the *N. crassa* functional genomics project (Dunlap et al., 2007). *Trichoderma reesei* QM9414 was a gift from Dr. Monika Schmoll (Vienna University of Technology). Strains were grown on Vogel's salts (Vogel 1956) with 2% (w/v) carbon source (*Miscanthus*, sucrose or Avicel (Sigma)). *Miscanthusxgiganteus* (milled stem to ~0.1 mm) was a gift from the University of Illinois.

Enzyme Activity Measurements

Total extracellular protein content was determined using a Bio-Rad DC Protein Assay kit (Bio-Rad). Endoglucanase activity in culture supernatants was measured with an azo-CMC kit (Megazyme SCMCL). β -glucosidase activity was measured by mixing 10-fold diluted culture supernatant with 500 μ M 4-nitrophenyl β -D-glucopyranoside in 50 mM sodium acetate buffer, pH 5.0, for 10 minutes at 40° C. The reaction was quenched with 5% w/v sodium carbonate, and the absorbance at 400 nm was measured. Avicelase activity was measured by mixing 2-fold diluted culture supernatant with 50 mM sodium acetate, pH 5.0, and 5 mg/mL Avicel at 40° C. Supernatants were analyzed for glucose content using a coupled enzyme assay with glucose oxidase/peroxidase. Fifty μ L of the avicelase reaction was transferred to 150 μ L of glucose detection reagent containing 100 mM sodium acetate pH 5.0, 10 U/mL horseradish peroxidase, 10 U/mL glucose oxidase, and 1 mM o-dianisidine. After 30 minutes absorption was measured at 540 nm. Cellobiose concentrations were determined using a coupled enzyme assay with cellobiose dehydrogenase (CDH) from *Sporotrichum thermophile*.

CDH was isolated from *S. thermophile* similar to previous reports (Canevascini 1988). Fifty µL of the avicelase reaction was transferred to 250 µL of cellobiose detection reagent containing 125 mM sodium acetate pH 5.0, 250 µM dichlorophenol indophenol, and 0.03 mg/mL CDH. After 10 minutes absorption was measured at 530 nm.

RNA Isolation, Microarray Analysis, and Signal Peptide Predictions

Mycelia were harvested by filtration and flash frozen in liquid nitrogen. Total RNA was isolated using trizol (Tian et al., 2007; Kasuga et al., 2005). Microarray hybridization and data analysis were as previously described (Tian et al., 2007). Normalized expression values were analyzed using BAGEL (Bayesian analysis of gene expression levels) (Townsend 2004; Townsend and Hartl 2002), which infers relative gene expression levels and credible intervals for each gene at each experimental time point. Signal peptides were predicted using the N-terminal 70 amino acid region of each predicted protein with the signalP3 program (webpage cbs.dtu.dk/services/SignalP-3.0/). Original profiling data is obtainable at (webpage yale.edu/townsend/Links/fldatabase/).

Protein Gel Electrophoresis

Except where otherwise noted, unconcentrated culture supernatants were treated with 5×SDS loading dye and boiled for 5 minutes before loading onto Criterion 4-15% Tris-HCl polyacrylamide gels. Coomassie dye was used for staining.

Preparation of Tryptic Peptides for Secretome Analysis

Culture supernatants were concentrated with 10 kDa MWCO PES spin concentrators. Cellulose binding proteins were isolated from the culture supernatant by addition of phosphoric acid swollen cellulose (PASC). Five mL of a suspension of 10 mg/mL PASC was added to 10 mL of culture supernatant. After incubation at 4° C. for 5 minutes, the mixture was centrifuged and the pelleted PASC was then washed with 20 pellet volumes of 100 mM sodium acetate pH 5.0. The supernatant after treatment with PASC was saved as the unbound fraction and concentrated. 36 mg of urea, 5 µL of 1M Tris PH 8.5, and 5 µL of 100 mM DTT were then added to 100 µL of concentrated culture supernatant or protein-bound PASC and the mixture was heated at 60° C. for 1 hour. After heating 700 µL of 25 mM ammonium bicarbonate and 140 µL of methanol were added to the solution followed by treatment with 50 µL of 100 µg/mL trypsin in 50 mM sodium acetate pH 5.0. For the PASC bound proteins, the PASC was removed by centrifugation after heating, and the supernatant was then treated with trypsin. The trypsin was left to react overnight at 37° C. After digestion the volume was reduced by speedvac and washed with MilliQ water three times. Residual salts in the sample were removed by using OMIX microextraction pipette tips according to the manufacturer's instructions.

Liquid Chromatography of Tryptic Peptides

Trypsin-digested proteins were analyzed using a tandem mass spectrometer that was connected in-line with ultraperformance liquid chromatography (HPLC). Peptides were separated using a nanoAcuity HPLC (Waters, Milford, Mass.) equipped with C18 trapping (180 µm×20 mm) and analytical (100 µm×100 mm) columns and a 10 µL sample loop. Solvent A was 0.1% formic acid/99.9% water and solvent B was 0.1% formic acid/99.9% acetonitrile (v/v). Sample solutions contained in 0.3 mL polypropylene snap-top vials sealed with septa caps (Wheaton Science, Millville, N.J.) were loaded into the nanoAcuity autosampler prior to analysis. Following sample injection (2 µL, partial loop), trapping was performed for 5 min with 100% A at a flow rate of 3 µL/min. The injection needle was washed with 750 µL each of solvents A and B after injection to avoid cross-con-

tamination between samples. The elution program consisted of a linear gradient from 25% to 30% B over 55 min, a linear gradient to 40% B over 20 min, a linear gradient to 95% B over 0.33 min, isocratic conditions at 95% B for 11.67 min, a linear gradient to 1% B over 0.33 min, and isocratic conditions at 1% B for 11.67 min, at a flow rate of 500 nL/min. The analytical column and sample compartment were maintained at 35° C. and 8° C., respectively.

Mass Spectrometry

The column was connected to a NanoEase nanoelectrospray ionization (nanoESI) emitter mounted in the nanoflow ion source of a quadrupole time-of-flight mass spectrometer (Q-T of Premier, Waters). The nanoESI source parameters were as follows: nanoESI capillary voltage 2.3 kV, nebulizing gas (nitrogen) pressure 0.15 mbar, sample cone voltage 30 V, extraction cone voltage 5 V, ion guide voltage 3 V, and source block temperature 80° C. No cone gas was used. The collision cell contained argon gas at a pressure of 8×10⁻³ mbar. The T of analyzer was operated in "V" mode. Under these conditions, a mass resolving power 1 of 1.0×10⁴ (measured at m/z=771) was routinely achieved, which is sufficient to resolve the isotopic distributions of the singly and multiply charged peptide ions measured in this study. Thus, an ion's mass and charge could be determined independently, i.e., the ion charge was determined from the reciprocal of the spacing between adjacent isotope peaks in the m/z spectrum. External mass calibration was performed immediately prior to analysis, using solutions of sodium formate. Survey scans were acquired in the positive ion mode over the range m/z=450-1800 using a 0.95 s scan integration and a 0.05 s interscan delay. In the data-dependent mode, up to five precursor ions exceeding an intensity threshold of 35 counts/second (cps) were selected from each survey scan for tandem mass spectrometry (MS/MS) analysis. Real-time deisotoping and charge state recognition were used to select 2+, 3+, 4+, 5+, and 6+ charge state precursor ions for MS/MS. Collision energies for collisionally activated dissociation (CAD) were automatically selected based on the mass and charge state of a given precursor ion. MS/MS spectra were acquired over the range m/z=50-2500 using a 0.95 s scan integration and a 0.05 s interscan delay. Ions were fragmented to achieve a minimum total ion current (TIC) of 30,000 cps in the cumulative MS/MS spectrum for a maximum of 3 s. To avoid the occurrence of redundant MS/MS measurements, real time exclusion was used to preclude re-selection of previously analyzed precursor ions over an exclusion width of ±0.25 m/z unit for a period of 180 s.

Mass Spectrometry Data Analysis

The data resulting from LC-MS/MS analysis of trypsin-digested proteins were processed using ProteinLynx Global Server software (version 2.3, Waters), which performed background subtraction (threshold 35% and fifth order polynomial), smoothing (Savitzky-Golay2 10 times, over three channels), and centroiding (top 80% of each peak and minimum peak width at half height four channels) of the mass spectra and MS/MS spectra. The processed data were searched against the *N. crassa* database (Broad Institute) using the following criteria: tryptic fragments with up to five missed cleavages, precursor ion mass tolerance 50 ppm, fragment ion mass tolerance 0.1 Da, and the following variable post-translational modifications: carbamylation of N-termini and Lys side chains, Met oxidation, and Ser/Thr dehydration. The identification of at least three consecutive fragment ions from the same series, i.e., b or y-type fragment ions, was required for assignment of a peptide to an MS/MS spectrum.

The MS/MS spectra were manually inspected to verify the presence of the fragment ions that uniquely identify the peptides.

Quantitative RT-PCR

The RT-PCR was performed in an ABI7300 with reagents from Qiagen (SYBR-green RT-PCR kit (Cat No. 204243)). The primers for CBH(I) (NCU07340) were: forward 5'-ATCTGGGAAGCGAACAAAG-3' (SEQ ID NO: 16) and reverse 5'-TAGCGGTGCGTCGGAATAG-3' (SEQ ID NO: 17). The primers for CBH(II) (NCU09680) were: forward 5'-CCCATCACCCTACTTAC-3' (SEQ ID NO: 18) and reverse 5'-CCAGCCCTGAACACCAAG-3' (SEQ ID NO: 19). Actin was used as a control for normalization. The primers for actin were: forward 5'-TGA TCT TAC CGA CTA CCT-3' (SEQ ID NO: 20) and reverse 5'-CAG AGC TTC TCC TTG ATG-3' (SEQ ID NO: 21). Quantitative RT-PCR was performed according to Dementhon et al., (2006).

Example 5

Discussion of Transcriptome and Secretome Studies

Degradation of plant biomass requires the production of many different enzymatic activities, which are regulated by the type and complexity of the available plant material (FIG. 9) (Bouws et al., 2008). The first systematic analyses of plant cell wall degradation by a cellulolytic fungus are described here, which include transcriptome, secretome, and mutant analyses. Profiling data showed that *N. crassa* coordinately expresses a host of extracellular and intracellular proteins when challenged by growth on *Miscanthus* or Avicel (FIG. 9). Many of the most highly expressed genes during growth on cellulosic substrates encode proteins predicted to be involved in the metabolism of plant cell wall polysaccharides, many of which were identified by MS analyses. Genome comparisons of filamentous fungi show a large number of glycosyl hydrolases (~200) with varying numbers of predicted cellulases, from 10 in *T. reesei* (Martinez et al., 2008) to 60 in *Podospora anserina* (Espagne et al., 2008), a dung-degrading species closely related to *N. crassa*. A comparison between these results and a recent transcriptome/secretome study on the white rot basidiomycete fungus, *Phanerochaete chrysosporium*, (Wymelenberg et al., 2009) showed little overlap in regulated genes (18 genes) and secreted proteins (2 proteins) when both species were grown on pure cellulose. These data suggest that different fungi may utilize different gene sets for plant cell wall degradation. However, one aspect that both studies had in common was the high number of uncharacterized genes/proteins associated with cellulose degradation. Other cellulolytic fungi, including *P. chrysosporium*, do not have the genetic and molecular tools that are readily available with *N. crassa*. Using the functional genomic tools available with *N. crassa*, both the function and redundancy of plant cell wall degrading enzyme systems can be addressed to create optimal enzyme mixtures for industrial production of liquid fuels from lignocellulose biomass.

In this study, it was found that cellobiohydrolase(I) (CBH(I)) in *N. crassa* is the most highly produced extracellular protein during growth on Avicel or *Miscanthus*, and deletion of this gene caused the most severe growth deficiencies on cellulosic substrates. These results are similar to those reported in *T. reesei* (Suominen et al., 1993; Seiboth et al., 1997). Deletion of cellobiohydrolase(II) also caused growth deficiencies on cellulosic substrates, but to a much lesser extent than CBH(I), suggesting that exoglucanase activity in *N. crassa* is predomi-

nantly from CBH(I) and that cellulases and other CBHs do not compensate for the loss of CBH(I). Here, it was shown that the three most highly produced endoglucanases during growth on cellulosic substrates are the proteins encoded by NCU05057, NCU00762, and NCU07190. These proteins have homology to endoglucanases EG1, EG2, and EG6, respectively. Deletion of these genes did not affect growth on Avicel, although differences in the secreted protein levels and endoglucanase activity were observed. Unexpectedly, in the ΔNCU05057 strain, extracellular protein levels were much higher, especially CBH(I), suggesting that to maintain the wild-type growth phenotype on crystalline cellulose the mutant was forced to increase production of other cellulases or that the products of NCU05057 catalysis may repress cellulase production. It was concluded that no one endoglucanase in *N. crassa* is required for growth on crystalline cellulose and that the different endoglucanases have overlapping substrate specificities.

The glycoside hydrolase family 61 enzymes are greatly expanded in *N. crassa* compared to *T. reesei* (Martinez et al., 2008). These enzymes have poorly defined biological function, but their general conservation and abundance in cellulolytic fungi suggests an important role in plant cell wall metabolism. Here, genes for 10 of the 14 GH61 enzymes were identified in the *N. crassa* transcriptome, suggesting that these enzymes are utilized during growth on cellulosic biomass. The four GH61 deletion strains tested showed only small differences compared to wild type in the secreted protein levels, endoglucanase, and total cellulase activities. However, analyses of additional GH61 mutants and the capacity to create strains containing multiple mutations in *N. crassa* via sexual crosses will address redundancy and expedite functional analysis of this family.

In addition to predicted cellulase genes, genes encoding hemicellulases, carbohydrate esterases, β-glucosidases, β-xylanidases, and other proteins predicted to have activity on carbohydrates were identified in the *N. crassa* transcriptome from both *Miscanthus* and Avicel. The fact that Avicel contains no hemicellulose components suggests that cellulose is probably the primary inducer of genes encoding plant cell wall degrading enzymes in *N. crassa*. However, genes encoding some hemicellulases and carbohydrate esterases were only expressed during growth on *Miscanthus*. Similarly, in other cellulolytic fungi such as *T. reesei* and *Aspergillus niger*, genes encoding some cellulases and hemicellulases are coordinately regulated, while others are differentially regulated (Stricker et al., 2008). As expected, deletions of non-cellulase genes had little effect on growth on Avicel or cellulase activity, with the exception of NCU05137 and gh3-4. The ΔNCU05137 strain secreted more protein, had higher cellulase activity, and showed higher expression of cbh-1 (CBH(I)) and gh6-2 (CBH(II)) than wild type. NCU05137 encodes a secreted hypothetical protein that has no homology to proteins of known function, but is highly conserved in other cellulolytic fungi (FIG. 2; E value 0.0). NCU05137 also has more distant homologs, but also of unknown function, in a number of bacterial species. The protein product of NCU05137 may interfere with signaling processes associated with induction of cellulase gene expression in *N. crassa* (FIG. 9). Similarly, mutations in gh3-4 (NCU04952) also increased CBH(I) activity. Deletion of this gene completely removed PNPGase activity and cellobiose accumulated in in vitro cellulase assays using Δgh3-4 culture filtrates. All the data together suggested that NCU04952 encodes the primary extracellular β-glucosidase in *N. crassa*. These data were consistent with catabolite repression of cellulase production by glucose.

77

Extracellular degradation of cellulose and hemicellulose results in the formation of soluble carbohydrates that are subsequently transported into the cell (FIG. 9). In this study, 10 genes encoding permeases/transports were identified which showed significantly increased expression when *N. crassa* was grown on *Miscanthus* or Avicel, suggesting their involvement in transport of plant cell wall degradation products into the cell. The major degradation products by cellulases and hemicellulases in vitro are cellobiose, glucose, xylobiose, and xylose. Some of these transporters may be functionally redundant or capable of transporting oligosaccharides. The function of these putative transporters was further explored (see Examples 7-9). Construction of downstream processing strains capable of transporting oligosaccharides by heterologous expression of *N. crassa* transporters may improve industrial fermentation of biomass hydrolysis products. None of these transporters or what they may transport has been characterized at the molecular or functional level in any filamentous fungi.

Many genes that showed increased expression levels during growth on *Miscanthus* and Avicel encode proteins of unknown function that are conserved in other cellulolytic fungi. By assessing the phenotype of only 16 strains, a mutant in a gene encoding a protein of unknown function that significantly affects cellulase activity was identified. The well-understood genetics and availability of functional genomic resources in *N. crassa* make it an ideal model organism to determine the biological function of these proteins, as well as regulatory aspects of cellulase and hemicellulase production, and to dissect redundancies and synergies between extracellular enzymes involved in the degradation of plant cell walls.

78

Example 6

Screening of Mutants of Genes Upregulated during Growth on *Miscanthus*

In order to analyze additional genes identified in the transcriptional profiling experiment, the phenotypes of mutants of 188 genes that were upregulated in *Neurospora* grown on *Miscanthus* for 16 hours were analyzed (see Example 1). A knockout mutant of each gene was grown on minimal Vogel's medium for 10-14 days. Conidia were harvested with 2 mL ddH₂O and inoculated into 100 mL media in 250 mL flasks at a concentration of 10⁶ conidia per mL. One of three different carbon sources was added to each flask: 2% sucrose, 2% Avicel, or 2% *Miscanthus* (1 mm particles from Calvin Laboratory, University of California, Berkeley, Calif.). Cultures were grown at 25° C. with 220 rpm of shaking for 4 days.

Table 10 lists the phenotypes of the mutants that showed a significant difference in cellulase activity and growth on Avicel or *Miscanthus* compared to wild-type. Growth on Avicel or *Miscanthus* was evaluated by eye with a “+” scoring system. Wild-type growth was set at “++”. Total protein in the culture supernatant was measured by Bradford assay (100 µl supernatant to 900 µl Bradford dye). Endoglucanase activity was measured with the Azo-CMC kit from Megazyme and indicated in Table 10 as the percentage of endoglucanase activity in the mutant compared to wild type. Total cellulase activity was measured by detecting cellobiose levels in the supernatant as described in Example 4. Results are indicated in Table 10 as a percentage of wild-type.

US 8,431,360 B2

79

80

NCU#	FGSC #	Broad Annotation (Domains)	Pfam*	Up-Regulation	Growth (Avi, Mis)	% WT Bradford (Avi)	% WT Bradford (Mis)	% WT endo (Avi)	% WT endo (Mis)	% WT cellobiose (Avi)	% WT cellobiose (Mis)
NCU00130.2	FGSC 11823	beta-glucosidase (GH1)	Glycosyl Hydrolase 1 (2.5e-196) no significant hit	394.6	++, ++	203.2477947	118.3987972	152.2858378	129.3547494	n/a	n/a
NCU00248.2	FGSC 12214	Predicted Protein		9.74	+, ++	86.96013289	86.04471858	30.39187506	156.5050144	93.05143946	89.27698219
NCU00326.2	FGSC 15868	Conserved Hypothetical (SMP_30/gluconolactonase)	SMP-30/Glucuronolactonase/ LRE-like region (3.5e-82)	7.7	+, ++	33.02879291	144.1210486	39.91568458	227.0366809	89.76872415	79.05154639
NCU00762.2	FGSC 16747	Endoglucanase-3 precursor (GH5, CBD1)	Cellulase (1.4e-69), Fungal cellulose binding domain (9.2e-14) Glycosyl hydrolases family 2 (1.7e-145)	29.6	++, ++	104.3504411	84.57056944	26.92790756	39.26890058	n/a	n/a
NCU00810.2	FGSC 11285	Similar to Glycosyl Hydrolase (GH2, beta-galactosidase)		5.3	++, ++	163.805047	123.5564757	161.2908993	159.4983744	102.2745211	91.73345664
NCU00890.2	FGSC 16749	Similar to beta-mannosidase (GH2)	Glycosyl hydrolases family 2 (4.1e-06)	20.45	+, +	47.57417803	101.5974441	43.25546345	164.0819718	n/a	n/a
NCU03328.2	FGSC 16589	Conserved Hypothetical (GH61)	Glycosyl hydrolase family 61 (2.3e-10)	26.4	++, ++	100.1752848	109.9667248	142.6962073	167.0075481	n/a	n/a
NCU03415.2	FGSC 12922	Aldehyde Dehydrogenase	Aldehyde dehydrogenase family (2.5e-267) haloacid dehalogenase-like hydrolase (9.2e-21)	9.8	++, ++	104.2278304	96.61435373	96.4633125	63.45523329	76.96643943	103.1273983
NCU03731.2	FGSC 18653	Similar to HAD Superfamily Hydrolase		2.7	++, ++	131.3691128	110.5801446	145.0235135	134.627995	230.1450412	100.4172375
NCU03753.2	FGSC 16379	cgg-1 (clock controlled gene)	no significant hit	10.5	++, ++	107.6792892	111.3481086	74.42402278	129.2196777	n/a	n/a

-continued

NCU#	Broad Annotation (Domains)	Pfam*	Up-Regulation	Growth (Avi, Mis)	% WT Bradford (Avi)	% WT Bradford (Mis)	% WT endo (Avi)	% WT endo (Mis)	% WT cellobiose (Avi)	% WT cellobiose (Mis)
NCU04197.2	FGSC Conserved	no significant hit	5.04	++, ++	103.0668127	99.08305414	108.9737808	89.86128625	75.17285531	96.05075054
NCU04249.2	FGSC Hypothetical protein	no significant hit	5.3	++, ++	93.29682366	106.1012167	79.0053469	84.16141236	64.07989522	100.124185
NCU04287.2	FGSC Predicted Protein	no significant hit	4.7	++, ++	115.5157859	102.2361065	125.5086234	127.9282577	202.516129	183.8679245
NCU04349.2	FGSC Similar to 18634	BCDHK_Adon3 (4.7e-78), HistidineATPase_c (6.9e-14)	2.9	++, ++	87.87776465	89.36205196	71.41381803	145.2415813	208.4329349	101.2993763
NCU04475.2	FGSC Predicted Protein	no significant hit	76.7	++, ++	98.10205352	122.2034851	156.3643221	127.0676692	n/a	n/a
NCU04997.2	FGSC Similar to 15623	Glycosyl hydrolase family 10 (3.3e-148), Fungal cellulose binding domain (2.1e-16)	25.6	++, ++	105.3520176	114.5840184	123.3295466	231.6983895	136.4189483	102.5403983
NCU05057.2	FGSC Endoglucanase HG-1 precursor (GH7)	Glycosyl hydrolase family 7 (3.3e-189) Cutinase (3.4e-110), Fungal cellulose binding domain (7.4e-14)	8.7	++, ++	137.5316563	95.69220651	133.5226686	174.2679556	182.023775	97.81330657
NCU05159.2	FGSC esterase precursor (Cutinase, CBD1)	Cutinase (3.4e-110), Fungal cellulose binding domain (7.4e-14)	34.8	++, ++	86.18543871	39.51658235	92.2873845	67.11779449	n/a	n/a
NCU05493.2	FGSC Predicted Protein	no significant hit	4.5	+, ++	73.25266013	104.4102564	102.3841739	116.8954593	70.37185126	99.42837929
NCU05519.2	FGSC Similar to Tna1 (MFS transporter)	Major Facilitator Superfamily (3.7e-40) GDSL-like Lipase/Acylhdrolase	2.8	++, ++	85.31191321	101.0666667	118.8447721	87.77719113	51.6886931	97.87501655
NCU05751.2	FGSC Conserved	Hypothetical	3.9	+, ++	97.01648237	111.4051282	114.7202911	136.3780359	87.71492649	105.5920583

-continued

NCU#	Broad Annotation (Domains)	FGSC #	Pfam*	Up-Regulation	Growth (Avi, Mis)	% WT Bradford (Avi)	% WT Bradford (Mis)	% WT endo (Avi)	% WT endo (Mis)	% WT cellobiose (Avi)	% WT cellobiose (Mis)
	(GDSL-like lipase)		(1.3e-11)								
NCU05770.2	FGSC Peroxidase/ Catalase 2	11532	Peroxidase (9.4e-195)	11.9	++, ++	109.8630989	86.73412029	69.1872525	146.2155388	n/a	n/a
NCU05853	FGSC Sugar Transporter	13771	Sugar Transporter	130.7	+	40.27924687	24.41259790	n/a	n/a	n/a	n/a
NCU05897.2	FGSC Similar to L-fucose permease (MFS transporter)	13717	Major Facilitator Superfamily (3.8e-16)	20.9	+, ++	33.78464142	34.72754541	26.3266891	86.25954198	n/a	n/a
NCU05932.2	FGSC Predicted protein	19952	no significant hit	38.2	++, ++	70.89826428	76.87132044	80.78910753	117.9596823	58.07431478	96.7108463
NCU06009.2	FGSC Similar to aldo/keto reductase family (4.8e-63)	14922	Aldo/zero reductase family (4.8e-63)	6.9	+, ++	148.6633726	74.06784413	120.602266	99.48075748	70.89513625	97.00573241
NCU06490.2	FGSC Conserved Hypothetical	15539	no significant hit	13.8	+, ++	77.46104143	80.26332677	76.95289207	79.38301772	59.91109168	99.99371385
NCU07340.2	FGSC Exoglucanase-1 precursor, CBHI (GH7)	15630	Glycosyl hydrolase family 7 (le-999), Fungal cellulose binding domain (4.9e-18)	426.4	+, ++	21.09634551	95.21973786	35.54661301	96.99134496	93.62619808	78.44902553
NCU07853.2	FGSC Uricase	19036	Uricase (1.7e-119)	4.3	++, ++	n/a	n/a	120.9286362	168.2340648	65.7594546	99.14659177
NCU07997.2	FGSC Predicted protein	18273	no significant hit	4.5	++, ++	n/a	n/a	148.127436	98.111912226	60.65548063	93.78704271
NCU08114.2	FGSC Similar to MFS hexose transporter (MFS transporter)	17869	Sugar (and other) transporter (5.1e-88), Major Facilitator Superfamily (3.8e-24)	6.7	+, ++	81.69263905	79.22624054	85.18187239	92.97495418	58.83068556	93.1432252
NCU08744.2	FGSC Predicted protein, possible	11387	n/a	2.3	++, ++	n/a	n/a	168.8527368	110.7628004	136.2451567	97.44134197

-continued

NCU#	Broad Annotation (Domains)	FGSC #	Pfam*	Up-Regulation	Growth (Avi, Mis)	% WT Bradford (Avi)	% WT Bradford (Ms)	% WT endo (Avi)	% WT endo (Ms)	% WT cellobiose (Avi)	% WT cellobiose (Mis)	
	TF (basic region leucine zipper)											
NCU08746.2	FGSC Conserved 18358		Hypothetical (starch binding domain)	Starch binding domain (5.3e-54)	6	++, ++	98.69504624	79.111410149	111.0713576	120.2504582	447.2796518	100.5753667
NCU08760.2	FGSC Predicted 15664		Protein (CBDD1)	Fungal cellobiose binding domain (1.9e-11), Glycosyl hydrolase family 61 (1.3e-9) no significant hit	107.5	++, ++	158.1395349	86.17964534	208.2590783	81.00013738	97.32646961	84.34251774
NCU09108.2	FGSC 19207		Conserved Hypothetical		4.1	++, ++	n/a	93.22148788	111.8077325	60.89420655	97.24517906	
NCU09495.2	FGSC 12411		set-6, histone methyltransferase	SET domain (6.9e-5)	26.2	++, ++	109.3300111	122.5327679	129.9223915	130.8971013	152.7495439	92.25216554
NCU09680.2	FGSC 15633		Exoglucanase-2 precursor, CBH2 (GH6, CBD1)	Glycosyl hydrolases family 6 (1.1e-152), Fungal cellulose binding domain (1.2e-13)	230.9	+, ++	102.7131783	95.20046261	89.54680464	102.6789394	94.61873756	83.87661343
NCU110045.2	FGSC 18480		pectinesterase precursor	Pectinesterase (4.4e-22)	10.9	+, ++	105.3085012	101.5138772	109.8886901	132.5290165	83.25906421	103.6151641

*Note: All sequences were searched against Pfam 1s models and hits were accepted with an e-value <0.001

Further Analyses of Transporter Genes

As described in Example 1, ten genes encoding predicted sugar transporter proteins showed increased expression levels when *Neurospora* was grown on *Miscanthus* and Avicel: NCU00801, NCU00988, NCU01231, NCU04963, NCU05519, NCU05853, NCU05897, NCU06138, NCU08114 and NCU10021. Deletion strains for nine of these genes were available from the Fungal Genetics Stock Center. A deletion strain of NCU10021 was not available.

Deletion mutations of NCU05853, NCU05897, or NCU08114 resulted in strains that showed a growth defect on

Miscanthus or Avicel and/or had a cellulase enzyme defect (see Example 6; Table 10). ΔNCU05853 showed reduced growth on Avicel and reduced endoglucanase activity compared to wild-type. ΔNCU05897 showed reduced growth on Avicel and reduced endoglucanase activity compared to wild-type, and ΔNCU08114 showed reduced growth on Avicel and reduced cellobiose levels compared to wild-type. Notably, in a comparison with expression analysis of *Sporotrichum thermophile*, another filamentous fungus, the homologs of NCU05853 (ST8454) and NCU08114 (ST5194) were also upregulated when *S. thermophile* was grown on Avicel compared to glucose (see Example 8, Table 11), further indicating their importance in cellulose utilization.

Gene Name	Like NCU#	Gene Length	Glu	Avi	Cot	Glu_norm
jgi Spoth1 108890	NCU00988	1937	322	370	293	42.97830583
estExt_fgenesh1_pg.C_60848						
jgi Spoth1 48439	NCU00132	1539	113	59	56	15.08244894
e_gw1.3.3367.1						
jgi Spoth1 79030	NCU01231	1776	1171	1206	469	156.2968824
estExt_Genewise1Plus.C_31624						
jgi Spoth1 116270	NCU05519	1680	103	78	54	13.74771895
estExt_fgenesh1_pm.C_50266						
jgi Spoth1 84164	NCU05853	1706	2703	20760	14284	360.7775176
estExt_Genewise1Plus.C_62100						
jgi Spoth1 102977	NCU05897	1446	1510	546	322	201.5442292
fgenesh1_pm.5.#_763						
jgi Spoth1 84305	NCU06138	1605	1131	1330	2376	150.9579624
estExt_Genewise1Plus.C_70023						
jgi Spoth1 114107	NCU08114	1945	2246	22423	10779	299.7803568
estExt_fgenesh1_pm.C_20669						
jgi Spoth1 112305	NCU10021	2026	6204	5287	5619	828.0664888
estExt_fgenesh1_kg.C_60263						
jgi Spoth1 43941	NCU00801	1614	41	71	159	5.472392979
e_gw1.2.4209.1						
jgi Spoth1 62521	NCU04963	2204	799	1548	641	106.6449266
estExt_Genewise1.C_21757						
Gene Name		Avi_norm		Cot_norm	Avi/Glu	Cot/Glu
jgi Spoth1 108890		60.2250594		48.07756207	1.149068	0.9099379
estExt_fgenesh1_pg.C_60848						
jgi Spoth1 48439		9.60345542		9.188885583	0.522124	0.4955752
e_gw1.3.3367.1						
jgi Spoth1 79030		196.30114		76.95691676	1.029889	0.4005124
estExt_Genewise1Plus.C_31624						
jgi Spoth1 116270		12.6960936		8.860711098	0.757282	0.5242718
estExt_fgenesh1_pm.C_50266						
jgi Spoth1 84164		3379.11414		2343.822173	7.680355	5.2844987
estExt_Genewise1Plus.C_62100						
jgi Spoth1 102977		88.8726553		52.8360921	0.361589	0.213245
fgenesh1_pm.5.#_763						
jgi Spoth1 84305		216.484673		389.8712883	1.17595	2.1007958
estExt_Genewise1Plus.C_70023						
jgi Spoth1 114107		3649.80137		1768.696388	9.983526	4.7991986
estExt_fgenesh1_pm.C_20669						
jgi Spoth1 112305		860.567268		922.006216	0.852192	0.905706
estExt_fgenesh1_kg.C_60263						
jgi Spoth1 43941		11.5567006		26.08987157	1.731707	3.8780488
e_gw1.2.4209.1						
jgi Spoth1 62521		251.968627		105.1799225	1.937422	0.8022528
estExt_Genewise1.C_21757						

In order to narrow down the identity of each predicted transporter's substrate, strains containing deletion mutations of NCU05853 or NCU08114 were cultured on glucose, xylose, cellobiose, xylan and Avicel (Table 12). The culturing medium contained Vogel's medium plus 2% of the carbon source. Both mutants showed greatly reduced growth on Avicel but not on xylan, glucose, xylose, or cellobiose.

Gene Name	Growth on Sucrose	Growth on Avicel	Growth on Mis	Growth on Xylan	Growth on Glucose	Growth on Xylose	Growth on Cellobiose
NCU00801	***	***	***				
NCU00988	***	***	***				
NCU01231	***	***	***				
NCU04963	***	***	***				
NCU05519	***	***	***				
NCU05853	***	*	**	***	***	***	***
NCU05897	***	*	**				
NCU06138	***	***	***				
NCU08114	***	*	**	***	***	***	***
NCU10021	No deletion strain						
wt	***	***	***	***	***	***	***

To investigate the role of these transporters in utilization of hemicellulose, the expression of the ten transporter genes was examined when *Neurospora* was grown on xylan. Methods were used as described in Example 4, except that strains were grown on Vogel's salts with 2% (w/v) xylan. Expression of all ten transporters was upregulated during growth on xylan (Table 13), suggesting that they can transport sugars derived from hemicellulose degradation (e.g., xylobiose, xylose, arabinose, xylo-oligosaccharides) as well as from cellulose degradation (e.g., cellobiose, glucose, cello-oligosaccharides). The mutant growth results and expression analyses suggested that at least two of the predicted transporters, NCU05853 and NCU08114, can transport disaccharides (cellobiose, xylobiose) and/or oligosaccharides (cellooligosaccharides).

Gene Name	wt-Xylan 4 h	Fold change in St-Avicel-4 h/Glucose-4 h
NCU00801	~6	10
NCU00988.2	31.1	NO CHANGE
NCU01231.2	732.1	NO CHANGE
NCU04963.2	96.5	NO DETECT
NCU05519.2	3.9	NO CHANGE
NCU05853.2	71.2	8.5
NCU05897.2	122.3	NO CHANGE
NCU06138.2	141.0	NO CHANGE
NCU08114.2	10.0	11
NCU10021.2	44.7	NO CHANGE

Example 8

Expression Analysis of *Sporotrichum thermophile* Homologs of *N. crassa* Transporters During Growth on Various Carbon Sources

In order to compare the expression of homologous genes from a different filamentous fungus, the expression profile of *Sporotrichum thermophile* was analyzed from cultures grown on glucose, Avicel, or cotton. cDNA was isolated from cultures grown on minimal media with a carbon source of glucose, Avicel, or cotton for 16-30 hours.

First, in order to identify homologs of *Neurospora* transporter proteins in the *S. thermophile* genome, each *Neurospora* sequence was compared against a database of *S. thermophile* proteins with BLAST. The sequences of *S. thermophile* proteins found by this method were then compared to a database of *Neurospora* proteins with BLAST. These results are listed in FIG. 10. The amino acid sequences for all of the *S. thermophile* homologs of putative *Neurospora* transporters that were identified can be found in SEQ ID NOS: 22-32.

Next, the expression profile of the *S. thermophile* homologs was examined. The data is presented in Table 11. 20 The first column contains the *S. thermophile* gene name from the Joint Genome Institute *S. thermophile* assembly. The second column contains the NCU number for the most closely related putative transporter in *Neurospora*. The third column contains the gene length of the *S. thermophile* gene in nucleotides. The fourth to sixth columns contain the expression level (number of reads, comparable to absolute expression level) during growth on Vogel's minimal media supplemented with 2% of glucose, Avicel, or cotton balls as the carbon source. The seventh to ninth columns contain the normalized expression data (the # of reads divided by the total reads in the dataset). The final two columns contain the relative expression level data for each gene as a ratio of Avicel/glucose or cotton/glucose. Homologs of NCU5853, NCU8114, and NCU0801 were upregulated when grown on both Avicel and cotton. The homolog of NCU6138 was upregulated when grown on cotton, and the homolog of NCU4963 was upregulated when grown on Avicel. These 25 data provided further support that putative transporters NCU5853, NCU8114, NCU0801, NCU6138, and NCU4963 30 are important for the utilization of cellulose. 35

Example 9

Identification and Analysis of Cellooligosaccharide Transporters

When grown on pure cellulose, *N. crassa* was shown to increase transcription of seven Major Facilitator Superfamily sugar transporters as well as an intracellular β -glucosidase 50 (Ex. 1; also see Supplemental Data, Dataset S1, page 6 in Tian et al., *PNAS*, 2009). Notably, knockout strains lacking individual transporters from this set grew more slowly on crystalline cellulose, suggesting that they may play a direct role in cello-oligosaccharide uptake under cellulolytic conditions 55 (Ex. 7; Tables 10, 12). For example, deletion of NCU08114 resulted in severely retarded *N. crassa* growth (FIG. 11), and reduced *N. crassa* consumption of cellobiose (FIGS. 12-13). In this example, transporter genes NCU00801/cbt1 and 60 NCU08114/cbt2 were further analyzed and identified to encode transporters of cellooligosaccharides.

To assay the function of each transporter individually, the fact that cellobiose is not catabolized by *S. cerevisiae* and is not accumulated in its cytoplasm was exploited (FIG. 14). It was reasoned that expression of a functional cellobiose transporter in conjunction with an intracellular β -glucosidase would allow *S. cerevisiae* to grow when cellobiose is presented as the sole carbon source. Yeast strains were engi-

neered to express the transporters NCU00801 or NCU08114 fused to Green Fluorescent Protein (GFP), and the putative intracellular β -glucosidase, NCU00130. Both transporters were expressed and localized correctly to the plasma membrane (FIG. 15). The strains expressing NCU00801 or NCU08114 allowed yeast to grow with specific growth rates of 0.0341 hr^{-1} and 0.0131 hr^{-1} , respectively (FIG. 16A). These growth rates correspond to 30% and 12% of the growth rate on glucose, respectively (FIG. 17). Growth could not be explained by the extracellular hydrolysis of cellobiose to glucose followed by transport, as a strain expressing only the putative intracellular β -glucosidase grew at a rate of 0.0026 hr^{-1} (FIG. 16A), and did not grow in large-scale cultures (FIG. 18). Based on these observations, NCU00801 and NCU08114, which were named CBT1 and CBT2, were determined to function as cellobiose transporters.

To directly assay transporter function, the uptake of [^3H]-cellobiose into yeast cells was measured. Both CBT1 and CBT2 were found to be high-affinity cellobiose transporters, with K_m values of $4.0 \pm 0.3 \mu\text{M}$ and $3.2 \pm 0.2 \mu\text{M}$, respectively (FIG. 19). The expression-normalized V_{max} of CBT1 was 2.2 times that of CBT2, a fact that explained differences seen in the yeast growth assays. Notably, celldextrin molecules longer than cellobiose supported the growth of yeast expressing cbt1 and cbt2 (FIG. 20; FIG. 16B), suggesting that celldextrin molecules are transported by CBT1 and CBT2. In agreement, cellobiose transport by CBT1 and CBT2 was inhibited by excess cellotriose, and CBT1 activity was also inhibited by cellotetraose (FIG. 21). Furthermore, upon purification, the β -glucosidase, NCU00130 (FIG. 22), was found to hydrolyze cellobiose, cellotriose, and cellotetraose (FIG. 16C).

Orthologs of cbt1 and cbt2 were identified and found to be widely distributed in the fungal kingdom (FIG. 23). Recent expression data shows their importance to various interactions between fungi and plants. For example, when the ascomycete, *Tuber melanosporum*, or the basidiomycete, *Laccaria bicolor*, interacts symbiotically with root tips to form ectomycorrhizas, the ortholog of cbt1 is upregulated in both (Martin et al., 2010). Likewise, the saprophytes, *Aspergillus oryzae* (Noguchi et al., 2009), *Postia placenta* (Vanden Wymelenberg et al., 2010), and *Phanerochaete chrysosporium* (Vanden Wymelenberg et al., 2010), upregulate orthologs of cbt2 when in contact with plant wall material. Certain yeasts, such as *Kluveromyces lactis* and *Pichia stipitis* grow on cellobiose (Freer, 1991; Preez et al., 1986), and cellobiose transport has been reported in *Clavispora lusitaniae* (Freer and Greene 1990). It was determined in this study that all of these yeasts contain orthologs of cbt1, cbt2, or both (see below for methods). Cellobiose transport has been observed in *Hypocrea jecorina* (*Trichoderma reesei*), but since the transporter was not identified, it is not clear if this activity can be ascribed to orthologs of cbt1 or cbt2 (Kubicek et al., 1993).

The use of cellobiose transporters by cellulolytic fungi suggests that they are essential for their optimal growth on cellulose. To test whether cellobiose catabolism could improve yeast ethanol production, the yeast strains constructed above were grown under fermentation conditions. With little optimization, yeast with a complete cellobiose catabolism pathway ported from *N. crassa* were shown to ferment cellobiose to ethanol efficiently (FIG. 24A), with an ethanol yield of 0.47, 86% of the theoretical value (Bai et al., 2008). This was comparable to industrial yields from glucose of 90-93% (Basso et al., 2008). The high affinity of CBT1 and CBT2 for cellobiose compared to the hexose transporters of *S. cerevisiae* (Reifenberger et al., 1997), and reported extra-

cellular β -glucosidases (Chauve et al., 2010), suggested that a cellobiose/celldextrin transport system would be particularly useful during SSF. For example, cellobiose/celldextrin transport would lower the requirement for full hydrolysis of cellulose to glucose, decrease cellobiose-mediated inhibition of cellulolytic enzymes, and reduce the risk of contamination by glucose-dependent organisms. Indeed, yeasts expressing a cellobiose/celldextrin transport system markedly improved the efficiency of SSF reactions by reducing the steady state concentration of both cellobiose and glucose, and increasing the ethanol production rate (FIG. 24B, C).

Biofuel production from cellulose requires efficient and economical depolymerization of plant biomass to sugars coordinated with fuel production by improved host strains (Kumar et al., 2008). Here it was shown that cellulolytic fungi use cello-oligosaccharide transport pathways for optimal growth on plant biomass. Furthermore, reconstitution of these pathways in yeast revealed that they can be ported in a modular fashion to improve cellobiose catabolism, with a minimal pathway composed of a transporter and an intracellular cellobiose hydrolase (FIG. 25). The use of celldextrin transport in biofuel-producing strains of yeast and other organisms is critical for making cellulosic biofuel processes more economically viable.

25 Transporter and β -Glucosidase Orthologs

GenBank accession numbers or Joint Genome Institute (JGI) protein ID (PID) numbers for celldextrin transporters are as follows: *Tuber melanosporum*, CAZ81962.1; *Pichia stipitis*, ABN65648.2; *Laccaria bicolor*, EDR07962; *Aspergillus oryzae*, BAE58341.1; *Phanerochaete chrysosporium*, PID 136620 (JGI) (Martinez et al., 2004); *Postia placenta*, PID 115604 (JGI) (Martinez et al., 2009). The GenBank accession number for *Saccharomyces cerevisiae* HXT1 and *Kluveromyces lactis* LACP are DAA06789.1 and CAA30053.1, respectively. The *P. chrysosporium* and *P. placenta* genomes can be accessed at genome.jgi-psf.org/Phchr1/Phchr1.home.html and genome.jgi-psf.org/Pospl1/Pospl1.home.html, respectively.

GenBank accession numbers for celldextrin hydrolases 40 that are orthologs of NCU00130 are as follows: *T. melanosporum*, CAZ82985.1; *A. oryzae*, BAE57671.1; *P. placenta*, EED81359.1; and *P. chrysosporium*, BAE87009.1. The other 45 organisms that contain celldextrin transporter orthologs contain genes in the GH3 family predicted to be intracellular β -glucosidases (Bendtsen et al., 2004; Cantarel et al., 2009), as follows: *Kluveromyces lactis*, CAG99696.1; *Laccaria bicolor*, EDR09330; *Clavispora lusitaniae*, EEQ37997.1; and *Pichia stipitis*, ABN67130.1.

50 Strains and Media

The yeast strain used in this study was YPH499 (Sikorski et al., 1989), which has the genotype: MAT α ura3-52 lys2-801_amber ade2-101_ochre trp1-Δ63 his3-Δ200 leu2-Δ1. It was grown in YPD media supplemented to 100 mg/L adenine hemisulfate. Transformed strains (Becker et al., 2001) were 55 grown in the appropriate complete minimal dropout media, supplemented to 100 mg/L adenine hemisulfate. *Neurospora crassa* stains used in this study were obtained from the Fungal Genetics Stock Center (McCluskey 2004) and include WT (FGSC 2489) and two cellobiose transporter deletion strains (FGSC 16575, ΔNCU00801.2 and FGSC 17868, ΔNCU08114.2 (Colot et al., 2006)).

Plasmids and Cloning

Transporters were cloned into the 2 μ plasmid, pRS426, which was modified to include the *S. cerevisiae* PGK1 promoter inserted between SacI and SpeI using the primers, ATATATGAGCTCGTGAGTAAGGAAGAGTGAGGA-65 ACTATC (SEQ ID NO: 53) and ATATAT

ACTAGTTGTTTATATTGTTGAAAAAGTAGATAAT
TACTTCC (SEQ ID NO: 54). (In all primers above and below, restriction sites are underlined). NCU00801 with a C-terminal Myc-tag and optimized Kozak sequence (Miyasaka 1999) was then inserted between BamHI and EcoRI using the primers, AT GGATCCAAAATGTCGTCACGGCTCC (SEQ ID NO: 55) and ATGAATTCTACAAATCTTCTCAGAA-ATCAATTGTTGTCAGCAACGATAGCTTCGGAC (SEQ ID NO: 56), and NCU08114 with a C-terminal Myc-tag and optimized Kozak sequence was inserted between SpeI and Clal using the primers, AT ACTAGTAAAATGGGCATCTCAACAAGAACG (SEQ ID NO: 57) and GCATATCGATCTACAAATCTTCTT-CAGAAATCAATTGTTGTCAGCAACA-GACTTGCCCTCAT G (SEQ ID NO: 58). To make GFP fusions, superfolder GFP (Pedelacq et al., 2006) with an N-terminal linker of Gly-Ser-Gly-Ser was first inserted between the Clal and Sall sites of the PGK1 promoter-containing pRS426 plasmid with the primers, TATTAA ATCGATGGTAGTGGTAGTGTGAGCAAGGGCGAGGAG (SEQ ID NO: 59) and TATTAAAGTCGACCTACTT-GTACAGCTCGTCCATGCC (SEQ ID NO: 60). Transporters were then fused to GFP as follows: NCU00801 was inserted between BamHI and EcoRI using the primers, GCAT GGATCCATGTCGTCACGGCTCC (SEQ ID NO: 61) and TATAATGAATTCTCAGCAACGATAGCTTCGGAC (SEQ ID NO: 62), and NCU08114 was inserted between SpeI and EcoRI using the primers, TATTAA ACTAGTATGGGCATCTCAACAAGAACG (SEQ ID NO: 63) and TTATAAGAATTCTCAGCAACAGAC-TTGCCTCATG (SEQ ID NO: 64).

The β -glucosidase, NCU00130, was cloned into the 2μ plasmid, pRS425, modified to include the PGK1 promoter described above. NCU00130 with an optimized Kozak sequence and a C-terminal 6xHis tag was inserted between SphI and PstI using the primers, GCAT ACTAGTAAAAATGTCTCTCCTAAGGATTCCTCT (SEQ ID NO: 65) and ATACTGCAGTTAATGATGA-TGATGATGATGGTCCTCTTGATCAAAGAGTCAAAG (SEQ ID NO: 66). All constructs included the Cyc transcriptional terminator between XbaI and KpnI. All *N. crassa* genes were amplified by PCR from cDNA synthesized from mRNA isolated from *N. crassa* (FGSC 2489) cultured on minimal media with pure cellulose (Avicel) as the sole carbon source.

Yeast Growth Assays

To monitor growth on cello-oligosaccharides, engineered strains were grown in 5 mL of complete minimal media with appropriate dropouts overnight. These starter cultures were washed three times with 25 mL of ddH₂O, and resuspended to an OD (at 600 nm) of 0.1 in Yeast Nitrogen Base (YNB) plus the appropriate Complete Supplemental Media (CSM) and 1% (w/v) of cellobiose, or 0.5% (w/v) of either cellotriose or cellotetraose. Assays were performed in a Bioscreen CT™ with constant shaking at maximum amplitude at 30° C. and a final assay volume of 0.4 mL. The change in OD was measured either at 600 nm or using a wideband filter from 450–580 nm. Growth rates were taken from the linear portion of each growth curve, and are reported as the mean of three independent experiments±the standard deviation between these experiments. Cellotriose and cellotetraose were obtained from Seikagaku Biobusiness Corporation (Tokyo, Japan).

Purification of NCU00130 and Assay of its Activity

A 1 L culture of *S. cerevisiae* expressing cbt1 and NCU00130 was grown to an OD of 2.0 in complete minimal media. Cells were harvested by centrifugation and resus-

pended in 30 mL of lysis buffer (50 mM Na₂HPO₄ [pH 8.0], 300 mM NaCl, 10 mM imidazole, 2 mM β-ME, Complete™ Mini, EDTA free protease inhibitor cocktail). Cells were lysed by sonication, and the lysate was cleared by centrifugation at 15,000 g for 30 minutes. The lysate was bound to 1 mL of nickel-NTA resin by gravity flow, and washed three times with 25 mL wash buffer (identical to lysis buffer but with 20 mM imidazole). NCU00130 was eluted with 5 mL of elution buffer (identical to lysis buffer but with 250 mM imidazole), and the appropriate fractions were pooled, exchanged into storage buffer (Phosphate Buffered Saline (PBS), 2 mM DTT, 10% glycerol), aliquoted, frozen in liquid nitrogen, and stored at -80° C. Purity was determined by SDS-PAGE (FIG. 22), and protein concentration was determined from the absorbance at 280 nm, using an extinction coefficient of 108,750 M⁻¹cm⁻¹.

Purified NCU00130 was assayed from hydrolysis activity with different celldextrin substrates. Activity was measured by incubating 5 pmol of enzyme with 500 μ M of each sugar in 150 μ L PBS plus 3 mM DTT. Reactions proceeded for 40 minutes at 30° C. before 100 μ L was removed and quenched in 400 μ L of 0.1 M NaOH. The results were analyzed by ion chromatography with a Dionex ICS-3000, with CarboPac PA200 column. Peaks were detected with an electrochemical detector.

Phylogenetic Analysis of Transporter Orthologs

Amino acid sequences of orthologs of CBT1 and CBT2 were obtained from online databases. Multiple sequence alignments were performed using T-Coffee (Notredame et al., 2000). A maximum likelihood phylogeny was determined using PhyML version 3.0 (Guindon and Gascuel 2003) with 100 Bootstraps. Both programs were accessed through Phylogeny.fr (webpage phylogeny.fr/). The resulting tree was visualized with FigTree v.1.2.1 (webpage tree.bio.ed.ac.uk/).

Fermentation and SSF

In fermentation and SSF experiments, comparisons were made between yeast expressing NCU00130 and either Myc-tagged cbt1, or no transporter. These strains were grown aerobically overnight in complete minimal media, washed three times with 25 mL water, and resuspended to a final OD₆₀₀ of 2.0 in 50 mL YNB plus the appropriate CSM, and either 2% (w/v) cellobiose or 3% (w/v) pure cellulose (Avicel), in sealed serum flasks. The SSF reactions also included 50 Filter Paper Units/g cellulose of filter-sterilized Celluclast (Sigma C2730), without β-glucosidase supplementation. Reactions were carried out anaerobically at 30° C. with shaking. At indicated time points, 1 mL samples were removed and filtered through a 0.2 μm syringe filter. The ethanol, glucose, and cellobiose concentration in the filtrate was determined by HPLC with an Aminex HPX-87H column and refractive index detection.

N. crassa Growth and Alamar Blue® Assays

WT *N. crassa* (FGSC 2489), and the homokaryotic NCU08114 (FGSC 17868) (Colot et al., 2006) were acquired from the Fungal Genetics Research Center (McCluskey 2003), and grown at 25° C. in 50 mL of Vogel's salts plus 2% of either sucrose or pure cellulose (Avicel) in a 250 mL unbaffled flask. After 16 or 28 hours, respectively, 100 µL of Alamar Blue® was added, and cultures were incubated at room temperature for 20 minutes. At this time, 1 mL samples were removed, debris pelleted, and the fluorescence of 100 µL of the supernatant determined with excitation/emission wavelengths of 535/595 nm in a Beckman Coulter Paradigm plate reader.

N. crassa Cellobiose Transport Assays

WT *N. crassa* (FGSC 2489), and homokaryotic deletion lines (Colot et al., 2006) of NCU00801 (FGSC 16575) and NCU08114 (FGSC 17868) were acquired from the Fungal Genetics Stock Center (McCluskey 2003), and grown for 16 hours in 50 mL of Vogel's salts plus 2% (w/v) sucrose at 25° C., starting with an inoculum of 10⁶ conidia/mL. Mycelia were harvested by centrifugation, washed three times with Vogel's salts, and transferred to Vogel's salts plus 0.5% (w/v) pure cellulose (Avicel) for 4 hours to induce the transporter expression. Ten mL of the culture was harvested by centrifugation, washed three times with Vogel's salts, and resuspended in 1 mL ddH₂O plus cycloheximide (100 µg/mL) and 90 µM of the respective celldextrin (cellobiose, cellotriose, or cellotetraose). To measure celldextrin consumption, 100 µL was removed after 15 minutes, clarified by centrifugation, and transferred into 900 µL of 0.1 M NaOH. The amount of sugar remaining in the supernatant was determined by HPLC with a Dionex ICS-3000, using a CarboPac PA200 column. Peaks were detected with an electrochemical detector.

GFP Fluorescence and Confocal Fluorescence Microscopy

Bulk-cell GFP fluorescence measurements were made in a Beckman Coulter Paradigm plate reader with excitation/emission wavelengths of 485/535 nm. Confocal fluorescence microscopy was performed with cells at an OD (at 600 nm) of 0.8-1.2, using a 100×1.4 NA oil immersion objective on a Leica SD6000 microscope attached to a Yokogawa CSU-X1 spinning disc head with a 488 nm laser and controlled by Metamorph software. Z series were recorded with a 200 nm step size and analyzed using ImageJ.

[³H] Cellobiose Transport Assays and Kinetic Parameters

Transport assays were performed using a modification of the oil-stop method (Arendt et al., 2007). Yeast strains expressing either cbt1 or cbt2 fused to GFP were grown to an OD (at 600 nm) of 1.5-3.0 in selective media, washed three times with ice cold assay buffer (30 mM MES-NaOH [pH 5.6] and 50 mM ethanol), and resuspended to an OD of 20. To start transport reactions, 50 µL of cells were added to 50 µL of [³H] cellobiose layered over 100 µL of silicone oil (Sigma 85419). Reactions were stopped by spinning cells through oil for 1 minute at 17,000 g, tubes were frozen in ethanol/dry ice, and tube-bottoms containing the cell-pellets were clipped off into 1 mL of 0.5 M NaOH. The pellets were solubilized overnight, 5 mL of Ultima Gold scintillation fluid added, and CPM determined in a Tri-Carb 2900TR scintillation counter. [³H] cellobiose was purchased from Moravek Biochemicals, Inc. and had a specific activity of 4 Ci/mmol and a purity of >99%. Kinetic parameters were determined by measuring the linear rate of [³H] cellobiose uptake over 3 minutes for a range of cellobiose concentrations. V_{max} and K_m values were determined by fitting a single rectangular, 2-parameter hyperbolic function to a plot of rates vs. cellobiose concentration by non-linear regression in SigmaPlot®. V_{max} values were normalized for differences in transporter abundance by measuring the GFP fluorescence from 100 µL of cells at OD 20 immediately before beginning transport assays. Kinetic parameters reported in the text are mean±the standard deviation from three separate experiments. Competition assays were performed by measuring transport of 50 µM [³H]-cellobiose over 20 seconds in the percent of 250 µM of the respective competitors.

Large Scale Yeast Growth

To monitor growth on different carbon sources, engineered strains were grown in 5 mL of complete minimal media with appropriate dropouts overnight. These starter cultures were washed three times with 25 mL of ddH₂O and resuspended to

an OD (at 600 nm) of 0.1 in 50 mL Yeast Nitrogen Base (YNB) plus the appropriate Complete Supplemental Media (CSM) and 2% (w/v) cellobiose. Cultures were grown in 250 mL unbaffled flasks at 30° C., with shaking at 200 rpm. The change in OD (at 600 nm) was monitored by periodically removing samples.

Example 10

Identification of Critical Residues for Celldextrin Transporter Function

In this example, sequence analysis and mutagenesis studies were used to identify conserved and functionally important residues in the celldextrin transporters. In addition, additional celldextrin transporters were identified.

The growth rates of yeast strains expressing various mutants of the celldextrin transporter NCU00801 (cbt1) or NCU08114 (cbt2) and the wild-type β-glucosidase NCU00130 were grown with cellobiose as the sole carbon source. Amino acid residues at 96 positions of NCU00801 and at 96 positions of NCU08114 were individually mutated to alanine using QuickChange® II Site-directed Mutagenesis Kit (Stratagene, La Jolla, Calif.) as per the manufacturer's instructions. Strains were grown in synthetic defined media-ura-leu 100 mg/L adenine with 2% cellobiose. Cultures were started from two independent colonies.

As the results shown in FIG. 26 (a, b) indicate, mutant strains that expressed NCU00801 with substitutions at W66, L73, Y74, N87, Y89, D90, Q104, F107, G113, F120, Y123, D139, G142, K144, M147, G150, Q169, F170, G173, R174, G178, G180, P189, Y191, E194, P198, R201, Y208, W235, R236, Q242, ²⁵⁷PESPRF²⁶² (SEQ ID NO: 67), Y279, G283, E296, D307, K308, W310, D312, R325, G336, Y345, N369, D385, F462, P468, E476, T480, or G486 showed at least a 25% growth defect compared to wild-type strain.

The alanine scanning experiment on NCU08114 indicated the following residues as being functionally important: L38, Y39, G54, D56, F73, G91, P100, D104, G107, R108, M118, R139, F144, Q150, P154, E159, P163, H165, R166, Y173, N174, W199, Q214, ²²²PESP²²⁵ (SEQ ID NO: 68), Y244, H245, D249, E258, E268, Q302, W303, S304, N306, Y312, F359, L360, F402, Y403, S404, Y414, E417, P420, Y421, K426, N442, N446, P447, W459, K460, E482, T483, L488, E489, E490, D496, and G497 (FIG. 26b).

In particular, the motifs ⁷³LYF⁷⁵, ²⁵⁷PESP²⁶⁰ (SEQ ID NO: 69), and ²⁷⁸KYH²⁸⁰ (residue numbering of NCU00801) appeared to be functionally important in both transporters (residues ²⁵⁷PESP²⁶⁰ (SEQ ID NO: 69) of NCU00801 and residues ²²²PESP²²⁵ (SEQ ID NO: 68) of NCU08114), which have an amino acid sequence identity of 29% (FIG. 26b, c). Several residues that are conserved in transporters in general (italicized in FIG. 26b, c), or in β-linked transporters in particular (double-underlined), were experimentally shown to be important for transporter function (underlined), e.g., D90 (NCU00801) and D56 (NCU08114), and L73 (NCU00801) and L38 (NCU08114). Results of the mutagenesis experiment also implicated residues conserved in the NCU00801/NCU08114 clade (capped) as being functionally important, e.g., Q168 (NCU00801) and Q214 (NCU08114). Moreover, multiple residues determined to be functionally important in this experiment were previously shown to be conserved in the *S. cerevisiae* sugar transporters (Hxt1/Hxt3), e.g., L73 (NCU00801) and L38 (NCU08114).

Orthologs of *N. crassa* celldextrin transporters from different organisms were also studied (FIG. 27). Representative orthologs were synthesized by Genescript and cloned into the

expression vector, pRS426 containing the Cup1 promoter using the sites BamHI and HindIII. These constructs were transformed into the yeast strain, YPH499 along with the intracellular β -glucosidase, NCU00130. Transporter activity was determined by measuring the growth rates of these strains when cellobiose was present as the sole carbon source.

Alternatively, different fungal strains containing putative orthologs were cultivated in rich media supplemented with cellobiose. Total RNA was isolated and reverse transcribed into cDNA. Polymerase chain reaction (PCR) was used to amplify the putative transporter genes directly from cDNA. However, because the regulation mechanism and expression pattern were unknown for cellobextrin transporters in fungal species, cDNAs encoding the putative transporters were not always obtainable despite alteration of cultivation condition. In this case, primers were designed according to the corresponding cDNA sequences from GenBank and used to amplify the exons using genomic DNA as a template. Overlap-extension PCR was then used to assemble the exons into the full-length genes. The resulting PCR products were cloned into the pRS424 shuttle vector containing a HXT7 promoter and a HXT7 terminator using the DNA assembler method. Yeast plasmids isolated from transformants were retransformed into *E. coli* DH5 α , and isolated *E. coli* plas-

mids were first checked by diagnostic PCR using the primers used to amplify the original transporter genes. The entire open reading frames were submitted for sequencing to confirm the correct construction of the plasmids. In the orthologs LAC2, LAC3, HXT2.1, and HXT2.6 from *P. stipitis*, one or more alternative codons (CUG) substitute Ser for Leu. Most of the cloning work was carried out using the yeast homologous recombination mediated DNA assembler method. pRS424-HXT7-GFP plasmid was used for cloning of putative cellobextrin transporters. In this plasmid, the HXT7 promoter, the GFP gene flanked with the EcoRI sites at both ends, and the HXT7 terminator were assembled into the pRS424 shuttle vector (New England Biolabs) linearized by Clal and BamHI. PCR products of the putative transporters flanked with DNA fragments sharing sequence identity to the HXT7 promoter and terminator were co-transferred into CEN.PK2-1C with EcoRI digested pRS424-HXT7-GFP using the standard lithium acetate method. The resulting transformation mixture was plated on SC-Trp plates supplemented with 2% D-glucose to recover transformants. Yeast expressing putative cellobextrin transporter orthologs and NCU00130 were tested for growth on cellobiose as the sole carbon source.

A listing of the putative cellobextrin transporter orthologs and results obtained from the study are shown in Table 14.

TABLE 14

Listing of putative cellobextrin transporter orthologs and summary of results.

<i>N. crassa</i> ortholog	Species	NCBI Reference Sequence/ NCBI GI Number/JGI number ¶	Aver. Growth Rate	Growth Rate error	Seq results*
NCU00809	<i>Chaetomium globosum</i> CBS148.51	XP_001220480	—	—	OK
NCU00809	<i>Podospora anserina</i>	XP_001912722	—	—	—
NCU00809	<i>Nectria haematococca</i> mpVI77-13-4	EEU41662	—	—	—
NCU00809	<i>Aspergillus nidulans</i> FGSC A4	XP_660803	—	—	1 intron and 50 bp insertion
NCU00809	<i>Aspergillus terreus</i> NIH2624	XP_001218592	—	—	—
NCU00809	<i>Talaromyces stipitatus</i> ATCC 10500	XP_002341594	—	—	—
NCU00809	<i>Aspergillus niger</i>	XP_001395979	—	—	Ala > Val
NCU00809	<i>Aspergillus fumigatus</i> Af293	XP_747891	—	—	—
NCU00809	<i>Aspergillus terreus</i> NIH2624	XP_00120996	—	—	—
NCU00809	<i>Aspergillus oryzae</i> RIB40	XP_001817400	—	—	OK
NCU08114	<i>Podospora anserina</i>	XP_001908539	—	—	N/A
NCU08114	<i>Penicillium chrysogenum</i> Wisconsin 54-1255	XP_002568019	—	—	N/A
NCU08114	<i>Aspergillus terreus</i> NIH2624	XP_001209810	—	—	Wrong
NCU08114	<i>Aspergillus oryzae</i> RIB40	XP_001820343	—	—	OK
NCU08114	<i>Aspergillus terreus</i> NIH2624	XP_001210859	—	—	N/A
NCU08114	<i>Neurospora crassa</i> OR74A	XP_001728155	—	—	N/A
NCU08114	<i>Aspergillus oryzae</i> RIB40	XP_001826848	—	—	N/A
NCU08114	<i>Aspergillus nidulans</i> FGSC A4	XP_657617	—	—	OK
NCU08114	<i>Talaromyces stipitatus</i> ATCC 10500	XP_002487579	—	—	N/A
NCU08114	<i>Chaetomium globosum</i> CBS 148.51	XP_001227497	—	—	Wrong
NCU08114	<i>Trichoderma atroviridae</i>	215408	0.000836364	0.00064871	I, D
NCU08114	<i>Chaetomium globosum</i>	XP_001220290.1	0.004036364	0.00047168	OK
NCU08114	<i>Aspergillus nidulans</i>	ANID_08347	0.011109091	0.000072727	Other
NCU08114	<i>Pleurotus ostreatus</i>	51322	0.00390303	0.00018212	—

TABLE 14-continued

Listing of putative cellobextrin transporter orthologs and summary of results.

<i>N. crassa</i> ortholog	Species	NCBI Reference Sequence/ NCBI GI Number/JGI number [¥]	Aver. Growth Rate	Growth Rate error	Seq results*
NCU08114	<i>Sporotrichum thermophile</i>	114107	0.009569697	0.00216366	—
NCU00801	<i>Aspergillus nidulans</i>	XP_660418.1	0.000860606	0.000438	P
NCU00801	<i>Magnaporthe grisea</i>	XP_364883.1	0.005090909	0.00138313	OK
NCU00801	<i>Aspergillus fumigatus</i>	XP_753099.1	0.003975758	0.00211951	OK
NCU00801	<i>Trichoderma atroviridiae</i>	211304	0.002678788	0.00031193	D
NCU00801	<i>Chaetomium globosum</i>	XP_001220469.1	0.005890909	0.00010285	OK
NCU00801	<i>Tremella mesenterica</i>	63529	0.004381818	0.00115751	D
NCU00801	<i>Heterobasidion annosum</i>	105952	0.002751515	0.00068763	D
NCU00801	<i>Cryphonectria parasitica</i>	252427	0.02250303	0.00021692	D
NCU00801	<i>Trichoderma reesei</i>	67752	0.003672727	0.00066233	D
NCU00801	<i>Aspergillus clavatus</i>	XP_001268541.1	0.014381818	0.00059613	OK
NCU00801	<i>Neurospora discreta</i>	77429	0.007060606	0.00110566	D
NCU00801	<i>Trichoderma reesei</i>	3405	0.003264646	0.001033998	D
NCU00801	<i>Sporotrichum thermophile</i>	43941	0.013654545	0.00431534	—
NCU00801	<i>Neurospora crassa</i>	XP_963801.1	0.048754872	0.00354017	—
NCU05853	<i>Chaetomium globosum</i>	XP_001226269.1	0.003593939	0.00062306	OK
NCU05853	<i>Trichoderma reesei</i>	46819	0.002042424	0.000085924	D
NCU05853	<i>Mycosphaerella graminicola</i>	68287	0.00290101	0.00060123	D
NCU05853	<i>Aspergillus flavus</i>	AFLA_000820A	0.003078788	0.00209132	—
—	None	—	0.0026	0.0001	—
NCU00809	<i>Pichia stipitis</i> CBS6054 (LAC1)	XP_001383110.1/ GI: 126133170	See FIG. 27	—	—
NCU00809	<i>Pichia stipitis</i> CBS6054 (LAC2)	XP_001387231.1/ GI: 126276337	See FIG. 27	—	—
NCU00809	<i>Pichia stipitis</i> CBS6054 (LAC3)	XP_001383677.2/ GI: 150864727	See FIG. 27	—	—
NCU08114	<i>Pichia stipitis</i> CBS6054 (HXT2.1)	XP_001386873.1/ GI: 126275571	See FIG. 27	—	—
NCU05853	<i>Pichia stipitis</i> CBS6054 (HXT2.3)	XP_001382754.1/ GI: 126132458	See FIG. 27	—	—
NCU08114	<i>Pichia stipitis</i> CBS6054 (HXT2.4)	XP_001387757.1/ GI: 126273939	See FIG. 27	—	—
NCU08114	<i>Pichia stipitis</i> CBS6054 (HXT2.5)	XP_001385684.1/ GI: 126138322	See FIG. 27	—	—
NCU08114	<i>Pichia stipitis</i> CBS6054 (HXT2.6)	XP_001384653.2/ GI: 15086543	See FIG. 27	—	—

*Wrong = difference between tested sequence and sequence in NCBI or JGI databases; I = insertion in tested sequence; D = deletion in tested sequence; P = point mutation in tested sequence; OK = no difference between tested sequence and sequence deposited in NCBI or JGI databases; Other = other problems in sequencing, excluding insertion, deletion, and point mutations in tested sequence; “—” = results not yet available (study in progress).

[¥] When accession numbers were not available, the JGI number was used. The JGI number allows access to the gene sequence via the JGI genome portal for this organism (accessible from the following page: genome.jgi-psf.org/programs/fungi/index.jsf). The *A. flavus* and *A. nidulans* identifiers allow access to the genes through their genome portals at webpage.cadre-genomes.org.uk/ and webpage.broadinsti-tute.org/annotation/genome/aspergillus_group/MultiHome.html, respectively.

In certain cases, the sequences of the cloned orthologs were determined to be correct, and the yeast expressing those clones were able to utilize cellobiose. Thus, these clones, LAC2 from *Pichia stipitis* and XP_001268541.1 from *Aspergillus clavatus* were confirmed to be functional cellobiose transporters. Testing of the cellobiose transporting function of other clones is still in progress. Cloned orthologs with sequences different from the published sequences in databases (e.g., ones with insertions, deletions, etc.) (Table 14) will be re-cloned, re-sequences, and similarly tested for cellobiose transport activity by expressing them in *S. cerevisiae* and monitoring growth rates.

An alignment of NCU00801, NUC08114, and functional orthologs of these transporters is shown in FIG. 28. The alignment in FIG. 28a includes both putative and confirmed cellobextrin transporters, whereas the alignment in FIG. 28b includes only confirmed cellobextrin transporters. In addition, FIG. 28c shows an alignment of NCU00801 and NCU08114. The two transporters share 29% amino acid sequence identity.

50 Motifs critical for cellobextrin transporter function were identified by visual inspection of multiple sequence alignments between sugar transporters. Specifically, motifs common to cellobextrin transporters were identified from multiple sequence alignments produced in T-COFFEE between putative cellobextrin transporter orthologs and confirmed cellobextrin transporters. To ensure that these motifs were largely unique to cellobextrin transporters, their absence was confirmed from a multiple sequence alignment between the hexose transporters of *S. cerevisiae*, the human glucose transporter, Glut1, and two *N. crassa* monosaccharide transporters produced in T-COFFEE.

60 The identified motifs are described below. In the motifs, residues that were found to be critical to the function of NCU00801 are underlined. The residues that were critical for the function of NCU08114 are marked with the superscript “†”. The residues that were critical to the function of both transporters are marked with the superscript “*”. All motifs were defined using PROSITE notation. As an example of how

101

to read a PROSITE motif, the following motif, [AC]-x-V-x(4)-{ED}, is translated as: [Ala or Cys]-any-Val-any-any-any-any-{any but Glu or Asp} (SEQ ID NO: 13)

Cellobextrin transporters, like all sugar transporters, have 12 transmembrane α -helices. The N- and C-terminus of cellobextrin transporters are both intracellular.

The sequence before transmembrane helix 1 had no distinguishing features.

Transmembrane helix 1 contained the motif, [L*¹⁰IVM]-Y*--[FL]-x(13)-[YF]-D* (SEQ ID NO: 1).

Transmembrane helix 2 contained the motif, [YF]x(2)-G † -x(5)-[PVF]-x(6)-[DQ]* (SEQ ID NO: 2).

The loop connecting transmembrane helix 2 and transmembrane helix 3 contained the motif, G*-R † -[RK]* (SEQ ID NO: 3).

Transmembrane helix 3 had no distinguishing features.

Transmembrane helix 4 had no distinguishing features.

Transmembrane helix 5 contained the motif, R*-x(6)-[YF]*-N † (SEQ ID NO: 4).

Transmembrane helix 6 contained the motif, W*-R-[IVLA]-P-x(3)-Q (SEQ ID NO: 5).

The sequence between transmembrane helix 6 and transmembrane helix 7 contained the motif, P*-E-S*-P*-R-x-L-x(8)-A-x(3)-L-x(2)-Y*-H † (SEQ ID NO: 6).

Transmembrane helix 7 contained the motif, F † -[GST]Q*-x-S † -G-N † -x-[LIV] (SEQ ID NO: 7).

Transmembrane helix 8 had no distinguishing features.

Transmembrane helix 9 had no distinguishing features.

Transmembrane helix 10 and transmembrane helix 11 and the sequence between them contained the motif, L-x(3)-[YIV] † -x(2)-E*-x-L-x(4)-R-[GA]K † -G (SEQ ID NO: 8).

Transmembrane helix 12 had no distinguishing features.

The sequence after transmembrane helix 12 had no distinguishing features.

Homology models of NCU00801 and NCU08114 were produced from the primary amino acid sequences of NCU00801 and NCU08114 using the I-TASSER server at:

102

zhanglab.ccmb.med.umich.edu/I-TASSER/ (Roy et al., 2010). The top structural models produced by I-TASSER were visualized in PYMOL ([webpage pymol.org/](http://www.pymol.org/)). Mapping of the motifs was also performed in PYMOL. The homology models of NCU00801 and NCU08114 with the cellobextrin transporter motifs marked are shown in FIG. 29 (a, b). FIG. 29 (c) shows the predicted secondary structures of NCU00801 and NCU08114.

Example 11

Characterization of Novel Pentose-Specific Transporters from *Neurospora crassa* and *Pichia stipitis* in *Saccharomyces cerevisiae*

In this example, a bioinformatics approach was taken to identify novel pentose-specific transporters in *N. crassa* and *P. stipitis*.

Genome Mining of Pentose-Specific Transporters Bioinformatics Study

To discover novel D-xylose-specific transporters, the genes encoding the D-glucose/D-xylose symporter Gxs1 from *C. intermedia* (Leandro et al., 2006) and the uncharacterized putative L-arabinose-proton symporter Aut1 from *P. stipitis* (locus tag PICST_87108) were used as probes in BLAST searches ([webpage ncbi.nlm.nih.gov/](http://www.ncbi.nlm.nih.gov/)) against the sequenced genomes of two efficient xylose-utilizing species, *N. crassa* and *P. stipitis* (Galagan et al., 2003; Jeffries et al., 2007). Any proteins with known D-glucose transport activity or activity other than sugar transport were eliminated from the analyses. Using a cut-off of 25% minimal sequence identity, 17 putative pentose transporter genes were identified (Table 15), in addition to AUT1 from *P. stipitis*. These putative pentose transporter genes shared 25-50% identity with either GXS1 from *C. intermedia* or AUT1 from *P. stipitis*. All 17 putative pentose transporters were annotated as either sugar-transport proteins or hypothetical proteins with unknown activity. The D-glucose transporter genes SUT1 and SUT2 from *P. stipitis* were also cloned for comparison.

35

a. BLAST search results using AUT1 as a probe.

Name	Origin	% identity with AUT1	Annotation from NCBI	Length (cDNA)	Locus Tag
Ap31/SUT2	<i>P. stipitis</i>	31	sugar uptake (tentative)	1653	ABN66266
Ap26/XP_001387242	<i>P. stipitis</i>	26	sugar transporter	1404	XP001387242
AN49/NCU01494	<i>N. crassa</i>	49	hypothetical protein NCU01494, similar to MFS sugar transporter	2025	EAA26691
AN41/NCU09287	<i>N. crassa</i>	41	hypothetical protein NCU09287, similar to galactose-proton symporter	1968	EAA28903
AN29-2/NCU04963	<i>N. crassa</i>	29	hypothetical protein NCU04963, similar to MFS monosaccharide transporter	1584	EAA30175
AN28-3/NCU02188	<i>N. crassa</i>	28	hypothetical protein NCU02188, conserved hypothetical protein	1458	EAA30346
AN25/NCU00821	<i>N. crassa</i>	25	sugar transporter	1689	EAA35128

b. BLAST search results using GSX1 as a probe.

Name	Origin	% identity with GSX1	Annotation from NCBI	Length (cDNA)	Locus Tag
Xy50/NCU04537	<i>N. crassa</i>	50	hypothetical protein NCU04537 similar to monosaccharide transporter	1626	EAA26741

-continued

Xy31/NCU06138	<i>N. crassa</i>	31	hypothetical protein NCU06138, similar to MFS monosaccharide transporter	1757	EAA30764
Xy33/NCU00988	<i>N. crassa</i>	33	hypothetical protein NCU00988, similar to MFS quinate transporter	1614	EAA34662
Xyp37/SUT3	<i>P. stipitis</i>	37	sugar uptake (tentative)	1653	ABN67990
Xyp33/XUT3	<i>P. stipitis</i>	33	sugar transporter; putative xylose uptake (tentative); predicted transporter (major facilitator superfamily)	1656	EAZ63115
Xyp32/XUT1	<i>P. stipitis</i>	32	sugar transporter, high affinity, putative; xylose uptake (tentative)	1701	ABN67554
Xyp30/STL1	<i>P. stipitis</i>	30	sugar transporter, strongly conserved	1590	ABN65745
Xyp31/XUT2	<i>P. stipitis</i>	31	sugar transporter, xylose transporter (tentative) similarly to GXSI (STL1)	1407	AAVQOIOOOO02
Xyp29/STL12/ XUT6	<i>P. stipitis</i>	29	sugar transporter, putative (STL12); .xylose uptake (tentative)	1641	ABN68560
Xyp30-1/HGT3	<i>P. stipitis</i>	30	high affinity xylose transporter (putative), xylose uptake (tentative)	1587	ABN68686
Xyp28/XUT7	<i>P. stipitis</i>	28	xylose transporter, high affinity, putative similarity to STL13, high affinity sugar transporters	1257	EAZ63044

Cloning of Putative Pentose Transporters

N. crassa and *P. stipitis* were cultivated in rich media supplemented with either D-xylose or L-arabinose as carbon sources. Total RNA was isolated and reverse transcribed into cDNA. Polymerase chain reaction (PCR) was used to amplify the putative transporter genes directly from cDNA. However, because the regulatory mechanism and expression patterns of pentose transporters in fungal species were unknown, cDNAs encoding the putative pentose transporters were not always obtainable despite alteration of cultivation conditions. In those cases, primers were designed according to the corresponding cDNA sequences from GenBank and used to amplify the exons with genomic DNA as templates. Overlap-extension PCR was then used to assemble the exons into full length genes. The resulting PCR products were cloned into the pRS424-HXT7-GFP shuttle vector using the yeast homologous recombination-mediated DNA assembler method (Shao et al., 2009). In this plasmid, an HXT7 promoter, a GFP gene flanked with the EcoRI sites at both ends, and an HXT7 terminator were assembled into the pRS424 shuttle vector (New England Biolabs) linearized by ClaI and BamHI. PCR products of the putative pentose transporters flanked with DNA fragments, sharing sequence identity with the HXT7 promoter and terminator (FIG. 30a) were co-transferred into *S. cerevisiae* CEN.PK2-1C strain (MAT α leu2-3, 112 ura3-52, trp1-289, his3-Δ1 MAL2-8c) purchased from Euroscarf (Frankfurt, Germany) with EcoRI digested pRS424-HXT7-GFP using the standard lithium acetate method. The resulting transformation mixture was plated on SC-Trp plates supplemented with 2% D-glucose.

Yeast plasmids isolated from transformants using Zymo-prep Yeast Plasmid Miniprep II (Zymo Research, Orange, Calif.) were re-transferred into *Escherichia coli* DH5 α cells (Cell Media Facility, University of Illinois at Urbana-Champaign, Urbana, Ill.). The plasmids were isolated using the QIAprep Spin Miniprep Kit (QIAGEN, Valencia, Calif.) and then checked by diagnostic PCR with the primers used to amplify the original transporter genes. The entire open reading frames were also submitted for DNA sequencing to confirm correct construction (Core Sequencing Facility, Univer-

sity of Illinois at Urbana-Champaign, Urbana, Ill.). The DNA sequencing results were compared to gene sequences in databases using Sequencher 4.7 (Gene Codes Corporation, Ann Arbor, Mich.). All sequences of cloned putative transporters are listed in SEQ ID NOs: 33-52.

Yeast strains were cultivated in synthetic dropout media to maintain plasmids (0.17% Difco yeast nitrogen base without amino acids and ammonium sulfate, 0.5% ammonium sulfate, 0.05% amino acid drop out mix). YPA media supplemented with 2% of sugar was used to grow yeast strains harboring no plasmids (1% yeast extract, 2% peptone, 0.01% adenine hemisulfate). *S. cerevisiae* strains were cultured at 30° C. and 250 rpm for aerobic growth and at 30° C. and 100 rpm for oxygen-limited conditions. Yeast strains were grown under aerobic conditions for cell manipulation unless specified otherwise. *E. coli* strains were cultured at 37° C. and 250 rpm in Luria broth (LB) (Fisher Scientific, Pittsburgh, Pa.). All restriction enzymes were purchased from New England Biolabs (Ipswich, Mass.). All chemicals were purchased from Sigma Aldrich (St. Louis, Mo.) or Fisher Scientific.

Transporter Activity Assay for Cloned Putative Transporters

Intracellular Accumulation of Pentose Sugars

The cloned putative pentose transporters were over-expressed in an *S. cerevisiae* sugar transporter deletion strain, and uptake of pentose sugars was measured. The D-xylose-uptake ability of putative pentose transporters was determined by summation of intracellular D-xylose and xylitol concentrations. D-xylose accumulated within *S. cerevisiae* cells can be partially converted to xylitol due to the presence of endogenous aldose reductase. Both D-xylose and xylitol were extracted using osmosis and analyzed using high performance liquid chromatography (HPLC).

The sugar transporter knock-out *S. cerevisiae* strain EBY.VW4000 (CEN.PK2-1c Δhxt1-17, Δstl1, Δagt1, Δydl247w, Δyjr160c, Δgal2), which was a gift from Professor E. Boles' laboratory (Institut für Mikrobiologie, Heinrich-Heine-Universität, Universitätsstr. 1, Geb. 26.12.01, D-40225 Düsseldorf, Germany), had concurrent knock-outs of more than 20 sugar transporters and sensors including

105

HXT1-17 and GAL2. Growth on D-glucose as the sole carbon source was completely abolished in this strain, whereas uptake of maltose through a different sugar transport system was retained. The EBY.VW4000 strain also exhibited minimal pentose-uptake under HPLC assay conditions, which made it a suitable host for testing recombinant D-xylose uptake. Plasmids over-expressing the cloned putative pentose transporter genes were transferred into the EBY.VW4000 strain using the standard lithium acetate method, and single colonies were used for measuring sugar uptake activity.

Cells were first cultured in 2 mL SC-Trp medium supplemented with 2% maltose. Seed culture was then used to inoculate a 50 mL culture in a 250 mL flask. The cells were harvested by centrifugation after 24 hours of growth and re-suspended in YPA medium supplemented with 2% D-xylose or L-arabinose to a final OD₆₀₀ of 10. At 30 min, 60 min, 120 min, and 24 hours, 5 mL cultures were taken for measuring intracellular sugar concentrations. Culture samples were washed twice with ice-cold water and re-suspended in 3 mL of deionized water. Cell suspensions were incubated at 37°C. with 250 rpm agitation for 2 days to extract intracellular sugars. The resulting cell suspension was filtered through a 0.22 µm PES filter (Corning, Lowell, Mass.) before HPLC analysis. The concentrations of sugar and corresponding sugar alcohol (discussed below) were determined using Shimadzu HPLC equipped with a BioRad HPX-87C column (BioRad Laboratories, Hercules, Calif.) and Shimadzu ELSD-LTII low temperature-evaporative light scattering detector (Shimadzu) following the manufacturer's protocol. The sugar-uptake activity was calculated as mg of sugar extracted through osmosis per mL of cell culture at OD~10.

Several putative pentose transporters were identified to be active in uptake of D-glucose or D-xylose or both. Since D-glucose can be metabolized once inside yeast, the D-glucose transport activity could not be determined by measuring intracellular D-glucose concentration. However, because the EBY.VW4000 strain normally cannot grow on media containing D-glucose as the sole carbon source, growth of the strain transformed with a putative pentose transporter on D-glucose indicated that the putative transporter has D-glucose transport activity.

Introduction of SUT3 (Xyp37), XUT3 (Xyp33), SUT2 (Ap31), NCU04963 (An29-2), and NCU06138 (Xy31) restored growth of the EBY.VW4000 strain on D-glucose and, thus, enabled glucose transport activity. SUT3, XUT3, SUT2, and NCU04963 also had xylose transport activity, whereas NCU04963 and NCU06138 showed arabinose transport activity (FIG. 31). The rest of the putative transporters failed to enable growth on D-glucose, and most of them also did not show any pentose transport activity. However, NCU00821 and STL12/XUT6 showed xylose transport activity, and XUT1 exhibited arabinose transport activity, indicating they may be sugar transporters specific for pentoses (FIG. 32).

To further confirm that STL12/XUT6 and XUT1 from *P. stipitis* and NCU00821 from *N. crassa* were actually pentose-specific transporters with no D-glucose-uptake activity, the sugar-uptake assay was performed using ¹⁴C-labeled D-glucose, D-xylose, and L-arabinose as substrates. It was found that D-glucose- and L-arabinose-uptake activities of the EBY.VW4000 strain over-expressing only STL12/XUT6 and NCU00821 were too low to be measured under assay conditions used to determine D-xylose-uptake kinetics of both transporters.

¹⁴C-labeled D-glucose, L-arabinose, and D-xylose were purchased from American Radiolabeled Chemicals (St. Louis, Mo.) as solutions in 90% ethanol. Radiolabeled sugars were first dried in a chemical hood and then re-suspended in water.

106

Sugar solutions at concentrations of 1.33 M and 1 M with specific radioactivity of approximately 40,000 dpm/µL, and at concentrations of 500 mM, 350 mM, 250 mM, 100 mM, and 50 mM with specific radioactivity of about 20,000 dpm/µL were used for the sugar-uptake assay. Cell culture at the exponential phase was harvested and washed twice with ice-cold water and re-suspended to about 60 mg dry cell weight (DCW) per mL in 100 mM Tris-citrate buffer at pH 5. Three aliquots of 160 µL cell suspension were dried at 65°C. for 24 hours to determine the DCW. The rest of the cell suspension was kept on ice before use. For the sugar-uptake assay, cell suspension was equilibrated at 30°C. for 5 min before the assay. In a 50 mL conical tube, 160 µL of cell suspension was mixed with 40 µL of radio-labeled sugar solution for 40 or 60 seconds (accurately timed). The reaction was stopped by adding 10 mL of ice-cold water delivered by a syringe. The zero-time-point sample was obtained by adding ice-cold water and cell suspension simultaneously in a culture tube containing the radio-labeled solution. The mixture was then filtered immediately through a Whatman GF/C filter (Whatman, Florham Park, N.J.) pre-soaked in 40% sugar solution and washed with 15 mL of ice-cold water. The filter was placed in 3 mL of Econo I scintillation cocktail (Fisher Scientific) and counted using a Beckman LS6500 scintillation counter (Beckman Coulter, Brea, Calif.) for 1 min. All data points were measured in three independent experiments. The sugar-uptake rate was calculated as mmol sugar transported per hour per gram of dry cell weight.

Intracellular accumulation of both D-xylose and L-arabinose in EBY.VW4000 strains over-expressing STL12/XUT6, NCU00821, or XUT1 was also measured using HPLC. Cell cultures incubated with pentose sugars for 30 min, 60 min, 120 min, and 24 hours were analyzed by HPLC. The EBY.VW4000 strains over-expressing STL12/XUT6 or NCU00821 exhibited D-xylose uptake activity, whereas the strain over-expressing XUT1 exhibited L-arabinose-uptake activity after a 24-hour incubation (FIG. 33).

The ¹⁴C-labeled sugar uptake assay together with HPLC analysis of intracellular sugar accumulations confirmed that among the three most abundant monosaccharides in lignocellulosic hydrolysates, D-glucose, D-xylose, and L-arabinose, STL12/XUT6 and NCU00821 were responsible for D-xylose uptake and XUT1 was responsible for L-arabinose uptake. Of note, most sugar transporters studied in yeast for D-xylose uptake have higher uptake activity towards D-glucose than towards D-xylose. Only Trxlt1 from *Trichoderma reesei* after adaptive evolution exhibited D-xylose-specific uptake activity (Saloheimo et al., 2007). This data indicated that STL12/XUT6 from *P. stipitis*, NCU00821 from *N. crassa* are the first two experimentally confirmed naturally-occurring D-xylose-specific transporters introduced into *S. cerevisiae*. Similarly, XUT1 from *P. stipitis* is the first experimentally confirmed naturally-occurring L-arabinose-specific transporter introduced into *S. cerevisiae*.

55 Kinetic Parameters

Using the ¹⁴C-labeled sugar-uptake assay, kinetic parameters of D-xylose transport through NCU00921, STL12/XUT6, and XUT1 were determined. It was observed that under the assay conditions, sugar uptake was within a linear range for the first 60 seconds (FIG. 34). The EBY.VW4000 strains over-expressing NCU00821, STL12/XUT6, or XUT1 were incubated with labeled D-xylose or L-arabinose for 40 or 60 seconds followed by addition of ice-cold water to stop further sugar uptake. The reaction mixture was then filtered and washed before measurement using a liquid scintillation counter. The sugar-uptake rates and substrate concentrations were fitted into a Michaelis-Menten equation by non-linear

107

regression using the Origin software (OriginLab Corporation, Northampton, Me.). The K_m values for D-xylose uptake by the EBY.VW4000 strain harboring only NCU00821 or STL12/XUT6 were 175.7 ± 21.4 mM and 56.0 ± 9.4 mM, respectively. The corresponding V_{max} values were 36.7 ± 2.9 and 41.5 ± 2.3 $\mu\text{mol/h/gram DCW}$, respectively. Similarly, the K_m and V_{max} values for L-arabinose uptake by the EBY.VW4000 strain harboring XUT1 were 48.0 ± 13.2 mM and 5.6 ± 1.6 $\mu\text{mol/h/gram DCW}$ respectively.

In naturally-occurring D-xylose-assimilating fungal species, both the high affinity D-xylose-proton symport system and the low affinity D-xylose facilitated diffusion system are present. The K_m values of these two systems were determined to be 0.4-4 mM for the symport system and around 140 mM for the facilitated diffusion system (Leandro et al., 2006; Stambuk et al., 2003). These values are close to the affinity of the D-glucose-uptake system in *S. cerevisiae*, which has a K_m of 1.5 mM for the high affinity system and 20 mM for the low affinity system (Lang and Cirillo 1987; Ramos et al., 1988). Unfortunately, the D-xylose uptake affinity of wild-type *S. cerevisiae* is two orders of magnitude lower than its affinity for D-glucose. The K_m values for D-xylose uptake in *S. cerevisiae* are only 190 mM for the high affinity system and 1.5 M for the low affinity system (Kötter and Ciriacy, 1993). The affinities of the newly discovered D-xylose-specific transporters were lower when compared to the high affinity D-xylose-uptake system in naturally occurring D-xylose-assimilating yeasts. However, compared to the D-xylose-uptake system in wild-type *S. cerevisiae*, NCU00821 and STL12/XUT6 showed higher affinity towards D-xylose. In particular, the K_m of D-xylose uptake by STL12/XUT6 and XUT1 were only one-fourth of the K_m of xylose uptake by the transporter in wild-type *S. cerevisiae*. The K_m values of the D-xylose-specific transporters were also close to those of Gxf1 (K_m 88 mM) and Sut1 (K_m 145 mM), which have been shown to improve D-xylose fermentation in recombinant *S. cerevisiae* (Runquist et al., 2009; Katahira et al., 2008). Thus, D-xylose fermentation may be improved by introducing these newly discovered D-xylose-specific transporters into *S. cerevisiae*.

Cellular Localization of Sugar Transporters

Sugar transporters are transmembrane proteins, and correct folding and localization in the cell membrane is required for them to be functional. Since no signal peptide was specifically added when the putative pentose transporters were cloned, it was important to ensure that the D-xylose-specific transporters were correctly localized to the cell membrane. This was particularly true for putative pentose transporters like NCU00821 cloned from the filamentous fungi *N. crassa*, which exhibits a very different physiology compared to *S. cerevisiae*. To study the cellular localization of D-xylose-specific transporters in *S. cerevisiae*, NCU00821, STL12/XUT6, and XUT1 were fused with Green Fluorescent Protein (GFP) at the C-termini via linkers, and their localization was monitored by fluorescent imaging.

The fusion proteins of the pentose-specific transporters with the GFP at the C-terminus were constructed for the transporter localization study. A GS-linker (Gly-Gly-Gly-Gly-Ser-Gly-Gly-Gly-Ser (SEQ ID NO: 70)) was introduced between the transporter and the GFP. The GS-linker was added to the N-terminus of the GFP open reading frame by a PCR primer, resulting in a PCR product of GS-linker-GFP flanked with nucleotide sequence homologous to the transporters at the 5'-end and the HXT7 terminator at the 3'-end. Transporter genes were amplified from the original pRS424-HXT7-transporter constructs to generate DNA fragments of the transporters flanked with nucleotide sequence identical to the HXT7 promoter at the 5'-end and GS-linker-

108

GFP at the 3'-end. These two fragments were then co-transferred into the *S. cerevisiae* strain CEN.PK2-1C with pRS424-HXT7-GFP digested with EcoRI (FIG. 30b). The resulting transformation mixture was plated on SC-Trp plates supplemented with 2% D-glucose.

Single colonies were inoculated into 2 mL of SC-Trp liquid medium supplemented with 2% maltose. Cell culture was harvested at the exponential phase. In a centrifuge tube, 250 μL of cell culture was stained with 10 μL Hoechst 33342 nuclei dye (Invitrogen, Carlsbad, Calif.) for 10 minutes at room temperature. A small droplet of cell culture was then transferred onto a piece of cover glass and fluorescent images were taken using an Andor Technology Revolution System Spinning Disk Confocal Microscope (Core facilities, Institute for Genomic Biology, University of Illinois at Urbana-Champaign, Urbana, Ill.). Images were processed using Imaaris image analysis and visualization software (Bitplane, Saint Paul, Minn.).

Yeast strains over-expressing pentose-specific transporters showed a distinctive fluorescent halo at the cell periphery (FIG. 35). For NCU00821 and XUT1, almost all the GFP fluorescence appeared in the cell membrane, while a large portion of fluorescence in STL12/XUT6-over-expressing cells remained in the cytoplasm. This could indicate inefficient export of the STL12/XUT6 transporter due to elevated expression of the membrane protein. It was also noticed that not all the cells showed fluorescence, indicating that expression of the transporter was not optimal. Further improvements of transporter expression can be achieved through altering the expression level and/or integrating the transporter genes into the genome of recombinant *S. cerevisiae*.

Determination of the Type of Pentose Transporters

There are two types of sugar transporters in *S. cerevisiae*, symporters and facilitators. For symporters, sugar uptake is coupled to proton uptake. Sugar symporters usually exhibit high affinity towards sugar. Meanwhile, sugar uptake through facilitators is not coupled to proton transport, and facilitators usually exhibit low sugar-uptake affinities (Leandro et al., 2006). Symporter assays were performed for NCU00821, STL12/XUT6, and XUT1 expressed in the EBY.VW4000 strain.

To determine the type of transporters, pH change of the EBY.VW4000 over-expressing pentose-specific transporters was measured in un-buffered cell suspension containing D-xylose, L-arabinose, or maltose using a Seven Multi pH meter equipped with an USB communication module and Direct pH software (Mettler Toledo, Columbus, Ohio). Plasmids encoding pentose-specific transporters were transferred into EBY.VW4000 strain followed by plating on the SC-Trp plates supplemented with 2% maltose. Single colonies were inoculated in 2 mL SC-Trp medium supplemented with 2% maltose. Seed culture was then used to inoculate a 400 mL culture in 2 L flasks. The culture was harvested at OD~1 and washed twice with ice-cold water. Cell pellets were re-suspended in 4 mL of water and kept on ice before use. For the symporter assay, the pH electrode was immersed in a water-jacketed beaker of 50 mL capacity kept at 25°C. and provided with magnetic stirring. To the beaker, 23 mL of deionized water and 1 mL of cell suspension equilibrated at 25°C. was added. The pH was adjusted to 5, and a base line was obtained. The pH change was recorded with addition of 1 mL of 50% sugar solution at pH 5.

FIG. 36 shows pH changes in un-buffered cell suspension after the addition of maltose. As was reported, pH in un-buffered *S. cerevisiae* cell suspension went up with the addition of maltose. One mL of 50% maltose solution was added to the un-buffered cell suspension to ensure that the pH

109

recording system was functional. The pH elevations observed in all samples indicated that the pH recording system could monitor transient pH changes in the experimental setting.

No elevation of pH in un-buffered cell suspensions was observed for any of the pentose-specific transporters, indicating that pentose uptake through these transporters is not coupled with proton transport (FIG. 37). Thus, NCU00821, STL12/XUT6, and XUT1 were determined to be pentose facilitators.

This result was consistent with the fact that the kinetic parameters of NCU00821 and STL12/XUT6 were similar to those of the low affinity D-xylose facilitated diffusion system in naturally-occurring D-xylose-assimilating yeasts. Despite the fact that symporters have higher affinities towards D-xylose, over-expression of symporters may not always facilitate sugar utilization by D-xylose-assimilating strains due to the ATP requirement to create the proton gradient. In fact, most of the transporters shown to be beneficial for D-xylose fermentation are facilitators (Runquist et al., 2009; Katahira et al., 2008).

Heterologous Over-Expression of D-Xylose-Specific Transporters

The over-expression of active heterologous D-xylose-specific transporters in *S. cerevisiae* strains containing the D-xylose utilization pathway was also investigated to determine whether their over-expression could improve xylose utilization. Xylose utilization was studied using a shake-flask under aerobic conditions. Plasmids expressing the xylose transporters NCU00821, NCU04963, XUT1, STL12/XUT6, and Hxt7 were introduced into strain HZE63 (CEN.PK2 ura3::xylose utilization pathway). This strain had a xylose utilization pathway integrated into the URA3 site onto the chromosome. It was constructed using a plasmid from previous work that contained xylulose reductase (XR) and xylitol dehydrogenase (XDH) from *N. crassa* and xylulokinase (XKS) from *P. stipitis*. This plasmid was digested with Apal and transformed into yeast strain CEN.PK2 to yield the strain HZE63.

The HZE63 strain transformed with the xylose transporter-encoding plasmids was selected by plating on SC-Ura plates supplemented with 2% glucose. The transformed strain was pre-cultured in SC-Trp-Ura with 2% glucose and then inoculated into SC-Trp-Ura supplemented with 0.5% or 5% of xylose to an initial OD₆₀₀=1.0. Cell cultures were grown in a 125 mL shake-flask containing 50 mL of culture at 30°C. and 250 rpm (FIG. 38).

Yeast plasmids of transformants were transformed into *E. coli* DH5α cells. The plasmids were then isolated and checked by diagnostic PCR and submitted for sequencing to confirm correct construction. Plasmid maps can be found in FIG. 39.

Unfortunately, the advantage of pentose-specific transporter over-expression could not be observed despite alteration of expression strategies, cultivation conditions, and choice of the D-xylose utilization pathway. There are several possible reasons. Firstly, the over-expression of membrane proteins, such as sugar transporters, could affect the integrity of the cell membrane and consequently hamper cell growth (Wagner et al., 2006). It was observed that transporter over-expression strains displayed a slower growth rate even when D-glucose was used as a carbon source. The final OD of 2-day cultures of strains carrying transporters grown in glucose-containing SC-ura media was only 4, whereas the OD of the negative control was around 6. Secondly, the D-xylose-uptake activity of the wild-type *S. cerevisiae* through hexose transporters is much higher than the D-xylose-uptake activity of a certain D-xylose transporter over-expressed in a hexose transporter knockout strain. The low sugar transport activity of

110

newly discovered D-xylose-specific transporters may make it hard to observe the improvement of sugar uptake ability. Thirdly, even if the introduction of new D-xylose-specific transporters could improve the uptake of D-xylose into *S. cerevisiae* cells, the benefit of D-xylose utilization can only be observed when the D-xylose utilization pathway is efficient enough to make sugar-uptake the limiting step. It was shown that the effect of over-expression of sugar transporters depends on the strain background and cultivation conditions (Runquist et al., 2010). Examples 12-15 below describe the optimization of the xylose utilization pathway in yeast.

Cloning of Additional Pentose-Specific Transporters

Orthologs of NCU00821, STL12/XUT6, and XUT1 were cloned and tested for pentose uptake. Different fungal strains were cultivated in rich media supplemented with glucose or pentoses. Total RNA was isolated and reverse transcribed into cDNA. Polymerase chain reaction (PCR) was used to amplify the putative transporter genes directly from cDNA. However, because the regulation mechanism and expression pattern were unknown for pentose transporters in fungal species, cDNAs encoding the putative pentose transporters were not always obtainable despite alteration of cultivation condition. In this case, primers were designed according to the corresponding cDNA sequences from GenBank and used to amplify the exons using genomic DNA as a template. Overlap-extension PCR was then used to assemble the exons into the full-length genes. The resulting PCR products were cloned into the pRS424 shuttle vector containing a HXT7 promoter and a HXT7 terminator using the DNA assembler method. Yeast plasmids isolated from transformants were retransformed into *E. coli* DH5α, and isolated *E. coli* plasmids were first checked by diagnostic PCR using the primers used to amplify the original transporter genes. The entire open reading frames were submitted for sequencing to confirm the correct construction of the plasmids.

Most of the cloning work was carried out using the yeast homologous recombination mediated DNA assembler method. pRS424-HXT7-GFP plasmid was used for cloning of putative pentose transporters. In this plasmid, the HXT7 promoter, the GFP gene flanked with the EcoRI sites at both ends, and the HXT7 terminator were assembled into the pRS424 shuttle vector (New England Biolabs) linearized by ClaI and BamHI. PCR products of the putative pentose transporters flanked with DNA fragments sharing sequence identity to the HXT7 promoter and terminator were co-transferred into CEN.PK2-1C with EcoRI digested pRS424-HXT7-GFP using the standard lithium acetate method. The resulting transformation mixture was plated on SC-Trp plates supplemented with 2% D-glucose. Transformants were then tested for pentose transport activity.

The results are shown below in FIG. 40 and Table 16. Among the eight putative pentose specific transporters [XP_960000 (NC52), CAG88709 (DH48), XP_457508 (DH61), XP_681669 (32-10), XP_001487429 (29-6), XP_001727326 (29-9), XP_657854 (32-8), XP_720384 (29-4)], only NC52 enabled cell growth on a glucose plate, which suggested that the other seven transporters may be pentose-specific or inactive. Using the HPLC-based pentose uptake assay, four xylose-specific transporters were found, including XP_457508 (DH61), XP_001727326 (29-9), XP_720384 (29-4), and XP_681669 (32-10). In addition, one arabinose-specific transporter, XP_657854 (32-8) was identified (FIG. 40; Top). Five additional putative pentose specific transporters (XP_002488227, AB070824.1, XP_001389300, XP_002488227, EEQ43601.1) were also tested, none of which enabled cell growth in a glucose plate. Further pentose uptake assays indicated that XP_002488227 and AB070824.1 were xylose specific transporters (FIG. 40; Bottom). The summary of these results are shown in Table 16D.

TABLE 16A

Cloning of xylose-specific transporter NCU00821 orthologs				
NCBI Reference Sequence	Origin	Sequence Results*	Uptake Assay	Status
XP_002488227	<i>Talaromyces stipitatus</i>	Correct	Yes	Cloned
XP_001400900	<i>Aspergillus niger</i>	Correct	Yes	Cloned
XP_001220481	<i>Chaetomium globosum</i> CBS 148.51	No	No	Sequenced, one intron
XP_001912725	<i>Podospora anserina</i>	No	No	OE-PCR, no PCR product
XP_660079	<i>Aspergillus nidulans</i> FGSC A4	Correct	Yes	Cloned
AAL89823	<i>Aspergillus niger</i>	Correct	Yes	Cloned
XP_002382573	<i>Aspergillus flavus</i> NRRL3357	Wrong	Yes	Cloned
XP_459386	<i>Debaryomyces hansenii</i> CBS767	No	No	Genomic DNA, no PCR product
XP_001825132	<i>Aspergillus oryzae</i> RIB40	Correct	Yes	Cloned
XP_001389300	<i>Aspergillus niger</i>	Correct	Yes	Cloned

*“Correct” = Sequence of clone matched sequence in database(s); “Wrong” = Sequence of clone did not match sequence in database(s); “No” = Results not available (work in progress)

TABLE 16B

Cloning of xylose-specific transporter STL12/XUT6 orthologs				
NCBI Reference Sequence	Origin	Sequence Results*	Uptake Assay	Status
XP_457508 (DH61)	<i>Debaryomyces hansenii</i> CBS767	Correct	No	Cloned
XP_002551364	<i>Candida tropicalis</i> MYA-3404	Wrong	No	No
XP_001523322	<i>Lodderomyces elongisporus</i> NRRL	Wrong	No	No
XP_720384 (29-4)	<i>Candida albicans</i> SC5314	Correct	No	Cloned
XP_456868	<i>Debaryomyces hansenii</i> CBS767	Wrong	No	No
XP_001487429 (29-6)	<i>Pichia guilliermondii</i> ATCC 6260	Wrong	No	Cloned
XP_961039 CAG88709 (DH48)	<i>Neurospora crassa</i>	Wrong	No	No
XP_001727326 (29-9)	<i>Debaryomyces hansenii</i> CBS767	Correct	No	Cloned
XP_001816757	<i>Aspergillus oryzae</i>	Correct	No	Cloned

*“Correct” = Sequence of clone matched sequence in database(s); “Wrong” = Sequence of clone did not match sequence in database(s); “No” = Results not available (work in progress)

TABLE 16C

Cloning of arabinose-specific transporter XUT1 orthologs				
NCBI Reference Sequence	Origin	Sequence Results*	Uptake Assay	Status
XP_002545773	<i>Candida tropicalis</i> MYA-3404	Correct	Yes	Cloned
EEQ43601	<i>Candida albicans</i> WO-1	Correct	Yes	Cloned
XP_001818631	<i>Aspergillus oryzae</i> RIB40	No	No	No PCR product
XP_002558275	<i>Penicillium chrysogenum</i> Wisconsin 54-1255	Wrong	Yes	Cloned
XP_001390883	<i>Aspergillus niger</i>	No	No	No PCR product
XP_750103	<i>Aspergillus fumigatus</i> Af293	Wrong	No	No
XP_960000 (NC52)	<i>Neurospora crassa</i> OR74A	Wrong	No	Cloned
XP_657854 (32-8)	<i>Aspergillus nidulans</i> FGSC A4	Correct	No	Cloned

25

30

35

40

45

50

60

65

TABLE 16C-continued

Cloning of arabinose-specific transporter XUT1 orthologs				
NCBI Reference Sequence	Origin	Sequence Results*	Uptake Assay	Status
XP_001825068	<i>Aspergillus oryzae</i> RIB40	Correct	No	Cloned
XP_681669 (32-10)	<i>Aspergillus nidulans</i> FGSC	Correct	No	Cloned
**“Correct” = Sequence of clone matched sequence in database(s); “Wrong” = Sequence of clone did not match sequence in database(s) (e.g., because of mutation in clone) “No” = Results not available (work in progress)				
TABLE 16D				
Listing of new xylose-specific transporters and one arabinose-specific transporter.				
NCBI Reference Sequence	Origin	Xylose-specific	Arabinose-specific	
XP_457508 (DH61)	<i>Debaryomyces hansenii</i> CBS767	Yes		
XP_001727326 (29-9)	<i>Aspergillus oryzae</i>	Yes		
XP_720384 (29-4)	<i>Candida albicans</i> SC5314	Yes		
XP_681669 (32-10)	<i>Aspergillus nidulans</i> FGSC A4	Yes		
XP_657854 (32-8)	<i>Aspergillus nidulans</i> FGSC A4		Yes	
XP_002488227 AB070824.1	<i>Talaromyces stipitatus</i> Aspergillus oryzae	Yes	Yes	

The orthologs with sequences inconsistent with the sequences in databases (e.g., ones with mutations) will be re-cloned, sequenced, expressed in yeast strains, and tested for sugar uptake function. Similarly, the orthologs for which there is no sequencing results will also be tested for transporter function.

Sequence alignments of the pentose transporter orthologs were analyzed to identify conserved residues, which could have potential roles in transporter function. Alignments of a sample of xylose transporters (NCU00821, STL12/XUT6, XP_002488227.1, and XP_002382573.1) and arabinose transporters (XUT1 and EEQ43601.1) are shown in FIG. 41 (a, b) respectively. Several residues are specifically conserved in xylose transporters whereas others are specifically con-

served in the arabinose transporters. These residues may have critical roles in transporting the specific pentose. An overall comparison of the sequences of the xylose and arabinose transporters (FIG. 41c) shows that there are also residues that are conserved in both types of pentose transporters, indicating functional roles in uptake of pentoses in general.

Examples 12-15 relate to optimization of the xylose utilization pathway in yeast.

Example 12

Engineering Pentose-Utilizing *S. cerevisiae* Strain

An efficient xylose metabolic pathway was reconstituted by exploiting the concept of isoenzymes. Isoenzymes catalyze the same chemical reaction with different kinetic or regulatory properties, and are known to confer fine-tuned control of metabolic fluxes in response to dynamic changes in the cytosolic environment. However, no prior metabolic engineering approaches had employed isoenzymes to increase fluxes of interest. This study demonstrated that simultaneous expression of both wild-type and mutant xylulose reductase (XR) isozymes could decrease xylitol accumulation and increase the overall xylose fermentation rate.

Inspired by the prevalence of isoenzymes in living systems, wild type XR and mutant XR (R276H) were co-expressed in *S. cerevisiae* along with xylitol dehydrogenase (XDH) and xylulokinase (XK) in order to construct a functional xylose metabolic pathway in *S. cerevisiae*. The XR mutant had been reported to exhibit much lower preference for NADPH over NADH whereas wild type XR showed 116 two-fold higher preference for NADPH over NADH (Watanabe et al., 2007).

The xylose-metabolizing genes (wild-type XYL1, 2, and 3 and mutant XYL1) from *P. stipitis* were PCR-amplified and placed under the control of constitutive promoters (PGK1 and TDH3) to construct expression cassettes. These integration cassettes were integrated into the genome of the D452-2 strain.

Transformation of expression cassettes for constructing xylose metabolic pathways was performed using the yeast EZ-Transformation kit (BIO 101, Vista, Calif.). To select transformants using an amino acid auxotrophic marker, yeast synthetic complete (YSC) medium was used, which contained 6.7 g/liter yeast nitrogen base plus 20 g/liter glucose, 20 g/liter agar, and CSM-Leu-Trp-Ura (BIO 101), which supplied appropriate nucleotides and amino acids. Yeast strains were routinely cultivated at 30° C. in YP medium 234 (10 g/liter yeast extract, 20 g/liter Bacto peptone) with 20 g/liter glucose.

The effect of *S. cerevisiae* strain background on xylose-metabolizing efficiency was also tested by expressing identical constructs containing optimized xylose utilization pathway enzymes in several different yeast strains. The three laboratory strains used were D452-2 (MAT_a, leu2, his3, ura3, can1), L2612 (MAT_a, leu2-3, leu2-112, ura3-52, trp1-298, can1, cyn1, gal+), and CEN.PK. Production of xylitol, acetate, and ethanol was monitored together with use of xylose and OD₆₀₀. The results indicated that the D452-2 strain was the best amongst the three tested strains (FIG. 42-44). *S. cerevisiae* D452-2 was used for engineering of the xylose-metabolizing enzymes in yeast. Strains and plasmids used in this study are described in Table 17.

TABLE 17

Strain and plasmids used in study		
Strain or plasmid	Description	Reference
Strain		
D452-2	MAT _a , leu2, his3, ura3, can1	Hosaka et al., (1992)
D801-130	D452-2 expressing β-glucosidase (NCU00130) and cbt1 (NCU00801)	In this study
D809-130	D452-2 expressing β-glucosidase (NCU00130) and NCU00809	In this study
D8114-130	D452-2 expressing β-glucosidase (NCU00130) and cbt2 (NCU08114)	In this study
DA24	D452-2 expressing XYL1, mXYL1, XYL2, and XKs1 (Isogenic of D452-2 except for leu2::TDH3P-XYL1-TDH3T, ura3::URA3-PGKP-mXYL1-PGKT-PGKP-XYL2-PGKT, Ty3::neo-TDHP-XKS1-TDHT)	In this study
DA24-16	Evolved strain of DA24 in xylose containing media	In this study
DA24-16BT3	DA24-16 expressing β-glucosidase (NCU00130) in a multi-copy plasmid and cbt1 (NCU00801) though single-copy integration	In this study
DA24-16BT-M	DA24-16 expressing β-glucosidase (NCU00130) and cbt1 (NCU00801) in multi-copy plasmids	In this study
Plasmid		
pRS425	LEU2, a multi copy plasmid	Christianson et al., (1992)
pRS426	URA3, a multi copy plasmid	Christianson et al., (1992)
pRS403	HIS3, an integrative plasmid	Sikorski et al., (1989)
pRS405	URA3, an integrative plasmid	Sikorski et al., (1989)
pRS425-β-glucosidase	β-glucosidase (NCU00130) under the control of PGK promoter in pRS425	Submitted
pRS426-cbt1	cbt1 under the control of PGK promoter in pRS426	Submitted

TABLE 17-continued

Strain and plasmids used in study		
Strain or plasmid	Description	Reference
pRS426-cbt2	cbt2 under the control of PGK promoter in pRS426	Submitted
pRS426-NCU00809	NCU00809 under the control of PGK promoter in pRS426	Submitted
pRS403-cbt1	cbt1 under the control of PGK promoter in pRS403	In this study

The engineered xylose-fermenting *S. cerevisiae* strain (DA24) consumed xylose and produced ethanol with negligible amounts of xylitol accumulation. When 40 and 80 g/L of xylose were used as a sole carbon source, the DA24 strain produced ethanol with consistent yields ($Y_{Ethanol/Xylose} = 0.31\text{--}0.32 \text{ g/g}$) in both shaker-flask and bioreactor fermentation experiments (FIG. 45). However, the DA24 strain consumed xylose slower than the naturally existing xylose-fermenting yeast, *P. stipitis*. Xylose fermentation capability of DA24 was further improved using an evolutionary engineering approach (Sauer 2001). One of the strains (DA24-16) isolated after repeated sub-cultures of the DA24 on xylose-containing medium showed much faster xylose fermentation rates as compared to the parental strain under various culture conditions (Table 18).

Carbon source	Strains	Sugar consumption			
		Produced Ethanol (g/L)	rate (g/L/h)	Yield (g/g)	Productivity (g/L · h)
Xylose (80 g/L)	DA24	24	1.16	0.34	0.40
	DA24-16	28	1.32	0.35	0.47
Glucose (70 g/L) and xylose (40 g/L)	DA24	34	1.45	0.39	0.74
	DA24-16	45	1.78	0.42	0.96

Interestingly, the DA24-16 strain consumed xylose as fast as *P. stipitis*, the fastest xylose-fermenting yeast known. However, ethanol yield by DA24-16 was slightly lower than that by *P. stipitis* (FIG. 46).

A screen was set up using *S. cerevisiae* strain L2612 expressing the xylose-utilizing enzymes (strain YSX3) transformed with a genomic library. Transformation was followed by serial culture transfer in 40 g/L xylose under oxygen-limiting conditions to enrich for strains that are efficient in utilizing xylose. Fermentations were performed in 50 mL YPX media under oxygen-limited conditions and 0.1% (50 µL) of a fully grown cell culture was transferred to the next serial culture when $OD_{600}=10$ was reached. After 10 serial cultures, cells were spread with serial dilution on YPX (40 g/L) agar media. Through fermentation experiments using 5 mL of YPX media, colonies were screened for low xylitol and high ethanol formation. DNA sequencing revealed that the two most efficient strains contained integrated copies of XYL2, which was then cloned into a multi-copy plasmid through homologous recombination and transformed into YSX3 cells.

The XYL2 gene was placed in integration vectors under the control of promoters of different strength, e.g., TDH or PGKp, and transformed into YSX3 cells (FIG. 47). Studies were conducted to monitor the effect of these plasmids on xylitol and ethanol formation in the transformed yeast cells.

The results indicated that the YSX3 cells expressing higher levels of XYL2 (under the PGKp) were more efficient at ethanol production and in addition, produced lower amounts of xylitol (FIG. 48). When additional XYL3 was expressed in these cells (termed SR1 strain), the amount of xylitol produced was further decreased in the resulting strain SRu-23 (FIG. 49). Therefore, it appeared that XYL2 expression level in engineered *S. cerevisiae* strains is a key factor for implementing xylose fermentation, and when expression is under a strong promoter, the strain has less xylitol accumulation as well as high ethanol yield. Simultaneous over-expression of XYL2 and XYL3 can further decrease the amount of xylitol accumulation. However, when XYL1 was further over-expressed in a strain over-expressing XYL2 and XYL3, there was considerable xylitol accumulation and consequently decreased xylose fermentation (FIGS. 50-51). Therefore, it appeared that there was an optimal level of XYL1 for efficient xylose fermentation.

Experiments were also carried out to test if over-expression of endogenous GRE3 in *S. cerevisiae* expressing XYL2 and XYL3 could facilitate xylose fermentation. For the construction of pRS403-GRE3, GRE3 gene was amplified from *S. cerevisiae* D452-2 and inserted into pR403 vector with TDH3 promoter and CYC terminator. After linearization of pRS403-GRE3, it was integrated into the genome of D452-2. The xylose-utilizing genes were introduced into the yeast strain D452-2 (FIG. 52), and xylose fermentation parameters were monitored. The results indicated that over-expression of GRE3 was as effective as the over-expression of XYL1 in ethanol production and xylitol accumulation, particularly when cells were grown in 80 g/L of xylose at high OD inoculations (FIGS. 53-54).

Example 13

Engineering LAD and XDH

L-arabinitol and xylitol accumulation, thought to be caused by cofactor imbalance between NADPH-dependent XR and NAD⁺-dependent XDH and LAD, has been regarded as a major bottleneck during xylose fermentation in engineered *S. cerevisiae* expressing the pentose-utilizing enzymes. While the imbalance between XR and XDH has been corrected by engineering enzymes with reversed cofactor preferences (Watanabe et al., 2007; Matsushika et al., 2008; Bengtsson et al., 2009), this approach resulted in reduced flux, as the modified enzymes had reduced specific activities. The *P. stipitis* XR mutant had been reported to exhibit much lower preference for NADPH over NADH whereas wild type psXR showed two-fold higher preference for NADPH (Watanabe et al., 2007).

In this study, similar studies were done on L-arabinitol 4-dehydrogenase (LAD) and XDH from *N. crassa* to alter cofactor specificity and hence improve xylose fermentation in engineered *S. cerevisiae*.

Identification of Putative LAD-Encoding Genes

Methods of identifying putative LAD-encoding genes and of cloning LAD-encoding and putative LAD-encoding genes are described.

Identification of Putative LAD-Encoding Genes

From a protein BLAST search using ncLAD (EAA36547.1) as a probe, two putative genes were identified in *P. chrysogenum* (XP_002569286.1) and *P. guilliermondii* (EDK37120.2), respectively. The amino acid sequence identities of these two proteins with ncLAD were 71% and 46%, respectively.

Cloning LAD-Encoding and Putative LAD-Encoding Genes

A. niger (NRRL 326), *P. guilliermondii* (NRRL Y2075), and *P. chrysogenum* (NRRL 807) were obtained from the United States Department of Agriculture Agricultural Research Service Culture Collection (Peoria, Ill.). *T. longibrachiatum* (*T. reesei*, YSM 768) was obtained from the German Resource Centre for Biological Material (DSMZ).

A. niger, *T. longibrachiatum*, *P. chrysogenum*, and *P. guilliermondii* were grown in liquid media or on agar plates containing 1% yeast extract, 2% peptone, and 2% L-arabinose. Cells were frozen in liquid nitrogen for the isolation of total RNA or genomic DNA. Reverse transcription-PCR(RT-PCR) was performed on mRNAs isolated from *T. longibrachiatum*, *P. chrysogenum*, and *P. guilliermondii* to obtain cDNA, and PCR was used to obtain the genes encoding (putative) LADs. For *A. niger*, the putative LAD gene could not be amplified from cDNA due to unknown reasons. Thus, overlap extension-PCR (OE-PCR) was used to clone this intron-containing gene from the isolated genomic DNA. Note that all primer sequences used to clone these genes are listed in Table 19.

TABLE 19

Primers used for the cloning of wild type LADs. Restriction enzyme sites are in bold and italicized.			
	Restriction Enzyme	Primer	Sequence
anLAD	NdeI	Fwd-fragment1 ^a	5'-GACATCGATGA CATA TG ^c GCTACCGAAC-3' SEQ ID NO: 71
		Rev-fragment1	5'-GTGCACGT ^c GGACCCGAGATTCC-3' SEQ ID NO: 72
	BamHI	Fwd-fragment2 ^b	5'-GGAATCTGGGGTCCGACGTGCAC-3' SEQ ID NO: 73
		Rev-fragment2	5'-CAGAAGATTAA GGAT CCTGAACGTAGA-3' SEQ ID NO: 74
tLLAD	NdeI	Fwd	5'-GACATCACTGA CATA TGTCGCCCTCC-3' SEQ ID NO: 75
		BamHI	5'-CCTGGATTGAG GGAT CCTGAACGTATA-3' SEQ ID NO: 76
	EcoRI	For	5'-GACATCGATGA CATA TG ^c GCTCCGAAAC-3' SEQ ID NO: 77
		Rev	5'-CCAGAA ^c TATTGAG AAATT C ^c TGAACGTAGA-3' SEQ ID NO: 78
pgLAD	NdeI	Fwd	5'-GACATCGATGA CATA TG ^c GCGACTCTGC-3' SEQ ID NO: 79
		BamHI	5'-GGATACAGAATGAG GGAT CCTGAACGTAGA-3' SEQ ID NO: 80

^{a,b}Fragment 1 and 2 indicate the upstream and downstream exons flanking the intron.

^cSequences in bold (italicized) indicate restriction enzyme sites.

PCR products were subcloned into pET-28a vector and the constructs were used to transform into two *E. coli* strains, DH5 α and BL21 (DE3), by electroporation for cloning and expression, respectively. NdeI/BamHI restriction sites were used for the subcloning of the predicted genes from *A. niger*,

T. longibrachiatum, and *P. guilliermondii*, and NdeI/EcoRI sites were used for *P. chrysogenum*. The constructs encoded (putative) LADs as N-terminal His₆-tagged fusions. Plasmids were sequenced using BIGDYE™ Terminator sequencing method and analyzed with 3730xL Genetic Analyzer (Applied Biosystems, Foster City, Calif.) at the Biotechnology Center at the University of Illinois at Urbana-Champaign (Urbana, Ill.).

Protein Expression and Purification

Genes encoding pcLAD (XP_002569286.1), pgLAD (EDK37120.2), anLAD (CAH69383.1), and tLLAD (AAL08944.1) were cloned into the pET-28a vector and expressed in *E. coli* BL21 (DE3). *E. coli* BL21 (DE3) containing the LAD genes were grown overnight at 30° C. on a rotary shaker at 250 rpm. Overnight culture (50 μ L) was used to inoculate a fresh culture (5 mL), which was grown at 30° C. with shaking at 250 rpm until the optical density at 600 nm (OD₆₀₀) reached 0.6-1.0. The cultures were then induced with 0.3 mM IPTG at 30° C. for 3-4 hrs or at 18° C. for 20 hrs.

The induced cells (1 mL) were lysed by re-suspending them in 1 mL of 50 mM potassium phosphate buffer (pH 7.0) with 1 mg/mL lysozyme and shaking at 30° C. and 250 rpm for 30 min. Cells were kept at -80° C. overnight and thawed at room temperature. The resulting cell lysates were centrifuged at 13,200 rpm for 15 min, and the supernatant and precipitate were analyzed for protein expression by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE).

For protein purification, the induced cells (400 mL) were treated with 15 mL of Buffer A (20 mM Tris, 0.5 M NaCl, 20% glycerol, pH 7.6) with 1 mg/mL lysozyme and shaken at 30° C. and 250 rpm for 30 min. After a freeze-thaw cycle, the

resulting product was further lysed by sonication followed by centrifugation for 20 min at 12,000 rpm to remove cell debris. The supernatants were applied to a column packed with Co²⁺-immobilized metal affinity chromatography resin to purify His₆-tagged proteins following the manufacturer's instruc-

tions. The purified proteins were desalted by ultrafiltration (Amicon Ultra, Millipore, Billerica, Mass.) and washed with HEPES buffer (pH 7.0) containing 150 mM NaCl and 15% glycerol and kept at -20° C. Protein concentrations were determined by the Bradford method (Bradford 1976) according to the manufacturer's protocol.

Characterization of LAD Proteins

The steady-state kinetics, molecular weight, quaternary structure, temperature dependence, pH dependence, L-arabinitol dehydrogenase activity, and metal content of LAD enzymes were analyzed.

L-Arabinitol Dehydrogenase Activity

Lysates were prepared from host cells expressing LAD from *P. chrysogenum*, *P. guilliermondii*, *A. niger*, and *T. longibrachiatum*. Ten microliters of cell lysate were used for an activity assay with 200 mM L-arabinitol and 2 mM NAD⁺ as the substrates in 50 mM potassium phosphate buffer (pH 7.0). NADH production was monitored by measuring absorbance at 340 nm ($\epsilon=6.22 \text{ mM}^{-1} \text{ cm}^{-1}$) using a Cary 300 Bio UV-vis spectrophotometer (Varian, Cary, N.C.).

Steady-State Kinetics

Kinetic parameters of different LAD enzymes were determined. Initial rates were determined by measuring the absorbance change at 340 nm using a UV-vis spectrophotometer at room temperature in 50 mM potassium phosphate buffer (pH 7.0). Initial rates were measured at various concentrations of the substrate (L-arabinitol) and cofactors (NAD⁺/NADP⁺) (5 to 320 mM for L-arabinitol, 0.5 to 3.2 mM for cofactors). Enzyme kinetics for the substrate and cofactors were analyzed using Michaelis-Menten kinetics, and kinetic parameters were determined by fitting data to the Lineweaver-Burk plot. The parameters for substrate were determined by measuring initial rates at saturated cofactor concentrations (3.2 mM) and those for cofactors were determined at saturated substrate concentrations (320 mM). Assays were performed in triplicate.

The cloned LADs showed different binding affinities and catalytic activities for L-arabinitol: K_m differed by two fold and k_{cat} by about three fold amongst the LADs. For L-arabinitol, the K_m values of anLAD, tILAD, and pcLAD were 25±1, 18±1, and 37±2 mM, and the k_{cat} values were 507±22, 346±41, and 1085±71 min⁻¹, respectively (Table 20). The tILAD enzyme had the lowest K_m while pcLAD showed the highest catalytic activity (k_{cat}) and efficiency (k_{cat}/K_m) despite having the highest K_m (Table 20). For cofactor NAD⁺ kinetics, the cloned LADs showed K_m values in the range of 0.2-0.3 mM and catalytic efficiencies in the range of 2526 to 3460 mM⁻¹·min⁻¹ (Table 21). All cloned LADs showed minimal activities toward NADP⁺ (Tables 20, 21). The initial rates were not saturated at highest substrate and cofactor concentration (320 mM for L-arabinitol and 3.2 mM for NADP⁺) due to the large K_m. Therefore, only the catalytic efficiency of the enzyme was determined using 0.1 or 0.2 mM for NADP and 10 or 20 mM for L-arabinitol (K_m>>[S]) (Tables 20, 21).

TABLE 20

Kinetic parameters of LADs for L-arabinitol at saturated cofactor concentrations.					
	Specific activity (U/mg protein)	K _m (mM)	k _{cat} (min ⁻¹)	k _{cat} /K _m (mM ⁻¹ ·min ⁻¹)	
anLAD	NAD ⁺	11.7 ± 0.3 ^a	25 ± 1	507 ± 22	20.0 ± 0.8
	NADP ⁺	— ^b	—	—	0.04 ± 0.01
tILAD	NAD ⁺	8.7 ± 0.1	18 ± 1	346 ± 41	19.0 ± 0.8
	NADP ⁺	—	—	—	0.13 ± 0.02
pcLAD	NAD ⁺	25.3 ± 1.4	37 ± 2	1085 ± 71	29 ± 1
	NADP ⁺	—	—	—	0.04 ± 0.02

5

10

15

20

30

35

40

50

60

65

TABLE 20-continued

Kinetic parameters of LADs for L-arabinitol at saturated cofactor concentrations.				
	Specific activity (U/mg protein)	K _m (mM)	k _{cat} (min ⁻¹)	k _{cat} /K _m (mM ⁻¹ ·min ⁻¹)
tILAD	NAD ⁺	8.7 ± 0.1	18 ± 1	346 ± 41
	NADP ⁺	—	—	0.13 ± 0.02
pcLAD	NAD ⁺	25.3 ± 1.4	37 ± 2	1085 ± 71
	NADP ⁺	—	—	0.04 ± 0.02

^aError indicates standard deviation from the mean, n = 3^bDash indicates not determined due to high K_m for indicated cofactor

TABLE 21

Kinetic parameters of LADs for NAD ⁺ and NADP ⁺ at saturated L-arabinitol concentration.			
	K _m (mM)	k _{cat} (min ⁻¹)	k _{cat} /K _m (mM ⁻¹ ·min ⁻¹)
anLAD	NAD ⁺	0.20 ± 0.01 ^a	494 ± 11
	NADP ⁺	— ^b	20 ± 9
tILAD	NAD ⁺	0.2 ± 0.1	436 ± 96
	NADP ⁺	—	17 ± 9
pcLAD	NAD ⁺	0.3 ± 0.1	1039 ± 165
	NADP ⁺	—	3460 ± 505

^aError indicates standard deviation from the mean, n = 3^bDash indicates not determined due to high K_m for indicated cofactor

Molecular Weight and Quaternary Structure

Calculated molecular weights of the subunits of the four proteins were 43 kDa (anLAD), 41 kDa (tILAD), 42 kDa (pcLAD), and 42 kDa (pgLAD). The molecular weights of the proteins were determined using a Bio-Sil SEC-250 column (300×7.8 mm, Bio-Rad, Hercules, Calif.) on a Shimadzu HPLC system (Shimadzu, Kyoto, Japan). The mobile phase consisted of 50 mM Na₂HPO₄, 50 mM NaH₂PO₄, 150 mM NaCl, and 10 mM NaN₃ (pH 6.8) and the flow rate was 1.0 mL/min. The molecular weights were calculated by comparing the retention times with those of protein molecular weight standard.

The quaternary structures were determined based on the molecular weights observed by HPLC and the molecular weights of monomeric subunits which were determined by SDS-PAGE analysis. Molecular weights of an-, tl-, and pcLAD were determined to be 178, 194, and 173 kDa, respectively. Comparing to the molecular weights of the subunits determined by SDS-PAGE, results suggested that the LADs were non-covalently linked tetramers in their native forms.

Temperature and pH Dependence

The optimal temperatures of the proteins were determined by assaying enzyme activities at temperatures ranging from 10 to 70° C. Thermal inactivation was determined by measuring enzyme activity after various incubation times at 50° C. in phosphate buffer. Enzyme activity was measured with 2 mM NAD⁺ and 200 mM L-arabinitol. Half-life of enzyme activity was determined using a first-order exponential decay function. Temperature was controlled by a Cary temperature controller connected to the UV-vis spectrophotometer (Varian, Cary, N.C.). pH-dependent enzyme activity was determined by measuring activity at pH between 5.0 and 11.0 at saturated concentrations of NAD⁺ (2 mM) and L-arabinitol (200 mM) in a universal buffer (50 mM morpholineethanesulfonic acid/50 mM Tris/50 mM glycine) (Ellis and Morrison 1982).

121

The optimal temperatures of anLAD and pcLAD were between 40 and 50° C., whereas tLAD showed higher optimal temperature between 55 and 65° C. (FIG. 55a). Catalytic activities of the LADs exponentially decreased with the length of incubation time at 50° C. and were almost completely deactivated after 100 min (FIG. 55b). tLAD was the most thermally stable with a half-life of 20 min at 50° C., and anLAD was least stable with a half-life of less than 5 min at 50° C. All characterized LADs showed activity in the pH range of 7 to 11 with maximum activity around pH 9.4 (FIG. 55c). In the pH range outside of 9 to 10, activity was significantly reduced and approximately 20% of activity remained at pH 7.0 (FIG. 55c). No activity was detected at or below pH 5.0.

Metal Analysis

Duplicate samples for metal analysis were prepared in phosphate buffered saline (PBS) by buffer exchange and lyophilization. Each sample contained 1-2 mg of protein in 1 mL buffer solution. The identity and content of the metal were analyzed by inductively coupled plasma atomic emission spectrometry (OES Optima 2000 DV, Perkin Elmer, Boston, Mass.) in the Microanalytical Laboratory at the University of Illinois at Urbana-Champaign (Urbana, Ill.).

Measured weight percentages of Zn²⁺ were close to those calculated based on the 1:1 molar ratio (Table 22).

TABLE 22

Calculated and measured Zn ²⁺ contents.		
	Calculated Weight ^a (%)	Measured weight (%)
anLAD	0.027	0.027 ± 0.003 ^b
tLAD	0.047	0.048 ± 0.003
pcLAD	0.048	0.061 ± 0.013

^aCalculated molecular weights were determined based on the buffer composition, protein concentration, and 1:1 molar ratio of LAD monomer subunit and Zn²⁺. Buffer solution (1 L) contained NaCl (8 g), KCl (0.2 g), Na₂HPO₄ (1.44 g), and KH₂PO₄ (0.24 g).

^bAll samples were analyzed in duplicate and errors were standard deviations.

122

Engineering of LAD Enzymes with Altered Cofactor Specificity

Methods of altering the cofactor specificity of LADs were determined, and mutated LADs were analyzed for altered cofactor specificity and other characteristics.

Development of LADs with Altered Cofactor Specificity

10

Site-directed mutagenesis was performed to alter the cofactor specificity of anLAD, tLAD, and pcLAD from NAD⁺ to NADP⁺. Amino acid numbers 224, 225, and 362 of naturally occurring tLAD were substituted with serine, arginine, and threonine, respectively, to generate the tLAD with altered cofactor specificity. The amino acid sequences of cloned anLAD and pcLAD were aligned with the *T. longibrachiatum* LAD (tLAD) sequence, and the amino acids that correspond to tLAD amino acid numbers 224, 225, and 362 were mutated. For all of the LADs with altered cofactor specificity, two amino acid residues within the β-α-β motif of the coenzyme binding domain were replaced with serine and arginine, respectively: D213 and I214 for anLAD, D224 and I225 for tLAD, and D212 and I213 for pcLAD (Korkhin et al., 1998; Pauly et al., 2003; Watanabe et al., 2005), and the third mutation was introduced at A359 for anLAD, A362 for tLAD, and S358 for pcLAD and replaced with threonine (For primer sequences, see Table 23). Megaprimer PCR method was used to introduce site-specific mutations using wild type LAD constructs as the templates (Sarkar and Sommer 1990). Correct mutations were confirmed by DNA sequence analysis.

TABLE 23

Primers used for site directed mutagenesis by the megaprimer PCR method. ^a		
Fwd-T7-pro		5' -TAATACGACTCACTATAAGGG-3'
Rev-T7-term		SEQ ID NO: 81 5' -GCTAGTTATTGCTCAGCGG-3' SEQ ID NO: 82
anLAD Fwd-D213S/I214R		5' -CCTATCGTCATTACCTCACGT ^b GACGAGGGCGGCTG-3' SEQ ID NO: 83
Rev-D213S/I214R		5' -CAGCCGCCCTCGTCACGTGAGGTAATGACGATAGG-3' SEQ ID NO: 84
Fwd-A359T		5' -CCT TCGAAACGGCTACAAACCCAAGACG-3' SEQ ID NO: 85
tLAD Fwd-D214S/I215R		5' -GCTTGTCAATCACATCACGTTCAGAGAGCCGCTG-3' SEQ ID NO: 86
Rev-D214S/I215R		5' -CAGACGGCTCTCTGAACGTGATGTGATGACAAGC-3' SEQ ID NO: 87
Fwd-S362T		5' -GCATTTGAGACGTCAACAGATCCAAAGAGC-3' SEQ ID NO: 88
pcLAD Fwd-D212S/I213R		5' -CCTATTGTCATCACTTCACGTGACGAGGGCGCTTG-3' SEQ ID NO: 89
Rev-D212S/I213R		5' -CAAGCGCCCTCGTCACGTGAAGTGATGACAATAGG-3' SEQ ID NO: 90
Fwd-S358T		5' -CCTTGAGACTGCCACAAACCTAAGACCAGGTG-3' SEQ ID NO: 91

^aTo create mutant LADs, fragments 1 and 2 were amplified using Fwd-T7-pro and Rev-D213S/I214R and Fwd-A359T and Rev-T7-term primers, respectively. Fragment 3 was amplified using Fwd-D123S/I214R and fragment 2 (Rev megaprimer). Full mutant genes were amplified by overlap extension of fragment 1 and 3. Template DNA was pET-28a plasmid.

^bSequences underlined were the mutation sites.

Kinetic Analysis of LADs with Altered Cofactor Specificity

In this example, “tlLAD mutant” is defined as tlLAD with the mutations D224S/I225R/A362T; “anLAD mutant” is defined as anLAD with the mutations D213S/I214R/A359T; and “pcLAD mutant” is defined as pcLAD with the mutations D212S/I213R/S358T. The tlLAD mutant showed significantly altered cofactor specificity from NAD⁺ to NADP⁺. It also demonstrated the highest catalytic activity. The K_m and k_{cat} of the tlLAD mutant for L-arabinitol with NADP⁺ were 46±4 mM and 170±9 min⁻¹, respectively (Table 24). In all assays including the tlLAD mutant with saturated NAD⁺, a plateau of reaction rate was not observed in the tested concentration range, so catalytic efficiencies were determined at 0.8 mM for NAD⁺ and 80 mM for L-arabinitol (Tables 24, 25). For cofactors, anLAD and tlLAD mutants showed significantly higher preference for NADP⁺ over NAD⁺ (Table 25). The K_m values of the anLAD and tlLAD mutants were 0.46±0.09 and 0.10±0.01 mM, and the k_{cat} values were 55.7±6.4 and 90.5±9.2 min⁻¹, respectively (Table 25). The catalytic efficiencies of anLAD and tlLAD mutants were 130±32 and 934±72 mM⁻¹·min⁻¹, and the ratios of the catalytic efficiencies with NADP⁺ to NAD⁺ were 100 and 161, respectively. For the tlLAD mutant, the ratio of catalytic efficiency for NADP⁺ to NAD⁺ was increased by 2.5×10⁴ fold (Tables 21, 25). The pcLAD mutant showed no activity with NAD⁺.

		Specific activity (U/mg protein)	K _m (mM)	k _{cat} (min ⁻¹)	k _{cat} /K _m (mM ⁻¹ ·min ⁻¹)
anLAD	NAD ⁺	— ^a	—	—	0.010 ± 0.002 ^b
mutant	NADP ⁺	—	—	—	0.45 ± 0.20
tlLAD	NAD ⁺	—	—	—	0.050 ± 0.007
mutant	NADP ⁺	3.9 ± 0.2	46 ± 4	170 ± 9	3.7 ± 0.2
pcLAD	NAD ⁺	—	—	—	—
mutant	NADP ⁺	—	—	—	0.02 ± 0.02

^aDash indicates not determined due to high K_m for indicated cofactor

^bError indicates standard deviation from the mean, n = 3

		K _m (mM)	k _{cat} (min ⁻¹)	k _{cat} /K _m (mM ⁻¹ ·min ⁻¹)
anLAD mutant	NAD ⁺	— ^a	—	1.3 ± 0.3 ^b
	NADP ⁺	0.46 ± 0.09	55.7 ± 6.4	130 ± 32
tlLAD mutant	NAD ⁺	—	—	5.8 ± 0.8
	NADP ⁺	0.097 ± 0.011	90.5 ± 9.2	934 ± 72

Enzyme	NAD ⁺			NADP ⁺			Source
	k _{cat} (min ⁻¹)	K _m (mM)	k _{cat} /K _m (mM ⁻¹ ·min ⁻¹)	k _{cat} (min ⁻¹)	K _m (mM)	k _{cat} /K _m (mM ⁻¹ ·min ⁻¹)	
ncXDH-wt	2160	0.127	17000	-a	~5.6	~68	This work
ncXDH-ARS	-a	~3.5	~165	2080	0.325	6400	This work
psXDH	1050	0.381	2760	110	170	0.65	Watanabe et al. (2005b)
psXDH-ARS	240	1.3	181	2500	0.897	2790	Watanabe et al. (2005b)

^aNot determined, cofactor saturation not reached.

All assays were performed at 25° C. in 50 mM Tris, pH 8.0.

-continued

		K _m (mM)	k _{cat} (min ⁻¹)	k _{cat} /K _m (mM ⁻¹ ·min ⁻¹)
5	pcLAD mutant	NAD ⁺	—	—

^aDash indicates not determined due to high K_m for indicated cofactor

^bError indicates standard deviation from the mean, n = 3

10 Engineering of *N. crassa* XDH (ncXDH) with Altered Cofactor Specificity

Cloning and Characterization of Putative ncXDH

A putative *N. crassa* xylitol dehydrogenase (ncXDH) sequence was found using a protein BLAST search on the National Center for Biotechnology Information website (webpage ncbi.nlm.nih.gov) using the *P. stipitis* xylitol dehydrogenase (psXDH) enzyme as a query sequence. The two enzymes were aligned fully using a ClustalW algorithm and found to share 44% identity and 60% similarity (FIG. 56). The whole-genome sequence of *Neurospora crassa* has been published (Galagan et al., 2003) and it was utilized to design primers for cloning of the putative xylitol dehydrogenase (XDH) gene.

RT-PCR performed on total RNA isolated from D-xylose-induced *N. crassa* 10333 showed the expected size of gene product (~1.1 kb). The RT-PCR product was cloned into the pET-28a vector using NdeI and SacI restriction sites and was transformed into *E. coli* BL21 (DE3). This construct (pET-28a ncXDH) expressed ncXDH as an N-terminal His6-tagged fusion with a thrombin cleavage site. Cell lysates of IPTG-induced cultures of these cells were prepared, analyzed by SDS-PAGE, and assayed for XDH activities. The XDH was then purified by immobilized metal ion affinity chromatography (IMAC) using Talon® Co2+ Superflow resin (Clontech, Mountain View, Calif.) according to manufacturer's protocol. The purified protein was desalted by ultrafiltration with several washes of 50 mM 4-(2-hydroxyethyl)piperazine-1-ethanesulfonic acid (HEPES) buffer (pH 7.25)+15% glycerol and stored frozen at -80° C. Protein concentrations were determined by the Bradford method (Bradford 1976).

40 ncXDH is a strictly NAD⁺-preferring enzyme. ncXDH also displays high stability (half-life of ~200 min at 50° C.) and expression. Previous work by Watanabe et al. (2005b) was aimed at reversing the cofactor specificity of psXDH.

Development of ncXDH with Altered Cofactor Specificity

45 Through sequence alignment, residues D204, I205, and V206 of ncXDH were targeted for site-directed mutagenesis to alanine, arginine, and serine, respectively, to create ncXDH-ARS. Table 26 shows that ncXDH-ARS has completely reversed cofactor specificity, now preferring NADP⁺. The affinity for substrate xylitol did not suffer very much from the affinity-change for the co-factor.

125

Kinetic Analysis of ncXDH Mutant

The mutant ncXDH had a dramatic reversal of cofactor specificity. The K_m of the mutant ncXDH for NADP⁺ was only about 2.5-fold higher than the K_m of wild-type ncXDH for NAD⁺ whereas the k_{cat} values were similar (Table 27).

Enzyme	k_{cat} (min ⁻¹)	K_m (mM)	k_{cat}/K_m (mM ⁻¹ min ⁻¹)
ncXDH-wt	2170 ± 135	6.6 ± 2.0	330
ncXDH-ARS	2090 ± 35	4.3 ± 0.3	490

a Not determined, cofactor saturation not reached.

All assays were performed at 25° C. in 50 mM Tris, pH 8.0.

All enzymes were purified and characterized with N-His₆-tag

As shown in FIG. 57, XDH activity exhibits a higher tolerance to more acidic conditions with activity extending down to pH 4.0, whereas LAD activity is abolished at pH 5.0 in the in vitro activity assay.

Example 14

Expression of Xylose Isomerase from *Bacteroides stercoris* in *S. cerevisiae*

Bacterial xylose isomerase (XI) is involved in converting xylose into xylulose. Recently, three successful cases of expressing active XI from two species of anaerobic fungi (*Piromyces* sp. and *Orpinomyces* sp.) and from the anaerobic bacteria (*Clostridium phytofermentans*) have been reported. A fungal XYLA gene from *Piromyces* sp. E2 was functionally expressed in *S. cerevisiae* and a maximum 1.1 U/mg-protein of XI activity was obtained at 30° C. (Kuyper et al., 2003). The second fungal XYLA gene from *Orpinomyces*, which has 94% identity with that from *Piromyces* sp., was also functionally expressed in *S. cerevisiae* (Madhavan et al., 2009). Recently, the first prokaryotic xylA gene from *Clostridium phytofermentans* was functionally expressed in *S. cerevisiae* (Brat et al., 2009).

The isomerase gene xylA from the anaerobic bacteria *Bacteroides stercoris* (BtXI) shares high sequence identity with the isomerase gene from *Piromyces* sp. (82%). BtXI was cloned into the pRS424TEF vector and transformed into the *S. cerevisiae* L2612 strain. The gene was also integrated into the *S. cerevisiae* D452-2 strain by using the pRS403TEF vector. Ethanol production was observed in both strains expressing BtXI (5 g/L in L2612 and 7.8 g/L in D452-2) (FIG. 58-59). However, rates of production were relatively low compared to that of engineered strains expressing the XYL genes.

The low ethanol production could be attributed to the inhibitory effect of any accumulated xylitol (formed from xylose by endogenous yeast aldose reductase). To decrease xylitol accumulation, XDH and XK were expressed in BTXI-expressing yeast strain (DBtXI). The resulting strain had slightly improved ethanol yield and decreased xylitol production (FIG. 60). Co-expression of these two XYL genes in DBtXI resulted in ethanol production even under aerobic conditions.

Example 15

Over-Expression of Enzymes in Pentose Phosphate Pathway (PPP)

The PPP enzymes glucose-6-phosphate dehydrogenase (ZWF1), 6-phosphogluconate dehydrogenase (GDN1), tran-

126

saldolase (TAL1), and transketolase (TKT1) from *P. stipitis* were cloned into an integration vector (pRS406) under the control of a strong promoter (P_{GPD}). The plasmid was linearized by the enzyme StuI and integrated into the chromosome of *S. cerevisiae*.

However, to get the beneficial effects of over-expressing the PPP enzymes, there also had to be over-expression of XYL3 (XK) (FIG. 61). Expression of XYL3 and the PPP enzymes also improved ethanol production in YP-xylulose media.

Example 16

Expression of Aldose-1-Epimerase

Hydrolysis of cellobiose by β -glucosidase releases β -D-glucose. However, yeast hexokinases prefer (or exclusively use) α -D-glucose, and the rate of mutarotation of β -D-glucose to α -D-glucose could effectively slow down metabolic rate.

One way of enhancing the conversion was to over-express the predicted aldose-1-epimerase NCU09705. This hypothesis was tested by over-expressing NCU09705 homologs: galM in *E. coli*; GAL10, YHR210C, and YNR071c in *S. cerevisiae*; and GAL 10 in *P. stipitis*. The strains were then tested for cellobiose consumption and ethanol production (FIG. 62). The results indicated that over-expression of the homologs in *S. cerevisiae* caused a slight increase in cellobiose consumption and ethanol production.

Example 17

Co-Fermentation of Xylose and Cellobiose

In this example a new strategy was used to overcome glucose repression in which a dimer of glucose, cellobiose, was co-fermented with xylose (a pentose). Cellobiose is an intermediate product from enzymatic hydrolysis of cellulose, which is further converted to glucose by β -glucosidases in the cocktail of cellulases including exocellulases, endocellulases, and β -glucosidases, whereas pentose sugars are the products of dilute acid hydrolysis of hemicellulose. Wild type *S. cerevisiae* cannot assimilate cellobiose because it lacks both a cellobiose transporter and a β -glucosidase capable of hydrolyzing cellobiose into glucose. Hence, the newly discovered cellobextrin transporter genes described in Example 9 and a β -glucosidase gene from *N. crassa* were co-expressed in *S. cerevisiae* and a mixture of xylose and cellobiose was used as carbon source (FIG. 63). Similar approaches have employed either secretion, or cell surface display, of β -glucosidases to allow cellobiose fermentation by *S. cerevisiae* (van Rooyen et al., 2005; Skory et al., 1996; Kotaka et al., 2008; Katahira et al., 2006). In those cases, cellobiose was hydrolyzed into glucose extracellularly before being transported by the endogenous hexose transport system of *S. cerevisiae*. In contrast, in this strategy, cellobiose was hydrolyzed intracellularly following transport.

In the conventional methods for mixed sugar fermentation in *S. cerevisiae*, a mixture of glucose and pentose sugars derived from lignocellulose is used. However, in this new strategy, a mixture of cellobiose and pentose sugars was used. The cellobiose was transported inside yeast cells via the heterologous cellobextrin transporters while pentose sugars were transported inside yeast cells by endogenous hexose transporters, thus removing the direct competition between glucose and pentose sugars for the same transporters, a phenomenon that is partly responsible for glucose repression. Once inside yeast cells, cellobiose was converted to glucose

127

by β -glucosidase and immediately consumed by yeast cells, which resulted in low intracellular glucose concentration, thereby further alleviating glucose repression.

The engineered xylose-utilizing yeast strain L2612 was used as a host to co-express celldextrin transporter and β -glucosidase genes. In this strain, the D-xylose utilization pathway consisting of xylose reductase, xylitol dehydrogenase, and xylulokinase from *Pichia stipitis* was integrated into the chromosome. The celldextrin transporters from *Neurospora crassa* including NCU008011, NCU08114, and, NCU00809, and two β -glucosidase genes, one from *Neurospora crassa* and the other from *Aspergillus aculeatus*, were evaluated.

S. cerevisiae L2612 (MAT α , leu2-3, leu2-112, ura3-52, trp1-298, can1, cyn1, gal+) was cultivated in synthetic drop-out media to maintain plasmids (0.17% of Difco yeast nitrogen base without amino acids and ammonium sulfate, 0.5% of ammonium sulfate, 0.05% of amino acid dropout mix). YPA medium (1% yeast extract, 2% peptone, 0.01% adenine hemisulfate) with 2% of sugar was used to grow yeast strains.

To integrate the D-xylose utilization pathway consisting of D-xylose reductase, xylitol dehydrogenase, and xylulokinase from *Pichia stipitis*, the corresponding genes were PCR-amplified and cloned into the pRS416 plasmid using the DNA assembler method (Shao et al., 2009). BamHI and HindIII were used to remove the DNA fragment encoding the D-xylose utilization pathway and then ligated into the pRS406 plasmid digested by the same two restriction enzymes. The resulting plasmid was then linearized by ApaI and integrated into the URA3 locus on the chromosome of L2612.

The pRS425 plasmid (New England Biolabs, Ipswich, Mass.) was used to co-express a celldextrin transporter gene and a β -glucosidase gene. As shown in FIG. 64, the pRS425 plasmid was digested by BamHI and ApaI. The PYK1 promoter and the ADH1 terminator were added to N-terminus and C-terminus of the celldextrin transporter, respectively, while the TEF1 promoter and the PGK1 terminator were added to the N-terminus and C-terminus of the β -glucosidase, respectively. These DNA fragments were assembled into the linearized pRS425 shuttle vector using the DNA assembler method (Shao et al., 2009). Three celldextrin transporter genes NCU00801 (XM_958708), NCU08114 (XM_958780), and NCU00809 (XM_959259) from *Neurospora crassa* and two β -glucosidase genes NCU00130 (XM_951090) from *Neurospora crassa* and BGL1 (D64088) from *Aspergillus aculeatus* were used. There were six combinations in total, each with one celldextrin transporter gene and one β -glucosidase gene.

Yeast plasmids were then transferred into *E. coli* DH5 α , which were used for recombinant DNA manipulation. The transformants were plated on Luria broth plates containing 00 mg/L ampicillin. Single colonies of *E. coli* transformants were then inoculated into the liquid Luria broth media (Fisher Scientific, Pittsburgh, Pa.) and grown at 37° C. and 250 rpm. Plasmids were isolated from *E. coli* using the QIAprep Spin Miniprep Kit (QIAGEN). These plasmids were transformed into the L2612 strain individually to yield the following strains: SL01 (contained the plasmid harboring the NCU00801 celldextrin transporter gene and the NCU00130 β -glucosidase gene from *Neurospora crassa*), SL02 (contained the plasmid harboring the NCU00809 celldextrin transporter gene and the NCU00130 β -glucosidase gene from *Neurospora crassa*), SL03 (contained the plasmid harboring the NCU08114 celldextrin transporter gene and the NCU00130 β -glucosidase gene from *Neurospora crassa*), SL04 (contained the plasmid harboring the NCU00801 celldextrin transporter gene and the BGL1 gene from *Aspergil-*

128

lus aculeatus), SL05 (contained the plasmid harboring the NCU00809 celldextrin transporter gene and the BGL1 gene from *Aspergillus aculeatus*), and SL06 (contained the plasmid harboring the NCU08114 celldextrin transporter gene and the BGL1 gene from *Aspergillus aculeatus*). The empty pRS425 plasmid was transformed into the L2612 strain to yield the SL00 strain, which was used as a negative control. Yeast transformation was carried out using the standard lithium acetate method (Gietz et al., 1995). The resulting transformation mixtures were plated on SC-Ura-Leu medium supplemented with 2% D-glucose.

To confirm the proper construction of plasmids using the DNA assembler method, plasmids were isolated from yeast cells using the Zymoprep Yeast Plasmid Miniprep II kit (Zymo Research, Orange, Calif.) and then transferred into *E. coli* DH5 α cells. The resulting cells were spread on LB plates containing 100 mg/L ampicillin. Single *E. coli* colonies were inoculated into the LB liquid media. Plasmids were isolated from *E. coli* using the QIAprep Spin Miniprep Kit (QIAGEN, Valencia, Calif.) and checked by diagnostic PCR or restriction digestion using ClaI and HindIII. All restriction enzymes were obtained from New England Biolabs (Ipswich, Mass.). All chemicals were purchased from Sigma Aldrich or Fisher Scientific.

For each yeast strain, single colony was first grown up in 2 mL SC-Ura-Leu medium plus 2% glucose, and then inoculated into 50 mL of the same medium in a 250 mL shake flask to obtain enough cells for mixed sugar fermentation studies. After one day of growth, cells were spun down and inoculated into 50 mL of YPA medium supplemented with 4% cellobiose and 5% D-xylose, or 4% cellobiose, 5% xylose, and 0.5% glucose, or 4% cellobiose, 5% xylose, and 1% glucose in a 250 mL unbaffled shake-flask. Starting from an initial OD₆₀₀~1, cell culture was grown at 30° C. at 100 rpm for fermentation under oxygen limited condition. OD₆₀₀ reading and cell culture sample were taken at various time points. Sugar concentrations were analyzed using HPLC, while ethanol formation was analyzed using the Ethanol Kit (R-biopharm, Darmstadt, Germany). For each data point, triplicate samples were taken. The mixed sugar fermentation results for the strains ranging from SL00 to SL06 are shown in FIG. 65. The best strain SL01 was selected for further characterization.

A total of six different strains, ranging from SL01 to SL06, were constructed by introducing a pRS425 plasmid harboring one of the celldextrin transporter genes and one of the β -glucosidase genes into the L2612 strain. In each plasmid, the celldextrin transporter gene and the β -glucosidase gene were added with a yeast promoter and terminator, respectively, and assembled into the pRS425 multi-copy plasmid by the DNA 10 assembler method (Shao et al., 2009) (FIG. 64). The empty pRS425 plasmid was introduced into the L2612 strain to yield the SL00 strain, which was used as a negative control. All strains were cultivated with a mixture of 40 g/L cellobiose and 50 g/L D-xylose in shake-flasks, and their sugar consumption rates, cell growth rates, and ethanol titers were determined (FIG. 65). Amongst all strains, the SL01 strain containing the β -glucosidase from *Neurospora crassa* and the celldextrin transporter NCU00801 showed the highest sugar consumption rate and ethanol productivity. Thus, this strain was selected for further characterization.

Both SL01 and SL00 were cultivated using a mixture of 40 g/L cellobiose and 50 g/L D-xylose in both shake-flasks and bioreactors (FIG. 66). In the shake-flask cultivation (FIG. 66a-b), 83% cellobiose was consumed in 96 hours by SL01, with 41.2% higher average D-xylose consumption rate compared to SL100 (from 0.33 g/L/h to 0.46 g/L/h). Consistent

129

with the enhanced sugar consumption rate, 1.32-fold increased average biomass growth rate was observed (from 0.031 g dry cell weight/L/h to 0.072 g dry cell weight/L/h). The ethanol productivity was increased by more than 2.1-fold, from 0.07 g/L/h to 0.23 g/L/h. The highest ethanol yield of 0.31 g per g sugar was reached in 48 hours, and the average ethanol yield was 0.28 g per g sugar, representing a 23% increase compared to the SL00 strain. In the SL01 cultivation, a faster D-xylose consumption rate was observed, without the lag phase that is the hallmark of glucose repression in co-fermentation of glucose and D-xylose. Moreover, enhanced biomass growth and ethanol production were also observed.

The Multifors system (Infors-HT, Bottmingen, Switzerland) was used for mixed sugar fermentation in bioreactors. Each vessel had a total capacity volume of 750 mL. For each vessel, there was one individual set of pO₂ sensor, air sparger, exit gas cooler, temperature sensor, inoculation port, spare port, dip tube, antifoam sensor, pH sensor, drive shaft, heater block, rotameter, and peristaltic pumps system. The whole bioreactor system was equipped with a cooling system, ThermoFlex900 (Thermo Scientific, Waltham, Mass.).

Single colonies of yeast strains were first grown up in 2 mL SC-Ura-Leu medium plus 2% glucose, and then inoculated into 50 mL of the same medium in a 250 mL shake flask to obtain enough cells for mixed sugar fermentation studies. After one day of growth, 10 mL saturated culture were inoculated in 400 mL YPA medium supplemented with 4% cellobiose and 5% D-xylose, or 4% cellobiose, 5% xylose, and 0.5% glucose, or 4% cellobiose, 5% xylose, and 1% glucose. The temperature was maintained at 30° C. and the pH was maintained at 5.5, adjusted by addition of either 2 N H₂SO₄ or 4 N NaOH. In the first 48 hours, the air flow rate was maintained at 0.5 L/min, with the impeller speed at 250 rpm. Afterwards, the air flow rate was adjusted to 0.2 L/min to achieve high ethanol production under oxygen limited condition. Triplicate samples were taken at various time points and the OD₆₀₀, sugar concentration, and ethanol concentration were determined as described above.

In the bioreactor cultivation (FIG. 66c-d), almost all cellobiose and 66% D-xylose were consumed in 48 hours, representing 44% increased D-xylose consumption rate (from 0.47 g/L/h to 0.68 g/L/h) and 1.1-fold increased biomass growth rate (from 0.08 g dry cell weight/L/h to 0.17 g dry cell weight/L/h). The ethanol productivity was increased by more than 4.3-fold (from 0.09 g/L/h to 0.50 g/L/h), and the ethanol yield was 0.39 g per g sugar. Compared to shake-flask cultivations, sugar consumption rates in the first 24 hours were lower, which was due to the low cell density used in the beginning of batch cultivation.

Unexpectedly, a small amount of glucose was detected even though there was no glucose added in fermentation (FIG. 66a-b). The maximum glucose concentration was reached in approximately 24 hours in both shake-flasks (12.1 g/L) and bioreactors (17.5 g/L) and then dropped to a very low level. However, no obvious glucose repression was observed even in the presence of such glucose. Because no glucose was detected in the SL00 strain, the extracellular glucose may result from the slow conversion of β-glucose to its epimer α-glucose, the main form of glucose used in glycolysis. Typically, β-glucose can be efficiently converted to α-glucose either enzymatically or chemically because of its relatively low concentration in glucose (Bouffard et al., 1994). However, in the engineered SL01 strain, catalyzed by β-glucosidase, an excess amount of β-glucose is produced from cellobiose intracellularly and a small fraction may be secreted outside cells, similar to what was observed with β-galactose (Bouffard et al., 1994).

130

Because a small amount of glucose (less than 10% of total sugars) is typically present in lignocellulosic hydrolysates in industrial settings, the fermentation performance of the engineered SL01 strain was also investigated using a mixture of cellobiose, D-xylose, and glucose. Two concentrations of glucose, 5 g/L or 10 g/L, were combined with 40 g/L cellobiose and 50 g/L D-xylose as mixed carbon source in bioreactors. With 5 g/L glucose (FIG. 67a-b), 81.7% cellobiose was consumed by SL01, with 67.8% D-xylose consumed at 48 hours in batch cultivation. The D-xylose consumption rate was increased by 1.19-fold, from 0.32 g/L/h to 0.69 g/L/h. The ethanol productivity was increased by 3.3-fold (from 0.11 g/L/h to 0.46 g/L/h) while the ethanol yield was increased from 0.26 g per g sugar to 0.33 g per g sugar. With 10 g/L glucose (FIG. 67c-d), 83.8% cellobiose was consumed by SL01, with 74.7% D-xylose consumed at 48 hour in batch cultivation. The D-xylose consumption rate was increased by 68%, from 0.45 g/L/h to 0.76 g/L/h. The ethanol productivity was increased by 2.1-fold (from 0.16 g/L/h to 0.50 g/L/h) and the ethanol yield was increased from 0.30 g per g sugar to 0.33 g per g sugar. As expected, the engineered SL01 strain showed both a higher efficiency of sugar consumption and a higher rate of ethanol production than the SL00 wild type strain. More importantly, there was no significant glucose repression in the co-fermentation of three sugars even with glucose up to 10% of total sugars (FIG. 67c-d) suggesting that this approach may be viable for industrial applications.

A similar study was carried out in the *S. cerevisiae* strain D452-2, where the three *N. crassa* cellobextrin transporters NCU00801, NCU08114, and NCU00809 were introduced together with the β-glucosidase NCU00130. The transformants were selected on YSC medium containing 20 g/liter cellobiose expressing an intracellular β-glucosidase (NCU00130). Strains and plasmids used in this work are described in Table 17 (Ex. 12). The primers used are listed in Table 28.

	Name	Sequences
40	NCU00801-F	ATGGATCCAAAATGTCGCTCACGGCTCC SEQ ID NO: 92
45	NCU00801-R	ATGAATTCTACAAATCTTCTTCAGAAATCAATTTTGTCAGAACGATAGCTTCGGAC SEQ ID NO: 93
50	NCU08114-F	ATACTAGTAAAAATGGGCATCTTCAACAAGAAGC SEQ ID NO: 94
55	NCU08114-R	GCATATCGATCTACAAATCTTCTTCAGAAATCAATTTGTTCAGCAACAGACTGCCCTCATG SEQ ID NO: 95
60	NCU00130-F	GCATACTAGTAAAATGTCCTTCTTAAGGATTCCTCT SEQ ID NO: 96
65	NCU00130-R	ATACTGCAGTTAATGATGATGATGATGATGGTCCTTCTTGTCAAAGAGTC AAG SEQ ID NO: 97

Yeast were grown in YP medium containing 20 g/L of glucose or 20 g/L of cellobiose to prepare inoculums for xylose or cellobiose fermentation experiments, respectively. Cells at mid-exponential phase from YP media containing 20 g/L of glucose or cellobiose were harvested and inoculated after washing twice with sterilized water. All of the flask fermentation experiments were performed using 50 mL of YP medium containing 40 g/L or 80 g/L of xylose in 250 mL flask at 30° C. with initial OD₆₀₀ of 1.0 under oxygen limited

131

conditions. Bioreactor fermentations were performed in 400 mL of YP medium containing appropriate amounts of sugars using Sixfors Bioreactors (Appropriate Technical Resources, Inc) at 30° C. with an agitation speed of 200 rpm under oxygen limited 250 conditions. Initial cell densities were adjusted to OD₆₀₀=1.0.

Cell growth was monitored by optical density (OD) at 600 nm using UV-visible Spectrophotometer (Biomate 5, Thermo, N.Y.). Glucose, xylose, xylitol, glycerol, acetate, and ethanol concentrations were determined by high performance liquid chromatography 264 (HPLC, Agilent Technologies 1200 Series) equipped with a refractive index detector using 265a Rezex ROA-Organic Acid H+ (8%) column (Phenomenex Inc., Torrance, Calif.). The column was eluted with 0.005 N of H₂SO₄ at a flow rate of 0.6 mL/min at 50° C.

All three transformants were able to grow and produce ethanol when cellobiose was the sole carbon source (FIG. 68), but the three transformants exhibited different cellobiose fermentation rates (NCU00801>NCU08114>NCU00809). The fastest cellulose-fermenting transformant (D801-130), expressing both NCU00801 and NCU00130, consumed 40 g/L of cellobiose within 4 hours, producing 16.8 g/L of ethanol. The volumetric productivity of cellobiose fermentation ($P_{Ethanol/Cellobiose} = 0.7 \text{ g/L/h}$) was lower than that of glucose fermentation ($P_{Ethanol/Glucose} = 1.2 \text{ g/L/h}$), and ethanol yield from cellobiose ($Y_{Ethanol/Cellobiose} = 0.42 \text{ g/g}$) was about the same as ethanol yield from glucose ($Y_{Ethanol/Glucose} = 0.43 \text{ g/g}$) under the same culture conditions. However, the observed cellobiose consumption rate and ethanol yield by D801-130 were an improvement over *S. cerevisiae* strains engineered to ferment cellobiose through surface display of β-glucosidase (Kotaka et al., 2008; Nakamura et al., 2008). These results suggest that simultaneous expression of NCU00801 and NCU00130 in *S. cerevisiae* can result in efficient cellobiose fermentation.

After developing the efficient xylose fermenting strain DA24-16 (described in Example 13), genes coding cellobextrin transporter and β-glucosidase (NCU00801 and NCU00130) enzyme were introduced into the strain enabling it to consume cellobiose and xylose simultaneously. It was hypothesized that glucose repression of xylose utilization may be alleviated in this strain, due to the intracellular hydrolysis of cellobiose. The NCU00801 gene was integrated into the genome of DA24-16, and NCU00130 was expressed from a multi-copy plasmid. The resulting transformant, DA24-16-BT3, was selected on an agar plate containing cellobiose as the sole carbon source.

The DA24-16-BT3 strain grown in media containing various amounts of cellobiose and xylose co-consumed cellobiose and xylose, and produced ethanol with yields of 0.38-0.39 g/g in all conditions tested (FIG. 69). The potential synergistic effects of co-fermentation were tested by culturing DA2416-BT3 under three different conditions: 40 g/L of cellobiose, 40 g/L of xylose, and 40 g/L of both sugars (total 80 g/L of sugars). Surprisingly, DA24-16BT3 was able to co-consume 80 g/L of a cellobiose/xylose mixture within the same period that was required to consume 40 g/L of cellobiose or 40 g/L xylose separately (FIG. 70). Moreover, DA24-16BT3 produced ethanol with a higher yield (0.39 g/g) from a mixture of cellobiose and xylose as compared to ethanol yields (0.31~0.33 g/g) from single sugar fermentations (cellobiose or xylose). Ethanol productivity also drastically increased from 0.27 g/L/h to 0.65 g/L/h during co-fermentation. These results demonstrated that co-fermentation of cellobiose and xylose can enhance overall ethanol yield and productivity. Fermentation experiments were also done to

132

compare this engineered *S. cerevisiae* strain (DA24-16BT3) to *P. stipitis*, which is capable of co-fermenting cellobiose and xylose efficiently.

A simulated hydrolysate (10 g/L of glucose, 80 g/L of cellobiose, 40 g/L of xylose) based on the composition of energycane was used. The composition of different lignocellulosic plants varies in a broad range. For instance, the US Department of Energy biomass database lists the composition of more than 150 biomass samples (webpage eere.energy.gov/biomass/m/feedstock_databases.html). The cellulose-to-hemicellulose ratios of these samples are between 1.4 and 19, and the average is 2.3. Energy crops typically have higher hemicellulose content than woody biomass. The average cellulose to hemicellulose ratios of sugarcane bagasse, corn stover, sorghum are 2.0, 1.85 and 2.14, respectively. We therefore used a glucan/xylan ratio of 2 in our simulated sugar experiment design. The engineered yeast will likely be used in conjunction with traditional cellulase cocktails that are deficient in β-glucosidase activities for the biofuels production. The biomass hydrolysis process may result in small amounts of glucose in the lignocellulosic hydrolysates as 6-30% glucan-to-glucose conversions with incomplete cellulase cocktails were reported (Medve et al., 1998). Considering all the above factors, a sugar combination of 10 g/L glucose, 80 g/L cellobiose, and 40 g/L xylose was chosen in the simulated sugar experiments.

The DA24-16BT3 consumed glucose first before co-consuming cellobiose and xylose rapidly. A total of 130 g/L of sugars was consumed within 60 hours even though small inoculums were used (OD₆₀₀=1). In contrast, *P. stipitis* could not finish fermenting the sugar mixture within the same period under identical culture conditions (FIG. 71). DA24-16BT3 produced 48 g/L of ethanol within 60 hours ($Y_{Ethanol/Sugars} = 0.37 \text{ g/g}$ and $P_{Ethanol/Sugars} = 0.79 \text{ g/L/h}$).

A transient accumulation of cellobextrins in the medium during cellobiose consumption was observed (FIG. 72-73). The accumulated cellobiose and cellobetraose were again consumed after depletion of cellobiose. It is likely that the accumulated cellobextrins were generated by the trans-glycosylation activity (Christakopoulos et al., 1994) of β-glucosidase (NCU00130), and secreted by the cellobextrin transporter (NCU00801), which might facilitate the transport of cellobextrins in both directions (intracellular ↔ extracellular). This transient cellobextrin accumulation would probably not reduce product yields since the accumulated cellobextrins would eventually be consumed by the engineered yeast. However, it might decrease productivity because the transport rates of cellobiose and cellobetraose might be slower than that of cellobiose.

Small amounts of glucose were constantly detected in the medium during co-fermentation. Since even low amounts of glucose accumulation can repress xylose fermentation, glucose levels have to be kept at a minimum. It can be hypothesized that the relative expression levels of the cellobextrin transporter and β-glucosidase are likely to affect glucose accumulation. In support of this, it was observed that more glucose was accumulated in the medium when NCU00801 was introduced on a multi-copy plasmid than when NCU00801 was integrated into the yeast genome. The strain (DA24-16-BT), containing both NCU00801 and NCU00130 on multi-copy plasmids, had relatively slower xylose utilization rates than those observed in DA24-16-BT3, a potential reason being glucose repression (FIG. 74). Further adjustments of the cellobextrin transporter and β-glucosidase expression levels, or the identification of β-glucosidases with

133

reduced trans-glycosylation activities, may be able to reduce the accumulation of glucose and cellobextrin during co-fermentation.

Co-fermentation of xylose and cellobiose could also be achieved by mixed cultivation of two different yeast strains: the xylose-fermenting DA24-16 strain and the cellobiose-fermenting DA452BT (FIG. 75). As explained above, the yeast strain DA24-16 expressed the xylose-utilizing enzymes wild type xylose reductase (XYL1), mutant xylose reductase R276H (mXYL1), xylitol dehydrogenase (XYL2), and xylulokinase (XKS1) (Ex. 12; Table 17). D452BT was formed by engineering D452 to express the cellobextrin transporter NCU00801 and the β -glucosidase NCU00130. In the mixed culture, the DA24-16 strain took up xylose (xylose molecule shown as a green pentagon in FIG. 75a) and metabolized it using the enzymes XYL1 (wild type and mutant), XYL2, and XYL3, whereas the other strain D452BT was able to take up cellobiose (cellobiose molecule shown as two red hexagons in FIG. 75a) using the transporter NCU00801 and convert the cellobiose into glucose using the enzyme NCU00130. Hence, the mixed culture was able to co-ferment both xylose and cellobiose to produce ethanol (FIG. 75b).

This study demonstrated a novel strategy to allow co-fermentation of hexose and pentose sugars by *S. cerevisiae*. By combining an efficient xylose utilization pathway with a cellobextrin transport system, the problem caused by glucose repression was over-come. As a result, the engineered yeast co-fermented two non-metabolizable sugars in cellulosic hydrolysates synergistically into ethanol. The new co-fermentation method described herein advances lignocellulosic technologies on both the saccharification and fermentation fronts. Most traditional fungal cellulase cocktails are deficient in β -glucosidase and end the cellulose hydrolysis with cellobiose that is not fermented efficiently by yeast. As a result, extra β -glucosidase enzyme must be added to convert cellobiose into glucose. The cellobiose/xylose co-fermentation yeast makes it possible to use these cellulase cocktails with limited β -glucosidase activities, lowering enzyme usage and cost associated with the cellulose saccharification process. Further, the synergy between cellobiose and xylose co-fermentation significantly increases ethanol productivity, thus improving fermentation economics. The presence of a small amount of glucose from the pre-treatment and hydrolysis of lignocellulosic materials does not affect the capacity of the engineered yeast to convert hexose and pentose sugar mixtures into ethanol.

This study involved measuring the capacity of an engineered *S. cerevisiae* strain to ferment various mixtures of sugars meant to mimic hydrolysates from plant biomass. The ability of this strain to co-ferment cellobextrins and xylose is particularly useful during the simultaneous saccharification and co-fermentation (SSCF) of pre-treated plant biomass. During SSCF, hemicellulose would first be hydrolyzed by acid pre-treatment, resulting in formation of xylose and still-crystalline cellulose. Then, fungal cellulases and the yeast strain described herein would be added, allowing the cellulases to co-convert xylose and cellobiose into ethanol. Because of the limited extracellular glucose production in this scheme, there will be reduced repression of xylose utilization and co-fermentation will proceed rapidly and synergistically.

Although the *S. cerevisiae* strain used in this study was a laboratory strain, the fermentation performance of the engineered strain was very impressive when compared to published results. The key fermentation parameters (yield and productivity) may be further improved by the use of industrial yeast strains as a platform. Applications of this co-fermenta-

134

tion strategy would not be limited to ethanol production. Since it is a foundational technology, the strategy presented here can be combined with any other product diversification technologies to produce commodity chemicals and advanced biofuels.

Example 18

Transcriptome Analysis of *N. crassa* Grown on Xylan

Lignocellulosic biomass is composed of cellulose, hemicellulose, and lignin. Examples 1-3 describe the discovery of genes critical for growth on cellulose through transcriptome and secretome analysis of *N. crassa*. In this example the expression profile of the *N. crassa* genome was examined during growth on xylan to determine which genes are important for utilization of hemicellulose.

Ten day old conidia of WT or Δ lnR strains were inoculated at 10^6 conidia/mL on 100 mL 1 \times Vogel's salts minimal medium (2% sucrose), grown for 16 hours at 25° C. with constant light, and washed with 1 \times Vogel's only medium. Conidia were then transferred into 100 mL 1 \times Vogel's salts with 2% sucrose or 2% Beechwood xylan as the sole carbon source in the medium and allowed to grow for 4 hours. Mycelia were harvested by filtration and immediately flash frozen in liquid nitrogen. Total RNA was isolated using TRIzol (Invitrogen) according to the manufacturer's instructions and treated with DNase (Turbo DNA-free kit; Ambion) (Kasuga, Townsend et al., 2005).

For cDNA synthesis and labeling, the Pronto kit (Catalog No. 40076; Corning) was used according to the manufacturer's specifications except that the total RNA used was 10 μ g per sample.

Microarray hybridization and data analysis were performed as previously described (Tian, Kasuga et al., 2007). A GenePix 4000B scanner (Axon Instruments) was used to acquire images, and GenePix Pro6 software was used to quantify hybridization signals and collect the raw data. Normalized expression values were analyzed by using the BAGEL (Bayesian analysis of gene expression levels) software program (Townsend and Hartl 2002; Townsend 2004). 354 genes were found to be induced greater than 2-fold in *N. crassa* grown on xylan. The list is shown in FIG. 76.

Example 19

Secretome Analysis of *N. crassa* Grown on Xylan

The secretome of *N. crassa* during growth on xylan was analyzed using a shotgun proteomics approach. Supernatants from xylan cultures were digested with trypsin and analyzed by liquid chromatography nano-electrospray ionization tandem mass spectrometry.

Mass spectrometry samples were prepared as follows. *N. crassa* wild type strain was grown on 2% xylan media for 4 or 7 days. Culture supernatants were isolated by centrifugation, filtered through 0.22 μ m filters, and concentrated 10 times with 10 kDa MWCO PES spin concentrators. 3.36 mg of urea, 5 μ L of 1M Tris pH 8.5, and 5 μ L of 100 mM DTT were then added to 100 μ L of concentrated culture supernatant, and the mixture was heated at 60° C. for 1 hour. After heating, 700 μ L of 25 mM ammonium bicarbonate and 140 μ L of methanol were added to the solution followed by treatment with 50 μ L of 100 μ g/mL trypsin in 50 mM sodium acetate pH 5.0. The trypsin was left to react overnight at 37° C. with inverting for about 8-9 hours at basal pH. After digestion the volume was

135

reduced to dryness by speedvac and washed with 300 μ L MilliQ water three times. The final volume was 100 μ L. TFA was added at 0.1-0.3% v/v. Residual salts in the sample were removed by using OMIX microextraction pipette tips according to the manufacturer's instructions. The acetonitrile was removed by evaporation. The sample solution was an aqueous solution with 0.1%-1% TFA, and the final volume was 10 microliters or greater.

Example 20

Analysis of Xylan-Induced Genes Predicted to Encode Secreted Proteins

The transcriptome and secretome analysis results indicated a total of 71 genes, of which 55 were predicted to be secreted. The list of these genes is in Table 29. Deletion strains were available for 46 out of 69 genes. Out of these 46, six of the strains were heterokaryons, thus the remaining 40 deletion strains were analyzed for total secreted protein, amount of xylose present, and azo-endo-xylanase activity. Results are shown in FIG. 77.

Gene Name	Signal P Data	Annotation
NCU00642	Y	Transcription probable beta-galactosidase
NCU00695	Y	Transcription putative protein
NCU00798	MS	hypothetical protein
NCU00937	Y	Transcription conserved hypothetical protein
NCU01517	Y	Transcription glucan 1,4-alpha-glucosidase
NCU02136	MS	probable transaldolase
NCU02252	MS	probable phosphoglyceromutase
NCU02343	Y	Transcription related to alpha-L-arabinofuranosidase A precursor
NCU02455	Y	Transcription FK506-binding protein 2 precursor (Peptidyl-prolyl cis-trans isomerase)
NCU02583	Y	Transcription probable Alpha-glucosidase precursor (Maltase)
NCU03013	Y	Transcription related to cytosolic Cu/Zn superoxide dismutase
NCU03222	Y	Transcription putative protein
NCU03636	Y	Transcription
NCU03639	Y	Transcription probable triacylglycerol lipase precursor
NCU04202	MS	nucleoside-diphosphate kinase
NCU04265	Y	Transcription related to beta-fructofuranosidase
NCU04388	Y	Transcription probable phosphatidylglycerol/phosphatidylinositol transfer protein
NCU04395	MS	beta-1,6-glucanase Neg1 NEG-1
NCU04415	Y	Transcription related to brefeldin A resistance protein
NCU04431	Y	MS related to endo-1,3-beta-glucanase
NCU04475	Y	Transcription probable lipase B precursor
NCU04482	MS	hypothetical protein
NCU04623	Y	Transcription related to beta-galactosidase
NCU04674	Y	Transcription related to alpha-glucosidase b
NCU04675	Y	Transcription putative protein
NCU04930	Y	Transcription related to triacylglycerol lipase
NCU05137	Y	Transcription conserved hypothetical protein
NCU05143	Y	Transcription related to Rds1 protein
NCU05159	Y	Transcription probable acetylxyran esterase precursor
NCU05275	MS	probable ubiquitin fusion protein (ubiquitin/ribosomal protein)
NCU05315	Y	Transcription hypothetical protein
NCU05395	Y	Transcription conserved hypothetical protein
NCU05686	MS	probable cell wall protein UTR2
NCU05751	Y	Transcription related to acetylxyran esterase
NCU05924	Y	Transcription probable endo-beta-1,4-D-xylanase
NCU05965	Y	Transcription related to putative arabinase

136

-continued

Gene Name	Signal P Data	Annotation
NCU05974	MS	related to cell wall protein (putative glycosidase)
NCU06364	Y	Transcription hypothetical protein
NCU06380	Y	Transcription related to catecholamines up protein
NCU06650	Y	Transcription conserved hypothetical protein
NCU06781	MS	probable beta (1-3) glucanosyltransferase
10 NCU06961	Y	Transcription probable exopolysaccharide
NCU07067	MS	related to class I alpha-mannosidase 1B
NCU07143	Y	Transcription
NCU07190	Y	Transcription related to cellulose 1,4-beta-cellulobiosidase II precursor
15 NCU07200	Y	MS related to metalloprotease MEP1
NCU07225	Y	Transcription probable endo-1,4-beta-xylanase B precursor
NCU07281	MS	probable glucose-6-phosphate isomerase
20 NCU07787	Y	MS probable SnodProt1 precursor
NCU08131	Y	Transcription probable alpha-amylase precursor
NCU08171	Y	MS conserved hypothetical protein
NCU08189	Y	Transcription related to endo-1,4-beta-xylanase
NCU08384	MS	probable D-xylose reductase
NCU08418	MS	related to tripeptidyl-peptidase I
NCU08457	Y	Transcription hydrophobin Ccg-2 CCG-2
25 NCU08516	Y	Transcription related to aldoze 1-epimerase
NCU08750	Y	Transcription related to isoamyl alcohol oxidase
NCU08752	Y	Transcription related to esterase
NCU08755	Y	Transcription hypothetical protein
NCU08909	Y	MS probable beta (1-3) glucanosyltransferase gel3p
30 NCU08936	MS	related to sporulation-specific gene SPS2
NCU09024	Y	MS related to choline dehydrogenase
NCU09133	Y	Transcription putative protein
NCU09170	Y	MS probable alpha-N-arabinofuranosidase
35 NCU09175	Y	Transcription related to glucan 1,3-beta-glucosidase precursor
NCU09267	MS	related to glyoxal oxidase precursor
NCU09491	MS	feruloyl esterase B precursor (subclass of the carboxylic acid esterases)
40 NCU09923	Y	Transcription related to xylan 1,4-beta-xylosidase
NCU09924	Y	Transcription conserved hypothetical protein
NCU10040	Y	Transcription
NCU10045	Y	Transcription

45 Samples were prepared as follows. 10 day old conidia were grown in 100 mL 2% xylan Vogel's media at 10^6 conidia/mL. Two replicates were prepared for each strain. Cultures were grown at 25° C. with constant light and 220 rpm. Samples were harvested on day 4. Supernatants were isolated by centrifugation and used in assays.

Bradford protein concentrations were measured to determine the total amount of secreted protein. Stocks were prepared with BSA standards: 0 μ g/mL, 50 μ g/mL, 100 μ g/mL, 55 250 μ g/mL, and 500 μ g/mL. Bradford solution was diluted 1:4. A multichannel pipette was used to pipette 200 μ L of Bradford solution into a 96-well plate. 10 μ L of sample and 10 μ L of each standard were added. Samples were incubated at room temperature for 10 minutes. The absorbance was read at 595 nm, and the protein concentration was determined.

60 The assay used to measure xylose was modified from Bailey et al., 1992 (*J Biotech* 23: 257-270). Xylose standards were prepared in H₂O. For concentrated 0.8 M xylose (1.2 g in 10 mL), the standards included 0 mM, 8 mM (1:100 dilution; 990 μ L+10 μ L), 20 mM (1:100 dilution; 975 μ L+25 μ L), 40 mM (1:100 dilution; 950 μ L+50 μ L), 80 mM (1:100 dilution; 900 μ L+100 μ L), and 160 mM (1:100 dilution; 800 μ L+200 μ L).

137

A multichannel pipette was used to add 900 μ L of substrate solution to a deep well 96-well plate. The substrate was allowed to incubate at 50° C. for 10 minutes. One hundred μ L of culture supernatant and the standards were added and allowed to incubate at 50° C. for 5 minutes. Samples were centrifuged for 10 minutes at 3,400 rpm. A multichannel pipette was used to pipette 75 μ L DNS solution into a 96-well PCR plate. Five μ L of solution was removed from the reaction and added to the PCR plate containing DNS solution. The plate was heated at 99° C. in the PCR machine for 5 min. After the samples cooled, they were transferred to clear flat-bottomed plates, and the absorbance was read at 540 nm. Substrate solution (500 mL) contained beechwood xylan (5 g; 10 mg/mL), 3M NaOAc, pH 5.0 (8.33 mL; 50 mM), water (491 mL), and was autoclaved for 20 minutes. DNS solution (100 mL) contained 3,5-dinitrosalicylic acid (707 mg), NaOH (1.32 g), Rochelle salts (Na K tartrate) (20.4 g), Sodium meta-bisulfate (553 mg), phenol (507 μ L), and water (94.4 mL).

Azo-endo-xylanase activity was measured with a kit from Megazyme. This assay indirectly measures the amount of endo-xylanase activity in a sample by spectrophotometrically measuring the amount of dye liberated from a xylan chain complexed with the dye. The more enzymes that are present, the more dye will be released. All supernatant samples were diluted 1:10 by adding 50 μ L of supernatant to 450 μ L of Na Acetate buffer (50 mM, pH 4.5) in separate 15 mL Falcon tubes. Next, Falcon tubes were pre-warmed about 10 minutes. Substrate solution was added for all samples (500 μ L/sample) to the tubes. Samples and substrate solutions were added into a 40° C. water bath for 10 minutes to pre-equilibrate them. Five hundred μ L substrate solution was added to each 1:10 diluted sample, vortexed for 10 seconds, and incubated at 40° C. for 10 minutes. The reaction was terminated by adding 2.5 mL of precipitant solution (95% ethanol) to each sample and vortexing for 10 seconds. Tubes were allowed to stand at room temperature for 10 minutes. Tubes were vortexed for 10 seconds and then centrifuged at room temperature for 10 minutes at 1,000 g. One mL of supernatant solution from each tube was placed directly into a cuvette, and the absorbance was measured at 590 nm. The blank used for this procedure was the supernatant from 500 μ L substrate solution added to 2.5 mL of precipitant solution.

In conclusion, it is anticipated that the modulation of genes identified here that affect the degradation of hemicellulose in *N. crassa* will facilitate engineering strains that have enhanced capacity for plant cell wall breakdown and growth on plant cell wall components such as hemicellulose. Genes of interest include NCU01517, which encodes a predicted glucamylase; NCU02343, which encodes a predicted arabinofuranosidase; NCU05137, which encodes a conserved hypothetical protein; NCU05159, which encodes a predicted acetylxyran esterase precursor; NCU09133, which encodes a conserved hypothetical protein; and NCU10040, which encodes a hypothetical protein.

The growth of a cell on hemicellulose will be increased by providing a host cell that contains a recombinant polynucleotide that encodes a polypeptide encoded by NCU01517, NCU09133, or NCU10040. The host cell will be cultured in a medium that contains hemicellulose such that the recombinant polynucleotide is expressed. The host cell will grow at a faster rate in this medium than a cell that does not contain the recombinant polynucleotide.

Example 21

Further Analysis of the ΔNCU05137 Strain

As described in Examples 1-3 and 18-20, NCU05137 is a predicted secreted protein that was overexpressed during

138

growth of *N. crassa* on any of *Miscanthus*, Avicel, or xylan. A deletion strain of *N. crassa* lacking NCU05137 grown on Avicel showed increased endoglucanase, β -glucosidase, and Avicelase activity. An NCU05137 deletion strain grown on xylan showed increased azo-endo-xylanase activity. As described in this example, the complementation of ΔNCU05137 was performed in order to verify that the phenotypes observed in the ΔNCU05137 strain were due to the loss of the NCU05137 gene.

10 A plasmid containing NCU05137 with a C-terminal GFP tag under the control of the *ccg1* promoter was generated. *N. crassa* conidia were transformed with the NCU05137-GFP construct. Experiments were performed according to standard *Neurospora* procedures (webpage fgsc.net/Neurospora/ 15 NeurosporaProtocolGuide.htm).

The total secreted protein and carboxymethyl cellulase (CMC) activity of wild-type, ΔNCU05137, and ΔNCU05137-NCU05137-GFP strains was measured. Total secreted protein was measured by taking 100 μ L of supernatant from a culture of each strain, adding it to 900 μ L Bradford Dye, and measuring absorbance at 595 nm. CMC activity was measured with 20 \times diluted supernatant from each strain culture and an azo-CMC kit (Megazyme SCMCL). ΔNCU05137 knockout strains displayed increased levels of secreted protein and CMC activity. Introduction of the GFP-tagged NCU05137 into ΔNCU05137 strains reduced these levels back to wild-type levels (FIG. 78).

In addition, the localization of NCU05137-GFP in complemented strains was observed. NCU05137-GFP localized to 30 the cell wall of conidia and to the hypha tip (FIG. 79-80). These data indicate that the GFP-tagged NCU05137 protein is fully functional and can be used for purification and experiments addressing the biochemical activity of this protein.

Thus, the normal function of NCU05137 may be to inhibit 35 signaling processes associated with induction of cellulase and hemicellulase gene expression. Reduction of expression of NCU05137 or a homolog of NCU05137 in a cell is likely to increase cellulase and hemicellulase activity in that cell and, consequently, growth of the cell on cellulose or hemicellulose. The growth of a cell on cellulose or hemicellulose will 40 be increased by providing a host cell that contains an endogenous polynucleotide that encodes a polypeptide encoded by NCU05137. The expression of the endogenous polynucleotide will be inhibited, and the cell will be cultured in a medium containing cellulose and/or hemicellulose. The host cell will grow at a faster rate in the medium than a cell in 45 which expression of the endogenous polynucleotide is not inhibited.

50

Example 22

Further Analysis of NCU07705

Expression of NCU07705 was found to be upregulated 55 during growth of *N. crassa* on cellulose. BLAST analysis of the polypeptide encoded by NCU07705 revealed that the polypeptide has high similarity to many C6 zinc finger domain containing transcription factors (FIG. 1). To further investigate the role of NCU07705 in the utilization of cellulose, the phenotype of a deletion strain lacking NCU07705 60 was evaluated.

The ΔNCU07705 strain was unable to grow on 2% cellulose (Avicel), PASC, or CMC as a sole carbon source (Table 30) but grew with similar kinetics to wild-type strain on 65 sucrose, xylan, and xylose. In order to determine whether NCU07705 plays a role in regulating expression of cellulases, the expression of cellulase and hemicellulase genes was

139

examined during growth of ΔNCU07705 on cellulose. Ten-day-old conidia from wild-type (FGSC 2489) and ΔNCU07705 strains were inoculated into Vogel's liquid MM (2% sucrose) (Vogel 1956) and grown for 16 hours. Mycelia were centrifuged, washed with 1× Vogel's salts, and then

140

ΔNCU07705. The host cell will be cultured in a medium that contains cellulose such that the recombinant polynucleotide is expressed. The host cell will grow at a faster rate in this medium than a cell that does not contain the recombinant polynucleotide.

	7705-switch ²	WT-switch ¹	Gene/locus name	GH Family	Class	up in Avi ³
No	15	NCU00762	5	endo-	31.5	
No	No	NCU03996	6	CBHII like		
No	168	NCU07190	6	CBHII like	119	
No	26	NCU09680	6	CBHII	251.3	
No	18	NCU04854	7	CBHII like	10.8	
No	3.8	NCU05057	7	CBHII like	7.4	
No	No	NCU05104	7	CBHII like		
No	93	NCU07340	7	CBHII	382.2	
No	2	NCU05121	45	endo-	17.2	
No	5.8	NCU00836	61	endo-	31	
No	3.7	NCU01050	61	endo-	382.1	
No	No	NCU01867	61	endo-		
No	49	NCU02240	61	endo-	84	
No	No	NCU02344	61	endo-	4.1	
No	6.1	NCU02916	61	endo-	17.7	
No	No	NCU03000	61	endo-		
No	17	NCU03328	61	endo-	23.8	
No	No	NCU05969	61	endo-	12.7	
No	No	NCU07520	61	endo-		
No	No	NCU07760	61	endo-		
No	103	NCU07898	61	endo-	230	
No	No	NCU07974	61	endo-		
No	25	NCU08760	61	endo-	44.7	

¹Expression levels of predicted cellulase genes from an *N. crassa* (ΔNCU07705) culture grown in Vogel's/sucrose for 16 hours, filtered, and resuspended in Vogel's/Avicel for 4 hours prior to RAN extraction.

²Expression levels of predicted cellulase gene from an *N. crassa* (wild type FGSC 2489) culture grown in Vogel's/sucrose for 16 hours, filtered, and resuspended in Vogel's/sucrose for 4 hours prior to RNA extraction.

³Expression levels derived from microarray analyses of wild type (FGSC 2489) cells grown for 30 hours in Avicel (Tian et al., 2009).

transferred into either Vogel's media with 2% sucrose or 2% Avicel and grown in constant light for 4 hours. They were harvested by filtration and immediately frozen in liquid nitrogen. Total RNA was isolated using TRIzol (Invitrogen, Carlsbad, Calif.) according to the manufacturer's instructions and treated with DNase (Turbo DNA-free kit, Ambion/Applied Biosystems, Foster City, Calif.) (Kasuga et al., 2005). Chip-Shot™ Indirect Labeling/Clean-Up System (Catalog No. Z4000, Promega, Madison, Wis.) and CyDye Post-Labeling Reactive Dye Pack (Catalog No. RPN5661, GE Healthcare, Piscataway, N.J.) were used to synthesize and label cDNA according to the manufacturer's instructions except the amount of RNA used was 10 µg. The Pronto! Hybridization Kit (Catalog No. 40076, Corning, Lowell, Mass.) was used for microarray hybridization according to the manufacturer's specifications.

Data analyses were performed as previously described (Tian et al., 2007). A GenePix 4000B scanner (Axon Instruments, Union City, Calif.) was used to acquire images, and GenePix Pro6 software was used to quantify hybridization signals and collect the raw data. Normalized expression values were analyzed by using BAGEL (Bayesian Analysis of Gene Expression Levels) (Townsend and Hartl, 2002). None of the predicted cellulase genes were induced in the ΔNCU07705 strain, whereas induction of predicted hemicellulase genes was unaffected (see Table 30 below). Thus, ΔNCU07705 has been named cdr-1, cellulose degradation regulator 1.

Therefore, the growth of a cell on cellulose will be increased by providing a host cell that contains a recombinant polynucleotide that encodes a polypeptide encoded by

- 35
- 40
- 45
- 50
- 55
- 60
- 65
- 70
- 75
- 80
- 85
- 90
- 95
- 100
- 105
- 110
- 115
- 120
- 125
- 130
- 135
- 140
- 145
- 150
- 155
- 160
- 165
- 170
- 175
- 180
- 185
- 190
- 195
- 200
- 205
- 210
- 215
- 220
- 225
- 230
- 235
- 240
- 245
- 250
- 255
- 260
- 265
- 270
- 275
- 280
- 285
- 290
- 295
- 300
- 305
- 310
- 315
- 320
- 325
- 330
- 335
- 340
- 345
- 350
- 355
- 360
- 365
- 370
- 375
- 380
- 385
- 390
- 395
- 400
- 405
- 410
- 415
- 420
- 425
- 430
- 435
- 440
- 445
- 450
- 455
- 460
- 465
- 470
- 475
- 480
- 485
- 490
- 495
- 500
- 505
- 510
- 515
- 520
- 525
- 530
- 535
- 540
- 545
- 550
- 555
- 560
- 565
- 570
- 575
- 580
- 585
- 590
- 595
- 600
- 605
- 610
- 615
- 620
- 625
- 630
- 635
- 640
- 645
- 650
- 655
- 660
- 665
- 670
- 675
- 680
- 685
- 690
- 695
- 700
- 705
- 710
- 715
- 720
- 725
- 730
- 735
- 740
- 745
- 750
- 755
- 760
- 765
- 770
- 775
- 780
- 785
- 790
- 795
- 800
- 805
- 810
- 815
- 820
- 825
- 830
- 835
- 840
- 845
- 850
- 855
- 860
- 865
- 870
- 875
- 880
- 885
- 890
- 895
- 900
- 905
- 910
- 915
- 920
- 925
- 930
- 935
- 940
- 945
- 950
- 955
- 960
- 965
- 970
- 975
- 980
- 985
- 990
- 995
- 1000
- 1005
- 1010
- 1015
- 1020
- 1025
- 1030
- 1035
- 1040
- 1045
- 1050
- 1055
- 1060
- 1065
- 1070
- 1075
- 1080
- 1085
- 1090
- 1095
- 1100
- 1105
- 1110
- 1115
- 1120
- 1125
- 1130
- 1135
- 1140
- 1145
- 1150
- 1155
- 1160
- 1165
- 1170
- 1175
- 1180
- 1185
- 1190
- 1195
- 1200
- 1205
- 1210
- 1215
- 1220
- 1225
- 1230
- 1235
- 1240
- 1245
- 1250
- 1255
- 1260
- 1265
- 1270
- 1275
- 1280
- 1285
- 1290
- 1295
- 1300
- 1305
- 1310
- 1315
- 1320
- 1325
- 1330
- 1335
- 1340
- 1345
- 1350
- 1355
- 1360
- 1365
- 1370
- 1375
- 1380
- 1385
- 1390
- 1395
- 1400
- 1405
- 1410
- 1415
- 1420
- 1425
- 1430
- 1435
- 1440
- 1445
- 1450
- 1455
- 1460
- 1465
- 1470
- 1475
- 1480
- 1485
- 1490
- 1495
- 1500
- 1505
- 1510
- 1515
- 1520
- 1525
- 1530
- 1535
- 1540
- 1545
- 1550
- 1555
- 1560
- 1565
- 1570
- 1575
- 1580
- 1585
- 1590
- 1595
- 1600
- 1605
- 1610
- 1615
- 1620
- 1625
- 1630
- 1635
- 1640
- 1645
- 1650
- 1655
- 1660
- 1665
- 1670
- 1675
- 1680
- 1685
- 1690
- 1695
- 1700
- 1705
- 1710
- 1715
- 1720
- 1725
- 1730
- 1735
- 1740
- 1745
- 1750
- 1755
- 1760
- 1765
- 1770
- 1775
- 1780
- 1785
- 1790
- 1795
- 1800
- 1805
- 1810
- 1815
- 1820
- 1825
- 1830
- 1835
- 1840
- 1845
- 1850
- 1855
- 1860
- 1865
- 1870
- 1875
- 1880
- 1885
- 1890
- 1895
- 1900
- 1905
- 1910
- 1915
- 1920
- 1925
- 1930
- 1935
- 1940
- 1945
- 1950
- 1955
- 1960
- 1965
- 1970
- 1975
- 1980
- 1985
- 1990
- 1995
- 2000
- 2005
- 2010
- 2015
- 2020
- 2025
- 2030
- 2035
- 2040
- 2045
- 2050
- 2055
- 2060
- 2065
- 2070
- 2075
- 2080
- 2085
- 2090
- 2095
- 2100
- 2105
- 2110
- 2115
- 2120
- 2125
- 2130
- 2135
- 2140
- 2145
- 2150
- 2155
- 2160
- 2165
- 2170
- 2175
- 2180
- 2185
- 2190
- 2195
- 2200
- 2205
- 2210
- 2215
- 2220
- 2225
- 2230
- 2235
- 2240
- 2245
- 2250
- 2255
- 2260
- 2265
- 2270
- 2275
- 2280
- 2285
- 2290
- 2295
- 2300
- 2305
- 2310
- 2315
- 2320
- 2325
- 2330
- 2335
- 2340
- 2345
- 2350
- 2355
- 2360
- 2365
- 2370
- 2375
- 2380
- 2385
- 2390
- 2395
- 2400
- 2405
- 2410
- 2415
- 2420
- 2425
- 2430
- 2435
- 2440
- 2445
- 2450
- 2455
- 2460
- 2465
- 2470
- 2475
- 2480
- 2485
- 2490
- 2495
- 2500
- 2505
- 2510
- 2515
- 2520
- 2525
- 2530
- 2535
- 2540
- 2545
- 2550
- 2555
- 2560
- 2565
- 2570
- 2575
- 2580
- 2585
- 2590
- 2595
- 2600
- 2605
- 2610
- 2615
- 2620
- 2625
- 2630
- 2635
- 2640
- 2645
- 2650
- 2655
- 2660
- 2665
- 2670
- 2675
- 2680
- 2685
- 2690
- 2695
- 2700
- 2705
- 2710
- 2715
- 2720
- 2725
- 2730
- 2735
- 2740
- 2745
- 2750
- 2755
- 2760
- 2765
- 2770
- 2775
- 2780
- 2785
- 2790
- 2795
- 2800
- 2805
- 2810
- 2815
- 2820
- 2825
- 2830
- 2835
- 2840
- 2845
- 2850
- 2855
- 2860
- 2865
- 2870
- 2875
- 2880
- 2885
- 2890
- 2895
- 2900
- 2905
- 2910
- 2915
- 2920
- 2925
- 2930
- 2935
- 2940
- 2945
- 2950
- 2955
- 2960
- 2965
- 2970
- 2975
- <p

141

- Bouffard, G. G., K. E. Rudd, and S. L. Adhya (1994). "Dependence of lactose metabolism upon mutarotase encoded in the gal operon in *Escherichia coli*." *J Mol Biol* 244(3):269-278.
- Bouws, H. A. Wattenberg and H. Zorn (2008). "Fungal secretomes—nature's toolbox for white biotechnology." *Appl Microbiol Biotechnol* 80(3):381-388.
- Brat, D., E. Boles, and B. Wiedemann (2009). "Functional Expression of a Bacterial Xylose Isomerase in *Saccharomyces cerevisiae*." *Appl Environ Microbiol* 75(8):2304-2311.
- Canevacini, G. (1988). "Cellulobiose dehydrogenase from *Sporotrichum thermophile*." *Method Enzymol* 160:443-448.
- Cantarel, B. L., P. M. Coutinho, et al. (2009). "The Carbohydrate-Active EnZymes database (CAZy): an expert resource for glycogenomics." *Nucleic Acids Res* 37(Database issue):D233-238.
- Chauve, M., H. Mathis, et al. (2010). "Comparative kinetic analysis of two fungal beta-glucosidases." *Biotechnol Biofuels* 3(1):3.
- Christakopoulos, P., M. K. Bhat, et al. (1994). "Enzymatic synthesis of trisaccharides and alkyl beta-D-glucosides by the transglycosylation reaction of beta-glucosidase from *Fusarium oxysporum*." *Int J Biol Macromol* 16(6):331-334.
- Christianson, T. W., R. S. Sikorski, et al., (1992). "Multi-functional yeast high-copy number shuttle vectors." *Gene*, 110(1):119-122.
- Chu, B. C. and H. Lee (2007). "Genetic improvement of *Saccharomyces cerevisiae* for xylose fermentation." *Bio-technol Adv* 25(5):425-441.
- Colot, H. V., G. Park, et al. (2006). "A high-throughput gene knockout procedure for *Neurospora* reveals functions for multiple transcription factors." *Proc Natl Acad Sci USA* 103(27):10352-10357.
- Cullen, D., L. J. Wilson, et al. (1987). "Sequence and centromere proximal location of a transformation enhancing fragment ans1 from *Aspergillus nidulans*." *Nucleic Acids Res* 15(22):9163-9175.
- Davies, G. J., A. M. Brzozowski, et al. (2000). "Structure and function of *Humicola insolens* family 6 cellulases: structure of the endoglucanase, Ce16B, at 1.6 Å resolution." *Biochem J* 348 Pt 1:201-207.
- Davis, R. H. (2000). "*Neurospora*: Contributions of a model organism." New York, Oxford University Press.
- Davis, R. H. and D. D. Perkins (2002). "Timeline: *Neurospora*: a model of model microbes." *Nat Rev Genet* 3(5):397-403.
- de Groot, M. J., W. Prathumpai, et al. (2005). "Metabolic control analysis of *Aspergillus niger* L-arabinose catabolism." *Biotechnol Prog* 21(6):1610-1616.
- Dementhon K, G. Iyer, and N. L. Glass (2006). "VIB-1 is required for expression of genes necessary for programmed cell death in *Neurospora crassa*." *Eukaryot Cell* 5(12):2161-2173.
- Doran-Peterson, J., A. Jangid, et al. (2009). "Simultaneous saccharification and fermentation and partial saccharification and co-fermentation of lignocellulosic biomass for ethanol production." *Methods Mol Biol* 581:263-280.
- Driissen, R. E. T., R. H. W. Mass, et al. (2009). "Modelling ethanol production from cellulose: separate hydrolysis and fermentation versus simultaneous saccharification and fermentation." *Biocatal Biotransfor* 27(1):27-35.

142

- Dunlap, J. C., K. A. Borkovich, et al. (2007). "Enabling a Community to Dissect an Organism: Overview of the *Neurospora* Functional Genomics Project." *Adv Genet* 57:49-96.
- du Preez, J. C., M. Bosch, and B. A. Prior (1986). "The fermentation of hexose and pentose sugars by *Candida shehatae* and *Pichia stipitis*." *Appl Microbiol Biotechnol* 23(3):228-233.
- Eberhart, B. M., R. S. Beck, et al. (1977). "Cellulase of *Neurospora crassa*." *J Bacteriol* 130(1):181-186.
- Ellis, K. J. and J. F. Morrison (1982). "Buffers of constant ionic strength for studying pH-dependent processes." *Methods Enzymol* 87:405-426.
- Espagne, E., O. Lespinet, et al. (2008). "The genome sequence of the model ascomycete fungus *Podospora anserina*." *Genome Biol* 9(5):R77.
- Fonseca, C. R. Romao, et al. (2007). "L-Arabinose transport and catabolism in yeast." *FEBS J* 274(14):3589-3600.
- Freer, S. N. (1991). "Fermentation and aerobic metabolism of cellobextrins by yeasts." *Appl Environ Microbiol* 57(3):655-659.
- Freer, S. N. and R. V. Greene (1990). "Transport of glucose and cellobiose by *Candida wickerhamii* and *Clavispora lusitaniae*." *J Biol Chem* 265(22):12864-12868.
- Galagan, J. E., S. E. Calvo, et al. (2003). "The genome sequence of the filamentous fungus *Neurospora crassa*." *Nature* 422(6934):859-868.
- Gems, D., I. L. Johnstone, and A. J. Clutterbuck (1991). "An autonomously replicating plasmid transforms *Aspergillus nidulans* at high frequency." *Gene* 98(1):61-67.
- Guindon, S., and O. Gascuel (2003). "A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood." *Syst Biol* 52(5):696-704.
- Hahn-Hagerdal, B., K. Karhumaa, et al. (2007). "Towards industrial pentose-fermenting yeast strains." *Appl Microbiol Biotechnol* 74(5):937-953.
- Hector, R. E., N. Qureshi et al. (2008). "Expression of a heterologous xylose transporter in a *Saccharomyces cerevisiae* strain engineered to utilize xylose improves aerobic xylose consumption." *Appl Microbiol Biotechnol* 80(4):675-684.
- Himmel, M. E., S. Y. Ding, et al. (2007). "Biomass recalcitrance: engineering plants and enzymes for biofuels production." *Science* 315(5813):804-807.
- Hosaka, K., J. Nikawa, et al., (1992). "A dominant mutation that alters the regulation of INO1 expression in *Saccharomyces cerevisiae*." *J Biochem* 111(3):352-358.
- Jacobson, D. J., A. J. Powell, et al. (2004). "*Neurospora* in temperate forests of western North America." *Mycologia* 96(1):66-74.
- Jeffries, T. W., I. V. Grigoriev, et al. (2007). "Genome sequence of the lignocellulose-bioconverting and xylose-fermenting yeast *Pichia stipitis*." *Nat Biotechnol* 25(3):319-326.
- Kasuga, T. and N. L. Glass (2008). "Dissecting colony development of *Neurospora crassa* using mRNA profiling and comparative genomics approaches." *Eukaryot Cell* 7(9):1549-64.
- Kasuga, T., J. P. Townsend, et al. (2005). "Long-oligomer microarray profiling in *Neurospora crassa* reveals the transcriptional program underlying biochemical and physiological events of conidial germination." *Nucleic Acids Res* 33(20):6469-85.
- Katahira, S., M. Ito, et al. (2008). "Improvement of ethanol productivity during xylose and glucose co-fermentation by xylose-assimilating *S. cerevisiae* via expression of glucose transporter Sut1." *Enzyme Microb Tech* 43(2):115-119.

- Katahira, S., A. Mizuike, et al. (2006). "Ethanol fermentation from lignocellulosic hydrolysates by a recombinant xylose- and cellobiosaccharide-assimilating yeast strain." *Appl Microbiol Biotechnol* 72(6):1136-1143.
- Korkhin, Y., A. J. Kalb(Gilboa), et al. (1998). "NADP-dependent bacterial alcohol dehydrogenases: crystal structure, cofactor-binding and cofactor specificity of the ADHs of *Clostridium beijerinckii* and *Thermoanaerobacter brockii*." *J Mol Biol* 278(5):967-981.
- Kotaka, A., H. Bando, et al. (2008). "Direct ethanol production from barley beta-glucan by sake yeast displaying *Aspergillus oryzae* beta-glucosidase and endoglucanases." *J Biosci Bioeng* 105(6):622-627.
- Kötter, P. and M. Ciriacy (1993). "Xylose fermentation by *Saccharomyces cerevisiae*." *Appl Microbiol Biotechnol* 38(6):776-783.
- Kubicek, C. P., R. Messner, et al. (1993). "Triggering of cellulase biosynthesis by cellulose in *Trichoderma reesei*. Involvement of a constitutive, sophorose-inducible, glucose-inhibited beta-diglucoside permease." *J Biol Chem* 268(26):19364-19368.
- Kumar, R., S. Singh, and O. V. Singh (2008). "Bioconversion of lignocellulosic biomass: biochemical and molecular perspectives." *J Ind Microbiol Biotechnol* 35(5):377-391.
- Kuyper, M., H. R. Harhangi, et al. (2003). "High-level functional expression of a fungal xylose isomerase: the key to efficient ethanolic fermentation of xylose by *Saccharomyces cerevisiae*?" *FEMS Yeast Res* 4(1):69-78.
- Lang, J. M. and V. P. Cirillo (1987). "Glucose transport in a kinaseless *Saccharomyces cerevisiae* mutant." *J Bacteriol* 169(7):2932-2937.
- Leandro, M. J., P. Gonçalves, and I. Spencer-Martins (2006). "Two glucose/xylose transporter genes from the yeast *Candida intermedia*: first molecular characterization of a yeast xylose-H⁺ symporter." *Biochem J* 395(3):543-549.
- Linder, M. and T. T. Teeri (1996). "The cellulose-binding domain of the major cellobiohydrolase of *Trichoderma reesei* exhibits true reversibility and a high exchange rate on crystalline cellulose." *Proc Natl Acad Sci USA* 93(22): 12251-12255.
- Lynd, L. R., M. S. Laser, et al. (2008). "How biotech can transform biofuels." *Nat. Biotechnol.* 26(2):169-172.
- Lynd, L. R., P. J. Weimer, et al. (2002). "Microbial cellulose utilization: fundamentals and biotechnology." *Microbiol Mol Biol Rev* 66(3):506-77, table of contents.
- Madhavan, A., S. Tamalampudi, et al. (2009). "Xylose isomerase from polycentric fungus orpinomyces: gene sequencing, cloning, and expression in *Saccharomyces cerevisiae* for bioconversion of xylose to ethanol." *Appl Microbiol Biotechnol* 82(6):1067-1078.
- Martin, F., A. Kohler, et al. (2010). "Périgord black truffle genome uncovers evolutionary origins and mechanisms of symbiosis." *Nature* 464(7291):1033-1038.
- Martinez, D., R. M. Berka, et al. (2008). "Genome sequencing and analysis of the biomass-degrading fungus *Trichoderma reesei* (syn. *Hypocrea jecorina*)."*Nat Biotechnol* 26(5):553-60.
- Martinez, D., J. Callacombe, et al. (2009). "Genome, transcriptome, and secretome analysis of wood decay fungus *Postia placenta* supports unique mechanisms of lignocellulose conversion." *Proc Natl Acad Sci USA* 106(6):1954-1959.
- Martinez, D., L. F. Larondo, et al. (2004). "Genome sequence of the lignocellulose degrading fungus *Phanerochaete chrysosporium* strain RP78." *Nat Biotechnol* 22(6): 695-700.

- Matsushika, A., S. Watanabe, et al. (2008). "Expression of protein engineered NADP+-dependent xylitol dehydrogenase increases ethanol production from xylose in recombinant *Saccharomyces cerevisiae*." *Appl Microbiol Biotechnol* 81(2):243-255.
- Matteucci, M. D. and M. H. Caruthers (1980). *Tetrahedron Lett* 21:719-722.
- McCluskey, K. (2003). "The Fungal Genetics Stock Center: from molds to molecules." *Adv Appl Microbiol* 52:245-262.
- Medve, J., J. Karlsson, et al. (1998). "Hydrolysis of microcrystalline cellulose by cellobiohydrolase I and endoglucanase II from *Trichoderma reesei*: adsorption, sugar production pattern, and synergism of the enzymes." *Biotechnol Bioeng* 59(5):621-34.
- Miyasaka, H. (1999). "The positive relationship between codon usage bias and translation initiation AUG context in *Saccharomyces cerevisiae*." *Yeast* 15(8):633-637.
- Nair, N. and H. Zhao (2007). "Biochemical characterization of an L-Xylulose reductase from *Neurospora crassa*." *Appl Environ Microbiol* 73(6):2001-2004.
- Nakamura, N., R. Yamada, et al. (2008). "Effective xylose/cellobiose co-fermentation and ethanol production by xylose-assimilating *S. cerevisiae* via expression of β-glucosidase on its cell surface." *Enzyme Microb Tech* 43(3): 233-236.
- Noguchi, Y., M. Sano, et al. (2009). "Genes regulated by AoXlnR, the xylanolytic and cellulolytic transcriptional regulator, in *Aspergillus oryzae*." *Appl Microbiol Biotechnol* 85(1):141-154.
- Notredame, C., D. G. Higgins, and J. Heringa (2000). "T-Coffee: A novel method for fast and accurate multiple sequence alignment." *J Mol Biol* 302(1):205-217.
- Pandit, A. and R. Maheshwari (1996). "Life-history of *Neurospora intermedia* in a sugar cane field." *J Biosci (Bangalore)* 21(1):57-79.
- Pauly, T. A., J. L. Ekstrom, et al. (2003). "X-ray crystallographic and kinetic studies of human sorbitol dehydrogenase." *Structure* 11(9):1071-1085.
- Pedelacq, J. D., S. Cabantous, et al. (2006). "Engineering and characterization of a superfolder green fluorescent protein." *Nat Biotechnol* 24(1):79-88.
- Perkins, D. D., B. C. Turner, et al. (1976). "Strains of *Neurospora* collected from nature." *Evolution* 30: 281-313.
- Ramos, J., K. Szkutnicka, and V. P. Cirillo (1988). "Relationship between low- and high-affinity glucose transport systems of *Saccharomyces cerevisiae*." *J Bacteriol* 170(11): 5375-5377.
- Rawat, U. and M. Rao (1997). "Site and significance of cysteine residues in xylose reductase from *Neurospora crassa* as deduced by fluorescence studies." *Biochemical and Biophysical Research Communications* 239(3): 789-93.
- Reifenberger E., E. Boles, and M. Ciriacy (1997). "Kinetic characterization of individual hexose transporters of *Saccharomyces cerevisiae* and their relation to the triggering mechanisms of glucose repression." *Eur J Biochem* 245(2): 324-333.
- Romero, M. D., J. Aguado, et al. (1999). "Cellulase production by *Neurospora crassa* on wheat straw." *Enzyme Microb Tech* 25: 244-250.
- Roy, A., A. Kucukural, and Y. Zhang (2010). "I-TASSER: a unified platform for automated protein structure and function prediction." *Nat Protoc* 5(4):725-738.
- Rubin E. M. (2008). "Genomics of cellulosic biofuels." *Nature* 454(7206):841-845.

145

Ruepp, A., A. Zollner, et al. (2004). "The FunCat, a functional annotation scheme for systematic classification of proteins from whole genomes." *Nucleic Acids Res* 32(18):5539-5545.

Runquist, D., C. Fonseca, et al. (2009). "Expression of the Gxf1 transporter from *Candida intermedia* improves fermentation performance in recombinant xylose-utilizing *Saccharomyces cerevisiae*." *Appl Microbiol Biotechnol* 82(1):123-130.

Runquist, D., B. Hahn-Hägerdal, and P. Rådström (2010). "Comparison of heterologous xylose transporters in recombinant *Saccharomyces cerevisiae*." *Biotechnol Biofuels* 3:5.

Saha, B. C. (2003). "Hemicellulose bioconversion." *J Ind Microbiol Biotechnol* 30:279-291.

Saloheimo, A., J. Rauta, et al. (2007). "Xylose transport studies with xylose-utilizing *Saccharomyces cerevisiae* strains expressing heterologous and homologous permeases." *Appl Microbiol Biotechnol* 74(5):1041-1052.

Sarkar, G. and S. S. Sommer (1990). "The "megaprimer" method of site-directed mutagenesis." *Biotechniques* 8(4): 404-407.

Sauer, U. (2001). "Evolutionary engineering of industrially important microbial phenotypes." *Adv Biochem Eng Biotechnol* 73:129-169.

Scarborough, G. A. (1973). "Transport in *Neurospora*." *Int Rev Cytol* 34:103-122.

Seibold, B., S. Hakola, et al. (1997). "Role of four major cellulases in triggering of cellulase gene expression by cellulose in *Trichoderma reesei*." *J Bacteriol* 179(17): 5318-5320.

Seidl, V., C. Gamauf, et al. (2008). "The *Hypocrea jecorina* (*Trichoderma reesei*) hypercellulolytic mutant RUT C30 lacks a 85 kb (29 gene-encoding) region of the wild-type genome." *BMC Genomics* 9:327.

Shao, Z., H. Zhao, and H. Zhao (2009). "DNA assembler, an in vivo genetic method for rapid construction of biochemical pathways." *Nucleic Acids Res.* 37(2):e16.

Sikorski, R. S., and P. Hieter (1989). "A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in *Saccharomyces cerevisiae*." *Genetics* 122(1):19-27.

Skory, C. D., S. N. Freer, and R. J. Bothast (1996). "Expression and secretion of the *Candida wickerhamii* extracellular beta-glucosidase gene, bglB, in *Saccharomyces cerevisiae*." *Curr Genet* 30(5):417-422.

Smith, M. L., O. C. Micali, et al. (2000). "Vegetative incompatibility in the het-6 region of *Neurospora crassa* is mediated by two linked genes." *Genetics* 155(3):1095-1104.

Stambuk, B. U., M. A. Franden, et al. (2003). "D-Xylose transport by *Candida succiphila* and *Kluyveromyces marxianus*." *Appl Biochem Biotechnol* 105-108:255-263.

Stephanopoulos, G. (2007). "Challenges in Engineering Microbes for Biofuels Production." *Science* 315(5813): 801-804.

Stricker A. R., R. L. Mach and L. H. de Graaff (2008). "Regulation of transcription of cellulases- and hemicellulases-encoding genes in *Aspergillus niger* and *Hypocrea jecorina* (*Trichoderma reesei*)." *Appl Microbiol Biotechnol* 78(2):211-220.

Sullivan, R. and H. Zhao (2007). "Cloning, characterization, and mutational analysis of a highly active and stable L-arabinitol 4-dehydrogenase from *Neurospora crassa*." *Appl Microbiol Biotechnol* 77(4):845-52.

Sun Y. and J. Cheng (2002). "Hydrolysis of lignocellulosic materials for ethanol production: a review." *Bioresource Technol* 83(1):1-11.

146

Suominen P. L., A. L. Mantyla, et al. (1993). "High frequency one-step gene replacement in *Trichoderma reesei*. II. Effects of deletions of individual cellulase genes." *Mol Gen Genet* 241(5-6):523-530.

Tian, C., W. T. Beeson, et al. (2009). "Systems analysis of plant cell wall degradation by the model filamentous fungus *Neurospora crassa*." *Proc Natl Acad Sci USA* 106(52): 22157-22162.

Tian, C., T. Kasuga, et al. (2007). "Transcriptional profiling of cross pathway control in *Neurospora crassa* and comparative analysis of the Gcn4 and CPC1 regulons." *Eukaryot Cell* 6(6): 1018-1029.

Townsend J. P. (2004). "Resolution of large and small differences in gene expression using models for the Bayesian analysis of gene expression levels and spotted DNA microarrays." *BMC Bioinformatics* 5:54.

Townsend J. P. and D. L. Hartl (2002). "Bayesian analysis of gene expression levels: statistical quantification of relative mRNA level across multiple strains or treatments." *Genome Biol* 3(12):research0071.1-0071.16.

USDA, The Economic Feasibility of Ethanol Production from Sugar in the United States, United States Department of Agriculture (2006).

van Rooyen, R., B. Hahn-Hägerdal, et al. (2005). "Construction of cellobiose-growing and fermenting *Saccharomyces cerevisiae* strains." *J Biotechnol* 120(3):284-295.

Vogel, H. J. (1956). "A convenient growth medium for *Neurospora*." *Microbiol. Genet. Bull.* 13:42-46.

Wagner, S., M. L. Bader, et al. (2006). "Rationalizing membrane protein overexpression." *Trends Biotechnol* 24(8): 364-371.

Watanabe, S., A. Abu Saleh, et al. (2007). "Ethanol production from xylose by recombinant *Saccharomyces cerevisiae* expressing protein-engineered NADH-preferring xylose reductase from *Pichia stipitis*." *Microbiology* 153 (Pt 9):3044-3054.

Watanabe, S., T. Kodaki, and K. Makino (2005a). "L-Arabinose 1-dehydrogenase: a novel enzyme involving in bacterial L-arabinose metabolism." *Nucleic Acids Symp Ser (Oxf)* 49:309-310.

Watanabe, S., T. Kodaki, and K. Makino (2005b). "Complete reversal of coenzyme specificity of xylitol dehydrogenase and increase of thermostability by the introduction of structural zinc." *J Biol Chem* 280(11):10340-10349.

Wiedemann, B. and E. Boles (2008). "Codon-optimized bacterial genes improve L-arabinose fermentation in recombinant *Saccharomyces cerevisiae*." *Appl Environ Microbiol* 74(7):2043-2050.

Wisselink, H. W., M. J. Toirkens, et al. (2007). "Engineering of *Saccharomyces cerevisiae* for efficient anaerobic alcoholic fermentation of L-arabinose." *Appl Environ Microbiol* 73(15):4881-4891.

Wisselink, H. W., M. J. Toirkens, et al. (2009). "Novel Evolutionary Engineering Approach for Accelerated Utilization of Glucose, Xylose, and Arabinose Mixtures by Engineered *Saccharomyces cerevisiae* Strains." *Appl Environ Microbiol* 75(4):907-914.

Woodyer, R., M. Simurdia, et al. (2005). "Heterologous expression, purification, and characterization of a highly active xylose reductase from *Neurospora crassa*." *Appl Environ Microbiol* 71(3): 1642-7.

Wymelenberg, A. V., et al. (2009). "Transcriptome and Secretome Analyses of *Phanerochaete Chrysosporium* Reveal Complex Patterns of Gene Expression." *Appl Environ Microbiol* 75(12):4058-4068.

Xin, Z., Q. Yinbo, and G. Peiji. (1993). "Acceleration of ethanol production from paper mill waste fiber by supplementation with β-glucosidase." *Enzyme Microb Tech* 15(1):62.

SEQUENCE LISTING

<160> NUMBER OF SEQ ID NOS: 97

<210> SEQ ID NO 1
<211> LENGTH: 18
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Sequence motif
<221> NAME/KEY: VARIANT
<222> LOCATION: 1
<223> OTHER INFORMATION: Xaa = Leu or Ile or Val or Met
<221> NAME/KEY: VARIANT
<222> LOCATION: 3
<223> OTHER INFORMATION: Xaa = Phe or Leu
<221> NAME/KEY: VARIANT
<222> LOCATION: 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16
<223> OTHER INFORMATION: Xaa = Any Amino Acid
<221> NAME/KEY: VARIANT
<222> LOCATION: 17
<223> OTHER INFORMATION: Xaa = Tyr or Phe

<400> SEQUENCE: 1

Xaa Tyr Xaa
1 5 10 15

Xaa Asp

<210> SEQ ID NO 2
<211> LENGTH: 17
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Sequence motif
<221> NAME/KEY: VARIANT
<222> LOCATION: 1
<223> OTHER INFORMATION: Xaa = Tyr or Phe
<221> NAME/KEY: VARIANT
<222> LOCATION: 2, 3, 5, 6, 7, 8, 9, 11, 12, 13, 14, 15, 16
<223> OTHER INFORMATION: Xaa = Any Amino Acid
<221> NAME/KEY: VARIANT
<222> LOCATION: 10
<223> OTHER INFORMATION: Xaa = Pro or Val or Phe
<221> NAME/KEY: VARIANT
<222> LOCATION: 17
<223> OTHER INFORMATION: Xaa = Asp or Gln

<400> SEQUENCE: 2

Xaa Xaa Xaa Gly Xaa
1 5 10 15

Xaa

<210> SEQ ID NO 3
<211> LENGTH: 3
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Sequence motif
<221> NAME/KEY: VARIANT
<222> LOCATION: 3
<223> OTHER INFORMATION: Xaa = Arg or Lys

<400> SEQUENCE: 3

Gly Arg Xaa
1

<210> SEQ ID NO 4
<211> LENGTH: 9
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:

-continued

<223> OTHER INFORMATION: Sequence motif
 <221> NAME/KEY: VARIANT
 <222> LOCATION: 2, 3, 4, 5, 6, 7
 <223> OTHER INFORMATION: Xaa = Any Amino Acid
 <221> NAME/KEY: VARIANT
 <222> LOCATION: 8
 <223> OTHER INFORMATION: Xaa = Tyr or Phe

<400> SEQUENCE: 4

Arg	Xaa	Asn						
1								5

<210> SEQ ID NO 5
 <211> LENGTH: 8
 <212> TYPE: PRT
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: Sequence motif
 <221> NAME/KEY: VARIANT
 <222> LOCATION: 3
 <223> OTHER INFORMATION: Xaa = Ile or Val or Leu or Ala
 <221> NAME/KEY: VARIANT
 <222> LOCATION: 5, 6, 7
 <223> OTHER INFORMATION: Xaa = Any Amino Acid

<400> SEQUENCE: 5

Trp	Arg	Xaa	Pro	Xaa	Xaa	Xaa	Gln
1							5

<210> SEQ ID NO 6
 <211> LENGTH: 24
 <212> TYPE: PRT
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: Sequence motif
 <221> NAME/KEY: VARIANT
 <222> LOCATION: 6, 8, 9, 10, 11, 12, 13, 14, 15, 17, 18, 19, 21, 22
 <223> OTHER INFORMATION: Xaa = Any Amino Acid

<400> SEQUENCE: 6

Pro	Glu	Ser	Pro	Arg	Xaa	Leu	Xaa	Xaa	Xaa	Xaa	Xaa	Xaa	Ala
1					5		10					15	

Xaa	Xaa	Xaa	Leu	Xaa	Xaa	Tyr	His
						20	

<210> SEQ ID NO 7
 <211> LENGTH: 9
 <212> TYPE: PRT
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: Sequence motif
 <221> NAME/KEY: VARIANT
 <222> LOCATION: 2
 <223> OTHER INFORMATION: Xaa = Gly or Ser or Thr
 <221> NAME/KEY: VARIANT
 <222> LOCATION: 4, 8
 <223> OTHER INFORMATION: Xaa = Any Amino Acid
 <221> NAME/KEY: VARIANT
 <222> LOCATION: 9
 <223> OTHER INFORMATION: Xaa = Leu or Ile or Val

<400> SEQUENCE: 7

Phe	Xaa	Gln	Xaa	Ser	Gly	Asn	Xaa	Xaa
1					5			

<210> SEQ ID NO 8
 <211> LENGTH: 18
 <212> TYPE: PRT
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: Sequence motif

-continued

<221> NAME/KEY: VARIANT
<222> LOCATION: 2, 3, 4, 6, 7, 9, 11, 12, 13, 14
<223> OTHER INFORMATION: Xaa = Any Amino Acid
<221> NAME/KEY: VARIANT
<222> LOCATION: 5
<223> OTHER INFORMATION: Xaa = Tyr or Ile or Val
<221> NAME/KEY: VARIANT
<222> LOCATION: 16
<223> OTHER INFORMATION: Xaa = Gly or Ala

<400> SEQUENCE: 8

Leu	Xaa	Xaa	Xaa	Xaa	Xaa	Glu	Xaa	Leu	Xaa	Xaa	Xaa	Xaa	Arg	Xaa
1	5					10							15	

Lys Gly

<210> SEQ ID NO 9
<211> LENGTH: 5
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Transmembrane helix

<400> SEQUENCE: 9

Pro	Glu	Ser	Pro	Arg
1	5			

<210> SEQ ID NO 10
<211> LENGTH: 4
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Transmembrane helix

<400> SEQUENCE: 10

Pro	Glu	Thr	Lys
1			

<210> SEQ ID NO 11
<211> LENGTH: 18
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Sequence motif
<221> NAME/KEY: VARIANT
<222> LOCATION: 1
<223> OTHER INFORMATION: Xaa = Leu or Ile or Val or Met or Ser or Thr or Ala or Gly
<221> NAME/KEY: VARIANT
<222> LOCATION: 2
<223> OTHER INFORMATION: Xaa = Leu or Ile or Val or Met or Phe or Ser or Ala or Gly
<221> NAME/KEY: VARIANT
<222> LOCATION: 3
<223> OTHER INFORMATION: Xaa = Any Amino Acid But Ser or His
<221> NAME/KEY: VARIANT
<222> LOCATION: 4
<223> OTHER INFORMATION: Xaa = Any Amino Acid But Arg or Asp or Glu
<221> NAME/KEY: VARIANT
<222> LOCATION: 5
<223> OTHER INFORMATION: Xaa = Leu or Ile or Val or Met or Ser or Ala
<221> NAME/KEY: VARIANT
<222> LOCATION: 6
<223> OTHER INFORMATION: Xaa = Asp or Glu
<221> NAME/KEY: VARIANT
<222> LOCATION: 7
<223> OTHER INFORMATION: Xaa = Any Amino Acid But Thr or Asp
<221> NAME/KEY: VARIANT
<222> LOCATION: 8
<223> OTHER INFORMATION: Xaa = Leu or Ile or Val or Met or Phe or Tyr or Trp or Ala
<221> NAME/KEY: VARIANT
<222> LOCATION: 11
<223> OTHER INFORMATION: Xaa = Arg or Lys

-continued

<221> NAME/KEY: VARIANT
<222> LOCATION: 12, 13, 14, 15, 16, 17
<223> OTHER INFORMATION: Xaa = Any Amino Acid and up to two of them can
be present or absent
<221> NAME/KEY: VARIANT
<222> LOCATION: 18
<223> OTHER INFORMATION: Xaa = Gly or Ser or Thr or Ala
<400> SEQUENCE: 11

Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Gly Arg Xaa Xaa Xaa Xaa Xaa Xaa
1 5 10 15

Xaa Xaa

<210> SEQ ID NO 12
<211> LENGTH: 26
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Sequence motif
<221> NAME/KEY: VARIANT
<222> LOCATION: 1
<223> OTHER INFORMATION: Xaa = Leu or Ile or Val or Met or Phe
<221> NAME/KEY: VARIANT
<222> LOCATION: 2, 6, 9, 10, 11, 12, 13, 14, 15, 17, 18, 20, 21, 22, 23,
24, 25
<223> OTHER INFORMATION: Xaa = Any Amino Acid
<221> NAME/KEY: VARIANT
<222> LOCATION: 4
<223> OTHER INFORMATION: Xaa = Leu or Ile or Val or Met or Phe or Ala
<221> NAME/KEY: VARIANT
<222> LOCATION: 5
<223> OTHER INFORMATION: Xaa = Any Amino Acid But Val
<221> NAME/KEY: VARIANT
<222> LOCATION: 8
<223> OTHER INFORMATION: Xaa = Any Amino Acid But Lys or Pro
<221> NAME/KEY: VARIANT
<222> LOCATION: 16
<223> OTHER INFORMATION: Xaa = Leu or Ile or Phe or Tyr
<221> NAME/KEY: VARIANT
<222> LOCATION: 19
<223> OTHER INFORMATION: Xaa = Glu or Gln
<221> NAME/KEY: VARIANT
<222> LOCATION: 26
<223> OTHER INFORMATION: Xaa = Arg or Lys
<400> SEQUENCE: 12

Xaa Xaa Gly Xaa Xaa Xaa Gly Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa
1 5 10 15

Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa
20 25

<210> SEQ ID NO 13
<211> LENGTH: 8
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthesized Construct
<221> NAME/KEY: VARIANT
<222> LOCATION: 1
<223> OTHER INFORMATION: Xaa = Ala or Cys
<221> NAME/KEY: VARIANT
<222> LOCATION: 2, 4, 5, 6, 7
<223> OTHER INFORMATION: Xaa = Any Amino Acid
<221> NAME/KEY: VARIANT
<222> LOCATION: 8
<223> OTHER INFORMATION: Xaa = Any Amino Acid But Glu or Asp
<400> SEQUENCE: 13

Xaa Xaa Val Xaa Xaa Xaa Xaa
1 5

<210> SEQ ID NO 14

-continued

```

<211> LENGTH: 9
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Sequence motif
<220> FEATURE:
<221> NAME/KEY: VARIANT
<222> LOCATION: 1
<223> OTHER INFORMATION: Xaa = Leu or Ile or Val or Met or Phe or Ser or
    Thr or Cys
<220> FEATURE:
<221> NAME/KEY: VARIANT
<222> LOCATION: 2
<223> OTHER INFORMATION: Xaa = Leu or Ile or Val or Phe or Tyr or Ser
<220> FEATURE:
<221> NAME/KEY: VARIANT
<222> LOCATION: 3
<223> OTHER INFORMATION: Xaa = Leu or Ile or Val
<220> FEATURE:
<221> NAME/KEY: VARIANT
<222> LOCATION: 4
<223> OTHER INFORMATION: Xaa = Leu or Ile or Val or Met or Ser or Thr
<220> FEATURE:
<221> NAME/KEY: VARIANT
<222> LOCATION: 8
<223> OTHER INFORMATION: Xaa = Leu or Ile or Val or Met or Phe or Ala or
    Arg
<220> FEATURE:
<221> NAME/KEY: VARIANT
<222> LOCATION: 9
<223> OTHER INFORMATION: Xaa = Cys or Ser or Ala or Gly or Asn

<400> SEQUENCE: 14

Xaa Xaa Xaa Xaa Glu Asn Gly Xaa Xaa
    1           5

```

```

<210> SEQ_ID NO 15
<211> LENGTH: 18
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Sequence motif
<220> FEATURE:
<221> NAME/KEY: VARIANT
<222> LOCATION: 1, 2
<223> OTHER INFORMATION: Xaa = Leu or Ile or Val or Met
<220> FEATURE:
<221> NAME/KEY: VARIANT
<222> LOCATION: 3
<223> OTHER INFORMATION: Xaa = Lys or Arg
<220> FEATURE:
<221> NAME/KEY: VARIANT
<222> LOCATION: 4, 6, 7, 8, 9, 16, 17
<223> OTHER INFORMATION: Xaa = Any Amino Acid
<220> FEATURE:
<221> NAME/KEY: VARIANT
<222> LOCATION: 5
<223> OTHER INFORMATION: Xaa = Glu or Gln or Lys or Arg or Asp
<220> FEATURE:
<221> NAME/KEY: VARIANT
<222> LOCATION: 11
<223> OTHER INFORMATION: Xaa = Leu or Ile or Val or Met or Phe or Thr or
    Cys
<220> FEATURE:
<221> NAME/KEY: VARIANT
<222> LOCATION: 12
<223> OTHER INFORMATION: Xaa = Leu or Ile or Val or Thr
<220> FEATURE:
<221> NAME/KEY: VARIANT
<222> LOCATION: 13
<223> OTHER INFORMATION: Xaa = Leu or Ile or Val or Met or Phe
<220> FEATURE:
<221> NAME/KEY: VARIANT
<222> LOCATION: 14
<223> OTHER INFORMATION: Xaa = Ser or Thr
<220> FEATURE:
<221> NAME/KEY: VARIANT
<222> LOCATION: 18

```

-continued

<223> OTHER INFORMATION: Xaa = Ser or Gly or Ala or Asp or Asn or Ile or
Thr

<400> SEQUENCE: 15

Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Gly Xaa Xaa Xaa Xaa Asp Xaa
1 5 10 15

Xaa Xaa

<210> SEQ ID NO 16

<211> LENGTH: 19

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Synthesized Construct

<400> SEQUENCE: 16

atctggaaag cgaacaaag 19

<210> SEQ ID NO 17

<211> LENGTH: 18

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Synthesized Construct

<400> SEQUENCE: 17

tagcggtcgt cggaatag 18

<210> SEQ ID NO 18

<211> LENGTH: 18

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Synthesized Construct

<400> SEQUENCE: 18

cccatcacca ctactacc 18

<210> SEQ ID NO 19

<211> LENGTH: 18

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Synthesized Construct

<400> SEQUENCE: 19

ccagccctga acaccaag 18

<210> SEQ ID NO 20

<211> LENGTH: 18

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Synthesized Construct

<400> SEQUENCE: 20

tgatcttacc gactacct 18

<210> SEQ ID NO 21

<211> LENGTH: 18

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Synthesized Construct

<400> SEQUENCE: 21

US 8,431,360 B2

159

-continued

cagagcttct ccttgatg

18

<210> SEQ_ID NO 22
<211> LENGTH: 533
<212> TYPE: PRT
<213> ORGANISM: Sporotrichum thermophile

<400> SEQUENCE: 22

Met	Gly	Leu	Ser	Thr	Lys	Ile	Leu	Gln	Lys	Ile	Val	Arg	Asn	Glu	Ala
1						5			10					15	
Met	Ala	Ser	Asp	Pro	Pro	Glu	Ile	Tyr	Gly	Trp	Arg	Val	Tyr	Leu	Leu
	20					25				30					
Ala	Cys	Ser	Ala	Cys	Phe	Gly	Ala	Met	Ser	Phe	Gly	Trp	Asp	Ser	Ser
	35				40				45						
Val	Ile	Gly	Gly	Val	Ile	Val	Leu	Pro	Pro	Phe	Ile	Arg	Asp	Phe	Asn
	50				55				60						
Leu	Gly	Asp	Pro	Lys	Ser	Gln	Ala	Ser	Ala	Asn	Leu	Ser	Ala	Asn	Ile
65				70					75				80		
Val	Ser	Thr	Leu	Gln	Ala	Gly	Cys	Phe	Leu	Gly	Ala	Leu	Val	Ala	Ser
	85				90				95						
Pro	Met	Thr	Asp	Arg	Phe	Gly	Arg	Lys	Trp	Cys	Leu	Ile	Gly	Val	Ser
	100				105				110						
Leu	Ile	Ile	Ile	Ile	Gly	Ile	Ile	Met	Gln	Ala	Ala	Ala	Ser	Gly	Asn
	115				120				125						
Leu	Gly	Pro	Met	Tyr	Ala	Gly	Arg	Phe	Ile	Ala	Gly	Ala	Gly	Val	Gly
	130				135				140						
Ala	Ala	Ser	Thr	Ile	Asn	Pro	Ile	Tyr	Val	Ser	Glu	Asn	Ala	Pro	Arg
145					150				155				160		
Ala	Ile	Arg	Gly	Leu	Leu	Thr	Gly	Leu	Tyr	Gln	Leu	Phe	Ile	Val	Thr
	165				170				175						
Gly	Gly	Met	Ile	Ala	Phe	Trp	Ile	Asn	Tyr	Ser	Val	Ser	Ile	His	Phe
	180				185				190						
Pro	Glu	Thr	Lys	Ile	Met	Tyr	Val	Phe	Pro	Leu	Ala	Ile	Gln	Ala	Leu
	195				200				205						
Pro	Ala	Ala	Leu	Leu	Cys	Leu	Cys	Met	Leu	Leu	Cys	Gln	Glu	Ser	Pro
	210				215				220						
Arg	Trp	Leu	Ala	Arg	Arg	Asp	Arg	Trp	Glu	Asp	Thr	Lys	Arg	Val	Leu
225				230				235					240		
Ser	Arg	Ile	Arg	Asn	Leu	Pro	Pro	Asp	His	Pro	Tyr	Ile	Gln	Asp	Glu
	245				250				255						
Phe	Gln	Glu	Ile	Val	Ala	Gln	Leu	Glu	His	Glu	Arg	Arg	Leu	Ile	Gly
	260				265				270						
Asp	Ala	Ser	Phe	Trp	Asn	Leu	Gln	Arg	Glu	Met	Trp	Thr	Ile	Ala	Gly
	275				280				285						
Asn	Arg	Arg	Arg	Val	Leu	Ile	Ser	Ile	Ile	Leu	Met	Ile	Cys	Gln	Gln
	290				295				300						
Met	Thr	Gly	Thr	Asn	Ala	Ile	Asn	Thr	Tyr	Ala	Pro	Thr	Ile	Phe	Lys
305					310				315				320		
Asn	Leu	Gly	Leu	Thr	Gly	Thr	Ser	Thr	Ser	Leu	Phe	Ser	Thr	Gly	Val
	325				330				335				335		
Tyr	Gly	Ile	Val	Lys	Val	Thr	Ser	Cys	Ile	Ile	Phe	Leu	Leu	Phe	Met
	340				345				350						
Ala	Asp	Ser	Leu	Gly	Arg	Arg	Arg	Ser	Leu	Leu	Trp	Thr	Ser	Ile	Ala
	355				360				365						
Gln	Gly	Leu	Ala	Met	Phe	Tyr	Ile	Gly	Leu	Tyr	Val	Arg	Ile	Ala	Pro

160

US 8,431,360 B2

161

-continued

370	375	380
Pro Lys Glu Gly Glu Ser Val Pro Pro Ala Gly Tyr Phe Ala Leu Val		
385	390	395
		400
Cys Ile Phe Leu Phe Ala Ala Phe Phe Gln Phe Gly Trp Gly Pro Ala		
405	410	415
Cys Trp Ile Tyr Ala Ser Glu Ile Pro Ala Ala Arg Leu Arg Ser Leu		
420	425	430
Asn Val Ala Tyr Ala Ala Ala Thr Gln Trp Leu Phe Asn Phe Val Val		
435	440	445
Ala Arg Thr Val Pro Val Met Ile Val Thr Met Gly Glu Gly Gly Tyr		
450	455	460
Gly Thr Tyr Leu Leu Phe Gly Ser Phe Cys Phe Ser Met Phe Val Phe		
465	470	475
		480
Val Trp Phe Phe Val Pro Glu Thr Lys Gly Val Ser Leu Glu Ala Met		
485	490	495
Asp Lys Leu Phe Gly Val Thr Asp Glu Ser Ser Lys Ser Leu Thr Val		
500	505	510
Asp Glu Asp Ala Lys Glu Lys Glu Lys Asp Gly Pro His Ala Arg Gln		
515	520	525
Thr Glu Val Val Ala		
530		

<210> SEQ ID NO 23

<211> LENGTH: 512

<212> TYPE: PRT

<213> ORGANISM: Sporotrichum thermophile

<400> SEQUENCE: 23

Met Lys Lys Phe Leu Gly Leu Arg Gly Gln Ala Leu Asn Leu Ala Val	1	15
5	10	
		15
Gly Thr Ile Ala Gly Cys Asp Phe Leu Leu Phe Gly Tyr Asp Gln Gly		
20	25	30
Val Met Gly Gly Ile Leu Thr Leu Lys Val Phe Leu Asp Ala Phe Pro		
35	40	45
Met Ile Asn Pro Glu Ala Ala Gly Leu Ser His Asp Glu Ser Ser Thr		
50	55	60
Arg Ser Thr Tyr Gln Gly Ile Ala Val Ala Ser Tyr Asn Leu Gly Cys		
65	70	80
Phe Leu Gly Ala Ile Ile Thr Ile Phe Ile Gly Asn Pro Leu Gly Arg		
85	90	95
Lys Arg Val Ile Met Leu Gly Thr Ser Val Met Val Ile Gly Ala Ile		
100	105	110
Leu Gln Ala Ser Ser Thr Thr Leu Pro Gln Phe Ile Val Gly Arg Ile		
115	120	125
Ile Thr Gly Leu Gly Asn Gly Gly Asn Thr Ser Thr Val Pro Thr Trp		
130	135	140
Gln Ser Glu Thr Ser Lys Ala His Lys Arg Gly Lys Met Ile Phe Phe		
145	150	155
		160
Cys Ala Ile Ile Leu Ala Phe Ile Pro Phe Leu Pro Glu Ser Pro Arg		
165	170	175
Trp Leu Ile Leu Lys Gly Arg Glu Asp Glu Ala Arg Glu Val Ile Ala		
180	185	190
Ala Leu Glu Asp Thr Asp Thr Ser Asp Arg Ile Val Glu Asn Glu Phe		
195	200	205
Leu Ala Ile Lys Glu Thr Val Leu Glu Met Ser Lys Gly Thr Phe Arg		

US 8,431,360 B2

163**164**

-continued

210	215	220
Asp Leu Phe Thr Met Asp Lys Asn Arg Asn Leu His Arg Thr Leu Leu		
225	230	235
240		
Ala Tyr Phe Asn Gln Val Phe Gln Gln Ile Ser Gly Ile Asn Leu Ile		
245	250	255
255		
Thr Tyr Tyr Ala Ala Val Ile Tyr Lys Gly Leu Gly Met Ser Asp Phe		
260	265	270
270		
Leu Ser Arg Leu Leu Ala Ala Leu Asn Gly Thr Glu Tyr Phe Leu Ala		
275	280	285
285		
Ser Trp Pro Ala Val Phe Leu Val Glu Arg Val Gly Arg Arg Asn Leu		
290	295	300
300		
Met Leu Phe Gly Ala Val Gly Gln Ala Ala Thr Met Ala Ile Leu Ala		
305	310	315
320		
Gly Val Asn Ser Arg Gln Glu Thr Gly Phe Gln Ile Ala Gly Ile Val		
325	330	335
335		
Phe Leu Phe Val Phe Asn Thr Phe Phe Ala Val Gly Trp Leu Gly Met		
340	345	350
350		
Thr Trp Leu Tyr Pro Ala Glu Ile Val Pro Leu Arg Ile Arg Ala Pro		
355	360	365
365		
Ala Asn Ala Leu Ser Thr Ser Ala Asn Trp Ile Phe Asn Phe Leu Val		
370	375	380
380		
Val Met Ile Thr Pro Val Ala Phe Asn Asn Ile Gly Tyr Gln Thr Tyr		
385	390	395
400		
Ile Ile Phe Ala Val Ile Asn Ala Phe Met Val Pro Cys Val Tyr Phe		
405	410	415
415		
Phe Tyr Pro Glu Thr Ala Tyr Arg Ser Leu Glu Glu Met Asp Asn Ile		
420	425	430
430		
Phe His Lys Val Ala Asp Gly Trp Lys Gly Val Phe Thr Val Val His		
435	440	445
445		
Gln Ala Lys Val Glu Pro Arg Trp Tyr Gly Lys Asn Gly Glu Leu Leu		
450	455	460
460		
Val Asp Tyr Gln Gln Thr Glu Glu His Arg Arg His Leu Gln Gln Gln		
465	470	475
480		
Glu Gly Ala Val Thr Ala Ser Glu Lys Arg Ser Val Glu Gly Ala Gly		
485	490	495
495		
Ser Gly Ser Gly Ser Gly Asp Val Lys Gln Asp Glu Tyr Lys Asp Cys		
500	505	510
510		

<210> SEQ_ID NO 24

<211> LENGTH: 519

<212> TYPE: PRT

<213> ORGANISM: Sporotrichum thermophile

<400> SEQUENCE: 24

Met Glu Ser Thr His Glu Pro Ala Asp Pro Ile Ala Lys Gly Val Leu		
1	5	10
15		
Ala Thr Ala Lys Gln Ser Trp His Asp Leu Phe Ile Phe Lys Gln Arg		
20	25	30
30		
Val Val Val Thr Asn Glu Leu Gly Glu Thr Ser Thr Glu Trp Ala Arg		
35	40	45
45		
Pro Val Pro Leu Arg Asn Pro Ile Ser Leu Leu Ala Gln Leu Ser Ala		
50	55	60
60		
Arg Asn Trp Leu Phe Phe Ile Val Gly Phe Leu Ala Trp Val Ala Asp		
65	70	75
80		
Ala Tyr Asp Phe His Ala Leu Ser Ile Gln Gln Val Lys Leu Ala Glu		

US 8,431,360 B2

165

-continued

166

85	90	95
Phe Tyr Asn Thr Thr Lys Thr Asn Ile Ser Thr Ala Ile Thr Leu Thr		
100	105	110
Leu Leu Leu Arg Ser Val Gly Ala Ala Phe Phe Gly Leu Ala Gly Asp		
115	120	125
Lys Trp Gly Arg Lys Trp Pro Met Val Ala Asn Met Ile Val Leu Gly		
130	135	140
Val Leu Gln Ile Gly Thr Ile Tyr Ser Val Thr Phe Ser Asp Phe Leu		
145	150	155
Ala Val Arg Ala Leu Phe Gly Leu Phe Met Gly Gly Val Tyr Gly Asn		
165	170	175
Ala Ile Ala Met Ala Leu Glu Asn Ser Pro Pro Asp Ala Arg Gly Leu		
180	185	190
Met Ser Gly Ile Leu Gln Gln Gly Tyr Ser Leu Gly Tyr Val Ile Ala		
195	200	205
Ala Cys Ala Asn Leu Gly Val Gly Gly Asp Asn Ser Trp Lys Thr		
210	215	220
Val Phe Trp Ile Gly Ala Gly Leu Ser Ile Gly Val Gly Leu Val Arg		
225	230	235
Cys Phe Phe Pro Glu Ser Gln Gln Phe Leu Glu Ala Arg Ala Ala Gly		
245	250	255
Lys Gly Gln Ala Ser Ala Ser Ala Phe Trp Lys Glu Thr Lys Met Met		
260	265	270
Leu Ala Gln Glu Trp Lys Met Cys Val Tyr Cys Ile Ile Leu Met Thr		
275	280	285
Trp Phe Asn Tyr Tyr Ser His Thr Ser Gln Asp Ser Tyr Thr Thr Phe		
290	295	300
Met Leu Thr Gln Lys Glu Leu Asp Asn Asp Gly Ala Ser Arg Ala Ser		
305	310	315
Ile Leu Met Lys Val Gly Ala Cys Val Gly Gly Thr Ile Ile Gly Tyr		
325	330	335
Ile Ser Gln Trp Phe Gly Arg Arg Arg Thr Ile Ile Val Ala Ala Leu		
340	345	350
Ile Ser Gly Leu Ile Ile Pro Ala Trp Ile Leu Pro Glu Gly Glu Arg		
355	360	365
Ser Leu Ser Val Thr Gly Phe Phe Met Gln Phe Phe Val Gln Gly Ala		
370	375	380
Trp Gly Val Ile Pro Ile His Leu Asn Glu Leu Ser Pro Pro Ala Phe		
385	390	395
Arg Ser Ser Phe Pro Gly Leu Thr Tyr Gln Leu Gly Asn Met Ile Ser		
405	410	415
Ser Pro Ser Ala Gln Ile Val Asn Ala Ile Ala Glu Ser His Ser Val		
420	425	430
Thr Ser Lys Ser Gly Lys Ser Val Asn Ala Tyr Gly Pro Thr Met Gly		
435	440	445
Ile Ala Thr Ala Ile Ile Ala Thr Gly Ile Ala Val Thr Thr Ala Leu		
450	455	460
Gly Pro Glu Lys Arg Gly Arg Glu Phe Glu Lys Thr Leu Pro Ala Gly		
465	470	475
Met Asn Ile Ile Gln Gly Gly Lys Ala Val Asp Asp Leu Glu Lys Gly		
485	490	495
Asp Ser Arg Asp Glu Lys Pro Val Val Gly Glu Val Glu Gly Gly Asn		
500	505	510

US 8,431,360 B2

167

-continued

Asp Gly Ser Gly Glu Lys Lys
515

<210> SEQ ID NO 25
<211> LENGTH: 488
<212> TYPE: PRT
<213> ORGANISM: Sporotrichum thermophile

<400> SEQUENCE: 25

Met Ala Asp Glu Lys Arg Met Gly Ser Ser Asp Ser Asp Lys Ala Ala	10	15
1	5	

Val Gln His Ser Glu Thr Leu Pro Gly Val Ser Ser Thr Ala Ala Glu	30	
20	25	

Arg Gly Phe Ala Ala Thr Asp Gln Asn Gly Gln Pro Ile Val Gln Phe	45	
35	40	

Asp Leu Lys Ala Glu Ala Arg Leu Arg Arg Lys Leu Asp Leu Phe Ile	60	
50	55	

Val Pro Thr Val Ser Leu Leu Tyr Leu Phe Cys Phe Ile Asp Arg Ala	80	
65	70	75

Asn Ile Gly Asn Ala Arg Ile Ala Gly Leu Glu Lys Asp Leu Asn Leu	95	
85	90	

Thr Gly Tyr Asp Tyr Asn Ala Leu Leu Ser Val Phe Tyr Ile Ser Tyr	110	
100	105	110

Ile Val Phe Glu Ile Pro Ser Asn Ile Ala Cys Lys Trp Ile Gly Pro	125	
115	120	

Gly Trp Phe Ile Pro Ala Ile Ser Leu Gly Phe Gly Val Val Ser Leu	140	
130	135	

Ala Thr Ala Phe Val Asp Asn Phe Ala Gln Ala Ala Gly Val Arg Phe	160	
145	150	155

Leu Leu Gly Val Phe Glu Ala Gly Met Met Pro Gly Ile Ala Tyr Tyr	175	
165	170	

Leu Ser Arg Trp Tyr Arg Arg Ala Glu Leu Thr Phe Arg Leu Ser Leu	190	
180	185	

Tyr Ile Val Met Ala Pro Met Ala Gly Ala Phe Gly Gly Leu Leu Ala	205	
195	200	205

Ser Gly Ile Leu Ser Leu Asp His Val Gly Gly Val Thr Gly Trp Arg	220	
210	215	

Met Ile Phe Val Val Glu Gly Ile Ile Thr Ile Gly Leu Ser Val Ile	240	
225	230	235

Ser Phe Ile Thr Leu Thr Asp Arg Pro Glu Thr Ala Arg Trp Leu Thr	255	
245	250	

Gln Glu Glu Lys Asp Leu Ala Ile Ala Arg Val Lys Ser Glu Arg Val	270	
260	265	

Ala Thr Thr Glu Val Leu Asp Arg Met Asp Thr Lys Lys Leu Ile Gln	285	
275	280	

Gly Ile Leu Ser Pro Val Thr Leu Ala Thr Ser Phe Met Phe Leu Leu	300	
290	295	

Asn Asn Ile Thr Gln Leu Phe Thr Val Pro Pro Tyr Val Val Gly Gly	320	
305	310	315

Phe Phe Thr Leu Ala Leu Pro Leu Leu Ser Trp Tyr Leu Asp Arg Arg	335	
325	330	

Gln Ile Ile Ile Leu Leu Ser Thr Pro Leu Val Ile Val Gly Tyr Ser	350	
340	345	

Met Phe Leu Gly Thr Thr Asn Pro Ser Ala Arg Tyr Gly Ala Thr Phe	365	
355	360	

168

US 8,431,360 B2

169**170**

-continued

Leu	Leu	Ser	Ser	Ser	Leu	Phe	Ala	Val	Gly	Ala	Leu	Ser	Asn	Ser	Gln
370					375				380						
Val	Ser	Ala	Asn	Val	Val	Ser	Asp	Thr	Ala	Arg	Ser	Ser	Ala	Ile	Gly
385					390				395						400
Leu	Asn	Val	Met	Met	Gly	Asn	Val	Gly	Gly	Leu	Ile	Ala	Thr	Trp	Ser
			405					410					415		
Tyr	Leu	Pro	Trp	Asp	Gly	Pro	Asn	Tyr	Lys	Ile	Gly	Asn	Gly	Leu	Asn
			420					425					430		
Leu	Ala	Ala	Cys	Cys	Thr	Val	Leu	Ile	Leu	Ser	Ala	Val	Thr	Leu	Leu
			435				440					445			
Trp	Met	Lys	Trp	Asp	Asn	Arg	Arg	Glu	Gly	Arg	Asn	Ala	Glu	Glu	
			450			455			460						
Glu	Leu	Ala	Gly	Met	Ser	Arg	Gln	Glu	Ile	Gln	Asp	Leu	Asp	Trp	Lys
			465			470			475						480
His	Pro	Ala	Phe	Arg	Trp	Arg	Pro								
			485												

<210> SEQ_ID NO 26

<211> LENGTH: 546

<212> TYPE: PRT

<213> ORGANISM: Sporotrichum thermophile

<400> SEQUENCE: 26

Met	Pro	Lys	Ala	Arg	Ser	Arg	Val	Pro	Val	Lys	Val	Asn	Val	Gly	Thr
1							5			10			15		
Ser	Ala	Asp	Pro	Ile	Val	Thr	Arg	Leu	Val	Gln	Glu	Asp	Lys	Ile	Pro
				20				25					30		
Trp	Tyr	Lys	Lys	Pro	Asn	Leu	Arg	Arg	Met	Tyr	Ile	Phe	Leu	Phe	Leu
				35			40				45				
Cys	Cys	Met	Gly	Val	Glu	Met	Thr	Ser	Gly	Phe	Asp	Ser	Gln	Leu	Ile
				50			55				60				
Asn	Thr	Leu	Gln	Tyr	Ala	Glu	Thr	Phe	His	Lys	Tyr	Leu	Gly	Asn	Gly
				65			70			75			80		
Arg	Lys	Asp	Glu	Asp	Gly	Asn	Tyr	Ala	Ile	Glu	Pro	Gly	Leu	Leu	Gly
				85			90						95		
Phe	Val	Asn	Ser	Cys	Tyr	Gln	Leu	Gly	Ser	Ile	Phe	Ala	Val	Pro	Ile
				100			105				110				
Ala	Pro	Trp	Phe	Ala	Gln	Arg	Phe	Gly	Arg	Arg	Trp	Ser	Ile	Met	Leu
				115			120				125				
Gly	Ser	Leu	Ile	Met	Val	Gly	Gly	Ala	Ile	Ile	Gln	Gly	Phe	Ala	Gln
				130			135				140				
His	Val	Ala	Met	Tyr	Ile	Ile	Ala	Arg	Met	Ile	Leu	Gly	Met	Gly	Ile
				145			150			155			160		
Leu	Phe	Cys	Ile	Ile	Ser	Gly	Ala	Ala	Leu	Ile	Gly	Glu	Leu	Gly	His
				165			170				175				
Pro	Lys	Glu	Arg	Ala	Val	Leu	Thr	Ser	Leu	Phe	Asn	Ser	Ser	Tyr	Phe
				180			185				190				
Ile	Gly	Gln	Ile	Leu	Ala	Ser	Ala	Ile	Thr	Ile	Gly	Thr	Thr	Glu	Met
				195			200				205				
Lys	Thr	Asn	Trp	Ala	Trp	Arg	Leu	Pro	Ser	Leu	Leu	Gln	Ile	Cys	Pro
				210			215				220				
Ser	Leu	Leu	Gln	Ile	Val	Thr	Val	Phe	Phe	Leu	Pro	Glu	Ser	Pro	Arg
				225			230				235			240	
Phe	Leu	Ile	Ser	Lys	Asp	Arg	Asp	Asp	Asp	Ala	Lys	Glu	Val	Leu	Ile
				245			250				255				

-continued

Lys Tyr His Ala Glu Gly Asp Ala Ser Ser Leu Leu Val Gln Ala Glu
260 265 270

Ile Val Gln Ile Arg Glu Thr Ile Arg Thr Glu Met Glu Val Ser Asn
275 280 285

Gln Ser Trp Met Glu Leu Val Ser Thr Tyr Gly Met Arg Arg Arg Leu
290 295 300

Val Ile Thr Leu Phe Ile Gly Leu Phe Thr Gln Leu Ser Gly Asn Thr
305 310 315 320

Leu Leu Ser Tyr Tyr Ser Gly Lys Leu Phe Glu Met Met Gly Tyr Thr
325 330 335

Glu Ala Ser Val Lys Thr Arg Ile Asn Val Ala Asn Ala Cys Trp Ser
340 345 350

Leu Leu Asn Ala Thr Thr Ile Ala Phe Leu Val Pro Tyr Phe Lys Arg
355 360 365

Arg His Met Phe Met Thr Ser Ala Leu Ser Met Cys Ala Val Phe Ile
370 375 380

Ala Ile Thr Val Ser Leu Glu Arg Thr Gln Ala Ala Gln Asp Ala Gly
385 390 395 400

Phe Lys Asn Thr Ala Ala Gly Ile Ser Gly Leu Phe Trp Tyr Phe Ala
405 410 415

Phe Ala Pro Cys Tyr Asn Met Gly Asn Asn Ala Leu Thr Tyr Thr Tyr
420 425 430

Leu Val Glu Leu Trp Pro Tyr Ser His Arg Ser Arg Gly Ile Gly Val
435 440 445

Gln Gln Ile Phe Gly Lys Leu Gly Gly Phe Phe Ser Thr Asn Val Asn
450 455 460

Ser Ile Ala Leu Asp Ala Ile Arg Trp Lys Tyr Met Ala Ile Tyr Cys
465 470 475 480

Gly Trp Ile Phe Phe Glu Phe Leu Ile Val Phe Phe Leu Tyr Pro Glu
485 490 495

Thr Ser Gly Arg Thr Leu Glu Glu Leu Ala Phe Leu Phe Glu Asp Ala
500 505 510

Ser Leu Asn Glu Lys Ala Ala Ala Val Glu Lys Gln Ile His Tyr
515 520 525

Gly Asp Glu Lys Val Val His Glu Glu Gly Gln Pro Ala Ala Lys Ser
530 535 540

Val Val
545

<210> SEQ ID NO 27
<211> LENGTH: 481
<212> TYPE: PRT
<213> ORGANISM: Sporotrichum thermophile

<400> SEQUENCE: 27

Met Leu Ser Ser Gly Phe Trp Lys Arg Arg Ser Leu Arg Val Pro Asp
1 5 10 15

Asn Gln Arg Thr Lys Ala Ala Glu Leu Thr Leu Arg Glu Ser Leu Tyr
20 25 30

Pro Leu Ser Leu Val Thr Ile Leu Phe Phe Leu Trp Gly Phe Ser Tyr
35 40 45

Gly Leu Leu Asp Thr Leu Asn Lys His Phe Gln Asn Thr Leu Gly Ile
50 55 60

Thr Lys Thr Arg Ser Ser Gly Leu Gln Ala Ala Tyr Phe Gly Ala Tyr
65 70 75 80

-continued

Pro Leu Ala Ser Leu Gly His Ala Ala Trp Ile Leu Arg His Tyr Gly
 85 90 95
 Tyr Arg Ala Val Phe Ile Trp Gly Leu Phe Leu Tyr Gly Leu Gly Ala
 100 105 110
 Leu Leu Ala Ile Pro Ser Ile Met His His Ser Phe Ala Gly Phe Cys
 115 120 125
 Val Cys Ile Phe Ile Ile Gly Asn Gly Leu Gly Ser Leu Glu Thr Ala
 130 135 140
 Ala Asn Pro Tyr Ile Thr Val Cys Gly Pro Pro Lys Phe Ser Glu Ile
 145 150 155 160
 Arg Ile Asn Val Ala Gln Ala Phe Asn Gly Ile Gly Thr Val Val Ala
 165 170 175
 Pro Val Leu Gly Ser Tyr Val Phe Phe Thr Phe Asp Asp Gln Thr Ala
 180 185 190
 Leu Arg Asn Val Gln Trp Val Tyr Leu Ala Ile Ala Cys Phe Val Phe
 195 200 205
 Leu Leu Ala Gly Val Phe Phe Leu Ser Val Ile Pro Glu Ile Thr Asp
 210 215 220
 Ala Asp Met Ala Phe Gln Ala Ala Glu Thr His Ala Gly Ala Asp Asp
 225 230 235 240
 Arg Pro Phe His Thr Gln Tyr Arg Leu Phe His Ala Ala Phe Ala Gln
 245 250 255
 Phe Cys Tyr Thr Gly Ala Gln Val Ala Ile Ala Gly Tyr Phe Ile Asn
 260 265 270
 Tyr Ala Thr Glu Thr Arg Pro Asn Thr Asp Ser Ser Leu Gly Ser Lys
 275 280 285
 Phe Leu Ala Gly Ser Gln Ala Gly Phe Ala Val Gly Arg Phe Gly Gly
 290 295 300
 Ala Ala Met Met Gln Phe Ile Lys Pro Arg Lys Val Phe Ala Leu Phe
 305 310 315 320
 Met Thr Met Cys Ile Val Phe Ser Ala Pro Ala Ile Thr Gln Arg Gly
 325 330 335
 Asn Ala Gly Leu Ser Met Leu Tyr Leu Val Met Phe Phe Glu Ser Ile
 340 345 350
 Cys Phe Pro Thr Ile Ile Ala Leu Gly Met Arg Gly Leu Gly Arg His
 355 360 365
 Thr Lys Arg Gly Ser Gly Trp Ile Val Ala Gly Val Leu Gly Gly Ala
 370 375 380
 Cys Val Pro Pro Leu Met Gly Ala Ala Ala Asp Ala Arg Gly Thr Gly
 385 390 395 400
 Phe Ser Met Leu Val Pro Leu Cys Phe Phe Val Ala Ala Trp Thr Tyr
 405 410 415
 Ala Leu Ala Val Asn Phe Ala Pro Pro Tyr Arg Ser Val Val Asp Ala
 420 425 430
 Phe Ser Thr Thr Asp Val Gly Leu Arg Glu Lys Gln Arg Glu Asp Val
 435 440 445
 Gly Ala Glu Lys Gly Gly Glu Ala Gly Gly Lys Gly Val Thr Gly
 450 455 460
 Pro Glu Asp Ala Ser Glu Asp Lys Pro Asp Val Val Asn Ser Glu Lys
 465 470 475 480

Val

US 8,431,360 B2

175

-continued

<212> TYPE: PRT

<213> ORGANISM: Sporotrichum thermophile

<400> SEQUENCE: 28

Met	Leu	Ser	Ser	Leu	Arg	Ile	Ala	Ser	Arg	Arg	Ala	Ala	Val	Ala	Arg
1				5				10					15		

Asn	Phe	Ser	Ala	Val	Arg	Ala	Ala	Ser	Thr	Trp	Ala	Asn	Val	Pro	Gln
				20				25				30			

Gly	Pro	Pro	Val	Cys	Ile	Thr	Glu	Ala	Phe	Lys	Ala	Asp	Pro	Phe	Glu
				35			40			45					

Lys	Lys	Ile	Asn	Leu	Gly	Val	Gly	Ala	Tyr	Arg	Asp	Asp	Lys	Gly	Lys
				50			55			60					

Pro	Tyr	Val	Leu	Pro	Ser	Val	Arg	Lys	Ala	Glu	Glu	Lys	Val	Ile	Ala
65				70			75			80					

Ser	Arg	Leu	Asn	Lys	Glu	Tyr	Ala	Gly	Ile	Thr	Gly	Val	Pro	Glu	Phe
				85			90			95					

Thr	Lys	Ala	Ala	Ala	Val	Leu	Ala	Tyr	Gly	Lys	Asp	Ser	Ser	Ala	Leu
				100			105			110					

Asp	Arg	Leu	Ala	Ile	Thr	Gln	Ser	Ile	Ser	Gly	Thr	Gly	Ala	Leu	Arg
				115			120			125					

Ile	Gly	Ala	Ala	Phe	Leu	Ser	Arg	Phe	Tyr	Pro	Gly	Ala	Lys	Thr	Ile
				130			135			140					

Tyr	Ile	Pro	Thr	Pro	Ser	Trp	Ala	Asn	His	Ala	Ala	Val	Phe	Lys	Asp
145				150			155			160					

Ser	Gly	Leu	Gln	Val	Glu	Lys	Tyr	Ala	Tyr	Tyr	Asn	Lys	Asp	Thr	Ile
				165			170			175					

Arg	Leu	Asp	Phe	Glu	Gly	Met	Ile	Ala	Asp	Ile	Asn	Lys	Ala	Pro	Asn
				180			185			190					

Gly	Ser	Ile	Phe	Leu	Phe	His	Ala	Cys	Ala	His	Asn	Pro	Thr	Gly	Val
				195			200			205					

Asp	Pro	Thr	Gln	Glu	Gln	Trp	Lys	Glu	Ile	Glu	Ala	Ala	Val	Lys	Ala
				210			215			220					

Lys	Gly	His	Phe	Ala	Phe	Phe	Asp	Met	Ala	Tyr	Gln	Gly	Phe	Ala	Ser
225				230			235			240					

Gly	Asp	Ile	His	Arg	Asp	Ala	Phe	Ala	Val	Arg	Tyr	Phe	Val	Glu	Lys
				245			250			255					

Gly	His	Asn	Ile	Cys	Leu	Ala	Gln	Ser	Phe	Ala	Lys	Asn	Met	Gly	Leu
				260			265			270					

Tyr	Gly	Glu	Arg	Thr	Gly	Ala	Phe	Ser	Ile	Val	Cys	Ala	Asp	Ala	Glu
				275			280			285					

Glu	Arg	Lys	Arg	Val	Asp	Ser	Gln	Ile	Lys	Ile	Leu	Val	Arg	Pro	Met
				290			295			300					

Tyr	Ser	Asn	Pro	Pro	Ile	His	Gly	Ala	Arg	Ile	Ala	Ala	Glu	Ile	Leu
305				310			315			320					

Asn	Thr	Pro	Glu	Leu	Tyr	Asp	Gln	Trp	Leu	Val	Glu	Val	Lys	Glu	Met
				325			330			335					

Ala	Asn	Arg	Ile	Ile	Thr	Met	Arg	Ala	Leu	Leu	Lys	Glu	Asn	Leu	Glu
				340			345			350					

Lys	Leu	Gly	Ser	Lys	His	Asp	Trp	Ser	His	Ile	Thr	Ser	Gln	Ile	Gly
				355			360			365					

Met	Phe	Ala	Tyr	Thr	Gly	Leu	Thr	Pro	Glu	Gln	Met	Glu	Lys	Leu	Ala
370				375			380			385					

Lys	Glu	His	Ser	Val	Tyr	Ala	Thr	Arg	Asp	Gly	Arg	Ile	Ser	Val	Ala
385				390			395			400					

176

US 8,431,360 B2

177

178

-continued

Gly Ile Thr Thr Asp Asn Val Gly Arg Leu Ala Glu Ala Ile Phe Lys
 405 410 415

Val Lys Gly

<210> SEQ ID NO 29

<211> LENGTH: 522

<212> TYPE: PRT

<213> ORGANISM: Sporotrichum thermophile

<400> SEQUENCE: 29

Met Gly Ile Phe Ala Phe Asn Lys Gln Lys Pro Asn Ala Glu Ala Thr
 1 5 10 15

Ala Val Ala Gln Glu Glu Ala Pro Gln Phe Glu Arg Val Asp Trp Lys
 20 25 30

Arg Asp Pro Gly Leu Arg Lys Leu Tyr Phe Tyr Ala Phe Val Leu Cys
 35 40 45

Ile Ala Ser Ala Thr Thr Gly Tyr Asp Gly Met Phe Phe Asn Ser Val
 50 55 60

Gln Asn Phe Glu Thr Trp Glu Asn Tyr Phe Asn His Pro Thr Gly Ser
 65 70 75 80

Lys Leu Gly Val Leu Gly Ala Leu Tyr Gln Ile Gly Ser Leu Ala Ser
 85 90 95

Ile Pro Leu Val Pro Ile Ile Ala Asp Arg Val Gly Arg Lys Ile Pro
 100 105 110

Ile Ala Ile Gly Cys Val Ile Met Ile Val Gly Ala Val Leu Gln Ala
 115 120 125

Ala Cys Arg Asn Leu Gly Thr Phe Met Gly Gly Arg Phe Leu Leu Gly
 130 135 140

Phe Gly Asn Ser Leu Ala Gln Leu Cys Ser Pro Met Leu Leu Thr Glu
 145 150 155 160

Leu Ala His Pro Gln His Arg Gly Arg Leu Thr Thr Val Tyr Asn Cys
 165 170 175

Leu Trp Asn Val Gly Ala Leu Val Val Ala Trp Val Ser Phe Gly Thr
 180 185 190

Asp Tyr Leu Lys Ser Asp Trp Ser Trp Arg Ile Pro Ala Leu Ile Gln
 195 200 205

Ala Phe Pro Ser Val Ile Gln Leu Leu Phe Ile Phe Trp Val Pro Glu
 210 215 220

Ser Pro Arg Tyr Leu Met Ala Lys Asp Lys His Glu Arg Ala Leu Ala
 225 230 235 240

Ile Leu Ala Lys Tyr His Ala Asn Gly Asp Ala Asn His Pro Thr Val
 245 250 255

Gln Phe Glu Tyr Arg Glu Ile Lys Glu Thr Leu Arg Leu Glu Phe Glu
 260 265 270

Ala Ser Lys Ser Ser Ser Tyr Leu Asp Phe Val Arg Thr Arg Gly Asn
 275 280 285

Arg Tyr Arg Leu Ala Val Leu Ile Ser Leu Gly Ile Phe Ser Gln Trp
 290 295 300

Ser Gly Asn Ala Ile Ile Ser Asn Tyr Ser Ser Lys Leu Tyr Asp Thr
 305 310 315 320

Ala Gly Val Thr Gly Ser Thr Gln Lys Leu Gly Leu Ser Ala Gly Gln
 325 330 335

Thr Gly Leu Ser Leu Ile Ile Ser Val Thr Met Ala Leu Leu Val Asp
 340 345 350

Lys Phe Gly Arg Arg Pro Met Phe Leu Thr Ser Thr Ala Gly Met Phe

US 8,431,360 B2

179**180**

-continued

355

360

365

Cys Thr Phe Ile Phe Trp Thr Leu Thr Ser Gly Leu Tyr Glu Glu His
 370 375 380

Asn Ala Asp Gly Ala Arg Tyr Ala Met Ile Leu Phe Ile Trp Ile His
 385 390 395 400

Gly Ile Phe Tyr Ser Ile Ser Trp Ser Gly Leu Leu Val Gly Tyr Ala
 405 410 415

Ile Glu Val Leu Pro Tyr Lys Leu Arg Ala Lys Gly Leu Met Ile Met
 420 425 430

Asn Leu Thr Val Gln Ala Ala Leu Thr Leu Asn Thr Tyr Ala Asn Pro
 435 440 445

Val Ala Phe Asp Ala Phe Glu Gly His Ser Trp Lys Leu Tyr Ile Ile
 450 455 460

Tyr Thr Ile Trp Ile Phe Leu Glu Leu Cys Phe Val Trp Lys Met Tyr
 465 470 475 480

Ile Glu Thr Lys Gly Pro Thr Leu Glu Glu Leu Ala Lys Ile Ile Asp
 485 490 495

Gly Asp Glu Ala Ala Val Ala His Val Asp Ile Lys Gln Val Glu Lys
 500 505 510

Glu Thr His Ile Asn Glu Glu Lys Ser Val
 515 520

<210> SEQ ID NO 30

<211> LENGTH: 554

<212> TYPE: PRT

<213> ORGANISM: Sporotrichum thermophile

<400> SEQUENCE: 30

Met Ser Ser Ser Glu Lys Glu Ala Thr Gly Pro Val Ala Ala His Val
 1 5 10 15

Gly Asn Leu Ala Thr Thr Gln Asp Val Glu Lys Ile Glu Ala Pro Val
 20 25 30

Thr Trp Lys Ala Tyr Leu Ile Cys Ala Phe Ala Ser Phe Gly Gly Ile
 35 40 45

Phe Phe Gly Tyr Asp Ser Gly Tyr Ile Asn Gly Val Leu Ala Ser Lys
 50 55 60

Leu Phe Ile Asn Ala Val Glu Gly Ala Gly Lys Asp Ala Ile Ser Glu
 65 70 75 80

Ser His Ser Ser Leu Ile Val Ser Ile Leu Ser Cys Gly Thr Phe Phe
 85 90 95

Gly Ala Leu Ile Ala Gly Asp Leu Ala Asp Phe Ile Gly Arg Lys Tyr
 100 105 110

Thr Val Ile Leu Gly Cys Leu Ile Tyr Ile Ile Gly Cys Val Ile Gln
 115 120 125

Ile Ile Thr Gly Leu Gly Asn Ala Leu Gly Ala Ile Val Ala Gly Arg
 130 135 140

Leu Ile Ala Gly Ile Gly Val Gly Phe Glu Ser Ala Ile Val Ile Leu
 145 150 155 160

Tyr Met Ser Glu Ile Cys Pro Arg Lys Val Arg Gly Ala Leu Val Ala
 165 170 175

Gly Tyr Gln Phe Cys Ile Thr Ile Gly Leu Met Leu Ala Ser Cys Val
 180 185 190

Val Tyr Gly Thr Gln Asn Arg Gln Asp Thr Gly Gln Tyr Arg Ile Pro
 195 200 205

Ile Gly Ile Gln Phe Ile Trp Ala Leu Ile Leu Gly Gly Leu Leu

US 8,431,360 B2

181

-continued

210	215	220
Cys Leu Pro Asp Ser Pro Arg Tyr Phe Val Lys Arg Gly Arg Leu Ala		
225	230	235
Asp Ala Thr Ser Ala Leu Ser Arg Leu Arg Gly Gln Pro Glu Asp Ser		
245	250	255
Glu Tyr Ile Gln Val Glu Leu Ala Glu Ile Val Ala Asn Glu Glu Tyr		
260	265	270
Glu Arg Gln Leu Ile Pro Ser Thr Thr Trp Phe Gly Ser Trp Ala Asn		
275	280	285
Cys Phe Lys Gly Ser Leu Phe Lys Ala Asn Ser Asn Leu Arg Lys Thr		
290	295	300
Ile Leu Gly Thr Ser Leu Gln Met Met Gln Gln Trp Thr Gly Val Asn		
305	310	315
Phe Ile Phe Tyr Tyr Ser Thr Pro Phe Leu Lys Ser Thr Gly Ala Ile		
325	330	335
Asp Asp Pro Phe Leu Met Ser Met Val Phe Thr Ile Ile Asn Val Phe		
340	345	350
Ser Thr Pro Ile Ser Phe Tyr Thr Val Glu Arg Phe Gly Arg Arg Thr		
355	360	365
Ile Leu Phe Trp Gly Ala Leu Gly Met Leu Ile Cys Gln Phe Leu Val		
370	375	380
Ala Ile Val Gly Val Thr Val Gly Phe Asn His Thr His Pro Ala Pro		
385	390	395
Thr Ala Asp Asp Pro Glu Ala Thr Leu Ala Asn Asn Ile Ser Ala Val		
405	410	415
Asn Ala Gln Ile Ala Phe Ile Ala Ile Phe Ile Phe Phe Ala Ser		
420	425	430
Thr Trp Gly Pro Gly Ala Trp Ile Val Ile Gly Glu Ile Phe Pro Leu		
435	440	445
Pro Ile Arg Ser Arg Gly Val Gly Leu Ser Thr Ala Ser Asn Trp Leu		
450	455	460
Trp Asn Thr Ile Ile Ala Val Ile Thr Pro Tyr Met Val Gly Glu Asp		
465	470	475
Arg Gly Asn Met Lys Ser Ser Val Phe Phe Val Trp Gly Gly Leu Cys		
485	490	495
Thr Cys Ala Phe Val Tyr Thr Tyr Phe Leu Val Pro Glu Thr Lys Gly		
500	505	510
Leu Ser Leu Glu Gln Val Asp Lys Met Met Glu Glu Thr Thr Pro Arg		
515	520	525
Thr Ser Ala Lys Trp Lys Pro Thr Thr Thr Phe Ala Ala Ser His Pro		
530	535	540
Thr Asp Leu Lys Gln Gly Glu Ala Ala Val		
545	550	

<210> SEQ ID NO 31

<211> LENGTH: 537

<212> TYPE: PRT

<213> ORGANISM: Sporotrichum thermophile

<400> SEQUENCE: 31

Met Gly Thr Ser Arg Asp Glu Lys Glu Thr Val Val Ala Asp His Ala		
1	5	10
15		

Asp Asp Asp Ala Leu Arg Glu Ala Asp Leu Ala Val Gln Val Ala His		
20	25	30

Asp Ala Asp Gly Thr Val Tyr Ser Pro Trp Ser Leu Arg Met Ile Arg

182

-continued

-continued

Ile Gly Tyr Tyr Phe Tyr Leu Val Phe Val Phe Trp Asp Leu Ile Glu
 465 470 475 480

Gly Ala Ile Met Tyr Phe Tyr Phe Pro Glu Thr Lys Asp Arg Thr Leu
 485 490 495

Glu Glu Leu Glu Glu Val Phe Ser Ala Pro Asn Pro Val Lys Lys Ser
 500 505 510

Leu Glu Lys Arg Ser Ala Gln Thr Val Leu Asn Thr Val Gly Ala Ala
 515 520 525

Gln Asn Glu Lys Leu Ala Arg Asp Val
 530 535

<210> SEQ_ID NO 32

<211> LENGTH: 566

<212> TYPE: PRT

<213> ORGANISM: Sporotrichum thermophile

<400> SEQUENCE: 32

Met Ala Val Phe Ala Met Gly Trp Gln Lys Pro Asp Asn Val Ala Gly
 1 5 10 15

Ser Ser Ala Pro Ala Ile Met Val Gly Leu Phe Val Ala Thr Gly Gly
 20 25 30

Leu Leu Phe Gly Tyr Asp Thr Gly Ala Ile Asn Gly Ile Leu Ala Met
 35 40 45

Asp Thr Phe Lys Glu Asp Phe Thr Thr Gly Tyr Thr Asp Lys Gln Gly
 50 55 60

Lys Pro Gly Leu Tyr Ala Ser Glu Val Ser Leu Ile Val Ala Met Leu
 65 70 75 80

Ser Ala Gly Thr Ala Thr Gly Ala Leu Leu Ser Ala Pro Met Gly Asp
 85 90 95

Arg Trp Gly Arg Arg Leu Ser Leu Ile Val Ala Ile Gly Val Phe Cys
 100 105 110

Val Gly Ala Ile Ile Gln Val Cys Ala Thr Asn Val Ala Met Leu Val
 115 120 125

Val Gly Arg Thr Leu Ala Gly Ile Gly Val Gly Val Val Ser Val Leu
 130 135 140

Val Pro Leu Tyr Gln Ser Glu Met Ala Pro Lys Trp Ile Arg Gly Thr
 145 150 155 160

Leu Val Cys Ala Tyr Gln Leu Ser Ile Thr Ala Gly Leu Leu Ala Ala
 165 170 175

Ala Thr Val Asn Ile Leu Thr Tyr Lys Leu Lys Ser Ala Ala Ala Tyr
 180 185 190

Arg Ile Pro Ile Gly Leu Gln Leu Thr Trp Ala Leu Val Leu Ala Leu
 195 200 205

Gly Leu Val Ile Leu Pro Glu Thr Pro Arg Tyr Leu Val Lys Arg Gly
 210 215 220

Leu Lys Glu Ala Ala Ala Leu Ser Leu Ser Arg Leu Arg Arg Leu Asp
 225 230 235 240

Ile Thr His Pro Ala Leu Ile Glu Glu Leu Ala Glu Ile Glu Ala Asn
 245 250 255

His Glu Tyr Glu Met Ala Leu Gly Pro Asp Thr Tyr Lys Asp Ile Ile
 260 265 270

Phe Gly Glu Pro His Leu Gly Arg Arg Thr Leu Thr Gly Cys Gly Leu
 275 280 285

Gln Met Leu Gln Gln Leu Thr Gly Val Asn Phe Ile Met Tyr Tyr Gly
 290 295 300

-continued

Thr Thr Phe Phe Tyr Gly Ala Gly Ile Gly Asn Ala Phe Thr Val Ser
 305 310 315 320
 Leu Ile Met Gln Val Ile Asn Leu Val Ser Thr Phe Pro Gly Leu Phe
 325 330 335
 Val Val Glu Ser Trp Gly Arg Arg Lys Leu Leu Ile Val Gly Ser Val
 340 345 350
 Gly Met Ala Ile Cys Gln Leu Leu Ile Ala Ser Phe Ala Thr Ala Ser
 355 360 365
 Gly Asn Asp Asn Lys Pro Thr Gln Asn Gln Ile Leu Ile Ile Phe Val
 370 375 380
 Ala Ile Tyr Ile Phe Phe Ala Ala Ser Trp Gly Pro Val Val Trp
 385 390 395 400
 Val Val Thr Ser Glu Ile Tyr Pro Leu Lys Val Arg Ala Lys Ser Met
 405 410 415
 Ser Ile Ser Thr Ala Ser Asn Trp Val Leu Asn Phe Gly Ile Ala Tyr
 420 425 430
 Gly Thr Pro Tyr Leu Val Asp Thr Ser Asp Gly Ser Pro Asp Leu Gly
 435 440 445
 Ser Arg Val Phe Phe Val Trp Gly Ala Phe Cys Ile Leu Ser Ile Ala
 450 455 460
 Phe Val Trp Tyr Met Val Tyr Glu Thr Ser Lys Ile Ser Leu Glu Gln
 465 470 475 480
 Ile Asp Glu Met Tyr Glu Arg Val Ala His Ala Trp Asn Ser Arg Ser
 485 490 495
 Phe Glu Pro Ser Trp Ser Phe Gln Gln Met Arg Asp Phe Gly Phe Ser
 500 505 510
 Asp Ser Gly Ile Pro Pro Ala Glu Pro Gln Leu Glu Leu Gln Gln Ser
 515 520 525
 Asn Ala Ser Thr Ser Gln Ser Asp Thr Gly Gly Ser Ser Ala Thr His
 530 535 540
 Ala Thr Ala Ala Asn Pro Gln Asp Ala Lys Met Val Ser Gln Leu Ala
 545 550 555 560
 Asn Ile Asp Leu Ser Tyr
 565

<210> SEQ ID NO 33
 <211> LENGTH: 1404
 <212> TYPE: DNA
 <213> ORGANISM: Pichia stipitis
 <400> SEQUENCE: 33

atgaagtatt ttcaaattctg gaaatcaggc aaacaagtaa	60
tgtgaattgg catttattct ttttgttatt gaacagggtt	120
aaccaggact tcctaaacac tttggaaac cccaccggta	180
tctatctata ccttagggtg ttttttggt tggatgtatga	240
atgggcagaa gaagcaaat tgcttcctca atgacagtta	300
caatgttagtt cttttcagt tgaacaattt gatgttggaa	360
actgggttggg aaacttctac ttgtccaatg tatcaggcag	420
agaggacgtt tggatgtgtc agaagcattt tttgttggag	480
tgggttggatt atgcttttc ttctacttct ggtccttattt	540
tctcagattt tggtcgccct tggtgttttc tggttcaatc	600
agatacatgt ttacaaagg agagaaagaa gaagccaaaa	660

-continued

ggaaagccag gagatcatcc tgacattttt aaggaatgga atgatattaa ttagtgcgtt	720
attttgaaaa cttcagaagg agctttctcg tgggcaaaac ttttcaagcc cgataaggca	780
agaactggat acagagtctt cttggcatac atgagcatgt ttgcgcaca gttgagtgg	840
gttaatgttag ttaattacta tattacattt gtttggatta acagtgttgg catcgaagac	900
aacctggccc taattcttgg tgggtgttgc gtcatctgtt tcactgttgg ttctttagtt	960
cctactttct ttgctgatag gatgggaaga agattgcctt cagcagttgg agctttggc	1020
tgtgggttt gtatgtgtt aatttcaatc ttatatagtt ttcaagacaa tccaaagttt	1080
aagaagagca gtggagctgg tgctgtggct ttcttttcg ttttccaact tgccttcggc	1140
tccactggta attgtattcc atggctgatg atttcagacg ttatccccct tcatgcacgt	1200
gtctaaaggat cttcatttagc tacatcaagt aactggctt ggaattttt tgggtgttgg	1260
atcactccaa ctatcattga aaagttgaag tggaaagcat atttgatctt tatgtgctgc	1320
aacttctccct tcgtaccaat gtttacttt ttctttcccg agacaaagaa ccttacttta	1380
gaagccattt acgatttgc ctca	1404

<210> SEQ ID NO 34

<211> LENGTH: 1653

<212> TYPE: DNA

<213> ORGANISM: Pichia stipitis

<400> SEQUENCE: 34

atgtcctcac aagatttacc ctcgggtgct caaaccccaa tgcgtgggtt ttccatcctc	60
gaagataaaag ttgagcaaaat ttgcgttca aatagccaaa gtgatttagc ttccatttca	120
gcaacagggta tcaaaaggcta tctcttgggtt tgggtttctt gcatgttgg tgcctttgg	180
ggcttcgtat tcgggttgcg taccgggtaca atttccgggtt tccttaatat gtctgatttc	240
ctttccagat ttggtaaaga tgggtctgaa ggaaaatattt tgcgtatcat cagagtccgt	300
ttgattgttt ccatttttaa cattgggtgtt gcaattgggtt gtatttctt ttcttaagata	360
ggagatgttt acggtagaaat aattgggtatc atttcagctt tgggtgttca cgtcgccgt	420
attatcatcc agatctcgcc ccaagacaag tggtaaccaac ttacaattttt acgtggagtt	480
acaggattttt ctgttgggtac ttgttctgtt ttgtctccaa tgggttcaat tggatgtgt	540
ccaaaggattt tgagaggtac ttgggtatac tggtaaccaat tatgtatcac cttaggtt	600
ttcattgggtt actgtgtcac ttatggaaacc aaagattttt atgatcaag acaatggaga	660
gttcctttgg gcttatgtt cctttgggtt attttcttag ttgtcggtat gttggctatg	720
ccagaatccc caagattctt aattggaaatgg aagagaatcg aagaagccaa gaagccctt	780
gcaagatcca acaagtttac tccggaaatgg ccagggtgtt acactgaact tcaatttgc	840
caggctggta ttgacagaga agctgtgtca ggttctgtt cgtggatgg attgtatct	900
ggtaagccag ctattttca gaaatggatc atggaaatttt tcttgcagtc ttgtcaacaa	960
ttaactgggt tcaacttattt cttcttattac ggaacttacaa tcttccaaac tgggttttgc	1020
caagatttccct tccagacttc catcatcttta ggtacagtca actttttttc tacatttgc	1080
ggtattttggg ccattggaaatgg atttggaaatgg agacaatgtt tgggttgcgg ttctgttgg	1140
atgttgcgtt ttgttcatcat ttactccgtt attggatcaa ctcattttttt cattgtatgg	1200
gttagtagata acgacaacac ccgtcaactg tctggtaatg ctatgtatctt tatcaatgtt	1260
ttgttcatct tcttcttgc ctgtacatgg gctggaggtt tttttccat cattttccgaa	1320
tcatatccat tgagaatca gatccaaggca atgtcttatttgc tcatgtgtc taactggat	1380

-continued

tggggcttct tgatttcctt ctgcactcca ttcatgttta atgcccataa cttcaaggtc	1440
ggctttgtgt ttactggttt tttactcttt tcgttcttct atgtctactt ctttgtcagc	1500
gaaaccaaaag gtttgcgtt ggaagaagtt gatgagtgt acgctgaagg tattgcacca	1560
tggaaagtctg gtgcattgggt tcctcccttct gcccaacaac aaatgcaaaa ctccacttat	1620
ggtgccgaag caaaaagagca agagcaagtt tag	1653

<210> SEQ_ID NO 35
<211> LENGTH: 2025
<212> TYPE: DNA
<213> ORGANISM: Neurospora crassa

<400> SEQUENCE: 35	
atggcgctcgaa acccaacgaa caccgcggcc cctacgggtg gccttaccga gaagaagcat	60
gaccgcgtt caacatcgtc cgaatccgtc tcgggaaccg ggtttgcgga acatgcagac	120
cgcacccggca ctttaacca gaacgctcgta cttagggctt caaaaaagat agcgaatcct	180
ttggccggtc taagccctca gcgtctcgag gccatgggag aagaatatgc aatgatggcc	240
ggtctcacca gcgaggagga catcagggcc tttcgaactcg gagccagaat cgccggcgat	300
gagagcaact acgacccatc cccggagctt actgaacggg agaaagaggt gttggtgcbc	360
gaaacaactc acaagtggtc taacccaccc atgctttact gggttgggtt catttgcct	420
ctatgcgcgc ccgtccaaagg aatggacgag acggcgtc acggcgccca gctttctac	480
aaggacaagt ttggcattgg tactgatagc cagagagaca cttggcttct gggtctcgtc	540
aactcagcgc cttacctttt ctgtgccttt atcggctgtc ggctcaactga accgatgaac	600
agaatcttgc cagacgagg caccatctt gtttcttgc tcatctcagc cgtagcttgc	660
ttccaccagg ctttaccaa cacgtgggtg cacatgttca tcggccgtt ctacctcgcc	720
cttggcatacg gtcccaagtc agccaccacc cccatcttcg ccggcgaatg ctccccctccc	780
aagctcccggt gtgegctgg catgcaatgg cagatgtgg ccgccttcgg tatcatggtc	840
ggctacatttgc cccatcttcg tttctacttc gttcccgtc acggcatcgg cttgggtctg	900
aactggcgtc tgatgatggg ctccgcccattt attcccgcgg tcatcgtcgt ctgcctcgcc	960
ttccctctgcc cggagtcgccc cctgggttac ctcagcaagg ccgcacacca agacgccttc	1020
ggggcgctct gcccgcgtcg tttcgaaaag gtccaaaggcc cccgcgcacct cttctacacc	1080
cacaccctcc tagaagccga gaagcaagcc atgtcgccgg tcaagaaggg taaccgcctt	1140
aaggagctt tcaaccgtcg tctgttccgc aacgcgggtca ttgcctcgac gggactcatg	1200
ttcatgcagc agttctcggtt cgtcaacatc atcgcctactt actcctcgcc ggtttccgaa	1260
gacgcggcgt tcaaggacgt ctcagcaactg ccgcgcgtcg tgggtttgg cgtcgtaac	1320
tggctgttttgc cccatcccggtt catgttacacc atcgacactt tggccggcgcaacgtgt	1380
ctgaccaccc tcccgctcat gtcctcttc ctttcttca ccggcttcag ctttggatc	1440
cccgaggactt cccaaaggccca ctcggcgtc atcgcgttag gcattttactt gttcgccatg	1500
gtctactccc cccgtgaagg gcccgtcccg tttacttactt cggccggaggc tcaaccgtcg	1560
tacatccggc caatcggtt gtcctctcc acggcgacta cctgggttctt caatttcat	1620
ctttccatca cctggccttag gatggtcacg gccttcaagc cgcaggccgc gtttggctgg	1680
tatgcagggtt ggaatatcat tgggtttctt tttaccgtt tcttggtccc cgagaccaag	1740
ggcaagacgc tggaggagct cgtcacgtt tttgacgtgc cgttgaagaa gttggcaga	1800
tacggggccgg atcagacgtt tgggttttc cacaggggaa agaatggaaa tggaaatggagg	1860

-continued

ccgacggcgc ctagtgcgga gatgtatcat gggatgcgg agcggatgaa cgaggtggtt 1920
 agcgggcagc agctgggaa gggtgagagg gagaagagg ggaacaagga acaagagagg 1980
 gaagggggaa ttatgggacg agggatgct gctggaaagg tgttag 2025

<210> SEQ ID NO 36
 <211> LENGTH: 1458
 <212> TYPE: DNA
 <213> ORGANISM: Neurospora crassa

<400> SEQUENCE: 36

atgtccgcca tcgtcggtac cgaccaatac ctcacctact tcaacaaccc ccatgatatac 60
 atccaaggag ccateggctc tgcccttgc gctggctccg tcgtcggttc cgccatcgcc 120
 ggtcctcttt ccgacaagat cggtcgttgt gactccatct ttttcgcctg cttcttcgtt 180
 ctcatggta cctccgtcca gggtgcgtgc aagaactatg gccagctcat cgccggccgt 240
 gtgctcaacg gctttaccgt cggcatcaact tcctcccagg ttcccggtta ctttgcgag 300
 atcgccaagg cagagaagcg tggttcccttg gtcatcatcc agcaactcgc catcgagtt 360
 ggtatcttga tcatgtactt ttcggctac ggctgtgcgt cgatcgaggg ccctgtttcg 420
 ttccggaccg cttggggcat tcagtttatac ctttgccttt tcctcatgtt cggctttccc 480
 ttcttgcccta ggtcgccagc atggctggcc aaggtcggtt gggaccagga ggccattgtct 540
 gtcctggcta acatccaggc tcatggcaac gttgatgacc cgagagtcgt tgctgagtgg 600
 gaggagattg tcaccgttat gaacgcccag cgtgaggccg gtaaggatg gaggaagttt 660
 gtcaagaacg gcatgtggaa gcgaaccatg gctggcatga ctgtacaggc ttggcagcaa 720
 ctgcggccgc ccaacgtaat cgtctactac ctaacctaca tcgccccaaat ggccggactc 780
 acaggcaacg tcgccatggt gacctcggtc atccaatacg ccgttttcat catttcacc 840
 ggcgtcatgt ggctttcat cgacaagacc ggctgtcgca cccttttagt ttacggcc 900
 ttgggaatgg ctttgccttgc ctttgcgttc ggccgggtca tggggcgccaa ccacgacaac 960
 gttccggacg ggtcgccgg caacgccaac attgtcatta ggtgcacaa gggcgccccc 1020
 gecaacacgg tcatactgtt ctcgtacctg ctcattgtcg tctacgcctt gacgtcgct 1080
 cccgtctgtt ggtctacgc cccggaggcc tgggtcggtt gcaactcgcc ttcggccatg 1140
 tccatggctt ccatgtccaa ctgggtgttc aacttgcgc tgggcattttt caccggccg 1200
 ggtttgtca atattacgtt gaaatgtttt atcattttcg ggtgttttgc cgtcacggcg 1260
 ggggtctgtt tctgggtt ttacccggag acgtgtggta agacgctgga ggagattgag 1320
 atcctgtttt gtgatcaggg tcctaagccg tggaaagacaa agaaggccg gtcgagactt 1380
 acggccggaga ttgaggctgt caaggccgagg aagacgggtgg agcacgagat tgaggtgcatt 1440
 gagcatgaga aggttttag 1458

<210> SEQ ID NO 37
 <211> LENGTH: 1407
 <212> TYPE: DNA
 <213> ORGANISM: Pichia stipitis

<400> SEQUENCE: 37

atgaagtatt ttcaaatactg gaaatcaggc aaacaagtaa gtcacgtgt tacattca 60
 tgtgaattgg catttttctt ttttggattt gaaacagggtt ttatggtaa tcttattaa 120
 aaccaggact tcctaaacac ttttggaaac cccaccggta gttatttagg tattatcgtt 180
 tctatctata ctttaggggtt ttttttgggt tttgttatgta acttcttcat tgggtatcga 240

-continued

<210> SEQ ID NO 38
<211> LENGTH: 1653
<212> TYPE: DNA
<213> ORGANISM: *Pichia stipitis*

<400> SEQUENCE: 38

atgtccctac aagattacc ctcgggtgc caaacccaa tcatggttc ttccatcctc 60
gaagataaag ttgagcaaag ttcttgcctca aatagccaa gtgatttagc ttccattcca 120
gcaacaggta tcaaaggcta tctcttggtt tgtttcttct gcatgttggc tgccttggc 180
ggatcgat tcggttcga taccggtaca atttccgggtt tccttaatat gtctgatttc 240
ctttccagat ttggtaaga tggttctgaa ggaaaatatt tgcgtatc cagactcggt 300
ttgattgttt ccatttttaa cattgggtgt gcaatttggc gtatttcct ttctaaagata 360
ggagatgttt acggtagaaag aattggtatac atttcageta tggttgtcta cgctgtcggt 420
attatcatcc agatctcgac ccaagacaag tggtaccaac ttacaattgg acgtggatt 480
acaggattag ctgttggtag tggttcagtg ttgtctccaa tggtcattag tgaaagtgc 540
ccaaaggatt tgagaggtag tttggtatac tggttaccaat tatgtatcac cttaggtatt 600
ttcattgggtt actgtgtcac ttatggaaacc aaagattaa atgattcaag acaatggaga 660
gttccttgg gtttatgttt ccttgggtt attttcttag ttgtcggtat gttggctatg 720
ctgtaatccc caagattttt aattgaaaaag aagagaatcg aagaagccaa gaagccctt 780
gcaagatcca acaaggatcc tccagaagat ccaggtgtc acactgaagt tcaattgatt 840
caggctggta ttgacagaga agctgctgca ggttctgttt catggatggaa attgatca 900
ggtaaggccat ctatccatggaaatc tcttacagtc tttggcaaa 960

-continued

ttaactggtg tcaactattt cttctattac ggaactacaa tcttccaagc tggtggttg 1020
 caagattcc tccagacttc catcatctt ggtacagtca acttttttc tacatttgt 1080
 ggtatggg ccattgaaag atttgaaaga agacaatgtg tgtagtgcg ttctgctgg 1140
 atgttcgtt gtttcatcat ttactccatt attggtaaca ctcatttgtt cattgtgg 1200
 gtagtagata acgacaacac ccgtcaactg tctggtaatg ctatgtctt tatcaactgt 1260
 ttgttcatct tcttcttgc ctgtacatgg gctggagggtg ttttaccat cattccgaa 1320
 tcataatccat tgagaatcag atccaaggca atgtctattt ctactgctgc taactggatg 1380
 tggggcttct tgatttcctt ctgcactcca ttcatgttta atgcccataa cttcaaggtc 1440
 ggctttgtgt ttactgggtg tttactctt tctgttctt atgtctactt ctttgcagc 1500
 gaaaccaaaag gtttgcgtt ggaagaagtt gatgagttgt acgctgaagg tattgcacca 1560
 tggaaagtctg gtgcattgggt tcctccctt gccccacaac aaatgcaaaa ctccacttat 1620
 ggtgccgaag caaaaagagca agagcaagtt tag 1653

<210> SEQ_ID NO 39
 <211> LENGTH: 1641
 <212> TYPE: DNA
 <213> ORGANISM: Pichia stipitis

<400> SEQUENCE: 39

atggaattct ccagtgttga aaaaagtgtc gaaaactgctt cctatacgcc gcaggtcagc 60
 gcaagcggct ctgcaaagac caaacagctac cttggcctca gaggccacaa acttaatttt 120
 gctgtctctt gttttgttgg tgttggtttc ttactttcg gttacgtatca aggtgtcatg 180
 ggttcattgt tgaccttgcc atccttcgaa aacactttcc cggccatgaa ggcttagcaac 240
 aacgctacct tacaaggcgc cggttattgtca ctttatgaaa tcgggtgtat gtcttcttct 300
 tttagcaacca ttaccccttgg tgacagatttgg ggttagatttga agatcatgtt tattggctgt 360
 gtaattgtct gtattgggtc tgctttgcaaa gtttctgtt tcactattgc tcacttgact 420
 gttgctgaa ttatcacttgg ttttaggtaca ggtttcatca cttctactgt tccagtttac 480
 caatcggagt gctctccagc caagaaaaga ggacagttga tcatgtatggaa aggttcttt 540
 atcgccttgc gattgcccattt ctcataactggg attgacttttgc gattttactt tttggaaac 600
 gatgggttgc actcctcgcc ttcttggaga gcacctatcg cgcttcaatg tgccttcgct 660
 gttttgttgc ttccacagt ctcttcttcc ccagaatctca caagatgggtt gctcaacaaa 720
 ggttaggaccg aagaagcttag agaagttttt tctgtctttt acgacttgcc agccgactct 780
 gaaaagattt ctattcaaat tgaagaaattt caagctgtatca tagattttaga aagacaagcc 840
 ggagaaggtt tcgtactttaa ggaattgttc actcaggccc cagccagaaa cttgcagcgt 900
 gtggccttgc tcatgttggtc tcaaaataatcg caacaaatcg ctggattttt cattattacg 960
 tactatgtctg gtactatttt tgaatcatac attggatgttgc gtccattttt gtcaagaatc 1020
 ttggcgttgc tgaacggta tgaatatttc cttgtctctt ttattgtttt ctacaccgtc 1080
 gaaagatttgc gtagaaatgtt cttttttttt tgggggttgc tggccatggc tcttgcattg 1140
 gctgggtttaa ctgttacgt taaaacttgcg ggtgaaggcc acacccatgc tgggtcggt 1200
 gctgtgttgc ttttgcatttgc attcaactca ttcttcggcg tctcctgggtt aggtggatcc 1260
 tgggttgcatttgc cacctgttgcatttgc gttgtttttt aaatttgcatttgc tctcctgtc 1320
 accgcttcta actggggctttt taacttcatttgc gttgtatgttgc tcatgttgc tgggttccaa 1380
 agtattgggtt cctacaccta ctttatctttt gctggccatca atttgcatttgc ggtccggc 1440

US 8,431,360 B2

199**200**

-continued

```

atctacttct tgcgtttccaa aaccaagggt agatcggtgg aagaaatgga tatcatttc 1500
accatatgtc ctgtttggaa gccatggaaag gttgtccaaa ttgccagaga cctccctatt 1560
atgcactcg aagtcttga ccacgaaaag aatgtcatta taaaaaaatc tagaatagag 1620
catgtcgaaa acatcagcta a 1641

```

<210> SEQ ID NO 40

<211> LENGTH: 1701

<212> TYPE: DNA

<213> ORGANISM: Pichia stipitis

<400> SEQUENCE: 40

```

atgcacggtg gtgggtacgg taacgatatac acagaaatatta ttgcagccag acgtctccag 60
atcgctggta agtctggtgt ggctggttta gtcgcaaact caagatcttt cttcatcgca 120
gtctttgcattt ctcttgggtt attggtctac gggtacaatc aaggtatgtt cggtcaaatt 180
tccggatgtt actcattctc caaagctatt ggtgttggaa agattcaaga caatcctact 240
ttgcaagggtt tggtgacttc tattcttcaa cttgggtgcct gggttgggtt cttgtatgaaac 300
ggttacattt ctgatagattt gggtcgttaag aagtcagttt ttgtcggtt tttcttcttc 360
ttcatcggtt tcattgttaca agtctgggtt cttgggtggta actacgacta catcttaggt 420
ggtagattttt tcgtcggtat ttgggtgggtt attctttctaa ttgggtgtgcc attgtacaat 480
gctgaaatttt ctccaccaga aattcgttggta tctttgggtt ctttgcacaa atttggctatt 540
actttcggtt ttatgtttt tcactggattt acctacggta ccaactacat ttgggtgtact 600
ggctctggtc aaagtaaaagc ttcttgggtt gttccttattt gtatccaattt gggtccagct 660
ttgtctttgggtt ttgttgcattt ctttgcattt ctttgcattt ttgggtgtt gatgttgc 720
gacagagaagt acgaatgtttt gtccgttctt tccaacttgc gttccttgcgta taaggaaat 780
actcttggttt aatggaaattt ctttgcattt aaggcacaaa aatgttgcgta aagagaactt 840
ttgttgcattt ttgttgcattt ctttgcattt ctttgcattt ttgggtgtt gatgttgc 900
ttcaaccaat acaagtccat gattactcac tacccaaacctt tcaagctgtt tgcaatgtt 960
ttgttgcattt ttgttgcattt ctttgcattt ctttgcattt ttgggtgtt gatgttgc 1020
ttcatcttcattt ttgttgcattt aacaccattt ctcttttgcattt ttgttgcattt 1080
gtcggtatcg tcatgttgcattt tgcttgcattt ccagctgttgcattt ttgggtgtt gatgttgc 1140
agaaagccat ttgttgcattt ctttgcattt aatgttgcattt ttgttgcattt ttgggtgtt gatgttgc 1200
gtcaatcttgcattt ttgttgcattt gtcgttgcattt ttgttgcattt ttgggtgtt gatgttgc 1260
gttgcattt ttgttgcattt ctttgcattt ctttgcattt ttgggtgtt gatgttgc 1320
tttgcattt ttgttgcattt ctttgcattt ctttgcattt ttgggtgtt gatgttgc 1380
ttttgcattt ttgttgcattt ctttgcattt ctttgcattt ttgggtgtt gatgttgc 1440
gtcaagttcg tgcgttgcattt ttgttgcattt ttgttgcattt ttgggtgtt gatgttgc 1500
caattcttcattt ttgttgcattt ctttgcattt ctttgcattt ttgggtgtt gatgttgc 1560
gacacctctg gtacttccaa gatggaaaag gaaatccatg agcaaaaagct taaggaaat 1620
ggtttgcattt aattgttgcattt tgaagaaaat gcttgcattt ttgggtgtt gatgttgc 1680
gtcttaccacg ttgaaaaataa a 1701

```

<210> SEQ ID NO 41

<211> LENGTH: 1656

<212> TYPE: DNA

<213> ORGANISM: Pichia stipitis

-continued

<400> SEQUENCE: 41

atgagagaag ttggattct tcatgttgc catggcaacg ttgttaactat aatgtatgaaa	60
gatccagtag tatttttgtt gattttattt gcattcccttg gaggtttgct ttttgttat	120
gatcaagggg ttatttagtgg cattgtcaca atggaatctt ttggtgcaaa attccccaga	180
atttttatgg atgcccatta caagggttgg tttgtgtcta ctttttgtct atgcgcattgg	240
tttggctcta ttatataac tccaattgtt gataggtttga agaacgtga ttctatcaca	300
atctcttgc ttattttgtt cattggttct gcgttccaaat gtgtggcat taatacaagt	360
atgttatttg gtggggcgtgc tggtgctgtt ctgcgcgtcg gtcaattaac catggtagtt	420
ccaatgtaca tgcggaaatt ggctccctca tcggtgagag gtgggttgggt tgtaattcag	480
caacttcga tatacaattgg tatcatgatt tcctattgtt tggattatgg cactcatttt	540
atggaggtt ctagatgtgc tcctagtcac ccataccaag gtgaaacttt taaccctaat	600
gtggatgttc ctccagggtgg ctgctatgtt caaagtgtatgc ccagttggag aattccttt	660
ggtgttcaga ttgctccagc agtgttggat ggtattggaa tgatattttt cccaaagatct	720
cccgatgtt tactctctaa aggtcgccgc acggaaactt ggagctctt gaaatatctc	780
agaagaaaga gtcatgagga tcaagtgcac agagatgg ctgaaattaa ggcagaggc	840
gtttatgaag acaagtacaa ggaaaagaga ttccctggta agactggagt tgcttaaca	900
cttactggat actggatata tcttactact aaatctact tcaagagagt ttatggat	960
tcagctgtca tggcttcac acaattcatt ggctgcataatg caataattttt ttacgcacct	1020
acaattttca cacaattggg aatgaactct acaactactt cttgtttgg tactggcttt	1080
tatggtattt ttaattgtct ttccaccctt ccaggcgtgt tctgtatcga tagatgtgg	1140
agaaagactt tgtaatggc aggtgttattt ggaacttttta ttcccttggat tattgtcgcc	1200
gcaatcggtt gcaagtatgg cgatcgatca tctgaattca agacagcagg gagaactgca	1260
attgtttca ttttcatat ttttcatat ttttcatat ttttcatat ttttcatat ttttcatat	1320
gttttacccctt cagatgtttt cccaaatcgcc atcagatccaa atgcacatctc cataactacc	1380
tcatctactt ggatgaataa ttatgttattt ggatgttattt ctccacatat gtttagaaaca	1440
atgaagtggg gcacttacat ttttttgtca ggtttgttca ttattgtgtt cttttttact	1500
tggcttatca tcccgaaac caagggagtt ccattggaaag aaatggatgc cgtgtttggc	1560
gataactgcacg cattgcagga aaagaatttg gttaccattt cgtcagtttc tgaatctgac	1620
gcaaggatc gcaactcgat taaaatgtca gaataa	1656

<210> SEQ ID NO 42

<211> LENGTH: 1590

<212> TYPE: DNA

<213> ORGANISM: Pichia stipitis

<400> SEQUENCE: 42

atggcatatc ttgattggtt aacagctaga accaacactt tcgggttggag gggcaagaag	60
ttgagagcct tcatctactgt agtggctgtc actggttctt cattattcgg atatgtatcaa	120
gggttgtatgtt ccggaaattat tactgtgtat caattcaact ctgagtttcc cgccactaga	180
aataacagta ctatccaagg tgccgtcacc tcctgttacg agcttgggtt tttttttgtt	240
gctgtgtttt cttgtttaag aggtgaaaga attggaaagaa gacctttgtt gctttgtggc	300
tctgttattt tcatcttggg aacagttattt tctgttaccc cttccatcc acactggatca	360
ttaggtcagt ttgttattgg tagagttatc actggatttg gtaatggat gaataactgcc	420

-continued

accatttccag	tttggcaatc	ggaaatgtca	agagctgaaa	acagaggaag	attggtcac	480
tttggaaagg	tttgcgtcg	tgtgggtaca	tgtattgc	actgggttga	tttcggtttg	540
tctttagtgc	acaattcagt	ttccctggaga	tttccagttg	ctttccaaat	agtgtttgt	600
tccgttttat	tttgtggaaat	gttcaatttgc	cccgcacttc	caagatggtt	ggttgttaac	660
cacagaagag	cagaggctct	tcaagtgttgc	tctgttttgc	aagacttgcc	cgaagacgac	720
gaagaaaattc	ttaatgaagc	tgaagtttatt	caggaaagtgc	tagacaagtt	tgctggacat	780
gtttccgtca	aggaagtgtt	tactgggtgt	aagacccagc	actggcaagag	aatggttatt	840
ggatccagca	cccaatttctt	tcagcagttc	actgggttgc	acgctgccat	ttactattcc	900
actgtcttgc	ttcaagacac	tattggtttgc	gaaagaagaa	tggcattgtat	tatcggttgt	960
gttttcgcaa	ctgtctacgc	cattttcaca	attccttcctt	tcttcttgcgt	cgatacttttgc	1020
ggacgttagaa	acttggtttgc	gattgggtgt	atgggacaag	gtattgcatt	cactatcacc	1080
tttgcctgtt	tgatttgacga	tactgaaaac	aacgcaccaagg	gtgcccgcgt	tggttttttc	1140
ttgttttattt	gttttttcgc	cttcaccatc	ttgcattgc	catgggtata	cccaccagaa	1200
atcaatcctt	tgagaaactag	aactatacg	tctgcatttgc	ccacttgcac	caactggatc	1260
tgtaacttttgc	ctgtttgttat	gttcacccct	gtctttgtca	ctaaccacca	atggggagcc	1320
tatcttttctt	ttgtgtgtat	gaacttcctt	ttcggttcttgc	ttattttttttgc	cttctaccca	1380
gaaacagctg	gaagatcggtt	ggaagaaatc	gatatcatct	ttgcgaaggc	attcggttgc	1440
aaaagacagc	catggagagt	tgctgcaccc	atgcacaaatgt	tgtccaaatcc	cgaaattgaa	1500
gacgaaagccaa	acagattggg	cttgggttgc	gatggatcat	tgcacaagga	agcatttgc	1560
acccaaqaqaa	acqcatccaaq	caqctctttaa				1590

<210> SEQ ID NO 43
<211> LENGTH: 1689
<212> TYPE: DNA
<213> ORGANISM: Neurospora crassa

<400> SEQUENCE: 43
atggcgccctc caaagttcctt gggccctctca ggccgcacccgc tctctcttagc tgtctcgact 60
gtagccaccac cgggcttcct tctcttcggc tatgaccaaag gtgtcatgag cggccatcatt 120
accggcccccg cttaacaaca ctttttcaca ccaacccaaag acaactcgac catgcagggt 180
cttcatactg ccatctacga aattggatgc ttgtatggtg ccatgttcgt cctctggacc 240
ggcgatttgt tgggttagacg caggaacatc atggtgccggc ctttcattat ggctctcggt 300
gttcattttc aggttacctg tcaggcttgc tccaaccctt ttgtctcagct gtgtcgcc 360
agagtcgtca tgggtattgg caacggcatg aacacttcga ccattccac ttatcaagcc 420
aatgtctcaa agacatcgaa ccgcggcttt ttgtatctgca ttgaaggccgg tgtcattggc 480
tttggtaactt tgattgctta ttggatcgac tatggtgcat cttacggtcc cgatgacctc 540
gtttggcgct tccccatcgc tttccagctt ctcttcgcata ttttcattctg cgtccccatg 600
ttttacccctt ccgagtcgcc cagatggctc ctcagccatg gcccggaccac agaagctgac 660
aaggtcattt ctgcggcccg tggctacccg atcgatggtc cggagaccat ccaagagcgc 720
aaccttcattt ttgactccctt gctgtccctt ggagggttccg gccaaaagag cactcccttc 780
aaggccctct tcactggccgg caagaccacg catttcgcgt gtctcttgc cggttccagc 840
tcccaaggatca tgcagcaagt tggtggttgc aacggccgtca tctactactt cccattctg 900
ttccaggattt ctatggcgat gtcggccaaac atgtccatgt tgctggccgg tatcaacatg 960

-continued

atcgctact ccatctcgc taccgttcc tggttcgcca ttgagcggtg cggtcgtcgt 1020
cgctgttct tgatcgac cgttgccag atgctctcca tggtcatcgt ctgcgcgtc 1080
ttgatccccg acgaccctat gaaggccccg ggtgcccgcc tcggtctt cacttacatt 1140
gccttttcg gtgccacttgc ttcccccctc ccctggctct accccgccc ggttaacccc 1200
atccgcacac gtggaaaggc taacgcccgc tccacctgcgt ccaactggat gttcaacttc 1260
ctcategtca tggtcacccc catcatggtc gacaagattt gctggggAAC ttacctttc 1320
ttcgcggtca tgaacggctg ctteccccc atcatttact tcttctaccc cgagactgc 1380
aaccgcgcgc tcgaggagat cgacatcatc ttgcgcagg gttcgctcgaa acatgtcg 1440
tacgttactg ccccaaggaa gctgcctcac ctcactgccc aggagatcgaa gtcctatgcc 1500
aacaagtatg gcctcgctcgaa cccgcgttcc aacggcgagg gccggcaacccg ccatgacgag 1560
gagaagacgc gcgaccggcc cgaccagatg gacagcgact ccccgctca cgtcgagatt 1620
gtatgttgcg acgagcacgg tgcgttgcggt ggcttcgggtt atggtattaa caccaggaa 1680
acacgttaa 1689

<210> SEQ ID NO 44
<211> LENGTH: 1626
<212> TYPE: DNA
<213> ORGANISM: Neurospora crassa

<400> SEQUENCE: 44

atggaattcg agcacatca ctccgcctcc gacattgaga aggaggccgt cactgtggcc 60
cggccacagg gcgatgtcac ccgcgtttag gctcccgta ccctcaaggc gtacatgtatg 120
tgcgttttgc cccgtttcggtt cggtatcttc ttggctacg attcaggtta catctctgg 180
gtcatgggca tgaagtactt tatcgaaacc atcaacggac cccggccac cttccgtcc 240
tccaaggaaa agtcgtcat cacccattt ctctctggcc gaaccttctt tggcccccctc 300
atgggggttgc atctcgctga ctgggttggc cgtcgctcta ccatacatttt cggctgtcc 360
gtcttcatcg tcgggttgtt tctccagact gctcccgaa gcttgggtctt cattgtggcc 420
ggccgtctcg tcgctggttt cgggtcggtt ttcgtctcggtt ccattatcat cctgtacatg 480
tctgagatcg cggcccgcaaa ggtcccgccgtt gctatgggtt cgggttacca gttctgtatc 540
tgcctgggtc tgctccgtgc ctgcgtcggtt gactacggca cccagaacccg caccgacagc 600
ggctttaaca gaatcccgat tggctcccgat atggccctggg cccttattttt tgctactgg 660
atcttttcc ttccgtatc ccctcgcttc ttctgtcaaga agggcaagctt cgacaaggcc 720
ggcggtgtc tctcccgctt ggcggaccag ccgcgtcgattt ccgactacgtt cagggacgaa 780
tttgcggaga tcgttgcggaa ccacgaattt gaaatgaccg tggccatccat cggcaactac 840
ttccagcgtt gggccaaactt gttcccgccgc tccatctggc aggggtgggtt ttaccccgcc 900
cgccacattt cccgttccatc gatgcagatg atgcagcgtt ggacggaaat caactttatc 960
ttttacttttgc gaaaccaccc ttcccgccgtt ctcggccatca ttgacaaccc cttccgtatc 1020
tctctgggtca ctactttgtt caacgtctgc tccacccccc tctcccttca caccatggag 1080
aagctcgccgc gtcgttccatc cctcatctgg ggcgtctcg gcatgtgtat ctgcgttgc 1140
atcgctcgccca tcgttggtaatc ctgcggccgtt gatgataccca tggccatcaa ggccatgtc 1200
gccttcatctt gcatctacat cttttttttt gtcaccatc gggccatccgc ttccctgggtc 1260
gtcatacgccg aggttttccc ttcccccattt cgtgccaagg gttttggccctt ttccaccggcc 1320
tccaaactggc tctgttaacttgc catcatcgcc gtcatacttc cctacatggt cgacgaggac 1380

-continued

aaggcaacc tggcccaa ggtttctac atctgggtg gctctgcac ctgctgctc	1440
atctacgcct acctgcttgc gcccagacc aaggcccta cgctcgacca ggtcgacaa	1500
atgcttcgg agtctacccc ccgcacctcg accaagtgg acccctcacac cacttatgt	1560
gctgagatgg gcatgaccga gaagactgtt gctggccacg ctgagaaccg cagcgatagc	1620
gagtaa	1626

<210> SEQ ID NO 45
<211> LENGTH: 1614
<212> TYPE: DNA
<213> ORGANISM: Neurospora crassa

<400> SEQUENCE: 45	
atgggtcttt cgataggaaa taggatcctc cgaaaaattt taaaaatga ggccatggca	60
gaagatcccc cagagatcta tggctggcgt gtctatctcc tagcgtgctc tgcctgctc	120
ggcgccatgt ctccggctg ggattccctcc gtcatacgcc gcgtcatcga actcgaaacc	180
ttaaacacg actttggctt catcgcaac gataaagcca aggccaaacct gggcgccaa	240
atcgctctca ccctccaagc cggctgctc ctcggtgccg tgatcgccctc acctataacc	300
atcgcttcg gccgcaagtg gtgtctcatc gtgtctccc tggtcgtcat catcggtatc	360
atcatgcaag cggccgcctc aggcaaccc gcacccatgt acattggccg ttcgtcgcc	420
ggcggtggcg tcggccgc cagctgcata aacccgtct ttgtgtctga gaacgctccc	480
cgtcgatcc gcggtctgtt gacggccctc taccactt tcatgttgcac cggcgccatg	540
atcgcatttt ggatcaacta ctccgtctct ctgcacttca agggcaaaatc catgtacatc	600
ttcccgctcg ccatccaagg tctccgc ggccttttgtt gctctgcat gctctctgc	660
cacgaaagcc cgcgtggct ggcccgctgt gaccgtggg aagaatgcaa gtctgtgt	720
gcgcgcatcc gcaacccccc cccagaccac cctgatcg tcgacgagtt cgcgagatc	780
caggaccage tcgaacagga ggtcgatc caggccgacg ccacttactg ggacttgacc	840
cgcgatatgt ggaccgtcgc cggcaaccgc aagcgcgcgc tgatttagtat tttctgtat	900
atctgccagc aaatgacggg caccaacgc atcaacacgt acgcgcctac catttcaag	960
aacttggta tcaacggcac gtgcactgc ttgttagta cggcatcta tggtagttgtc	1020
aaggctgtta gctgegtcat ttcttgcgt ttcttggccg actcgctggg tcgttagacgt	1080
tgcgtgtgtt ggacgtcgat tgcgcagggt ctgtatgtt ttatattgg cttttatgtc	1140
cgcatttcgc cggcgattga tggccagccg gtggccgcctg cgggttatgt agcggttgt	1200
tgcattttc tggccgc tttttccaa tttggctggg gtctgcctg ctggatctac	1260
gcctcgaaa tccccccgc cgcgtgcgc tccctcaacg tgcctacgc cggcgacg	1320
cagtggatgt tcaatttcgt cgtggccgc gccgtgcata ctatgttgtt caccggccgc	1380
ccccacgggtt acggcaccta cctcatctt ggcagcttgc gctcagcat gtttgtttt	1440
gtctgggtt tgcgtccgc gacaaagggt atctcgcttgc agcacaatggg tgatgtcggg	1500
ggcggttactg atggccctgc cgctgagaag tgcgtggc atggtgagaa tgatgtcggg	1560
tggagatgg ggaaggggga tcagaagtgc aagcatgtgg aggtttatgt ttaa	1614

<210> SEQ ID NO 46
<211> LENGTH: 1587
<212> TYPE: DNA
<213> ORGANISM: Pichia stipitis

<400> SEQUENCE: 46

-continued

atgtcttcgt tattgactaa cgaatacttc aaagactact accacaaccc gactctgtt
gaagtgggta ctatgattgc tatcttagag atcggcgac tttttctc cttcatagct 120
ggaagagtag gtgacatcgt tggcagaaga agaaccatta gatacgggtc tttcatttt
gtagtagggcg gtcttgatac agctacttcg gtcaatattg tcaatctctc acttaggaaga 240
ttgattgccc gtattgccat tggcttttg acaaccatca tcccatgcta ccagtctgaa 300
atcagcccc cagacgatag aggtttctat gcctgttgg agttcacccg aaatatcatt 360
ggatatgcta gtagtatttgc ggttagactac gggtttcat ttttagacaa tgattcage 420
tggaggagcc cattgtatgt tcaggttgtt attggctcca tgttatttt tggttcattc 480
cttattgttag aaacccctag atggcttgc gatcacaacc atgatatcga agggatgatt 540
gtcatttctg acttgtatgc agatgggtat gtggaaagacg atgatgtat tgctgagtag 600
agaaacataa agggaaagtgt cttgatagcc agagttgaag gcggagagag atcgtaccag 660
tattttttca ccaaataatac caagagactt tctgtggcat gctttcgca aatgtttgcc 720
cagatgaatg gtataaacat ggtatcttac tatgcttca tgatctcgat atctgctggc 780
tgggttggta gacaagctat cttagtact ggtatcaact ccattatcta catcttagt 840
accatttctc catggtaact agttgattct tggggcagaaa aacctttgct tttatctgg 900
tctgtgttca tgggttcc gctttaacc attgcttgcgttattttttaaacaacaca 960
tacacacccg ggggttggta atcgatattca atgctgtttt tggatacagt 1020
tgggttccaa ttccctggct catgagcgaa gtgttccctt actcagtttag atcaaaaggt 1080
gtgtccatgt ctactgcaac caactggctc tttaacttta ttgttggaga gatgacacct 1140
attttggta atacaatttttac ctggagaact tacttgcattcc cggcaacttc gtgtgttattttttaa 1200
tctgttttttgcgttattttttaaacaacaca gagaccaagg gtttagcatt ggaggatatg 1260
gggtccatgt tcgatgataa ttccgttcaata ttttcatatc actcaacttc ttccactggg 1320
tatgggtcgca ccgagtctaa cagtaatgcc aggagagcaat gtgtcatctc ttccatgg 1380
taccaggata gtttgcattca gacagcgctc tcattggctt gaaatcttca aagcatgagg 1440
cctgattacatggcataat cacaggatgtc gtttgcattca gaaatcttca aagcatgagg 1500
ccaataaaaca tttccagcaat tattccgcag gaaatttgcac caccaacccat tttatcttca 1560
tttaagtaca aqttqaatqaa qatqgaa 1587

<210> SEQ ID NO 47
<211> LENGTH: 1257
<212> TYPE: DNA
<213> ORGANISM: *Pichia stipitis*

<400> SEQUENCE: 47

atgacttttg cagtttactt gtatgtttt gcagttggta gagtgcttc tgggggggt	60
gttaggaggttc tatcgactat ggtgcgtcc tatcaatgcg aaattagtcc cagcgaagaa	120
agaggcaagt tggtgtgtgg agagttcacg gggaaatatac ctggttatgc tctcgtgt	180
tggggcgatt acttctgcta cttaattcaa gatataggtg atgcaaggga gaaggctcat	240
agcttctttt cccacttgc ctggcgattt cctctattca tccagggtt gatagcggct	300
gttctttttt ttgggggatt tttaatttgc gagtcaccc gttggttatt agatgttagac	360
caggaccacaa aaggattcca tgtattagecg ttgcctatg attcacatct agatgataac	420
aaaccacgtt aagagttctt tatgtatcaaa aactccatct tgtagaaag agaaaactaca	480
cctaagagcc aacgaaactt gaaacatgtt tcagaactt acatgaccgg agtggttata	540

-continued

gcttggtag cactggcgt tgcacagttc aacggcataaa atatcatttc gtactatgcc 600
 cccatggtat ttgaagaagc aggcttcaac aactccaagg cttaacttat gacaggcatc 660
 aactctatag tatattgggt cagtagcatt cctccgtggt ttctcggttgc tcattgggt 720
 agaaagccaa ttttgataatc cgggggttta tctatggaa tatgtattgg tttgattgcg 780
 gtggtaattc tactagacaa gtcgttcaca cctgttatgg ttgcggattt ggtgataatc 840
 tacaatgcat cttttggcta cagttgggtt cctatcgat tcttgcatttccc gccggaggtg 900
 atgcccattgg cagtttagatc gaaaggtt tctatattctt cggctacaaa ctggtttgc 960
 aattttgttg tgggtcagat gacgccaattt ctacagcaga gattgggcgtt gggacttat 1020
 ctatccccgg cttggtagttt tatcatctcg gtgtatgtt tgatatttttctt ctatccagag 1080
 acaaagggtt cagagctaga ggatgtggac tctgtgttgc agagcttttcaactacaag 1140
 tctccgttca agatttcacg aaagagacac cagaatgtatc gccaggcgta ccaaagggtt 1200
 gagaacgata tccgcccacaa cgtatgttgcgaa atggacgattt tggacgattt ggactaa 1257

<210> SEQ_ID NO 48
 <211> LENGTH: 1757
 <212> TYPE: DNA
 <213> ORGANISM: Neurospora crassa

<400> SEQUENCE: 48

atggaattcg gtggcgagg cggctccggc gcagctgggtt tttatgtatc ggcttttag 60
 aggctgttgggg cagtgtatgggg gaagaggtggc cctgcagcac ttgtcaagaa cttccgggtc 120
 ttttctatttgc atgtatcggtt ggtgtgtctt atgggtacaa ccaagggtt 180
 ttttcgggttgc ttctcgccat gccagccctt cagaaacaca tggggcaata cgtatccgata 240
 gacgagaacg cgagtcagac aaagaaggc tggcttaaccg caatttttgc gctcggtgt 300
 tggcttggttgc cgttttgc tgggttcatgc gcaagggttgc tctcgagaaa gtacgggtgt 360
 ctatgtgggttgc tttatgtgggtt gttgtatcc aagccacgtt tttatcttgc 420
 ggacatgaga ccatttttgc cggacgggtt atcacgggttgc tgggtgttgc atccttagcc 480
 atgatcatc ccatttacaa ctcggaaatgc gcaccacgtt agtccgttgc agtctttgtt 540
 gtcgttcacgc agttggctat ctgtttcggtt atcatgttgc gtttctggat tgactacgg 600
 accaactata tcggggcac caagctcgat acccaatccg acgcccgcgtt gttgtaccc 660
 gtcgttcacgc aacttcgttgc tggcttcattt ctgttttgc gtcgtatgtt catgccttc 720
 tccccacgtt ggttcatccca ccatggccgc gaggcggaaatgc ctgttttgcacc 780
 cttcgccgtt tccccacgtt ccacgttgcgtt gtcgtatgtt catgccttc 840
 cagtcgttgc tggcttcattt ctgttttgc gtcgtatgtt catgccttc 900
 gtcgttcacgc aacttcgttgc tggcttcattt ctgttttgc gtcgtatgtt catgccttc 960
 atgttccgac gtcgttgcgtt gtcgtatgtt accatgttgc ttcacgttgc gtcgttcacgc 1020
 aatgcgttgc ttcactacgc cccgcaatc ttcaagcgcg ttggacttgc gtcgtatgtt 1080
 acctcactcc tggcttcattt ctgttttgc gtcgtatgtt catgccttc 1140
 gtcgttcacgc tggcttcattt ctgttttgc gtcgtatgtt catgccttc 1200
 gtcgttcacgc atatcatcat cgttgcattt gttgttgc gtcgtatgtt catgccttc 1260
 cataaggcttgc tggatgggc tggatgttgc atgttgcgtt ttcacgttgc gtcgtatgtt 1320
 tattcatggg gtcgttcacgc tggatgttgc atgttgcgtt ttcacgttgc gtcgtatgtt 1380
 ccatatgggttgc ttcacttcagg agtccgttgc aactggatgc acaactttat cgttcgttgc 1440

-continued

gttcacgcggg atatgttaaa ggcgatcccg tacggaaacgt atatcatctt cgggttgttg	1500
actatatggg tgccgccttt atttggttct ttgtgccgga aacgaagaga ttaaccttgg	1560
aagagatgga catgatctc ggatccgaag gcactgcaca agccgacaat gagcgcattgg	1620
aggagatcaa tgctgagatt ggtcttaccc gattcctgca aggtggtagt ggtgcaaacc	1680
aagggtctgc tgatggaaagc gatactggtt atgatgcgga gaagggcaag agcgaacact	1740
attctcaqca tgtctaa	1757

<210> SEQ ID NO 49
<211> LENGTH: 1584
<212> TYPE: DNA
<213> ORGANISM: Neurospora crassa

<400> SEQUENCE: 49

<210> SEQ ID NO 50
<211> LENGTH: 1968
<212> TYPE: DNA
<213> ORGANISM: *Neurospora crassa*

-continued

<400> SEQUENCE: 50

atgggcacaca atccagacct ggacagttagc ggactgccc gagaacccaa aggtgtcacc	60
ggctcgacaca ttgaacaaac ctcgtccaac ctgcgaagcca acatcaaccc cgaagccaaag	120
ctcaagaacc cgctcgacgg ccttccccgc gtgcagctcc tgtcacgcgt cgagaccttt	180
tgcgcgaaaa agaacctaacc cgagcaccc cctctttcc gtaagggagc actcatcgcc	240
cagtccccgg acagctatgc gtccatctcg ggcccgaaag cttggacga tgaggagaag	300
gcagttttt tgaaggaggt cgaacacaag tggcggtcg cggcaagact gttctgacg	360
attgctactt gctcgatcg tgctgctgca caagggtggg atcagacggg cacgaatggc	420
gcaaatatct tctttccaa ctattacggt atcggaggcg acactgcgag ggagaagttg	480
cttgatggat tgatcaatgc tggcccttat attgggagc cattcatcg ttgctggctt	540
tctgatccca tcaacaactg gattggctgt cgtgggtta tctttgtctc tgctcacttc	600
tgtatctggc ccgtcatcggt ttctgcttgc tgcacacat ggccccagca actggcctgc	660
cgtctgctga tgggtatcggt tatgggtgtg aaggcatcaa cggtgccgat ctatgcccg	720
aaaaactcgc ctgcttctat tcgagggtcg ctggcatgt catggcagat gtggcagaccc	780
ttcggcatct tctgggacac tgccttaac ctggcgctc tccacgcccag ctccaaacgtt	840
aactggcgcc tcatgctcg tggcccttc attcccgccg taccctgtct tctgctcatc	900
tatcttgcc ccgagtcggc gcgctggta atgaagaagg gcccgtaccc agaagctgg	960
aaatccatgg tcaagctgcg caaccacccc atccaagttt cccgcgcacat gttctacatc	1020
cactcgcaat tggaaagtgcg gcaccagctc ctgcggcgcget ccaactatgc caagcgcttc	1080
gtcgagctt tcaccgtccc tcgtgttcgc cgcgcaccc tcgcgcgttt caccgtcatg	1140
attggccagc agatgtgcgg aatcaacatc atgcctttt acagcaccac catttcaag	1200
gattccggctt ccaccaattt ccaaggccctg ctgcgcgtt tcggcgttcgg tctagtcaat	1260
tggcttttg cttcccccgc ctctggact atcgacactt ttggcggggc ctctgtgtt	1320
ctttttacat tcccgaaat gatgtggacc ctgcttagcg cggccctttt cacctgtctc	1380
gacatgggtc ccgcgggac cgggctcgcc gccttattcg tttctctttt cggcggttc	1440
tactcaccgg gtgaagggtcc tgccttc acctactcgcc cggaaagtctt cccctctct	1500
cacagagaag taggcatggg ctgcggcgtc gccacactgccc tttctgggc atctgtttt	1560
gggttattactt tcccttctt gcttgcactt ctgcggcaccg tggcgcctt tggctgtac	1620
gggggcttca accttagtggc gtttattggc atcttcttgg tctgtgcggg gacgaagcag	1680
aagacgcgtcg aggagtttgg aatgtgtttt gctgtgaaga cggcaagttt catgtcgat	1740
cagtgcacca aggcgcgtgcc gtgggtcata aagagggtgg tggctggca gaggaatgca	1800
aagctggagc cactgtatga gtttgcgttcc atcaaggagg cggagaagga gaggagagca	1860
gaggaggaga gaaggggaaa ggagacggga acgtcaccc ctactgtac agggagctgag	1920
ttggatgaga agaaggact gactgtatgtt aatgtctctt attcttag	1968

<210> SEQ ID NO 51

<211> LENGTH: 1725

<212> TYPE: DNA

<213> ORGANISM: *Saccharomyces cerevisiae*

<400> SEQUENCE: 51

atggcagttt aggagaacaa tatgcctgtt gtttccacagg aaccccaagc tggtaagac	60
gtgatctctt cactcagtaa agattcccat ttaagcgcac aatctcaaaa gtattctaat	120

-continued

gatgaaatgttga aagccgggtga gtcagggtct gaaggctccc aaagtgttcc tatagagata 180
cccaagaagc ccatgtctga atatgttacc gttcccttgc tttgtttgtg tggtgccttc 240
ggcgccgttca tgtttggctg ggataccggt actatttctg ggtttggtgtt ccaaacagac 300
tttttgagaa ggtttggtat gaaacataag gatggtaccc actatttgc aaacgtcaga 360
acaggtttaa tcgtcgccat tttcaatatt ggctgtgc ttgggtgtat tatactttcc 420
aaaggtggag atatgtatgg ccgtaaaaag ggtcttgcgat ttgtcgctc ggtttatata 480
gttggattat tcattcaaat tgccctctatc aacaagtggt accaatattt cattggtaga 540
atcatatctg gtttgggtgt cggcggcatac gccgtcttat gtcctatgtt gatctctgaa 600
attgctccaa agcacttgag aggcacacta gtttctgtt atcagctgtat gattactgca 660
ggtatctttt tgggctactg tactaattac ggtacaaaaga gctattcgaa ctcagttcaa 720
tggagagttc cattagggtct atgtttcgt tggtcattat ttatgattgg cgcttgacg 780
ttagttctg aatccccacg ttatttatgt gaggtgaata aggtagaaga cgccaagcgt 840
tccattgtca agtctaacaa ggtgtcacca gaggatctg cggtccaggc agagtttagat 900
ctgatcatgg ccggataga agctgaaaaa ctggctggca atgcgtctg gggggatta 960
ttttccacca agaccaaagt attcacacgt ttgttgcatttgg tggtgtttgt tcaaatagttc 1020
caacaattaa ccggtaacaa ttatTTTTC tactacggta ccgttatttt caagtcaattt 1080
ggcctggatg attcccttga aacatccatt gtcatggtg tagtcaactt tgcctccact 1140
ttcttttagtt tggtggactgt cgaaaacttg ggacatcgta aatgtttact ttggggcgct 1200
gccactatgtat tggcttgcatttggatccatcgta aatgtttact ttggggcgct 1260
cacggtaaaaa gccageccatc ttctaaaggt gceggtaact gtatgattgt cttaacctgt 1320
ttttatattt tctgttatgc cacaacctgg ggcggcgttgc cttgggtcat cacagcagaa 1380
tcattccac tgagagtcaaa gtcgaaatgt atggcggttg cctctgttc caattggta 1440
tgggggttct tgattgcatt ttccacccca ttcatcacat ctgcccattaa ctctactac 1500
ggttatgtct tcatgggtctg tttgggtgcctt atgtttttt atgtttttt cttgttcca 1560
gaaaactaaag gocatcgatc agaagaattt caagaattat ggaaagaagg tggtttacct 1620
tggaaatctg aaggctggat tccttcatcc agaagaggta ataattacga ttttagaggat 1680
ttacaacatq acqacaaccc qttgtacaag qccatqctag aataaa 1725

<210> SEQ ID NO 52
<211> LENGTH: 1908
<212> TYPE: DNA
<213> ORGANISM: *Pichia stipitis*

<400> SEQUENCE: 52

atgagtgctg acgaaaaagt cgctgctgcc ggccaggacg gcttgttga acacaacagt	60
tccacttcga gcacatcgagga caagaagccc tccaaagagct ccgatgtcga ttccgtgaac	120
tcgcaattag tagacaactc ggttagagggc aacatcttgt cccagttacac cgaaagtcag	180
gtgtatgcaga tgggttagaaag ctatgccacc aagcacggct tggaccctaga attgttcgcc	240
aaggcagctg ctgttgcacag aactccttgg ttcaact ccatgccctt cttgacagag	300
gaagagaagg ttgggttggaa tgccgaagcc actaataagt ggcacattcc acccagattg	360
atcgggggtta ttgccttggg ttctatggcc gctgctgtgc agggtatggaa cgaatcggtc	420
attaacggtg ccaacttggt ctaccccaag gctttcgag tcgacaccat gcacaattcg	480
gacttgattt aagggttggat caatgggtct ccttacccctt gctgtggat tctttcttgt	540

-continued

tggttgtctg acgcttgtaa ccgtcgctt ggtagaaaat ggaccattt ctggtgtgt	600
gtcatttctg ccatcacctg tgtctggca ggtctgtca acaactggta ccattgttc	660
attgctcggt tcttccttgg atttgggtt ggtatcaagt cgcactgt tcctgcctac	720
tctgccgaat gtactcctaa acacatcaga ggtcgttag tcatgttg gcaattctc	780
acagctgttg gtattatgtt tggttatgtt gttcccttgg ctttctacaa tgctggagat	840
agaggaatcc attacgggtt gaactggaga ttgatgttg gttcggccgc tattctgct	900
gtcatcatct tggccaaat tccttcgtt cctgaatctc cacgtgggtt aatggtaag	960
gacagacacc ttgaaggcctt tgagtccctt aagcaattga gatacgaaga acttgctgct	1020
gctcgtgact gtttctacca gtacgtctt ttagctgaag aaggttctta caagatccc	1080
acccttcacca gatttaagga aatgttcacc aagagaagaa acagaaacgg tgccatcggt	1140
gcatttattt tcatgttcat gcaacagtcc tgggttatca acgtcattgc ttactactct	1200
tctgtctatct ttgtccaatc tggtttctct caaacttctg ctttgcgc ttcttgggt	1260
ttcggatgc ttaacttcac ctttgcatt cctgccttct tcacaatcga tcgttcggt	1320
agaagatcct tattgttggg tacttcccc ttgatggcta ttttcttatt gattggcggt	1380
ttcggtttct tgataaacga agaaacaaac tccaaggaa gattggaaat gatcatcatc	1440
ggtatctata tggccaccat ctgttactct tccggtaag gtccagttcc tttcacctac	1500
tctggccaaat cttccattt gtacatcaga gactgggtt tggatgttgc tactgccacc	1560
tgttggactt tcaacttcat cttggccttc acctggaaaca gattggtaa tgcattcaca	1620
tctactggtg ctttcggctt ctacgtgtt tggaaacatca ttggatgtt cttggatctt	1680
tggatgttgc cagaaaccaa gggcttgacc ttgaaagaat tggacgaat ctgcgggtt	1740
tccggcgcc aacacgcca a taccaaacc aagagtttga tcaacttcat ccaaagatac	1800
gttttacgtt ccaagggtgc tccattgtct ccattgtacg accaccagag attggctgtc	1860
accaacccag aatggAACGA caagccagaa gtctttatg tggatgtt	1908

<210> SEQ ID NO 53

<211> LENGTH: 40

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Synthesized Construct

<400> SEQUENCE: 53

atatatgagc tcgtgagtaa ggaaagagtg aggaactatc 40

<210> SEQ ID NO 54

<211> LENGTH: 50

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Synthesized Construct

<400> SEQUENCE: 54

atataatacta gttgttttat atttgggtta aaaagtagat aatttacttcc 50

<210> SEQ ID NO 55

<211> LENGTH: 30

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Synthesized Construct

<400> SEQUENCE: 55

-continued

atggatccaa aaatgtcgtc tcacggctcc 30

<210> SEQ ID NO 56
<211> LENGTH: 60
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthesized Construct
<400> SEQUENCE: 56

atgaattcct acaaatcttc ttcagaaatc aatttttgtt cagcaacgat agtttcggac 60

<210> SEQ ID NO 57
<211> LENGTH: 34
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthesized Construct
<400> SEQUENCE: 57

atactatgtaa aaatgggcat cttcaacaag aagc 34

<210> SEQ ID NO 58
<211> LENGTH: 64
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthesized Construct
<400> SEQUENCE: 58

gcatatcgat ctacaaatct tcttcagaaa tcaatttttg tttagcaaca gacttgccct 60

catg 64

<210> SEQ ID NO 59
<211> LENGTH: 42
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthesized Construct
<400> SEQUENCE: 59

tattaaatcg atggtagtgg tagtgtgagc aaggcgagg ag 42

<210> SEQ ID NO 60
<211> LENGTH: 36
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthesized Construct
<400> SEQUENCE: 60

tattaaatcg acctacttgt acagtcgtc catgcc 36

<210> SEQ ID NO 61
<211> LENGTH: 28
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthesized Construct
<400> SEQUENCE: 61

gcatggatcc atgtcgctc acggctcc 28

<210> SEQ ID NO 62
<211> LENGTH: 31

-continued

<212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: Synthesized Construct

<400> SEQUENCE: 62

tataatgaat tcagcaacga tagttcgga c

31

<210> SEQ ID NO 63
 <211> LENGTH: 34
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: Synthesized Construct

<400> SEQUENCE: 63

tattaaacta gtatgggcat cttcaacaag aagc

34

<210> SEQ ID NO 64
 <211> LENGTH: 33
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: Synthesized Construct

<400> SEQUENCE: 64

ttataagaat tcagcaacag acttgccctc atg

33

<210> SEQ ID NO 65
 <211> LENGTH: 39
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: Synthesized Construct

<400> SEQUENCE: 65

gcatactagt aaaaatgtct cttcctaagg atttcctct

39

<210> SEQ ID NO 66
 <211> LENGTH: 55
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: Synthesized Construct

<400> SEQUENCE: 66

atactgcagt taatgatgat gatgatgatg gtccttcttg atcaaagagt caaag

55

<210> SEQ ID NO 67
 <211> LENGTH: 6
 <212> TYPE: PRT
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: Synthesized Construct

<400> SEQUENCE: 67

Pro Glu Ser Pro Arg Phe
1 5

<210> SEQ ID NO 68
 <211> LENGTH: 4
 <212> TYPE: PRT
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: Synthesized Construct

<400> SEQUENCE: 68

-continued

Pro Glu Ser Pro
1

<210> SEQ ID NO 69
<211> LENGTH: 4
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthesized Construct

<400> SEQUENCE: 69

Pro Glu Ser Pro
1

<210> SEQ ID NO 70
<211> LENGTH: 10
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: GS-linker

<400> SEQUENCE: 70

Gly Gly Gly Gly Ser Gly Gly Gly Ser
1 5 10

<210> SEQ ID NO 71
<211> LENGTH: 29
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthesized Construct

<400> SEQUENCE: 71

gacatcgatg acatatgcgc taccgcaac

29

<210> SEQ ID NO 72
<211> LENGTH: 24
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthesized Construct

<400> SEQUENCE: 72

gtgcacgtcg gacccgcaga ttcc

24

<210> SEQ ID NO 73
<211> LENGTH: 24
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthesized Construct

<400> SEQUENCE: 73

ggaatctgctg ggtccgacgt gcac

24

<210> SEQ ID NO 74
<211> LENGTH: 28
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthesized Construct

<400> SEQUENCE: 74

cagaagattt aaggatcctg aacgtaga

28

<210> SEQ ID NO 75
<211> LENGTH: 26

-continued

<212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: Synthesized Construct

<400> SEQUENCE: 75

gacatcagtg acatatgtcg cttcc

26

<210> SEQ ID NO 76
 <211> LENGTH: 26
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: Synthesized Construct

<400> SEQUENCE: 76

ctggattga ggatcctgaa cgata

26

<210> SEQ ID NO 77
 <211> LENGTH: 28
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: Synthesized Construct

<400> SEQUENCE: 77

gacatcgatg acatatggct tccgcaac

28

<210> SEQ ID NO 78
 <211> LENGTH: 29
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: Synthesized Construct

<400> SEQUENCE: 78

ccagaagtat tgagaattct gaacgtaga

29

<210> SEQ ID NO 79
 <211> LENGTH: 27
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: Synthesized Construct

<400> SEQUENCE: 79

gacatcgatg acatatggcg actctgc

27

<210> SEQ ID NO 80
 <211> LENGTH: 29
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: Synthesized Construct

<400> SEQUENCE: 80

ggatacagaa tgaggatcct gaacgtaga

29

<210> SEQ ID NO 81
 <211> LENGTH: 20
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: Synthesized Construct

<400> SEQUENCE: 81

taatacgtact cactatagg

20

-continued

<210> SEQ ID NO 82
<211> LENGTH: 19
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthesized Construct

<400> SEQUENCE: 82

gctagttatt gctcagcgg

19

<210> SEQ ID NO 83
<211> LENGTH: 36
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthesized Construct

<400> SEQUENCE: 83

cctatcgta ttacctcacg tgacgagggg cggctg

36

<210> SEQ ID NO 84
<211> LENGTH: 36
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthesized Construct

<400> SEQUENCE: 84

cagccgcccc tcgtcacgtg aggtaatgac gatagg

36

<210> SEQ ID NO 85
<211> LENGTH: 29
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthesized Construct

<400> SEQUENCE: 85

ctttcgaaac ggctacaaaac cccaagacg

29

<210> SEQ ID NO 86
<211> LENGTH: 34
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthesized Construct

<400> SEQUENCE: 86

gtttgtcatc acatcacgtt cagagagccg tctg

34

<210> SEQ ID NO 87
<211> LENGTH: 34
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthesized Construct

<400> SEQUENCE: 87

cagacggctc tctgaacgtg atgtgatgac aagc

34

<210> SEQ ID NO 88
<211> LENGTH: 30
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthesized Construct

-continued

<400> SEQUENCE: 88

gcatttgaga cgtcaacaga tcccaagaga

30

<210> SEQ ID NO 89
<211> LENGTH: 36
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthesized Construct

<400> SEQUENCE: 89

cctattgtca tcacttcacg tgacgaggcg cgcttg

36

<210> SEQ ID NO 90
<211> LENGTH: 36
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthesized Construct

<400> SEQUENCE: 90

caagcggccc tcgtcacgtg aagtgtatgc aatagg

36

<210> SEQ ID NO 91
<211> LENGTH: 33
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthesized Construct

<400> SEQUENCE: 91

cctttgagac tgccacaaac cctaagaccg gtg

33

<210> SEQ ID NO 92
<211> LENGTH: 30
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthesized Construct

<400> SEQUENCE: 92

atggatcaa aatgtcgta tcacggctcc

30

<210> SEQ ID NO 93
<211> LENGTH: 60
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthesized Construct

<400> SEQUENCE: 93

atgaattcct acaaatcttc ttcatggaaatc aatttttgtt cagcaacgat agcttcggac

60

<210> SEQ ID NO 94
<211> LENGTH: 34
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthesized Construct

<400> SEQUENCE: 94

atactagtaa aatgggcat cttcaacaag aagc

34

<210> SEQ ID NO 95
<211> LENGTH: 64

-continued

<212> TYPE: DNA		
<213> ORGANISM: Artificial Sequence		
<220> FEATURE:		
<223> OTHER INFORMATION: Synthesized Construct		
<400> SEQUENCE: 95		
gcataatcgat ctacaaatct tcttcagaaa tcaatttttg ttccagcaaca gacttgcct	60	
catg	64	
<210> SEQ ID NO 96		
<211> LENGTH: 39		
<212> TYPE: DNA		
<213> ORGANISM: Artificial Sequence		
<220> FEATURE:		
<223> OTHER INFORMATION: Synthesized Construct		
<400> SEQUENCE: 96		
gcatactagt aaaaatgtct cttcctaagg atttcctct	39	
<210> SEQ ID NO 97		
<211> LENGTH: 55		
<212> TYPE: DNA		
<213> ORGANISM: Artificial Sequence		
<220> FEATURE:		
<223> OTHER INFORMATION: Synthesized Construct		
<400> SEQUENCE: 97		
atactgcagt taatgatgat gatgatgatg gtccttcttg atcaaagagt caaag	55	

The invention claimed is:

1. A method of increasing transport of cellobextrin into a cell, comprising:

culturing a host cell which comprises a recombinant polynucleotide encoding a cellobextrin transported polypeptide in a medium such that the recombinant polynucleotide is expressed, said cellobextrin transporter having a structure of a Major Facilitator Superfamily protein and comprising transmembrane α -helix 1, α -helix 2, α -helix 3, α -helix 4, α -helix 5, α -helix 6, α -helix 7, α -helix 8, α -helix 9, α -helix 10, α -helix 11, α -helix 12, said transmembrane α -helix 1 characterized by:

a leucine, isoleucine, valine or methionine at the position corresponding to amino acid 1 of SEQ ID NO:1;

a tyrosine at the position corresponding to amino acid 2 of SEQ ID NO:1;

a phenylalanine or leucine at the position corresponding to amino acid 3 of SEQ ID NO:1;

a tyrosine or phenylalanine at the position corresponding to amino acid 17 of SEQ ID NO:1; and

an aspartate at the position corresponding to amino acid 18 of SEQ ID NO:1,

wherein expression of the recombinant polynucleotide results in increased transport of cellobextrin into the cell compared with a cell that does not comprise the recombinant polynucleotide.

2. The method of claim 1 wherein the polypeptide comprises an amino acid sequence having at least 29%, at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 99%, or at least 100% amino acid identity to NCU00801 or NCU8114.

3. The method of claim 1 wherein the host cell further comprises a second recombinant polynucleotide encoding at least a catalytic domain of a β -glucosidase.

4. The method of claim 3 wherein the β -glucosidase is from *Neurospora crassa*.

5. The method of claim 4 wherein the β -glucosidase is encoded by NCU00130.

6. The method of claim 1 wherein the host cell further comprises one or more recombinant polynucleotides wherein the one or more polynucleotides encode one or more enzymes selected from one or more of the group consisting of L-arabinose isomerase, L-ribulokinase, L-ribulose-5-P 4 epimerase, xylose isomerase, xylulokinase, aldose reductase, L-arabinitol 4-dehydrogenase, L-xylulose reductase, and xylitol dehydrogenase.

7. The method of claim 1, wherein the host cell further comprises a second recombinant polynucleotide wherein the second recombinant polynucleotide encodes a pentose transporter.

8. The method of claim 7, wherein the pentose transporter is selected from the group consisting of NCU00821, NCU04963, NCU06138, STL12/XUT6, SUT2, SUT3, XUT1, and XUT3.

9. The method of claim 1 wherein the medium comprises a cellulase-containing enzyme mixture from an altered organism, wherein the cellulase-containing mixture has reduced β -glucosidase activity compared to a cellulase-containing mixture from an unaltered organism.

10. The method of claim 1, wherein the host cell is selected from the group consisting of *Saccharomyces* sp., *Saccharomyces cerevisiae*, *Saccharomyces monacensis*, *Saccharomyces hayanus*, *Saccharomyces pastorianus*, *Saccharomyces carlsbergensis*, *Saccharomyces pombe*, *Kluyveromyces* sp., *Kluyveromyces marxianus*, *Kluyveromyces lactis*, *Kluyveromyces fragilis*, *Pichia stipitis*, *Sporotrichum thermophile*, *Candida shehatae*, *Candida tropicalis*, *Neurospora crassa*, *Zymomonas mobiles*, *Clostridium* sp., *Clostridium phytofermentans*, *Clostridium thermocellum*, *Clostridium beijerinckii*.

235

inckii, *Clostridium acetobutylicum*, *Moorella thermoacetica*, *Escherichia Klebsiella oxytoca*, *Thermocenaerobacterium saccharolyticum*, and *Bacillus subtilis*.

11. The method of claim 1 wherein the celldextrin is selected from one or more of the group consisting of cello- 5 biose, cellotriose, and cellotetraose.

12. The method of claim 2, wherein the polypeptide comprises an amino acid sequence having at least 85% amino acid identity to NCU00801.

13. The method of claim 2, wherein the polypeptide comprises an amino acid sequence having at least 85% amino acid identity to NCU08114. 10

14. The method of claim 2, wherein the polypeptide comprises an amino acid sequence having at least 90% amino acid identity to NCU00801. 15

15. The method of claim 2, wherein the polypeptide comprises an amino acid sequence having at least 90% amino acid identity to NCU08114.

16. The method of claim 2, wherein the polypeptide comprises an amino acid sequence having at least 95% amino acid 20 identity to NCU00801.

17. The method of claim 2, wherein the polypeptide comprises an amino acid sequence having at least 95% amino acid identity to NCU08114.

* * * * *

25

236

UNITED STATES PATENT AND TRADEMARK OFFICE
CERTIFICATE OF CORRECTION

PATENT NO. : 8,431,360 B2
APPLICATION NO. : 12/843844
DATED : April 30, 2013
INVENTOR(S) : N. Louise Glass et al.

Page 1 of 1

It is certified that error appears in the above-identified patent and that said Letters Patent is hereby corrected as shown below:

In the Claims:

In column 233, line 36, in claim 1, delete “transported” and insert -- transporter --, therefor.

In column 233, line 38, in claim 1, delete “having a” and insert -- having the --, therefor.

In column 233, line 46, in claim 1, delete “corresponding,” and insert -- corresponding --, therefor.

In column 233, line 64, in claim 2, delete “NCU8114” and insert -- NCU08114 --, therefor.

In column 234, line 61, in claim 10, delete “*hayanus*” and insert -- *bayanus* --, therefor.

In column 234, line 66, in claim 10, delete “*mobiles*” and insert -- *mobilis* --, therefor.

In column 235, line 2, in claim 10, delete “*Escherichia*” and insert -- *Escherichia coli*, --, therefor.

In column 235, line 2, in claim 10, delete “*Thermocenaerobacterium*” and
insert -- *Thermoanaerobacterium* --, therefor.

Signed and Sealed this
Eighth Day of October, 2013



Teresa Stanek Rea
Deputy Director of the United States Patent and Trademark Office